A Temporal Analysis of Peer-to-Peer Lending Platforms: A Case Study of Lending Club

A thesis submitted to the
Graduate School of Natural and Applied Sciences

by

Parinaz Toufani

in partial fulfillment for the
degree of Master of Science

in

Industrial and Systems Engineering

İSTANBUL ŞEHİR UNIVERSITY

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Industrial and Systems Engineering.

APPROVED BY:

Asst.Prof.Dr Murat Küçükvar (Thesis Advisor)	....
Asst.Prof.Dr Nuri Cihat Onat (Thesis Co-advisor)	....
Asst.Prof.Dr Onur Güzey	....
Asst.Prof. Dr.Barış Arslan	....
Assoc.Prof.Dr. Ali Fuat Alkaya	....

This is to confirm that this thesis complies with all the standards set by the Graduate School of Natural and Applied Sciences of İstanbul Şehir University:

DATE OF APPROVAL: 23/5/2017

SEAL/SIGNATURE:

Declaration of Authorship

I, Parinaz TOUFANI, declare that this thesis titled, 'A Temporal Analysis of Peer-to-Peer Lending Platforms: A Case Study of Lending Club' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 23/05/2017

A Temporal Analysis of Peer-to-Peer Lending Platforms: A Case Study of Lending Club

Parinaz Toufani

Abstract

Peer to Peer (P2P) lending constructs a new way in lending market which provides certain benefits for individuals who attend this market either as borrowers or lenders. Lending Club (LC) is one of the pioneer online social lending platforms. This paper explores loan data gained from LC with the aim of helping lenders by discovering temporal patterns of borrowers' characteristics. Each loan application contains detailed information of borrowers' personal information, credit history and loan characteristics. This study consists of three main parts. First, we perform a temporal analysis to extract trends and seasonal variations of applications' characteristics, those which are highly correlated with loan status between 2007 and 2016. Second, we detect the likely causes of existing trends and seasonality for future extrapolation. Finally, we apply predictive statistics to estimate how the existing trend and seasonality of each attribute will change. According to our analyses, future applicants' annual income, debt to income ratio, and credit revolving balance have a rising trend, indicating increase in the applicants' financial quality. However, the stable trend of FICO score (credit score introduced by Fair Isaac Corporation) and the decreasing trend of the grade LC assigns to each application are negative indicators for creditworthiness of future applications.

Keywords: Social lending, peer-to-peer lending, temporal analysis, future extrapolation, predictive statistics

Birebir Eşleşme Yontemiyle Borç Verme ve Borç Alan Platformlarının Zamansal Analizi: Lending Club Bir Örnek Olay İncelemesi

Parinaz Toufani

Öz

Birebir eşleşme yontemiyle (P2P) borç verme ve borç alan olarak bu piyasaya katılan bireylere belirli avantajlar sağlayan yeni bir kredi piyasası yontemi oluşturur. Lending Club (LC) oncu online sosyal kredi platformlarından biridir. Bu yazıda, borç verenin ozelliklerinin zamansal kalıplarını keşfederek kredi verenlere yardımcı olmak amacıyla LC'den edinilen kredi verileri incelenmektedir. Her kredi başvurusu, borçlunun kişisel bilgileri, kredi geçmişi ve kredi ozellikleri hakkında ayrıntılı bilgiler içerir. Bu çalışma uç ana bolumden oluşmaktadır. İlk olarak, 2007 ve 2016 yılları arasındaki kredi durumuyla yuksek derecede ilişkili olan uygulamaların ozelliklerinin eğilim ve mevsimsel varyasyonlarını bulmak için zamansal bir analiz yapıyoruz. İkinci olarak, gelecekteki ekstrapolasyon için mevcut eğilimlerin ve mevsimselliğin olası nedenlerini tespit ediyoruz. Son olarak, her bir ozelliğin mevcut eğiliminin ve mevsimsellik oranının nasıl değişeceğini tahmin etmek için tahmini istatistikler uyguluyoruz. Yaptığımız analizlere gore, gelecek başvuru sahiplerinin yıllık geliri, Borç-Gelir Oranı ve kredi devir bakiyesi, başvuranların finansal kalitesinde artışa işaret eden yukselen bir eğilime sahiptir. Bununla birlikte, FICO skorunun istikrarlı eğilimi ve LC'nin her bir uygulamaya verdiği azalan eğilim, gelecekteki uygulamaların kredibilitesi için negatif gostergelerdir.

Anahtar Sözcükler: Sosyal kredilendirme, eşler arası borç verme, zamansal analiz, gelecekteki ekstrapolasyon, ongorusel istatistikler

Acknowledgments

I will forever be thankful to my advisor, Asst. Prof. Murat Küçükvar for his support during last year of my graduate studies at Istanbul Sehir University. He is the one who accepted me as a student whose previous advisor left university. I have been fortunate to be under his supervision. He is the reason I was able to pursue my research and studies here.

I am very grateful to my co-advisor, Asst. Prof. Nuri Cihat Onat for all his patience and insightful suggestions. He is an example of excellence as a researcher and instructor.

I would like also to thank Assoc. Prof. Vural Aksakallı for encouraging me to start this work and for all his valuable ideas and support.

I would like to thank my thesis committee members for all of their invaluable discussion and feedback through this process.

And finally, I would especially like to thank my parents for their love, support, and constant encouragement. I undoubtedly could not have done this without them.

Contents

List of Figures

List of Tables

Abbreviations

AIC	**A**kaike **I**nformation **C**riterion
AR	**A**utoregressive **M**oving
ARFIMA	**A**utoregressive **F**ractionally **I**ntegrated **M**oving **A**verage
ARIMA	**A**utoregressive **I**ntegrated **M**oving **A**verage
ARMA	**A**utoregressive **M**oving **A**verage
BATS	**B**ox-Cox transformation **A**RMA errors **T**rend and **S**easonal components
DTI	**D**ebth to **I**ncome
ETS	**E**rror **T**rend and **S**easonality
IPO	**I**nitial **P**ublic **O**ffering
IRS	**I**nternal **R**evenue **S**ervice
LC	**L**ending **C**lub
MA	**M**oving **A**verage
P2P	**P**eer to **P**eer
SARIMA	**S**easonal **A**utoregressive **I**ntegrated **M**oving **A**verage
STL	**S**easonal and **T**rend decomposition using **L**oessa
TBATS	**T**rigonometric **B**ox-Cox transformation **A**RMA errors **T**rend and **S**easonal components
UCC-1	**U**niform **C**ommercial **C**ode-1

Chapter 1

Introduction

1.1 Overview

Social lending, also known as P2P lending, is a financial trend that allows people to lend and or borrow money without the help of traditional intermediaries like banks. This financial platform provides a domain for lenders and potential borrowers in a way which is profitable for both sides. Unlike traditional lending, borrowers for either personal or business loans can borrow money with lower interest rates without offering any collateral. Once borrower's request is approved, the loan will be issued if enough funds are committed by investors. The whole process takes few days and there will not be any impact on borrowers credit score if the loan is not funded [1]. On the other hand, lenders can invest on hundreds of loans made to qualified borrowers and in return receive monthly payments of principal and interest. This payment can often exceed the interest that can be earned by other financial means such as saving accounts in banks [2].

All processes of this marketplace run via online service which adds much ease to the lending procedure in comparison to the time consuming and rigid procedure of traditional systems. This new financial sector has attracted a great attention and a significant growth [3]. Because of its increasing demand, it will be a preferred saving and investing choice that can take the place of traditional systems in near future [4].

The first social lending company was Zopa, which was founded in 2005 in United Kingdom. The popular social lending platforms are in developed countries like Lending Club Corp and Prosper in the USA, ZOPA ltd in the UK and Smava GmbH in Germany. All of these platforms are collaborating with credit reporting agencies like TransUnion LLC as a third party to gain the customers credit scores.

There are many studies contending the analysis of social lending under different approaches in recent decades. Several studies probe the reasons for the emergence of this online socio-economic market [5]. Ashta and Assadi [6] analyzed the effect of Web 2.0 on this lending market appearance and development. Galloway et al. [7] looked into the challenges and also potential benefits of P2P lending in community development finance. Berkovich [8] explored the reasons why people prefer to use online P2P lending rather than other systems at a given risk level.

Several studies have been published for evaluation of the factors of loan success or default due to the existing potential risks for lenders. Many of these researches focus on the financial determinants of applications [9, 10]. Klafft [11] suggests some rules for lenders to improve their investment profit. The factors that attract lenders' trust in P2P lending market are provided by Chen et al. [12]. Herding behavior is an issue that threaten the lenders when they do not have perfect information regarding to borrowers. Lee and Lee [13] and Wang and Greiner [14] show that herding behavior exists among lenders in P2P market and they follow each other. Some other studies present methods to evaluate the status of borrowers, good or bad, to shed light on the creditworthiness of a loan [15].

There are some studies concentrating on some other characteristics rather than finacial ones, such as race or age of applicants in funding success and even creditworthiness of an application [16, 17]. How applicants' appearance effect the loan foundation is also studied by Duarte et al. [18]. Even the relationship between the loan funding, loan success and the languages that applicants use to describe their financial situation is assessed [19]. Besides financial features and demographic ones, analyzing the social marginal effects of attending P2P lending, such as forming social communities by lenders to share the strategies or friendship bands among borrowers is another aspect of studies related to this field [20, 21]. Social interactions between peers in this industry have the potential power to lead to a successful funding and also lower default rates [22]. However, Weiss et al. [23] explored the negative consequences that social groups may have for borrowers' chance of getting the requested funds.

On the other side, there are many studies devoted to time series forecasting. A time series forecasting is used in wide range of applications from weather forecasting [24] to economic time series forecasting such as stock price [25]. A large number of articles have analyzed the empirical properties of different forecasting models [26, 27]. Many researches have been published to propose several modifications of original forecasting methods [28–30]. Many studies yielded the comparative advantages of their developed methods over original ones [31]. Different measurement approaches have been used to evaluate the accuracy of different forecasting methods [32, 33]. Thompson [34] and Wun and Pearn [35] examined the statistical characteristics of these measures. Besides these performance measurements, statistical tests are used to compare the performance of several forecasting models [36].

The aim of this paper is a comprehensive analysis of loan applications in LC platform. This analysis focuses on major features of applications related to loan default. Emekter et al. [9] examined LC between May 2007 and June 2012. The results of this study showed that LC credit grade, debt to income ratio, FICO score, and revolving line utilization have the highest correlation with loan default or success. Besides, Malekipirbazari and Aksakalli [15] analyzed LC applications between January 2012 and September 2014. The research found that among different applicant' attributes, annual income, debt to income ratio, and FICO score are highly correlated with loan default. In this research, we used the findings of these two studies and provide temporal analysis for the main attributes highly correlated with loan status. A temporal analysis applied to extract trend and seasonal variations, which are the main features of time series. Without justifying the likely causes of time series characteristics, forecasting is not significant [37]. Hence, we try to find possible causes for existing trend and seasonality afterwards. Finally, we apply several forecasting methods. Once the best model is found, it is used to extrapolate the future situation of applications' characteristics.

This study provides useful information for lenders who decide to attend LC lending industry as lenders and even are evaluating different P2P lending markets. This exploratory research helps them by illuminating important attributes related to loan default, using forecasting methods to present the future scheme of each attribute. The importance of this analysis is because of the fact that in P2P lending, lenders deal with risk, instead of financial institutions. Additionally, this research is useful for LC and enables LC policy makers to make informed decisions regarding to their existing policies.

1.2 Organization of the Thesis

The rest of this thesis is organized as follows:

Chapter 2 presents temporal analysis of each attribute related to borrowers and loans. Their existing trend and seasonal variation are extracted.

Chapter 3 provides a comprehensive statistical forecasting of each attribute. In this chapter, we applied different forecasting methods and selected the best one based on AIC. Our goal is to forecast the loan applications' situation for the following three years.

And our detailed summary and conclusions are presented in chapter 4. We also propose several directions for future research.

Chapter 2

Temporal Analysis of Data

2.1 Lending Club review

LC was established as one of Facebook applications for the first time. It was developed into a full-scale lending company when 10.26 million dollars funding was received in August 2007. LC is currently the largest online lender, having more than 13 billion dollars in loans since 2007. Residents of all states are eligible to apply for LC, excluding Iowa or West Virginia. In application process, applicants fill a form and provide their personal information and details related to their credit report. LC verifies the given information to approve or reject a request. It assigns a grade, known as LC grade, a measurement which calculates the likelihood of default on the loan and interest rate to the requested loan. After LC approval, applicants wait for loan to get funded. All of applicants' information including personal details, bank accounts report, desired loan amount, and loan purpose are recorded. These attributes are categorized in four groups of borrowers' characteristics, their credit history, loan characteristics, and attributes contributed to borrowers' assessment. Tables 2.1-2.4 define attributes related to borrowers characteristics, credit history, loan characteristics, and borrowers assessment, respectively.

Among the recorded information, we review FICO score, LC credit grade, annual income, debt to income ratio, and credit revolving line utilization as they play an important role in loan status. The rest of attributes related to borrowers' and Loans' characteristics are examined thoroughly in a supporting file.

Table 2.1: Borrowers' characteristics

Attribute	Definition
Annual Income	Annual income of applicants which is reported by themselves.
Debt to Income Ratio	Monthly payments (e.g. credit cards, student loans, car loans) on the total debt obligations, excluding mortgage, divided by stated monthly income. This attribute expresses the borrower's ability to manage monthly payments and repay debts.
Home Ownership	Home ownership status is provided by the applicants. LC categorizes the home ownership status of borrowers as rent, mortgage, own, none, and any.
Employment Length	The length of time (in years) that an applicant has been employed within his/her current job. Possible values are between 0 and 10, where 0 represents being employed for less than one year and 10 serves as ten or more years.

Table 2.2: Borrowers' credit history

Attribute	Definition
Credit Lines	Total number of credit lines and open credit lines of applicants.
Credit Card Age	Applicant credit card age (in months) is calculated by using issue date (date the loan is funded) and the date when the borrower's earliest reported credit line was opened.
Revolving Line Utilization	Indicates how much the applicant is using the credit currently relative to all available revolving credit.
Number of Inquiries	Number of applicants' hard credit inquires within the past 6 months.
Number of Delinquencies	Number of delinquencies in applicants' credit file during past 2 years.

Table 2.3: Loan characteristics

Attribute	Definition
Loan Amount	Amount of loan that borrowers are allowed to apply for. There are minimum and maximum amounts for two kinds of loans offered by LC, namely personal loans and business loans.
Loan Purpose	The purpose for which the loan will be used for. There are fourteen categories as: Car, Credit Card, Debt Consolidation, Educational, Home Improvement, House, Major Purchase, Medical, Moving, Renewable Energy, Small Business, Vacation, Wedding, and Other.

