Proceedings of the 2nd International Conference on Business and Policy Studies
DOI: 10.54254/2754-1169/13/20230735

A Critical Analysis of the Efficacy of SBA Loans

Yujia Peng[1,a,*] [0000-0002-8717-8324]

[1]Indiana University Bloomington, 107 S Indiana Ave, Bloomington, IN 47405, USA
3820 E Barrington Dr.,47408, USA
a. yujpeng@iu.edu
**corresponding author*

Abstract: This paper analyzes the efficacy of the SBA loans from different perspectives, including diversity, supported employment rate, and loan size. This paper also predicts the SBA loans' future performance using historical data. To evaluate the efficacy of the SBA loans and make performance predictions, this paper used Excel, R, and Tableau to build and visualize models with large sample data. The results showed that the diversity, supported employment rate, and loan size of the SBA loans continue to decrease, which means the SBA loans can't produce a desired or intended result from the SBA. From another perspective, this study emphasizes that The U.S. Small Business Administration should not hesitate to implement actions to improve the effectiveness of SBA loans.

Keywords: SBA loans; small business loans; AI Machine Learning; Business Decision; The U.S. Small Business Administration

1. Introduction

The U.S. Small Business Administration (SBA) offers loans to small businesses for many business purposes. The SBA loans make up a large part of the total volume of small-business loans and play a significant role in promoting the development of small businesses in the United States. This study investigates the efficacy of SBA loans through different factors such as loan amounts, jobs supported, diversity. Additionally, the study predicts future performance to explore methods that can improve the SBA loans program.

1.1. Background

In the U.S., small businesses provide 50 percent of private-sector jobs. With the rapid growth of small businesses, their prominence and impact have proliferated. Various countries gradually realize small businesses' indispensable social and economic significance, making it imperative to advance the growth of small businesses [1]. SBA loans are a great help in meeting working capital requirements, maintaining financial stability, and expanding the business for small businesses. Thus, SBA loans' efficacy is a crucial element of small businesses' growth [2]. This study aimed to evaluate SBA loans' efficacy through factor analysis and predict SBA loans future performance. This research can provide detailed information for small businesses to better understand SBA loans and decide whether they should apply for SBA loans. Moreover, the study might give effective advice and inspiration to relevant government agencies on improving loan programs.

DOI: 10.54254/2754-1169/13/20230735

1.2. Problem Statement

Most studies focus on a single factor to evaluate the efficacy of SBA loans and lack the predictive analysis of SBA loans. The goals of this study are to use specific factors to evaluate whether SBA loans are successful or not, predict their future performance, and find practical approaches to improve SBA loans based on the results of the evaluation.

1.3. Purpose of the study

According to a study [3], small firms that receive an SBA loan in their first four years of operation have higher survival rates than the general population of small new firms. Since he only used one factor (survival rates) to verify the efficacy of SBA loans, this paper should consider if SBA loans showed their effectiveness on other factors such as jobs retained, loan amount, and diversity. Since there's not enough study to investigate the efficacy of SBA loans from multiple perspectives, this research aims to study the effectiveness of SBA loans comprehensively.

1.4. Research Questions

1. How is the efficacy of the SBA loans?
2. How to predict SBA loans future performance?

1.5. Significance of the study

Although SBA loans have played a positive role in the innovation and entrepreneurship of small businesses in the U.S. for the past 20 years, the efficacy of SBA loans still needs to be carefully evaluated through different criteria to have an in-depth understanding of SBA loans' performance. In addition, using historical data to predict future performance can help improve SBA loans.

With this study, small businesses with loan needs can have a comprehensive insight into SBA loans and decrease the probability of choosing an inappropriate loan program as each small business has a different size, industry type, market value, and so on. To be an effective loan program, SBA loans' policy, such as application requirements, approval process, and offer amount, needs to be adjusted regularly because of market fluctuations, political changes, and social developments. This study can help SBA workers improve SBA loans with an in-depth analysis of SBA.

