Ca' Foscari University of Venice

Master's Degree Program

Of Economics and Finance

Final Thesis

P2P Loans and States heterogeneity in the US marketplace

Supervisor
Ch. Prof Loriana Pelizzon

Assistant Supervisor
Ch. Prof Michele Costola

Student
Michele Meggiolaro
Matriculation number 989061

Academic Year
2016/2017

Abstract:

This thesis studies whether lenders' perceptions can affect the loan decision process in the peer-to-peer (P2P) lending marketplace. The analysis is performed considering the world's largest P2P lending platform, Lending Club, a US company based in San Francisco, California. To provide consistent results we consider all applications submitted to the platform from its origin to nowadays while the focus of analysis is on each Member State of the United States of America. We develop a series of models to estimate which effect can have the State in which borrowers live concerning their likelihood to have a loan granted using variables such as personal income, unemployment rate, gross domestic product, personal consumption expenditure, total amount of non-performing loans and the number of the financial institutes for each State. We find that, once individual's financial and employment characteristics are taken into account, the regionals more than the States heterogeneity effectively affect lenders' decision process, at least in the US. Besides, using the regional clusterization provided by the Bureau of Economics Analysis, we find that individuals who live in the Far West tendentially have a higher probability of getting a loan concerning other US citizens who live in other regions. On the contrary, southeastern people are less likely to get a loan. The reason for this statistical discrimination might be related to the prejudices and stereotypes associated with those the lenders consider alien to them rather than only economic factors.

.

Table of Contents

Introduction

Since its origins, the credit market has been the best way for financial intermediaries to gain revenues exploiting the broad needs for liquidity.

In the past, one who wanted to ask for liquidity inevitably had to take into considerations the financial intermediaries, which offered the unique way to have a loan. There were no other alternatives. If someone would have wanted to develop his business or buying its own house, a car or simply make it easier to reach his desires, he had to deal with conditions imposed by financial institutions. Moreover, with the onset of the global financial crisis, the relationship between banks and their customers have been degraded by even stringent constraints imposed by the former. These are only a few reasons why, in recent years, a new form of finance and lending have been emerging in the marketplace.

Thanks to the development of the internet network, individuals, nowadays, have seen their possibility of connecting each other grow up. Also, technology innovations and the growing misalignment in the market boosted the creation of many alternatives to conventional financial intermediaries. Through this thesis, we will analyze one of these new forms of finance: the P2P lending.

Since its origin, P2P lending has experienced rapid and unpredicted growth and, to date, has obtained a significantly share in the credit market. The distinctive and innovative feature of P2P lending is that it provides loans without recurring to any types of financial intermediaries, in a much simpler and quicker way compared to typical bank loans. The basic idea is to match borrowers and lenders directly. P2P lending platforms, exploiting the internet network, allow potential borrowers to compile an easy online application form and, after evaluating for borrowers' creditworthiness and information, to reach a wider number of retail and institutional investors to fund their loan requests. The purpose of this thesis is strictly connected to this aspect of P2P lending, and the final goal is to analyze the effects of borrowers' characteristics on their likelihood of getting a loan. Afterwards, the focus loves on the State, where the borrowers live, and how its' economic characteristics may affect this likelihood, once all the other traits are taken into consideration.

The results of this study want to provide a new point of view on the loan decision process over the US countries. Besides, the focus is on Lending Club, the world's largest P2P lending platform.

Chapter 1 – The P2P overview

1.1 History: The P2P Industry

The scope of this analysis is to argue how individuals and States' characteristics can affect lenders' decisions. Therefore, to better understanding these mechanisms it is necessary to describe how P2P industry is structured and, above all, which is the lenders approach towards the increasing amount of requests submitted to lending platforms. This chapter provides an overview of the P2P finance with a focus on P2P Lending starting from its origins.

1.1.1 The origins of P2P industry

The term peer-to-peer has been introduced in recent years to describe the connection between two parties with the absence of a central intermediary. In the beginning, "peer-to-peer" was used in computer networking for describing "a network where any computer con act as either a client or a server to other computers on the network, without considering the need of a central intermediary" (Milne e Parboteeah 2016). The recent growth of the internet and digital transformation in all aspects of our lives improve the development of a broad diffusion of P2P activities. The first activity to become widely adopted is file sharing where using a software platform, users could connect directly to other users on the same platform to share files such as photo or movies. The worldwide adoption of such kind of P2P file sharing has had a huge impact on the music and film industry and in particular on the sales of physical product. Whether the overall impact on these industries has been a positive or negative is an open debate and is out of the scope of this thesis.

In the US, the origin of P2P finance can be traced back to 2006 and is strictly connected with the launch of Prosper marketplace and Lending club in the next year. The mission of these companies is to provide facilitated peer-to-peer lending. The innovative idea is to connect borrowers and lenders directly providing them a central marketplace to deal with each other. The fundamental idea of this platform could be summarized in the quote of Chris Larsen, co-founder of Prosper, who described the offering of the company as an "eBay for credit." The divergence from the past is clear since, in this new context, there is no room for traditional financial insitutions.

Nowadays, Peer-to-Peer (P2P) lending platforms are constantly increasing their impact in the world of the financial services. To quote just two example, Lending Club originated over $11 billion in loans and Prosper $9 billion as of the first quarter of 2017 and, since its origin, the overall origination volumes of US P2P lending platforms have grown on average of 84% per quarter. (PWC LLP 2015)
For this reason alone, it is clear why P2P lending has gained attraction in recent years in the US and Europe. Since 2007, a wide range of alternative peer to peer financial services, operating outside of conventional banking and capital markets, have emerged. The independence of financial institutions, governments, and central banks is a defining feature for all of them. For example:

- Crowdfunding, where many smaller contributions from individuals are raised for a specific project or a company.
- Market Invoice for invoice finance.
- Alternative foreign exchange platforms, where individuals and businesses exchange foreign currencies without recurring to banks or financial intermediaries.
- Non-bank invoice discounting, where small firms can improve their cash flows by securing advances from investors against invoices due.
- Cryptocurrencies, such as Bitcoin, LiteCoin, and Ripple. Digital assets, characterized by the absence of a central issuer and which support instant online payments.

