Scotiabank Risk Analytics Student Engagement Data Science Competition Report

Daniel Zhang, Timmy Siu, Alice Hong, Candy Chow

Summary

Five models were built using the historical world, US, and Canadian macroeconomic data to predict interest rates and credit spread data under simulated stress scenarios. Visual representations of the predictions were interpreted according to historical patterns under stress. The model predictions generally show a possible recession in 2018 followed by slight economic growth and signs of economic downturn after 2020.

Introduction

The goal of this project is to use visual representation and statistical methods to model and predict credit spread data (CDS) and interest rates. A total of five models were produced to predict the CAD and USD interest rates as well as the CDS of AAA, BBB-, and CCC quality counterparties with respect to US, Canadian, and world macroeconomic data. Models were fitted using methods such as Multiple Linear Regression, Elastic Net, and LASSO regularization. They were then run on simulated market stress data to obtain predictions for interest rate and CDS.

Methodology

Data Processing: Interest (Yield) Rate

To convert daily data to quarterly without loss of information, we took the average of the daily data for each individual quarter. We found this to be a good way to condense the data after comparing plots of daily data to our averaged quarterly data (see appendix). In order to represent the yield curve, we chose the rate difference between 20-year and 1-year term (yr20-yr1) contracts as our response variable. We ended up with two models to represent the Canadian and US markets.

Canada Interest (Yield) Rate

A linear regression is performed on the historical data from 2004 to 2017 based on both dependent and independent variables. We compared the adjusted r-squared and plotted a graph with fitted and actual values to measure the fit of the analysis. Afterward, the model was applied to the stress scenario data from Scotiabank to compute a predicted interest (yield) rate difference between 20-year and 1-year maturity from 2018 to 2022. Then, we plotted the predicted values to see the trend and interpret potential recessions in the near future. The following variables were used:

- World Oil price (woil)
- World gold price (wgold)
- World gas price (wgas)
- World Price Index (wfood)
- The rate of Canadian treasury bill (ctbill)
- The rate of Canadian government bond of 5-year maturity (cgbond05)
- The rate of Canadian government bond of 10-year maturity (cgbond10)
- Canadian Relative GDP Growth (crgdp)
- Canadian Consumer Price Index (ccpi)
- Canadian Housing Starts (chst)
- Canadian Unemployment Rate (curate)

Variable selection was based on the economic relevance of the global and Canadian data, and by observing the historical trend to relate the Canada interest rate of different year maturity.

USA Interest (Yield) Rate

For the US market, we performed variable selection on the world and US data sets. Visually, many macroeconomic variables showed significant change when under stress but including all of them would produce an overfitting model. Additionally, many variables were highly correlated with each other. In response to this, we decided to perform variable selection and model using Elastic Net regularization.

In R, we performed cross-validation using the package glmnet() to find the optimal tuning parameters alpha and lambda. For lambda, we used the 1-SE rule and chose the largest value within 1-SE of the value that gives us the lowest error. Then, we ran the model on the stress scenario variables and produced a set of predictions for the rate difference. The following are the predictor variables chosen using Elastic Net:

- World Oil Price (woil)
- World Food Price (wffod)
- US Interest Rate 3-Month t-bill (utbill)
- US Real GDP Annual Growth (urgdp)
- US Inflation CPI (ucpi)
- US Government Interest Payments (uggip)
- US House Price Index (uhpi)
- US Commercial Real Estate Price Index (ucrepi)
- US Unemployment Rate (uurate)