2. Literature Review

Orzechowski found a positive relationship between the growth in SBA lending per capita and the change in the state's civilian employment rates from the early 1990s to 2013 [4], which means SBA loans may help with the public goal of assisting small businesses and may contribute to the finance-growth nexus. This hypothesis provides a general direction for this study of the efficacy of SBA loans and help answer the question, "Has the SBA small business loan program been successful?". However, the state's civilian employment rate might be too broad to be an appropriate factor to evaluate the efficacy of SBA loans because many other macroeconomics and microeconomics factors can affect the employment rate, and SBA loans might not necessarily cause the changes in the employment rate. Using jobs supported by SBA loans instead of state employment rates as a factor can increase the accuracy of the factor analysis.

Another similar factor in evaluating the efficacy of SBA loans is small firm survival rates. Galli-Debicalla mentioned his hypothesis that small firms that receive an SBA loan have higher survival rates than other small firms and testified in his study [3]. Even though the result is statistically significant, his sample size is too small to be persuasive enough. In addition, he didn't provide enough information about the industry types of small businesses in his sample pool. It is possible that his choice of small businesses is not diverse enough, and the results of his models might be biased. In this study, Increasing and diversifying the sample size and using different measurements to check if the model is a good fit are important.

Different from the above opinions, De Andrade and Lucas expressed the disadvantage of SBA loans [5]. Based on the data obtained from SBA, they found that large lenders charge borrowers relatively high rates. Since SBA loans have advantages and disadvantages, critical analysis of different factors can objectively measure the efficacy of SBA loans. Based on the studies collected in the reference list, various elements can be divided into three categories: SBA loans related (such as loan size, loan amount, loan popularity, diversity), small business-related (such as survival rate, job retained), and macro-level(such as regional economic growth).

Stadig used loan growth to predict bank performance in his study [6]. This research will not use loan growth to predict bank performance. Still, it will use predictive analysis to forecast SBA loans future performance and know whether the loans will increase, decrease, or stay unchanged in the future. With a thorough factor analysis of the efficacy of SBA loans and predictive analysis of SBA loans' performance, this research aimed to answer the question of has the SBA small business loan program been efficient and how to predict SBA loans future performance.

3. Methodology

The dataset used for the research comes from the U.S. SBA, which includes historical data from1962 through 2014 [7]. The dataset is comprehensive as it contains 899,164 observations for over 50 years. With a long-horizon dataset, evaluating the efficacy of SBA loans would be more persuasive and precise as the sample size is big enough. In addition, the dataset contains different variables such as borrower state, North American industry classification system code, the fiscal year of commitment, number of jobs retained, and SBA’s guaranteed amount of approved loan. Thus, the dataset with multifactor allows for analyzing the SBA loans from different perspectives.

Data reduction and transformation are inevitable steps to prepare the dataset for the SBA loan analysis. Since the original sample size is too large and contains outdated data that is ineffective, only historical data from 2005 to 2014 is maintained. In addition, multiple variables should be utilized from different perspectives to evaluate the efficacy of SBA loans comprehensively with less bias. However, the dataset contains 27 variables, some of which are not strongly relevant to the purpose of the research. The state and NAICS variables are good indicators of the diversity of SBA loans. The createjob and retainedjob variables can reflect the impact SBA loans bring to employment. The SBA loan amount variable can show the loan size. In this case, 7 variables should be focused on for the efficacy analysis. These variables are LoanNr_ChkDgt, State, NAICS, ApprovalFY, CreateJob, RetainedJob, and SBA_Appv. The LoanNr_ChkDgt is the identifier and rename it as LoanID to read this variable more easily. ApprovalFY is the fiscal year of commitment, and SBA_Appv is SBA’s guaranteed amount of approved loan.

Since CreateJob and RetainedJob are similar categories, to better understand the SBA loan’s impact on employment, these two variables are combined in a new column named JobSupported by adding the values of these two variables. Note that NAICS is a 2 through 6 digits hierarchical classification system used by federal statistical agencies in classifying business establishments. The first two digits of the NAICS classification represent the economic sector. In this case, 6 digits data should be transformed into the first 2 digits code of NAICS to analyze this variable more conveniently. Create a new column named NAICS2 to keep the first 2 digits of the NAICS code by the left function in Excel. Table 1 shows the 2-digit sectors and a corresponding description for each sector.

DOI: 10.54254/2754-1169/13/20230735

Table 1: 2-digit sectors with corresponding description.