Table 2.4: Borrowers' assessment

Attribute	Definition
FICO Score	FICO score which is provided by TransUnion is calculated by using applicants financial credit records. LC reports the applicants FICO score with two numbers as FICO low and FICO high. The FICO score which we use in this paper is the average of these two values.
LC Grade	LC assigns a credit grade ranging from A to G, and each letter grade has a sub-grade ranging from 1 to 5 for each loan using the risk assessment of applicants and loan characteristics.

2.1.1 FICO Score

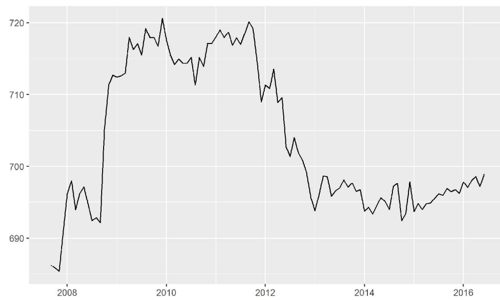
FICO score, which is received from TransUnion, is one of the most important factors in an applicant document file. Borrowers with a good FICO score, in general credit score, get lower interest rates and can apply for bigger loans. But even applications with high FICO score may be declined since there are other influential elements in application assessment. The minimum credit score required for applying to LC is 660.

Figures 2.1 and 2.2 display temporal analysis for FICO score regarding accepted and rejected applicants. Because there are many missing values in recent declined applications, the rejected credit score plot is displayed until 2014. As Figure 2.1(a) shows the average credit score is dropping since 2012. It can be explained by some changes into the LC underwriting standards such as [1]:

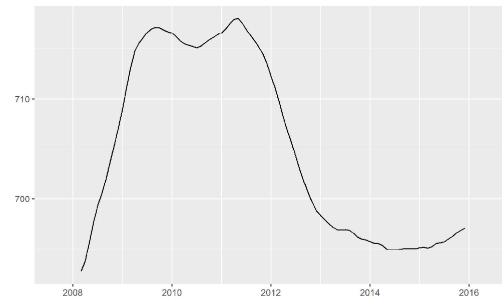
- Maximum number of credit inquiries has been increased to 6 for all borrowers (before, it was maximum of 3 inquiries for FICO scores of less than 740 and up to 8 inquiries for scores of 740 or more).
- Current delinquency allowed.
- Revolving credit balance maximum restriction removed.
- Major bankruptcy allowed.
- Maximum credit utilization of 98% restriction removed.

At first glance, it seems that LC is moving to a riskier phase. However, based on LC analysis, these changes lead to a more successful borrower ranking. According to LC

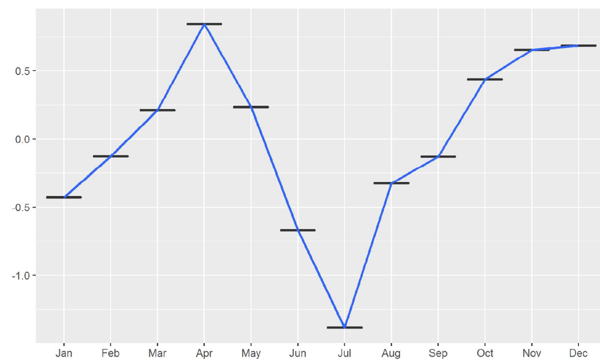
[1] www.lendingclub.com



((A)) Time Series



((B)) Trend



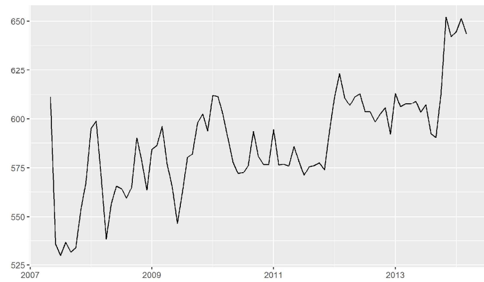
((C)) Seasonality

FIGURE 2.1: FICO score

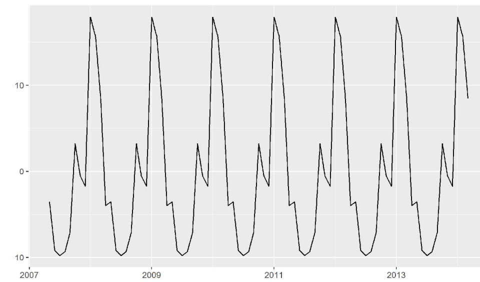
analysis, there was possibility of approving risky borrowers as well as rejecting best borrowers before applying the above changes.

Based on new criteria, many factors are taken into consideration in the evaluation of an application. It allows borrowers with low FICO score but better other credit variables to have a chance to attend LC lending family. Therefore, decreasing trend in borrowers' FICO score is not a negative sign of accepting the risky borrowers, but an outcome of some smart deviations in their selection criteria.

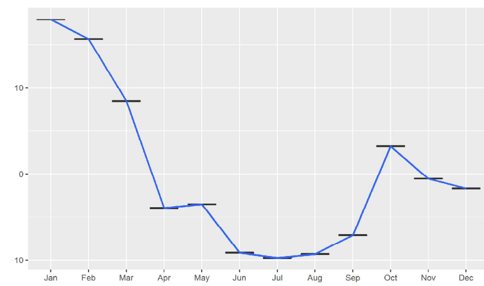
Based on Figure 2.1(c), borrowers with higher FICO score mostly apply in Spring and Winter, and those with lower FICO score apply during Summer time. Borrowers who apply in April have the the highest FICO score, however they do not get the sensible grade and they do not get the most acceptance among applicants. This argument signifies the fact that LC does not assign the grade to applicant solely based on FICO score.



((A)) Time Series



((B)) Trend



((C)) Seasonality

FIGURE 2.2: Rejected applicant risk score

2.1.2 LC Grade

For each loan, LC assigns a credit grade ranging from A to G, and each letter grade has a sub-grade ranging from 1 to 5. The assessment of the credit grade decision includes FICO score, loan term whether 36 or 60 months (shorter term is considered better), proprietary models, and the loan amount requested by the borrower[2]. According to this grade, LC appoints an interest rate for each borrower (Table 2.5).

First, LC assigns a loan grade from A1 through G5 for every borrower based on their credit score. For example, borrowers with FICO scores of 770 and above are assigned as an A1.[3] However, this is just a loan grade at initial evaluation, a borrower may be moved to a different grade based on other records, such as number of recent credit inquiries, age of credit history, number of open accounts, credit revolving utilization , and loan term. LC does not clarify exactly which attributes are considered in their grading formula. This information would effect the loan grade and consequently the interest rate.

[2]www.lendingclub.com
[3]www.lendingclub.com

Table 2.5: Interest rate for different LC grades

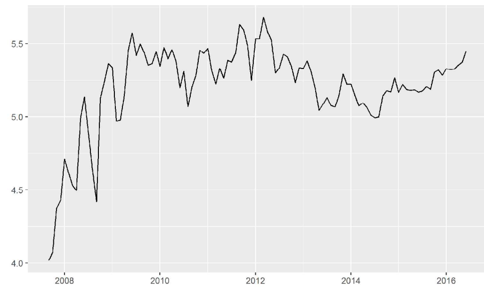
	A	B	C	D	E	F	G
1	5.32%	8.39%	11.99%	16.29%	19.99%	23.13%	27.34%
2	6.49%	9.16%	12.99%	17.27%	20.75%	24.11%	28.14%
3	6.97%	9.75%	13.67%	18.25%	21.18%	24.99%	28.34%
4	7.39 %	10.75%	14.46%	18.99%	21.97%	25.88%	28.67%
5	7.89%	11.47%	15.31%	19.53%	22.45%	26.57%	28.99%

Credit grade determines the interest rate charged for the loan. These rates are set based on the originator's underwriting assessment for the individual borrower. For determining an interest rate to a loan, LC assigns a base rate at first step. Then this rate accounts for the interest rate determination for each loan. No borrower can actually get that base rate, this is just where the calculations for loan interest rate begin. Besides, as mentioned, there is an adjustment rate for every loan grade from A1 to G5 which is added to the base rate to determine the final interest rate[4].

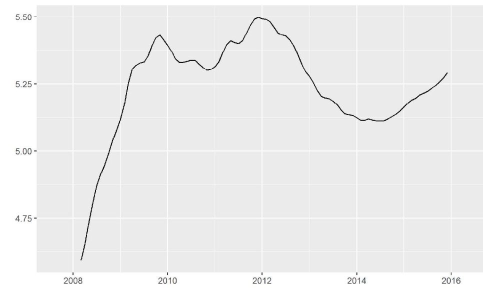
Interest rates on personal loans are between 5.49% and 28.69% which corresponds to A1 and G5 credit grade respectively. Origination fees range between 1% and 6% of the initial loan amount. This fee is subtracted from the loan when it is issued and as a result the money that applicant receives is less than the full amount of loan when it is issued. There is no application fee involved in the process. Lower interest rates are given to loans with better grades, which have potentially lower rate of being charged off. Each loan grade and its corresponding interest rate is displayed in Table 2.5.

Loan grades associate with both loan risk and potential return. It is a useful attribute for investors to check the potential return for each loan before investing. A-graded loans only have a 6-8% interest rate, so they earn less return in comparison to G-graded loans. On the other hand, investing on G-graded loans is more volatile. We make the grades and sub grades numeric to have better comparison. We assign 7 to A grade and 7 to G. Detailed information of application credit grade is displayed in Figure 2.3. Applicants in August get the lowest LC grade, in average 4.59 which corresponds to D grade. As 2.3(c) shows those who apply in January get the highest grades. However this amount is not that significant, since the average grade of applicants in January is 4.68 which is again under D grade. FICO score in August and January is neither perfect nor bad. Besides FICO score, many other application attributes which are not revealed by LC, are combined to determine the LC grade. The lower LC grade in August can be resulted

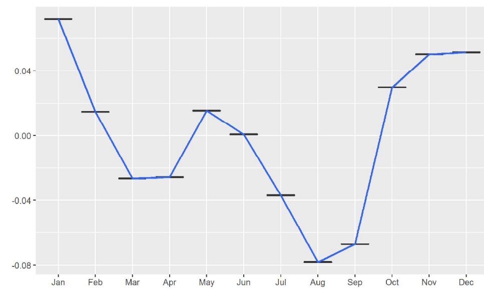
[4]www.lendingclub.com



((A)) Time Series



((B)) Trend

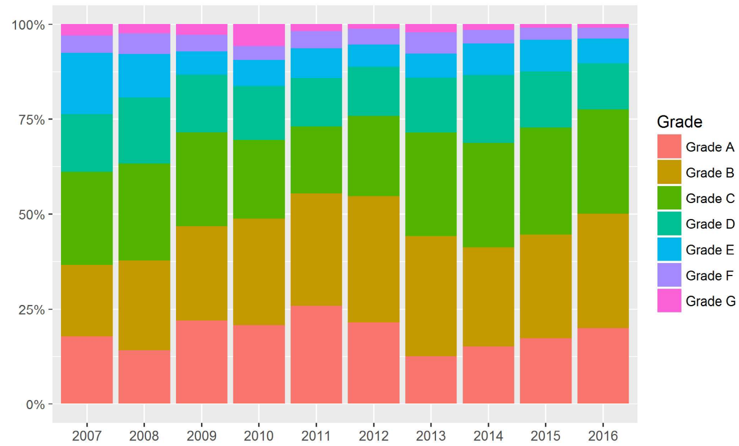


((C)) Seasonality

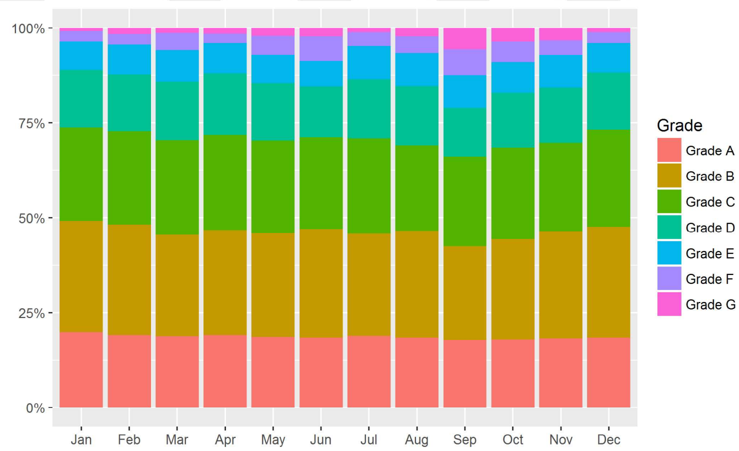
FIGURE 2.3: Applicant credit grade

as poor value of DTI, employment length, number of open accounts, number of inquiries, number of delinquencies and the number of applicants whose home ownership status is rent. In January, these features have more satisfactory level than August and it leads to a bit better LC grade. The analyses of these attributes are described thoroughly in the supporting file.

Figure 2.4 shows how LC grade fluctuates monthly and yearly. As Figure 2.4(a) displays, the number of applicants with grade of A is lower than what it is before this time. It is because of some changes in LC criterion in 2012 such as removing revolving credit balance maximum restriction or allowance of applicants with current delinquencies. It is not necessary to be high qualified applicant to be able to apply for LC. On the other hand, the number of applicants with grade G is decreasing as well. This fact signifies that changes in LC standards does not only bring LC to risky phase but also leads to an efficient evaluation of applications. The majority of applicants are those with grade of B and C. It is a good sign for lenders since there are many safe applications to invest on. As Figure 2.4(b) shows, the number of applicants who are assigned with grade A,



((A)) Yearly



((B)) Monthly

FIGURE 2.4: Bar plots for applicant credit grade

B, and C are nearly stable through months but applicants with the worst grade G are applying in Summer depending on low FICO score and other application features.

2.1.3 Annual Income

Annual income is among the records, which are provided by the applicants. However, they are required to submit related documents such as pay stubs, W-2 forms, or other tax records to verify the income stated in their loan request. Then, LC conducts a check to verify their claimed income. LC may also verify the income of the applicants through a third party provider. There are three outcomes of this verification:

- Income verified: Having the income within 10% of the claimed one is proved to LC.
- Income source verified: Both the claimed income and employer of the borrower were verified.
- Not verified: The claimed annual income could not be verified by LC. This does not mean that the application should be surely rejected. However, if LC finds out that it is a false claim the application will be declined.

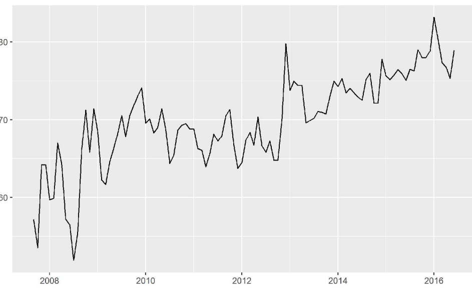
Figure 2.5(a) shows the borrowers' income after considering the inflation rate. Although inflation in the United States is not that high, it is included to make an unbiased comparison of borrowers' real wage. Usually people with high income and credit score do not apply for LC [9]. However, as Figure 2.5(b) indicates, the average income of those who apply for LC, is increasing by time. As of December 2015, the average personal income of LC borrowers is $75,055. This growing trend is a good sign for LC which was successful in attracting people with higher income in comparison to its starting years.

