Scotiabank Risk Analysis
--- Interest rate predictions over stress scenario

Shiyang Li
Yichun Liu
Yubin Lyu
Xinwei(Wendy) Kuang

I. Abstract

The goal is to use the given macroeconomic data to predict the interest rate in 4 different areas. Four regression models were constructed based on interest rates over the past years, and the prediction is based on stress scenario Macro variables from 2018-2022.

II. Introduction

The explanatory variables we are trying to use are all macroeconomic variables in datasets, including world GDP and commodity prices, Canadian, FV CAN Canada Corporate, Canadian retail sales, United States, and CBOE Market Volatility Index. The response variable is aggregated interest rates, ir127(CAD rates), ir164(USD rates), ir247(GBP rates) and ir302(EUR rates). Feature selection is proceed based on the fundamental understanding of the economy and the statistical significance of features toward the model. Four linear regression models are constructed and each model corresponds to the interest rate of a region. Based on the obtained models, forecasting is done on the period of 2018 to 2022. The prediction is an indicator of future interest rate under a stress event in order to reduce investment risks in economy.

III. Methodology

III.1. Variables transformations

In order to gain insight of the data, we have plotted all response variables, interest rates, versus the time index. By looking at the graphs, we got an intuition that the difference between interest rates of 2-year maturity and 20-year maturity changed the most during the recession period of 2008.

The explanatory variables and response variables are across different time periods; therefore, we only consider the data during the common time period, which is from March 2005 to June 2017. Also, the variables have different frequencies. Response variable has daily data and explanatory variables have quarterly data, so, aggregation on the response variable to turn data into quarterly data by taking the mean is necessary. Hence, the response variable for model fitting is the mean difference between interest rate of 2-year maturity and 20-year maturity over one quarter. This response variable makes sense to real world problem because by looking at Figure 1 we can see that for all four datasets, the response variable decreased during the economic crisis from 2017 to 2019, and increased then decreased again during the recession period around 2012.

Then, we plotted graphs of all explanatory variables over time index. It has shown that some of the variables increase over time and the transformations are taken by calculating the change rates.

Besides, we observed that the Macro variables given in the stress scenario dataset are inconsistent with our explanatory dataset; to make them consistent, we changed some variables into the same units and eliminated some variables that are missing for one of two datasets.

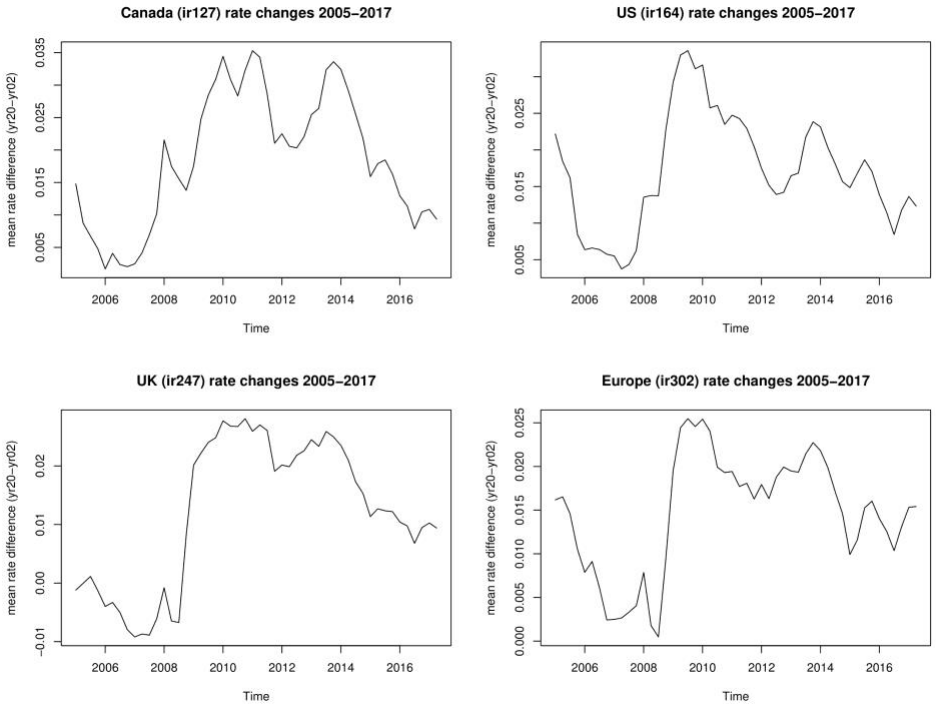

Figure 1.