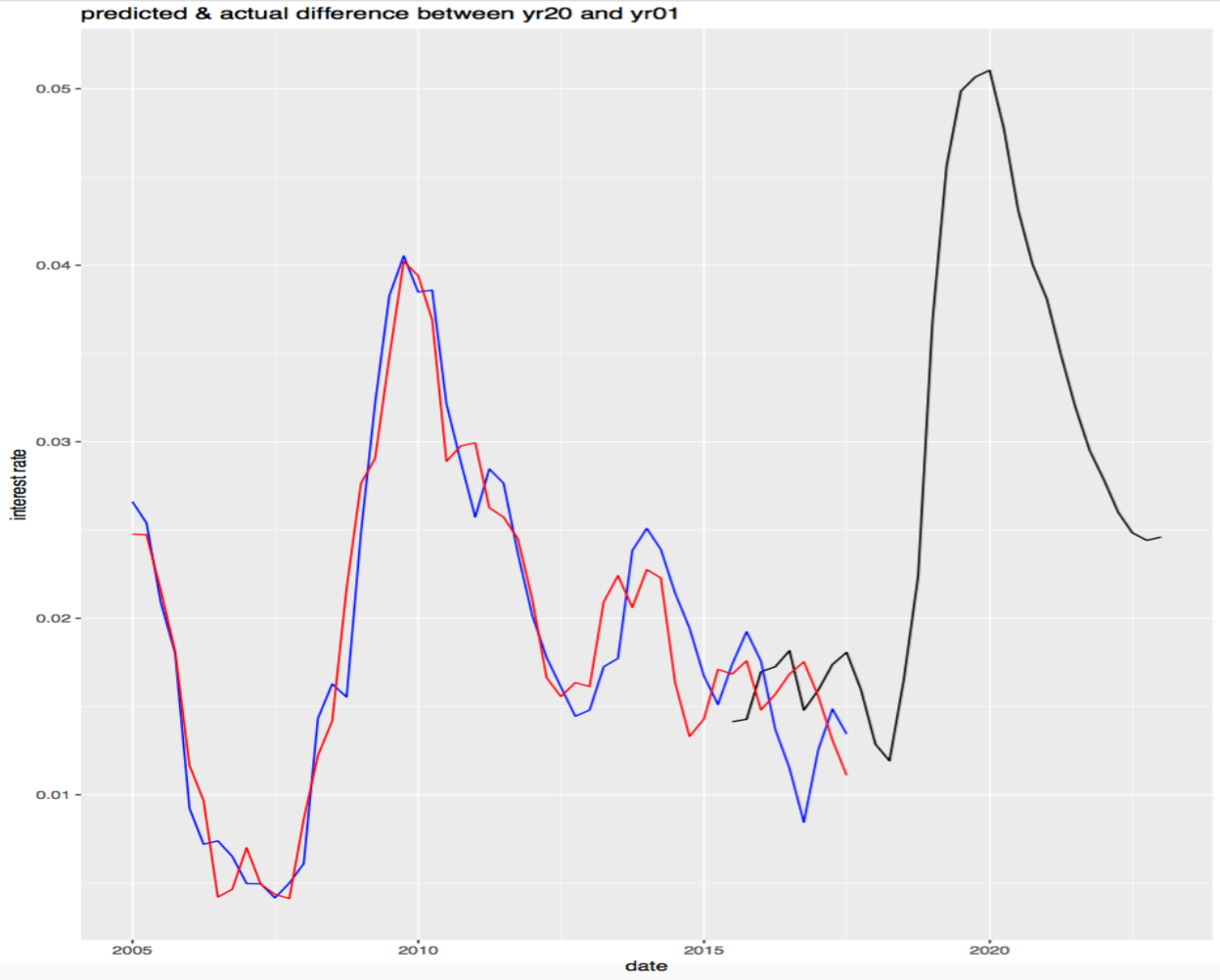
Credit Spread Data

The relative risk of bonds was determined to be response variable; the response was determined within each credibility level - AAA, BBB, and CCC - by differencing the maximum values of the quarterly CDS data of 20 and 1-year bonds to prevent the loss of outliers over quarters. As many of the macroeconomic data showed an increasing linear trend, the covariates that showed such a trend were transformed to the percentage of change relative to the previous quarter. All possible covariates were then collected into a single matrix and used to fit a LASSO shrinkage model. As each LASSO is randomized, we determined the optimal model by repeatedly fitted the model and minimizing the MSPE cross-validation. Each model was also cross-validated to determine the optimal tuning parameter based on the given data. The resulting best LASSO models for each of the three credibility levels were then used to predict the CDS difference of the respective credibility levels with the simulated stress data.

Results/Findings

Canada Interest (Yield) Rate

Figure 1.1. Blue plot: actual values from 2004 to 2017
Red plot: fitted values from 2004 to 2017

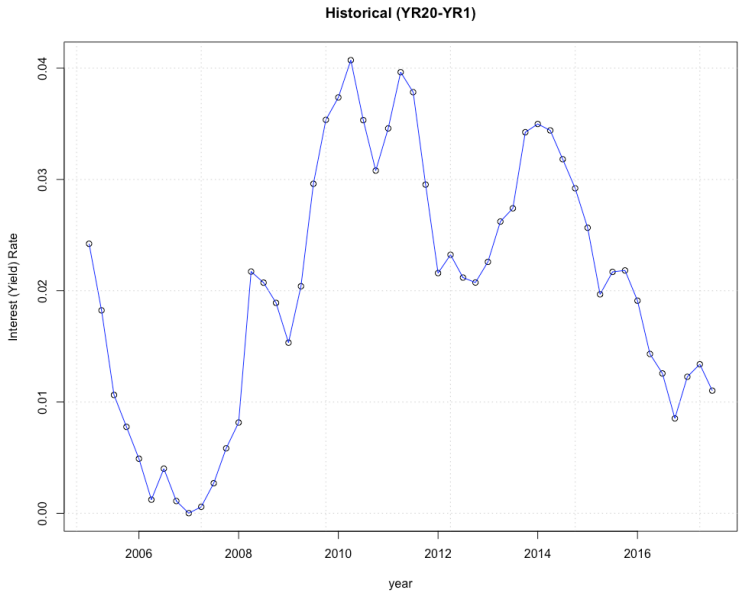


Black plot: predicted values from 2016 to 2022

From the blue plot, we can see that from the year of 2006-2008, the yield difference between long-term and short-term investment is close to 0, and there is a relative flattening yield curve. The curve begins to return back to normal in 2008. This is most likely due to the investors receiving fewer benefits holding a long-term investment than a short-term one. This may have been an indication of the 2008 financial crisis.

From the black plot showing the predicted values on the stress scenario data, we observe a slight decrease in the rate gap starting in 2017 followed by significant growth in 2018. A strong economic growth is to be expected in 2019, as demonstrated by the increasing trend of the difference between long-term and short-term investment. From now until 2020, long-term investments may be desired. Following 2020, there is a decrease again, however, due to the lack of data past 2022, we are not able to conclude with certainty the appearance of a recession in that time frame.

USA Interest (Yield) Rate



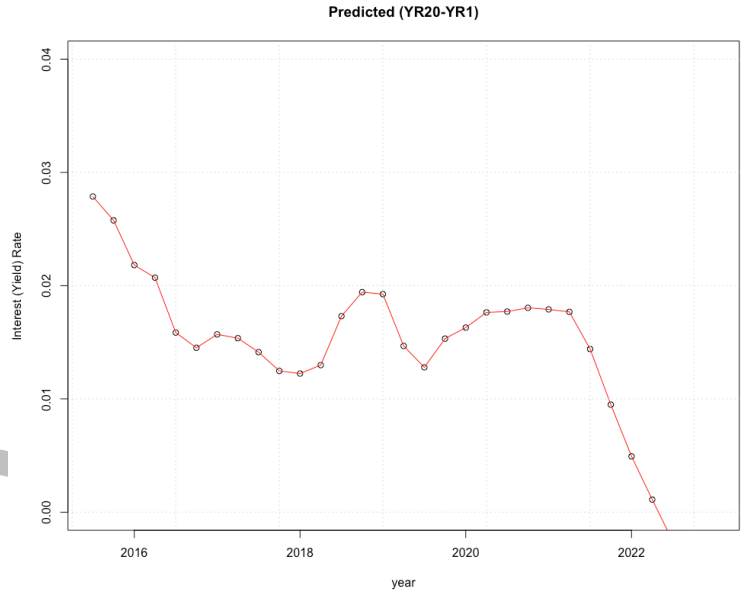


Figure 2.1. Historical rate difference data

Figure 2.2. predicted rate difference data

From Figure 2.1., the rate gap is decreasing in the years prior to the 2008 recession and then starts to recover in 2008 with a slight dip in 2009 before recovering again. It is known that a flat yield curve is an indication that investors and traders are worried about the macroeconomic outlook, and could possibly point to an upcoming recession. Future predictions using this model can be predicted by referencing this yield pattern.