Figure 2.5(c) represents there is a significant seasonality in income which is increasing in fall, and especially in December it reaches its highest value. This point can be explained by the fact that the shopping increases a lot around new year which results in a high income of majority of American population.

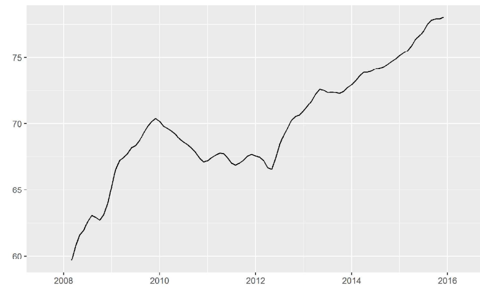
2.1.4 Debt-to-Income Ratio (DTI)

Borrowers' debt to income ratio (DTI) can be defined as monthly payments (e.g. credit cards, student loans, car loans) on the total debt obligations, excluding mortgage, divided by stated monthly income. This attribute expresses the borrower's ability to manage monthly payments and repay debts.

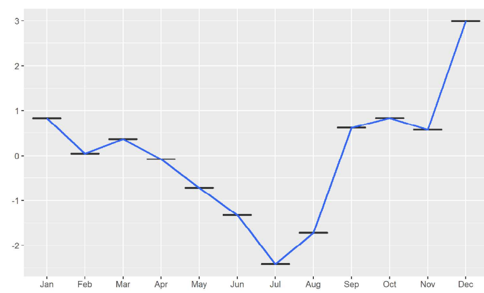
A lower debt-to-income ratio demonstrates a good balance between debt and income. Conversely, a high DTI is a demonstration that an individual has too much debt compared to their income. Although DTI is calculated based on self reported data by borrowers, this variable is among the potential variables which effects the default probability. The higher the borrower indebtedness, the higher the loan default risk.



((A)) Time Series



((B)) Trend

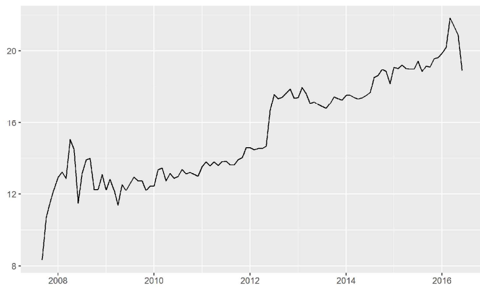


((C)) Seasonality

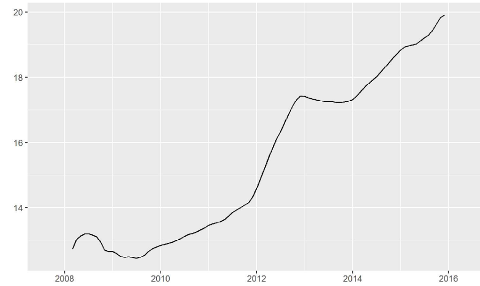
FIGURE 2.5: Annual income by considering inflation rate ($\times 1000$)

The average debt-to-income ratio for LC borrowers is 13.81%. LC has some rules related to DTI, for example the maximum debt to income ratio a borrower can have is 25% if their credit score is below 720 and it increases to 30% for borrowers with scores of 720 or more.

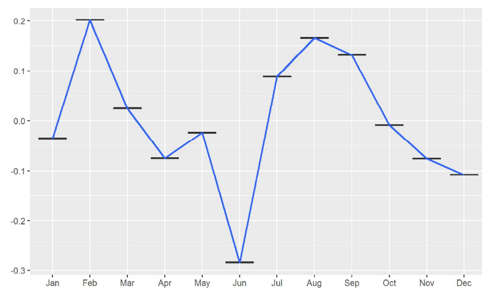
Figures 2.6 and 2.7 show the temporal analysis for DTI of accepted and rejected applicants respectively. As 2.6(b) and 2.7(b) show, DTI of both accepted and rejected applications have increasing trend since 2007. Increasing trend of borrowers' indebtedness is a warning for LC as well as lenders. However, this attribute along with other correlated attributes define the borrowers' total financial status. In February the applications have the highest DTI. As Figure 2.1(c) suggests the LC grade is in its lowest amount in August, when DTI is also in one of its highest values. It is an evidence of how important an effective DTI is regarding applicants' grading.



((A)) Time Series

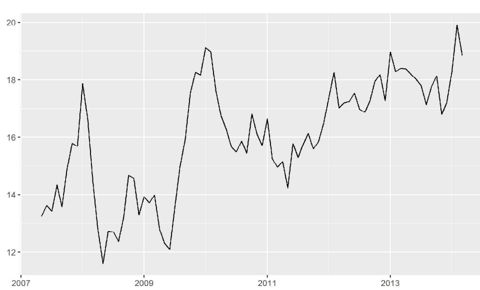


((B)) Trend

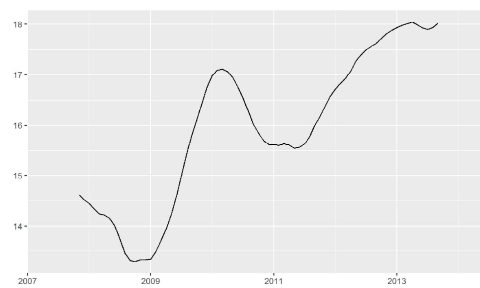


((C)) Seasonality

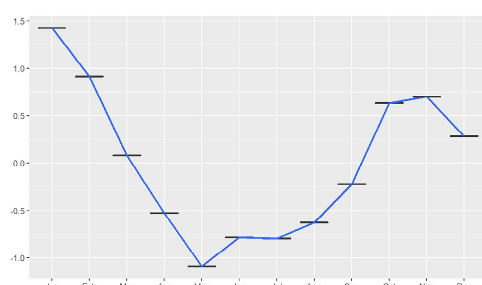
FIGURE 2.6: Accepted application Debt-to-Income ratio



((A)) Time Series



((B)) Trend



((C)) Seasonality

FIGURE 2.7: Rejected application Debt-to-Income ratio

2.1.5 Total Credit Revolving Balance

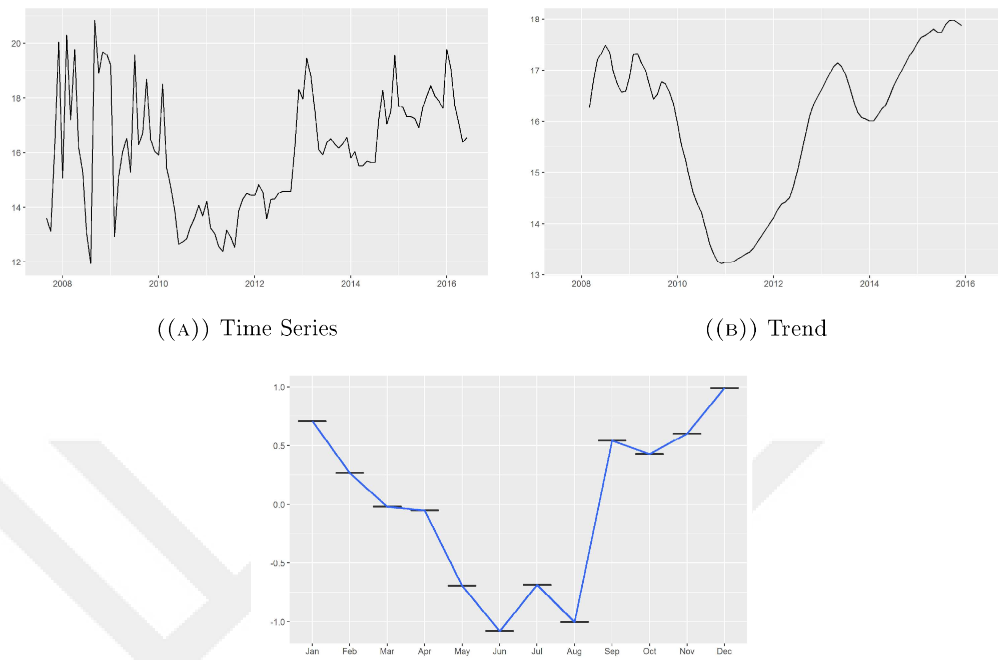
Revolving balance value indicates how much of credit card spent and is not paid at the end of each billing cycle. The revolving balance may increase or decrease based on the amount creditor borrowed and repaid. If credit or retail store card owner does not pay off the money at the end of grace period or even pay a portion of the money and revolve the rest to the next cycle, bank will charge them. The amount of the charge for revolving a balance depends on the size of the balance and the interest rate of the card. When the balance is paid off, the customer is no longer revolving the debt.

Credit card revolving balance is employed to estimate revolving utilization. Revolving utilization is defined as revolving balance divided by total credit limits, announced as a percentage. This percentage indicates how much the creditor is using the available credit currently. Creditors should try to keep this number to be as low as possible since it is one of the factors which influence the credit score.

In general, having too many accounts with high revolving balances is an indication of higher level of credit risk. Regularly and responsibly utilizing a few accounts, paying balances on time and keeping them low results in a high FICO score. Figure 2.8 displays temporal analysis for total credit revolving balance. According to Figure 2.8(b), the impact of this attribute on the acceptance rate increased until 2011, then decreased afterwards. the reason can be related to one of the changes LC made on its requirement which was removing the revolving credit balance maximum of $150,000 restriction.

2.1.6 Number of Applicants

Total number of applicants who applied to borrow money from LC is increasing every year (Figure 2.9). Until 2012, this trend was slow, but afterwards it went up drastically. The main reason of slow trend till 2012 is that this lending platform was launched on 2007 and until 2012 it was on start-up stage of its life cycle. After 2012, on the other hand, LC growth period has been started and it had a rapid extension since then. There are three major causes for this fast-track extension; most importantly, P2P lending platforms, in general, enable borrowers to get micro loans with lower rates in comparison to other financial institutions like banks or credit unions. It is much more appealing for borrowers with poor assets who want to apply for micro loans, and are not eligible to apply to



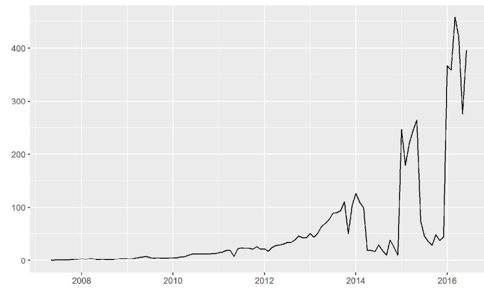
((A)) Time Series

((B)) Trend

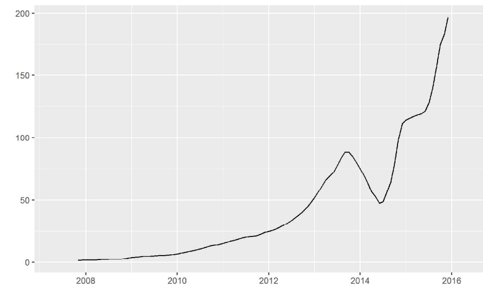
((C)) Seasonality

FIGURE 2.8: Total credit revolving balance ($\times$1000)

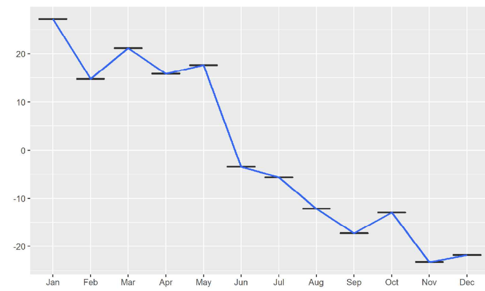
banks. Another force behind the growth of demands is instability of credit markets in 2008 which led to federal bailouts. Economic uncertainty in financial institutions scares people to use credit markets and as a consequent motivates them toward the new introduced lending industry [38]. In addition, LC has one of the best third-party investor ecosystems. Participation of many financial institutions and individual investors in this lending service helps to improve the value of LC which consequently helps this platform to be introduced more and become popular. As of June 2012, LC has been the largest peer-to-peer lending platform in the United States based on issued loan volume and revenue. In March 2013, LC had offered more than 100,000 loans with a total value of 1.5 billion dollars. In May 2013, this value has been increased nearly threefold when Google bought a stoke from LC. Moreover, LC began to collaborate with banks such as Titan Bank in 2013 and Union Bank in 2014. At this time, it has been filled for IPO (Initial Public Offering) and as a result transformed to a public company. Joining to IPO list indicated that LC is potential to be a better option than lending from banks or having credit card debt. The spread of web technologies is another effective factor in boosting the number of borrowers. Possibility of requesting online and funding the requested loan in a short time in the case of acceptance make these web-based platforms more popular.



((A)) Time Series



((B)) Trend



((C)) Seasonality

FIGURE 2.9: Number of applicants (×1000)

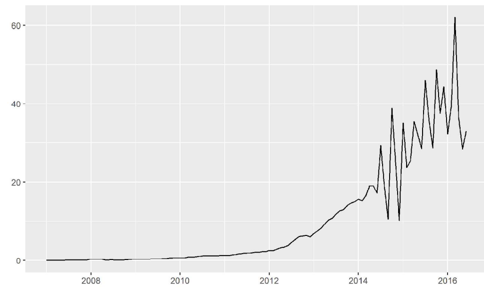
In Figure 2.9(c), seasonality plot shows that the average number of applicants is high in winter and spring and its lowest value is in November.

2.1.7 Number of Accepted Applicants

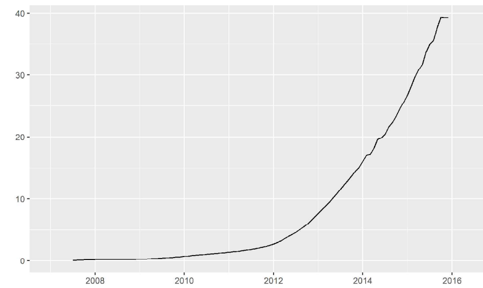
As P2P lending industry becomes more popular, many people tend to attend this construction. Figures 2.10(a) and 2.10(b) show there is an increasing trend in the number of accepted applications even though LC is strict in selection process. Many qualified borrowers based on LC criteria are attending this platform. In Figure 2.10(c), seasonality plot shows that there is no significant seasonal pattern in the number of accepted applicants. However, the average number of accepted applicants in October is the highest.

2.1.8 Number of Rejected Applicants

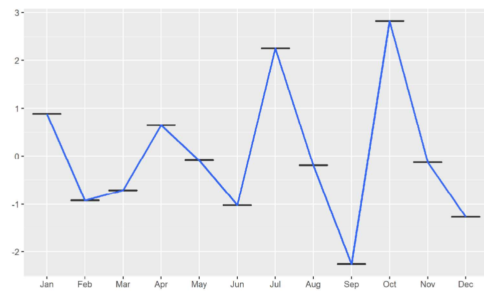
LC uses some information in order to determine if an applicant is qualified for a loan or not. This information includes the applicants' personal information such as income and employment status and the information provided by credit agencies like applicants'



((A)) Time Series



((B)) Trend



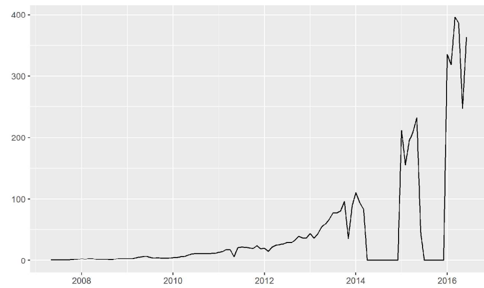
((C)) Seasonality

Figure 2.10: Number of accepted applicants ($\times 1000$)

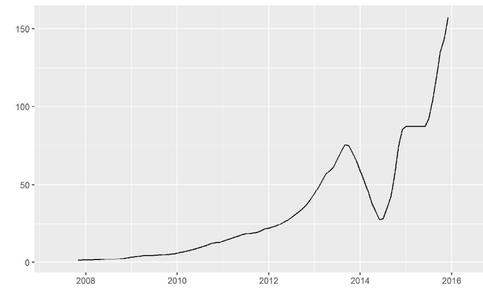
payment history, FICO score and credit activities. Figure 2.11 displays the temporal analysis of rejected applicants and their ratio with respect to the total applicants. As Figure 2.11(a) indicates many borrowers do not meet the minimum requirements of a successful application. Figure 2.11(b) exhibits the number of these applications are increasing and LC policy has remained selective.

In November 2013, approximately 90% of applicants were not approved for a loan (Orchard). The main reason for this issue is applicants' low FICO score. The applicants' lower FICO score strongly decreases their chances of approval. The other effective factor is applicants' high debt to income ratio (DTI). A loan request is difficult to be approved if the applicant DTI ratio is more than 50%. Higher DTI indicates lower applicant's power of payment since they have other payments on other credit cards or loans.

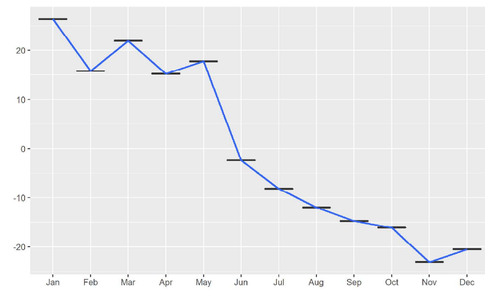
Same as DTI, applicants with high number of credit inquiries are likely to be declined because of low probability of returning their installments.



((A)) Time Series



((B)) Trend



((C)) Seasonality

FIGURE 2.11: Number of rejected applicant (×1000)

2.1.9 Loan Purpose

LC offers two kinds of loans, namely personal and business loans. Loan Purpose of personal loans indicates the purpose for which the loan will be used for. There are fourteen predefined purposes mentioned below[5], for which temporal analysis are displayed for the main six purposes in Figure 2.12- 2.17.

- Car
- Credit Card
- Debt Consolidation
- Educational
- Home Improvement
- House

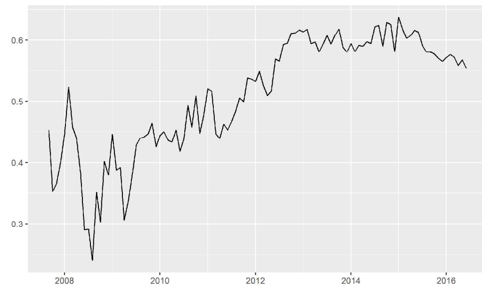
[5] www.lendingclub.com

- Major Purchase
- Medical
- Moving
- Renewable Energy
- Small Business
- Vacation
- Wedding
- Other

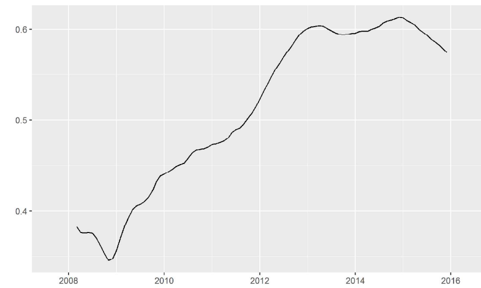
Almost 66% of the all loans issued since 2007 on LC are reported with the purposes of Debt Consolidation or Credit Card. Based on a LC survey, borrowers report that the interest rate on their loan was on average 35% lower than they were paying on their outstanding debt or credit cards. On the other hand, less than a percent of the loans are issued for Renewable Energy, Educational, Vacation and House Purchase purposes.

Loan purpose is significantly related to loan default. A loan with the purpose of wedding has not the same risk as the loan for funding a small business. However, we should consider that loan purpose is reported by the borrower and there is no secondary method to verify the intended use. Multiple intended use of the loan is also possible. To demonstrate the importance of loan purpose in our analysis, we can focus on the patterns happen to the important attributes in June. Figure 2.9 states that acceptance rate is one of the lowest in June while the number of applicants apply in this month is low as well. This fact clearly is the the result of decrease of good borrowers and increase of risky borrowers in June. Other seasonality patterns confirm this as well.

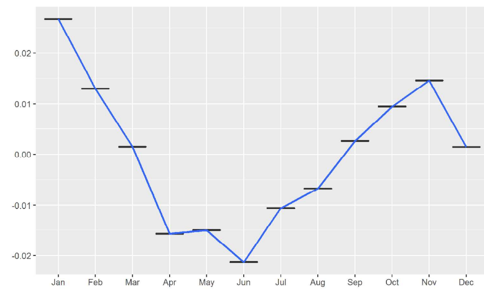
Besides the personal loans, LC offers a different product known as "Business Loans and Lines of Credit". This loan can be up to $100,000 for entrepreneurs who plan to build a business or business owners who intend to expand their business. This is a good opportunity for well-established businesses planning to expand or is in need of cash urgently with good business history and finances. Although LC lends to various types of businesses, business loan purposes cannot be financial investing, gambling or



((A)) Debt Consolidation



((B)) Trend



((C)) Seasonality

FIGURE 2.12: Loan purpose, debt consolidation

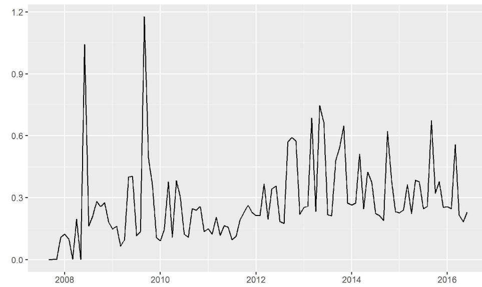
adult businesses. Business loans are not offered on consumer-focused platforms such as Peerform, Prosper, and Avant.

Examples of requirements for “business loan” are:

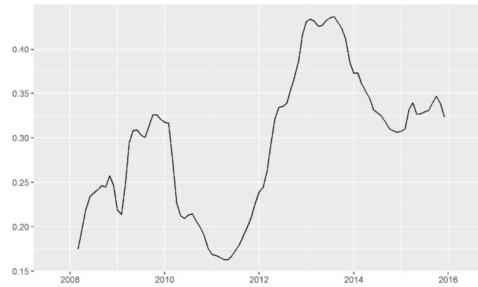
- Being the owner of at least 20% of the business
- Having a collateral with the value of higher than $100,000[6]
- Having personal credit score more than 600
- submitting three months of business bank account statements, an IRS Form 4506-T, and business tax

Figure 2.18 displays how the number of applicants with four main purposes of debt consolidation, credit card, others , and home improvement vary over years and months. As Figure 2.18(a) shows the majority of applicants are lending money with the purpose of debt consolidation or credit card. This fact confirms that LC is more successful than

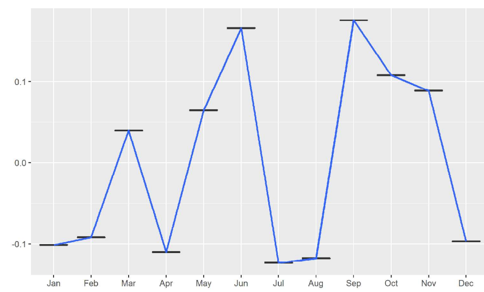
[6]Companies are required to fill a UCC-1 lien that allows LC to seize their business assets in the case of loan default.



((A)) Credit Card

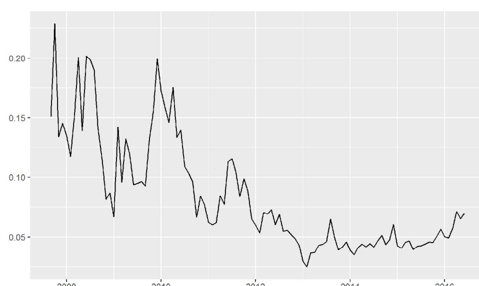


((B)) Trend

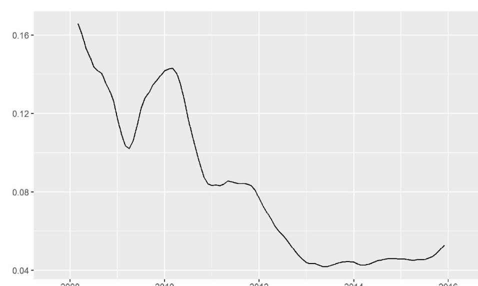


((C)) Seasonality

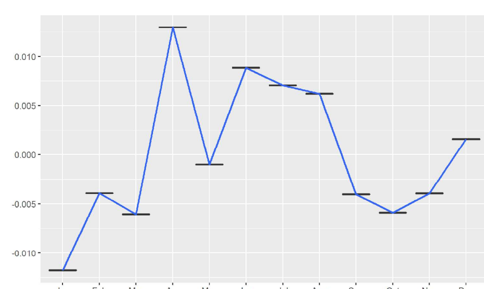
FIGURE 2.13: Loan purpose, credit card



((A)) Others

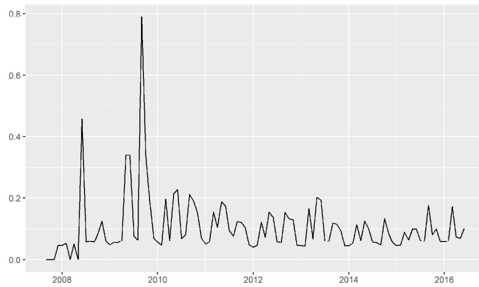


((B)) Trend

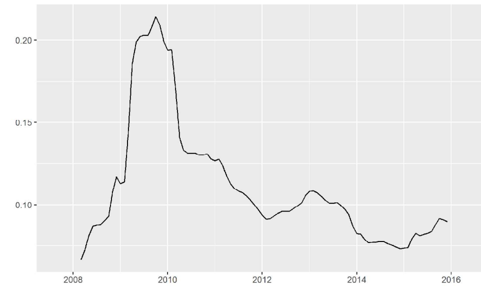


((C)) Seasonality

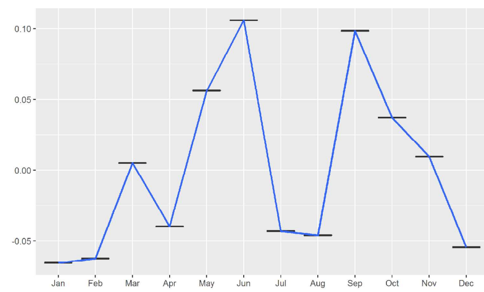
FIGURE 2.14: Loan purpose, others



((A)) Home Improvement

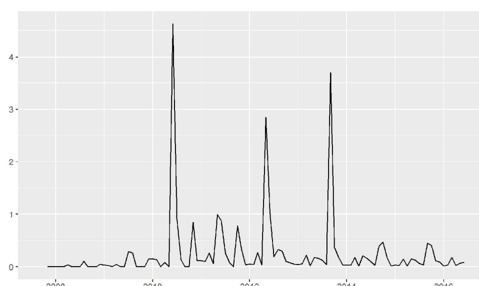


((B)) Trend

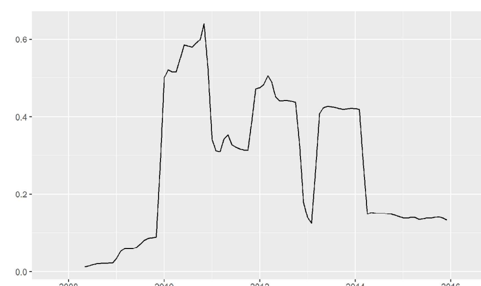


((C)) Seasonality

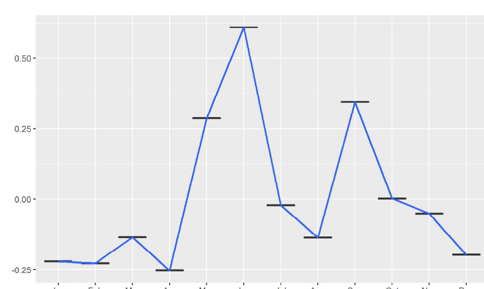
FIGURE 2.15: Loan purpose, home improvement



((A)) Major Purchase

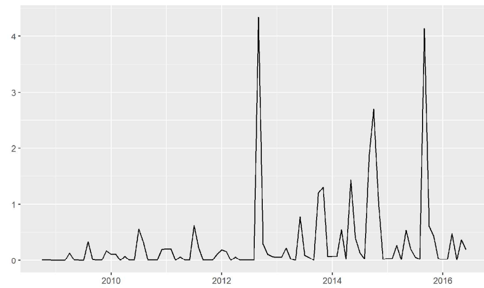


((B)) Trend

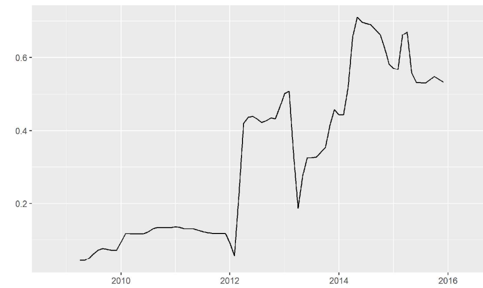


((C)) Seasonality

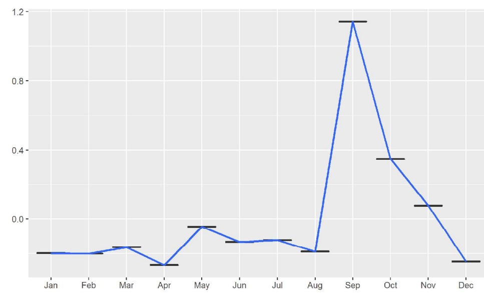
FIGURE 2.16: Loan purpose, major purchase



