CSBS Project Proposal

University/College Name: College of William & Mary, Mason. A School of Business

Participant List: 2022 graduates, MS in Business Analytics

Frank Wan, zwan02@email.wm.edu Layla Liang, zliang04@email.wm.edu

Kira Qian, qqian@email.wm.edu Hui Wang, hwang35@email.wm.edu

Lena Müller, lmueller@email.wm.edu

Faculty Sponsor: J. Alejandro Gelves, jagelves@wm.edu

Research Idea/Hypothesis:

During the pandemic, community banks had a critical function in supporting local businesses with the Paycheck Protection Program (PPP), especially the loan forgiveness. Considering its objective of sustaining small business and promoting employment, forgiveness amounts indicate loan usage aligned with program goals. We assessed the performance of community banks and non-community banks using *Loan Inefficiency Rate* shown below:

$$Loan\ \ Inefficiency\ Rate = \frac{\sum(Current\ Loan\ Approved - Loan\ Forgiveness\ Amount)}{\sum Current\ Loan\ Approved}$$

Based on the loan inefficiency rate defined above, we structured a hypothesis (with 5% significance level):

Ha: Community banks allocate PPP loans more efficiently compared to non-community banks

Ho: Community banks will not allocate PPP loans more efficiently compared to non-community banks

We analyze current efficiency of PPP loan relief between community banks and non-community banks during the pandemic and create visualizations and dashboards for pandemic loan modifications.

Based on the small business data in SBA and Census, we simulate two sets of sample data with size of 1 million for community banks and non-community banks to support our hypothesis t-testing. We can make experiments on the two sets of samples in every state, in different industries with NAICS Code and in Urban-rural indicators. Therefore, there will be a diversity of views from the hypothesis with different generated experiments.

Additionally, we conduct a regression analysis on PPP and small business dataset investigating the relationship between inefficiency rate (target) and factors including Urban-rural indicator, business area (NAICS), unemployment rate and household income. This is to find causal relationships between variables and provide insight for the PPP approval process.

Expected Data Sources:

Advocacy, O. of. (2021, August 30). 2020 Small Business Profiles for the states and territories. SBA's Office of Advocacy. Retrieved January 29, 2022, from https://advocacy.sba.gov/2020/05/20/2020-small-business-profiles-for-the-states-and-territories/

Bureau, U. S. C. (2021, October 8). SUSB tables. Census.gov. Retrieved January 29, 2022, from https://www.census.gov/programs-surveys/susb/data/tables.html

Explore census data. (n.d.). Retrieved January 29, 2022, from https://data.census.gov/cedsci/table?tid=ABSCS2019.AB1900CSA04&hidePreview=true

PPP foia - public_150k_plus_220102.csv - U.S. small business administration (SBA): Open data. U.S. Small Business Administration (SBA) | Open Data. (n.d.). Retrieved January 29, 2022, from https://data.sba.gov/dataset/ppp-foia/resource/c84fa84d-c047-4b66-8056-5748f6a2bfca

US Census Bureau. (2021b, November 22). Income Data. Census.Gov. https://www.census.gov/topics/income-poverty/income/data.html

Proposed Method(s) to Test Hypothesis:

To better understand the PPP loan forgiveness rate, we created interactive visualizations and published dashboards in Tableau. We first defined our new metric (Loan Inefficiency Rate) in Tableau and combined the Urban-rural indicator for the first sight of our hypothesis. To further investigate our hypothesis in multiple dimensions, we also created a map and bar chart by states for potential experiment directions.

In order to test our proposed hypothesis, we need the simulated sample sizes, sample means and the sample standard deviations with t-Test. If the p-value computed is less than our significance level, then we are rejecting our null hypothesis.

We made the assumption that the functional form of loan inefficiency is a polynomial regression model. After model selection, we separate PPP dataset into training and testing groups randomly and use the training data to train the model with forward and backward elimination. Therefore, the final model chosen is parsimonious. Lastly, this final model will be evaluated with the test dataset.

Deliverables:

The result of our simulation and model will indicate what significant factors will be affecting PPP loan efficiency rate. Stakeholders and decision makers can modify policies according to our hypothesis testing. Borrowers' states and industries, Urban-rural indicators and local unemployment rates can be considered as benchmarks while deciding whether or not to approve forgiveness loans. Furthermore, threshold relief, approval process management, and budget can be adjusted for different locations based on the prospective efficiency rates.