1.1.2 The advantages of P2P lending

As previously highlighted, an outstanding growth has characterized the P2P lending platforms in recent years, and their business doubles annually. "Their perceived cost and other advantages relative to established banks have led some analysts to make quite ambitious projections about the extent to which P2P lending" (Milne e Parboteeah 2016). Pricewaterhouse&Coopers' analysts indicate that the market could reach 150 $ billion or higher by 2025 and, for this reason, P2P lending could achieve a relevant market share in lending markets (PWC LLP 2015). There are many other reasons to support this forecasts. For instance, the exploiting of Internet opportunities eases the communication between parties is one fundamental reason. Moreover, some competitive advantages of the P2P lending platform over traditional financial vendor

strengthen the potential for growth. During the last years and particularly due to the 2008 financial crisis, regulations have become strict so that banks must account credit provisions and balance sheet issues. Consequently, as banks foresee a customer default, it is mandatory to write down the loan from their balance sheet and, whether recovered, accounted over again. In this context, lending for riskier clients may generate unexpected volatility on balance sheet size. Moreover, banks' credit provisions are strictly linked to relationship lending and react not only to market conditions.
The absence of this kind of issues strengthens the competitive advantage of P2P lending. The four categories of advantages are the following (Milne e Parboteeah 2016):

- P2P platforms provide better rates of return for investors than rates available on conventional bank saving deposits. Moreover, this aspect, together with relatively low fees, enhances the likelihood of a market match between demand and supply. This effect is due to the within nature of P2P activities that required relatively low overhead and administrative costs to be set. Another important difference from banking sector is the direct connection between borrowers and lenders that results in the absence of additional margins of interest. On the contrary, lenders in P2P platforms are exposed to greater risks (no deposit insurance and no promise of return), nevertheless are compensated by much higher rates of return.
- P2P lending allows broader access to the credit market. Banks and conventional lenders, since the onset of the global financial crisis, have been less willing to provide money to potential borrowers. Because of many individuals and small business were no longer able to satisfy the more stringent constraints imposed on granting loans. Through P2P platforms, these categories of borrowers can find alternative lenders who are willing to take on the risk of providing loans as well as better interest rates or conditions. (Milne and Parboteeah 2016)
- P2P lending platforms, connecting borrowers directly to lenders, are perceived to offer a more socially beneficial form of finance. There is a common understanding that banks and traditional financial intermediaries tend to exploit their market power and pursue profits without regard to their own consumers' interests. This aspect, however, is depleted by increasing presence of institutional investors such as lenders on P2P lending platforms. (Milne and Parboteeah 2016)

- The technological advantage of these platforms. P2P lending platforms are trying to compete with established banks design and implement operational systems that do not require continuity with older legacy systems. Therefore, P2P lending platforms can offer a better quality of service to both borrowers and lenders. On borrowers' side, they can offer simpler loan application process with a rapid decision and a transparent portal for monitoring outstanding commitments and repayments. On lenders' side, they provide portals for lending management and check for investment positions. All these innovative features are absent in traditional banks that are usually bound by huge and large legacy systems. These systems are difficult to replace due to the IT infrastructure built around them. (Milne and Parboteeah 2016)

Thanks to all these advantages, it is clear why P2P lending platforms have been hitting the market since its origins. In the next section, we will focus on the growth and outlook of P2P lending platforms in the US market.

1.1.3 The growth of P2P Lending in the US

Although several institutional and regulatory differences between P2P lending in US and Europe there are many similarities in the approach to the market. P2P lending in the US is much more focused on consumer credit. The US industry has further away evolved from the original concept of directly linking individual lenders and borrowers, becoming instead largely a mechanism for the sale of loans to institutional investors. Despite this rather different approach and orientation, the P2P lending marketplace, in the US, still, capture a relatively small share of the completely unsecured consumer loans. Since the origin of Lending Club and Prosper, the oldest and largest market-based platforms, the P2P lending market was oriented to offer unsecured customer lending and student loans' refinancing. After these "pioneers" of P2P lending, were established many other platforms such as Avant, which is focused on personal loans, and SoFi specialized in refinancing student loans. Regarding the market place for small businesses, OnDeck, CAN Capital and Kabbage. GroundFloor and Lending Home provide short-term bridge mortgage finance. The most were established after the global financial crisis exploiting favorable market conditions and lack of confidence in the traditional banking sector. (Milne e Parboteeah 2016)

In summary, to date, The US continues to be one of the world's top markets for advanced, technology-enabled, online alternative finance channels and instruments. The 2016 US market volume of $34.5 billion marked a 22% year-on-year increase from 2015. As mentioned previously, since its origin, the approach has changed, and institutional investors provided approximately $19 billion, or 55% of the total US alternative finance volume. Moreover, together with the enhancement of macroeconomic conditions, entries of new platforms have slowed while many smaller platforms have exited. This market selection is because older platforms are consolidating their market position increasing their services portfolios and quality. (Ziegler, et al. 2017)

From a regulatory point of view, the development of P2P lending in the U.S. is strictly related to the evolution of the laws and regulations applied to states. For example, in many states, there is a regulatory limit on consumer loan interest rates. To deal with this concern, usually, lenders work with partner banks who formally grant loans once they are agreed in P2P platforms. Furthermore, P2P lending platforms do not only need to comply with Security and Exchange Commission (SEC)[1] regulations, but they also have to be compliant with the respective State laws. The immediate priority of the regulators is an appropriate oversight on operational risks and customer protection. Competent authorities are extremely concerned about the need for consumer and prudential.