((A)) Small Business



((B)) Trend



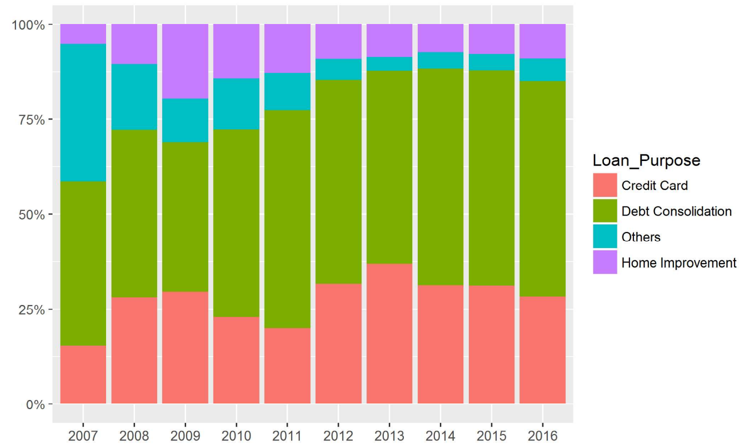
((C)) Seasonality

FIGURE 2.17: Loan purpose, small business

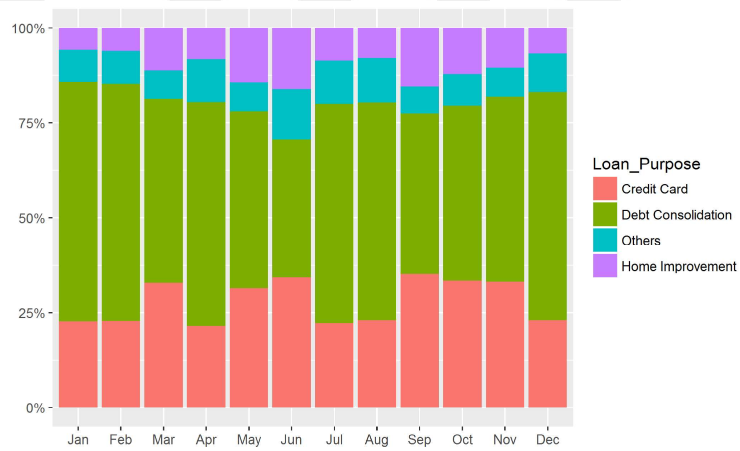
other lending platforms in a way that people prefer to borrow money from LC to pay their debts. In comparison to the first years of LC, frequency of borrowers with unidentified reasons, stated as Others in the form have been decreased. It can be as a result of either LC become more strict in approving loans with unknown purposes or lenders do not trust such loans. As it is clear in Figure 2.18(b), in June, there is a change in the pattern of loan purpose. The number of applicants with the purpose of Credit Consolidation decreases and the number of applicants with purpose of others increases significantly, which can be a sign for lenders to be more alert in this month since the loans with unknown purposes are risky.

2.1.10 Loan Amount

Personal loans are in the range of $1,000 up to $40,000. For small businesses, loans start at $15,000 and are capped at $100,000. Based on the assigned grade for each loan, borrowers are allowed to apply for an amount within the minimum and maximum amounts. Borrowers with grade A are qualified for the highest loan amount with the lowest interest rate. Those with grade G should pay the highest interest rate for the smallest amount



((A))



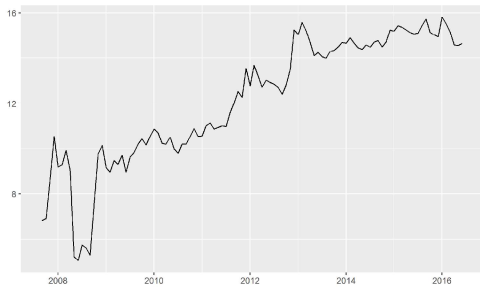
((B))

FIGURE 2.18: Bar plot of loan purposes

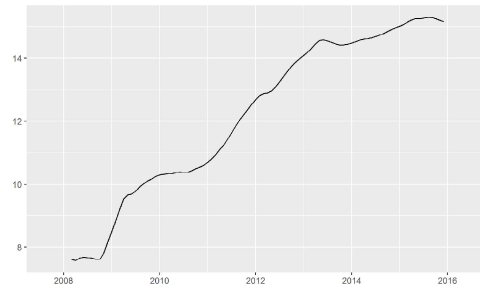
available. Figure 2.19 displays loan amount with the consideration of inflation rate. Although the number of accepted applicants is low in December, Figure 2.19(c) shows the highest amount of funded loans in this period. The reason is the applicants in December have higher higher FICO score (Figure 2.1(c)) and better LC grades (Figure 2.3).

2.1.11 Employment Length

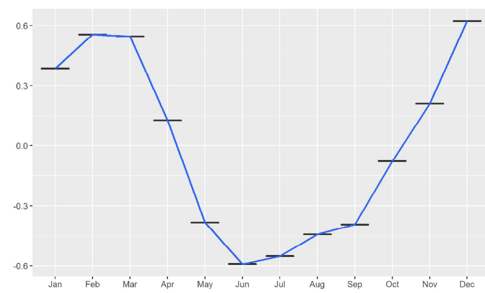
The length of time (in years) that an applicant has been employed within his/her current job is another question in the application. Possible values are between 0 and 10, where 0 represents being employed for less than one year and 10 serves as ten or more years. For



((A)) Time Series



((B)) Trend



((C)) Seasonality

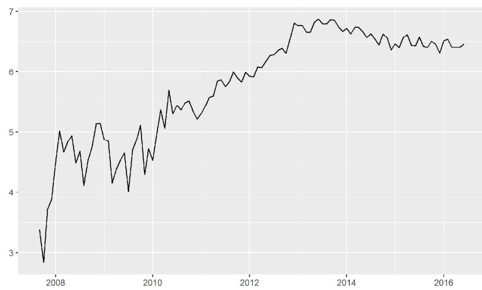
FIGURE 2.19: Loan amount by considering inflation rate (×1000)

individual borrowers, LC contacts with employers and checks the applicants employment status by requesting pay stubs or income tax statements. For self-employed applicants, LC may request more documentation about the applicant income and finances in order to check their employment status and length. This process can take up to 14 business days, which may be longer for self-employed applicants.

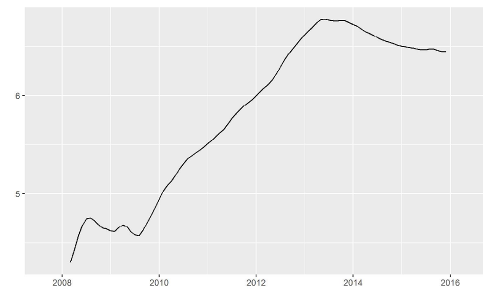
A majority of loan applications are rejected for applicants who have been in a firm for less than a year. About 84% of loan rejections were made for such employees. On the other hand, the acceptance rate reaches its highest value in the case the employee has been in a firm for more than ten years.

Temporal analysis of employment length for accepted and rejected applicants are shown in Figure 2.20 and 2.21. Increasing trend in Figure 2.20(b) reveals that employment length of applicants is increasing and LC is becoming more selective regarding this attribute.

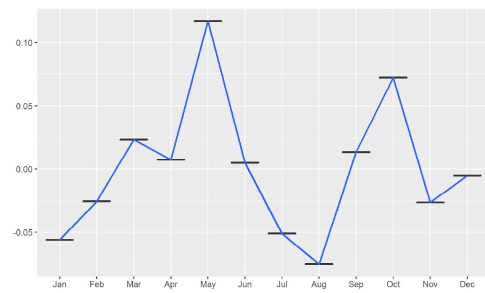
Figure 2.21(b) indicates LC rejects applicants even with moderate employment length at its early years, which was due to the fact that some other factors especially FICO score



((A)) Time Series



((B)) Trend



((C)) Seasonality

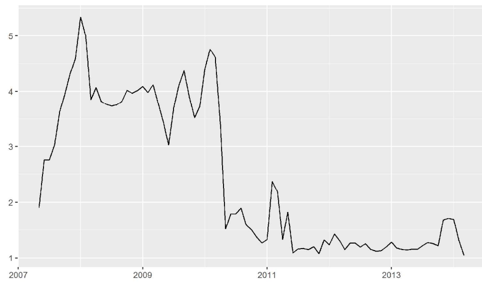
FIGURE 2.20: Accepted application employment length

and DTI were dominant in application evaluation rather than employment length.

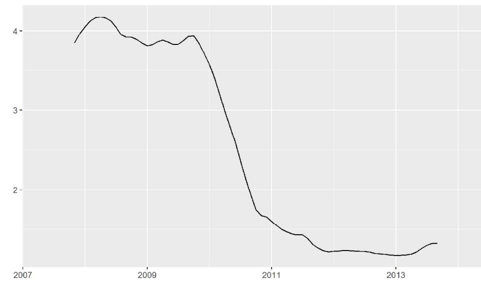
Figure 2.20(c) reveals that accepted applicants with the lowest employment length mostly applied in August. This observation is verified by noticing that accepted applicants in August had also the lowest LC grade (Figure 2.3). Rejected applicants in winter have the highest and in summer have the lowest length of employment (Figure 2.21(c)).

2.1.12 Credit Card Age

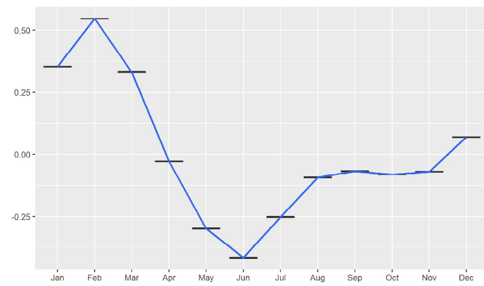
Applicant credit card age (in months) is calculated by using issue date (date the loan is funded) and the date when the borrower's earliest reported credit line was opened. The age of credit history has a significant impact on default rates. It is expected from fully paying borrowers to have older credit history since they had indicators of being good borrowers over past years which resulted in their account endurance. It is important to consider age of borrowers in the assessment of this attribute, due to their high correlation. In Figure 2.22, temporal analysis for age of applicants' credit history is shown. Although



((A)) Time Series



((B)) Trend



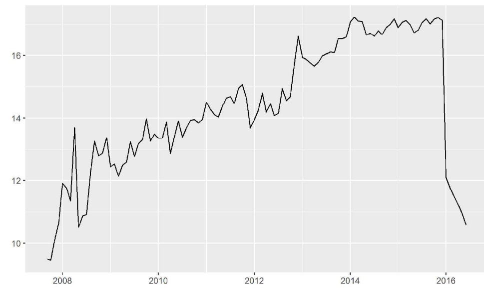
((C)) Seasonality

FIGURE 2.21: Rejected application employment length

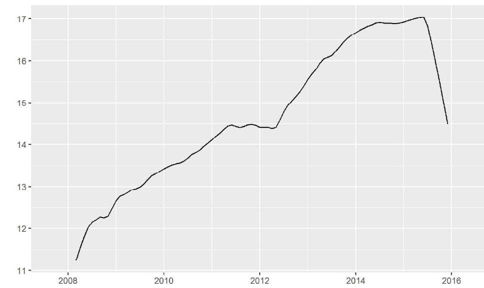
LC had become more selective toward the applicants with regard to their credit history, recently this attribute declined for accepted applicants (Figure 2.22(b)).

2.1.13 Credit Lines

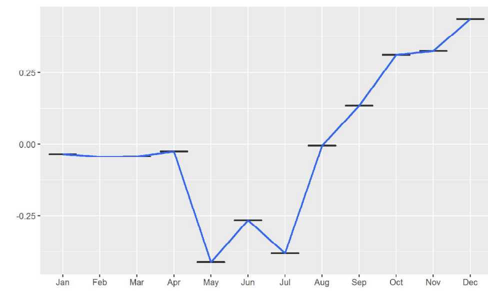
Total number of credit lines and open credit lines are two attributes under applicant credit history category. Total number of accounts and open accounts do not have notable impact on application acceptance [15]. It is interesting that fully paying borrowers with high FICO score tend to own slightly more accounts [39]. Figure 2.23 and Figure 2.24 demonstrate that the total number of credit lines and open credit lines of accepted applicants have an increasing trend. Although more credit lines decrease applicants' ability to manage their debts, it is not necessarily a negative mark and depends on borrowers' financial power and income. As mentioned before (Figure 2.5(b)), the applicants' average income is increasing over time and they are able to handle more credit lines accordingly. This can justify the increasing trend of applicants' credit lines over time. Additionally, some other correlated features of applicants, such as income, should be



((A)) Time Series



((B)) Trend



((C)) Seasonality

FIGURE 2.22: Age of credit history

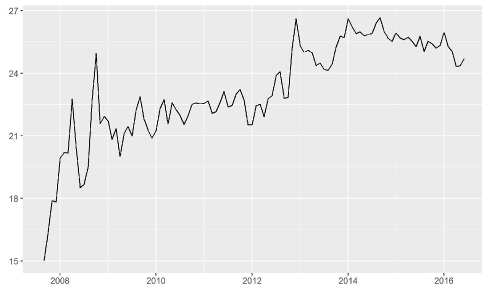
considered because they all have significant role of making a borrower able to apply for a credit card.

2.1.14 Number of Inquiries by Creditors

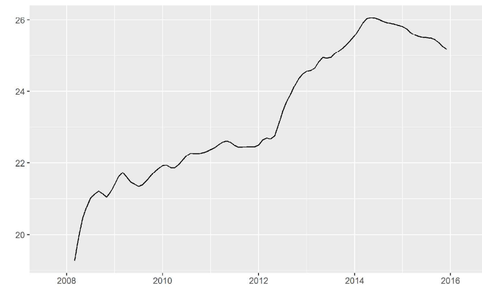
Credit history activities like the number of hard credit inquires within the last 6 months is another factor for LC to evaluate an application. Even though a soft pull is used by LC to pre-qualify the applicants, it does not count as a hard inquiry. This attribute may have dramatic impact on the interest rates as well as whether the applications are approved or not.

Stable and strong applications do not have too many hard inquiries especially in recent periods. When applicants are constantly applied for credits, their power of repaying the debts on time and being fully paid will decrease, and they may end up being declined.

