Understanding factors that lead to derogatory reports on credit history

Introduction

Derogatory reports in a person's credit history can affect their ability to qualify for a credit card, take out loans, and other financial activities. This report aims to understand what factors potentially lead to derogatory reports on a person's credit history. Using data from credit card applications, we will use a statistical model to tease out possible factors that contribute to the number of derogatory reports a person has. We chose a Negative Binomial regression model because it allows for easy interpretation of the model results, has a generally good fit with the data, and addresses the excess amount of applications with zero derogatory marks. Using this model, we determine home ownership, average monthly credit card expenditure, number of active credit accounts, and the number of months living at one's current address are the leading factors related to the number of derogatory reports present in a person's credit history.

Method

We will use a statistical model to better understand which variables are closely related to the number of derogatory reports in a person's credit history. Our choice of model should satisfy three criteria: the model should be easy to interpret the effect a variable has on the number of derogatory reports, the model should handle the excess number of applications with zero derogatory reports, and the model should be a good fit to the data, including that it's a model meant to handle count data as the outcome variable (in this case, the number of derogatory reports).

Before we do any modeling, we'll need to make some initial decisions on which variables and which observations to exclude from the final model. First, given that the age to legally apply for a credit card is 18 years old, we'll drop 7 applications that report an age of less than 18 years. Since this was a relatively small number of observations compared to the total sample size (1,319 applicants), we do not anticipate any negative impact on the results of the final model. Second, we exclude the variable for whether a credit card application was accepted (card) because this was created using all other variables in the dataset and we want to exclude variables that our outcome variable also influences. We also exclude the variable for the ratio of monthly credit card expenditure to yearly income (share) because this information is already present in two other variables (income and expenditure), and we want to avoid having redundant information in our model.

From these data decisions we move on to choose our model between three choices: Poisson Regression (P), Negative Binomial Regression (NB), and a Zero-Inflated Negative Binomial Regression (ZINB). All these choices are useful to model count data (the number of derogatory reports), and the zero-inflated model can further address the excess number of zeros in the outcome variable. In choosing the final model, we use a combination of AIC scores (between the P and NB model) and use the Vuong closeness test across the three models to test for which model has the best fit to the data. We chose to go with the NB model because it was the model that performed the best in terms of the Vuong test, and it performed better than the Poisson model in terms of AIC score (and addresses the overdispersion found in that model). The NB model should also do a fair job of addressing the zeroes in the outcome variable given that it fits the data better than the ZINB model and it provides better interpretation of the coefficient estimates than the ZINB model.

Data and Results

The full data consists of 1,312 credit card applications that contain information on an applicant's age, income level, credit card expenditure, number of dependents, number of active credit cards, number of months living at current address, and whether they are home owners (the full list of variables can be found in Table 2). There are no missing values for any of the variables in the data set. We exclude 7 observations from our model that had an age of less than 18 years old. In the initial data exploration phase, we found expenditure and number of active credit cards to be slightly correlated with number of derogatory reports. In Figure 1 below, we can see the shape of the relationship between these variables.

