Analysis of Interest Rate and CDS of US Under Stressed Scenario

Xueyang Zhao #50197145
Yue Chen #28166149
Xiaoce Wang #51430163

Abstract
The goals of our project are to estimate the interest rate and CDS of US using macroeconomic factors and market/credit risk-factors from 2005 to 2017, in order to forecast under stressed scenario. The methodology includes stepAIC, LASSO regression and regression tree to select features and fit models. Using the Mean Squared Prediction Power(MSPE) as our criteria, stepwise AIC has the smallest MSPE than the other methods meanwhile LASSO regression has a more sparse model. In conclusion, stepwise AIC has a better model for forecast stressed scenario from 2015 to 2022.

Introduction
During a recession period, the world economy undergoes a destructive break down. Predicting the trend of future economy under stress is a important but difficult topic. Therefore, we aim to build up a reasonable model with good prediction power using a combination of macroeconomic and market/credit data. The importance of this report is the choice of response variables, feature selections and model selections. We focus on the features related to US and only use the data set ir127(as interest rate) and igc037(as CDS) for regression.

Response Variables
The response variables for this project are interest rate and CDS. However, the data provided on explanatory variables are quarterly data while the interest rate and CDS are represented on a daily basis by maturity. After examining the patterns of daily zero curve interest rate and CDS during both stressed and unstressed period, we choose to average the differences of 10-year and 1-year maturities quarterly from 2005 to 2017.

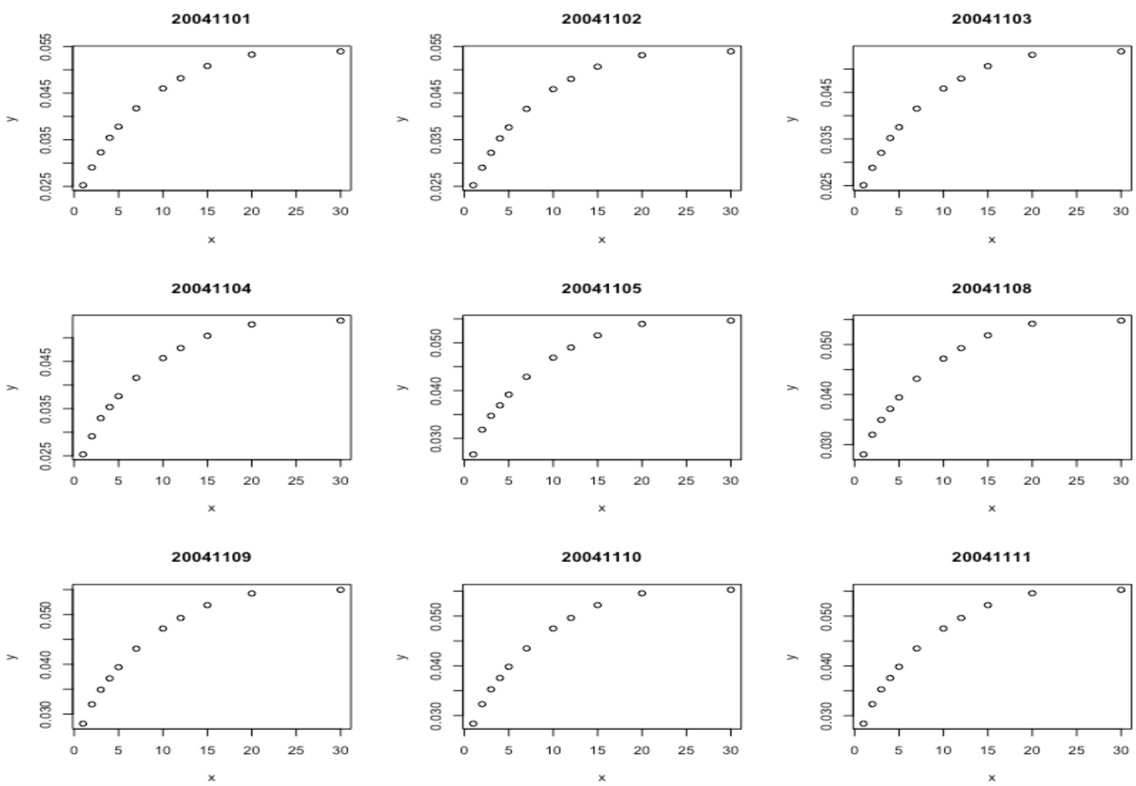

Figure 1: zero curve interest rate of ir127 from 20041101 to 2004 1103 unstressed.