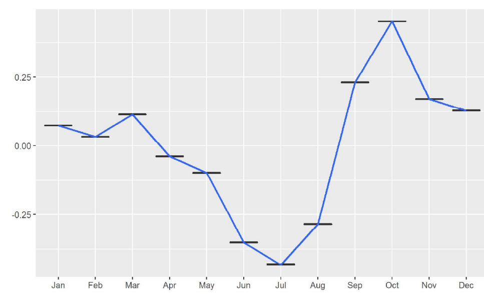
In Figure 2.25, temporal assessment of number of inquiries is presented. Figure 2.25(b) demonstrates that LC becoming more selective with regard to this attribute over time. It is clear from Figure 2.25(c) that applicants who applied in April have the lowest number



((A)) Time Series

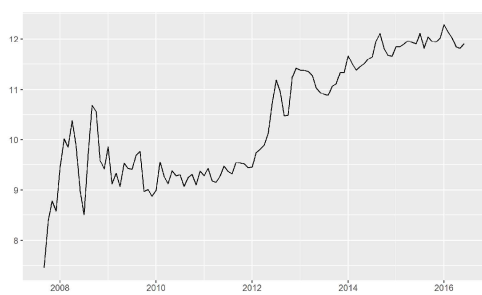


((B)) Trend

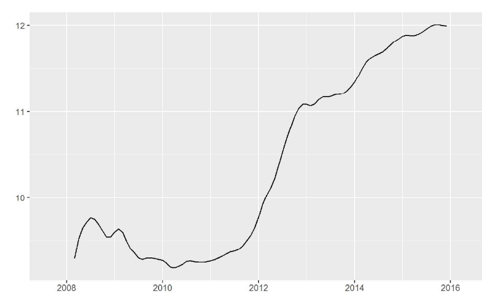


((C)) Seasonality

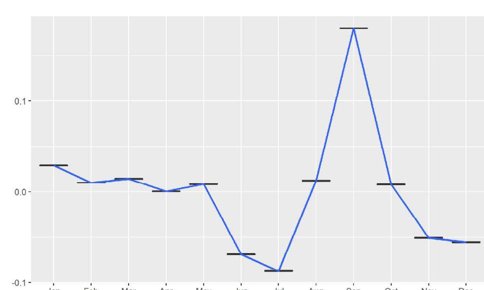
FIGURE 2.23: Total number of applicant credit lines



((A)) Time Series

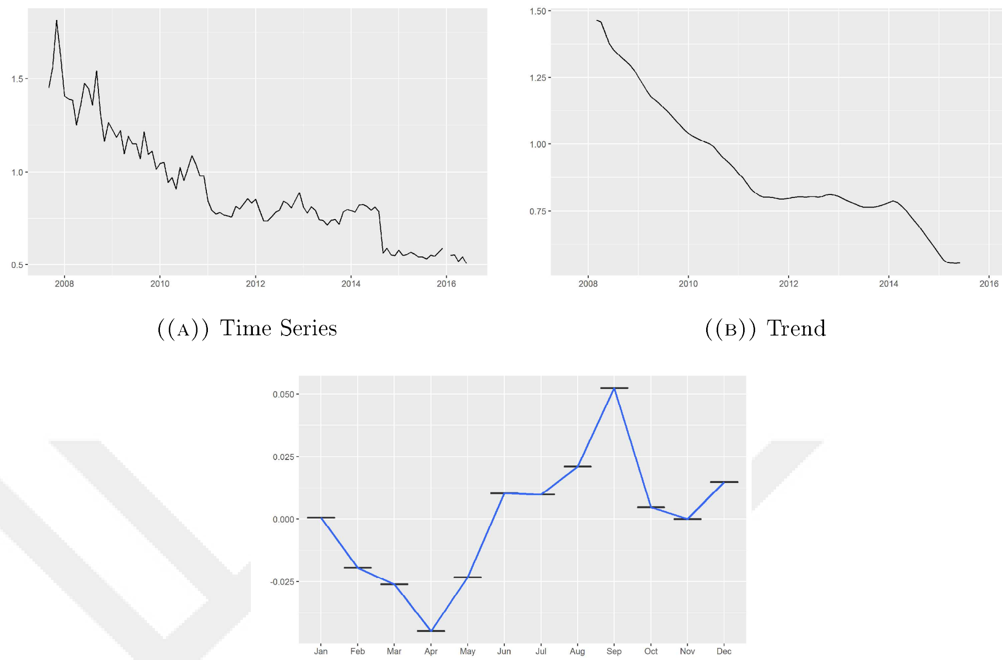


((B)) Trend



((C)) Seasonality

FIGURE 2.24: Number of applicant open credit lines



((A)) Time Series

((B)) Trend

((C)) Seasonality

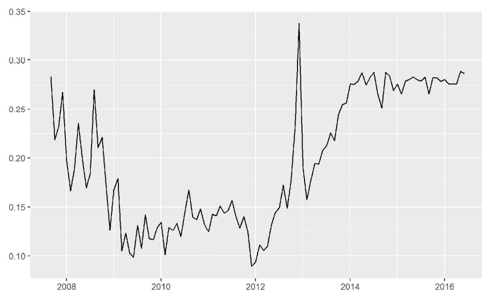
FIGURE 2.25: Number of inquiries by creditors during the past 6 months

of inquires. Since applicants with the highest FICO score also applied in April, there exists a good correlation between number of inquiries and FICO score for potentially good applicants.

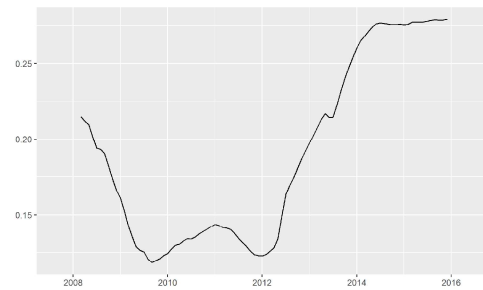
2.1.15 Number of Delinquencies

Borrowers who have a disappointing history of paying debt in the past show a higher rate of default because they have a tendency to repeat their same behavior in the future. Vast majority of the borrowers actually have zero delinquency on their credit bureau.

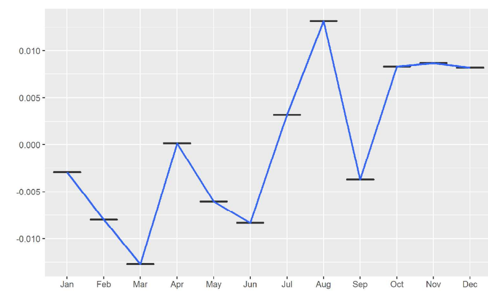
If someone has a record of delinquency at the time of application, they will most likely be rejected by LC. Temporal analysis for the number of delinquencies in borrowers' credit file is displayed in Figure 2.26. Figure 2.26(b) indicates the overall number of delinquencies for accepted applicants is low, however it has an increasing trend since 2012, after the changes in standards of LC. According to this change applicants with delinquencies do not get rejected because of their history before the evaluation.



((A)) Time Series



((B)) Trend



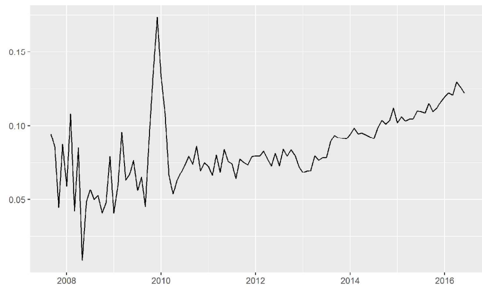
((C)) Seasonality

FIGURE 2.26: Number of delinquencies in the borrowers' credit file for the past 2 years

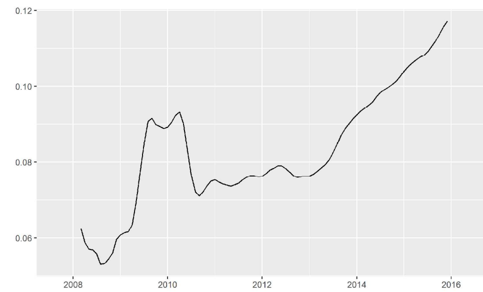
Figure 2.26 shows temporal analysis for number of delinquencies in the borrowers' credit file for the past 2 years. As it is obvious from Figure 2.26(c), accepted applicants who applied in August have the highest number of delinquencies in their credit history. Seasonality analysis of LC grades confirms this observation, i.e. the lowest LC grades for accepted applicants are assigned in August.

2.1.16 Home Ownership Status

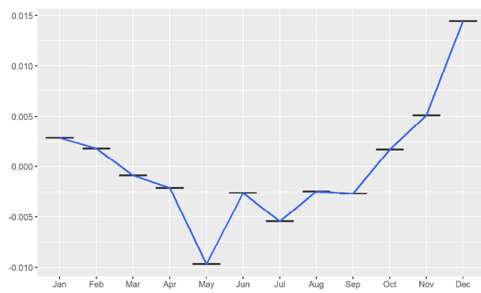
Home ownership status is provided by the applicants. LC categorizes the home ownership status of borrowers as Rent, Mortgage, Own, None, and Any. Figures 2.27, 2.28 and 2.29 illustrate temporal analysis for the ratio of applicants' number with home ownership status of Own, Mortgage, and Rent to the total number of applicants respectively. As Figure 2.27(b) shows, this ratio for applicants whose housing situation is Own is increasing, which confirms the previous notion indicating rich people are trusting this platform more over time. Ratio of accepted applicants who are under mortgage category is increasing as well (Figure 2.28(b)). And, the unusual increase in 2011 can be explained



((A)) Own



((B)) Trend

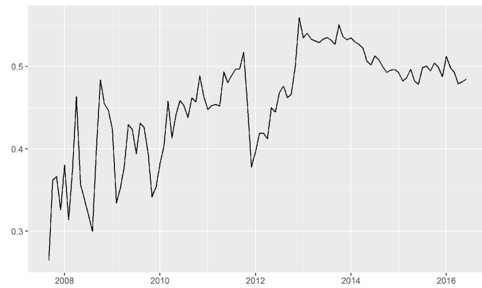


((C)) Seasonality

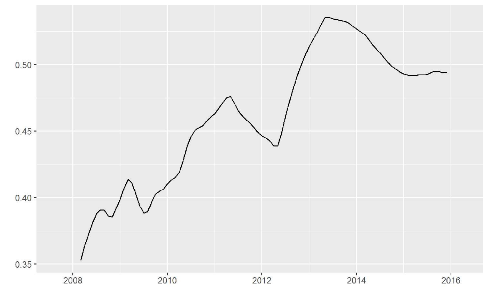
FIGURE 2.27: Applicant home ownership status, own

by the fact that American people went through a deep recession, primarily due to Real Estate bubble. According to Figure 2.29(b), ratio of applicants who rent has a decreasing trend.

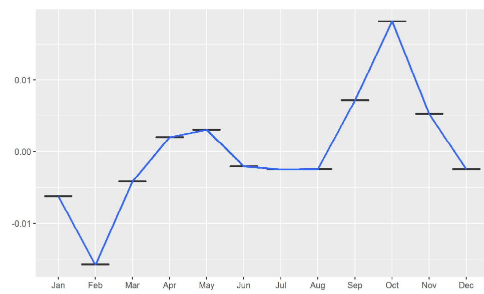
As shown in Figure 2.27(c), the ratio of qualified applicants who own a house has a significant value in December compare to the rest of year. Although the number of applicants in this month is in its lowest value, the ratio of accepted applicants reaches its maximum value at this time. According to the previous analyses, FICO score, annual income, LC grade, loan amount, age of credit history are also significant at the end of seasonal cycle. All in all, we can conclude that although the number of applicants decrease through the seasonal cycle, the quality of applicants improve significantly at the end of this period.



((A)) Mortgage

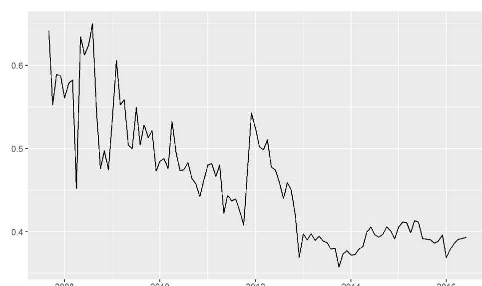


((B)) Trend

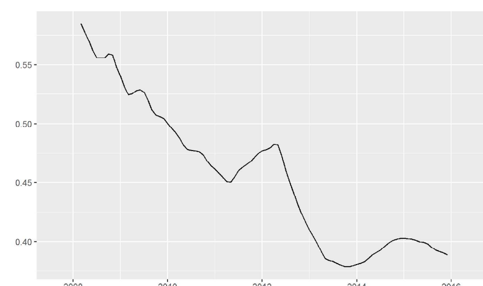


((C)) Seasonality

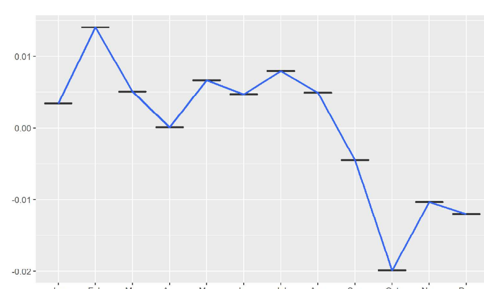
FIGURE 2.28: Applicant home ownership status, mortgage



((A)) Rent



((B)) Trend



((C)) Seasonality

FIGURE 2.29: Applicant home ownership status, rent

Chapter 3

Statistical Forecasting

3.1 Forecasting

3.1.1 Theory

Forecasting is a strategy to make extrapolations using past and present data. Forecasting analyze the existing trends in past and present data to project them into future. There are many different statistical methods in forecasting which based on the data can be either under qualitative forecasting methods or quantitative ones. When there is no available recorded data or even when the available data is irrelevant to the forecasting issue, qualitative forecasting techniques are used. Quantitative forecasting techniques are applicable when there is numerical history of data and the assumption that the past existing patterns of data will continue in the future is sensible. Data which are used in quantitative approaches may be either cross-sectional data, collected at a single point in time, or time series, collected regularly over time. LC applications which are recorded sequentially over time from 2007 up to June 2016 are in the latest group. In this section we discuss some quantitative forecasting methods to estimate how the sequence of applications will continue into the future.

3.1.1.1 Exponential Smoothing

In the Exponential Smoothing method, weighted averages of observations are used to estimate forecasts. The weights associated to recent observations are high and they are decreasing exponentially as the observations get older. This approach is useful when there is no systematic trend or seasonal effects in time series or they have been removed [40].

$$a_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + ... \tag{3.1}$$

where α is smoothing parameter and $0 < \alpha < 1$. When time series include complex seasonal patterns like non integer seasonality or calender effects, Exponential Smoothing is not practical anymore. Also, parameter estimation and consequently the calculations related to forecast become difficult with this method when the number of seasonal components increases in time series. Moreover, Exponential Smoothing method assumes that the error terms are uncorrelated which is not the case for most of the real time series. Because of these deficiencies, there are modifications of Exponential methods like BATS Method (Exponential Smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components). BATS removes the existing correlation among errors by considering an ARMA model for the error terms and using Box-Cox transformation. This modified method is able to deal with wider and complex seasonal patterns.

TBATS model is another flexible modification of Exponential methods which includes trigonometric representation of seasonal patterns. The number of seasonal parameters is less than BATS model. TBATS model is able to model the non-integer seasonal effects since it is based on trigonometric functions [41].

Innovations State Space models (ETS) are another general format of Exponential Smoothing method. ETS model can forecast intervals in addition to forecast the point while Exponential Smoothing methods are only able to generate point forecasts. This method has a three-character string (Error, Trend, Seasonal) based on the framework terminology represented in [29]. Each term has some possibilities as: Error $\{A, M\}$, Trend $\{N, A, A_d, M, M_d\}$ and Seasonal $\{N, A, M\}$. A refers to additive, M refers to multiplicative, A_d is for additive damped, M_d for multiplicative damped and N is for none. All these different models are classified into two groups; ETS model with additive errors and ETS models with multiplicative errors. The point forecasts produced by both kinds of models are identical if they use the same smoothing parameter values. However, they generate different prediction intervals [42].

