ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

MSc Economics & Business

Master Specialization Financial Economics

The Fintech Credit Market

Do proprietary credit models of marketplace lending platforms outperform traditional credit ratings? Evidence from Lending Club

Author: K.M.C. Chong

Student number: 437938

Thesis supervisor: dr. Haikun Zhu

Preface and Acknowledgements

I started researching P2P lending when at my summer internship in Suriname it was mentioned as a future project. A P2P lending platform in a developing country could be revolutionary in the way that borrowing and lending works in that country. I quickly learned that is something that developing countries need because of their high number of unbanked people. These unbanked have no access to formal credit. A marketplace lending platform could change this. After finding out I could choose it as a thesis topic I quickly jumped at the chance.

While writing the thesis, I had multiple opportunities come on my path. I started work at a bank in Suriname and started a company on the side. This delayed my thesis, but I was determined to finish it on time. I first would like to thank dr. Haikun Zhu for his guidance. He understood that I needed to finish quickly and helped me with that. His insights on the topic were invaluable. I would also like to thank my parents for giving me the time and freedom to finish my thesis.

NON-PLAGIARISM STATEMENT

By submitting this thesis, the author declares to have written this thesis completely by himself/herself, and not to have used sources or resources other than the ones mentioned. All sources used, quotes and citations that were literally taken from publications, or that were in close accordance with the meaning of those publications, are indicated as such.

COPYRIGHT STATEMENT

Abstract

This research paper researches whether Lending Club's proprietary credit rating model outperforms the traditional FICO credit score. It also researches whether adding more information variables improves the determination of default, showing the informational efficiency of Lending Club's credit model. And lastly, whether the updates Lending Club does to their proprietary credit model improves its grading decisions. This research is mainly done with logistic regressions and a dataset consisting of 1.2 million observations. Findings show that Lending Club's credit rating model does outperform the traditional FICO scores and the addition of borrower information increases the determination of default. Lending Club's update to their credit model in September of 2017 did improve compared to previous models.

Keywords: Fintech; P2P Lending; Information Asymmetry; default determination; P2P credit scoring.

Table of contents

List of tables

1. Introduction

One of the main functions of a bank is to act as an intermediary between borrowers and lenders. Lenders need somewhere to save their money, and borrowers need a place to borrow money from. But with the fast evolution of information technology and the rise of fintech companies, banks are falling more and more behind. These banks cannot move as quickly as these new fintech companies because of their existing legacy systems (Five Degrees, 2018). And switching from their legacy systems is not an easy task. Therefore, banks must compete differently with these fintech companies.

What is fintech actually? PWC states that “Fintech or financial technology describes the evolving intersection of financial services and technology. It refers to startups, tech companies, or even legacy providers” (PWC, 2016). Startups such as Venmo, Adyen, Apple Pay are considered Fintech companies and digital banks such as N26, Monzo, Revolut are also considered Fintech companies. What these companies have in common is that they offer innovation in an arguably stale industry. Other industry experts say that fintech started long before the term was even invented. They describe fintech as any type of technology that makes the financial sector run (Blomstrom, 2018). For this research, the definition of fintech proposed by Scheuffel (2016) will be followed. “Fintech is a new financial industry that applies technology to improve financial activities” (Schueffel, 2016).

Fintech covers a large part of the financial industry. Ranging from mobile payments to Robo-advisory to branchless banks. The area of fintech this research will cover is that of peer-to-peer (P2P) lending platforms. P2P lending platforms are online platforms that match borrowers and lenders. Individuals lending money to other individuals, with the platform only being the intermediary between the parties. P2P lending platforms are entirely online, have lower overhead and can offer their services for a lower price than traditional banks. These platforms use big data analysis and machine learning techniques to faster and more accurately issue loans. Another way they differ from banks is that their loans are not secured. That means the loan is not secured against the borrowers’ assets. Lending Club is an example of a P2P lending platform. Lending Club is the biggest P2P lending platform in the United States. Lending Club issued more than 38 billion USD of loans since its start in 2007, issuing around 8 billion in 2018 alone (Lending Club, 2019). The growth of P2P lending is not only apparent in the USA but also in the rest of the world. Zopa started the first P2P lending platform back in 2005 and is grown to one of the largest P2P lending business in Europe (Beioley & Megaw, 2019). P2P lending is not without its problems. China has been a hotspot for fraudulent platforms. Ezubao was shut down because of its similarities to a Ponzi scheme, stealing more than 50 billion renminbi from its investors (Gough, 2016).

A problem with lending, either online or with banks, is the information asymmetry between the borrowers and lenders. Lenders do not know the true nature of the creditworthiness of the borrowers. Problems such as moral hazard and adverse selection may arise. Banks try to mitigate these problems through several methods, bank's usage of guarantees, regular reporting and certified accounts to strengthen trust in borrowers. P2P online platforms do not have the luxury to do this due to the significant transaction costs, and because its services are only offered online. These P2P lending platforms, therefore, only rely on the information given by the borrower. And using that information they grade these loans according to the platform's algorithm. Traditionally, lenders make use of credit ratings like FICO credit score. The problem with these credit scores is that it merely looks at the borrowers' credit report. FICO is made up of five sources of information, the person's payment history (do you pay on time), utilization (balance to limit ratio), length of credit history, recent activity (new credit application), and credit mix (types of credit taken) (FICO, 2019). This is arguably the most critical information for issuing loans, but a lot of other borrower information is not included in the calculation of the FICO score. Next to the traditional credit scores, the P2P lending platforms use their algorithms to analyze the other data the borrowers provide. These algorithms are proprietary to the platforms, and their exact workings are not known.

This paper will use Lending Club to research the workings of its' credit rating model. Lending Club assigns grades and subgrades to the loans. The grades and subgrades are then used to determine an appropriate interest rate for the loan. From a rational perspective, Lending Club would include as much information as possible in their model to determine theses grades. The expectation then would be that Lending Club's credit model would be better than traditional credit ratings. This research will show whether this is the case. **Therefore, the research question is,** "Does Lending Club's proprietary credit rating model outperform traditional credit ratings like FICO score?''. Answering this research question will only give insight into the grades the credit model computes versus the traditional credit scores. Further research can be done on the informational efficiency of these grades. If Lending Club's credit rating model were perfect, all borrower information would be incorporated in it, and lenders could use only that credit rating to make investment decisions. Past research shows that additional information variables next to Lending Club’s credit rating improve its predictability of default. For that reason, the information efficiency of Lending Club will also be researched.

The third topic of research is the improvement of Lending Club’s proprietary credit model. These models are kept a secret, and when improvements arise, they usually are not advertised. Lending Club released a statement that their credit model was updated in September 2017 (Lending Club, 2019). Whether this update is indeed an improvement over the older versions will also be empirically researched.

Past research papers on the topic of P2P lending platforms mostly focus on the determinants of default. This research first looks at the efficiency of the credit rating model of Lending Club. There is also more information available. Past papers that were written in 2015 only have a few years of data they can use. This research will use more than ten years of data. The longer timeframe takes into account the changes in the economic environment, which in turn improves the empirical results.

The remainder of this paper is organized as follows. Chapter 2 presents an empirical literature review on P2P lending. It is split into P2P lending in general and following that the hypotheses development is described. Part 3 presents the data and methodology section, explaining the data used and the way the research will be done. Part 4 lists the results. And finally, the paper is closed off with a conclusion and ideas for further research.

2. Theoretical Framework

Research on P2P lending can be categorized into three topics. First, the emergence of P2P lending platforms and the reasons for it. The second topic lays the focus on the performance of online P2P loan portfolios. Lastly, researchers focus on the determinants of funding success and loan default. The focus of this research is partly on the last topic, more specifically, the determinants of default. Past research on determinants of default will be used to answer the part of the research question. The chapter is built up as follows. First research on P2P lending, in general, is listed, following that previous studies on the determinants of default will be described, and the relation to this research will be done. The essential papers on the topic will be used to support the hypotheses needed to answer the research questions.

2.1 Lending Club's Lending Process

Lending Club's loans are only issued online. The borrower can choose between 36- and 60-month loans. The borrower adds its information and finances to the loan application, which Lending Club processes. Lending Club assigns a grade and subgrade to the borrower with the corresponding interest rate. Lenders can then find the loan listed on the website and can then invest in a portion of the loan. Verification of the borrower's finances is done simultaneously to the funding process. If the loan is fully funded by lenders before that, the finances do not have to be verified. If the borrower's information cannot be verified, the loan application is cancelled.

2.2 P2P Lending

Research done by Bachmann et al. (2011) summarizes a literature overview of online P2P lending. This paper is used to get an overview of the history of P2P lending. The first P2P lending platforms started in 2005 and is thus a relatively new research field. In the past, P2P lending worked differently than it does now. Websites like Prosper.com had a more social networking element attached to the loan issuance. The interest rates were set through a Dutch auction. Borrowers had the option to add pictures to their loan application. Race, perceived happiness, military involvement were significant predictors of default (Pope & Sydnor, 2011). Now with SEC regulation, P2P lending platforms operate differently. Now instead of the Dutch auction method, P2P platforms accepts and evaluates the borrowers themselves. Researchers claim that P2P lending platforms are an alternative to traditional banking and not something completely new. That is, there are several similarities between P2P platforms and traditional banks. Both facilitate the supply and demand for money, can be classified as financial intermediaries and bear transaction costs. The difference occurs in other costs. P2P platforms are not subject to the same banking regulations, lowering their overall costs. Lower costs mean that borrowers benefit from lower interest rates and higher revenues for lenders. The interest margin namely comes from operating costs (Demirgüç-Kunt & Huizinga, 1999). Käfer (2018) compares P2P lending platforms to 'shadow banks' because they are riskier than traditional banking. The lenders run significant risks. They bear the credit risk of these loans. The P2P platforms just act as an intermediary. The loans do not require collateral, so if a loan defaults, the lender is not protected by collateral. The type of borrowers also differs from traditional banks. They are mostly the borrowers already rejected by banks (De Roure, Pelizzon, & Tasca, 2016). These underserved borrowers are willing to accept higher interest rates offered by the P2P lending platforms. Their loan amount required is also lower and for risky purposes, which banks typically avoid (Tang, 2019).

P2P lending platforms have the same fundamental problem the lending industry has, information asymmetry between the borrowers and the lenders. Information asymmetry problems appear because borrowers have more information about their ability and willingness to repay. Lenders are the group that is at a disadvantage. The P2P platforms try to decrease this asymmetry by collecting as much information from the borrower. Lending Club offers detailed loan information on active loans and historical information on past loans. Borrowers can use this information to better their loan selection.

In the next section, research on the determination of default topic of P2P lending will be reviewed, and hypotheses will be developed.

2.2 Determinants of Default

Multiple research on the topic of determinants of default in P2P lending platforms has been done. Information provided by the borrowers should be analyzed if they have any added significance over Lending Club's credit model. In table 1, an overview of research done on LendingClub data is listed.

Table 1: Summary of research on determinants of default

Study	*Data*	*Methodology*	*Determinants of Default*
Serrano-Cinca et al. (2015)	Loans: Jan 2008 to Dec 2011. Term: 36 Months	Survival Analysis (Cox regressions) and logistic regression	Annual income, credit grade, credit history length, debt-to-income ratio, delinquency past 2 years, homeownership, inquiries in last 6 months, loan purpose, open credit lines, revolving credit line utilization
Emekter et al. (2015)	Loans: May 2007 till June 2012. Term: 36 & 60 Months	Binary logistic regression	Credit grade, debt-to-income ratio, FICO, revolving credit utilization
Carmichael (2014)	Loans: June 2007 till Nov 2013. Term: 36 & 60 Months	Dynamic logistic regression	Annual income, credit grade, credit history length, FICO, inquiries in last 6 months, loan amount, loan description, loan purpose, months since last delinquency, revolving credit utilization, unemployment level, subgrade

These three papers all focused on the determinants of default. Serrano-Cinca, Gutierrez-Nieto, and Lopez-Palacios (2015) find that the subgrade assigned by Lending Club (based on a FICO credit score and other variables) is the most important variable in reducing the information asymmetry suffered by the lender. Other borrower characteristics are also a significant factor in the determination of default. Emekter, Tu, Jirasakuldech, and Lu (2015) find that not only the credit grade is significant, but also FICO score, debt to income ratio, revolving line utilization are important. Carmichael (2014) using a dynamic logistic regression, found that FICO score, credit inquiries, income and loan purpose are significant in determining default. He uses his models to forecast default and finds that it outperforms Lending Club subgrades. Meaning that there is more efficiency to be found for Lending Club's credit model. Funding success of loans is also a thoroughly researched topic. Lee and Lee (2012) find that herding behaviour is present in P2P loan funding and increase information asymmetry. Lin, Prabhala, and Viswanathan (2009) research whether borrower's online friendships increase their chances of funding success and find that online friendships are associated with lower ex-post default rates.

2.3 Hypotheses Development

With the borrowers' information provided, Lending Club assign grades to each loan it issues. If the grades are correctly determined, high-risk loans will receive a higher grade, and low risk loans a lower grade. Lenders must be compensated for this additional risk and will ask a premium over the risk-free rate. It is therefore important that these loan grades are correctly given. Either borrowers will be given an unfair higher interest, or lenders may invest in loans that have a higher risk than they are classified as. On the Lending Club website, it is stated that not only credit score (FICO) is considered when determining a loan grade, but also a combination of several indicators of credit risk from the credit report and loan application (Lending Club, 2019). From this statement it can be interpreted that FICO score is mainly used in the loan grade calculation. Meaning that Lending Club grade not only has FICO processed in the grade but also other variables. A comparison between FICO score and Lending Club grades is expected to show that the subgrades have better predictability of default. Subgrade is used because it is a more detailed score given to the loan. The first hypothesis is therefore as follows:

Hypothesis 1: Subgrade given by Lending Club to loans has significantly better predictability of default than the FICO scores.

LendingClub gathers a lot of information from the borrowers. Research shows that not all these information variables are equally important. Some variables carry more weight in the determination of default. Therefore, a deeper dive must be taken in the determinants of default. Hypothesis 2 will be split into four parts, each part analyzing different determinants. First is the loan characteristics (loan purpose, loan amount, loan term), second the borrower characteristics (annual income, housing situation, employment length), third the borrowers credit history (delinquency last 2 years, open accounts, revolving utilization etc.) and lastly the borrowers indebtedness (loan amount to annual income, annual installment to income, debt to income). The information variables are split to give a better overview of which variables are significant and which are not. Past research on the topic of determinants of default finds that most of the variables used next to subgrade, are significant. These hypotheses will show whether lenders should include more information into their investment decision than just the grades given by Lending Club. These hypotheses test whether the variables have a significant effect next to subgrade.

Hypothesis 2a: The addition of loan characteristics to subgrade significantly improves the determination of default.

$$Default = \beta_0 + \beta_1 Sub\ Grade + \beta_2 Loan\ Amount + \beta_3 Loan\ Term + \beta_{4-17} Loan\ Purpose + \varepsilon_i$$

Hypothesis 2b: The addition of borrower's characteristics to subgrade significantly improves the determination of default.

$$Default = \beta_0 + \beta_1 Subgrade + \beta_2 Annual\ Income + \beta_3 Employment\ Length + \beta_{4-7} Housing\ Situation + \varepsilon_i$$

Hypothesis 2c: The addition of borrower's credit history to subgrade significantly improves the determination of default.

$$Default = \beta_0 + \beta_1 Subgrade + \beta_2 Delinquency\ 2\ Years + \beta_3 Inquiries\ 6\ months + \beta_4 Public\ records + \beta_5 Revolving\ utilization + \beta_6 Open\ Accounts + \varepsilon_i$$

Hypothesis 2d: The addition of borrower's indebtedness next to subgrade significantly improves the determination of default.

$$Default = \beta_0 + \beta_1 Subgrade + \beta_2 Loan\ Amount\ to\ Annual\ Income + \beta_3 Annual\ Installment\ to\ Income + \beta_4 Debt\ to\ Income + \varepsilon_i$$

3. Data and Methodology

Lending Club had its data on loan applications and funded loans freely available on its website. That is not the case anymore. Although the data is available an account is required now, which only US residents can register for. The dataset was accessed in July 2019 when it was not required to have an account with Lending Club. The final dataset containing 1.256.558 observations is described in this chapter.

3.1 Data Cleanup

The original dataset had 2.2 million observations starting from June 2007 until December of 2018 and only consists of successfully funded loans. Not all this data is used. Loans that are still current are dropped from the dataset as these cannot be used for the analysis. There are 151 variables in the dataset with most of them being empty. Variables with a large number of empty observations are dropped. Literature listed in chapter 2 are used to create a variable list with the most important variables. These variables are kept, and the other variables are then dropped. With the selected variables further cleanup is needed. Loan status is broken down in nine types. Only loans that are in default or successful are used. Therefore, loans in default must be defined. For this research loans that are described as charged off, default, late (31-120 days) are considered in default. Loans that are late (31-120 days) are considered in default because around 74% are eventually charged off (Lending Club, 2019). Loans classified as Current, In Grace Period, and Late (16-30 days) are removed from the dataset, they are considered in progress loans. Finally, loans that do not have all observations for the selected variables are dropped from the dataset. The final dataset contains 1.256.558 observations with 19 variables. A description of the variables is given in the next section.

3.2 Variable Explanation

Variables used in this research are explained below. The dataset is broken down into borrower assessment, loan characteristics, borrower characteristics, credit history, and borrower indebtedness. The borrower filled in the information of these variables at the time of application.

Table 2: Explanation of the variables

Variable	*Definition*
Borrower Assessment	
Grade	LendingClub categorizes borrowers in seven different loan grades Ranging from A-G
Subgrade	Subgrade ranging from A1-G5.
FICO	The measure of consumer credit risk. Ranging from 610 to 845.
Interest Rate	The interest rate on the loan paid by the borrower
Loan Characteristic	
Loan Purpose	The loans are broken down in 14 types of purposes: Debt consolidation, credit card, home improvement, car, educational, house, major purchase, medical, moving, renewable energy, small business, vacation, wedding and other.
Loan Amount	Loan amount requested by the borrower
Loan Term	36 months or 60 months
Borrower Characteristics	
Annual Income	Annual income as stated by the borrower
Housing Situation	Own, rent, mortgage, and other
Employment Length	The length the borrower has been with its current employer
Credit History	
Delinquency 2 Years	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
Inquiries Last 6 Months	The number of inquiries by creditors during the past 6 months
Public Records	Number of derogatory public records
Revolving Utilization	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit
Open Accounts	The number of open credit lines in the borrower's credit file
Borrower Indebtedness	
Loan Amount to Annual Income	Loan amount to the annual income
Annual Installment to Income	The annual payment owed by the borrower divided by the annual income provided by the borrower during registration
Debt to Income	Borrower's debt-to-income ratio. Monthly payments on the total debt obligations, excluding mortgage, divided by self-reported monthly income

3.3 Methodology

This part will describe in detail how the hypotheses are tested. The statistical software STATA 15 is used for the analysis. Before any research is done, the data must first be described. The variables are split into discrete and continuous variables. The variables loan purpose, housing situation, and loan term are transformed into discrete variables. The hypotheses will be researched with a binary logistic regression. As the variable for default is either 0 or 1 (0 = fully paid, 1 = defaulted), logistic regression is a better fit than an OLS regression.

Binary logistic regression has several significant assumptions that need to be complied with:

- Dependent variable should be dichotomous
- No outliers should be present
- No multicollinearity among the independent variables (Tabachnick and Fidell (2013) suggest a correlation of less than 0.90.)

In the regressions run these assumptions are complied with.

Hypothesis 1 is researched by running 3 logistic regressions. $Default = \beta_0 + \beta_1 Sub\ Grade + \varepsilon_i$ (default against subgrade), $Default = \beta_0 + \beta_1 FICO + \varepsilon_i$ (default against FICO), and $Default = \beta_0 + \beta_1 Sub\ Grade + \beta_2 FICO + \varepsilon_i$ (default against subgrade and FICO). To compare the coefficients of subgrade and FICO, they first must be adjusted for scale. Their distribution differs so cannot be directly compared. The adjustment for scale is made by multiplying the coefficients against its own standard deviation. In for the last regression, the expectation is that FICO is insignificant next to subgrade. If that is the case, FICO will be shown in the regression as not significant.

Interpreting logistic regression coefficients is different from that of linear regressions. There is a difference in interpretation between continuous and discrete variables. The betas of the continuous variables can be transformed into odds by taking the exponential of the beta. If the odds ratio is above 1, then the beta can be interpreted as for each additional increase of that variable, the chance of, in this case, default, goes up by (odds ratio – 1) per cent. If the odds ratio is below 1, the beta can be interpreted as for each additional increase of the variable, the chance of default goes down by (1 – odds ratio) per cent. For discrete variables, the betas can be transformed into the odds ratio by taking the exponential of the beta. The odds ratio can then be interpreted as: this discrete variable has x times the chance of default than not the discrete variable (Tabachnick & Fidell, 2013).

Hypothesis 2 is analyzed by running 4 logistic regressions. Each regression is run with the respective information variables. The output of the regressions will show whether the variables are indeed significant next to subgrade. The coefficients are also used to interpret the effect of the variables.

To answer whether the credit rating model of Lending Club improved over time, the dataset must be split into two parts. One part 'after' the implementation of the fifth-generation credit model and one 'before' that. The new credit model was implemented starting 8th September 2017. Not all loans were immediately classified using the new model. Therefore, loan data from September 2017 is not used. The part of the data that is 'after' the implementation runs from October 2017 until the end of the dataset (December 2018). This part consists of 86.056 observations. The 'before' part of the dataset runs from June 2012 until August 2017 (again does not include September 2017). This part consists of 1.102.653 observations. Data from 2007 until June 2012 is not included in this part of the data. The recession in the US started December 2007 and officially ended June 2009 (Public Information Office, 2010). Loans issued in this period could have a 'wrong' subgrade given to the loan because of the economic conditions during that time. Loans issued in June of 2009 matured in June 2012 (only 36-month loans were issued back then). Therefore, loan data starting in June 2012 are used. The question states that the 'before' and 'after' explanatory power of subgrade must be compared. Lending Club claims that this fifth-generation model improves loan grade determination. This is tested by running two logistic regression. Default is the dependent variable and subgrade the independent variable in both regressions. The two coefficients of subgrade are then compared to see if they are significantly different. This is tested using an F test. If this hypothesis is rejected, it can be said that the coefficients a significantly different.

Direct comparison of the coefficients will not give any useful information. To test whether subgrade improved after the implementation of the new credit model, the explanatory power of FICO is tested before and after the implementation. If the explanatory power of FICO goes down after implementation of the new credit model, the new model has improved. Showing that the new credit model captures more information about the FICO score than before. The coefficients of subgrade and FICO are multiplied by its standard deviation. This is done to adjust for scale. This is done twice, once before the implementation and once after the implementation.

The following section lists the results of the tested hypotheses.

4. Results

This section goes over the results. The hypotheses are accepted or rejected here. Results are described. First, the data is described using tables and figures. The binary logistic regressions are then explained, and the hypotheses answered.

4.1 Data Description

Table 3 breaks down the fully paid and defaulted loans per grade, loan purpose, and housing situation. The default percentage goes up as the grade lowers. This hints towards the fact that the grades given by LendingClub do lower the asymmetric information between borrowers and lenders. The largest percentage of loans issued have grade B, and C. Small business as loan purpose also has a higher default rate.

Table 3: Exploratory study of the discrete variables

Loan Status %			
Variables	Fully Paid	Default	% of the total
Grade			
A	93.76%	6.24%	17.70 (222.388)
B	86.04%	13.96%	29.42 (369.631)
C	76.66%	23.34%	28.37 (356.518)
D	68.76%	31.24%	14.66 (184.153)
E	60.94%	39.06%	6.82 (85.697)
F	54.80%	45.20%	2.35 (29.563)
G	49.99%	50.01%	0.69 (8.608)
Loan Purpose			
Car	84.91%	15.09%	1.10 (13.793)
Credit Card	82.73%	17.27%	21.86 (274.686)
Debt Consolidation	78.39%	21.61%	58.06 (729.610)
Educational	79.71%	20.29%	0.03 (409)
Home Improvement	81.56%	18.44%	6.46 (81.144)
House	76.63%	23.37%	0.55 (6.912)
Major Purchase	80.41%	19.59%	2.22 (27.928)
Medical	78.07%	21.93%	1.12 (14.131)
Moving	76.13%	23.87%	0.71 (8.897)
Other	77.97%	22.03%	5.75 (72.303)
Renewable Energy	75.00%	25.00%	0.07 (872)
Small Business	68.89%	31.11%	1.21 (15.232)
Vacation	80.01%	19.99%	0.66 (8.328)
Wedding	87.78%	12.22%	0.18 (2.316)
Housing Situation			
Mortgage	82.25%	17.75%	49.59 (623.074)
Own	78.99%	21.01%	10.15 (127.490)
Rent	76.29%	23.71%	40.23 (505.496)
Other	78.31%	21.69%	0.04 (498)

Total loans analyzed: 1.256.558. Fully Paid: 999.179 (79.52%). Defaulted: 257.379 (20.48%).

Table 17 and 18 in Appendix I contain the Pearson correlation matrix of both the continuous and discrete variables. These matrices are created to give a quick overview of the relationships between the variables. They can also be used to find multicollinearity between the variables. In the correlation matrix for continuous variables, an important positive correlation (0, 9752) is between the subgrade and the interest rate. The higher the subgrade (lower rating), the higher the interest rate on the loan and vice versa. The correlation between the other variables is relatively low. The other variable that has a moderate correlation with subgrade is the FICO score. The negative relation (-0,426 with subgrade and -0,409 with interest rate) is to be expected. Higher credit grade borrowers receive lower subgrades (better grades) and lower interest rates. Subgrade and interest rate are both positively correlated with loan status. The higher the grade and interest rate, the higher the chance of default. Looking at the correlation matrix of the discrete variables, loan term and subgrade have a positive correlation. The correlation of 0.439 implies that the higher grade (worse loans) have is positively correlated with a higher loan term (60 months). Out of the loan purposes, small business and debt consolidation have a higher correlation with default. Although the correlations are weak, they are the highest between the loan purposes. This relationship also is seen between the small business, debt consolidation, and subgrade. A higher (worse) subgrade is given to those loan purposes. These matrices can give a fast overview of the variables that affect loan status, subgrade or the interest rate.

Table 4 contains the result of an exploratory study on the continuous variables. The average interest rate on defaulted loans is higher than that of the fully paid loans. Higher interest is given to riskier loans, again showing that the grading system of Lending Club lowers the information asymmetry.

Table 4: Exploratory study of the continuous variables

Variables	*All*		*Fully Paid*		*Defaulted*	
	Mean	St dev	Mean	St dev	Mean	St dev
Borrower Assessment						
Interest Rate	0.132	0.048	0.126	0.045	0.157	0.492
Loan Characteristic						
Loan Amount	14625.39	8763.01	14282.43	8691.68	15956.81	8910.38
Borrower Characteristics						
Annual Income	78486.51	72078.12	79717.79	72562.55	73706.53	69960.90
Employment Length	4.605	3.176	4.592	3.169	4.653	3.200
Credit History						
Delinquency 2 Years	0.322	0.885	0.313	0.867	0.359	0.951
Inquiries Last 6 Months	0.667	0.965	0.635	0.940	0.791	1.048
Public Record	0.210	0.602	0.202	0.584	0.241	0.668
Revolving Utilization	0.519	0.245	0.512	0.246	0.545	0.240
Open Accounts	11.61	5.471	11.53	5.418	11.93	5.66
Borrower Indebtedness						
Loan Amount to Annual Income	0.211	0.116	0.203	0.112	0.242	0.122
Annual Instalment to Income	0.006	0.004	0.006	0.003	0.007	0.004
Debt to Income	0.176	0.792	0.172	0.078	0.191	0.080

4.2 Regression Results: Hypothesis 1

The first hypothesis is answered with three binary logistic regressions. The regressions have default as the dependent variable. For the first part of the hypothesis, two separate logistic regressions are executed. One with subgrade as the independent variables and the other with FICO score. To finally answer the hypothesis, the two independent variables are added to the regression together. The results explained below.

Hypothesis 1: Subgrade given by Lending Club to loans has significantly better predictability of default than the FICO scores.

$$Default = \beta_0 + \beta_1 Subgrade + \varepsilon_i$$

Table 5 can be interpreted as such. The odds ratio equals exp (0.1001042) = 1.1053. Meaning for each additional subgrade the chance of default goes up with 10.53 per cent. This is in line with the expectation; a higher subgrade means that the loan is riskier, and the chance of default goes up.

Table 5: Logistic regression for default and subgrade

Dependent variable: Default				
Indicator	Coefficient	Standard Error	Z score	Probability
Subgrade	0.1001042	0.000351	285.21	0.000*

**Significant at the 1% level*

$$Default = \beta_0 + \beta_1 FICO + \varepsilon_i$$

Table 6 can be interpreted in the same way. The difference here is that the coefficient is negative, so the odds ratio is below 1. The odds ratio of the FICO coefficient is exp (FICO) = 0.9880. Meaning for each additional FICO score, the chance of default goes down by 1.20 (1-0.9880) per cent.

Table 6: Logistic regression for default and FICO score

Dependent variable: Default				
Indicator	Coefficient	Standard Error	Z score	Probability
FICO	-0.0120758	0.0000834	-144.73	0.000*

*Significant at the 1% level

$$Default = \beta_0 + \beta_1 Subgrade + \beta_2 FICO + \varepsilon_i$$

As stated in the methodology section, this hypothesis will be researched by adding both subgrade and FICO into logistic regression. The expectation is that FICO will not be significant because subgrade contains all the FICO information and more. The results in table 7 show that this is not the case. Both subgrade and FICO are significant at the 1% level. The FICO score variable is still a significant determinant of default. Hypothesis 1 is, therefore rejected. Lending Club does not use all of FICO in its determination of the loan subgrades. Odds ratio of subgrade equals exp (0.0941) = 1.098. Meaning that for an additional subgrade, chance of default goes up by 9.8%. Odds ratio of FICO equals exp (-0.0039) = 0.9961. Meaning for an additional FICO scores the chance of default goes down by 0.39%.

Table 7: Logistic Regression for subgrade and FICO score

Dependent variable: Default				
Indicator	**Coefficient**	**Standard Error**	**Z score**	**Probability**
Subgrade	0.0941359	0.000376	250.34	0.000*
FICO	-0.0039259	0.0000898	-43.72	0.000*

*Significant at the 1% level

Comparing the coefficients when adjusted for scale, subgrade does have higher explanatory power than FICO (table 8).

Table 8: Comparison of coefficient adjusted for scale

Variable	***Coefficient***	***Standard Deviation***	***Result (Absolute)***
Subgrade	0.0941359	6.440263	0.60626
FICO	-0.0039259	31.7078	0.12449

4.3 Regression Results: Hypothesis 2

Hypothesis 2 is also answered by logistic regressions with default as the dependent variable. The regressions contain subgrade and the other variables according to their divisions.

