FIN3210 Week 2 Assignment Report

Ma Kexuan 120090651

**Abstract**

This report provides a descriptive summary statistic for the dataset given, and construct several regressions to discover how borrower characteristics affect the outcome of default likelihood, the number of bids. Furthermore, we discover the relation between the platform default likelihood and platform characteristics.

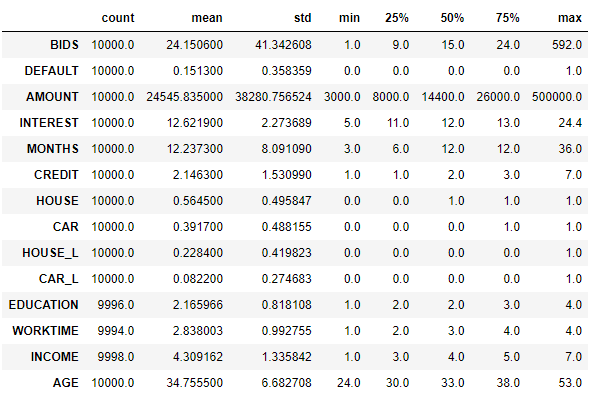
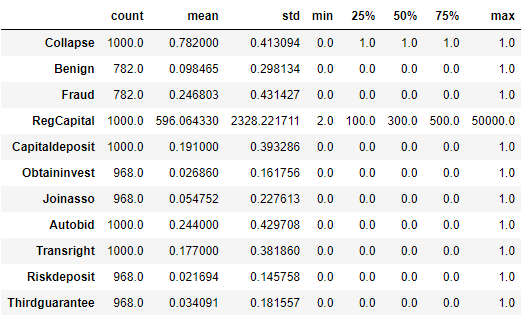
**Data Preprocessing**

The preprocessing procedures and some interpretations of the code are described in each code blocks in the appendix, please check.

**Questions**

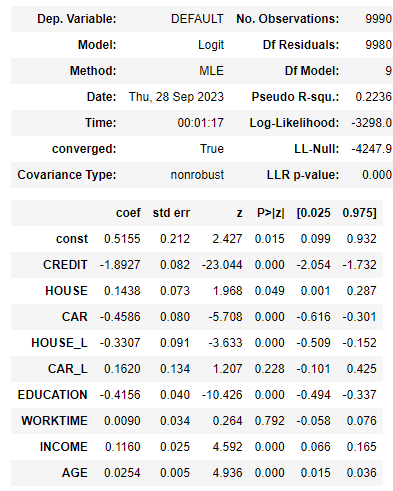
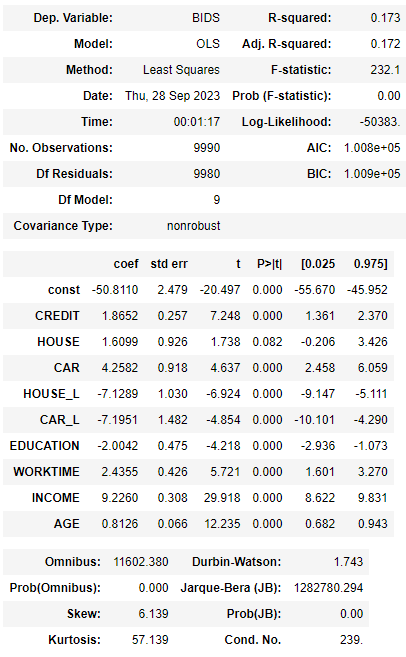
1) Present two tables for the summary statistics of the key variables in Renrendai loans.xlsx and p2p lending platforms.xlsx

In the table of Renrendai loans.xlsx, I choose 'BIDS', 'DEFAULT', 'AMOUNT', 'INTEREST', 'MONTHS','CREDIT', 'HOUSE', 'CAR', 'HOUSE\_L', 'CAR\_L', 'EDUCATION', 'WORKTIME', 'INCOME', 'AGE' as the key variables, the summary statistics is described as the left chart below. It’s worth noting that the mean value of default is 0.15, meaning that about 85% percent of people have been rejected from loaning the money. In the table of p2p lending platforms.xlsx, I choose 'OnlineTime\_YMD', 'Bankrupt\_WDZJ', 'Collapse','Benign', 'Fraud', 'RegCapital', 'Capitaldeposit', 'Obtaininvest', 'Joinasso', 'Autobid', 'Transright','Riskdeposit','Thirdguarantee' as the key variables, the summary statistics is described as the right chart below. The Collapse variable has a mean of 78%, meaning that 78% percent of the platforms have already collapsed, indicating a high risk of default.