3.1.1.2 Seasonal Linear Regression

Linear Regression models are suitable if there are deterministic trend and seasonal variations in time series.

Consider time series $X_t : t = 1, 2, ..., n$. A simple linear model can be expressed as

$$X_t = \alpha_0 + \alpha_1 U_{1,t} + \alpha_2 U_{2,t} + ... + \alpha_m U_{m,t} + Z_t \tag{3.2}$$

where $U_{i,t}$ is the value of ith predictor at time t, Z_t is error at time t and $\alpha_0, \alpha_1, ..., \alpha_m$ are model parameters which are estimated by least square error term.

In many real cases, time seasonal effects exist in time series because the observations are measured sequentially in time. Considering these seasonal effects in Linear models can improve the model efficiency. Seasonal effects in Linear models can be defined by an indicator for each season in a cycle like linear model with additive seasonal indicator variables.

For time series $X_t : t = 1, 2, ..., n$, we have $X_t = M_t + S_t + Z_t$, where M_t is trend, S_t is seasonal indicator and Z_t is error term. In this model each season has one parameter estimate. However, there is smooth variation over seasons in real time series. We can use smooth functions such as sine and cosine instead of separate seasonal indicators to make the estimation of parameters more precise. The obtained model is called Harmonic Seasonal Regression model [43].

$$X_t = M_t + \sum_{i=1}^{[S/2]} (s_i sin(2\pi it)/s + c_i cos(2\pi it)/s) + Z_t \tag{3.3}$$

For a time series X_t there are $[s/2]$ possible cycles where s is the number of seasons, for example for monthly data s is equal to 12. M_t is trend which is indeed a parameter for constant part of model, Z_t is error and s and c are unknown parameters [37].

3.1.1.3 Autoregressive Integrated Moving Average

The combination of Autoregressive (AR) model of order p and Moving Average (MA) model of order q result in ARMA model with order of (p, q). ARMA models use weighted

"q" most recent white noise terms and weighted "p" current values and take summation of them to forecast a value at time t.

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + W_t + \beta_1 W_{t-1} + \beta_2 W_{t-2} + ... + \beta_q W_{t-q} \quad (3.4)$$

where α and β are model parameters and W_t is white noise. ARMA models are for stationary time series, e.g. time series statistical properties are all constant over time. Many time series are non-stationary due to the existing trends or seasonal effects. Differencing time series can transform the non-stationary time series to stationary ones by removing the trend, whether the stochastic trend or deterministic one. A time series X_t follows an ARIMA model of order (p, d, q) if the dth differences of X_t is ARMA(p, q) model, d is the number of differences applied until time series become stationary. ARIMA models are applicable for wide range of forecasting situations [44] [45].

A seasonal ARIMA model(SARIMA) uses differencing equal to the number of seasons (s) to remove additive seasonal effects. SARIMA model is formed by including additional seasonal terms in the ARIMA models, ARIMA$(p, d, q)(P, D, Q)_s$ [44].

When d, the difference operator, can take non-integer value, ARIMA model transferred to ARFIMA model, Autoregressive Fractionally Integrated Moving Average. This model is useful for modeling persistent time series for long horizons [46].

3.1.1.4 STL decomposition and forecasting

For decomposing time series, STL method can be used as a robust model. It comes from “Seasonal and Trend decomposition using Loess”. In this model, Loess method is used to estimate the existing nonlinear relationships. The STL method was developed by Cleveland et al. [47].

Not only decomposition is useful for exploring time series changes over time but also it can be used for forecasting. In additive decomposition, the decomposed time series can be written as $X_t = S_t + T_t + E_t$, where S is for seasonal effect, T refers to trend and E is for error term. And for multiplicative decomposition we have $X_t = S_t T_t E_t$.

The summation of trend and error term in additive decomposition and the multiplication of trend and error in multiplicative one called seasonally adjusted component. Seasonal term and the seasonally adjusted term are estimated separately to forecast time series. In this method the seasonal term of last year is used to estimate the seasonal term since the seasonal part changes is extremely insignificant. To forecast the seasonally adjusted component, any non-seasonal forecasting model such as non-seasonal ARIMA model can be used.

3.1.2 Experiments

In this section, we present the results of implementing described forecasting models in our dataset. Performance of these models are evaluated based on Akaike Information Criterion (AIC) [48]. During last decades, AIC is used as one of the fundamental parameter for statistical models' evaluation. It has been widely used since its away from inferential and restrictive methods of evaluation. By using AIC and its modification, we are able to avoid over fitting and under fitting problems. AIC is based on maximizing the likelihood function (the probability of obtaining the data given the model) and penalizing models with too many parameters.

$$\text{AIC} = -2 \text{ log-likelihood} + 2 \text{ number of parameters} \tag{3.5}$$

Based on this evaluation, the lowest AIC provides the best model which fits to the existing data. After selecting the best model for each attribute, we perform forecasting for 30 months starting from July 2016 up to December 2018. Tables 3.1-3.4 present AIC values of different forecasting methods for borrowers' characteristics, credit history, loan characteristics, and borrower assessment respectively. At each table, the bold AIC value for each attribute corresponds to the optimal AIC and represents the selected forecasting method.

Table 3.1: AIC, Borrower Characteristics

Forecast Method	Annual Income	Employment Length	DTI	Own	Mortgage	Rent
Exponential Smoothing	743.13	205.96	422.98	-356.29	-227.60	-223.30
BATS	741.86	223.54	461.85	-362.11	-226.87	-242.59
TBATS	741.86	223.54	436.22	-362.11	-226.87	-242.59
Seasonal Linear Regression	584.51	151.32	327.08	-517.58	-347.53	-371.98
ARIMA	**533.01**	11.06	233.70	**-559.63**	-415.93	-396.34
ARFIMA	535.82	**-8.48**	**228.51**	-556.80	**-441.79**	**-416.62**
ETS	743.13	201.13	418.15	-356.29	-227.60	-249.28
STL	704.15	177.71	384.44	-389.91	-263.79	-277.85

Table 3.2: AIC, Credit History

Forecast Method	Delinquency	Inquiry	Open Accounts	Revolving Utilization	Credit Age
Exponential Smoothing	-254.22	-54.47	271.08	606.41	436.35
BATS	-254.52	-87.07	313.29	601.22	446.58
TBATS	-256.29	-87.06	285.20	603.88	446.58
Seasonal Linear Regression	-297.22	-128.77	204.88	471.10	414.79
ARIMA	-442.44	**-232.00**	70.35	411.76	**243.50**
ARFIMA	**-456.32**	-228.57	**65.88**	**385.51**	243.85
ETS	-258.57	-101.80	267.78	606.41	435.90
STL	-285.33	-127.39	237.24	564.99	402.08

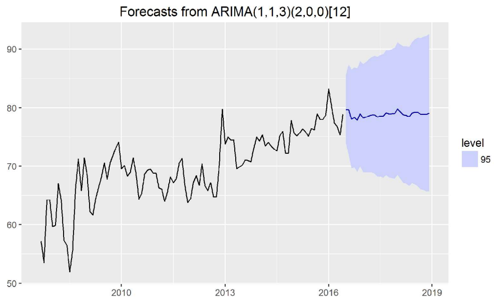
Table 3.3: AIC, Loan Characteristics

Forecast Method	Loan Amount	Credit	Debt Consolidation	Home Improvement	Others
Exponential Smoothing	421.29	149.59	-219.11	9.54	-310.08
BATS	420.41	91.81	-217.87	-93.42	-393.89
TBATS	418.96	92.67	-217.87	-98.00	-391.37
Seasonal Linear Regression	318.02	**-57.24**	-318.10	**-195.49**	-428.30
ARIMA	**210.55**	-49.69	-425.83	-186.27	-500.17
ARFIMA	226.84	-44.57	**-428.23**	-188.04	**-523.31**
ETS	421.29	69.83	-219.11	-169.54	-381.42
STL	361.97	87.52	-261.10	-53.82	-432.00

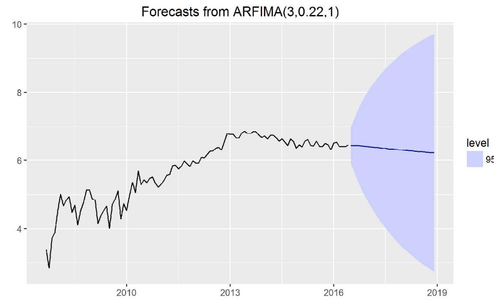
Table 3.4: AIC, Borrower Assessment

Forecast Method	FICO Score	LC Grade
Exponential Smoothing	703.98	88.96
BATS	705.34	85.29
TBATS	705.34	85.29
Seasonal Linear Regression	801.77	58.09
ARIMA	**504.69**	-100.28
ARFIMA	510.84	**-117.95**
ETS	703.98	88.96
STL	674.60	48.40

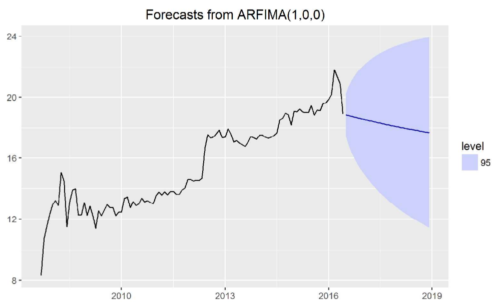
Figures 3.1-3.4 demonstrate 30 month forecasting values for borrowers' characteristics, credit history, loan characteristics and borrowers' assessment respectively. As Figure 3.1(a) indicates, the financial level of applicants would be relatively higher in the future which suggests future success of LC in attracting applicants with better financial status. Regarding Figures 3.1(b)-3.1(e), applicants' employment length, DTI, trend of borrowers who own a house and applicants of mortgage would slightly decrease but almost keep their constant trend. However, as Figure 3.1(f) indicates, number of applicants whose home ownership is rent is increasing significantly. According to these forecasting outputs, quality of borrowers' characteristics is not improving, which may not be good news for future lenders.



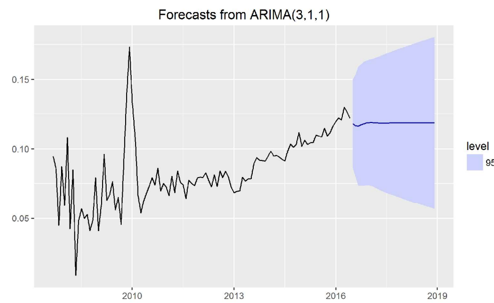
((A)) Annual Income(×1000)



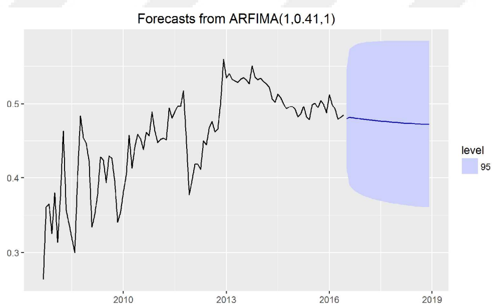
((B)) Employment Length



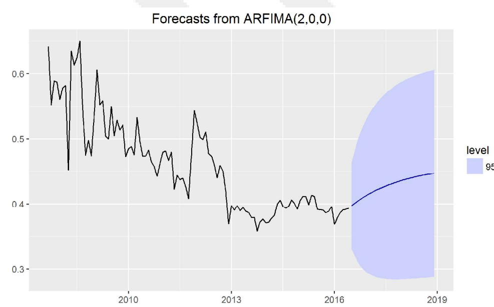
((C)) Debt to Income Ratio



((D)) Own



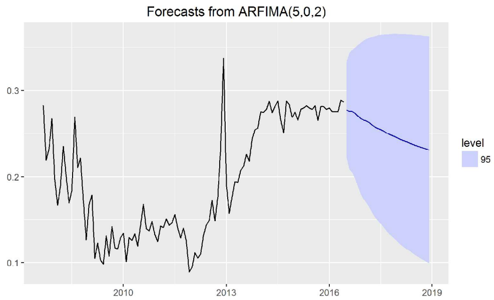
((E)) Mortgage



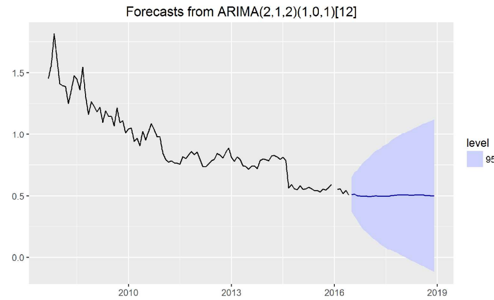
((F)) Rent

Figure 3.1: Forecasting plots of attributes related to borrowers' characteristics

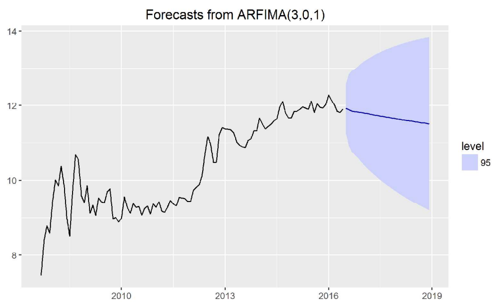
Figures 3.2(a) and 3.2(c) show a decreasing trend for future applicants' number of delinquencies and open accounts. According to Figures 3.2(b), 3.2(d), and 3.2(e), the future trends seem to be stable. Having steady and even decreasing trend of attributes related to applicants' credit file except the age of credit designate improvement of future applications' quality concerning these attributes.