Hypothesis 2a: The addition of loan characteristics to subgrade significantly improves the determination of default.

$$Default = \beta_0 + \beta_1 Subgrade + \beta_2 Loan\ Amount + \beta_3 Loan\ Term + \beta_{4-17} Loan\ Purpose + \varepsilon_i$$

The results from table 9 imply that hypothesis 2a can be accepted. All variables that fall under loan characteristics (loan amount, loan term, and loan purpose) are significant at the 1% level. Determination of default can thus be improved by the addition of loan characteristics. As expected, acquiring a loan for small business purposes has the highest chance of default than other purposes. The odds ratio of LP: Small Business equals exp (0.9969) = 2.71. The chance of default is 2.71 higher chance than other purposes. Economically this can be explained by the large risks associated with small business ownership compared to the other loan purposes. The lowest chance of default among loan purposes is that of car financing. The chance of default 1.74 times that of non-car loans. A longer loan of 60 months also has a higher chance of default than a shorter loan of 36 months of 1.48 times.

Table 9: Logistic regression with the loan characteristics

Dependent variable: Default				
Indicator	Coefficient	Standard Error	Z score	Probability
Subgrade	0.0871	0.0004	219.42	0.000*
Loan Amount	2.48e-06	2.86e-07	8.65	0.000*
Term	0.3911	0.0058	67.42	0.000*
LP: Car	0.5536	0.0702	7.88	0.000*
LP: Credit Card	0.7371	0.0660	11.17	0.000*
LP: Debt Consolidation	0.7643	0.0659	11.61	0.000*
LP: Educational	0.8205	0.1438	5.71	0.000*
LP: Home Improvement	0.6490	0.0665	9.76	0.000*
LP: House	0.6255	0.0723	8.65	0.000*
LP: Major Purchase	0.7646	0.0676	11.30	0.000*
LP: Medical	0.8047	0.0691	11.65	0.000*
LP: Moving	0.8024	0.0707	11.36	0.000*
LP: Other	0.7290	0.0664	10.97	0.000*
LP: Renewable Energy	0.8298	0.1050	7.90	0.000*
LP: Small Business	0.9969	0.0683	14.59	0.000*
LP: Vacation	0.7813	0.0716	10.92	0.000*
LP: Wedding	Omitted			

*Significant at the 1% level

Hypothesis 2b: The addition of borrower's characteristics to subgrade significantly improves the determination of default.

$$Default = \beta_0 + \beta_1 Subgrade + \beta_2 Annual\ Income + \beta_3 Employment\ Length + \beta_{4-7} Housing\ Situation + \varepsilon_i$$

Hypothesis 2b can partly be accepted. Running the initial logistic regression with all the borrower characteristics, housing situation: other, is not significant. Dropping this variable out of the regression analysis finds that the other housing situation variables are also insignificant. Only subgrade annual income and employment length stay significant. Therefore, hypothesis 2b is only partly accepted.

Table 10: Logistic regression with borrower characteristics

Dependent variable: Default				
Indicator	**Coefficient**	**Standard Error**	**Z score**	**Probability**
Subgrade	0.0990	0.0004	280.56	0.000*
Annual Income	-8.10e-07	4.86e-08	-16.68	0.000*
Employment Length	0.0021	0.0007	2.96	0.003*
H: Mortgage	-0.1598	0.0080	-20.08	0.000*
H: Rent	0.1219	0.0079	15.35	0.000*
H: Other	-0.0549	0.1143	-0.48	0.631
H: Own	Omitted			

*Significant at the 1% level

Hypothesis 2c: The addition of borrower's credit history to subgrade significantly improves the determination of default.

$$Default = \beta_0 + \beta_1 Subgrade + \beta_2 Delinquency\ 2\ Years + \beta_3 Inquiries\ 6\ months + \beta_4 Public\ records + \beta_5 Revolving\ utilization + \beta_6 Open\ Accounts + \varepsilon_i$$

When adding borrower's credit history, two variables are not significant; those are inquiries in the last 6 months and revolving utilization. Hypothesis 2c can, therefore only be partially accepted. An additional delinquency in the last 2 years increases the chance of default by 2.1% which is in line with the expectations.

Table 11: Logistic regression with borrower's credit history

Dependent variable: Default				
Indicator	Coefficient	Standard Error	Z score	Probability
Subgrade	0.0995	0.0004	269.16	0.000*
Delinquency 2 Years	0.0211	0.0025	8.59	0.000*
Inquiries Last 6 Months	0.0042	0.0023	1.80	0.071
Public Records	0.0518	0.0036	14.31	0.000*
Revolving Utilization	0.0185	0.0099	1.87	0.062
Open Accounts	0.0133	0.0004	31.85	0.000*

*Significant at the 1% level

Hypothesis 2d: The addition of borrower's indebtedness next to subgrade significantly improves the determination of default.

$$Default = \beta_0 + \beta_1 Subgrade + \beta_2 Loan\ Amount\ to\ Annual\ Income$$
$$+ \beta_3 Annual\ Installment\ to\ Income + \beta_4 Debt\ to\ Income + \varepsilon_i$$

All variables in table 12 are significant. Hypothesis 2d is therefore accepted. The coefficient of annual instalment to income is large because of the small values this variable has. Every per cent increase to debt to income increases the chance of default by around 3.9%. This is in line with the expected result.

Table 12: Logistic regression with borrower's indebtedness added

Dependent variable: Default				
Indicator	Coefficient	Standard Error	Z score	Probability
Subgrade	0.0926	0.0004	254.36	0.000*
Loan Amount to Annual Income	4.1115	0.0548	75.03	0.000*
Annual Instalment to Income	-94.8880	1.7971	-52.80	0.000*
Debt to Income	1.6063	0.0298	53.85	0.000*

*Significant at the 1% level

Summing up the findings of hypothesis 2 leads to the conclusion that the addition of variables does significantly improve the determination of default.

4.4 Lending Club Credit Model Update

Table 13 and 14 contain the logistic regression results of subgrade before and after the implementation of the new credit model. Just analyzing the coefficients of subgrade does not give a clear answer if these are different. An F-test must be done to research if there is a significant difference. Table 15 lists these results. There is a significant difference between the two subgrades at the 5 per cent level. Meaning there is an increase or decrease in the explanatory power of subgrade.