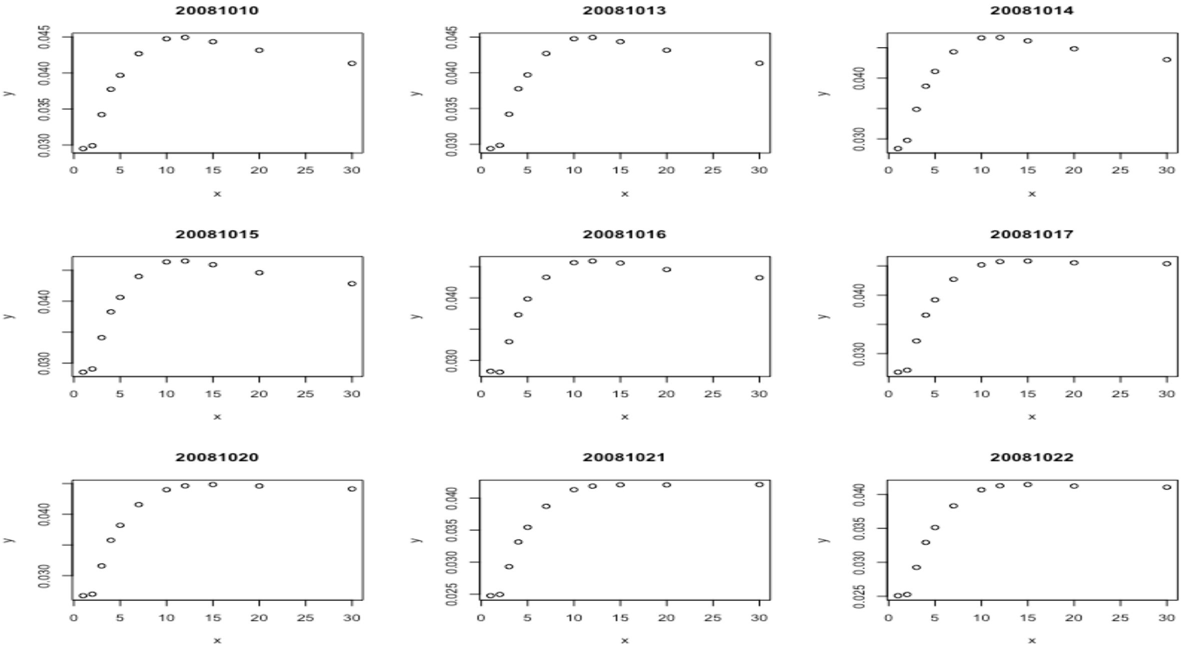

Figure 2: zero curve interest rate of ir127 from 20081010 to 20081022 stressed

From figure 1 and 2, we can see that the interest rate of stressed period tends to fall after reaching the top (around 10-year maturity). Therefore, choosing the difference between 10-year and 1-year maturity is reasonable to represent the slope of zero curve. After averaging, it is a good value for representing response variables. Similarly, after visualization tons of figures under stress and unstressed period respectively, we choose the same way of representing response variable for CDS data.

Feature Selection

The explanatory variables include macroeconomic factors, and market and credit risk-factors, which are 43 in total. Although all the data provided are time series, we choose to do regular regression after examining the plots of all the explanatory variables. Furthermore, since we aim to predict for US market, therefore all the features involving Canada are ignored.

The first step of feature selection is check the correlations between features and correlations between response variable and features. From the correlation plot, we eliminate features which have low correlation with y. And we keep only one of the highly correlated features to improve the regression results.

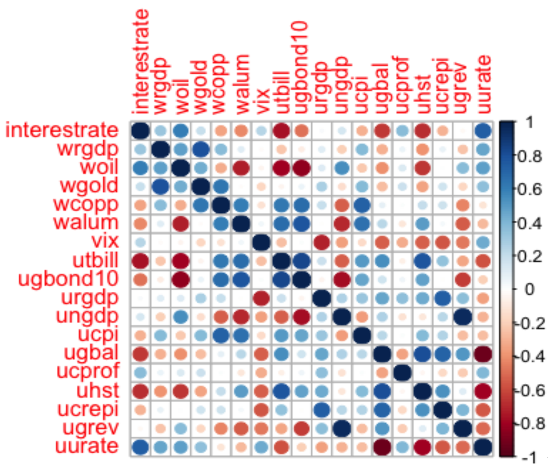

Figure 3: correlation plot of all variables

After eliminating according to correlation plot, the response variables remaining are wrgdp, woil, wgold, wcopp, walum, vix, utbill, ugbond10, urgdp, ungdp, ucpi, ugbal, ucprof, uhst, ucrepi, ugrev, uurate.