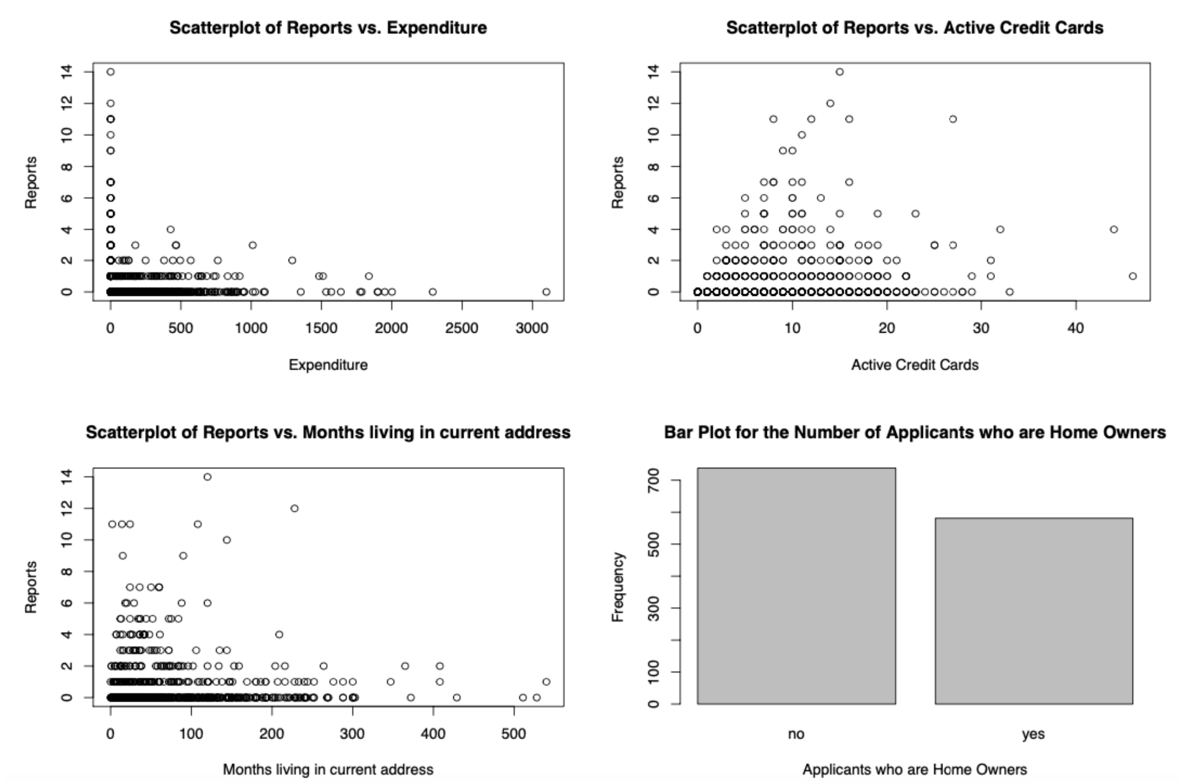

Figure 1: Scatter plots of number of derogatory reports against expenditure, active credit cards, and months living in current address. There is also a bar plot for the number of applicants that are home owners.

The data shows that about 80% of applications contain zero derogatory reports and the rest have about 1 to 14 reports in their credit history, with the most being 1 derogatory report. 7% of applicants are self-employed, 56% own a home, 78% were approved for a credit card, and 81% own other major credit cards. Below in Table 1, the median and IQR are listed for the numeric variables in the data (we chose these summary statistics because these variables were highly skewed).

Variable	Median (IQR)
Age	31.25 (14.00)
Income (Yearly)	2.90 (1.76)
Credit Card Expenditure	101.30 (244.45)
Number of Dependents	1.00 (2.00)
Months Living in Current Address	30.00 (60.00)
Number of Active Credit Cards	6.00 (9.00)

Table 1: Median and IQR for all numeric variables in the data.

Of the three models, we chose the Negative Binomial model for ease of interpretation and because it was the best fitted model using the AIC metric and Vuong closeness test, which test for which models are closer to the true data generating process (this test also performs well when the models are not nested i.e. all model comparisons have the same variables). Table 2 below provides the coefficient estimates of the NB model along with 95% confidence intervals and their corresponding p-values. The model shows that the factors that have a statistically significant ($p < 0.05$) relationship with the number of derogatory reports are home ownership, average monthly credit card expenditure, number of active credit accounts, and the number of months living at one's current address. Specifically, the model shows that for each unit increase in credit card expenditure, derogatory reports decrease by a factor of 1.00; for each unit increase in the number of active credit cards, derogatory reports increase by a factor of 1.13; and for each month living in one's current address, derogatory reports decrease by a factor of 1.00. Similarly, if an applicant was a homeowner, this shows a decrease in derogatory reports by a factor of 0.44.

Variable	Estimate (95% CI)	p-value
Home Owner	0.44 (0.31, 0.62)	< 0.001
Self-Employed	1.03 (0.59, 1.84)	0.911

Owns Major Credit Cards	1.02 (0.70, 1.48)	0.930
Age	1.00 (0.99, 1.02)	0.614
Income (Yearly)	1.09 (0.98, 1.20)	0.096
Credit Card Expenditure	1.00 (1.00, 1.00)	< 0.001
Number of Dependents	1.10 (0.97, 1.25)	0.144
Months Living in Current Address	1.00 (1.00, 1.00)	0.040
Number of Active Credit Cards	1.13 (1.10, 1.16)	< 0.001

Table 2: Coefficient estimates for all variables in the model, as well as their corresponding 95% confidence intervals and p-values.

Conclusion

This report aims to understand what factors could potentially contribute to derogatory reports in a credit card applicant's credit history. After assessing the fit of our data on multiple models, we chose a Negative Binomial regression model since it provided ease of interpretation, good fit to the data, and addressed the excess number of applications with zero derogatory reports. We recognize some possible limitations to our approach such as not taking into account other models that are designed to address excess number of zeros in the response variable. There could also be other factors that contribute to derogatory reports that were not available in our data, such as information on the number of missed payments, bankruptcy, repossessions, and foreclosures. Nonetheless, we believe that the NB model provides useful information on what factors are closely related to the number of derogatory reports. Specifically, the model highlighted leading factors that may contribute to derogatory reports such as home ownership, average monthly credit card expenditure, number of active credit accounts, and the number of months living at one's current address.