III.2. Feature Selection

After variable transformation, we used AIC as our criteria to select explanatory variables for the four different models. Then, using the selected variables and 5-fold cross validation methodology to find the best regression model. The following are the graphs of cross-validation predicted values for four models.

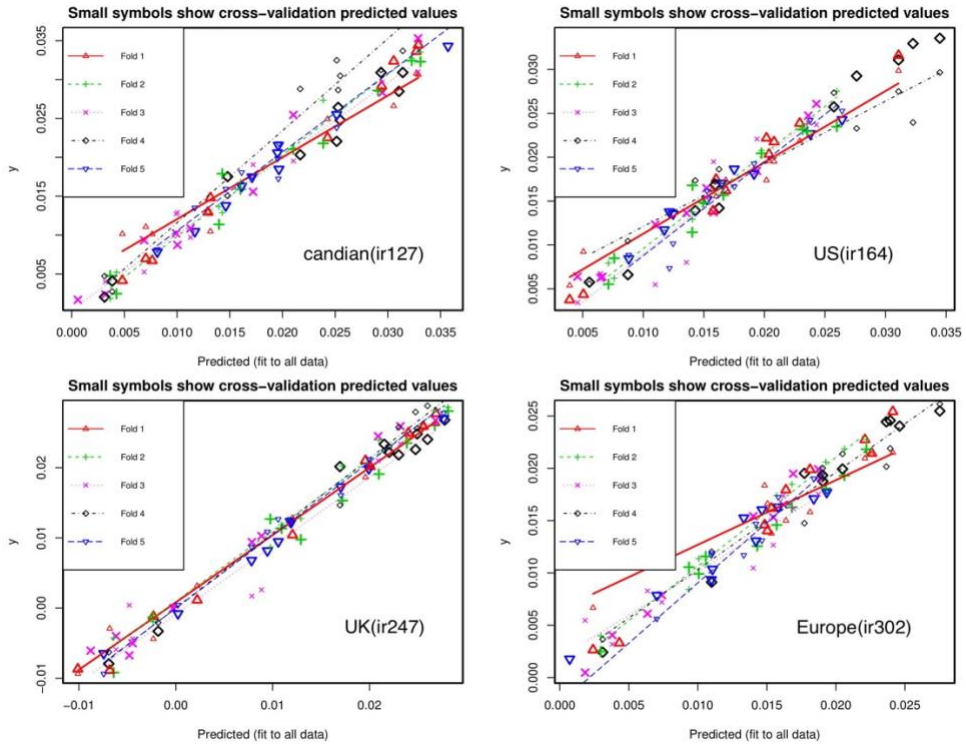

Figure 2

In order to measure how good our models are, we plot the predicted values along with actual data which are randomly selected from the dataset.