Sector	Description
11	Agriculture, forestry, fishing and hunting
21	Mining, quarrying, and oil and gas extraction
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale trade
44-45	Retail trade
48-49	Transportation and warehousing
51	Information
52	Finance and insurance
53	Real estate and rental and leasing
54	Professional, scientific, and technical services
55	Management of companies and enterprises
56	Administrative and support and waste management and remediation services
61	Educational services
62	Health care and social assistance
71	Art, entertainment, and recreation
72	Accommodation and food services
81	Other services (except public administration)
92	Public administration

Finally, the dataset changed from 899164 rows to 322286 rows columns. Figure 1 is the descriptive statistics for LoanID split by approval year.

To evaluate the diversity, supported employment rate, and loan size of the SBA loans, this research used Excel, R, and Tableau to build and visualize histograms and line charts to show the historical change. To make future performance predictions, I built two models, one is a single linear regression model, and the other is a multivariate linear regression model. I compared the performance statistics of these two models to find the more appropriate model for the SBA loans' future performance prediction. Here, I assume the variables in the dataset have a linear relationship and only consider the linear regression model. The relationship between variables may be exponential or other mathematics relationship, but the models would be much more complicated, so I would not discuss them here.

LoanID by Approval year ▾

Descriptive Statistics

	LoanID									
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Valid	77525	76040	71876	39540	19126	16848	12608	5997	2458	268
Missing	0	0	0	0	0	0	0	0	0	0
Mean	5.443e+9	2.820e+9	2.488e+9	3.035e+9	3.455e+9	3.968e+9	4.571e+9	5.132e+9	5.909e+9	6.620e+9
Std. Deviation	3.652e+9	2.587e+9	2.432e+8	1.682e+8	1.336e+8	1.741e+8	1.733e+8	1.518e+8	3.002e+8	1.056e+8
Minimum	1.075e+9	1.000e+9	2.051e+9	2.735e+9	3.232e+9	3.672e+9	4.281e+9	4.899e+9	5.458e+9	6.476e+9
Maximum	9.143e+9	9.899e+9	9.928e+9	3.448e+9	3.700e+9	4.281e+9	5.717e+9	5.455e+9	6.466e+9	6.921e+9

Figure 1: Descriptive statistics for LoanID split by approval year

DOI: 10.54254/2754-1169/13/20230735

4. Results

4.1. Diversity

Diversity is an essential element in evaluating whether a loan is successful or not. An efficient loan should have high diversity in the borrower's business region and business type to show fairness and comprehensiveness. If a loan can only focus on a small region or limited industry type, it is unreasonable to acknowledge that the loan has high efficacy. In this research, state and NAICS2 can be used to evaluate SBA loan diversity. Use the dataset in sheet1 to analyze these two variables. The countif function can help to calculate the total number of SBA loans in each state from 2005-2014. Table 2, Figure 2, 3, and 4 contain the histogram, map, and descriptive statistics of the total number of loans.

Table 2: Total number of SBA loans in each state.

State	Total Number of Loans
AK	450
AL	2356
AR	1665
AZ	6900
CA	42717
CO	7677
CT	3834
DC	640
DE	816
FL	17238
GA	8212
HI	1418
IA	2862
ID	3505
IL	13380
IN	6584
KS	3012
KY	2953

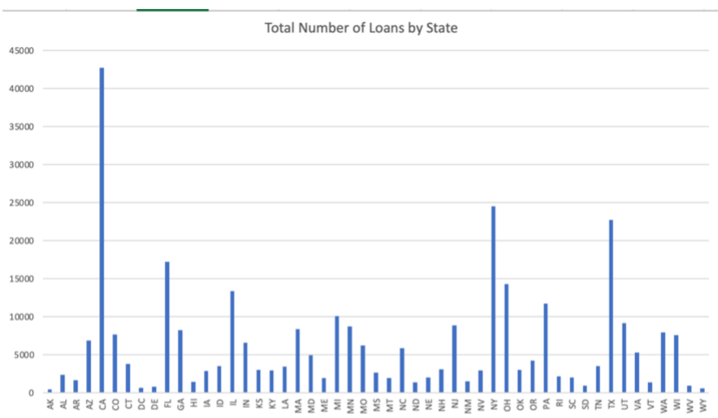
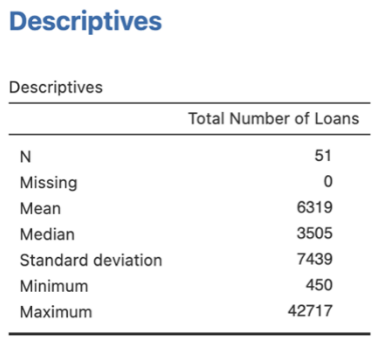


Figure 2: Histogram of total number of SBA loans in each state.