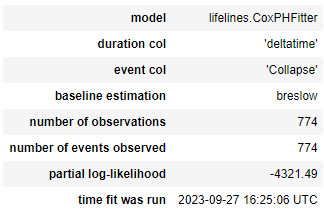
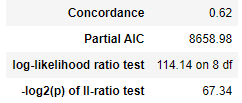
2) Perform a logit regression and examine the relation between the default likelihood and borrower characteristics such as credit, house, car, education, work time, etc.

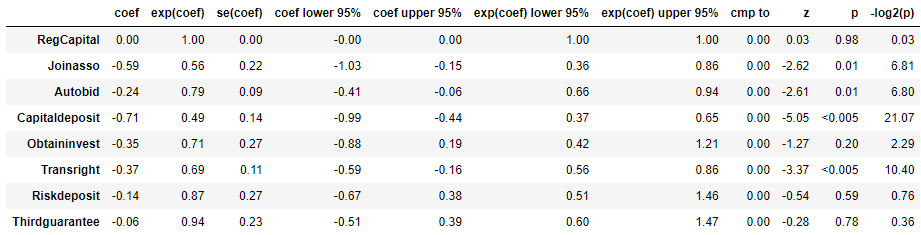
In this Logit regression, I use 'CREDIT', 'HOUSE', 'CAR', 'HOUSE\_L', 'CAR\_L', 'EDUCATION', 'WORKTIME', 'INCOME', 'AGE' as the independent variables, the result at the left below shows that all the variables chosen except for CAR\_L and WORKTIME have 99% significance level. Among the significant coefficients, CREDIT, CAR, HOUSE\_L and EDUCATION are negatively correlated with the dependent variable default. This can show that the platform has an accurate credit rating for users, and the higher the credit, the lower the default risk. Users with cars are likely to have good living conditions and a low probability of default, just as those with mortgages are. Higher education may mean higher quality and lower probability of default. However, for the positive ones, the larger the AGE, the more likelihood to default, since they may not earn enough money. But for the INCOME, it’s quite weird to get the result, to explain it, maybe we should do more research to figure out the logic underneath.

3) Perform an ols regression and examine the relation between the number of bids and borrower characteristics such as credit, house, car, education, work time, etc.

The OLS result is shown on the right above. By observing the p-value of all the independent variables, we find that they are all 99% significant except for HOUSE. For HOUSE\_L and CAR\_L, since there’s a loan on them, there may not be many investors to give money to them. For CREDIT, HOUSE, CAR, WORKTIME, INCOME, AGE, these characteristics describe the social status for certain person, thus if they’re larger, it implies that there should be larger probability for them to give the money back, hence more BIDS for them. Nevertheless, the negative relation between BIDS and EDUCATION is quite elaborate. We may need further research to figure out the abnormal phenomenon.



4) Perform the Cox model (Proportional hazards model) and examine the relation between the platform default (survival) likelihood and platform characteristics such as RegCapital, Joinasso, etc.

RegCapital has slight effect on the likelihood of collapse, the rest of the variables except Riskdeposit and Thirdguarantee have negative relationship with the likelihood of collapse, and the exp(coef) are all smaller than 0.8, indicating that there’s a significant impact. For Riskdeposit and Thirdguarantee, they have little impact on the likelihood. The concordance value is 0.62, indicating the model's ability to distinguish between different survival times. The p-value for the log-likelihood ratio test is very close to zero, reflecting the strong statistical significance of the model.

Below is the appendix for the code: