<u>Capstone Project</u>

COVID-19 Induced PPP Loans and Their Implications

Xinhui (Michael) Chen, PhD, CFA

Federal Reserve Bank of San Francisco[1]

February 1, 2022

Final Deadline: March 01, 2022

ABA Stonier Graduate School of Banking

[1] Disclaimer: the views expressed in the paper solely represent those of the author's own opinions based on publicly available data. They do not necessarily reflect the views of the Federal Reserve System or Federal Research Banks.

Contents

Executive Summary

After the onset of COVID-19 in the US in February 2020, the federal government initiated the Coronavirus Aid, Relief, and Economic Security (CARES) Act. The Small Business Administration (SBA) was authorized to administrate the act in the form of the Paycheck Protection Program (PPP) to help businesses survive the pandemic. The paper intends to present the analytical work to understand the impact, the effectiveness, and the contributing factors of the loans. According to the PPP loan data from SBA website as of July 6, 2021, the total of initial loan approval amount was approximately $803 bn and over 11 million loans.

Key Findings from the Analyses

In general, PPP loans aligned reasonably well with the COVID-19 cases at the state level, indicating that the loans went to the states where COVID-19-related financial help was most needed. The distribution of PPP loans by borrowers' state was roughly proportional to the economic activities and output. Overall, the average loan size was $68,220. On the per-employee basis, Black businesses had the highest loan amount at $10,986, followed by Unanswered group ($8,986) and White ($8,622). Asian businesses and Other had the lowest loan size per employee at $6,552 (60% of Black) and $5,429 (50%) respectively. Of all the PPP loans, Male accounts for 29% while Female 8% (the remaining with gender not specified). The average loan sizes per employee were $8,729 and $7,897 for Male and Female business owners respectively, representing a gender gap of about 10%. In terms of impacted sectors, the Construction industry had the largest initially approved loan amount at $98.68bn (roughly 12% of the entire PPP loan amount), followed by Health Care and Social Assistance, Accommodation and Food Services, Retail Trade etc. 94% of the loans were to for-profit organizations while 6% were to non-profit organizations. The average loans size was $189,959 for non-profit, which is 2.9x as high as the for-profit ($65,401). However, on the per-employee basis, the average non-profit loan is $8,114, which is about 10% less than that of for-profit ($8,950).

Recommendations

The author studied PPP loans along different dimensions such as HUB Zone, LMI loan types, Racial gap, Gender gap, Rural/Urban gap, etc. The findings are controlled results in that, it neutralizes the impact of other factors, i.e., given everything else equal.

The overall PPP loan sizes favored geographical areas with higher incomes. For example, in areas with higher average Household AGI, the overall PPP loan sizes are bigger. This is not intuitive. What is even more unintuitive is that the loan sizes per employee also favored higher income regions. The loans were expected to be tilted toward low-income area. Is it because the loan applicants' expectations for the lower income areas were lower, e.g., business costs were lower, therefore, they applied for smaller loan sizes, or, is it because the applicants were less informed? The author suggests that the federal government investigate the PPP loans' disparity with expectations. For example, study the loan amount applied vs. the actual approved loan amount.

For Asian American applicants, both overall loan sizes and per-employee loan sizes are the lowest among all major racial groups. The loan tape does not contain information such as primary business language, perhaps, English is presumed to be the language. For these businesses, is language a barrier? And consequently, the timeliness of PPP loan information was compromised. The federal government should look into the finding, and if there had provided multi-lingual loan application brochures and associated services, so as to reduce racial inequality for business owners with Asian background. A recent discussion with a banking industry practitioner suggests that Asians tend to avoid taking loans/debts. Therefore, the government should consider avoiding using the word "loan" in the application or providing more explanations of the nature of the loans, including multi-lingual brochures.

According to the US Department of Labor statistical data, women's annual earnings were 82% of the men's in 2020. This research confirms that the gender inequality permeates to the PPP loan program given all other factors equal. For female applicants, both overall PPP loan sizes and per-employee loan sizes are the lower than male applicants. The paper therefore recommends that the federal government investigate the gender disparity.

For Veteran's applicants, both overall loan size and per-employee loan size are lower than the non-Veteran applicants. Similarly, for rural loans, both overall loan size and per-employee loan size are the lower than the urban loans. This disparity does not surprise the author, as it could be an adjustment of living and business costs. It could also be due to lack of timely information and convenient services.

Of the nearly 5,000 loan originators and/or servicers, the top five are associated with nearly 16% of the PPP loans while top 50 with nearly 50% of the loans. The paper finds that the PPP loan originators tend to keep these loans for servicing. If this group had been more diversified, it may also have provided more services in more geographic regions and to more blocks of lives, which could have helped with other issues such as reducing rural/urban disparity. As a recommendation for similar situations in the future, the authority should diversify the originators and servicers, to better serve the impacted business sectors and communities.

Due to data limitations, especially, large percentages of missing responses to gender, race, and veteran status information, the author recommends that the applications make these fields required ones such that these data items will be captured more precisely.

These findings identify issues in the PPP loans in terms of how best distributing the loans and allocating to the neediest. This paper provides some areas in which the policy makers and loan approval authority can investigate, learn from the program, and take measures to eliminate or at least reduce such statistically confirmed biases in the future or compensate more for the underrepresented social groups (e.g., increasing the portion of their forgiveness). In particular, these recommendations could provide guidance, should similar situations arise in the future

1 Statement of the Problem

With the outbreak of the COVID-19 pandemic in the US in February 2020, the US federal government-initiated Paycheck Protection Program (PPP) loans in April 2020 to help businesses stay afloat. Since then, some industries/sectors have been hit harder than others. These sectors include travel, airlines, hospitalities, retails, restaurants, and other small businesses. For example, the unemployment rate in leisure and hospitality shot up from 5.0% in December 2019 to 16.7% in December 2020, representing 1.3 million job losses. For the same period, the unemployment rates were 2.6% and 8.4% respectively for travel and transportation; for construction 5.0% and 9.6% respectively.[2]

According to data from US Small Business Administration (SBA), the total of initial approval was about $803bn in value with over 11 million loans. The average loan size is $68,200 while the median is $20,687, less than 1/3 of the average, indicating some loans are very large. Indeed, the largest loan amount is $10MM and there are 765 of such loans (Table 3-5). In terms of jobs impacted, the average business that received a loan has 7.67 employees while the median is 1, implying that most of these businesses have only one employee (Figure 3-2), although the largest firms have 500 employees. As a matter of fact, of the 765 loans in the amount of $10MM, 542 businesses have 500 employees. In terms of loan maturity, the average is 46.4 months (almost 4 years) while the median is 60 months (or 5 years). The maximum and minimum terms are 65 and 6 months respectively.

The paper intends to present the analytical work in understanding the impact, the effectiveness, and the contributing factors of PPP loans. These inquiries also include whether the loans went to the most needed geographical locations (such as by income level, by COVID cases), whether the loans went to the hard-hit sectors (such as hospitalities and restaurant), and whether there are disparities in loan distribution to certain groups such as veterans, certain ethnic groups, gender, urban/rural etc.

[2] Source: Occupations Hit Hardest in 2020 by the Pandemic (aarp.org) (accessed November 15, 2021).

Furthermore, the analytics and the statistical models developed in the paper should explain if and how these government rescue efforts worked during the difficult times. For example, are there biases over gender, race, urban/rural regions? did the impacted industries get their fair shares? In addition, this paper provides insight why we should not reach a conclusion by only focusing on one data attribute without understanding the full picture. The research shows that, when only one-dimension of data is used, one may reach one conclusion. However, by taking into account multiple dimensions in the data, the conclusion is different, sometimes, opposite. Regression models provide such a tool which helps identify the contribution of one factor while keeping everything else the same (e.g., the what-if analysis). More importantly, from the forward-looking perspective, the research can help the government evaluate the performance/effectiveness of the PPP loans and provide some helpful hints for future decision making through the power of data science and merits of econometrical/statistical modeling.

The author expects the research paper to identify the most important contributing factors in PPP loans by using publicly available data. The author hypothesizes that these factors include size of the business (which is proxied by number of employees if such balance sheet or income statement items are not available), geographic regions such as state, urban/rural, gender, race, veteran status, the relationship with distribution of COVID-19 cases.

2 Research Methodology: Data Sources, Analytics and Results

All data used in the Capstone research project is from public sources. For example. PPP loan data was downloaded from US Small Business Administration (SBA) website[3]. COVID-19 data was from Johns

[3] PPP loan data from US Small Business Administration (SBA) website: PPP FOIA - Dataset - U.S. Small Business Administration (SBA) | Open Data (accessed July 6, 2021).

Hopkins Coronavirus Resource Center[4], income data was from Internal Revenue Service website[5]. NAICS data was from North American Industry Classification System[6], and state population data from the census bureau[7].

After these data files were downloaded to local drivers in their original formats, respectively, including comma separated file format and Excel file format. There are 13 PPP loan files with a total of 4.81GB data. The next step was to understand the data through reviewing the data dictionaries. After that, the author reviewed data in the raw format (such as Excel, csv, text) and sampled some the data to get a first glimpse of the overall quality of the data such as missing data, obvious data errors, and/or embedded data errors.

After that, the author wrote R code to load these data files into RStudio, a computing and modeling environment. The 13 loan files were loaded and combined in to one R object using a loop. In the future, if more loan files are added, the code can handle them with ease without duplicate efforts. Then, the R object, called loan tape, is saved back to the local drive for data scrubbing, data analytics, and modeling. Other data files were also load into R objects so that they can be linked (or technically called joined).

With the data cleaned up, the author found that some key data items have data missing. For example, 81% of the entries for Race was unanswered, 63% of did not respond to Gender entry, and 71% left Veteran status blank. Since the missing fraction is huge, it will be biased to impute them by

[4] COVID case data: https://www.kff.org/state-category/covid-19/covid-19-metrics/, which is from COVID-19 Map - Johns Hopkins Coronavirus Resource Center (jhu.edu) (accessed August 25, 2021).

[5] Income data source is the 2018 zipped data for all states, including adjusted gross income (AGI): SOI Tax Stats - Individual Income Tax Statistics - 2018 ZIP Code Data (SOI) | Internal Revenue Service (irs.gov) (accessed July 7, 2021).

[6] NAICS 2017 industry data from North American Industry Classification System (NAICS) U.S. Census Bureau, including 2-6 digit 2017 NAICS code file in Excel format (accessed July 13, 2021).

[7] State Level Population Data: State Population Totals: 2010-2019 (census.gov) (accessed August 30, 2021).

filling with median or mean of numerical data and most frequent for the categorical data. Instead, the author treated the missing data as its own category for these data items.

3 Findings and Conclusions

In this chapter, the author scrubbed the data, did data analysis, and built statistical models based on the data collected in Chapter 2.

3.1 Data Scrubbing and Integration

Since there are over 11 million PPP loans, removing those with unintuitive data values did not have material impact. The scrubbing analyses were done for each data source.

In the PPP loan tape records/loans with negative, zero, or missing Jobs Reported were excluded. Also excluded include loans with zero term (70 records) as the loan terms are expected to be positive, loans with Zip code of 99999, loans with non-positive Initial Approval amount (Table 3-1).

Table 3-1 Examples of Data Scrubbing

Jobs Reported	Record Count	Actions
Negative	1	Excluded
Zero	209	Excluded
Missing	7	Excluded
Positive	11,768,472	Keep

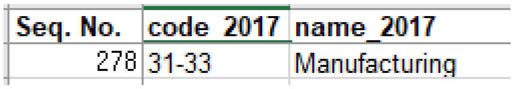
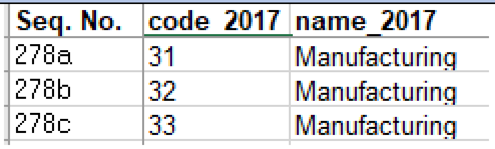
Adjusted Gross Income (AGI) data was downloaded from the IRS website. The purpose of this data is to understand how the PPP loans are represented in different income regions such as the state level (or as detailed as the zip code level). The vast majority of the AGIs are assigned with six AGI Stub numbers according to Table 3-2. AGI data outside its valid range defined in the table and AGI with zip code of 99999 were excluded.

Table 3-2 AGI Income Range for Stub

AGI Stub	**AGI (mnemonic A00100) Range**	**AGI Stub**	**AGI (mnemonic A00100) Range**
1	$1 under $25,000	4	$75,000 under $100,000
2	$25,000 under $50,000	5	$100,000 under $200,000
3	$50,000 under $75,000	6	$200,000 or more

The loan tape has business type to represent the business with which a loan is associated. It is less flexible than other industry classification methodology, e.g., Standard Industry Classification (SIC), North American Industry Classification System (NAICS), Global Industry Classification System (GICS). The reason NAICS was selected was that the code was provided as industry information in the PPP loan tape. The NAICS code is flexible such that it is between two-digit (i.e., most broad sector) to six-digit long (i.e., most granular industry or subindustry). The first two digits of the code are used to represent economic sectors. By joining 2-digit NAICS codes, industry names are retrieved and attached to the loan tape. In the raw NAICS data, group 31-33 is a single group. In order to join with the loan tape, it is separated into three separate rows (31, 32, and 33) as shown in Table 3-3.

Table 3-3 Two-Digit NAICS Code after Separating 31-33 Group into Three Groups

<table>
<tr><th>Raw NAICS Grouping</th><th>Enhanced NAICS Codes after Separation</th></tr>
<tr><td>
<table>
<tr><th>Seq. No.</th><th>code_2017</th><th>name_2017</th></tr>
<tr><td>278</td><td>31-33</td><td>Manufacturing</td></tr>
</table></td><td>
<table>
<tr><th>Seq. No.</th><th>code_2017</th><th>name_2017</th></tr>
<tr><td>278a</td><td>31</td><td>Manufacturing</td></tr>
<tr><td>278b</td><td>32</td><td>Manufacturing</td></tr>
<tr><td>278c</td><td>33</td><td>Manufacturing</td></tr>
</table></td></tr>
</table>

After data cleanup and scrubbing, loan data is summarized and provided in Table 3-4. There are over 11 million unique loans with an initial approval amount totaled $803bn. As of June 6, 2021, the current approval amount totaled $799 bn. In the report, the Initial Approval Amount was used to represent PPP loan amount because this data item is not expected to change over time. In addition, both draws (PPP and PPS) are combined as a single data set.

Table 3-4 Loan Summary after Cleanup

Processing Method/Draw/Batch	Loan Count	Initial Approval Amount ($bn)	Current Approval Amount ($bn)	Forgiveness Amount ($bn)
PPP	8,863,345	594	590	387
PPS	2,905,037	209	209	8
Total	11,768,382	803	799	395

Among all pieces of data described in Chapter 2, the center piece of the data assembly is the PPP loan data. All other pieces of data are integrated (or joined) into the loan tape as shown in Figure 3-1.

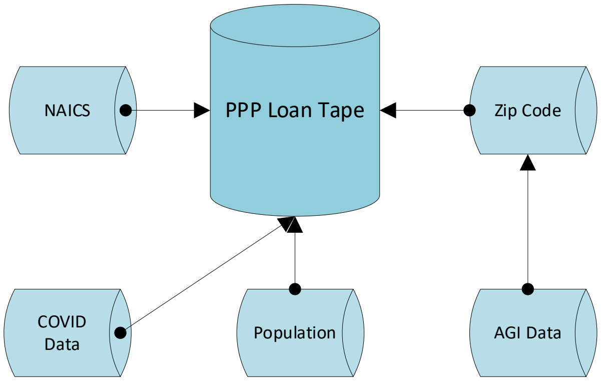


Figure 3-1 Data Integration – PPP Loan Tape as the Central Piece

3.2 Data Analysis and Research

The analysis sits not only at single-factor-level analytics (for example, certain groups seem favored), rather, the paper digs even further, such as two-factor-level analytics (this section) and multi-factor-level models (Section 3.3). These deep-dives explain the reasonableness of observations at multiple-factor-levels, e.g., these groups are favored because of the impact of second or third factor).

3.2.1 Analytical Tool

R/RStudio is selected as the analytical tool for data analysis and modeling. This is because, R has very powerful data analysis libraries, such as dplyr (for analytics), ggplot (for plotting), and also superior modeling packages (over e.g., Excel). Of course, automation is another key element. The R code for analytics and modeling can be easily rerun with the updated data or enhanced processes, and results are automatically stored when the code or underlying data changes.