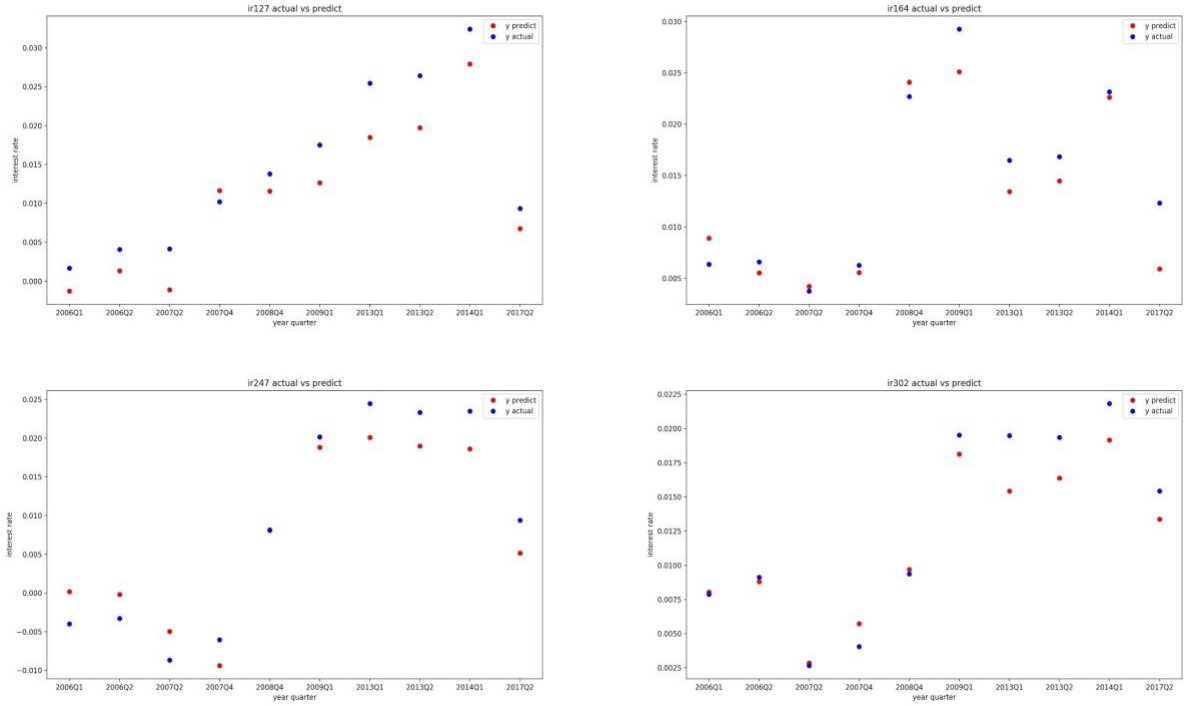

Figure 3

Model fitting results:

```
Call:
lm(formula = y127_training ~ ., data = X_training)

Residuals:
      Min        1Q    Median        3Q       Max
-0.003778 -0.001172  0.000134  0.001278  0.003900

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.43e-03   4.00e-03    0.61  0.54832
ccorp       -5.59e-03   1.23e-03   -4.54  0.00011 ***
crs          2.15e-03   6.59e-04    3.27  0.00306 **
vix          2.14e-04   9.55e-05    2.24  0.03355 *
urgdp        5.55e-03   1.78e-03    3.12  0.00436 **
ungdp       -4.77e-03   1.50e-03   -3.19  0.00373 **
ucpi        -2.44e-03   8.34e-04   -2.93  0.00694 **
ugbal        8.61e-04   6.26e-04    1.38  0.18058
ucrepi       1.72e-04   1.07e-04    1.61  0.11894
uurate       5.89e-03   7.82e-04    7.53  5.4e-08 ***
ccs         -4.87e-03   1.46e-03   -3.33  0.00260 **
ccpi         8.16e-03   1.15e-03    7.13  1.4e-07 ***
ccprof       8.03e-06   1.91e-06    4.19  0.00028 ***
wgas         2.08e-03   4.27e-04    4.87  4.8e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00247 on 26 degrees of freedom
Multiple R-squared:  0.96,      Adjusted R-squared:  0.939
F-statistic: 47.4 on 13 and 26 DF,  p-value: 9.86e-15
```

```
Call:
lm(formula = y164_training ~ ., data = X_training)

Residuals:
      Min        1Q    Median        3Q       Max
-0.005122 -0.001929 -0.000015  0.001590  0.004490

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.30e-02   1.83e-02   -3.44  0.00183 **
ccorp       -2.37e-03   1.24e-03   -1.91  0.06669 .
vix          2.55e-04   1.47e-04    1.73  0.09409 .
ungdp        1.59e-03   6.18e-04    2.57  0.01596 *
ugbal       -5.26e-04   6.77e-04   -0.78  0.44353
uurate       1.31e-04   9.84e-04    0.13  0.89536
ccpi         2.08e-03   9.75e-04    2.13  0.04170 *
cgbal        8.61e-04   1.12e-03    0.77  0.44781
ccprof       6.05e-06   2.20e-06    2.75  0.01028 *
curate       1.07e-02   2.88e-03    3.69  0.00095 ***
ceempl      -1.36e-03   9.76e-04   -1.39  0.17580
walum        5.88e-03   8.80e-03    0.67  0.50933
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00292 on 28 degrees of freedom