Descriptives

Descriptives

	Total Number of Loans
N	51
Missing	0
Mean	6319
Median	3505
Standard deviation	7439
Minimum	450
Maximum	42717

Figure 3: Descriptive statistics of the total number of loans.

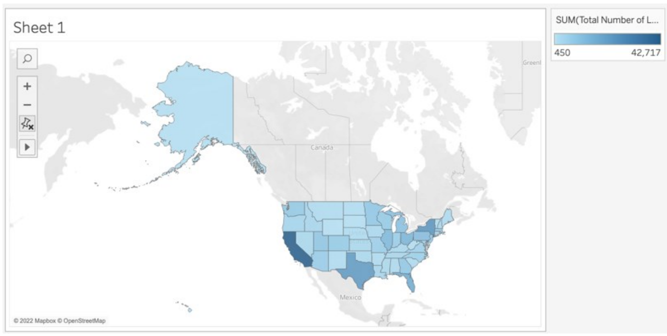


Figure 4: density map of the total number of loans.

These figures show that the standard deviation of the number of loans in each state is large. Some states, such as CA and NY, have a large number of SBA loans, while others, such as WV WY, only have a small number of SBA loans.

NAICS2 is another variable that can be used to evaluate SBA loans' diversity. Use the same method as the state variable to deal with the NAICS2 variable. Below are the table, histogram, and descriptive statistics of the total number of loans by each type of NAICS2.

Table 3: Total number of SBA loans in each type of NAICS2

NAICS2	Total Number
11	2548
21	756
22	246
23	35697
31-33	26105
42	19902
44-45	54251
48-49	14567
51	5900
52	5799
53	8038
54	34607
55	79

DOI: 10.54254/2754-1169/13/20230735

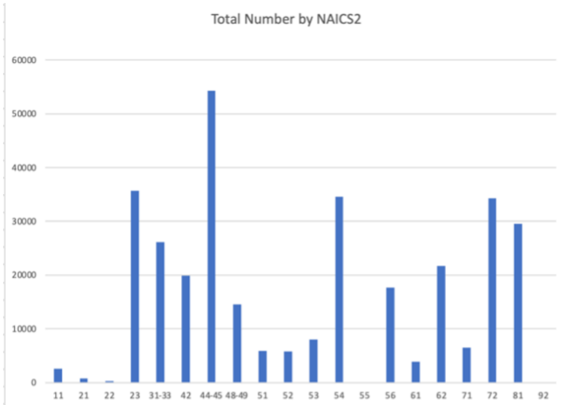


Figure 5: Histogram of total number of SBA loans in each type of NAICS2.

Descriptives

Descriptives

	Total Number
N	20
Missing	0
Mean	16114
Median	11303
Standard deviation	15462
Minimum	78
Maximum	54251

Figure 6: Descriptive Statistics of total number of SBA loans in each type of NAICS2.

The standard deviation of the total number of loans by NAICS2 is significant. From the histogram, I can find that some businesses such as utilities (22), Management of companies and enterprises (55), and Public administration (92) has a small number of SBA loans. Based on state and NAICS2 variables, the diversity of SBA loans is not outperforming as the standard deviation can show that SBA loans have a bias in regions and business types.

4.2. Employment

Since the dataset is still very large, data integration is required to reduce the data size before starting the analysis. As the purpose of the research is to analyze the number of jobs supported by SBA loans, integrating rows of data if they have the same approval years instead of analyzing the data based on LoanID would be a good choice. Using LoanID to analyze the data would have a super huge workload as each loan has a unique LoanID. Integrating data can reduce the data size and save the research time while not affecting the investigation of SBA loans. In this way, the integrated data can be used to analyze the trend of job supported and loan amount in time series.

Create a new sheet named Group By, and use the power query's group by function in Excel to combine rows of the data if the ApprovalFY column is the same, and sum the SBA loan amount and number of jobs supported. After the integration, the dataset only contains 10 rows with 3 variables in Figure 7.

DOI: 10.54254/2754-1169/13/20230735

ApprovalFY	SBA Loan Amount	Number of Job Supported
2005	8665698207	554220
2006	7515929145	611465
2007	6512106059	599945
2008	4377100434	357151
2009	2739492442	207348
2010	2800066981	191401
2011	2830214612	166502
2012	1226984156	75825
2013	517439050	29025
2014	36277570	2811

Figure 7: Main variables after data integration.

The dataset after grouping can be employed in the analysis can testify to the effectiveness of the SBA loan’s jobs supported. Data in 2014 should be excluded as it only contains the first half year of 2014 data. Then, draw a line chart to see the trend of the number of jobs supported from 2005 to 2013 in figure 8.

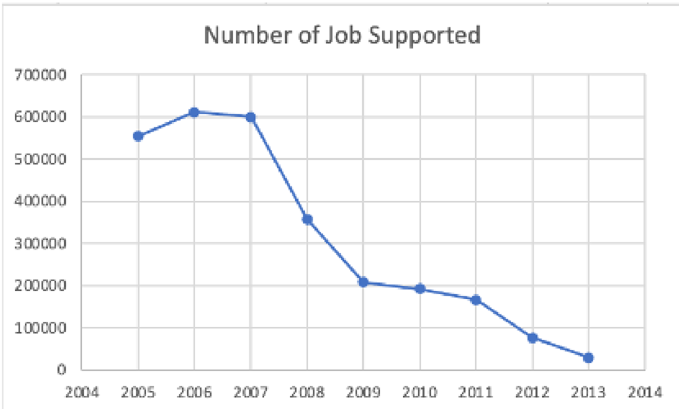


Figure 8: Trend of the number of jobs supported from 2005 to 2013.

The line chart shows the decreasing trend of the number of jobs supported. Even though the number increased in 2005 and the decreasing speed became slow from 2009-2011, the overall trend of the number of jobs supported by SBA loans is decreasing rapidly. The number of jobs supported significantly reduced because of the great recession between 2007 and 2009. From 2008-2013, the trend showed no signal of economic recovery, and the reducing trend didn’t stop. Based on the historical data, SBA loans have made a huge contribution to job support, the maximum number of jobs supported is 611465 per year, and the minimum number can reach 29025. However, analyzing the impact of SBA loans on employment from a long-term perspective, the efficacy of SBA loans is reduced.

4.3. Loan Size

Another element that can indicate the efficacy of SBA loans is the loan size. The loan amount guaranteed by SBA loans can reflect the loan size. The larger the amount, the more efficient the SBA loans are. Create a line chart to see the trend of the loan amount from 2005 to 2013 in figure 9.

DOI: 10.54254/2754-1169/13/20230735

Figure 9: Trend of the loan amount from 2005 to 2013.

The chart is similar to the graph of jobs supported. The trend from 2005 to 2013 decreased while it decreased most significantly during 2007-2009.

Based on the above analysis, the efficacy of SBA loans is reduced in three perspectives: diversity, employment, and loan size. Since limited factors are analyzed in this research, SBA loans may make progress in other factors such as default rates.

4.4. SBA loans' future performance prediction

To forecast the SBA loans' future performance, the loan amount guaranteed by SBA would be an appropriate target for the prediction. The loan amount guaranteed by SBA loans can reflect the loan size and development. If the loan amount is large, it proves that SBA loans have adequate funding to support the development of small businesses, and to some extent, it can also reflect that SBA loans have a low default rate or good performance. In this case, the loan amount is the target variable in the machine learning models to predict the SBA loan's future performance. This research applies linear regression and logistic models in RStudio for predictive analysis.

Import the dataset and divide it into 80% training and 20% test sets. The reason to partition the dataset is to avoid model overfitting. Use the correlation table between the target variable(SBA_Appv) and other variables in the data frame to decide on predictor variables.

	SBA_Appv
Zip	0.07890063
NAICS2	0.01678701
ApprovalFY	0.09423719
Term	0.54804331
NoEmp	0.15831419
NewExist	NA
Jobs.Supported	0.14861766
FranchiseCode	0.10230166
UrbanRural	0.02057854
GrAppv	0.98609178
SBA_Appv	1.00000000

Figure 10: Correlation between variables.