Regression Model and Further Feature Selection

The regression methods chosen are linear regression, stepwise AIC, LASSO and regression tree. First, we fit a linear regression to find out the significance of each explanatory variables. We can eliminate some insignificant variables. Then we use stepwise AIC to do further feature selection.

```
> summary(st)

Call:
lm(formula = interestrate ~ utbill + uurate + urgdp + woil +
    vix, data = ir127.clean)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0071073 -0.0023248 -0.0001462  0.0018830  0.0092289

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.429e-03  4.959e-03   0.893  0.37672
utbill      -4.321e-03  6.159e-04  -7.016 1.08e-08 ***
uurate       2.940e-03  4.039e-04   7.278 4.48e-09 ***
urgdp        3.014e-03  4.892e-04   6.160 1.96e-07 ***
woil        -9.744e-06  3.173e-06  -3.071  0.00365 **
vix          1.744e-04  9.854e-05   1.770  0.08368 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003949 on 44 degrees of freedom
Multiple R-squared:  0.856,     Adjusted R-squared:  0.8396
F-statistic:  52.3 on 5 and 44 DF,  p-value: < 2.2e-16
```

```
> coef(model, s=op.la)
18 x 1 sparse Matrix of class "dgCMatrix"
                         1
(Intercept)   4.937866e-04
interestrate  9.473186e-01
wrgdp         .
woil          .
wgold         .
wcopp         .
walum         .
vix           .
utbill       -9.670282e-05
ugbond10      .
urgdp         .
ucpi          .
ugbal         .
ucprof        .
uhst          .
ucrepi        .
ugrev         .
uurate        6.851557e-05
```

Figure 4: Stepwise AIC(left) and LASSO(right) result

From the result, stepwise AIC provides us a model with five important features. And the adjusted R-square is 0.8396 which is good. Then, we use LASSO regression in order to derive a more sparse model. Setting Alpha to 0.5, LASSO gives us a much more sparse model with only two explanatory variables -- utbill and uurate. The result is approximately similar to the Stepwise AIC model.

We also use regression tree which is a non-parametric regression model. Tree can give us the importance of features and solve the curse of dimensionality.

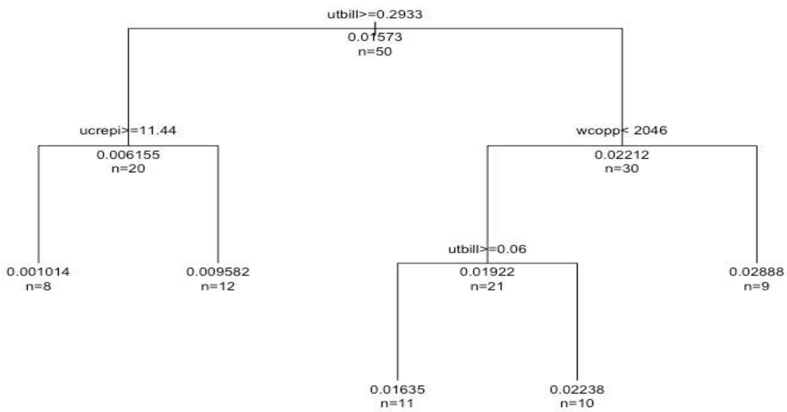

Figure 5: regression tree

From figure 5, tree gives us three important variables which are utbill, wcopp, ucrepi.

Five-Fold Cross Validation

All three of the regression models give us pretty good models. LASSO and stepwise AIC give similar result on feature importance. Regression tree is different from others. Therefore, in order to find the best model, we choose mean squared prediction error(MSPE) as our criteria and perform 5-fold cross validation on all models.

