Advanced Analytics Team Project

Spring 2019

[redacted], Andy Petersen, [redacted]

Project Description and Objective

A member of the project team is employed by a local insurance company. As with most insurance companies, this employer allocates a significant portion of their investment portfolio to corporate bonds. In the 1990's most corporate bonds were not callable, however over the last 20 or so years most have evolved into a callable structure with something called a makewhole fee to compensate the investor for the call risk.

A makewhole fee is equivalent to the sum of discounted future expected coupon payments at current market rates. Time value of money tells us that a bondholder should be indifferent between holding the bond to maturity or it being "made whole." However, due to accounting rules, the income from these two outcomes flows through the income statement in different ways. The income on a bond held to its maturity flows through evenly over its life in *Investment Income*. The income related to a bond that is made-whole flows through *Other Income* in a lumpy fashion prior to it being redeployed at current market rates.

Management does not like the lumpiness associated with the Makewhole Calls and has asked that the group provide an annual estimate for planning purposes. Historically, it has been widely held that this number is unpredictable because of the unobservable factors that drive the call feature. Observable bond characteristics that can be captured by a model include coupon rate, maturity, rating quality and whether or not the bond has a makewhole feature. Unobservable factors are events, individual company decision making, and treasury rate movement in the coming year.

This project had three main focuses related to this subject. The first objective was to see what measurable factors may be good predictors of call risk. Secondly, the group would try to build a predictive model to predict and rank the likelihood of calls on individual bonds based on the results from a previous year in a provided selection of roughly 900 bonds. Finally, this project aimed to see if this type of predictive modeling, using only the observable factors, is able to represent all of the primary drivers of bond calls.

Data

The data used to complete this project is comprised of nearly 900 individual corporate bond holdings and their individual data. The data set includes both publicly traded and non-publicly traded corporate bonds. The data used is publicly available via statutory filings and includes features of each security.

Some of the input variables available for each bond include the coupon rate, the number of months until maturity, NAIC Rating, the par and market values of the bond, and the ratings from each of three independent rating agencies. The coupon rate represents the annual cost of maintaining the existing debt. Months to Maturity tells us the time period that the expected future cash flows will be discounted over, which is a primary factor in the amount required to be paid as part of the makewhole fee. The NAIC Rating measures the quality of the issuer and their perceived ability to access capital markets. The ratings for each bond from each of the three rating agencies, Fitch, S&P, and Moody's, represent analyst

judgements about the credit risk of the issuer. These ratings were highly correlated to each other and to the NAIC rating. The NAIC rating is determined by the National Association of Insurance Commissioners and is based on the second lowest public rating. The NAIC rating is used in capital charge calculations. To eliminate correlation issues in the data, we used the NAIC rating to measure quality and access to markets.

We attempted to obtain additional data that could possibly be used to explain some of the other variables that are more difficult to measure. We pulled market capitalization information in an attempt to determine if company size mattered and was possibly a driver of sophistication and decision making. Due to structures and the non-public nature of many of the holding this information was overly burdensome to accumulate. We know that treasury rates are a major driver of call activity and none of our input variables capture this. We attempted to capture this through pulling a yield to maturity number for all bond holdings. Taking the coupon divided by the yield to maturity, we felt that we could get a good indication of the cost of the outstanding debt compared to the current market value of the debt. The problem with this technique is many of the holdings in the portfolio are illiquid and we were unable to get a quote for yield to maturity as of 1/1/2018 for many of the holdings. When we ran this through, it actually resulted in a poorer model because many of our yes outcomes were thrown out for incomplete data.

Analytic Techniques

To model the potential call risk for each bond, that is the likelihood that the bond would be called in the next 12 months, a logistic regression was created using JMP. To do this, a few modifications had to be made to the raw data.

To allow the model to include the ordinal nature of the bond ratings from each of the three rating agencies, these variable columns were recoded into numeric format. This was done ahead of time in r for simplicity, and involved ranking the different possible bond ratings from each agency and aligning them with the actual ratings given to each bond.

In addition, the Months to Maturity had to be calculated. The data included the date of maturity, but did not separately include the number of months until that date. This involved a simple formula that subtracted the current date from maturity date of each bond.

Next, the data was partitioned into training and validation data. This was done randomly to avoid inducing bias into the model. After this, a stepwise logistic regression was run on the training dataset using all of the available predictor variables. The least significant variable was removed and the model was rerun. This process was repeated until only significant variables remained. In the final model, the only variables which proved significant were makewhole, coupon rate, NAIC rating, and months to maturity.

Effect Summary

Source	LogWorth	PValue
MTM	8.499	0.00000
MAKEWHOLE	3.567	0.00027
Cpn	2.526	0.00298
NAIC	1.526	0.02979

Results and Interpretation

The results show that the model is not terribly accurate. The model tends to just classify nearly all outcomes as no. However, the bonds with the highest call probabilities do capture most of the yes outcomes, which is a positive outcome. Using the probability of a yes outcome from the model multiplied but the par value of the bonds, we arrived at an overall par callable amount of $80,372,325. This number was then inserted into a separate model that is used to calculate the actual makewhole fee, arriving at an estimated annual other income number of $3,277,923 attributable to the corporate bonds portion of the investment portfolio. This compares to actual Other Income attributable to corporate bonds of $7,380,877.

Fit Details

Measure	Training	Definition
Entropy RSquare	0.3659	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.3915	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0694	$\sum$ -Log(ρ[j])/n
RMSE	0.1354	$\sqrt{\sum(y[j]-\rho[j])^2/n}$
Mean Abs Dev	0.0368	$\sum \lvert y[j]-\rho[j]\rvert/n$
Misclassification Rate	0.0219	$\sum$ (ρ[j]≠ρMax)/n
N	913	n

Conclusion

Overall, it was concluded that the call risk of bonds can be modeled to some degree, however the model is not particularly good at making this prediction. The information the model is able to capture only represents approximately 37% of what is driving the called bonds, represented by the R^2 of .37. This supports the hypothesis that some of the inputs can be observed and quantified, but not all of them. There are clearly other primary drivers that play-in to the probability of a bond being called. The observable bond features that were found to be significant in predicting the likelihood of a call were Months to Maturity, makewhole, NAIC rating and Coupon rate, however these were of limited value because of the unobservable factors mentioned in the project description. We were able to put a call probability on each bond holding. We took this yes probability and multiplied it by the par holding to determine how much principal could be called. Historically management has just used estimates based on historical data. The par input number was more conservative than the previous estimate as well as the potential makewhole fee number calculated. Both numbers underestimated what was actually called in 2018 however management is eager to put better procedures around inputs and we believe they will find the model useful.