3.2.2 PPP Portfolio Summary

Table 3-5 provides a high-level summary of the PPP loans. There are over 11 million PPP loans. The average loan size is $68,200 while the median is $20,687, less than 1/3 of the average, indicating some loans are very large. Indeed, the largest loan amount is $10MM and there are 765 of such loans. In terms of jobs impacted, the average loan has 7.67 employees while the median is 1, implying that most of the businesses associated with the loans have only one employee (Figure 3-2) although the largest firms have 500 employees. As a matter of fact, of the 765 loans in the amount of $10MM, 542 businesses have 500 employees. This is intuitive that, for government rescue purpose, the more employees a business has, the larger the loan size. In terms of loan maturity, the average is 46.4 months (nearly 4 years) while the median is 60 months (or 5 years). The maximum and minimum terms are 65 and 6 months respectively. Figure 3-2 shows that

- the maturities of the loans are clustered around 2 years and 5 years
- over half of the loans have maturities 5 years or longer

Table 3-5 PPP Loan Summary

	Average	**Median**	**Maximum**	**Minimum**
Initial Loan ($)	68,220	20,687	10,000,000	1
Jobs Reported	7.67	1	500	1
Term (month)	46.4	60	65	6

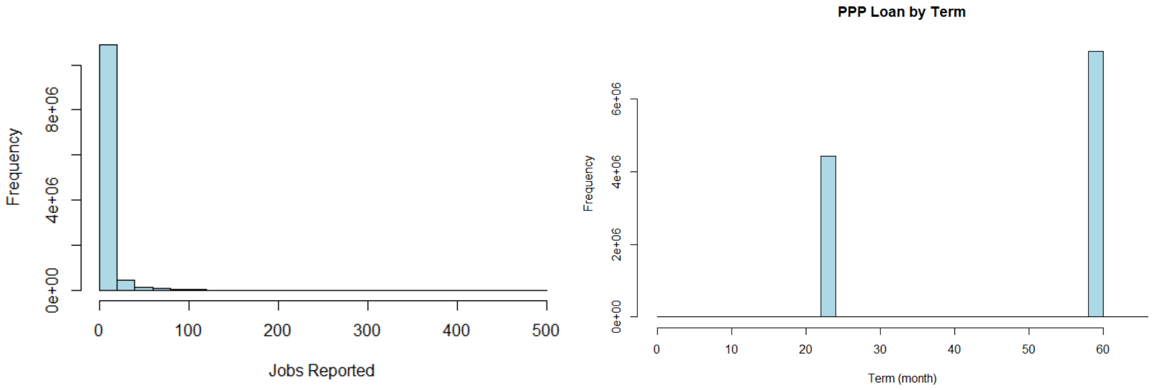


Figure 3-2 PPP Loan Distribution by Jobs Reported and Loan Term

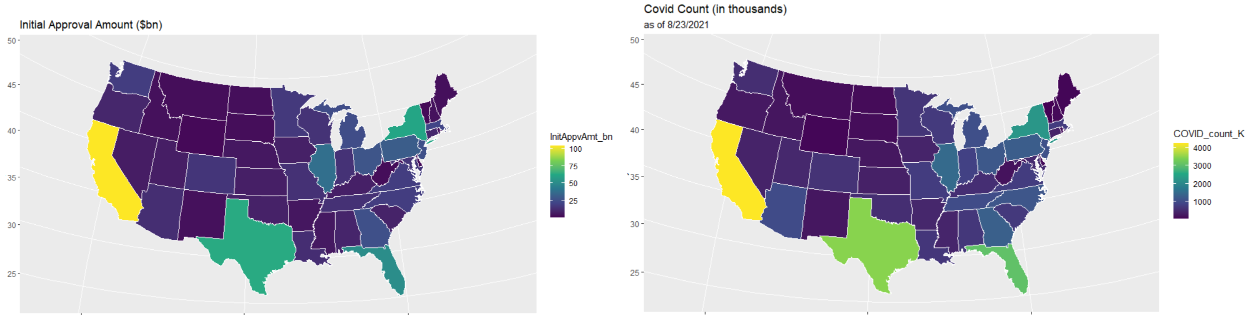
3.2.3 PPP Loans vs State-Level COVID-19 Data

The US COVID-19 cases data are downloaded from data source described in Chapter 2 as of August 23, 2021. The state-by-state COVID cases are provided in Table 3-6. The top five states with the most COVID cases are California, Texas, Florida, New York, and Illinois.

Table 3-6 Cumulative US COVID-19 Cases by State as of August 23, 2021

State	COVID Count	State	COVID Count	State	COVID Count
Alabama	665,653	Maine	73,660	Pennsylvania	1,274,337
Alaska	84,272	Maryland	487,893	Rhode Island	160,329
Arizona	988,714	Massachusetts	747,153	South Carolina	695,489
Arkansas	436,242	Michigan	1,044,958	South Dakota	128,626
California	4,247,528	Minnesota	635,222	Tennessee	997,479
Colorado	603,266	Mississippi	413,498	Texas	3,476,394
Connecticut	367,410	Missouri	745,930	USVI	5,638
Delaware	117,051	Montana	122,964	Utah	454,373
District of Columbia	53,898	Nebraska	237,492	Vermont	27,132
Florida	3,071,489	Nevada	381,766	Virginia	741,160
Georgia	1,328,156	New Hampshire	105,302	Washington	536,814
Hawaii	56,670	New Jersey	1,075,930	West Virginia	180,019
Idaho	214,010	New Mexico	225,994	Wisconsin	717,911
Illinois	1,491,582	New York	2,241,468	Wyoming	71,562
Indiana	825,549	North Carolina	1,161,818	American Samoa	-
Iowa	392,970	North Dakota	114,915	Guam	9,486
Kansas	357,177	Ohio	1,183,761	Nor. Mariana Is.	183
Kentucky	543,031	Oklahoma	530,594	Puerto Rico	164,759
Louisiana	660,804	Oregon	257,644	**Total**	**37,935,125**

Based the raw data, the initial approved PPP loans by states are summarized and the results are provided in Figure 3-3 (a). Similarly, using COVID-19 case data from Table 3-6, the COVID cases by state are plotted in Figure 3-3 (b). For consistency, both are plotted using the same color scale (i.e., legends), that means, in each respective data set, the maximum and minimum numbers, as well as the numbers in between, are presented by the same colors[8]. From visualization perspective, the two plots matching each other amazingly well, with the exceptions of TX, FL, and WA etc. Both TX and FL show lighter colors in COVID cases (the higher the number) as compared to PPP loan plot, implying higher COVID cases per loan amount, or inversely, lower loan amount per COVID case. The opposite color patterns are observed for WA. These observations are confirmed in Figure 3-4, in which both TX and FL (in red) have lower PPP loan amount than the national median line (dashed blue line) and WA (in green) has higher PPP loan amount than the national median.



(a) PPP Loans by State

(b) COVID Cases by State

Figure 3-3 PPP Loans and COVID-19 Cases by State

In general, PPP loans aligned reasonably well with the COVID cases at the state level. States with bars above the median line in Figure 3-4 have relative few cases per unit loan amount, or put it another way, these states have more loan amount per COVID case. These states include DC, HI, ME, OR, VT, and WA. Only three states are above the 2 standard deviations: DC, HI, and VT. The research also looked into the correlation between state's COVID rate (= COVID count/ population) and PPP loan fraction (=PPP loan amount/total PPP loan amount) and found the correlation coefficient is negligible (3.67%). This lack of relation at state level indicates that the PPP loans are distributed fairly across the states.

[8] Same color means the PPP loans are proportional to the COVID cases.

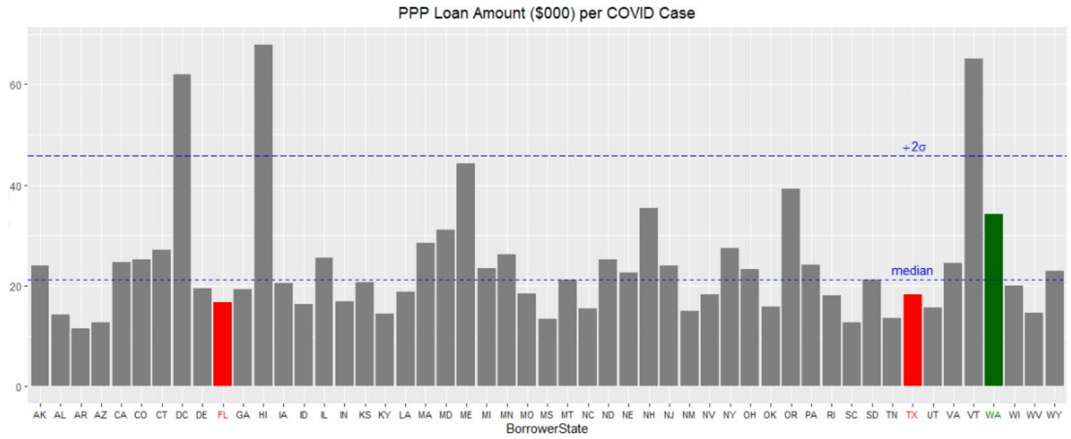


Figure 3-4 PPP Loan ($000) Per COVID Case by State[9]

3.2.4 Loan Statistics by Sectors

The loan tape provides the business type for a loan. However, the Business Type has some non-alphanumeric ASCII characters as shown in the first few rows in Table 5-2. As a result, the loan tape is joined with the NAICS data using two-digit NAICS code to represent industry sectors. The industry level loan statistics for the top 10 industries is provided in Table 3-7 and Figure 3-5.

Table 3-7 Top 10 Industries with PPP Loans

Rank	NAICS Code	NAICS Industry	Init. Approv. Amt ($bn)	% of Total	Cumul. % of Total
1	23	Construction	98.68	12%	12%
2	62	Health Care and Social Assistance	96.94	12%	24%
3	54	Professional, Scientific, and Technical Services	95.37	12%	36%
4	72	Accommodation and Food Services	83.95	10%	47%

[9] The distribution of the vertical axis skews to the right (i.e., fat right tail with skewness of 2.19. Zero skewness means a symmetric distribution).

5	81	Manufacturing	76.25	9%	56%
6	33	Other Services (except Public Administration)	58.87	7%	64%
7	44	Retail Trade	55.82	7%	70%
8	56	Administrative and Support and Waste Management and Remediation Services	39.60	5%	75%
9	42	Wholesale Trade	38.10	5%	80%
10	48	Transportation and Warehousing	33.20	4%	84%

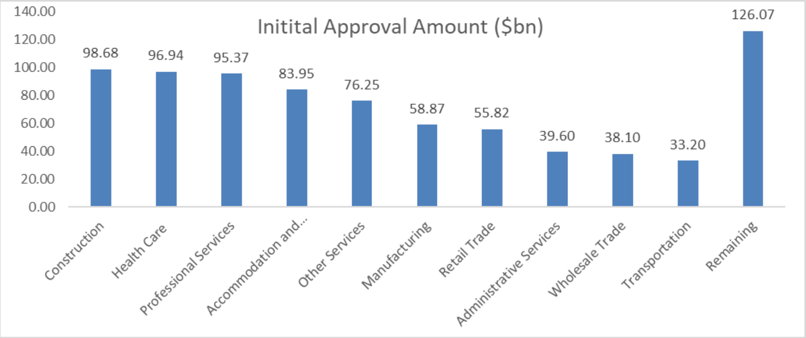


Figure 3-5 PPP Loans by Sector

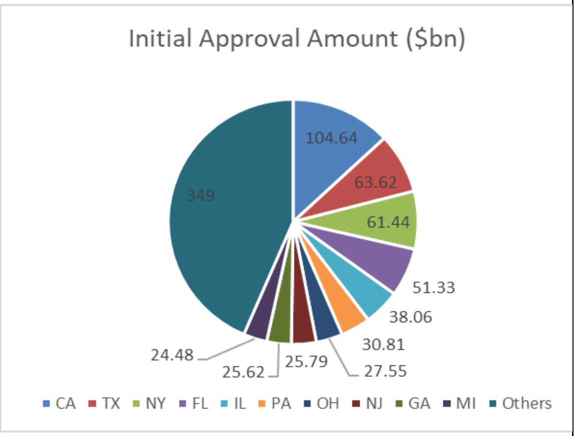
From Table 3-7, the Construction industry has the largest initially approved loan amount at $98.68bn (roughly 12% of all PPP loan amount). This is not surprising in that, after the outburst of COVID-19 in the US and shelter-in-home order, this industry was hit very hard, especially commercial real estate due to tremendous drops in hospitality and office space rental. Health Care and Social Assistance also hit hard such as hair salons, non-essential doctor's office businesses. Other familiar names include Accommodation and Food Services (such as the closure of restaurants), Retail Trade, Wholesale Trade, and Transportation (such as travels and leisure) etc.

In summary, this top 10 hard-hit industries make intuitive sense. Combinedly, they account for 84% of all PPP loans.

Loan Statistics by Top 10 States

Table 3-8 shows distribution of PPP loans by borrower's state. They are roughly proportional to the economic output with California, Texas, New York being the top three with $.10464bn (13%), $63.62bn (8%), $61.44bn (8%) of the total PPP loans respectively. The top 10 states account for 56% of the loans. The same data is also provided in the pie chart. In the downturn, they were hit hard at the state level.

Table 3-8 Top 10 States and Their Percentage

Borrower State	Loan Count	Initial Approval Amount ($bn)	Shares	Shares Pie
CA	1,302,235	104.64	13%	
TX	964,100	63.62	8%	
NY	755,582	61.44	8%	
FL	1,014,487	51.33	6%	
IL	650,446	38.06	5%	
PA	350,033	30.81	4%	
OH	363,682	27.55	3%	
NJ	307,827	25.79	3%	
GA	579,628	25.62	3%	
MI	302,087	24.48	3%	

3.2.5 Loan Statistics by Originators

There are two types of loan servicers in PPP loans: 4,891 Originating Lenders and 4,881 Service Lenders. However, not all servicers are the same. This is illustrated in Figure 3-6. The figure provides the percentages of total PPP loans from the top loan servicers. It is observed that

- the top 5 originators account for 15.5% of the PPP loans
- the top 10 account for 22.7%
- the top 50 account for 47.4%
- the top 100 account for 58.8%
- the originating lenders tend to keep servicing these PPP loans as indicted by the fact that the blue line (the originators) and the red line (the servicers) overlap very well. This may be because the forgiveness nature of the PPP loans backed by the federal government while earning essentially the return of loans.

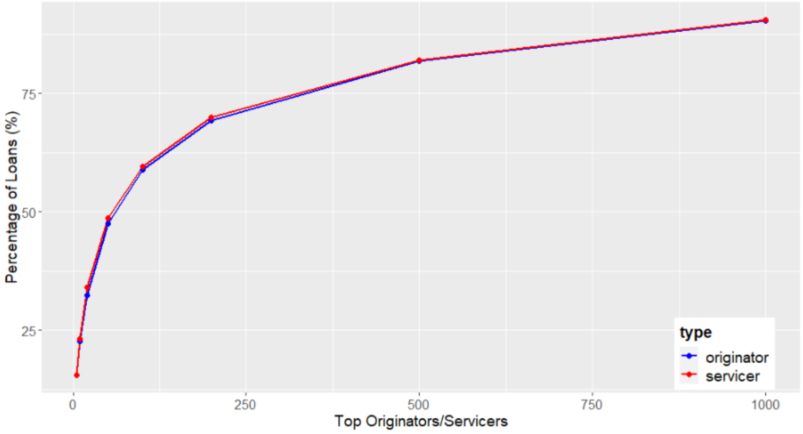


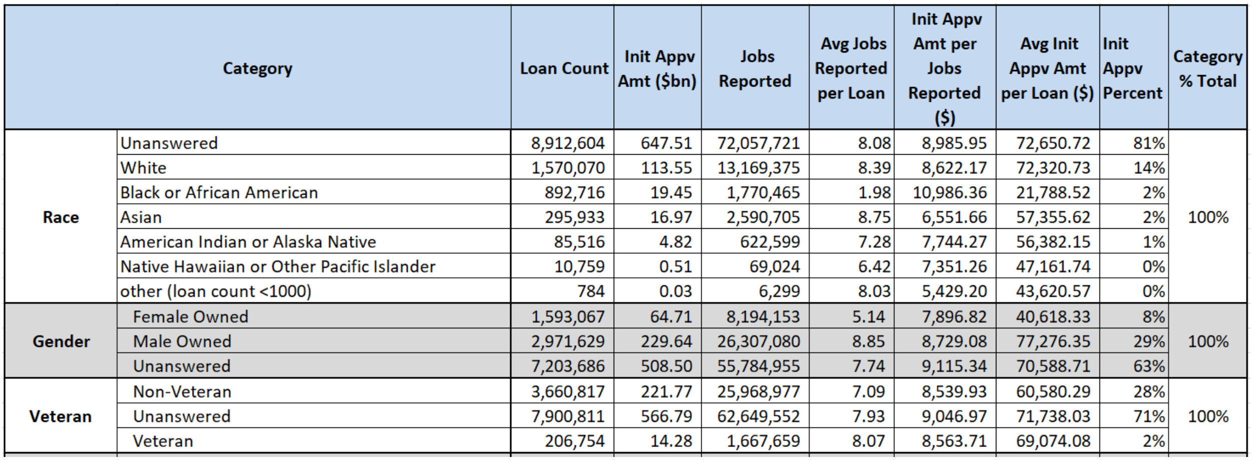
Figure 3-6 Percentage of Initial PPP Loans by Top Originators/Servicers

3.2.6 Loan Statistics by Other Dimensions

The PPP loan tape provides other characteristics for the loans such as race, gender, veteran status, non-profit status, rural/urban indicators etc. In this section, the paper analyzes the loans by these dimensions. The summary data is provided in Table 3-9 through Table 3-11 by category and the subcategory.