1.2 Regulations

The US Consumer Financial Protection Bureau (CFPB)[2] is increasingly involved in the oversight of P2P lending. Besides, its supervision has brought to a well-publicized

[1] The U.S. Securities and Exchange Commission (SEC) is an independent agency of the United Stated federal government that holds primary responsibility for enforcing the federal securities laws, proposing securities rules, and regulating the securities industry, the nation's stock and options exchanges, and other activities and organizations, including the electronic securities markets in the United States. (Wikipedia 2017)

[2] The Consumer Financial Protection Bureau (CFPB) is an agency of the United States government responsible for consumer protection in the financial sector. CFPB jurisdiction includes banks, credit unions, securities firms, payday lenders, mortgage-servicing operations, foreclosure relief services, debt collections and other financial companies operating in the United States. (Wikipedia 2017)

enforcement action against Lending Club for lack of clarity on interest rates paid by one group of borrowers.

The FDIC[3] has announced that it wishes to keep a close watch on developments in marketplace lending, including potential risks to insured bank partnerships. In general, US regulators are acting to ensure an adequate oversight without blocking the financial innovation and the use of P2P platforms.

1.3 The Business Model of P2P Lending

The focus of this section is on the business model of P2P Lending platforms and its main differences with traditional banks' business model.

The basic idea on which P2P lending platforms are built consists in the simplicity of their business. Every individual could submit his application for a loan amount between 1,000 $ and 40,000 $. The application format is very simple and asks for basic information such as age, employment, and income; while, each applicant is assessed for his creditworthiness after submission.

As already said, P2P lending platforms directly match borrowers and lenders without any form of intermediation. This innovative approach widely differs from the one of a traditional bank. In fact, banks lend their funds and consequently face funding issues. P2P platforms put borrowers, who are seeking a loan, to investors, who purchase notes or securities backed by notes issued by platforms; moreover, revenues framework is heavily different from the banks one.

P2P lending platforms generate revenues from fees that are charged with the different proportion of borrowers and lenders (as servicing fees). Furthermore, the remaining part of the interests, charged on borrowers on loan, constitutes the effective revenues collected by the investors. Another key feature and relevant difference between P2P lending platforms' business model and traditional banks' one is the very simple and

[3] The Federal Deposit Insurance Corporation (FDIC) is an independent agency created by the Congress to maintain stability and public confidence in the nation's financial system. Its main functions are the insurance of deposits, the examination and supervision of financial institutions for safety and soundness, consumer protection, the establishment of large and complex financial institutions resolvable and its receivership. (FDIC 2018)

quick online application process that introduce potential borrowers into a rapid assessment procedure with the possibility to follow the status of their loan application. The above business model it is different from the one of the conventional banking system. Indeed, traditional banks, whose main activity is the provision of liquidity, have a more structured model that offer a broad range of services like deposits, lending, guaranties and securities trading. This well-diversified business model allows banks to exploit economies of scale with a consequent increasing in their operative margin. Due to all these aspects, approaching the credit market, traditional banks need to incorporate many skills and competences to monitor the behavior of borrowers and manage all legal and administrative issues.

Therefore, it is evident the reason why P2P platforms in the US are attractive for borrowers rather than lenders and why the most of investors are institutional. Many consumers purse to borrow money from P2P platforms for two main reasons: benefit from comparatively low interest rates and access credits otherwise unavailable in the bank system. Conversely, for investors, above all risk-adverse ones enter into the P2P lending market rather than making a bank deposit means losing the deposit insurance protection and taking into consideration unpredictable risks on unconventional products. However, to make themselves more attractive to the investors, P2P lending usually offer better rates than bank deposits and guarantee a high level of traceability of the investments through a performance-monitoring panel. For these reasons, the overall impact of P2P lending should be seen as complementary to, rather than competitive with bank lending offerings.

As experiences confirm, traditional banks and P2P lending platforms have set up a sort of cooperation in recent years. This collaboration between banks and P2P platform allows marketing P2P borrowing to their customers and improving the availability of credit to a larger portion of the population.

1.4 Introduction to the Lending Club

For the objective of this thesis, which is analyzing how the P2P lending market manages and decide whether or not granting a loan request, Lending Club data are taken into consideration. In this section are exposed reasons of this choice and main characteristics of the world's largest online marketplace for P2P lending.

The Lending Club, founded in 2006, is headquartered in San Francisco, California. Since its origin as a Facebook application, the mission of the company is to provide an online platform for connecting borrowers and investors. Overall, as declared in its most recent annual report (Lending Club 2016), the company wants to transform the banking system by improving credit affordability and providing a new form of investments. The platform serves as an information provider for investors and delivers all information about potential borrowers' creditworthiness to investors. Customers have to complete a simple application form on the website inserting all their personal and financial information to request a loan. Vice versa, individual or institutional investors can select loans in which to invest according to their risk profile. Moreover, Lending club, after having analyzed the potential customers' profile defines grades and corresponding interest rates to be applied to loans. All these tasks are supported by the potentialities of the digital-powered marketplace that enhance the market matching between borrowers and investors (demand and supply).

Differently, from the traditional banking system, the company provides services that improve the customers' experience with ease of use and accessibility, eliminating the need for physical infrastructure and manual processes. Lending Club matches all the advantages of a P2P lending platform described before. Customers and small businesses can lower the interest rates applied on their loan requests and, on the other hand, investors, attracted by higher rates of returns, can diversify their portfolio with a brand new kind of investments. Through its market-based lending platform, Lending Club has made available more assets for more investors, including retail investors, high-net worth individuals and family offices, banks and finance companies, insurance companies, hedge funds, foundations, pension plans and universities endowments.

Furthermore, thanks to automation, Lending Club optimize many processes such as borrowers' application process, data gathering, credit scoring, loan funding, investing and servicing, regulatory compliance and fraud detection.

Similarly to other P2P lending platforms, Lending Club revenues come from transaction fees, both from borrowers and lenders and in particular from loan settlement procedure.

Lending Club was the first P2P lender to register its offerings as securities with the Security Exchange Commission (SEC) and thanks to an IPO (December 2014) the company raised $900 million. The stock ended the first trading day up 56%, valuing the company $8.5 billion. However, the post-IPO price performances disappointed investors in the long period. This fact is because investors might doubt that P2P lenders will be able to maintain this pace of growth in loan origination and supported their strong revenues increase. Nowadays, despite these investors' concerns, the relatively low market share of P2P lending and taking into account the recent consolidation of this sector, we can argue that there is room for expansion yet.

Chapter 2 – Literature review and theoretical background

2.1 Related literature and background of the quest

Although the phenomenon of P2P is relatively recent, in the last year, several studies have investigated into this form of financing using different models and approaches to figure it out. For instance, Faia & Paiella inquired "loans' spreads, proxying asymmetric information, decline with credit scores or hard information indicators and with indications from "group ties" (soft information from social networks). Also, an increase in the risk of a bank run in the traditional banking sector increases participation in the P2P markets and reduces their rates (substitution effect)." (Faia e Paiella 2008)

The outcomes of their studies depicted "several important implications. First, they show that transparency in debt markets helps to improve its liquidity and efficiency … Second, the importance of fostering the emergence and growth of markets offering alternative funding and investment opportunities concerning the traditional banking sector." (Faia e Paiella 2008).

Another research work from Tang found that "that P2P lending is a substitute to bank lending in that it serves infra-marginal bank borrowers, but also complements bank lending for small-size loans. These findings suggest that the credit expansion P2P lending brings about is likely to occur only among borrowers with access to bank credit." (Tang 2018)

The conclusion of the scholar suggests "…that the credit expansion opportunities brought by P2P lenders are likely to occur only for infra-marginal bank borrowers" and that "P2P platforms may not operate as substitutes to banks in the long run. That said, it is noteworthy that notwithstanding the rapid growth of the sector, Lending Club was the dominant player in the P2P unsecured consumer…" (Tang 2018). The abovementioned conclusions partially explain another reason why we have chosen Lending Club platform for our quest.

Even though not related, and using different models, the studies study of (Franks, Serran-Velarde and Sussman 2016) has inquired the and "the tradeoff between the information aggregation of auctions relationship and their susceptibility to liquidity shortages" (Franks, Serran-Velarde e Sussman 2016). The conclusion of the models' analysis has driven to the conclusion that the innovation introduced by P2P lending

platforms “may be… ultimately, just a reconfiguration of functions that were performed by financial markets for hundreds of years: generating information, aggregating information and providing market liquidity” (Franks, Serran-Velarde e Sussman 2016). More or less in the same period, other analysis, directed to verify the interaction between banks and this new financial means, identify “that P2P lending grows more at the expense of bank lending in these circumstances when borrowers display a greater awareness of P2P platforms”. De Roure et al., find that “… the advent of P2P lending may cause the banking sector to shrink, but be less risky and possibly more profitable in terms of risk-adjusted return on assets” (de Roure, Pelizzon e Thakor 2018).
On the other hand, our analysis differs from previous studies in that it focuses attention on the econometric characteristics of the States, within the USA, of belonging the requesting of loans. Moreover, our analysis only partially considers the influences of the institutional lenders contrarily to the earlier mentioned researchers.

2.2 Theoretical background and models

The focus of this section is the theoretical framework of the empirical analysis. The final goal of this thesis is to assess how and why personal and state characteristics affect loan granting. Thus, the possible outcomes of a loan request are two: granted or rejected. Considering the dichotomous nature of the dependent variable, we decide to use the logit regression model to verify the degrees of elasticity of personal and country variables on loan granting. Diving into the theoretic framework of the analysis, we are going to introduce a basic definition of categorical, binary variables and binomial distribution of a random variable. Then we proceed describing the logit transformation that is the main feature the model is built. Finally, we will describe the construction of the logit regression model through which the analysis is performed.
In statistics, the logistic regression, or logit regression or logit model is a regression model where the dependent variable is categorical. In the case under analysis, the dependent variable belongs to a specific subset of a categorical variable called binary variable. We define these concepts in the followings paragraphs.

2.2.1 Definition 1: Categorical Variable

A **categorical variable**, in statistics, is a variable that can take a limited, and usually fixed, number of values. The construction has been done by assigning each or another unit of observation to a particular group or nominal category determined by a qualitative point of view. Thus, each observation belongs to one specific group within groups considered. Categorical variables are usually adopted in computer science and some branches of mathematics referring as “enumerations” or “enumerated types.” The common adoption is to assign a level to each of the possible values of the variable. The probability distribution associated with a random categorical variable is called a categorical distribution. (Wikipedia 2017)

2.2.2 Definition 2: Binary data

In statistics, binary data is a type of statistical data described by dichotomous variables or binary variables. The main feature of these variables is that they can take only two possible values. Moreover, binary data represents the outcomes Bernoulli trials, i.e., statistical experiments with only two possible outcomes. It is a particular case under the group of categorical data, which are used to represents experiments with a determined number of possible outcomes. Despite being coded numerically as 0 and 1, in a binary variable, the two values are considered to exist on a nominal scale, which means that they represent, in a qualitative way, different values that cannot be compared numerically. From this point of view, binary data share similarities to categorical data but distinct from other types of numeric data (e.g., count data). Often, binary data is used to represent one of two conceptually opposed values, like:

- the result of an experiment ("success" or "failure")
- the answer to a yes-no question ("yes" or "no")
- presence or absence of some characteristic ("is present" or "is not present")
- the truth/false of a proposition ("true" or "false", "correct" or "incorrect", “head” or “tail”)

However, it can also be used for data that is assumed to have only two possible values, even if they are not conceptually opposed or conceptually represent all possible values in the space.”

2.3 Definition 3: Logit Transformation

The logit function is defined as the inverse of the logistic function and is widely used in statistical analysis to deal with binary data. If the variable of the function is the probability *p*, the corresponding logit function provides the log-odds. The basic idea is that the log-odds are equal to the logarithm of the odds $p/(1-p)$.

Thus, we can write the relationship in the following form:

$$logit(p) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p) = -\log(\frac{1}{p} - 1) \qquad (2.1)$$

The logarithm's base used has no importance in our case as long as it is greater than 1. For the sake of simplicity, the natural logarithm with base e is commonly used. The inverse formula for logit allows to obtain the "logistic" function for any number α as follow:

$$logit^{-1}(\propto) = \frac{1}{1+e^{-\propto}} = \frac{e^{\propto}}{e^{\propto}-1} \qquad (2.2)$$

Starting from equation 2.1, we can note that if *p* is a probability, then $p/(1-p)$ is the relative odds. Hence, the logit of the probability *p* is exactly the logarithm of the odds. In the same way, we can see that the difference between the logits of two probabilities results in the logarithm of the odds ratio, which, for the convention, is called R. Thus, to provide the correct combination of the odds ratio, exploiting the logarithm properties, the procedure is the following:

$$\log(R) = \log(\left(\frac{p_1}{1-p_1}\right)/(\frac{p_2}{1-p_2})) = \log(\frac{p_1}{1-p_1}) - \log(\frac{p_2}{1-p_2}) \qquad (2.3)$$
$$= logit(p_1) - logit(p_2)$$

From a statistical point of view, this short introduction allows us to define the model by analyzing the dependent variable. In fact, the response variable taken into account

y_i is binary and assumes only two different values "granted" and "rejected. In the following chapters, we will define the dependent variable as follow:

$$y_i = \begin{cases} 1 & if \text{ "}granted.\text{"} \\ 0 & if \text{ "}rejected.\text{"} \end{cases} \qquad (2.4)$$

Thus, we can view the variable y_i as the realization of a random variable Y_i that can take the value one or zero with probabilities π_i and $(\pi_i - 1)$, respectively. Consequently, the distribution of Y_i is the *Bernoulli* distribution with parameter π_i. We can write it in compact form as follow:

$$\Pr\{Y_i = y_i\} = \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \qquad (2.5)$$

For $y_i = 0{,}1$.

As we can see, if $y_i = 1$ we obtain π_i. On the other hand, if $y_i = 0$, we have $(1 - \pi_i)$. It is also easy to compute the expected value and variance of Y_i :

$$E(Y_i) = \mu_i = \pi_i, \qquad and \qquad (2.6)$$

$$Var(Y_i) = \sigma_i^2 = \pi_i(1 - \pi_i) \qquad (2.7)$$

It is important to note that the mean and the variance of Y_i depend on the underlying probability π_i. Due to that, any factor that affect the probability will modify not just the mean but also the variance of observations under analysis. Hence, a linear model that allows the predictors to influence the mean but assumes that the variance is constant will not be a good choice in order to analyze binary data.

To look for an adequate model we have to proceed in this way. First, we suppose that the factor of interest can divide the units under study into *k* groups. This partition can allow us to create *i* groups. All individuals that belong to a group have identical values of all covariates under analysis. Second, we denote with n_i the number of observation of the group *i* and y_i the number of individuals who have the characteristic of interest in group *i*.

In our case, let

$$y_i = \text{number of granted loan in group I} \qquad (2.8)$$

Hence, we can view y_i as the realization of the variable Y_i that takes the value $0,1,\dots,n_i$. If the n_i observations within each group are independent and have the same probability (π_i) of having the specific characteristic of interest, then the distribution of Y_i is binomial. We can write the distribution of Y_i with parameters π_i and n_i in the following way:

$$Y_i \sim B(n_i, \pi_i). \qquad (2.9)$$

In addition, The PDF (probability distribution function) of Y_i is:

$$\Pr\{Y_i = y_i\} = \binom{n_i}{y_i} \pi_i^{y_i}(1 - \pi_i)^{n_i - y_i} \qquad (2.10)$$

For $y_i = 0,1,\dots,n_i$.

In this case, $\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}$ is the probability of obtaining y_i "granted" and $n_i - y_i$ "rejected" in some specific order. It is important to point out that the combinatorial coefficient is the number of ways of obtaining y_i "granted" in n_i trials.

Following the same procedure saw before, we can compute the mean and the variance of Y_i. Hence,

$$E(Y_i) = \mu_i = n_i\pi_i, and \qquad (2.11)$$

$$Var(Y_i) = \sigma_i^2 = n_i\pi_i(1-\pi_i) \qquad (2.12)$$

Also, in this case, the mean and variance of Y_i depend on the underlying probability π_i. Anyway, this form is the most general one from a mathematical point of view. Thus, for our estimation, we can consider the binomial distribution as the right one.

The next step is to ensure that the probabilities π_i depend on a vector of observed covariates x_i, to build the structure of our model. The easiest way to reach our goal is to consider π_i a linear function of the covariates:

$$\pi_i = x_i'\boldsymbol{\beta} \tag{2.13}$$

Where $\boldsymbol{\beta}$ is a vector of regression coefficients.

The model defined above is usually called linear probability model. The first problem we have to manage is that the probability π_i on the left hand side is ranged, by definition, between zero and one. On the right hand side, the linear predictor $x_i'\boldsymbol{\beta}$ can take any real value. For this reason, unless complex restrictions imposed on estimated coefficient we will not be sure that predicted values will be in the correct range.

To deal with this issue, we decide to transform the probability. The procedure will allow us to remove the range restrictions in two different steps.

First, we transform the π_i in order to provide the *odd*, defined as the probability divided by its complement (e.g. the ratio of favorable to unfavorable cases). Hence,

$$odds_i = \frac{\pi_i}{1-\pi_i} \tag{2.14}$$

We note that if the probability of a specific event is one half, the odds are one-to-one. If the probability is 1/3, the odds are one-to-two. The main difference between probability and odds is that, as probability can take value only between 0 and 1, odds can take any positive value without restrictions. Also, we can easily translate probability into the odds and vice versa.

As the second step, we take logarithms and compute the logit or log-odds:

$$\eta_i = logit(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) \tag{2.15}$$

Thanks to that step we can remove the probability restrictions. In fact, we can note that as the probability decrease and fall to zero, the odds approach zero and the logit tends to $-\infty$. Conversely, as the probability goes to one the odds tend to $+\infty$, and so does the logit. To summarize this concept, we can state that the logit allows us to map probabilities from the range (0, 1) to the entire real line. So, probabilities below one-half are mapped to negative logit values. Vice versa, probabilities above one-half correspond to positive logit values. From a mathematical point of view, the logit transformation is a bijective function from the interval (0,1) to the real line. Exploiting the property of the function, we can take the inverse transformation (*antilogit*) to go back from logit values to probabilities. Solving for π_i in the equation 2.15 we have:

$$\pi_i = logit^{-1}(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \qquad (2.16)$$

The last step allows us to define to define the logistic regression model. In fact, we assume that the *logit* of the probability π_i, rather than the probability itself, follows a linear model.

The process shares some similarities seen before. First, we assume to have *k* independent observations$y_i, \dots, y_k$. Second, we impose that each *i*-th observation is the realization of a random variable Y_i, which has a binomial distribution. We write the binomial distribution of Y_i in this way:

$$Y_i \sim B(n_i, \pi_i) \qquad (2.17)$$

Where n_i is the binomial denominator and π_i is the probability. We set that individual data $n_i = 1$ for all *i* .

This step allows us to define the stochastic structure of the model. To define the structure of the model, we suppose that the *logit* of the underlying probability π_i is a linear function of the predictors:

$$logit(\pi_i) = x_i'\boldsymbol{\beta} \qquad (2.18)$$

Where x_i is a vector of covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients.

The model built is called generalized linear model with the binomial response and link logit. The regression coefficients $\boldsymbol{\beta}$ can be interpreted in the same way as in linear models, considering that the left-hand-side is a logit rather than a mean. As we will see, the β_j estimated coefficient represents the change in the logit of the probability associated with a unit change in the *j*-th predictor assuming all other predictors constant. We will discuss the interpretation of the outcome provided by the logit regression model further. (Rodríguez 2007)

Chapter 3 – Data Description and manipulations

Following the purpose of this thesis, in this chapter, we are going to describe the composition of the dataset under analysis. Therefore, to provide a clear idea of the variables used for the estimation, we divided the following section into four subparts. First, we focus on States' characteristics, giving an overview of them and explaining why we decide to perform such analysis. Second, we introduce briefly the structure of macroeconomic variables connected to states. Third, we focus on describing the main characteristics of data downloaded from Lending Club database, which refer to individuals' characteristics. Finally, we present our dataset and further manipulations on it.

3.1 States in the USA

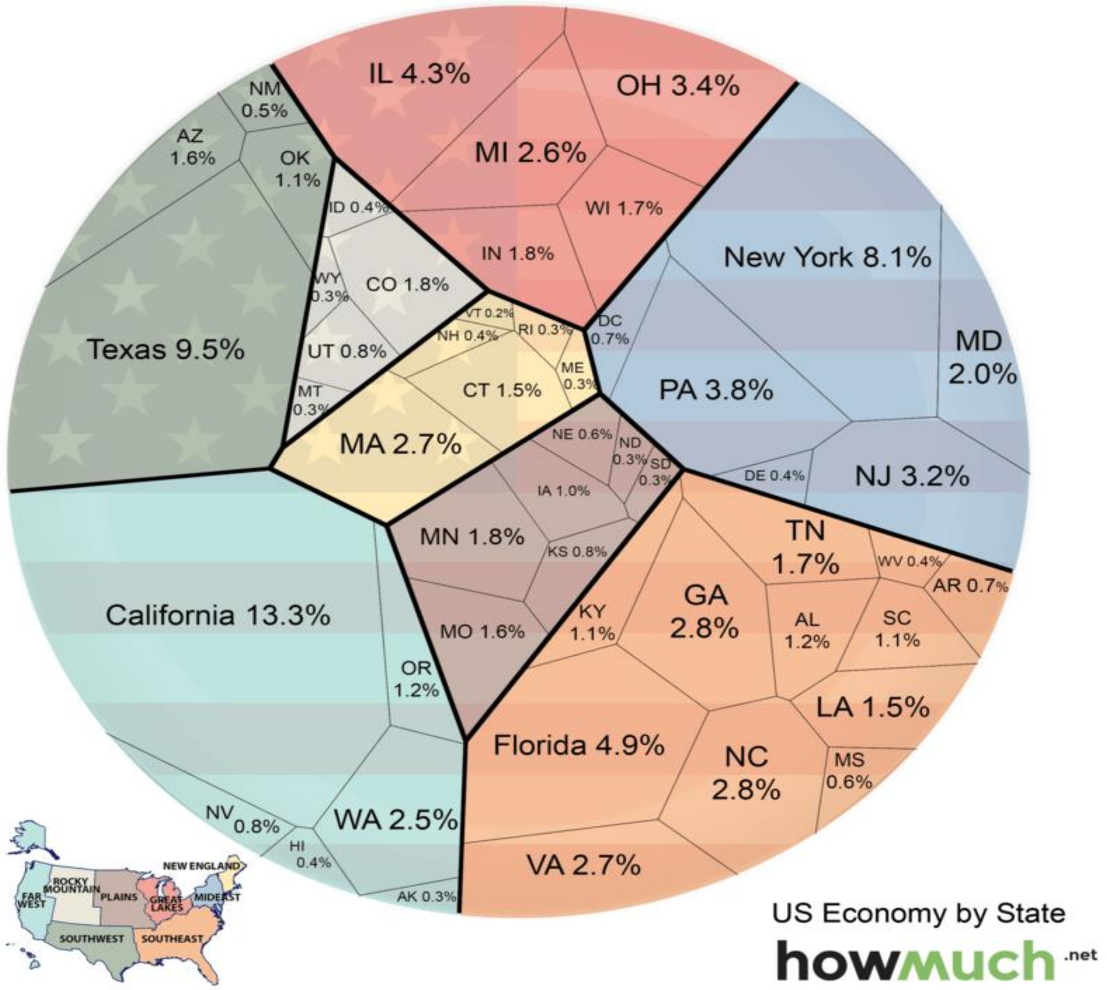
The United States moves more money throughout its economy than any other country. In fact, US generates a huge volume of international trade; however, since each States operate as separate entities, it is quite complicated to have a complete view of the overall US economy.