The rate difference begins to decrease slowly prior to 2016 until 2018, and recovers in mid-2018. Although not as quantically large as 2008, this can be understood as a possible recession in 2018-2019. Towards the end of 2021, the rate difference starts to approach zero more quickly. Although we do not have data after 2022, just from looking at this model alone says that perhaps there will be a bigger recession in the near future after the stress scenario data.

Credit Spread Data

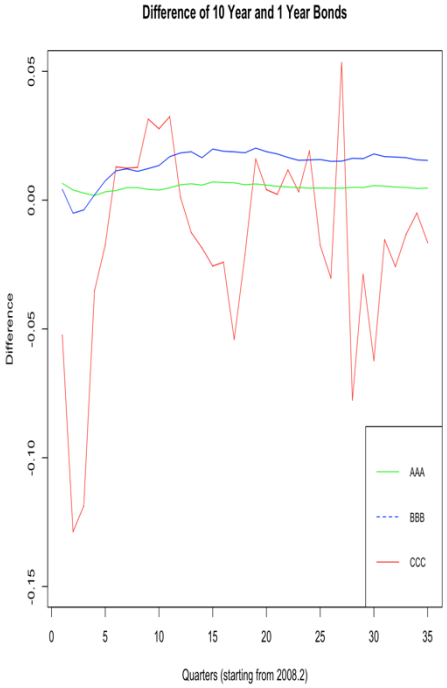


Figure 3.1 Difference of 10-year bonds and 1-year bonds

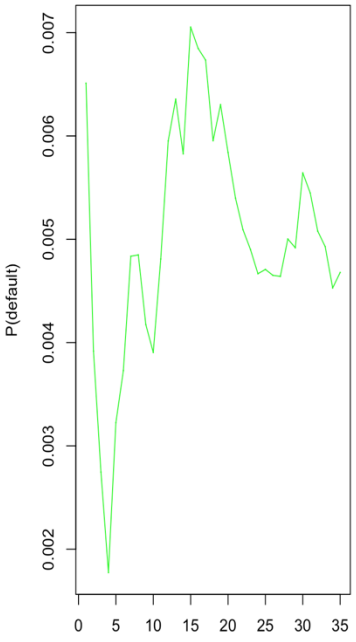


Figure 3.2 Difference of 10 Year and 1-year bonds (AAA)

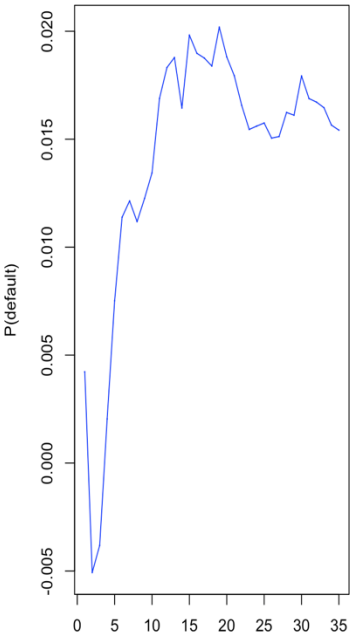


Figure 3.3 Difference of 10 year and 1-year bonds (BBB)

From historical data, it is noted that a severely negative slope of the relative risk indicates the possibility of crisis. This is observed across bonds of all credibility levels and aligns with the appearance of the recession in 2008. While all AAA, BBB, and CCC displayed the trend, the most drastically impacted bonds was that of the CCC credibility level. The CCC's difference drops below 0, indicating that the 1-year bonds hold a greater probability of default in comparison to the 20-year bonds. CCC's sensitivity to crisis times and its inversion of probability can be used as an indicator of an upcoming recession.