Multiple R-squared:  0.892,     Adjusted R-squared:  0.849
F-statistic:   21 on 11 and 28 DF,  p-value: 1.13e-10
```

```
Call:
lm(formula = y247_training ~ ., data = X_training)

Residuals:
      Min        1Q    Median        3Q       Max
-0.004490 -0.002749 -0.000323  0.002486  0.006645

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.005270   0.005103    1.03     0.31
ccorp       -0.007636   0.001113   -6.86  1.1e-07 ***
crs          0.000819   0.000491    1.67     0.11
urgdp       -0.000127   0.001810   -0.07     0.94
ungdp       -0.000431   0.001447   -0.30     0.77
ugbal        0.000718   0.000706    1.02     0.32
ucrepi       0.000163   0.000126    1.30     0.20
uurate       0.006998   0.001016    6.89  1.0e-07 ***
walum        0.004386   0.008757    0.50     0.62
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00365 on 31 degrees of freedom
Multiple R-squared:  0.93,      Adjusted R-squared:  0.912
F-statistic: 51.4 on 8 and 31 DF,  p-value: 9.28e-16
```

```
Call:
lm(formula = y302_training ~ ., data = X_training)

Residuals:
      Min        1Q    Median        3Q       Max
-0.00381 -0.00132 -0.00040  0.00110  0.00460

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.21e-02   1.83e-02   -2.85   0.0079 **
ccorp       -5.03e-03   8.47e-04   -5.94  1.7e-06 ***
urgdp        6.60e-04   1.06e-03    0.62   0.5390
ungdp        1.00e-04   8.48e-04    0.12   0.9065
ugbal        4.29e-04   4.32e-04    0.99   0.3282
uhpi         2.21e-04   6.20e-05    3.57   0.0012 **
uurate       6.29e-03   1.02e-03    6.14  9.3e-07 ***
ccprof       4.30e-06   1.59e-06    2.70   0.0113 *
walum        4.01e-03   5.44e-03    0.74   0.4673
wgas         8.46e-04   3.03e-04    2.79   0.0090 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00226 on 30 degrees of freedom
Multiple R-squared:  0.915,     Adjusted R-squared:  0.89
F-statistic:   36 on 9 and 30 DF,  p-value: 1.14e-13
```

Figure 4

III.3. Prediction

After getting the models, they predicted interest rates for 4 areas by plugging values of Canada and US from stress scenario datasets into our models. The below figure is the predicted values over 2018-2022.

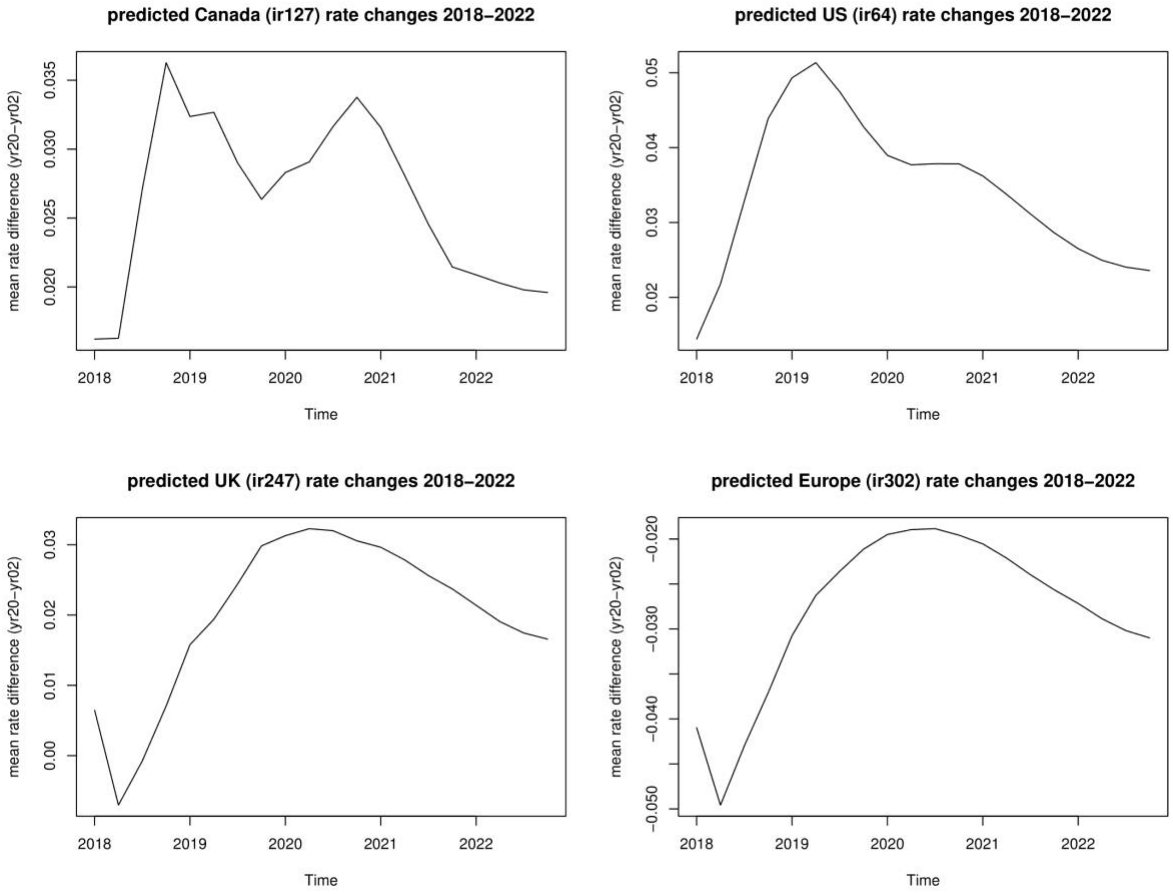

Figure 5

From Figure 5, it is clear that the trend of our predicted mean difference rate of Canadian and US is similar. They all increase dramatically in 2018, and then drop gradually when passing through 2019. As reported in 2018, the economy was resilient in 2018 but troubling signs ahead. This matches the predict of the models. Therefore, the models make good prediction for Canadian and US rates.

For UK and Europe, their predictions of mean difference rate before 2019 share more similarities in trend. This is because UK is a member of European union. But, the values differ a lot. It seems like the fitted models do not perform well for UK and Europe. This is because the given data are from Canada, US and world.

IV. Conclusion

Four models have been constructed for Canada, US, UK and Europe using Macro variables from March 2015 to June 2017. The mean change of interest rates under the stress scenario for Canada, US, UK and Europe were predicted. The US and Canadian prediction cases showed good accuracy, but may be less accurate for UK and EU cases.

IV.1. Limitations

The explanatory variables used in the model are from the given data, but there may be other variables not included. So, further data gathering is necessary. During the feature selection step, some features are eliminated based on our understanding of the economy, but this may not be accurate. Further consulting on this subject is required. We should try out different methods of calculating the response variable. This is because our model is subjective to the way of aggregating the interest rate to calculate response variable.

IV.2. Further Analysis

In the future, we will try out different response variables such as difference between interest rates of different years and the rate of change of interest rate of different years. In addition, we will add on more explanatory variables for model fitting as long as complete datasets were given.