The US economy is the largest in the world, with a gross domestic product of $17.3 trillion that is about $7 trillion ahead of China, this makes the United States of American the wealthiest nation in the world by far. However, as it is well known, the value of the economic activity is unevenly distributed throughout the country. States differ each other regarding culture, social activity, economic well-being and many other aspects. Furthermore, their economy and asset allocation regarding the products (goods and services) are far to be similar.

Considering this fact, the States can also be grouped by specific regions to analyse regional economic features. This clusterization, defined by the Bureau of Economic Analysis, could allow us to analyze States' economic characteristics in an easier way considering similarities within every single group. Furthermore, the BEA point many similarities within regional indicators like per capita income, employment, and industry earnings that allow us to consider this clusterization relevant to this study. (U.S. Department of Commerce 2017).

The following chart provides an overview of the economic value generated by each US state (How much cost information website 2015)

Figure 1: US economy by State



Looking at the figure 1 above, we can note that color groups associate each State according to its pertinence region. At a glance, the entire US territory is divided into eight macro-regions: Far West, Rocky Mountain, Plains, Southwest, Southeast, Great Lakes, Mideast and New England. States within a region share many peculiarities regarding economic and cultural structure. Moreover, the figure points out that most economic activity is concentrated in three regions: Far West (18.6%), Southeast (21.3%), and Mideast (18.2%). All of these regions contain major US States and cover the US coastline, which is where most large cities are located. Thus, it is not surprising that most of the economic value in the US is generated in these regions. On a State point of view, California (13.3%), Texas (9.5%), and New York (8.1%) have by far the largest economies. Menawhile, Vermont, Maine, Rhode Island, North and South Dakota, Montana, Wyoming, and Alaska are the smallest economies (all representing about 0,2-0,3% of the total US economy). In recent years, considering the ending of the global financial crisis, all States have increased their economic outputs, but some have grown faster than others. This circumstance happens because the sources of economic outcomes are tendentially diversified by region and have lead to a proportional percentage adjustements. For instance, thanks to mining and manufacturing sectors, Texas increased the size of its economy by almost $300 billion, more than any other state, growing from 8.8% of the US economy in 2011 to 9.5% in 2014. In the same period, California grew by just under $300 billion, but only increased its share of the total economy by 0.1%; because, at the initial time point (2011), California's share of the US economy was already large compared to other States'.

3.2 Macroeconomic Variables

Following our purpose and deepening insight into States' characteristics, we introduce five macroeconomic variables of interest in our analysis. All these variables are assumed significant in the estimation of our models and could effectively influence the decision of granting a loan. At first glance, it is not clear whether their effect towards the dependent variable is positive or negative. Thus, in this section, we just explain what could be our predictions from a qualitative point of view. It is important to underlying that, to enhance the precision of the estimates, every variable is built by downloading 51 time series, one for each state, and by merging all of them to create a pooled cross sectional dataset, which included all applications, submitted to the lending platform. However, data manipulation will be discussed later.

3.3 The Unemployment Rate

The first macroeconomic variable taken into considerations is the unemployment rate. By definition, it is the number of unemployed people computed as a percentage of the labor force. It is important to underline that unemployed people are those who are available for work and have been looking for a job in the last month. This variable could have a major role in our analysis because usually is strictly connected to economic growth, poverty, education, crime rate and many others States' economic and social aspects. Considering these aspects, usually, an increase in the unemployment rate discourage traditional banks to fund loan request.

3.4 Gross Domestic Product

The second macroeconomic variable considered is the gross domestic product; this variable represents a monetary measure of the market value of all industry divided by goods and service production in a certain period. For completeness, our dataset also includes its main components, i.e., goods and services. To have a more accurate figure of economic growth we consider the real gross domestic product for each State. This variable is adjusted for inflation and provides the value of all goods and services produced by an economy in a given period. The real gross domestic product is said to be expressed in base-year prices (year's average prices). Unlike the unemployment rate, increasing gross domestic product provides a good indicator of a healthy economy and its citizens' economic well-being.

3.5 Personal Income

The third macroeconomic variable included in our analysis is the personal income. This denomination, personal income, is defined as the total compensation received by an individual. It incorporates compensations from many sources such as salaries, wages, and bonuses received for employment of self-employment as well as dividends distribution received from investments or rental revenues from real estate investments. In our case, it refers to all of the income collectively received by all of the individuals or households in a specific State; for completeness, we also include in our dataset its two components: farm income and non-farm income. Seems to be clear which relationship may occur between personal income and loan granting. Indeed, we expect that the higher the personal income, the higher the likelihood of a loan to be granted.

3.6 Total non-performing loans

The fourth variable included in our dataset is the total amount of non-performing loans registered state by state during the period under analysis. The commonly used definition of non-performing loans is the amount of borrowed money upon which the debtor has not made any scheduled payments for at least 90 days. In recent years, this economic value has been assuming a growing interest in the financial sector since institutions, holding NPL in their portfolios, have begun to sell them to other investors to remove risky assets and clean their balance sheets. This mechanism, nowadays, is becoming standard practice for many banks and financial institutions and, at the same time, a new investment opportunity for investors. Thus, considering the total non-performing loans amount for each State, we can investigate the relationship between loan granting and a likely negative-correlated variable.