Table 13: Logistic regression. Before data sample

Dependent variable: Default				
Indicator	Coefficient	Standard Error	Z score	Probability
Subgrade	0.1020454	0.000378	269.99	0.000*

*Significant at the 1% level

Table 14: Logistic regression. After data sample

Dependent variable: Default				
Indicator	Coefficient	Standard Error	Z score	Probability
Subgrade	0.0987349	0.0013207	74.76	0.000*

*Significant at the 1% level

Table 15: F Test on the different subgrades

F – Test: 'before' subgrade = 'after' subgrade	
chi2 (1)	6.20
Prob > chi2	0.0127*

*Significant at the 5% level

Table 16 compares the explanatory power of subgrade and FICO before and after the implementation of the new credit model. Looking at the before part, FICO still has a large explanatory power compared to the after part. The explanatory power of FICO score goes down to 0.0662 from 0.142. From this, it can be concluded that the new credit model of Lending Club captures a larger part of FICO score information. This means that the new credit model improved the determination of subgrade for loans if we assume that Lending Club wants to incorporate as much information as possible in its determination of the grades.

Table 16: Comparison of the explanatory power of subgrade and FICO

Variable	Coefficient	Standard Deviation	Result (Absolute)
Subgrade before	0.0957322	6.412008	0.6138356
FICO before	-0.0046475	30.65203	0.1393901
Subgrade after	0.0944947	6.31062	0.5963201
FICO after	-0.0017902	37.00265	0.0662421

5. Conclusion

P2P lending platforms have enjoyed a large growth in the past years. As borrowers underserved by banks flock to these platforms, investors also have an increasing interest in these platforms. This paper researches whether the credit rating that these P2P platforms assign the loans are better compared to traditional loans. This is done by answering two hypotheses. The hypotheses together will give a sound answer to the research question.

Hypothesis 1 test whether subgrade has more information incorporated than the traditional FICO score, and whether FICO would be insignificant in determining default because subgrade has all that information already incorporated. The results show that subgrade does, in fact, have a better explanatory power when determining default compared to FICO score. But when both variables are added into a logistic regression model, FICO is still significant. It still has a significant explanatory power next to subgrade. This means that Lending Club does not fully incorporate FICO score information into their subgrade. The next part investigated the significance of the other borrower information. Variables were, according to their classification, separately added into a logistic regression with subgrade. Almost all the variables were significant. Which was in line with past research. Addition of variables next to subgrade bettered the determination of default. The last part of the research looked at the improvement of the Lending Club's credit model. The new credit model Lending Club introduced in September 2017 shows improvement over the older model in terms of including FICO score information.

With the research on the hypotheses complete, the research question can be answered. Research question goes as follows: "Does Lending Club's proprietary credit rating model outperform traditional credit ratings like FICO score?" The credit ratings Lending Club assigns the loans does indeed outperform the traditional FICO credit rating. This does not mean that it captures all information the FICO score has. FICO score is still significant when determining default. When predicting default, not only the subgrade should be considered, but other borrowers' information should also be added to the model. Lending Club updates its credit rating model over time. A comparison between old and new model shows that the credit rating model of Lending Club does improve. Although the Lending Club credit model does not capture all borrower' information completely, this does not mean that Lending Club's determination of its loan grades is wrong. From the data, it can clearly be seen that the loan grade does lower information asymmetry between the borrowers and lenders. It can be used as a good starting point when choosing loans to invest in. Lending Club's credit model is a secret; the process of establishing the grades is not known. It might be that all information is incorporated, but the weights given are different. Or that only certain information is considered. As long as their algorithm to determine loan grades is not public, a concrete conclusion cannot be drawn.

6. Further Research

In follow-up research, the prediction power of the models from this paper can be tested using machine learning techniques. It could be that the variables are all significant in a model, but it is not predicting default well compared to a model with only subgrade as a determinant. The same research regarding credit rating models could also be done for other P2P lending platforms outside the USA. It would be interesting to know whether the credit rating models of other platforms could be compared to the one of Lending Club.

Bibliography

Beioley, K., & Megaw, N. (2019, March 22). *Financial Times*. Retrieved from Peer-to-peer pressure: do the risks outweigh the rewards?: https://www.ft.com/content/6bf2c806-4a6b-11e9-8b7f-d49067e0f50d

Blomstrom, D. (2018). *Emotional Banking: Fixing Culture, Leveraging FinTech, and Transforming Retail Banks into Brands.* Palgrave Macmillan.

Carmichael, D. (2014). Modeling default for peer-to-peer loans. *Available at SSRN 2529240*.

De Roure, C., Pelizzon, L., & Tasca, P. (2016). How does P2P lending fit into the consumer credit market?

Demirgüç-Kunt, A., & Huizinga, H. (1999). Determinants of commercial bank interest margins and profitability: some international evidence. *The World Bank Economic Review*, *13*(2), 379-408.

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics, 47*(1), 54-70.

FICO. (2019). *Education – FICO® Score*. Retrieved from FICO Score: https://ficoscore.com/education/

Five Degrees. (2018). *making the switch. rewiring banking.* Five Degrees.

Gough, N. (2016, February 01). *Online Lender Ezubao Took $7.6 Billion in Ponzi Scheme, China Says*. Retrieved from The New York Times: https://www.nytimes.com/2016/02/02/business/dealbook/ezubao-china-fraud.html

Käfer, B. (2018). Peer-to-Peer Lending–A (Financial Stability) Risk Perspective. *Review of Economics, 69*(1), 1-25.

Lee, E., & Lee, B. (2012). Herding behavior in online P2P lending: An empirical investigation. *Electronic Commerce Research and Applications, 11*(5), 495-503.

Lending Club. (2019, March 31). *LendingClub Statistics*. Retrieved from LendingClub: https://www.lendingclub.com/info/demand-and-credit-profile.action

Lending Club. (2019). *Rate information - Lending Club*. Retrieved from Lending Club: https://www.lendingclub.com/foliofn/rateDetail.action

Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science, 59*(1), 17-35.

Pope, D. G., & Sydnor, J. R. (2011). What's in a Picture? Evidence of Discrimination from Prosper. com. *Journal of Human resources, 46*(1), 53-92.

Public Information Office. (2010, September 20). *US Business Cycle Expansions and Contractions*. Retrieved from The National Bureau of Economic Research: https://www.nber.org/cycles.html

PWC. (2016). *What is FinTech?* PWC's Financial Services Institute.

Schueffel, P. (2016). Taming the beast: a scientific definition of fintech. *Journal of Innovation Management, 4,* 23-54.

Serrano-Cinca, C., Gutiérrez-Nieto, B., & López-Palacios, L. (2015). Determinants of default in P2P lending. *PloS one, 10*(10).