It should be noted that, for the analyses in this section, two dimensions in the data will be analyzed combinedly**.** For example, by comparing in one dimension, it may seem that the results are unintuitive (e.g., some groups seem favored while others not favored). However, when a second dimension is added, the results can be intuitive due to the explanatory power embedded in the second dimension. This is the power the detailed two-level analysis or the multiple-level modeling approach.

Table 3-9 PPP Loans by Race, Gender, and Veteran Status



Category		Loan Count	Init Appv Amt ($bn)	Jobs Reported	Avg Jobs Reported per Loan	Init Appv Amt per Jobs Reported ($)	Avg Init Appv Amt per Loan ($)	Init Appv Percent	Category % Total
Race	Unanswered	8,912,604	647.51	72,057,721	8.08	8,985.95	72,650.72	81%	100%
	White	1,570,070	113.55	13,169,375	8.39	8,622.17	72,320.73	14%	
	Black or African American	892,716	19.45	1,770,465	1.98	10,986.36	21,788.52	2%	
	Asian	295,933	16.97	2,590,705	8.75	6,551.66	57,355.62	2%	
	American Indian or Alaska Native	85,516	4.82	622,599	7.28	7,744.27	56,382.15	1%	
	Native Hawaiian or Other Pacific Islander	10,759	0.51	69,024	6.42	7,351.26	47,161.74	0%	
	other (loan count <1000)	784	0.03	6,299	8.03	5,429.20	43,620.57	0%	
Gender	Female Owned	1,593,067	64.71	8,194,153	5.14	7,896.82	40,618.33	8%	100%
	Male Owned	2,971,629	229.64	26,307,080	8.85	8,729.08	77,276.35	29%	
	Unanswered	7,203,686	508.50	55,784,955	7.74	9,115.34	70,588.71	63%	
Veteran	Non-Veteran	3,660,817	221.77	25,968,977	7.09	8,539.93	60,580.29	28%	100%
	Unanswered	7,900,811	566.79	62,649,552	7.93	9,046.97	71,738.03	71%	
	Veteran	206,754	14.28	1,667,659	8.07	8,563.71	69,074.08	2%	

Nine racial subcategories are provided for Race column. Race seems a sensitive data item in PPP loan application and does not seem a critical item in loan approval in that 81% of loans (in $ term) did not provide race information. Since this subcategory is the dominant majority, the missing data is treated as its own subcategory, rather than being proportionally filled from other subcategories or by other imputation methods. In addition, races with less than 1,000 loans (by count) are combined as the "Other" category. The category includes Puerto Rican, Multi Group, and Eskimo & Aleut. Besides, the remaining racial groups are White, Black, Asian, American Indian/Alaska Native, and Native Hawaiian.

Due to disparity in number of loans approved for each racial group, the total loan amounts differ substantially across the group. To analyze on the equal footing, the average loan size and the loan size per Jobs Reported (employee count) are chosen to compare across different racial groups. The analysis shows

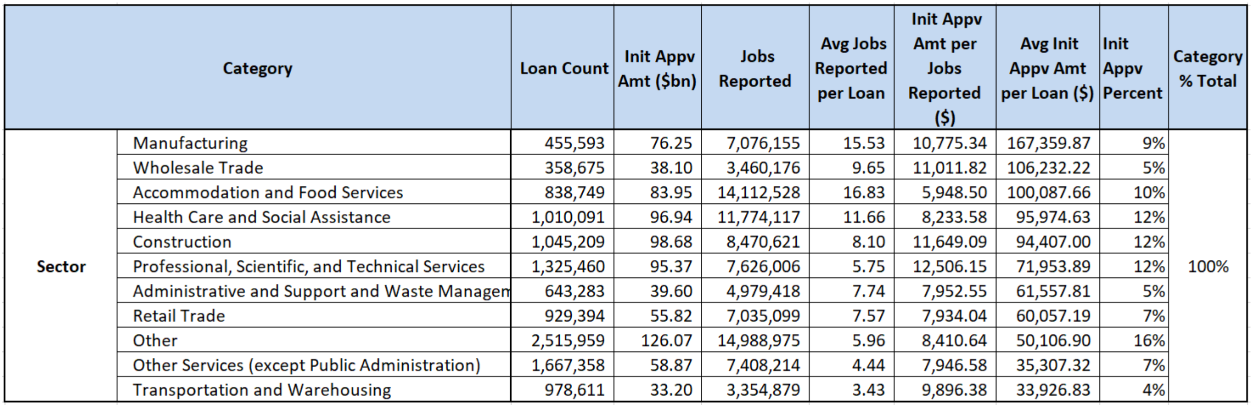
- Unanswered racial group and White have the largest average loan sizes at $72,651 and $72,321 each while the Black and Other have the lowest average loan sizes at $21,789 and $53,621 each. At first glance, this is shocking that the average loan size for the Black is less than 1/3 of that of White. Note that, overall, the average loan size is $68,220.
- An Asian business hired 8.75 employees on average, which is the highest among all groups, followed by White with 8.39 employees on average. Note that a Black business hired 1.98 employees on average, less than a quarter of that of Asian business or White. Note that, the employee count is 7.67 overall across all racial groups.
- In fact, Black has the highest loan amount per Jobs Reported among all racial groups at $10,986, followed by Unanswered group ($8,986) and White ($8,622). Asian businesses and Other have the lowest loan size per employee at $6,552 and $5,429 respectively. The Other business has only 784 loans (small sample). *Statistically, PPP loans for the Asian businesses are under-presented in the PPP loans on the per employee basis at 76% of White and 60% of Black*.

The dominant majority (63%) of loans (in $ term) did not provide gender information. Maybe some businesses were co-owned by both genders. Due to its dominance, any imputation may spoil the nature of the data, the missing data is kept as its own subcategory, rather than fill it proportionally from other subcategories or by other imputation methods. For those loans that did provide the information, male accounts for 29% of the PPP loans while female 8%. The average loan size for female is $40,618, which is about 52.5% that of a male business owner. At first glance, an obvious question is “does this imply the gender gap?” Next, let us take into account the number of employees impacted by the business. An average male business hired 8.85 employees while a female business 5.14. As a result, the average loan sizes per employee are $8,729 and $7,897 for male and female business owners respectively. *However, by controlling the number of employees, the gender gap still exists by about 10%.*

The dominant majority (71%) of loans (in $ term) did not provide veteran information. Due to this dominance, any imputation may spoil the nature of the data, the missing data is kept as its own subcategory, rather than fill it proportionally from other subcategories or by other imputation methods. For the data missing group, the average loan size is the highest ($71,738) and so is the average loan per employee ($9,047). *Veteran businesses have an average loan size of $69,074, which is 14% higher than those of non-Veteran’s ($60,580).*

There 21 NAICS sectors identified in the loan tape. For data analysis purpose, the largest 10 sectors are selected and the remaining sectors are combined into the “Other” sector. The sectors with the top 5 average loan size are Manufacturing, Wholesale, Accommodation and Food, Health Care, and Construction. The average employees are from 8.10 to 16.83. *After adjusting for the employee head count, the top loans are Professional Services ($12,506), Construction, Wholesale, Manufacturing, and Transportation.*

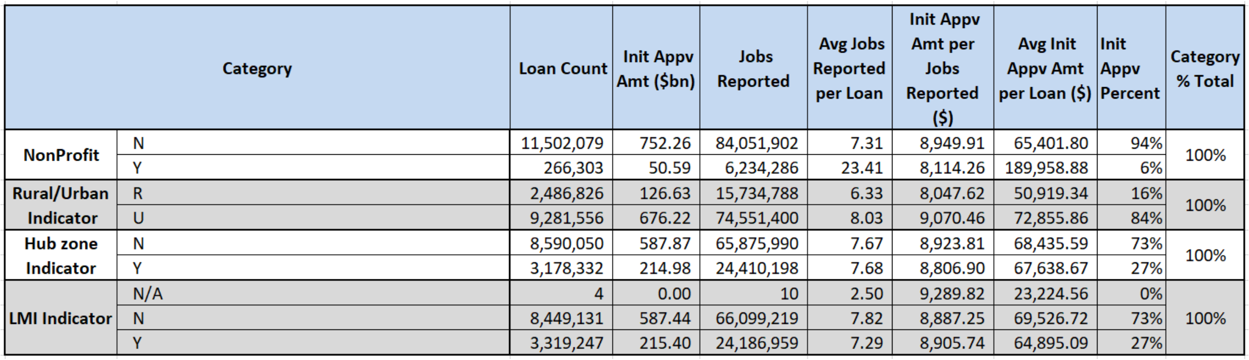
Table 3-10 PPP Loans by Sector



Category		Loan Count	Init Appv Amt ($bn)	Jobs Reported	Avg Jobs Reported per Loan	Init Appv Amt per Jobs Reported ($)	Avg Init Appv Amt per Loan ($)	Init Appv Percent	Category % Total
Sector	Manufacturing	455,593	76.25	7,076,155	15.53	10,775.34	167,359.87	9%	100%
	Wholesale Trade	358,675	38.10	3,460,176	9.65	11,011.82	106,232.22	5%	
	Accommodation and Food Services	838,749	83.95	14,112,528	16.83	5,948.50	100,087.66	10%	
	Health Care and Social Assistance	1,010,091	96.94	11,774,117	11.66	8,233.58	95,974.63	12%	
	Construction	1,045,209	98.68	8,470,621	8.10	11,649.09	94,407.00	12%	
	Professional, Scientific, and Technical Services	1,325,460	95.37	7,626,006	5.75	12,506.15	71,953.89	12%	
	Administrative and Support and Waste Managem	643,283	39.60	4,979,418	7.74	7,952.55	61,557.81	5%	
	Retail Trade	929,394	55.82	7,035,099	7.57	7,934.04	60,057.19	7%	
	Other	2,515,959	126.07	14,988,975	5.96	8,410.64	50,106.90	16%	
	Other Services (except Public Administration)	1,667,358	58.87	7,408,214	4.44	7,946.58	35,307.32	7%	
	Transportation and Warehousing	978,611	33.20	3,354,879	3.43	9,896.38	33,926.83	4%	

94% of the loans are from for-profit organizations while 6% are from non-profit organizations. The average loans size is $189,959 for non-profit, which is 2.9x as high as the for-profit ($65,401). However, non-profit hired 23.41 employees as compared with only 7.31 employees for the for-profit. *On the per-employee basis, the average non-profit loan is $8,114, which slightly (9%) smaller than that of for-profit ($8,950)*.

Table 3-11 PPP Loans by R/U, HUB Zone, and LMI



Category		Loan Count	Init Appv Amt ($bn)	Jobs Reported	Avg Jobs Reported per Loan	Init Appv Amt per Jobs Reported ($)	Avg Init Appv Amt per Loan ($)	Init Appv Percent	Category % Total
NonProfit	N	11,502,079	752.26	84,051,902	7.31	8,949.91	65,401.80	94%	100%
	Y	266,303	50.59	6,234,286	23.41	8,114.26	189,958.88	6%	
Rural/Urban Indicator	R	2,486,826	126.63	15,734,788	6.33	8,047.62	50,919.34	16%	100%
	U	9,281,556	676.22	74,551,400	8.03	9,070.46	72,855.86	84%	
Hub zone Indicator	N	8,590,050	587.87	65,875,990	7.67	8,923.81	68,435.59	73%	100%
	Y	3,178,332	214.98	24,410,198	7.68	8,806.90	67,638.67	27%	
LMI Indicator	N/A	4	0.00	10	2.50	9,289.82	23,224.56	0%	100%
	N	8,449,131	587.44	66,099,219	7.82	8,887.25	69,526.72	73%	
	Y	3,319,247	215.40	24,186,959	7.29	8,905.74	64,895.09	27%	

The Urban loans dominate the Rural loans by a ratio of 84:16 (≈5:1). Average Urban loan is $72,856, which is 43% higher Rural Loan ($50,919). The average employee count is 8.03 for Urban loan and 6.33 for Rural loan. Even after this is taken into account, *the average loan size per employee is $9,070 for an Urban business, which is 12.7% higher than a Rural business ($8,048), indicating some sort of disparity*.

The HUBZone program fuels small business growth in historically underutilized business (HUB) zones with a goal of awarding at least three percent of federal contract dollars to HUBZone-certified companies each year. These loans are identified by the HUB Zone Indicator in the PPP loan tape. Non-HUB Zone loans dominates HUB Zone loans by 73% vs 27% margin, roughly 3:1 ratio. The ratio exceeds the 3+% federal contract dollars threshold for the HUB Zones*. In terms of average employees hired, average loan size, and average loan per employee, HUB loans and non-HUB loans are very comparable*.

LMI indictor refers to loans associated with the Low- and Moderate-Income (LMI) communities. Non-LMI loans dominate over LMI loans by 73% vs 27% margin, roughly 3:1 ratio. There exist only 4 loans without the LMI indicator. Thus, their impact is negligible. The average loan size for non-LMI loans is $69,527 while that for the LMI loans is $64,895. *After adjusting for the employee head count, the two types of loans are comparable on the per employee basis.*

3.3 Model Development and Results

3.3.1 Methodology

The purpose of developing models is to identify combinations of independent variables that can explain the variation observed in the dependent variable, e.g., PPP loans. Compared with data analytics discussed in Section 3.2, the models provide more in-depth insight of the dynamics in the loan tapes, including the interaction with other factors and can be used to do what if analysis, sensitivity analysis. In this paper, the initial approval amount was of interest and therefore selected as the dependent variable to be modeled. Since it is a positive continuous variable, the Ordinary Least Square (OLS) Regression and the natural log of OLS are selected. The code for analytics and modeling can be easily rerun and results automatically stored for reviewing when the code or underlying data changes. The paper was intended to figure out what factors (aka the independent variables) impact the PPP loan size (aka the dependent variable) .

The analysis includes three distinct steps (1) the Univariate analysis (UVA) by using one factor (2) the Multivariate analysis (MVA) by using multiple factors, and (3) final model selection.

This chapter provides the most important results. For more detailed process, the readers are advised to reference the relevant sections in Chapter 5.

3.3.2 Modeling Assumptions

The author made these assumptions for model development purpose.

- Loan tapes from the first draw (PPP) and second draw (PPS) are combined and used.
- Initial Approval Amount is used as it is expected to be static over time, rather than Current Approval Amount which changes over time
- Consolidation of Race data for those with less than 1000 loans each as Other group
- Due to very large volumes of missing data/not-answered data both in terms of absolute counts and relative percentage in each category, they are treated as its own groups, instead of imputing such data

3.3.3 UVA

UVA is a modeling step that evaluate the explanatory power of a single independent variable on the dependent variable in model development. In OLS, the goodness of model fit is characterized by adjusted R^2 and the factor significance is determined by t-stat (>2.0) or p-value (<5%). In the paper, the dependent variables are JobsReported, Race, Gender, Veteran, non-Profit status, Sector etc. The results are provided in Table 5-3. The first six columns are directly pulled from the fitted regression models while the Explanations column summarizes the conclusions. From the table, there observations are made

- Almost all factors are statistically significant except LMI Indicator
- JobsReported is the most powerful explanatory variable in that it has the highest Adj. R^2 of 62.80%

- In terms of average loan size, this table is consistent with Table 3-9. Without loss of generality, the author used Gender as an example. Similar analysis can be done for other variables. From Table 5-3, GenderFemale is selected as the reference state. A reference state is a base line state that other values of the variable is compared with.
    - the intercept is 40618.33, matching its average Female loan size in Table 3-9
    - the slope of GenderMale Owned, a category variable, is 36658.02, meaning the average loan for Male applicant is 36658.02 more than that of Female (the reference), so the average loan for Male applicant is $77,276.35, matching its average Male loan size in Table 3-9
    - the slope of GenderUnanswered, a category variable, is 29970.39, meaning the average loan for GenderUnanswered applicant is 29970.39 more than that of Female, so the average loan for GenderUnanswered applicant is $70,588.71, matching that in Table 3-9

3.3.4 MVA

MVA is a modeling step that builds models with a combination of multiple independent variables. Sometimes, a variable with little explanatory power or statistical significance in UVA is kept in the model as they may have sizeable predictive power when combined with other factors. In this paper, OLS is used. The goodness of model fit is characterized by adjusted R^2 and the factor significance is determined by t-stat (>2.0) or p-value (<5%). In addition, for MVA, multicollinearity needs be checked though Variance Inflation Factor (VIF<5.0) or Generalized Variance Inflation Factor (GVIF)[10]. High level of multicollinearity can cause the factors unstable in both the statistical significance and/or the unintuitive signs of the coefficients, although it does not bias the projection. Two models are developed, one for

[10] When categorical variables are involved, GVIF is calculated otherwise VIF is used. More involved discussions on GVIF can be found at stackexchange.com (accessed August 23, 2021).

loan size and the other for loan size per employee. The final results are discussed below. For detailed development and thought processes, please refer to Appendix 5.4.

Modeling Loan Size With variables from the loan tape, sector and average AGI, the final model specification is provided in Table 5-10. Except Race2Asian, Race2Other, all other variables are statistically significant. The positive coefficient for JobsReported (8466) indicates that, for every additional employee, a loan is expected to increase by $8,466. The positive coefficient for avg Household AGI (0.06406) indicates that, for every $1 in AGI, the loan is expected to increase by $0.064, implying that for higher income region, the loan is expected to be higher, given everything else equal. Given everything else the same, the loan sizes from Black, Unanswered race, White, Other, Native Hawaiian/Pacific Islander are above the reference race (American India or Alaskan). A loan for a black applicant is $11,630 more than that of the reference. Only a loan from an Asian applicant is $338 less than that of the reference, indicating some racial disparity. In gender dimension, a loan from a Male applicant is $1,763 more than that of a female, indicating some degree of gender disparity unfavorable to Female. A Veteran's loan is $1,604 less than that of a non-Veteran, which is not intuitive at the first glance. However, this is confirmed by study done by the Federal Reserve Bank of New York[11]. In the paper, the authors concluded that the "veteran entrepreneurship is facing a generational decline" and their businesses "face greater difficulty in accessing capital relative to nonveteran-owned businesses". An urban loan is $4,510 more than a rural loan. A loan in HUB zone is $1,088 more than otherwise. The model selects Accommodation and Food Service sector as the reference sector. The top three loan sizes are Manufacturing, Construction, and Wholesale Trade, respectively $46,980, $39,080, and $37,570

1. [11] Federal Reserve Bank of New York, 2017-report-on-veteran-entrepreneurs-and-capital-access.pdf (newyorkfed.org) (accessed January 15, 2022).

higher than the reference. The results for the Sectors are intuitive in that these are the COVID-impacted sectors.

The model has a higher Adj. R^2 of 63.28%. The maximum Generalized Variance Inflation Factor (GVIF) is 4.01, which is below the threshold of multicollinearity (5.0) , implying that no obvious signs of multicollinearity exist in the independent variables. When the degrees of freedom are taken into account, the maximum factor is 1.42.

Modeling Loan Size Per Employee (log-linear) This model is similar to the first model except the model specification is based on the natural log of the dependent variable or ln(InitAppvAmtPerEmployee). The reason log-linear model is developed is to mathematically guarantee positive projected InitAppvAmtPerEmployee from the model. The model specification is provided in Table 5-11. Except Race2Other, all other variables are statistically significant. The negative coefficient for JobsReported (-0.003327) indicates that, for every additional employee, the natural log of loan per employee is expected to **decrease** by 0.003327 or the loan per employee is expected to reduce 0.33% (=$1 - e^{-0.003317}$) for each additional employee. _The positive coefficient for avg Household AGI (3.67*10^{-7}) indicates that for, higher income region, even the loan per employee is expected to be higher, given everything else equal._ The model has a slightly higher Adj. R^2 of 5.03%, almost doubling that in the model provided in Table 5-8. The maximum Generalized Variance Inflation Factor (GVIF)[12] is 4.70, which is below the threshold of multicollinearity (5.0) , implying that no obvious signs of multicollinearity exist in the independent variable. When the degrees of freedom are taken into account, the maximum factor is 1.47.

[12] When categorical variables are involved, GVIF is calculated. More discussions can be found at stackexchange.com (accessed August 23, 2021).

3.3.5 Final Model Selection

After studying the pros and cons of each model, the final models are selected as follows

- for Expected Loan Size Model, the model specified in Table 5-10
- for Expected Loan per Employee Model, the model specified in Table 5-11

3.3.6 Findings

The research analyzes PPP loan data from the perspectives of three levels of complexity with the increasing complexity and explanatory power.

- Analysis by Single Factor (e.g., Gender alone) – it analyses the impact of a single factor on the PPP loans with the explanatory powers of all other factors buried. This analysis is the most straightforward, yet the results can be deceiving
- Analysis by Two Factors (e.g., Gender and Number of Employees) – it analyses the PPP loans by two dimensions. Therefore, it is more comprehensive than Analysis by Single Factor and it can uncover facts from two different perspectives.
- Models with Multiple Factors (such as Gender, Race, Number of Employees, AGI, Sector, etc.) – they are based on mathematical models considering all statistically significant factors. It can do marginal analysis of each factor such as loan difference between Male and Female business owners and what-if analysis.

Implications from the Analyses

In general, PPP loans aligned reasonably well with the COVID-19 cases at the state level, indicating that the loans went to the states where financial help is most needed. For TX and FL, they have higher COVID cases per unit loan amount. On the other hand, states such DC, HI, ME, OR, VT, and WA have relative few cases per unit loan amount.

Overall, the average loan size is $68,220. Unanswered racial group and White have the largest average loan sizes at $72,000 while the Black has the lowest average loan sizes at $21,789 each which is less than 1/3 of that of White. Note that on average, an Asian business hired 8.75 employees, the highest among all races, followed by White with 8.39. The Black business had 1.98. Statistically, a Black business has the highest loan amount per employee at $10,986, followed by Unanswered group ($8,986) and White ($8,622). Asian businesses and Other have the lowest loan size per employee at $6,552 and $5,429 respectively. The Other business has only 784 loans (small sample size). Statistically, PPP loans for the Asian businesses are under-presented in the PPP loans on the per employee basis at 76% of White and 60% of Black.

Of all the PPP loans, Male accounts for 29% while Female 8%. The remaining loans did not specify the gender. The average loan size for Female is $40,618, which is about 52.5% that of a Male business. When the number of employees is taken into account, the average loan sizes per employee are $8,729 and $7,897 for Male and Female business owners respectively, representing a gender gap of about 10%.

In terms of impacted sectors, the Construction industry has the largest initially approved loan amount at $98.68bn (roughly 12% of all PPP loan amount). After outburst of COVID-19 in the US and shelter-in-home order, this industry was hit very hard, especially commercial real estate such as hospitality/hotels and office rentals. Health Care and Social Assistance also hit hard with the shutdown of business such as hair salons, non-essential doctor’s office businesses. Other familiar names include Accommodation and Food Services (such as the closure of restaurants), Retail Trade, Wholesale Trade, and Transportation (such as travels and leisure) etc. In summary, this top 10 list of industries combinedly account for 84% of all PPP loans.

The distribution of PPP loans by borrower's state is roughly proportional to the economic activities and economic output with California, Texas, New York being the top three with 13% ($104.64bn), 8%, 8% of the total PPP loans respectively.

Veteran's businesses have an average loan size of $69,074, which is 14% higher than those of non-Veteran's ($60,580).

94% of the loans are from for-profit organizations while 6% are from non-profit organizations. The average loans size is $189,959 for non-profit, which is 2.9x as high as the for-profit ($65,401). However, non-profit hired 23.41 employees as compared with only 7.31 employees for the for-profit. On the per-employee basis, the average non-profit loan is $8,114, which is about 10% less than that of for-profit ($8,950).

The average loan size per employee is $9,070 for an Urban business, which is 12.7% higher than a Rural business ($8,048), indicating some sort of disparity.

Implications of Loan Models

PPP Loan UVA models (Table 5-3) generates the same conclusions as the data analytics (Table 3-9). Based on the factors identified significant in UVA, the paper builds two multiple factor models (multivariate models). The model factors are all found to be statistically significant. One for explaining the PPP loan while the other for a PPP loan per employee.

- For numerical factors such as Jobs Reported and Average AGI, the sign of a coefficient represents the directionality of the factor's impact on the loan while its absolute value represents the magnitude or sensitivity of such impact. For examples, (1) the coefficient of Jobs Reported (8466) indicates that, for every additional employee, the loan is expected to increase by $8,466. (2) the positive coefficient for avg Household AGI (0.06406) indicates that, for every $1 in AGI, the loan is expected to increase by $0.064, implying that for higher income region, the loan is expected to be higher, given everything else equal.

- For categorical variables, the coefficient represents the impact relative to the reference state of the factor.

It should be emphasized that, the MVA model offers multi-faceted explanation of the PPP loans as compared to data analytics, since the former quantifies the observations not only by one or two dimensions used by the data analytics, but also other factors (e.g., third, fourth factors etc). For example, the average PPP loan for non-profit is $189,959 in Table 3-9 which is *a lot more than* that of a for-profit loan ($65,402). However, the difference between the two ($124,557) is not the impact purely from this variable. Indeed, it also includes the impact from different number of employers, different industries, different Urban/Rural areas. On the other hand, the model provides a much better dissection of the factor's contribution by controlling other factors. In this model in Table 5-10, when all other factors are the same, a non-profit loan is $5,568 *less than* that of a for-profit loan. A similar conclusion can be reached by controlling number of employees, e.g., when comparing both loans on a per-employee basis (Table 3-9). Therefore, it would be misleading to blindly conclude that a loan for non-profit business is almost three times as high as that of a for-profit business without simultaneously looking into other variables. These variables include number of employees which, in this case, plays a more significant role. Of the nearly 5,000 loan originators/servicers, the top 50 account for nearly 50% of the loans and the top 100 nearly 60%. And the originators tend to keep these loans for servicing.

4 Recommendations

In the analysis above, the author studied PPP loans along different dimensions such as HUB Zone, LMI loan types, Racial gap, Gender gap, Rural/Urban gap, etc. The discussions below are controlled results in that, it neutralizes the impact of other factors, i.e., given everything else equal.

The overall PPP loan sizes favored geographical areas with higher incomes. For example, in areas with higher average Household AGI, the overall PPP loan sizes are bigger. This is not intuitive. What is even more unintuitive is that the loan sizes per employee also favored higher income regions. The loans were expected to be tilted toward low-income area. Is it because the loan applicants' expectations for the lower income areas were lower, e.g., business costs were lower, therefore, they applied for smaller loan sizes, or, is it because the applicants were less informative? The author suggests that the federal government investigate the PPP loans' disparity with expectations for regions with different AGIs. For example, study the applied loan amount vs. the actual approved loan amount.

For Asian American applicants, both overall loan sizes and per-employee loan sizes are the lowest among all major racial groups. The loan tape does not contain information such as primary business language, perhaps, English is presumed to be the language. For these businesses, is language a barrier? And consequently, the timeliness of PPP loan information was compromised. The federal government should look into the findings, and if there had provided loan application brochures in multiple languages and associated services, so as to reduce racial inequality for business owners with Asian background. A recent discussion with a banking industry practitioner suggests that Asians tend to avoid taking loans/debts. Therefore, the government should consider avoiding using the word "loan" in the application or providing more explanations of the nature of the loans, including multi-lingual brochures.

According to the US Department of Labor statistical data in reference 9, women's annual earnings were 82% of that of the men's in 2020. This research confirms that the gender inequality permeates to the PPP loan program given all other factors equal. For female applicants, both overall PPP loan sizes and per-employee loan sizes are the lower than male applicants. The paper therefore recommends that the federal government investigate the gender disparity.

For Veteran's applicants, both overall loan size and per-employee loan size are the lower than the non-Veteran applicants. This is consistent with recent findings on veteran entrepreneurs[13].

Similarly, for rural loans, both overall loan size and per-employee loan size are the lower than the urban loans. This disparity does not surprise the author, as it could be an adjustment of business costs and living expenses. It could also be due to lack of timely information and convenient services.

Of the nearly 5,000 loan originators and/or servicers, the top 5 are associated with nearly 16% of the PPP loans while top 50 with nearly 50% of the loans. The paper finds that the originators tend to keep these loans for servicing. If this group had been more diversified, it may also have provided more services in more geographic regions and to more blocks of life, which led to helping with other issues such as reducing rural/urban disparity. As a recommendation for similar situations in the future, the authority should diversify the originators and servicers, to better serve the impacted sectors and communities.

Due to data limitations, especially, large percentages of missing responses of gender, race, and veteran status information, the author recommends that the applications make these fields compulsive ones. If this is done, similar programs in the future will have more accurate and complete data. As a result, research conducted on these programs will be more accurate to the finest combinations.

In term of research methodology, the author strongly recommends that Analysis by Two Factors or Models with Multiple Factors over Analysis by Single Factor. Furthermore, the order of preference would be favoring Models with Multiple Factors over Analysis by Two Factors, which is over Analysis by

[13] In 2017, an NY Fed paper, report-on-veteran-entrepreneurs-and-capital-access.pdf (newyorkfed.org), found that rate of veteran entrepreneurship was declining (accessed January 10, 2022).

Single Factor. This is because, the more complex methodologies will provide more comprehensive, more unbiased, and more fact-finding analyses by taking into account multiple factors simultaneously.

These findings identify issues in the PPP program in terms of how best distribute the loans and allocate them to the neediest. This paper provides some areas in which the policy makers and loan approval authority can investigate, learn from the program and take measures to reduce, if not eliminate, such statistically confirmed biases in the future, or compensate more for the underrepresented social groups (e.g., increasing the portion of their forgiveness). In particular, these recommendations could provide guidance, should similar crises arise in the future.

Even though the paper lists a handful of recommendations, the research concludes that, the PPP loans benefited the most impacted sectors such as construction, health care and social assistance, accommodation and food services, manufacturing, and retail trade. These sectors were hit the hardest by COVID-19. In addition, the PPP loan distribution was aligned with the COVID cases at the state level as of the time the author pulled the data. It helped 11+ millions of businesses with the much-needed liquidity to pay the expenses of running the business and/or to retain the employees.

5 Appendix

5.1 Limitation of Hardware

When the statistical models were built, the computer ran out of memory (Table 5-1). So non-critical data items in the dataset were pruned before the models were built.

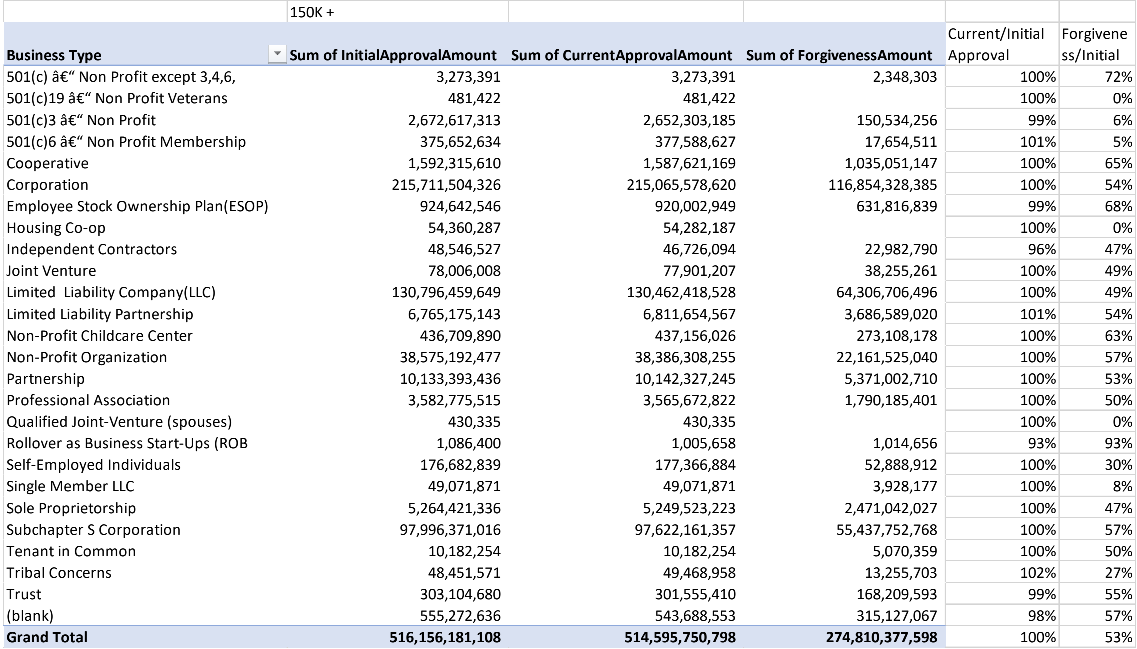
Table 5-1 Computer Ran out of Memory with All Data

```
> mdl2P = lm(InitialApprovalAmount ~ avg_HH_AGI, data = all_data_with_AGI_PPPmatchProject)

Error: cannot allocate vector of size 64.0 Mb

Error: no more error handlers available (recursive errors?); invoking 'abort' restart

> summary(mdl2P)

Error in summary(mdl2P) : object 'mdl2P' not found
```

5.2 PPP Loans by Business Type

The loan tape provides the business type for a loan. However, the Business Type has some non-alphanumeric ASCII characters as shown in the first few rows in Table 5-2 and does not mapped to sectors or industries.

Table 5-2 PPP Loan Statistics – by Business Type



	150K +				
Business Type	**Sum of InitialApprovalAmount**	**Sum of CurrentApprovalAmount**	**Sum of ForgivenessAmount**	Current/Initial Approval	Forgiveness/Initial
501(c) â€“ Non Profit except 3,4,6,	3,273,391	3,273,391	2,348,303	100%	72%
501(c)19 â€“ Non Profit Veterans	481,422	481,422		100%	0%
501(c)3 â€“ Non Profit	2,672,617,313	2,652,303,185	150,534,256	99%	6%
501(c)6 â€“ Non Profit Membership	375,652,634	377,588,627	17,654,511	101%	5%
Cooperative	1,592,315,610	1,587,621,169	1,035,051,147	100%	65%
Corporation	215,711,504,326	215,065,578,620	116,854,328,385	100%	54%
Employee Stock Ownership Plan(ESOP)	924,642,546	920,002,949	631,816,839	99%	68%
Housing Co-op	54,360,287	54,282,187		100%	0%
Independent Contractors	48,546,527	46,726,094	22,982,790	96%	47%
Joint Venture	78,006,008	77,901,207	38,255,261	100%	49%
Limited Liability Company(LLC)	130,796,459,649	130,462,418,528	64,306,706,496	100%	49%
Limited Liability Partnership	6,765,175,143	6,811,654,567	3,686,589,020	101%	54%
Non-Profit Childcare Center	436,709,890	437,156,026	273,108,178	100%	63%
Non-Profit Organization	38,575,192,477	38,386,308,255	22,161,525,040	100%	57%
Partnership	10,133,393,436	10,142,327,245	5,371,002,710	100%	53%
Professional Association	3,582,775,515	3,565,672,822	1,790,185,401	100%	50%
Qualified Joint-Venture (spouses)	430,335	430,335		100%	0%
Rollover as Business Start-Ups (ROB	1,086,400	1,005,658	1,014,656	93%	93%
Self-Employed Individuals	176,682,839	177,366,884	52,888,912	100%	30%
Single Member LLC	49,071,871	49,071,871	3,928,177	100%	8%
Sole Proprietorship	5,264,421,336	5,249,523,223	2,471,042,027	100%	47%
Subchapter S Corporation	97,996,371,016	97,622,161,357	55,437,752,768	100%	57%
Tenant in Common	10,182,254	10,182,254	5,070,359	100%	50%
Tribal Concerns	48,451,571	49,468,958	13,255,703	102%	27%
Trust	303,104,680	301,555,410	168,209,593	99%	55%
(blank)	555,272,636	543,688,553	315,127,067	98%	57%
Grand Total	**516,156,181,108**	**514,595,750,798**	**274,810,377,598**	100%	53%

5.3 Result Summary for Univariate Analysis (UVA)

The UVA results are provided in Table 5-3.

5.4 Detailed Process for Multivariate analysis (MVA)

MVA is a modeling step that builds models with a combination of multiple independent variables. Sometimes, variables with little explanatory power or statistical significance in UVA can be kept in the model as they may have sizeable predictive power when combined with other factors.

5.4.1 MVA Model on Initial Approval Amount

Loan Model with Factors from Loan Tape Only

When only variables from the loan tape are used, the model specification is provided in Table 5-4. Except Race2Native Hawaiian and LMI factors, all other variables are statistically significant. Since the t-stat is 0.046 and 0.058 for LMI factors, the statistically insignificant LMI factor group is removed from further analysis. The positive coefficient for JobsReported (8214.57) indicates that, for every additional employee, the loan is expected to increase by $8,214.57. The model has an Adj. R^2 of 62.84%. The maximum Generalized Variance Inflation Factor (GVIF) is 3.92, which is below the threshold of multicollinearity (5.0), implying that no obvious signs of multicollinearity exist in the independent variables. When the degrees of freedom are taken into account, the maximum factor is 1.41.

Table 5-3 UVA Result Summary

(a) Dependent Variable: Initial Approval Amount

Factor	Variable	Coefficient	t-stat	p-value	Adj. R^2	Explanations
Jobs Reported	(Intercept)	5193.38	108.47	0	62.80%	The positive coefficient of 8215.27 indicates that with every additional employee, the PPP loan amount increase by that $8,215.27. This conclusion is statistically significant with P-value is close to 0 (<5%). This factor explains 62.80% of the results.
	JobsReported	8215.27	4456.84	0		
Race	(Intercept)	56382.15	64.18	0	0.28%	**American Indian or Alaskan** is selected as the reference state. The intercept of 56382.15 represent its average loan size. This number matches its average loan size in Table 3-9. From this prospective, that of a White applicant is 15938.58 more than an average American Indian or Alaskan loan, pr $72,320.73. This number matches the average loan size for White in Table 3-9.
	Race2Asian	973.47	0.98	0.32905		
	Race2Black or African American	-34593.63	-37.62	0		
	Race2Native Hawaiian or Other Pacific Islander	-9220.40	-3.51	0.00045		
	Race2other	-12761.57	-1.38	0.16618		
	Race2Unanswered	16268.58	18.43	7.50E-76		
	Race2White	15938.58	17.67	7.38E-70		
Gender	(Intercept)	40618.33	199.47	0	0.19%	
	GenderMale Owned	36658.02	145.25	0		

Factor	Variable	Coefficient	t-stat	p-value	Adj. R^2	Explanations
	GenderUnanswered	29970.39	133.19	0		All factors are statistically significant. **GenderFemale** is selected as the reference state. The intercept matches its average loan size in Table 3-9.
Veteran	(Intercept)	60580.29	450.65	0	0.04%	All factors are statistically significant. **Non-Veteran** is selected as the reference state. The intercept matches its average loan size in Table 3-9.
	VeteranUnanswered	11157.73	68.61	0		
	VeteranVeteran	8493.78	14.61	2.47E-48		
Non-profit	(Intercept)	65401.80	864.44	0	0.52%	The factor is statistically significant. **For-Profit** is selected as the reference state. The intercept matches its average loan size in Table 3-9.
	NonProfitY	124557.08	247.65	0		
Rural/Urban	(Intercept)	50919.34	312.32	0	0.12%	The factor is statistically significant. **Rural** is the reference state. The intercept matches that in Table 3-9.
	RuralUrbanIndicatorU	21936.52	119.49	0		
HUB Zone	(Intercept)	68435.59	779.67	0	0.00%	The factor is statistically significant. **Non-HUB** is the reference state. The intercept matches that in Table 3-9. The factor does not contribute to Adj. R^2.
	HubzoneIndicatorY	-796.92	-4.72	2.38E-06		
LMI	(Intercept)	23224.53	0.18	0.8567	0.01%	The factor is NOT statistically significant. **N/A** (missing data) is the reference state. The intercept matches that in Table 3-9. The factor barely contributes to Adj. R^2.
	LMIIndicatorN	46302.20	0.36	0.7189		
	LMIIndicatorY	41670.56	0.32	0.7460		

Factor	Variable	Coefficient	t-stat	p-value	Adj. R^2	Explanations
Sector	(Intercept)	100087.66	358.905	0	1.44%	All sector dummies are statistically significant. **Accommodation and Food Service** is selected as the reference state. The intercept (100087.66) matches its average loan in Table 3-9. The average construction is \$94,407(= 100087.66 - 5680.66), matching its average loan in Table 3-9. All sub sectors are statistically significant, although the Adj. R^2 of 1.44% is pretty low.
	sector2Administrative and Support and Waste Management and Remediation Services	-38529.85	-91.027	0		
	sector2Construction	-5680.66	-15.173	5.36E-52		
	sector2Health Care and Social Assistance	-4113.02	-10.902	1.13E-27		
	sector2Manufacturing	67272.21	143.120	0		
	sector2Other	-49980.76	-155.21	0		
	sector2Other Services (except Public Administration)	-64780.34	-189.477	0		
	sector2Professional, Scientific, and Technical Services	-28133.76	-78.952	0		
	sector2Retail Trade	-40030.47	-104.07	0		
	sector2Transportation and Warehousing	-66160.83	-174.09	0		
	sector2Wholesale Trade	6144.57	12.059	1.74E-33		
	(Intercept)	30979.93	254.78	0	0.38%	

Factor	Variable	Coefficient	t-stat	p-value	Adj. R^2	Explanations
Avg. Household AGI	avg_HH_AGI	0.213658	165.92	0		The positive coefficient indicates that cities with high AGI tend to have more the PPP loan. This conclusion is statistically significant.

(b) Dependent Variable: Initial Approval Amount per Employee

Factor	Variable	Coefficient	t-stat	p-value	Adj. R^2	Explanations
Jobs Reported	(Intercept)	12285.99	2778.91	0	0.45%	The negative coefficient of -44.27 indicates that with every additional employee, the PPP loan amount per employee decreases by $44.27. This conclusion is statistically significant with P-value is close to 0. This factor explains small portion of the dependent variable.
	JobsReported	-44.78	-182.39	0		

Table 5-4 MVA Model with Factors from Loan Tape Only

```
Model Specification (Model No.: mdl_MVA)
lm(formula = InitialApprovalAmount ~ JobsReported + Race2 + Gender +
    Veteran + NonProfit + RuralUrbanIndicator + HubzoneIndicator +
    LMIIndicator, data = all_data)

Residuals:
     Min       1Q   Median       3Q      Max
-4116537   -13087    -2196     6475  9982533

Coefficients:
                                             Estimate Std. Error  t value Pr(>|t|)
(Intercept)                                -19400.517  78410.670   -0.247  0.80458
JobsReported                                 8214.572      1.858 4422.266  < 2e-16 ***
Race2Asian                                 -12422.451    609.500  -20.381  < 2e-16 ***
Race2Black or African American               9626.708    562.954   17.100  < 2e-16 ***
Race2Native Hawaiian or Other Pacific Islander  -2488.032   1604.299   -1.551  0.12094
Race2Other                                 -17825.636   5626.378   -3.168  0.00153 **
Race2Unanswered                              6396.788    546.596   11.703  < 2e-16 ***
Race2White                                   7556.921    551.044   13.714  < 2e-16 ***
GenderMale Owned                             6887.527    156.918   43.892  < 2e-16 ***
GenderUnanswered                             7642.804    213.469   35.803  < 2e-16 ***
VeteranUnanswered                            1488.628    188.249    7.908 2.62e-15 ***
VeteranVeteran                               -844.351    355.962   -2.372  0.01769 *
NonProfitY                                  -8366.915    309.697  -27.016  < 2e-16 ***
RuralUrbanIndicatorU                         8553.255    114.871   74.460  < 2e-16 ***
HubzoneIndicatorY                             281.497    115.209    2.443  0.01455 *
LMIIndicatorN                                4571.193  78408.717    0.058  0.95351
LMIIndicatorY                                3601.038  78408.771    0.046  0.96337
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 156800 on 11768365 degrees of freedom
Multiple R-squared:  0.6284,  Adjusted R-squared:  0.6284
F-statistic: 1.244e+06 on 16 and 11768365 DF,  p-value: < 2.2e-16
```

```
GVIF
                        GVIF Df GVIF^(1/(2*Df))
JobsReported        1.016805  1        1.008367
Race2               2.072753  6        1.062622
Gender              3.922458  2        1.407309
Veteran             3.729302  2        1.389655
NonProfit           1.015129  1        1.007536
RuralUrbanIndicator 1.052406  1        1.025869
HubzoneIndicator    1.252170  1        1.119004
LMIIndicator        1.268376  2        1.061236
```

Loan Model with Factors from Loan Tape and Sector

When only variables from the loan tape and sector are used, the model specification is provided in Table 5-5. Except Race2Native Hawaiian, all other variables are statistically significant. The positive coefficient for JobsReported (8241.58) indicates that, for every additional employee, the loan is expected to increase by $8,214.58. The model has a higher Adj. R^2 of 63.28%. The maximum Generalized Variance Inflation Factor (GVIF) is 4.01, which is below the threshold of multicollinearity (5.0), implying that no obvious signs of multicollinearity exist in the independent variables. When the degrees of freedom are taken into account, the maximum factor is 1.42.

Table 5-5 MVA Model with Factors from Loan Tape and Sector

```
Model Specification (Model No.: mdl_MVA2)
lm(formula = InitialApprovalAmount ~ JobsReported + Race2 + Gender + 
    Veteran + NonProfit + RuralUrbanIndicator + HubzoneIndicator + 
    sector2, data = all_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-4154652   -18174    -2109    10781  9990317 

Coefficients:
                                                 Estimate Std. Error  t value Pr(>|t|)    
(Intercept)                                    -55582.190    577.366  -96.269  < 2e-16 ***
JobsReported                                     8241.584      1.869 4410.425  < 2e-16 ***
Race2Asian                                      -2420.610    606.812   -3.989 6.63e-05 ***
Race2Black or African American                  10813.430    559.827   19.316  < 2e-16 ***
Race2Native Hawaiian or Other 
            Pacific Islander                    -1388.118   1594.905   -0.870  0.38411    
Race2Other                                     -16522.486   5593.479   -2.954  0.00314 ** 
Race2Unanswered                                  8066.290    543.428   14.843  < 2e-16 ***
Race2White                                       6691.082    547.830   12.214  < 2e-16 ***
GenderMale Owned                                 3466.109    157.574   21.997  < 2e-16 ***
GenderUnanswered                                 4244.666    212.922   19.935  < 2e-16 ***
VeteranUnanswered                                 474.894    187.183    2.537  0.01118 *  
VeteranVeteran                                  -1904.994    353.916   -5.383 7.34e-08 ***
NonProfitY                                      -3265.064    314.656  -10.377  < 2e-16 ***
RuralUrbanIndicatorU                             7696.370    114.318   67.324  < 2e-16 ***
HubzoneIndicatorY                                 813.793    103.457    7.866 3.66e-15 ***
sector2Administrative and Support and Waste 
       Management and Remediation Services      35092.449    260.185  134.875  < 2e-16 ***
sector2Construction                             65414.849    231.099  283.060  < 2e-16 ***
sector2Health Care and Social Assistance        37736.681    231.938  162.701  < 2e-16 ***
sector2Manufacturing                            77158.468    287.796  268.101  < 2e-16 ***
sector2Other                                    39537.988    199.853  197.835  < 2e-16 ***
sector2Other Services (except 
      Public Administration)                    36541.433    213.483  171.168  < 2e-16 ***
sector2Professional, Scientific, 
      and Technical Services                    61718.570    220.030  280.501  < 2e-16 ***
sector2Retail Trade                             35622.656    236.019  150.931  < 2e-16 ***
sector2Transportation and Warehousing           42526.005    234.857  181.072  < 2e-16 ***
sector2Wholesale Trade                          64157.460    311.936  205.675  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 155900 on 11768357 degrees of freedom
Multiple R-squared:  0.6328,  Adjusted R-squared:  0.6328 
F-statistic: 8.449e+05 on 24 and 11768357 DF,  p-value: < 2.2e-16
```

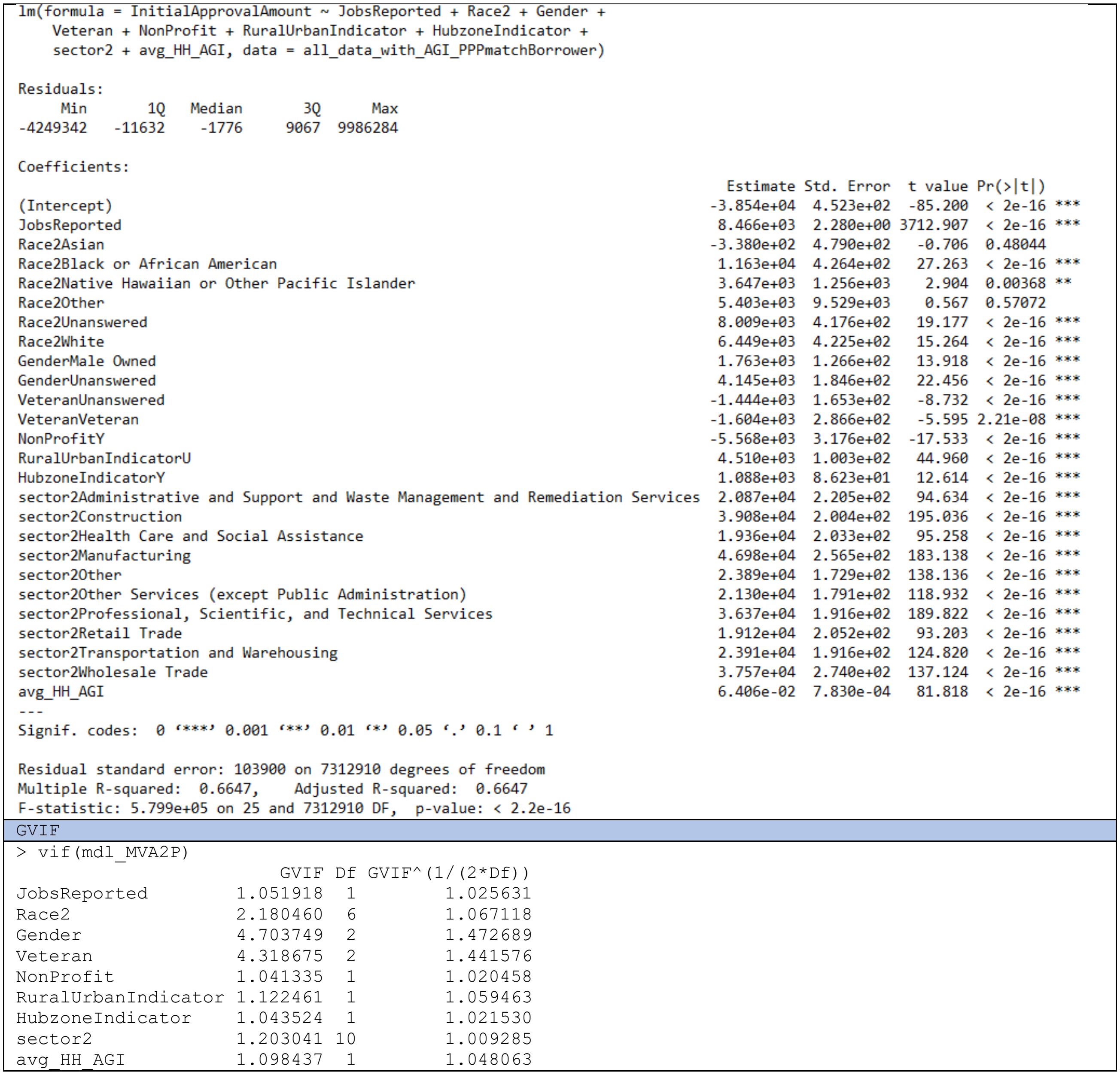
```
GVIF
                        GVIF Df GVIF^(1/(2*Df))
JobsReported        1.041179  1        1.020382
Race2               2.113617  6        1.064353
Gender              4.007907  2        1.414912
Veteran             3.731595  2        1.389868
NonProfit           1.060294  1        1.029706
RuralUrbanIndicator 1.054623  1        1.026949
HubzoneIndicator    1.021686  1        1.010785
sector2             1.167830 10        1.007788
```

Loan Model with Factors from Loan Tape, Sector and AGI

When variables from the loan tape, sector and average AGI are used, the model specification is provided in Table 5-7. Except Race2Asian, Race2Other, all other variables are statistically significant. The positive coefficient for JobsReported (8466) indicates that, for every additional employee, the loan is expected to increase by $8,466. *The positive coefficient for avg Household AGI (0.06406) indicates that, for every $1 in AGI, the loan is expected to increase by $0.064, implying that for higher income region, the loan is expected to be higher, given everything else equal.* The model has a higher Adj. R^2 of 63.28%. The maximum Generalized Variance Inflation Factor (GVIF) is 4.01, which is below the threshold of multicollinearity (5.0) , implying that no obvious signs of multicollinearity exist in the independent variables. When the degrees of freedom are taken into account, the maximum factor is 1.42.

Table 5-6 MVA with Factors from Loan Tape, Sector and Average AGI

Model Specification (Model No.: mdl_MVA2B)



```
lm(formula = InitialApprovalAmount ~ JobsReported + Race2 + Gender +
    Veteran + NonProfit + RuralUrbanIndicator + HubzoneIndicator +
    sector2 + avg_HH_AGI, data = all_data_with_AGI_PPPmatchBorrower)

Residuals:
     Min       1Q   Median       3Q      Max
-4249342   -11632    -1776     9067  9986284

Coefficients:
                                                                                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)                                                                      -3.854e+04  4.523e+02  -85.200  < 2e-16 ***
JobsReported                                                                      8.466e+03  2.280e+00 3712.907  < 2e-16 ***
Race2Asian                                                                       -3.380e+02  4.790e+02   -0.706  0.48044
Race2Black or African American                                                    1.163e+04  4.264e+02   27.263  < 2e-16 ***
Race2Native Hawaiian or Other Pacific Islander                                    3.647e+03  1.256e+03    2.904  0.00368 **
Race2Other                                                                        5.403e+03  9.529e+03    0.567  0.57072
Race2Unanswered                                                                   8.009e+03  4.176e+02   19.177  < 2e-16 ***
Race2White                                                                        6.449e+03  4.225e+02   15.264  < 2e-16 ***
GenderMale Owned                                                                  1.763e+03  1.266e+02   13.918  < 2e-16 ***
GenderUnanswered                                                                  4.145e+03  1.846e+02   22.456  < 2e-16 ***
VeteranUnanswered                                                                -1.444e+03  1.653e+02   -8.732  < 2e-16 ***
VeteranVeteran                                                                   -1.604e+03  2.866e+02   -5.595 2.21e-08 ***
NonProfitY                                                                       -5.568e+03  3.176e+02  -17.533  < 2e-16 ***
RuralUrbanIndicatorU                                                              4.510e+03  1.003e+02   44.960  < 2e-16 ***
HubzoneIndicatorY                                                                 1.088e+03  8.623e+01   12.614  < 2e-16 ***
sector2Administrative and Support and Waste Management and Remediation Services  2.087e+04  2.205e+02   94.634  < 2e-16 ***
sector2Construction                                                               3.908e+04  2.004e+02  195.036  < 2e-16 ***
sector2Health Care and Social Assistance                                          1.936e+04  2.033e+02   95.258  < 2e-16 ***
sector2Manufacturing                                                              4.698e+04  2.565e+02  183.138  < 2e-16 ***
sector2Other                                                                      2.389e+04  1.729e+02  138.136  < 2e-16 ***
sector2Other Services (except Public Administration)                              2.130e+04  1.791e+02  118.932  < 2e-16 ***
sector2Professional, Scientific, and Technical Services                           3.637e+04  1.916e+02  189.822  < 2e-16 ***
sector2Retail Trade                                                               1.912e+04  2.052e+02   93.203  < 2e-16 ***
sector2Transportation and Warehousing                                             2.391e+04  1.916e+02  124.820  < 2e-16 ***
sector2Wholesale Trade                                                            3.757e+04  2.740e+02  137.124  < 2e-16 ***
avg_HH_AGI                                                                        6.406e-02  7.830e-04   81.818  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 103900 on 7312910 degrees of freedom
Multiple R-squared:  0.6647,	Adjusted R-squared:  0.6647
F-statistic: 5.799e+05 on 25 and 7312910 DF,  p-value: < 2.2e-16
```

GVIF

```
> vif(mdl_MVA2P)
                        GVIF Df GVIF^(1/(2*Df))
JobsReported        1.051918  1        1.025631
Race2               2.180460  6        1.067118
Gender              4.703749  2        1.472689
Veteran             4.318675  2        1.441576
NonProfit           1.041335  1        1.020458
RuralUrbanIndicator 1.122461  1        1.059463
HubzoneIndicator    1.043524  1        1.021530
sector2             1.203041 10        1.009285
avg_HH_AGI          1.098437  1        1.048063
```

Loan Model with Factors from Loan Tape, Sector and AGI (log-linear)

This model is the same as that in Table 5-6, except the model specification is based on the natural log of the dependent variable or log(InitialApprovalAmount) in Table 5-7. The reason log-linear model is developed is to guarantee positive projected InitialApprovalAmount from the model. Except Race2Other, all other variables are statistically significant. The positive coefficient for JobsReported indicates that, the more the employee count, the higher the expected loan size. *The positive coefficient for avg Household AGI indicates that, the higher AGI, the larger the loan size, that is, for higher income region, the loan is expected to be higher, given everything else equal.* The maximum Generalized Variance Inflation Factor (GVIF) is 4.70, which is below the threshold of multicollinearity (5.0) , implying that no obvious signs of multicollinearity exist in the independent variables. When the degrees of freedom are taken into account, the maximum factor is 1.47. The model has a much lower Adj. R^2 of 30.74%. *As a result, this log-linear model is not selected*.

Table 5-7 Log-Linear MVA Model with Factors from Loan Tape, Sector and Average AGI

Model Specification (Model No.: mdl MVA2B-log)

```
lm(formula = log(InitialApprovalAmount) ~ JobsReported + Race2 +
    Gender + Veteran + NonProfit + RuralUrbanIndicator + HubzoneIndicator +
    sector2 + avg_HH_AGI, data = all_data_with_AGI_PPPmatchBorrower)

Residuals:
     Min       1Q   Median       3Q      Max
-19.2920  -0.5497   0.1467   0.5583   6.6397

Coefficients:
                                                                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)                                                       9.611e+00  4.428e-03 2170.679  < 2e-16 ***
JobsReported                                                      3.249e-02  2.232e-05 1455.644  < 2e-16 ***
Race2Asian                                                       -1.611e-02  4.689e-03   -3.436 0.000590 ***
Race2Black or African American                                    5.335e-02  4.174e-03   12.781  < 2e-16 ***
Race2Native Hawaiian or Other Pacific Islander                   -8.770e-02  1.229e-02   -7.135 9.69e-13 ***
Race2Other                                                       -2.929e-02  9.328e-02   -0.314 0.753492
Race2Unanswered                                                   1.153e-01  4.088e-03   28.214  < 2e-16 ***
Race2White                                                       -8.095e-03  4.135e-03   -1.957 0.050292 .
GenderMale Owned                                                  2.736e-01  1.240e-03  220.691  < 2e-16 ***
GenderUnanswered                                                 -6.542e-03  1.807e-03   -3.621 0.000294 ***
VeteranUnanswered                                                -1.131e-02  1.618e-03   -6.990 2.75e-12 ***
VeteranVeteran                                                   -4.948e-02  2.806e-03  -17.637  < 2e-16 ***
NonProfitY                                                        5.954e-01  3.109e-03  191.531  < 2e-16 ***
RuralUrbanIndicatorU                                              1.711e-01  9.818e-04  174.221  < 2e-16 ***
HubzoneIndicatorY                                                 2.061e-02  8.441e-04   24.419  < 2e-16 ***
sector2Administrative and Support and Waste Management and Remediation Services -5.758e-01  2.158e-03 -266.786  < 2e-16 ***
sector2Construction                                              -2.293e-01  1.961e-03 -116.902  < 2e-16 ***
sector2Health Care and Social Assistance                         -1.971e-01  1.990e-03  -99.045  < 2e-16 ***
sector2Manufacturing                                             -2.831e-02  2.511e-03  -11.275  < 2e-16 ***
sector2Other                                                     -5.502e-01  1.693e-03 -325.055  < 2e-16 ***
sector2Other Services (except Public Administration)             -6.163e-01  1.753e-03 -351.627  < 2e-16 ***
sector2Professional, Scientific, and Technical Services          -3.651e-01  1.875e-03 -194.675  < 2e-16 ***
sector2Retail Trade                                              -4.714e-01  2.008e-03 -234.739  < 2e-16 ***
sector2Transportation and Warehousing                            -7.923e-01  1.875e-03 -422.516  < 2e-16 ***
sector2Wholesale Trade                                           -2.110e-01  2.682e-03  -78.678  < 2e-16 ***
avg_HH_AGI                                                        1.605e-06  7.664e-09  209.457  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.017 on 7312910 degrees of freedom
Multiple R-squared:  0.3074,    Adjusted R-squared:  0.3074
F-statistic: 1.298e+05 on 25 and 7312910 DF,  p-value: < 2.2e-16
```

GVIF

```
                        GVIF Df GVIF^(1/(2*Df))
JobsReported        1.051918  1        1.025631
Race2               2.180460  6        1.067118
```

```
Gender              4.703749  2        1.472689
Veteran             4.318675  2        1.441576
NonProfit           1.041335  1        1.020458
RuralUrbanIndicator 1.122461  1        1.059463
HubzoneIndicator    1.043524  1        1.021530
sector2             1.203041 10        1.009285
avg_HH_AGI          1.098437  1        1.048063
```

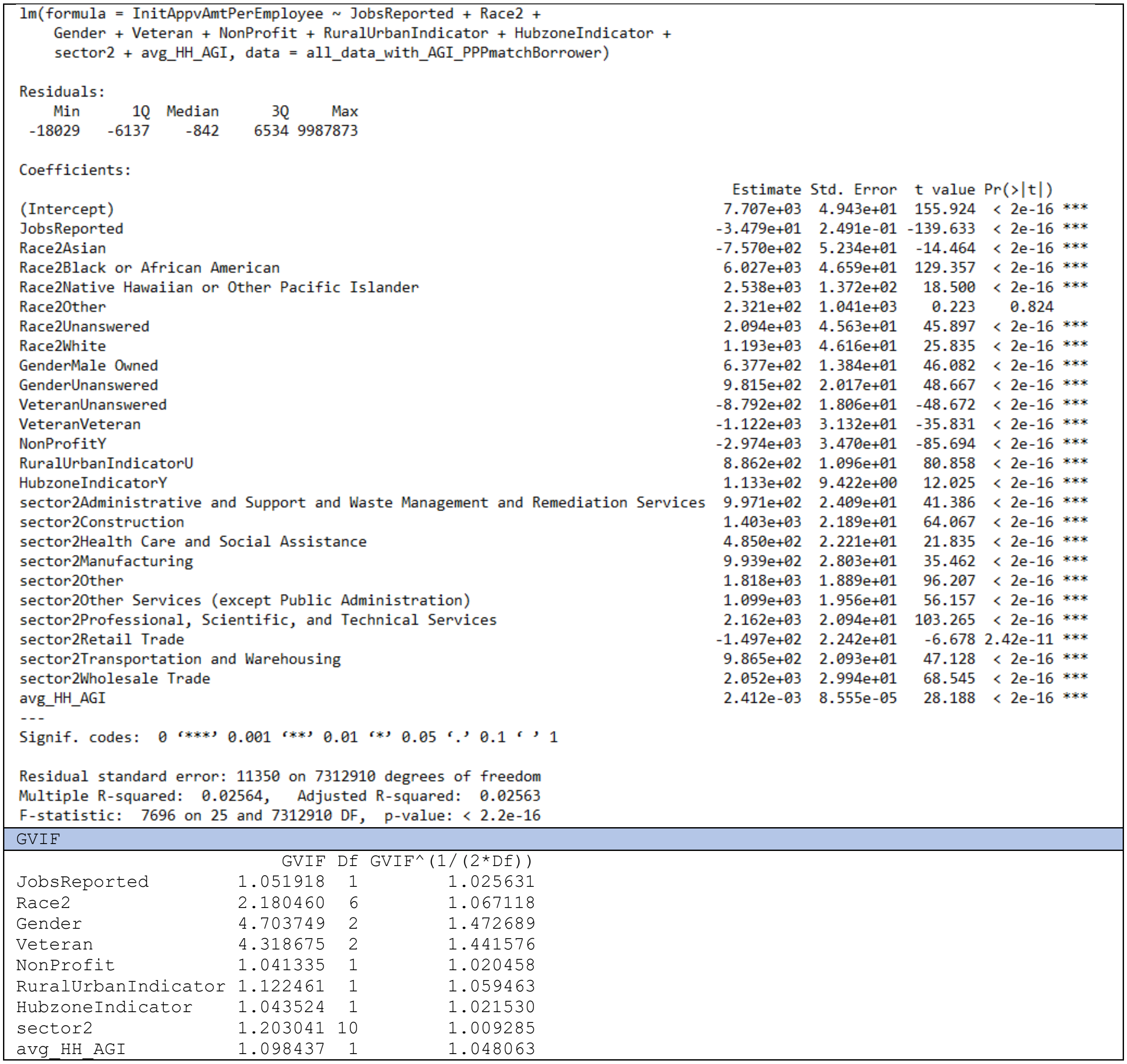
5.4.2 MVA Model on Initial Approval Amount Per Employee

Loan Per Employee with Factors from Loan Tape, Sector and AGI

So far the models are built at the loan level, where JobsReported has positively contributions to the projected loan size and a very high contribution to the models adj. R^2. The next question is, what factors contribute to the loan per JobsReported. The answer this question, a new dependent variable is defined as loan size divided by JobsReported. The same independent variables as in Table 5-6, i.e., variables from the loan tape, sector and average AGI, are used for modeling loan per employee. The model specification is provided in Table 5-8. Except Race2Other, all other variables are statistically significant. The negative coefficient for JobsReported (-34.79) indicates that, for every additional employee, the loan per employee is expected to **decrease** by $34.79. The slope is similar to that observed in Table 5-3 (b) in UVA but with a reduced magnitude due to the contributions from other factors in MVA here. *The positive coefficient for avg Household AGI (0.00224) indicates that, for every $1 increase in AGI, the loan per employee is expected to increase by $0.00224, implying that, for higher income region, even the loan per employee is expected to be higher, given everything else equal.* The model has a very low Adj. R^2 of 2.563%. The maximum Generalized Variance Inflation Factor (GVIF) is 4.70, which is below the threshold of multicollinearity (5.0) , implying that no obvious signs of multicollinearity exist in the independent variables. When the degrees of freedom are taken into account, the maximum factor is 1.47.

