UMN MFM Winter Modeling Workshop

CCAR Project

Instructors: Andy Mack, Christy Gwynn

Members: Hyun Joo Cho, Yingying Fang, Jiachen Ning,

Nan Shen, Ya Tu, Maohua Xu

Date: Jan. 20/2015
# Table of Contents

**Introduction**

**Methodology & Result**
- Justification for processes/methodology utilized
- Assumptions, limitations, weaknesses, and uncertainties

**Capital Plan Overview**

**MATLAB Code Utilized to Developing Models**

**Reference**
Introduction

In this project, we are doing the CCAR (Comprehensive Capital Analysis Review) for Cash Money Bank, a global bank holding company headquartered in Minneapolis. Under the CCAR guidance, the Dodd-Frank Wall Street Reform and Consumer Protection Act, we need to develop a stress test (how many negative factors can a bank hold before it goes bankrupt) under three different situations: baseline, adverse, and severe at least once a year. The purpose is to forecast the future operational losses and make sure that the company can maintain sufficient fund in different operating scenarios mentioned above. We combined both subjective and objective analysis in our project. This report will provide a detail description about our methodology, justification, and final capital plan.

Methodology & Result

Part I:

Goal: Create Aggregate Loss Distribution by using Loss Distribution Approach-(LDA) a statistical method for modelling. Calibrate severity and frequency distributions based on historical losses

Data Description: Provided with random dataset of historical monthly losses from 2009 Oct to 2014 Sep (60 quarters, 20 months), totally 2867 observations

Procedure & Justification: a) we count the total number of losses per quarter which forms our frequency data for operational losses. b) Because frequency data are not continuous, we try to fit the data to a discrete distribution (we have Poisson, Binomial, and Negative Binomial as the common choice for discrete distribution), we used “distribution fitting” toolbox in MATLAB to fit the data c) fit all the loss data to a severity distribution, because monthly losses are continuous, so we tried normal, lognormal, Weibull, generalized extreme value, generalized pareto distribution to fit the data. d) In general, there is no analytical expression of the compound distribution. Computing the loss distribution requires the numerical algorithms. The most used is the Monte Carlo method. Following Monte Carlo approach to generate an aggregate loss distribution, for the first iteration, we make a random frequency draw N1 from the frequency distribution, make N1 random draw from the severity distribution, and sum the distribution to get aggregate loss. Repeat the steps for 100,000 times. Sort all the data to generate the ECDF graph
Result: b) based on the statistical definition of the three distributions, Poisson distribution fits our data the best. However, we decided to choose the Negative Binomial distribution. Looking at the cumulative distribution function graph, it is clear that negative binomial (use NB below for simplification) distribution fits the frequency data the best (we also did the PDF, but CDF seems to be the better fit by looking at the graph). We also looked at the statistical data: the result turns out that the three distributions have the same mean 143.35, the variance of NB is most close to the variance of original frequency data, and the NB has the highest log likelihood. c) because the mean and variance for Generalized Extreme Value and generalized Pareto distribution are all infinity, so there is no comparability, we eliminate those two distributions. Now we compare the three left distributions, we can see from the above CDF, lognormal fit the severity distribution the best. d) below is a graph for ECDF, we have blue line as the actual loss ECDF, green line represents the ECDF compounded from NB & lognormal distribution, red line is compounded from Poisson & Normal distribution, we can see that blue line is more close to the real data, and that once again it proves that the distributions in part b and c are correct. We are going to use this ECDF in our latter part for forecasting.
Assumption, weakenss, uncertainty: **Assumptions:** after we generate the fit distributions, we are assuming the distribution for frequency and severity is going to be Negative Binomial and Lognormal respectively in the future. **Weakness** here is that we only use 20 quarters data of operational losses to try to forecast the future losses, the number of sampling is not big enough. And when we do the Monte Carlo simulation, for the visual test, we used less number of iterations for the actual projection. And we are not using the true losses data, so there exists some distortion in the distribution.

**Part II:**

**Goal:** Perform a sensitivity analysis of economic variables supplied by the Federal Reserve, select 3-8 variables using both subjective reasoning and objective reasoning (using regression/correlation testing). Once we have the appropriate economic variables, we want to integrate those into our operational loss projection. Then we extend the loss models based on three scenarios: baseline, adverse, and severely adverse outlooks.

**Data Description:** we are using historical economic data from Q4 2009 to Q3 2014 provided by the Federal Reserve to make the prediction about future losses. We then combine forecasted Q4 2014 to Q4 2017 economic data to determine our final operational loss.

**Procedure & Justification:** a) draw the histograms and kernel density estimations for each variables, and try to find general distributions. b) We used the Kendall’s tau Coefficient Test (this is a non-parametric hypothesis test based on our result from part a), our justification for using the test is that: it does not depend on the distribution of the sample data, which means the variables do not need to be normally distributed, that’s perfect for our data c) It was not easy for us to directly determine how the economic factors influence the severity and frequency of the losses, for example, the lower inflation means the recession, so the economy goes down, however, if the inflation is extremely high, there will be crisis, which means the losses will increase a lot. We mainly base our judgment on our subjective way. d) Mainly based on
Christy’s method. We need to figure out significant factors before doing it. After we pick some influential variables from the data, which are derived from the data analysis and the intuition of the economic factors, we look at a trend of frequency related to the variables. Let’s are economic factors with j quarters and i numbers of variables, and X is a frequency generated by the Monte Carlo simulator with parameters from the historical data in the Baseline scenario. X_new =\{([ w1*(\text{mean})] *X + w2*(\text{mean})] *X + + wi*(\text{mean})] *X We also need to find each weight of the variables on the frequency following the associated direction with losses. Then, now we update i numbers of X_new frequencies as the predictors for the adverse scenario. And, do repeat this and update a set of new frequencies to the severely adverse scenario as predictors. After getting all the frequencies, we put them into the Monte Carlo simulator to calculate an aggregate losses, which means frequency * losses in each quarter. It calculate on each scenarios. This will be our forecasting data from this process for 13 quarters in advance.

Result: a) the graphs looks like any shape and only one variable shows the normal distribution b) we could notice there is an odd result that the test statistics and P-value for frequency v.s. prime rate is NA, looking at the original data, we found the prime rate is a constant number 3.3 from Q4 2009 to Q3 2014, so there is no linear relationship between Prime Rate and frequency or average losses. Then we found 14 economic factors that has statistically significant influence on the Frequency. There is 9 variables with positive linear relationship: unemployment rate, ten year treasury rate, BBB corporate yield, mortgage rate, market volatility index, euro area inflation, developing Asia real GDP growth, developing Asia inflation, UK inflation, and we have 5 variables with negative linear relationship: Dow Jones total index, house price index, Japan exchange rate, UK exchange rate. c) Therefore, based on the correlation test and economic intuition with our assumption that the amount of losses (severity) and the number of losses happened (frequency) have the same impacts from economic factors, we included the economic factors that we were sure they must influence the losses and tried many different combinations of the economic factors when building the models. We got House Price index, Commercial Price Index, Market Volatility Index in the linear regression model. The way we determine which model is the best is according to AIC, since the smaller the AIC is, the better the model will be. From the summary of the regression model we found, although there is no economic factors that had significant influence on the frequency, but the House Price Index and Commercial Price Index had statistically significant positive influence on the severity. d) our result is the projected losses, they will be shown in the Capital Plan section.

Assumption, weakness, uncertainty: Assumptions: for the part b) we are assuming the alpha to be 0.05 significant level which means the P-value is less than 0.05. We assume that the economic factors are going to be linearly related in the next 13 quarters. And we also assume that the economic variable data from Q4 2014 to Q4 2017 are correct (actually the data from baseline scenario is the average projections from surveys of economic forecasts, the data from the adverse/severe scenario is forecasted by a global weakening in economic activity). We are also assuming a higher mean per loss when we run our model. Weakness: the historical data is small, and the economic variables are 28, and we were having difficulty to analyze how those variables are related to the frequency/aggregate loss distribution, it is a little hard to choose the best model. Uncertainty: the economy is evolving rapidly, so we are uncertain about the economic variables which relates back to our assumption.
Capital Plan Overview

This is our final results for forecasting the future operational losses under three different scenarios: baseline, adverse, and severely adverse. You can find in table that a general trend is that losses for severe is larger than adverse then larger than baseline. This meets our subjective judgment in financial fields. Of course, companies can incorporate assumptions about capital actions over the planning horizon into their company-run stress tests. They can calculate post-stress capital ratios using their own consideration under each scenarios. In here, we expect to fully utilize our net income, that’s why we set percentile to 100% to figure out our capital plan.

<table>
<thead>
<tr>
<th>Date</th>
<th>Baseline</th>
<th>Adverse</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4 2014</td>
<td>1,019,240.83</td>
<td>1,039,993.59</td>
<td>1,029,262.90</td>
</tr>
<tr>
<td>Q1 2015</td>
<td>984,553.70</td>
<td>1,066,762.64</td>
<td>1,094,108.88</td>
</tr>
<tr>
<td>Q2 2015</td>
<td>976,588.50</td>
<td>1,101,234.01</td>
<td>1,139,303.38</td>
</tr>
<tr>
<td>Q3 2015</td>
<td>963,620.14</td>
<td>1,126,157.65</td>
<td>1,220,169.18</td>
</tr>
<tr>
<td>Q4 2015</td>
<td>926,526.89</td>
<td>1,156,420.24</td>
<td>1,309,287.43</td>
</tr>
<tr>
<td>Q1 2016</td>
<td>914,723.97</td>
<td>1,190,361.08</td>
<td>1,366,674.70</td>
</tr>
<tr>
<td>Q2 2016</td>
<td>892,685.80</td>
<td>1,215,781.47</td>
<td>1,415,230.83</td>
</tr>
<tr>
<td>Q3 2016</td>
<td>871,577.30</td>
<td>1,235,215.99</td>
<td>1,443,053.22</td>
</tr>
<tr>
<td>Q4 2016</td>
<td>864,645.26</td>
<td>1,248,428.18</td>
<td>1,448,671.61</td>
</tr>
<tr>
<td>Q1 2017</td>
<td>849,362.50</td>
<td>1,251,470.41</td>
<td>1,449,421.20</td>
</tr>
<tr>
<td>Q2 2017</td>
<td>821,334.09</td>
<td>1,250,699.78</td>
<td>1,447,827.36</td>
</tr>
<tr>
<td>Q3 2017</td>
<td>788,384.99</td>
<td>1,245,176.31</td>
<td>1,438,787.66</td>
</tr>
<tr>
<td>Q4 2017</td>
<td>767,640.04</td>
<td>1,237,193.46</td>
<td>1,424,028.90</td>
</tr>
</tbody>
</table>
### MATLAB Code Utilized to Developing Models

1. **Monte Carlo ECDF Code:**

```matlab
% Parameters in distributions
N=100000;
R=7.58312;% Successes
P=0.0502416;% Probability
mu=7.65258;
sigma=3.274;
A=zeros(N,1);
tic
for i=1:N
    % Make random frequency using Negative Binomial distribution
    F=random('nbin',R,P);
    for j=1:F
        % Make random severities using Lognormal distribution
        S=random('lognormal',mu,sigma);
        while S<=0
            S=random('lognormal',mu,sigma);
        end
        A(i)=A(i)+S;
    end
end
toc
```

2. **Finding the best model for the economic variables**

```matlab
% Read the data as a table and transfer to the dataset to do the linear regression
raw = readtable('economic_factor.xlsx', 'ReadRowNames', true);
full = table2dataset(raw);
ex = 'Frequency_Quaterly ~ House_Price_Index+Commercial_Price_Index+Market_Volatility_Index';
reg_ex = fitlm(full, ex);
step_ex = step(reg_ex, 'NSteps', 20)
```

3. **Forecasting the model**

```matlab
freq = xlsread('whole.xlsx', 'AE1:AE20');
percentile = xlsread('whole.xlsx', 'AF1:AF20');
AAA = [freq,freq,freq];
BBB = [percentile, percentile, percentile];
results = zeros(13,3);
Fourteen_Frequency = zeros(13,3);
per = zeros(13,3);
Base_Frequency =zeros(13,3);
k=1;
for i = 1:3
    for j = k : 13*i
        results(j - k +1,i) =438.65 -0.62314*House_Price_Index(j) -1.1224*Commercial_Price_Index(j)+ 0.56964*Market_Volatility_Index(j);
    end
    k = 13*i + 1;
end
A = [AAA;results];
B = [BBB;per];
```
Reference