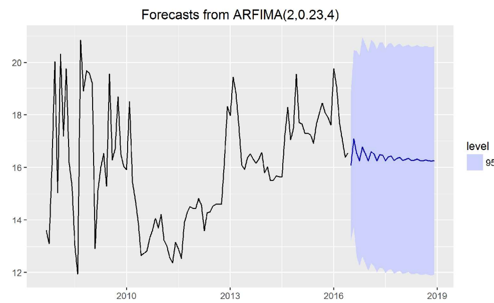
((A)) Number of delinquencies



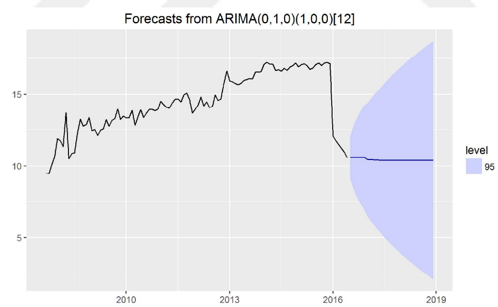
((B)) Inquiries in last 6 months



((C)) Open accounts



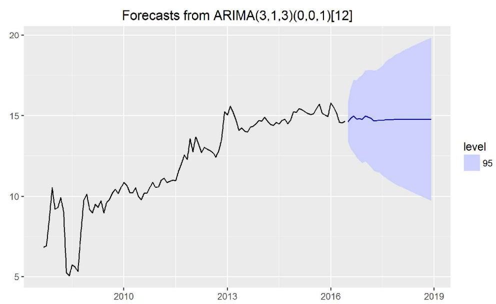
((D)) Credit revolving balance



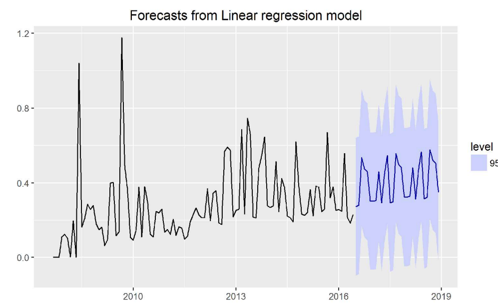
((E)) Credit age (years)

FIGURE 3.2: Forecasting plots of attributes related to applicants' credit history

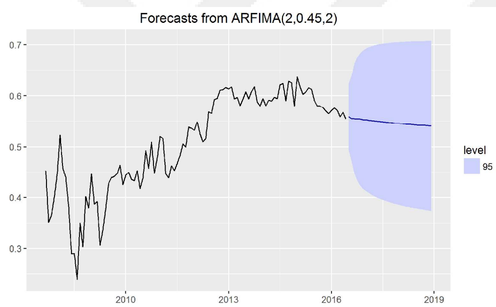
Our forecast indicates that the stable trend in the amount of loan that LC approves will not change significantly (see Figure 3.3(a)). Among loan purposes, the number of applicants with the loan purpose of credit card will increase (see Figure 3.3(b)), which signifies the success of LC to prove being a proper source for borrowers to pay their debts. But the increasing number of applicants whose purpose of borrowing is not among predefined 14 purposes may not be a good sign (see Figure 3.3(e)). Risks of loans with different purposes are not equal for lenders. Risk of investing on a loan with the purpose of debt consolidation is not the same as investing on a loan with wedding purpose. And this risk will be higher for applications with unknown purposes. As Figure 3.3(c) and Figure 3.3(d) show the number of applications with these two purposes will not have considerable fluctuations.



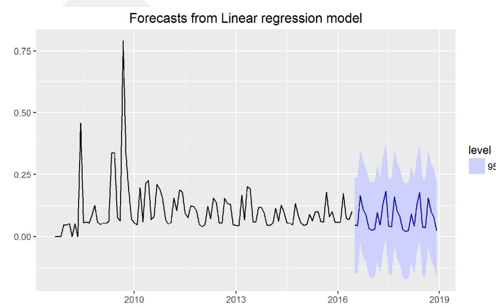
((A)) Loan amount (dollars)



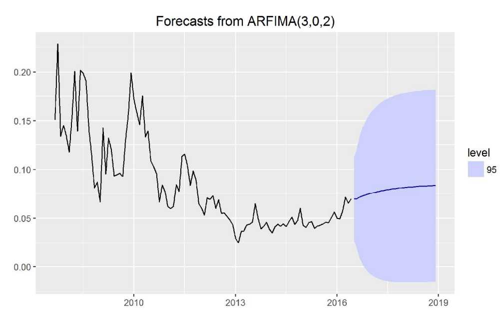
((B)) Loan purpose; Credit card



((C)) Loan purpose; Debt Consolidation



((D)) Loan purpose; Home improvement

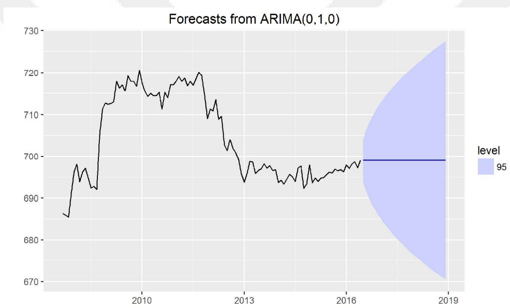


((E)) Loan purpose; Others

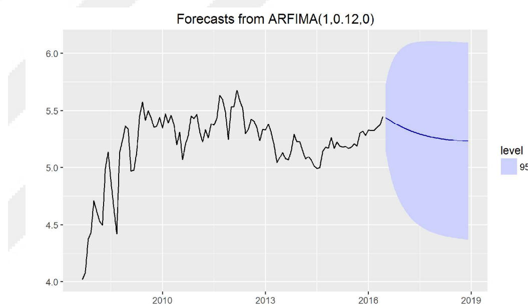
Figure 3.3: Forecasting plots of attributes related to loan characteristics

FICO score of future applicants is not expected to change as Figure 3.4(a) displays. However, LC grade will decay slightly (see Figure 3.4(b)). Although LC is strict in approving the applications, the decreasing trend of LC grade indicates that future applicants of LC will not be as qualified as they are now.

The results from Figure 3.4 could clarify that with the current policy how LC is going to select its borrowers up to December 2018. Our forecasting results show that lenders should be more careful in their future selection. Borrowers' income will be better, however, their DTI and credit revolving balance will drop and as a consequence the grade assigned by LC will get worse. It also indicates that the general place that LC will stand in future is a little bit threatening for lenders. In addition, they will need more thorough examination to find potentially good applications to invest on.



((A)) FICO score



((B)) Lending Club grade

FIGURE 3.4: Forecasting plots of attributes related to borrowers' assessment

Chapter 4

Summary, Conclusions, and Directions for Future Research

4.1 Summary and Conclusion

Social lending is one of the current widespread industries. Platforms providing this type of lending offer micro loans where all tasks and transitions happen online. People are eager to attend this industry whether as a borrower or lender. On one hand, borrowers can receive loans with lower interest rates in comparison to banks, and on the other hand, lenders obtain more benefits rather than investing their money in banks or other financial portfolios. Lenders are able to access the borrowers' files and decide whether to invest their money into this financial scheme or not. They can decrease the risk of losing money by dividing their investment among different good potential borrowers. Lending Club (LC) as one of the prominent social lending platforms is available in all the states of US except Iowa and West Virginia. In this platform, applications' features like borrower characteristics, credit history, and loan characteristics have been recorded over time since LC's infant stage. This research provides evaluations regarding main applicants' features recorded between 2007 and 2016 by looking into their temporal changes separately. Since it has been proven that these features are highly correlated with an application fault or success, this evaluation can be used as a compass for lenders who invest their money based on their analysis about candidate applications. This analysis is warranted for lenders because they are the only party in this industry who carry the whole burden when a loan becomes defaulted. In this regard, we conducted time series analysis for existing data and extracted the trends and seasonal variations for each important attribute. The main aspect of this study is proposing the likely causes of feature changes and justifying the existing trends and seasonality in order to forecast them properly. In the forecasting section, first we measured functioning level of some critical forecasting methods by AIC for our dataset. Then, employing the best method for each attribute, we modeled the trend and seasonal variations of borrowers attributes and extrapolated the fitted model into the future for the next thirty months. Our temporal analysis and predicted values reveal that among significant attributes, future applicants' annual income, DTI, and credit revolving balance will improve, indicating increase in the applicants financial quality. However, the stable trend of FICO score and the decreasing trend of the grade LC assigns to each application are negative indicators for creditworthiness of future applications. Consequently, it is clear that lenders should not blindly trust LC accepted applicants and they need to examine the future applications more carefully to find promising applications to invest on.

4.1.1 Future Works

As future works, same investigations with the help of temporal analysis need to be performed in other social lending platforms to help their lenders. On the other hand, although our work compared several methodologies in the forecasting section, there are still some other forecasting techniques not investigated such as multivariate analysis, neural network, multiple regression analysis, etc. Adding these techniques and finding another criteria for the comparison seems to be a good future research. Furthermore, some features of applicant related to their personality could definitely affect the predisposition to financial risk, such as impulsiveness, impatience, and mobility. Introducing social features can help us to capture personal characteristics of the platform users rather than relying solely on financial features. Also, ignoring the social features would exclude applicants without any prior financial history. Therefore, same analysis on LC's social attributes of borrowers, beside financial attributes, would definitely improve the quality of this investigation, which requires integration of borrowers' social media accounts.

Bibliography

[1] M. Greiner and H. Wang. Building consumer-to-consumer trust in e-finance marketplaces: An empirical analysis. *International Journal of Electronic Commerce*, 15 (2):105–136, 2010.

[2] Burkhardt B. Funk, Alexander Bachmann, Alexander Becker, Daniel Buerckner, Michel Hilker, Frank Kock, Mark Lehmann, and Phillip Tiburtius. Online peer-to-peer lending? a literature review. *The Journal of Internet Banking and Commerce*, 2011, 2015.

[3] Gartner. Gartner says social banking platforms threaten traditional banks for control of financial relationships. `http://www.gartner.com/newsroom/id/597907`, 2008. Accessed: 2016-07-30.

[4] B. Slavin. Peer-to-peer lending-an industry insight. *Retrieved from bradslavin. com*, 2007.

[5] M. K. Hulme and C. Wright. Internet based social lending: Past, present and future. *Social Futures Observatory*, 115, 2006.

[6] A. Ashta and D. Assadi. Do social cause and social technology meet? impact of web 2.0 technologies on peer-to-peer lending transactions. *Cahiers du CEREN*, 29: 177–192, 2009.

[7] I. Galloway et al. Peer-to-peer lending and community development finance. *Community Investments*, 21(3):19–23, 2009.

[8] E. Berkovich. Search and herding effects in peer-to-peer lending: evidence from prosper. com. *Annals of Finance*, 7(3):389–405, 2011.

[9] B. Jirasakuldech R. Emekter, Y. Tu and M. Lu. Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending. *Applied Economics*, 47(1):54–70, 2015.

[10] B. Gutiérrez-Nieto C. Serrano-Cinca and L. López-Palacios. Determinants of default in p2p lending. *PloS one*, 10(10):e0139427, 2015.

[11] M. Klafft. Online peer-to-peer lending: a lenders' perspective. In *Proceedings of the international conference on E-learning, E-business, enterprise information systems, and E-government, EEE*, pages 371–375, 2008.

[12] F. Lai D. Chen and Z. Lin. A trust model for online peer-to-peer lending: a lenderâs perspective. *Information Technology and Management*, 15(4):239–254, 2014.

[13] E. Lee and B. Lee. Herding behavior in online p2p lending: An empirical investigation. *Electronic Commerce Research and Applications*, 11(5):495–503, 2012.

[14] H. Wang and M. E. Greiner. Herding in multi-winner auctions. In *ICIS*, page 235, 2010.

[15] M. Malekipirbazari and V. Aksakalli. Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42(10):4621–4631, 2015.

[16] A. Ghose G. Burtch and S. Wattal. Cultural differences and geography as determinants of online pro-social lending. *MIS Quarterly, Forthcoming*, pages 14–021, 2013.

[17] U. M. Dholakia M. Herzenstein, R. L. Andrews and E. Lyandres. The democratization of personal consumer loans? determinants of success in online peer-to-peer lending communities. *Boston University School of Management Research Paper*, 14 (6), 2008.

[18] S. Siegel J. Duarte and L. Young. Trust and credit: the role of appearance in peer-to-peer lending. *Review of Financial Studies*, 25(8):2455–2484, 2012.

[19] J. Larrimore D. Markowitz L. Larrimore, L. Jiang and S. Gorski. Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, 39(1):19–37, 2011.

[20] C. R. Everett. Group membership, relationship banking and loan default risk: the case of online social lending. *Banking and Finance Review*, 7(2), 2015.

[21] S. Herrero-Lopez. Social interactions in p2p lending. In *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, page 3, 2009.

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

[23] K. Pelger G. NF. Weiss and A. Horsch. Mitigating adverse selection in p2p lending–empirical evidence from prosper. com. *Available at SSRN 1650774*, 2010.

[24] O. Malik S. S. Soman, H. Zareipour and P. Mandal. A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium (NAPS), 2010*, pages 1–8, 2010.

[25] K. Kim. Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1):307–319, 2003.

[26] J. Du Preez and S. F. Witt. Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting*, 19(3):435–451, 2003.

[27] R. Abdel-Aal. Univariate modeling and forecasting of monthly energy demand time series using abductive and neural networks. *Computers & Industrial Engineering*, 54(4):903–917, 2008.

[28] F. EH. Tay and LJ. Cao. Modified support vector machines in financial time series forecasting. *Neurocomputing*, 48(1):847–861, 2002.

[29] R. D. Snyder R. J. Hyndman, A. B. Koehler and S. Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002.

[30] M. Marchesi G. Armano and A. Murru. A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*, 170(1):3–33, 2005.

[31] S. R. Hejazi M. Khashei and M. Bijari. A new hybrid artificial neural networks and fuzzy regression model for time series forecasting. *Fuzzy Sets and Systems*, 159(7):769–786, 2008.

[32] R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

[33] C. WJ. Granger and Y. Jeon. Comparing forecasts of inflation using time distance. *International Journal of Forecasting*, 19(3):339–349, 2003.

[34] P. A. Thompson. An mse statistic for comparing forecast accuracy across series. *International Journal of Forecasting*, 6(2):219–227, 1990.

[35] L. Wun and W. L. Pearn. Assessing the statistical characteristics of the mean absolute error or forecasting. *International Journal of Forecasting*, 7(3):335–337, 1991.

[36] S. Leybourne D. Harvey and P. Newbold. Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291, 1997.

[37] P. SP. Cowpertwait and A. V. Metcalfe. *Introductory time series with R*. Springer Science & Business Media, 2009.

[38] E. C. Chaffee and G. C. Rapp. Regulating online peer-to-peer lending in the aftermath of dodd-frank: In search of an evolving regulatory regime for an evolving industry. *Wash. & Lee L. Rev.*, 69:485, 2012.

[39] Erin Lowry. 6 simple steps to improve your credit score. `http://www.magnifymoney.com`, 2014. Accessed: 2016-10-25.

[40] T. B. Fomby. Exponential smoothing models. *Southern Methodist University Dallas Department of Economics*, 2008.

[41] R. J. Hyndman A. M. De Livera and R. D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, 2011.

[42] Rob J Hyndman, Yeasmin Khandakar, et al. Automatic time series for forecasting: the forecast package for r. Technical report, Monash University, Department of Econometrics and Business Statistics, 2007.

[43] Michael D Intriligator. Econometric models, techniques, and applications. Technical report, Prentice-Hall Englewood Cliffs, NJ, 1978.

[44] G. C. Reinsel G. EP. Box, G. M. Jenkins and G. M. Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[45] M. Küçükvar M. Yalçıntaş, M. Bulu and H. Samadi. A framework for sustainable urban water management through demand and supply forecasting: The case of istanbul. *Sustainability*, 7(8):11050–11067, 2015.

[46] J. Xiu and Y. Jin. Empirical study of arfima model based on fractional differencing. *Physica A: Statistical Mechanics and its Applications*, 377(1):138–154, 2007.

[47] J. E. McRae R. B. Cleveland, W. S. Cleveland and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1): 3–73, 1990.

[48] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.