Based on the correlation table, the SBA_Appv variable has a 0.54 correlation coefficient with the Term variable and 0.99 with the GrAppv variable. The SBA_Appv variable has positive relationships

DOI: 10.54254/2754-1169/13/20230735

with both of these two variables, so the Term variable and GrAppcv variable can be the model predictors. Use the training dataset to build two linear regression models:

— Single factor with the GrAppv variable
— Two factors with the GrAppv and the Term variable

```
> summary(model0)

Call:
lm(formula = loan_data$SBA_Appv ~ loan_data$GrAppv, data = train_set)

Residuals:
     Min       1Q   Median       3Q      Max
-1452522    -8201     4644    11507   871512

Coefficients:
                     Estimate    Std. Error t value              Pr(>|t|)
(Intercept)     -13222.9430709    86.3284681  -153.2 <0.0000000000000002 ***
loan_data$GrAppv     0.8431490     0.0002503  3368.2 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43950 on 322284 degrees of freedom
Multiple R-squared:  0.9724,    Adjusted R-squared:  0.9724
F-statistic: 1.134e+07 on 1 and 322284 DF,  p-value: < 0.00000000000000022
```

Figure 11: Results of single factor model (0).

```
> summary(model1)

Call:
lm(formula = loan_data$SBA_Appv ~ loan_data$Term + loan_data$GrAppv,
    data = train_set)

Residuals:
     Min       1Q   Median       3Q      Max
-1428313    -7702     3769    11451   940969

Coefficients:
                     Estimate    Std. Error t value              Pr(>|t|)
(Intercept)     -25875.1990181   133.2822156  -194.1 <0.0000000000000002 ***
loan_data$Term     185.8802384     1.5157013   122.6 <0.0000000000000002 ***
loan_data$GrAppv     0.8246154     0.0002876  2867.3 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42950 on 322283 degrees of freedom
Multiple R-squared:  0.9736,    Adjusted R-squared:  0.9736
F-statistic: 5.945e+06 on 2 and 322283 DF,  p-value: < 0.00000000000000022
```

Figure 12: Results of two factors model (1).

For model 0 that only use the GrAppv predictor, the model is

$$y(SBA\ Appv) = -13222.9431 + 0.8431 * GrAppv \tag{1}$$

For model 1 that use the GrAppv and the Term predictors, the model is

$$y(SBA\ Appv) = -25875.1990 + 0.8246 * Grappv + 185.8802 * Term \tag{2}$$

Since all P-values are smaller than 0.05, the predictors are statistically significant. Then, use these two models to predict the loan amount in the test dataset and evaluate their predictive performance by accuracy.

```
> accuracy(pred_loan0, test_set$SBA_Appv)
               ME     RMSE      MAE       MPE     MAPE
Test set 66184.59 303943.3 129982.6 -222.9339 335.4943
```

Figure 13: Predictive performance by accuracy in test dataset for the single factor model (0).

```
> accuracy(pred_loan1, test_set$SBA_Appv)
               ME     RMSE      MAE       MPE     MAPE
Test set 67281.09 304553.3 130664.9 -215.6126 336.2844
```

Figure 14: Predictive performance by accuracy in test dataset for the two factors model (1).

Calculate the prediction errors for each model and describe the errors using histograms

DOI: 10.54254/2754-1169/13/20230735

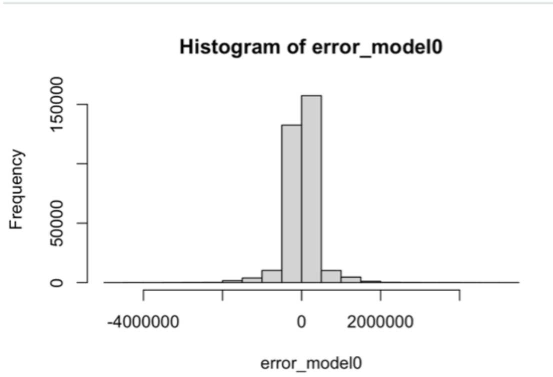


Figure 15: Histograms of prediction errors for single factor model (0).