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics: International edition. *Pearson2012*.

Tang, H. (2019). Peer-to-peer lenders versus banks: substitutes or complements? *The Review of Financial Studies, 32*(5), 1900-1938.

I. Appendix A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.Default	1.000															
2.Grade	0.2611	1.000														
3.Subgrade	0.2673	0.9755	1.000													
4.FICO	-0.1310	-0.4091	-0.4264	1.000												
5.Interest Rate	0.2618	0.9518	0.9752	-0.4069	1.000											
6.Loan Amount	0.0771	0.1474	0.1521	0.1029	0.1477	1.000										
7.Annual Income	-0.0337	-0.0630	-0.0660	0.0717	-0.0667	0.3046	1.000									
8.Employment History	0.0077	0.0032	0.0035	-0.0027	0.0024	-0.0348	-0.0336	1.000								
9.Delinquency 2 years	0.0212	0.0540	0.0555	-0.1758	0.0490	-0.0045	0.0391	-0.0255	1.000							
10.Inquiries 6 months	0.0652	0.2175	0.2230	-0.0861	0.2154	-0.0196	0.0297	0.0009	0.0213	1.000						
11.Public record	0.0263	0.0610	0.0622	-0.1865	0.0543	-0.0589	0.0009	-0.0100	-0.0186	0.0591	1.000					
12.Revolving Utilization	0.0543	0.2353	0.2458	-0.4602	0.2385	0.1036	0.0322	-0.0068	-0.0114	-0.0810	-0.0660	1.000				
13.Open Accounts	0.0293	-0.0025	-0.0038	0.0149	-0.0066	0.1805	0.1309	-0.0236	0.0519	0.1309	-0.0144	-0.1421	1.000			
14.Loan Amount to Annual Income	0.1372	0.2437	0.2520	0.0140	0.2496	0.5732	-0.2283	0.0199	-0.0605	-0.0640	-0.0605	0.0665	-0.0061	1.000		
15.Annual Instalment to Income	0.1184	0.2465	0.2550	-0.0425	0.2626	0.4954	-0.2457	0.0266	-0.0520	-0.0415	-0.0493	0.0824	-0.0232	0.9480	1.000	
16. Debt to Income	0.2213	0.1607	0.1665	-0.0842	0.1624	0.0357	-0.1615	-0.0134	-0.0089	0.0004	-0.0440	0.1846	0.3015	0.2193	0.2193	1.000

Table 17: Correlation matrix continuous variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1.Default	1.000																
2.Subgrade	0.2673	1.000															
3.Term	0.1834	0.4393	1.000														
4.H: Mortgage	-0.0671	-0.0710	0.1006	1.000													
5.H: Rent	0.0657	0.0693	-0.0926	-0.8136	1.000												
6.H: Other	0.0007	0.0030	-0.0049	-0.0198	-0.0164	1.000											
7.H: Own	0.0044	0.0048	-0.0159	-0.3333	-0.2757	-0.0067	1.000										
8.LP: Car	-0.0141	-0.0258	-0.0188	-0.0166	0.0103	0.0029	0.0107	1.000									
9.LP: Credit Card	-0.0422	-0.1675	-0.0392	-0.0199	0.0186	0.0012	0.0028	-0.0557	1.000								
10.LP: Debt Consolidation	0.0329	0.0935	0.0769	0.0015	0.0216	-0.0050	-0.0372	-0.1240	-0.6224	1.000							
11.LP: Educational	-0.0000	0.0009	-0.0088	-0.0081	0.0087	0.0107	-0.0014	-0.0019	-0.0096	-0.0213	1.000						
12.LP: Home Improvement	-0.0133	-0.0229	0.0054	0.1425	-0.1719	-0.0004	0.0433	-0.0277	-0.1390	-0.3092	-0.0048	1.000					
13.LP: House	0.0053	0.0344	0.0036	-0.0216	0.0135	0.0018	0.0138	-0.0078	-0.0393	-0.0875	-0.0013	-0.0195	1.000				
14.LP: Major Purchase	-0.0033	-0.0150	-0.0134	-0.0246	0.0193	-0.0000	0.0094	-0.0159	-0.0797	-0.1774	-0.0027	-0.0396	-0.0112	1.000			
15.LP: Medical	0.0038	0.0189	-0.0226	-0.0099	0.0069	0.0028	0.0050	-0.0112	-0.0564	-0.1255	-0.0019	-0.0280	-0.0079	-0.0161	1.000		
16.LP: Moving	0.0071	0.0359	-0.0260	-0.0534	0.0619	0.0002	-0.0120	-0.0089	-0.0447	-0.0994	-0.0015	-0.0222	-0.0063	-0.0127	-0.0090	1.000	
17.LP: Other	0.0095	0.0736	-0.0475	-0.0486	0.0410	0.0021	0.0137	-0.0260	-0.1307	-0.2908	-0.0045	-0.0649	-0.0184	-0.0373	-0.0264	-0.0209	1.000
18.LP: Renewable Energy	0.0029	0.0116	-0.0062	-0.0035	0.0031	0.0010	0.0008	-0.0028	-0.0139	-0.0310	-0.0005	-0.0069	-0.0020	-0.0040	-0.0028	-0.0022	-0.0065
19.LP: Small business	0.0292	0.0689	-0.0034	-0.0113	0.0105	0.0054	0.0013	-0.0117	-0.0586	-0.1303	-0.0020	-0.0291	-0.0082	-0.0167	-0.0118	-0.0094	-0.0274
20.LP: Vacation	-0.0010	0.0093	-0.0343	-0.0198	0.0167	-0.0002	0.0056	-0.0086	-0.0432	-0.0961	-0.0015	-0.0215	-0.0061	-0.0123	-0.0087	-0.0069	-0.0202
21.LP: Wedding	-0.0088	0.0079	-0.0076	-0.0155	0.0191	0.0001	-0.0054	-0.0045	-0.0227	-0.0506	-0.0008	-0.0113	-0.0032	-0.0065	-0.0046	-0.0036	-0.0106

Table 18: Correlation Matrix of discrete variables

	18.	19	20	21
18.LP: Renewable	1.000			
19.LP: Small Business	-0.0029	1.000		
20.LP: Vacation	-0.0022	-0.0090	1.000	
21.LP: Wedding	-0.0011	-0.0048	-0.0035	1.000