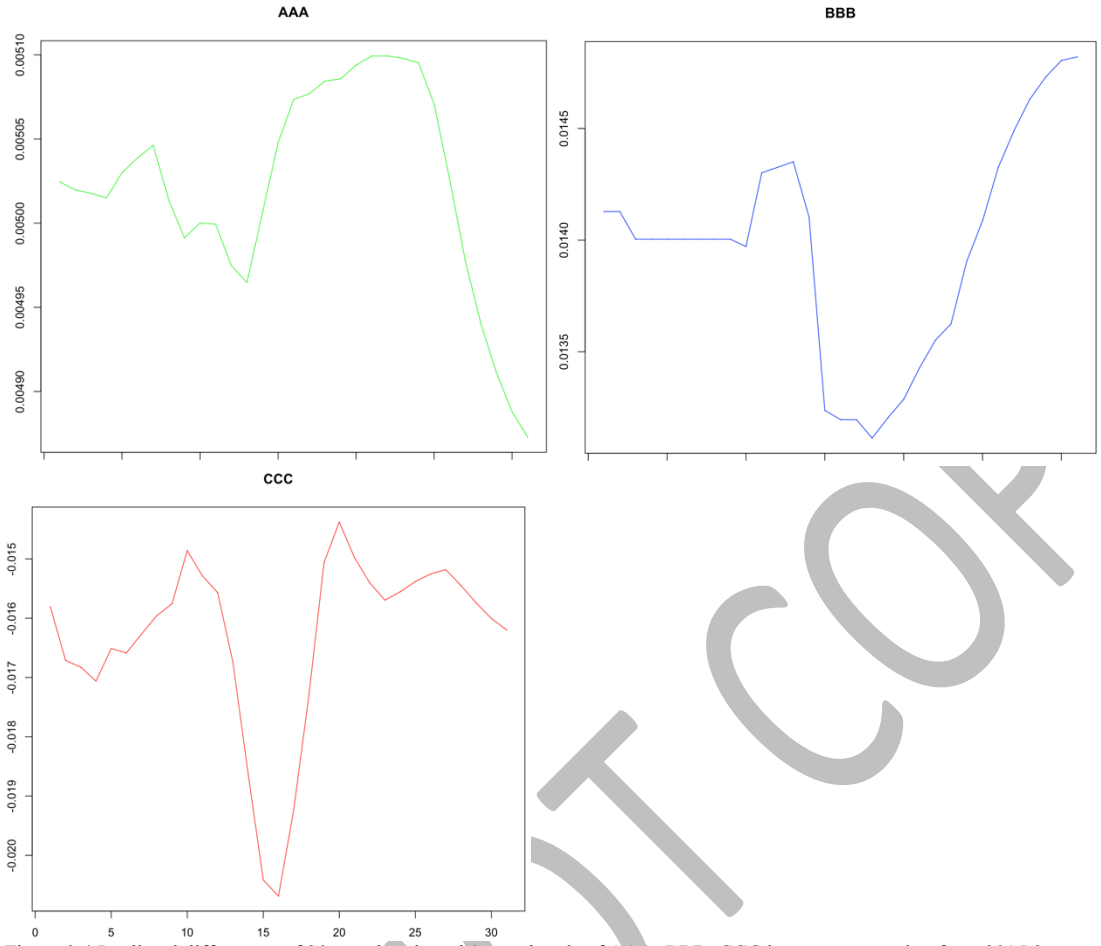


Figure 3.4 Predicted differences of 20-year bonds and 1-year bonds of AAA, BBB, CCC in quarters, starting from 2015.2

A significance of the CCC prediction is the difference is negative for the upcoming quarters. As the probability of default of 1-year bonds is higher than that of 20-year bonds signals an upcoming crisis event, the prediction for the next 31 quarters suggests the presence of a recession, given the simulated stress variables. This conclusion is further supported by the prediction of the AAA and BBB differences, as the largely negative slope during the 15th quarter (2018) indicates the approach of a crisis period. This result aligns with the predictions found by using multiple linear regression and elastic net methods on interest rates. This prediction data also suggests that long-term investments are a safer choice within the next 15 quarters.