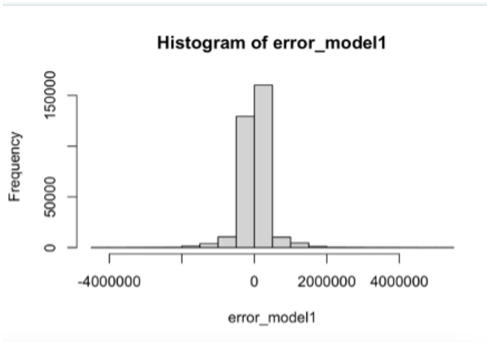


Figure 16: Histograms of prediction errors for two factors model (1).

To evaluate the model overfitting behavior, calculate the predictive performance for the training dataset.

```
> accuracy(pred_loan_train0, train_set$SBA_Appv)
               ME     RMSE      MAE       MPE     MAPE
Test set 17082.25 354941.1 164913.6 -560.8554 659.8312
```

Figure 17: Predictive performance by accuracy in training dataset for single factor model (0).

```
> accuracy(pred_loan_train1, train_set$SBA_Appv)
               ME   RMSE      MAE       MPE     MAPE
Test set 17487.59 355089 165845.5 -558.2392 663.8414
```

Figure 18: Predictive performance by accuracy in training dataset for two factors model (1).

For model 0, the RMSE for test data is 303943, and training data is 354941, which means it performs better on test data than on training data. Thus, model 0 is not over-fitted. For model 1, the RMSE for test data is also smaller than the training data, so model 1 is not overfitting. From the histogram, I can see that the errors of the two models are similar. If I check the accuracy data for these two models, all data, including RMSE for model 0, is smaller than model 1. The smaller the RMSE, the more concentrated the data around the best fit line. In this case, the linear regression model with a single factor would be better in predictive performance. With the model

$$y(SBA\ Appv) = -13222.9431 + 0.8431 * Grappv \tag{3}$$

I can predict future loan amount if I know the Gross Amount of Loan Approved by Bank (GrAppv).

5. Conclusions

In evaluating the efficacy of SBA loans, histograms and line charts are used to analyze the historical data of loan diversity, the number of jobs supported, and the amount. Based on the charts created above, the diversity, the supported employment rate, and the size of SBA loans continues to decrease. In this case, which means the SBA loans can't create a desired or intended result from the SBA, and the SBA loans doesn't have high efficacy. The U.S. Small Business Administration should not hesitate to implement actions to improve the effectiveness of SBA loans. The loan amount variable has positive relationships with the term and the gross amount of loan approved by bank variables. After testing, the gross amount of loan approved by bank can be a predictor in the linear regression model to forecast the loan amount for the future.

The missing data in some years may be the limitations of this research paper, as the loss of the data can cause bias and volatility in histograms, line charts, and linear regression models. In addition, linear regression models could be the limitation of this paper because, in most real-life scenarios, the relationship between the dataset's variables isn't linear. Still, I assume the variables in the dataset in this paper have linear relationships. However, this paper still contributes to evaluating the efficacy of the SBA loans comprehensively through different criteria and helps the audience have an in-depth understanding of SBA loans' performance. In addition, the models in this paper that used historical data to predict future performance are valuable tools for the loan while they can help SBA workers to improve SBA loans.

References

[1] *Craig, B. R., Jackson III, W. E., & Thomson, J. B. (2007). Small firm finance, credit rationing, and the impact of SBA-guaranteed lending on local economic growth. Journal of small business management, 45(1), 116-132.*

[2] *Wichmann Jr, H., Abramowicz, K., & Sparks, H. C. (2008). SBA helps small businesses think big. Strategic Finance, 90(4), 45-50.*

[3] *Galli-Debicella, A. (2020). The efficacy of SBA loans on small firm survival rates. Journal of Small Business Strategy, 30(2), 26-34.*

[4] *Orzechowski, P. E. (2020). US Small Business Administration loans and US state-level employment. Journal of Economics and Finance, 44(3), 486-505.*

[5] *De Andrade, F., & Lucas, D. (2009). Why do guaranteed SBA loans cost borrowers so much?. Unpublished working paper, Northwestern University and MIT.*

[6] *Stadig, D. (2017). Does fast loan growth predict poor bank performance?. Unpublished master's thesis.*

[7] *Li, M., Mickel, A., & Taylor, S. (2018). "Should This Loan be Approved or Denied?": A Large Dataset with Class Assignment Guidelines. Journal of Statistics Education, 26(1), 55-66.*