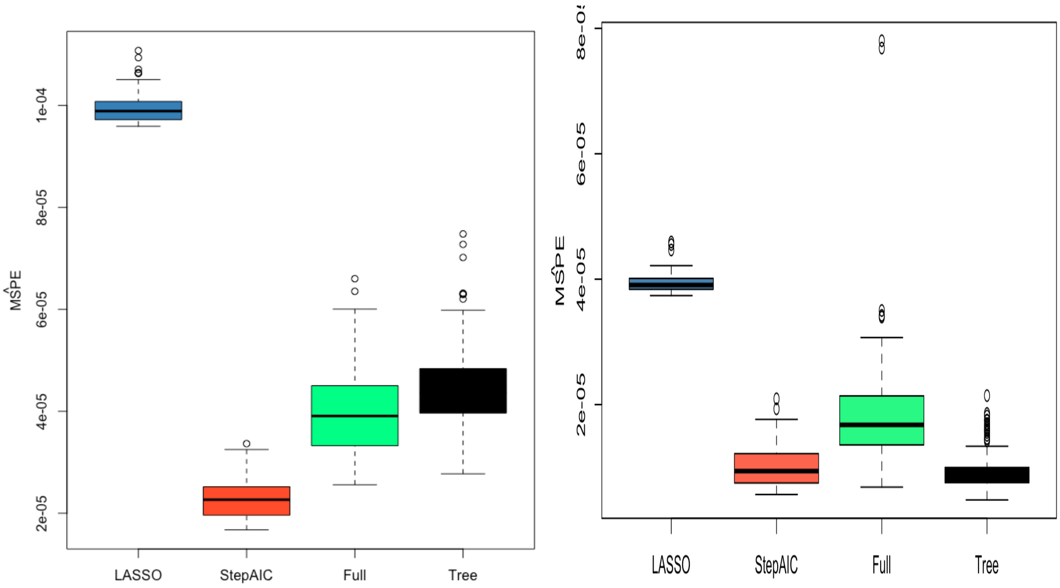

Figure 6: MSPE of four models for interest rate(left) and CDS(right)

From this figure, stepwise AIC has the lowest MSPE. And stepwise AIC provides a model with reasonable amount of features. Therefore, we choose stepwise AIC model to predict.

Conclusion

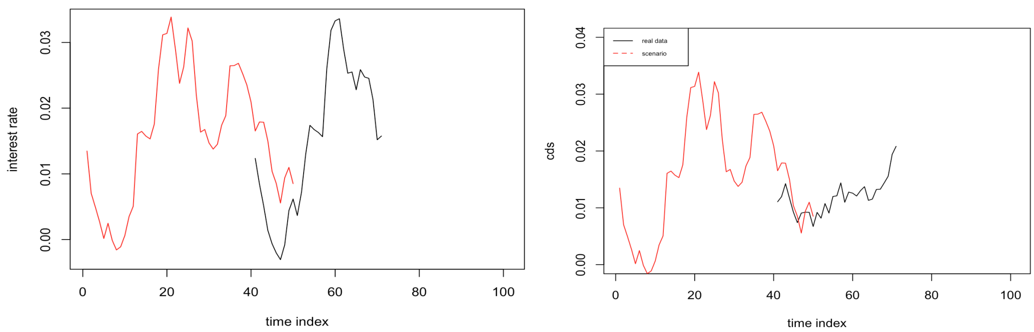

Figure 7: Plot of real data (2005-2017) and predict value(2015-2022) for interest rate(left) and CDS(right).

red line: quarterly average data of interest rate(cds) from 2005-2017 black line: prediction interest rate(cds) using Stepwise AIC model under stressed scenario from 2015-2022

According to the MSPE calculated from 5-fold cross validation and normality check, we derive two stepwise AIC regression model for interest rate and CDS, respectively.

*interest rate = -4.321e-03*utbill + 2.940e-03*uurate + 3.014e-03*urgdp - 9.744e-06*woil + 1.744e-04*vix + 4.429e-03*

*cds = 8.651e-03*utbill + 2.317e-04*uggip + 3.064e-04*ucprof - 2.615e-06*wgold + 7.229e-05*wgas - 1.819e-02*

The red line in Figure 7 is the quarterly average data of interest rate(CDS) from 2005-2017, and the black line is the prediction interest rate(CDS) using Stepwise AIC model under stressed scenario from 2015-2022. Although the scale of prediction is slightly different from real data, the future trend of interest rate and CDS agree well with our real data. From the trend shown in this figure, we can expect a brightening prospect in the future 2019-2021 financial market. However, the interest rate starts to show a decending trend at the end of 2021; meanwhile, CDS shows an increasing trend, which indicate that there will probably be another severe economic crisis starting at 2022.

Further Discussion

There are several ways to further improve our prediction. We can increase the number of observations by using more data sets. The choice of response variable also is important. For instant, we can also use average of daily interest rate instead of difference. With respect to methodology, we can improve on our current models or perform other models. For example, we can do bagging on regression trees to improve the stability, or perform random forest to get a list of important features. Furthermore, we eliminate the element time in our model, perhaps we should perform prediction with time series. Relevant method will be long short-term memory(LSTMs).