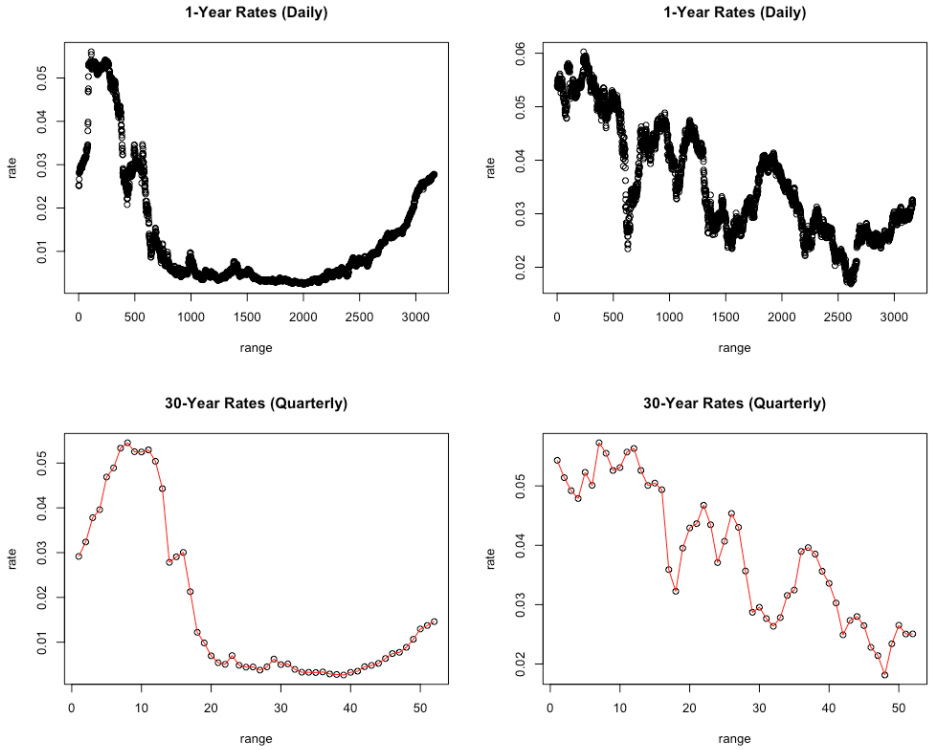
The negative slopes of the AAA and CCC data starting from the 25th quarter (2022) also indicates the approach of another recession. The discrepancy in the tail of the BBB data is unexpected, as a lowering in the difference should have been expected from the other predictions made. This could be due to the randomization of the LASSO model, leading to a smoother fit, or the appearance of outliers with greater leverage at the end of the period.

Conclusions

Through the various fittings of the macroeconomic data to both interest rates and CDS, we determined the timeline of the business cycle for the next 4 years, as well as the appearance of the next recession, given the simulated stress variables. In general, all the models fitted showed a downturn in economic activity following 2020. If additional stress data were given, continuing the current prediction trend, we would be able to conclude the possible presence of an upcoming recession after 2022.

The USA interest rate and CDS models showed similar predictive patterns under the stress data for periods 2018 to 2020 whereas the Canadian interest rate model differed. The difference may be caused by the variable selection of the shrinkage method, leading to a model of higher predictive power but lesser fit to the actual data. By contrast, the regression model was able to obtain an adjusted-R^2 of high 80%.

Appendix



Limitations

In this project, we faced several limitations, they include:

1. Insufficient data
    - Since we had to condense the data from daily to quarterly, the data set size has shrunk. The data may not be representative enough for interpretation.
2. Many zeros in the CDS dataset
    - Since there are many missing values in the CDS data from the 15-year bond to the 40-year bond, we are not able to fit the data for the long term bonds into the yield graph. Using the 10-year bond data may not reflect the best yield because the duration of the bond is closer to 1- year bond than the 15- year bond or above. As a result, the result of a 10-year bond may not be as significant as the ones with longer maturity years.
3. Losing information through converting data from daily to quarterly
    - By converting the data from daily into quarterly, the process eliminated out many variables in the dataset and some important information was pulled out during the procedure. Hence, the quarterly data may not reflect the real scenario.