Table 5-8 MVA Model of Loan Per Employee with Factors from Loan Tape, Sector and AGI

```
Model Specification (Model No.: mdl_MVA2B_PE)
```



```
lm(formula = InitAppvAmtPerEmployee ~ JobsReported + Race2 + 
    Gender + Veteran + NonProfit + RuralUrbanIndicator + HubzoneIndicator + 
    sector2 + avg_HH_AGI, data = all_data_with_AGI_PPPmatchBorrower)

Residuals:
    Min      1Q  Median      3Q     Max 
 -18029   -6137    -842    6534 9987873 

Coefficients:
                                                                                  Estimate Std. Error  t value Pr(>|t|)    
(Intercept)                                                                      7.707e+03  4.943e+01  155.924  < 2e-16 ***
JobsReported                                                                    -3.479e+01  2.491e-01 -139.633  < 2e-16 ***
Race2Asian                                                                      -7.570e+02  5.234e+01  -14.464  < 2e-16 ***
Race2Black or African American                                                   6.027e+03  4.659e+01  129.357  < 2e-16 ***
Race2Native Hawaiian or Other Pacific Islander                                   2.538e+03  1.372e+02   18.500  < 2e-16 ***
Race2Other                                                                       2.321e+02  1.041e+03    0.223    0.824    
Race2Unanswered                                                                  2.094e+03  4.563e+01   45.897  < 2e-16 ***
Race2White                                                                       1.193e+03  4.616e+01   25.835  < 2e-16 ***
GenderMale Owned                                                                 6.377e+02  1.384e+01   46.082  < 2e-16 ***
GenderUnanswered                                                                 9.815e+02  2.017e+01   48.667  < 2e-16 ***
VeteranUnanswered                                                               -8.792e+02  1.806e+01  -48.672  < 2e-16 ***
VeteranVeteran                                                                  -1.122e+03  3.132e+01  -35.831  < 2e-16 ***
NonProfitY                                                                      -2.974e+03  3.470e+01  -85.694  < 2e-16 ***
RuralUrbanIndicatorU                                                             8.862e+02  1.096e+01   80.858  < 2e-16 ***
HubzoneIndicatorY                                                                1.133e+02  9.422e+00   12.025  < 2e-16 ***
sector2Administrative and Support and Waste Management and Remediation Services  9.971e+02  2.409e+01   41.386  < 2e-16 ***
sector2Construction                                                              1.403e+03  2.189e+01   64.067  < 2e-16 ***
sector2Health Care and Social Assistance                                         4.850e+02  2.221e+01   21.835  < 2e-16 ***
sector2Manufacturing                                                             9.939e+02  2.803e+01   35.462  < 2e-16 ***
sector2Other                                                                     1.818e+03  1.889e+01   96.207  < 2e-16 ***
sector2Other Services (except Public Administration)                             1.099e+03  1.956e+01   56.157  < 2e-16 ***
sector2Professional, Scientific, and Technical Services                          2.162e+03  2.094e+01  103.265  < 2e-16 ***
sector2Retail Trade                                                             -1.497e+02  2.242e+01   -6.678 2.42e-11 ***
sector2Transportation and Warehousing                                            9.865e+02  2.093e+01   47.128  < 2e-16 ***
sector2Wholesale Trade                                                           2.052e+03  2.994e+01   68.545  < 2e-16 ***
avg_HH_AGI                                                                       2.412e-03  8.555e-05   28.188  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11350 on 7312910 degrees of freedom
Multiple R-squared:  0.02564,	Adjusted R-squared:  0.02563 
F-statistic:  7696 on 25 and 7312910 DF,  p-value: < 2.2e-16
```

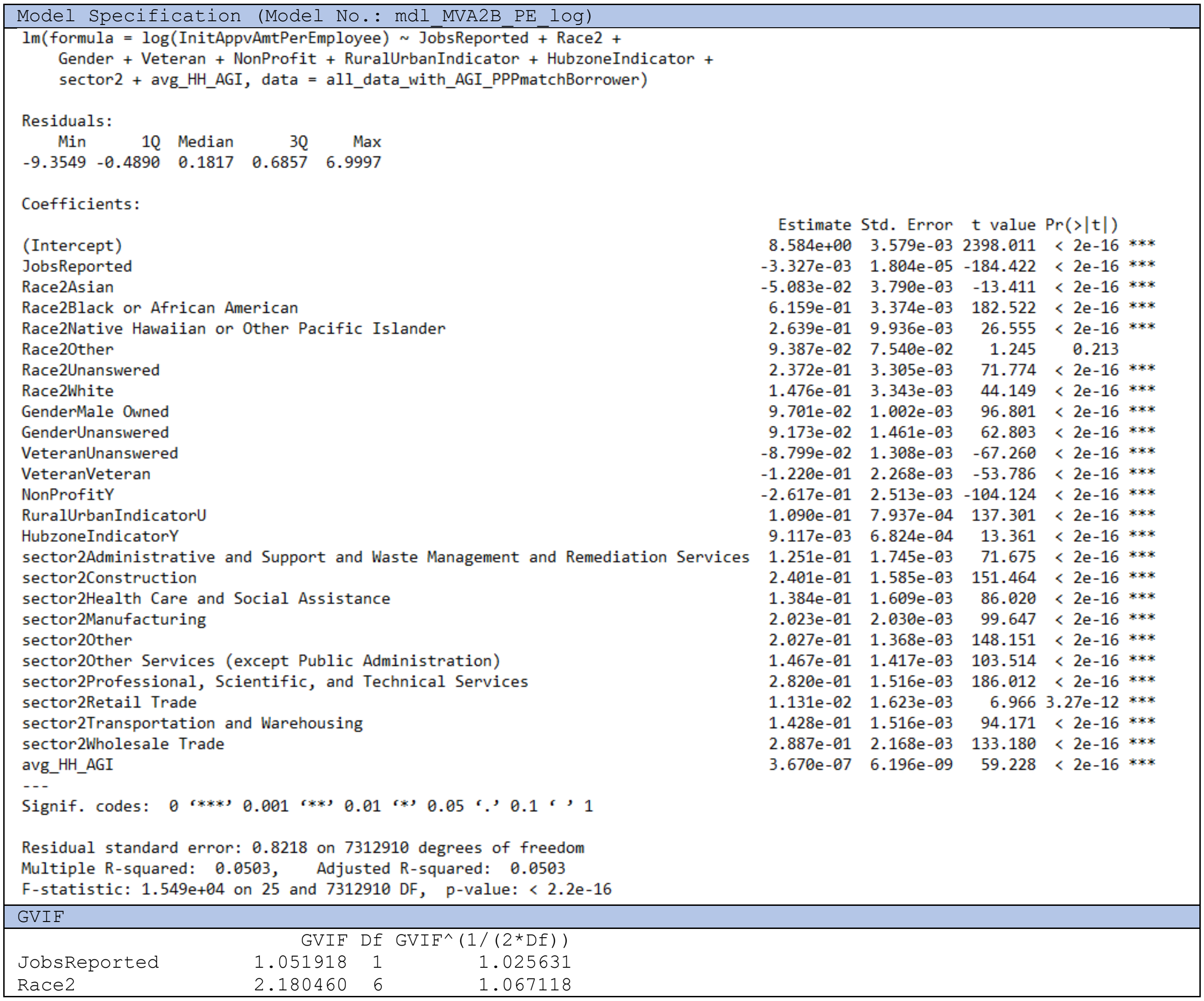
```
GVIF
                        GVIF Df GVIF^(1/(2*Df))
JobsReported        1.051918  1        1.025631
Race2               2.180460  6        1.067118
Gender              4.703749  2        1.472689
Veteran             4.318675  2        1.441576
NonProfit           1.041335  1        1.020458
RuralUrbanIndicator 1.122461  1        1.059463
HubzoneIndicator    1.043524  1        1.021530
sector2             1.203041 10        1.009285
avg_HH_AGI          1.098437  1        1.048063
```

Loan Per Employee with Factors from Loan Tape, Sector and Average AGI (log-linear)

This model is the same as that in Table 5-8 except the model specification is based on the natural log of the dependent variable or log(InitAppvAmtPerEmployee). The reason log-linear model is developed is to mathematically guarantee positive projected InitAppvAmtPerEmployee from the model.

The model specification is provided in Table 5-9. Except Race2Other, all other variables are statistically significant. The negative coefficient for JobsReported (-0.003327) indicates that, for every additional employee, the log of loan per employee is expected to **decrease** by 0.003327. *The positive coefficient for avg Household AGI (3.67*10^{-7}) indicates that for, higher income region, even the loan per employee is expected to be higher, given everything else equal.* The model has a slightly higher Adj. R^2 of 5.03%, almost doubling that in the model in Table 5-8. The maximum Generalized Variance Inflation Factor (GVIF) is 4.70, which is below the threshold of multicollinearity (5.0) , implying that no obvious signs of multicollinearity exist in the independent variables. When the degrees of freedom are taken into account, the maximum factor is 1.47.

Table 5-9 Log-linear MVA Model of Loan Per Employee with Factors from Loan Tape, Sector and AGI



Model Specification (Model No.: mdl_MVA2B_PE_log)

```
lm(formula = log(InitAppvAmtPerEmployee) ~ JobsReported + Race2 +
    Gender + Veteran + NonProfit + RuralUrbanIndicator + HubzoneIndicator +
    sector2 + avg_HH_AGI, data = all_data_with_AGI_PPPmatchBorrower)

Residuals:
    Min      1Q  Median      3Q     Max
-9.3549 -0.4890  0.1817  0.6857  6.9997

Coefficients:
                                                                   Estimate Std. Error  t value Pr(>|t|)
(Intercept)                                                       8.584e+00  3.579e-03 2398.011  < 2e-16 ***
JobsReported                                                     -3.327e-03  1.804e-05 -184.422  < 2e-16 ***
Race2Asian                                                       -5.083e-02  3.790e-03  -13.411  < 2e-16 ***
Race2Black or African American                                    6.159e-01  3.374e-03  182.522  < 2e-16 ***
Race2Native Hawaiian or Other Pacific Islander                    2.639e-01  9.936e-03   26.555  < 2e-16 ***
Race2Other                                                        9.387e-02  7.540e-02    1.245    0.213
Race2Unanswered                                                   2.372e-01  3.305e-03   71.774  < 2e-16 ***
Race2White                                                        1.476e-01  3.343e-03   44.149  < 2e-16 ***
GenderMale Owned                                                  9.701e-02  1.002e-03   96.801  < 2e-16 ***
GenderUnanswered                                                  9.173e-02  1.461e-03   62.803  < 2e-16 ***
VeteranUnanswered                                                -8.799e-02  1.308e-03  -67.260  < 2e-16 ***
VeteranVeteran                                                   -1.220e-01  2.268e-03  -53.786  < 2e-16 ***
NonProfitY                                                       -2.617e-01  2.513e-03 -104.124  < 2e-16 ***
RuralUrbanIndicatorU                                              1.090e-01  7.937e-04  137.301  < 2e-16 ***
HubzoneIndicatorY                                                 9.117e-03  6.824e-04   13.361  < 2e-16 ***
sector2Administrative and Support and Waste Management and Remediation Services 1.251e-01  1.745e-03   71.675  < 2e-16 ***
sector2Construction                                               2.401e-01  1.585e-03  151.464  < 2e-16 ***
sector2Health Care and Social Assistance                          1.384e-01  1.609e-03   86.020  < 2e-16 ***
sector2Manufacturing                                              2.023e-01  2.030e-03   99.647  < 2e-16 ***
sector2Other                                                      2.027e-01  1.368e-03  148.151  < 2e-16 ***
sector2Other Services (except Public Administration)              1.467e-01  1.417e-03  103.514  < 2e-16 ***
sector2Professional, Scientific, and Technical Services           2.820e-01  1.516e-03  186.012  < 2e-16 ***
sector2Retail Trade                                               1.131e-02  1.623e-03    6.966 3.27e-12 ***
sector2Transportation and Warehousing                             1.428e-01  1.516e-03   94.171  < 2e-16 ***
sector2Wholesale Trade                                            2.887e-01  2.168e-03  133.180  < 2e-16 ***
avg_HH_AGI                                                        3.670e-07  6.196e-09   59.228  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8218 on 7312910 degrees of freedom
Multiple R-squared:  0.0503,	Adjusted R-squared:  0.0503
F-statistic: 1.549e+04 on 25 and 7312910 DF,  p-value: < 2.2e-16
```

GVIF

```
                        GVIF Df GVIF^(1/(2*Df))
JobsReported        1.051918  1        1.025631
Race2               2.180460  6        1.067118
```

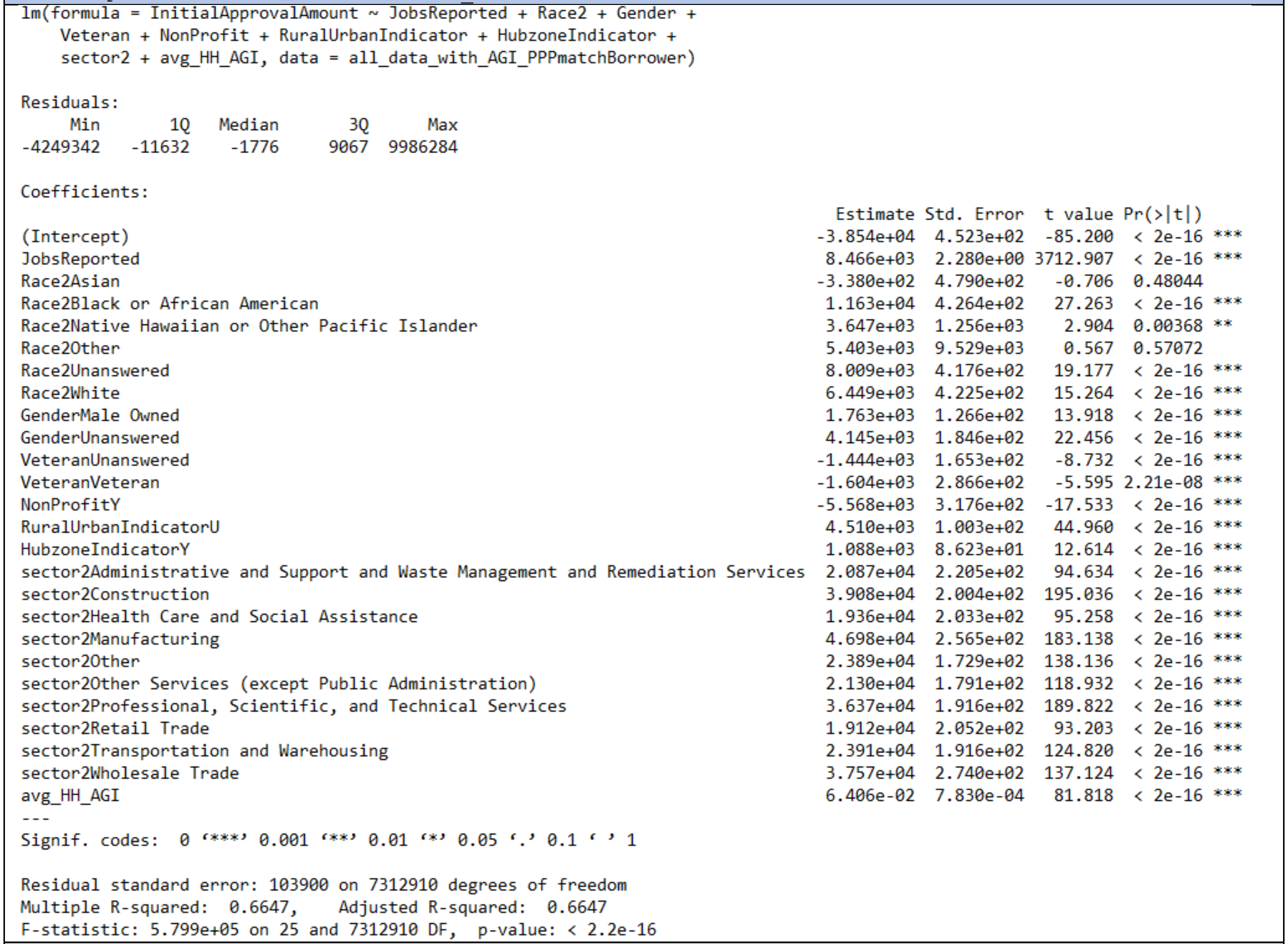
```
Gender               4.703749  2         1.472689
Veteran              4.318675  2         1.441576
NonProfit            1.041335  1         1.020458
RuralUrbanIndicator  1.122461  1         1.059463
HubzoneIndicator     1.043524  1         1.021530
sector2              1.203041 10         1.009285
avg_HH_AGI           1.098437  1         1.048063
```

5.5 Final Models

The final model for Initial Approval Amount is presented in provided in Table 5-6, reproduced here as Table 5-10 and that for Initial Approval Amount per Employee is Table 5-9, reproduced as Table 5-11 Table 5-11.

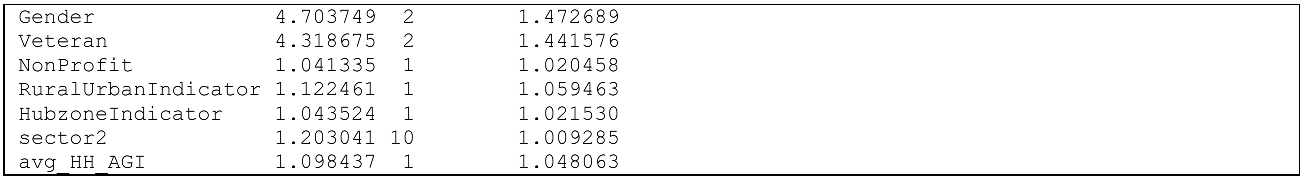
Table 5-10 Final Model for Initial Approval Amount (same as Table 5-6)

```
Model Specification (Model No.: mdl_MVA2B)
```



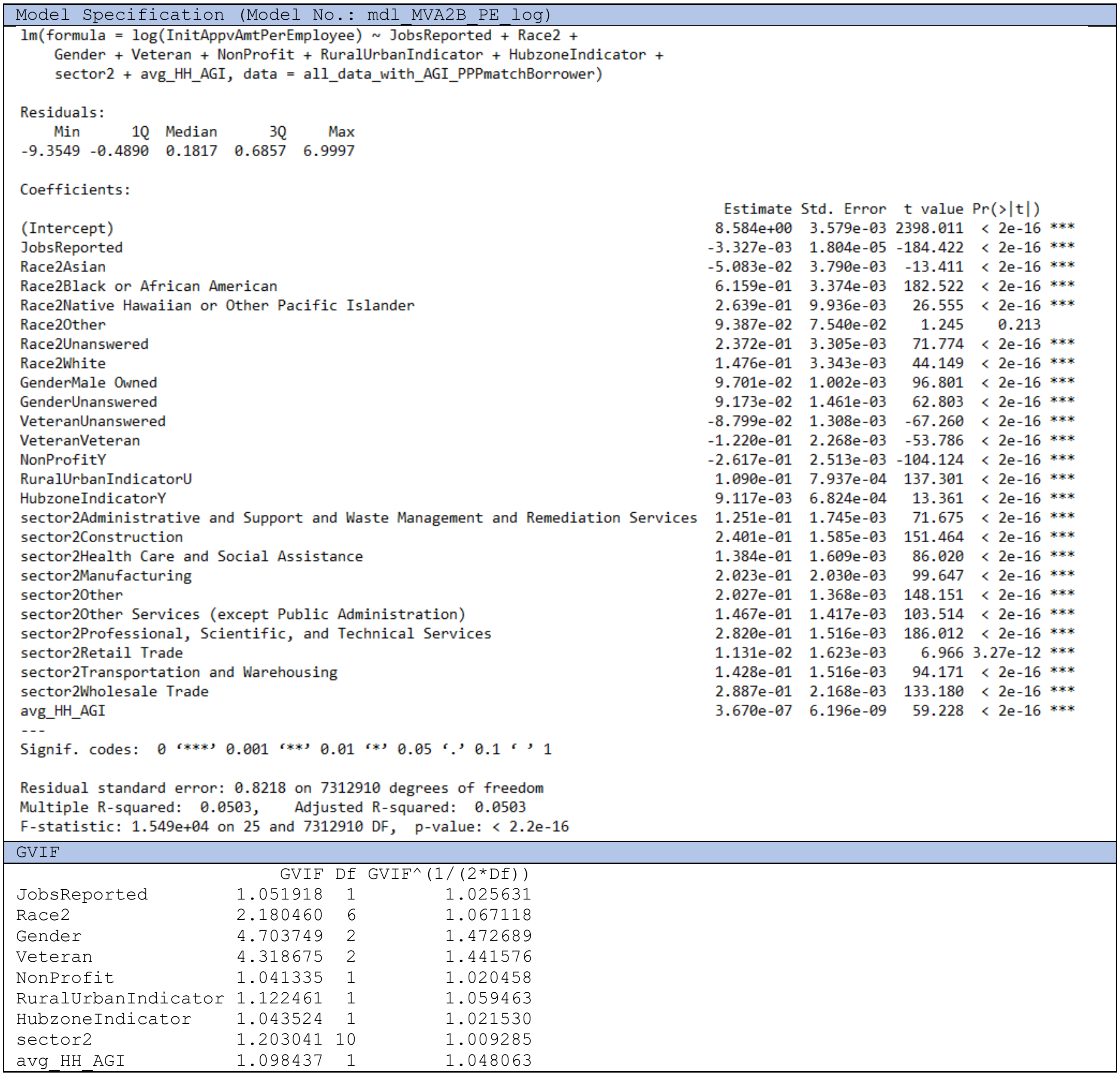
```
lm(formula = InitialApprovalAmount ~ JobsReported + Race2 + Gender + 
    Veteran + NonProfit + RuralUrbanIndicator + HubzoneIndicator + 
    sector2 + avg_HH_AGI, data = all_data_with_AGI_PPPmatchBorrower)

Residuals:
     Min       1Q   Median       3Q      Max 
-4249342   -11632    -1776     9067  9986284 

Coefficients:
                                                                     Estimate Std. Error  t value Pr(>|t|)    
(Intercept)                                                        -3.854e+04  4.523e+02  -85.200  < 2e-16 ***
JobsReported                                                        8.466e+03  2.280e+00 3712.907  < 2e-16 ***
Race2Asian                                                         -3.380e+02  4.790e+02   -0.706  0.48044    
Race2Black or African American                                      1.163e+04  4.264e+02   27.263  < 2e-16 ***
Race2Native Hawaiian or Other Pacific Islander                      3.647e+03  1.256e+03    2.904  0.00368 ** 
Race2Other                                                          5.403e+03  9.529e+03    0.567  0.57072    
Race2Unanswered                                                     8.009e+03  4.176e+02   19.177  < 2e-16 ***
Race2White                                                          6.449e+03  4.225e+02   15.264  < 2e-16 ***
GenderMale Owned                                                    1.763e+03  1.266e+02   13.918  < 2e-16 ***
GenderUnanswered                                                    4.145e+03  1.846e+02   22.456  < 2e-16 ***
VeteranUnanswered                                                  -1.444e+03  1.653e+02   -8.732  < 2e-16 ***
VeteranVeteran                                                     -1.604e+03  2.866e+02   -5.595 2.21e-08 ***
NonProfitY                                                         -5.568e+03  3.176e+02  -17.533  < 2e-16 ***
RuralUrbanIndicatorU                                                4.510e+03  1.003e+02   44.960  < 2e-16 ***
HubzoneIndicatorY                                                   1.088e+03  8.623e+01   12.614  < 2e-16 ***
sector2Administrative and Support and Waste Management and Remediation Services  2.087e+04  2.205e+02   94.634  < 2e-16 ***
sector2Construction                                                 3.908e+04  2.004e+02  195.036  < 2e-16 ***
sector2Health Care and Social Assistance                            1.936e+04  2.033e+02   95.258  < 2e-16 ***
sector2Manufacturing                                                4.698e+04  2.565e+02  183.138  < 2e-16 ***
sector2Other                                                        2.389e+04  1.729e+02  138.136  < 2e-16 ***
sector2Other Services (except Public Administration)                2.130e+04  1.791e+02  118.932  < 2e-16 ***
sector2Professional, Scientific, and Technical Services             3.637e+04  1.916e+02  189.822  < 2e-16 ***
sector2Retail Trade                                                 1.912e+04  2.052e+02   93.203  < 2e-16 ***
sector2Transportation and Warehousing                               2.391e+04  1.916e+02  124.820  < 2e-16 ***
sector2Wholesale Trade                                              3.757e+04  2.740e+02  137.124  < 2e-16 ***
avg_HH_AGI                                                          6.406e-02  7.830e-04   81.818  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 103900 on 7312910 degrees of freedom
Multiple R-squared:  0.6647,	Adjusted R-squared:  0.6647 
F-statistic: 5.799e+05 on 25 and 7312910 DF,  p-value: < 2.2e-16
```

```
GVIF
> vif(mdl_MVA2P)
                         GVIF Df GVIF^(1/(2*Df))
JobsReported         1.051918  1         1.025631
Race2                2.180460  6         1.067118
```



```
Gender              4.703749  2         1.472689
Veteran             4.318675  2         1.441576
NonProfit           1.041335  1         1.020458
RuralUrbanIndicator 1.122461  1         1.059463
HubzoneIndicator    1.043524  1         1.021530
sector2             1.203041 10         1.009285
avg_HH_AGI          1.098437  1         1.048063
```

Table 5-11 Final Model for log of Initial Approval Amount Per Employee (same as Table 5-9)



```
Model Specification (Model No.: mdl_MVA2B_PE_log)
lm(formula = log(InitAppvAmtPerEmployee) ~ JobsReported + Race2 +
    Gender + Veteran + NonProfit + RuralUrbanIndicator + HubzoneIndicator +
    sector2 + avg_HH_AGI, data = all_data_with_AGI_PPPmatchBorrower)

Residuals:
    Min      1Q  Median      3Q     Max
-9.3549 -0.4890  0.1817  0.6857  6.9997

Coefficients:
                                                                     Estimate Std. Error  t value Pr(>|t|)
(Intercept)                                                         8.584e+00  3.579e-03 2398.011  < 2e-16 ***
JobsReported                                                       -3.327e-03  1.804e-05 -184.422  < 2e-16 ***
Race2Asian                                                         -5.083e-02  3.790e-03  -13.411  < 2e-16 ***
Race2Black or African American                                      6.159e-01  3.374e-03  182.522  < 2e-16 ***
Race2Native Hawaiian or Other Pacific Islander                      2.639e-01  9.936e-03   26.555  < 2e-16 ***
Race2Other                                                          9.387e-02  7.540e-02    1.245    0.213
Race2Unanswered                                                     2.372e-01  3.305e-03   71.774  < 2e-16 ***
Race2White                                                          1.476e-01  3.343e-03   44.149  < 2e-16 ***
GenderMale Owned                                                    9.701e-02  1.002e-03   96.801  < 2e-16 ***
GenderUnanswered                                                    9.173e-02  1.461e-03   62.803  < 2e-16 ***
VeteranUnanswered                                                  -8.799e-02  1.308e-03  -67.260  < 2e-16 ***
VeteranVeteran                                                     -1.220e-01  2.268e-03  -53.786  < 2e-16 ***
NonProfitY                                                         -2.617e-01  2.513e-03 -104.124  < 2e-16 ***
RuralUrbanIndicatorU                                                1.090e-01  7.937e-04  137.301  < 2e-16 ***
HubzoneIndicatorY                                                   9.117e-03  6.824e-04   13.361  < 2e-16 ***
sector2Administrative and Support and Waste Management and Remediation Services 1.251e-01 1.745e-03 71.675 < 2e-16 ***
sector2Construction                                                 2.401e-01  1.585e-03  151.464  < 2e-16 ***
sector2Health Care and Social Assistance                            1.384e-01  1.609e-03   86.020  < 2e-16 ***
sector2Manufacturing                                                2.023e-01  2.030e-03   99.647  < 2e-16 ***
sector2Other                                                        2.027e-01  1.368e-03  148.151  < 2e-16 ***
sector2Other Services (except Public Administration)                1.467e-01  1.417e-03  103.514  < 2e-16 ***
sector2Professional, Scientific, and Technical Services             2.820e-01  1.516e-03  186.012  < 2e-16 ***
sector2Retail Trade                                                 1.131e-02  1.623e-03    6.966 3.27e-12 ***
sector2Transportation and Warehousing                               1.428e-01  1.516e-03   94.171  < 2e-16 ***
sector2Wholesale Trade                                              2.887e-01  2.168e-03  133.180  < 2e-16 ***
avg_HH_AGI                                                          3.670e-07  6.196e-09   59.228  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8218 on 7312910 degrees of freedom
Multiple R-squared:  0.0503,    Adjusted R-squared:  0.0503
F-statistic: 1.549e+04 on 25 and 7312910 DF,  p-value: < 2.2e-16
GVIF
                        GVIF Df GVIF^(1/(2*Df))
JobsReported        1.051918  1         1.025631
Race2               2.180460  6         1.067118
Gender              4.703749  2         1.472689
Veteran             4.318675  2         1.441576
NonProfit           1.041335  1         1.020458
RuralUrbanIndicator 1.122461  1         1.059463
HubzoneIndicator    1.043524  1         1.021530
sector2             1.203041 10         1.009285
avg_HH_AGI          1.098437  1         1.048063
```

6 Bibliography

During the Capstone project, the author reviewed literature quite extensively. The references are listed below. For website or data download, the date of visit is also provided.

1. PPP Loan Data Sources:
    a. PPP loan data from US Small Business Administration (SBA) website: PPP FOIA - Dataset - U.S. Small Business Administration (SBA) | Open Data. (Accessed July 6, 2021).
    b. PPP Data Dictionary, SBA website above. (Accessed July 6, 2021).
    c. Tracker: Paycheck Protection Program Loans - AAF (americanactionforum.org), report date: May 31, 2021. (Accessed September 22, 2021).
2. COVID case data: https://www.kff.org/state-category/covid-19/covid-19-metrics/, which is from COVID-19 Map - Johns Hopkins Coronavirus Resource Center (jhu.edu). (Accessed August 25, 2021).
3. Income Data Sources:
    a. Income data source is the 2018 zipped data for all states, including adjusted gross income (AGI): SOI Tax Stats - Individual Income Tax Statistics - 2018 ZIP Code Data (SOI) | Internal Revenue Service (irs.gov). (Accessed July 7, 2021).
    b. AGI data dictionary: 18zpdoc, IRS website above. (Accessed July 7, 2021).
4. NAICS 2017 industry data from North American Industry Classification System (NAICS) U.S. Census Bureau, including 2-6 digit 2017 NAICS code file in Excel format. (Accessed July 13, 2021).
5. State Level Population Data: State Population Totals: 2010-2019 (census.gov). (Accessed August 30, 2021).
6. Zip Code to City, State mapping: https://edelalon.com/blog/2013/09/zipcode-to-city-state-excel-spreadsheet/. (Accessed August 13, 2021).

7. Eight Occupations Hit Hardest by COVID-19, Occupations Hit Hardest in 2020 by the Pandemic (aarp.org). (Accessed October 29, 2021).

8. Modeling discussions on Generalized Variance Inflation Factor (GVIF) and Variance Inflation Factor (VIF): Which variance inflation factor should I be using: GVIF or $\text{GVIF}^{1/(2*df)}$? (stackexchange.com). (Accessed August 23, 2021).

9. Gender pay gap: 5 Facts About the State of the Gender Pay Gap | U.S. Department of Labor Blog (dol.gov). (Accessed December 12, 2021).

10. Federal Reserve Bank of New York, 2017-report-on-veteran-entrepreneurs-and-capital-access.pdf (newyorkfed.org). (Accessed January 15, 2022).