3.7 Number of banks and financial institutions

The fifth variable considered is the total number of banks and financial institutions in each State. Although it seems to be a time-invariant variable, considering the time span of the dataset and the worldwide impact of the global financial crisis we deem appropriate its inclusion for the analysis. However, unlike other variables, we do not take a position on how this variable could affect the loan granting.

3.8 Personal Consumption Expenditure (PCE)

The last variable considered is the personal consumption expenditure or PCE Index. It measures price variations of consumer goods and services. Through this index are measured data that refers to both durables and non-durables goods, as well as services. Unlike the Consumption Price Index (CPI), the PCE reveals changes in expenditures that fall within a pre-established fixed bucket. Though this indicator is seldom used in the most recent analysis, it includes a great variety of expenses of the households across the US so that it captures short-term changes in consumer behavior better than CPI; thus, resulting in a more comprehensive inflation metric. For these reasons, we include PCE index in our analysis to investigate whether exist a kind of relationship with loan granting.

3.9 Lending Club Data

The focus of this section is on Lending Club data, yielding a complete description of the dataset used for the analysis and providing some details on variables treated.

Lending Club, as a P2P Lending platform, operates in the United States and provides its services to a broad range of customers all over the country.

One of the reasons why we choose Lending Club as a reference for our analysis is that its website provides a great amount of information including those describing the borrowers' financial and personal characteristics. All these information are provided and made available to potential lenders/investors to ease investment decisions. Also, the website provides information about all granted and rejected loan applications from the very beginning of Lending Club activities.

Bearing in mind these aspects and the fact that Lending Club is, today, the world's largest P2P lending platform, we decide to consider it representative of the American P2P lending marketplace. Before describing the construction of the dataset under analysis, we introduce those variables downloaded from Lending Club website. All variables are collected from potential borrowers through their online applications.

3.9.1 Loan Amount

This variable describes the amount of money requested by the applicant. The value of the variable ranging from 2.000 $ to 40.000 $. It is important to underline that the platform also requires to including the reason why the money is requested.

3.9.2 Employment Length

Another information that has to be introduced in the online application is the employment span. By definition, it is the number of years a person has been employed in the current service. For the sake of simplicity, we set the correspondent value number to the number of employed years (e.g., 1= "1 year",…,10="10 years" and 11="10+ years").

3.9.3 States

The State is the most meaningful variable to perform our analysis. Every customer application has been connected with the State in which the potential borrower lives. Through this parameter, we will be able to assess whether exist a significant relationship between individual's State and his/her loan granting probability.

3.9.4 Debt to income ratio

The debt to income ratio is one of the most important personal financial indicators and is broadly used by financial intermediaries in the decision-making process in the mortgage market. It is computed as the ratio between individual's debt total payments to his/her overall income; this is one of the most important values used by lenders to assess the potential borrower's capability to manage monthly payment and repay debts. The debt to income ratio is expressed in percentage by the formula:

$$DTI = \frac{\text{Total recurring monthly debt}}{Gross\ monthly\ income} \tag{3.1}$$

A low debt-to-income ratio depicts a good balance between debt and income. Conversely, a high debt to income ratio can signal that an individual has too much debt concerning his/her income. This value is strictly correlated with the lending market and, overall, the lower the DTI, the higher the probability that an individual will be able to get a loan.

3.9.5 The dataset creation and further manipulations

To perform the analysis, we had to build a specific dataset for our purpose. In this section, we will present dataset's characteristics and its construction briefly.

First, we downloaded all available data about loan applications from the Lending Club website. Second, we downloaded macroeconomic and other variables of interest from FRED (Federal Reserve Bank of St. Louis 2017) and US BEA (U.S. Department of Commerce 2017). Third, after some arrangements, we built two separated datasets before merging them into the final one.

The first dataset, constituted by Lending Club variables, was built by merging all observations from granted applications and rejected ones. All data were downloaded from Lending Club website where they are divided by years into eleven files for granted applications and ten files for declined applications. These files are updated every four months, on the same day as the quarterly results of the company are released; this means that it is possible to access to the latest payment evidence. Information about almost all issued loans is available through the dataset we collected, except for those few loans for which Lending Club was not authorized to release the transactions' details publicly. Our objective is to create a unique dataset composed of all variables of interest described before (e.g., loan amount, debt-to-income). Therefore, we started by creating one dataset for granted loans and one for rejected ones. Then we kept only the variables of interest excluding the others and, finally, we merged the two datasets into one.

It is important to underline that, before this step, we also introduced a new variable that allows the detection of differences between grantin and rejected requests. Hence, before merging the two datasets, we defined this new variable called "granted"; the variable is a dummy that assigns value 1 to granted applications and value 0 to rejected ones. The value "granted" will be the dependent variables in our regression models and the main driver of our analysis.

Thanks to this new feature in our dataset, we could give a first flavor on which we are talking about in the following table:

Table 3.1: The granted variable

Granted		
0	**1**	**Total**
14,124,429	1,524,077	15,648,506
90.26%	**9.74%**	**100.00%**

Table 3.1 points out the dimension of the "Lending Club" dataset regarding total applications submitted and loan applications granted/declined. The most relevant indication found is that only the 9,74% of applications had been accepted while the 90,26% had been declined during the period considered.

Also, this variable allows providing another result in terms loan of granting on the P2P platform over the years as highlighted in the following:

Table 3.2: Applications accepted vs. declined (increments over years)

Variation		2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	Mean	Variance
Granted	0	1.9%	0.1%	-1.7%	1.1%	-5.0%	-1.6%	4.9%	-2.2%	5.2%	**0.3%**	**0.1%**
	1	-16.7%	-0.8%	18.2%	-9.5%	50.6%	10.2%	-27.8%	18.2%	-35.0%	**0.8%**	**6.3%**

Table 3.2 represents all variations divided by years together with their mean and variance. Results provide us another clue about the behavior of the variable. In fact, the overall mean variation of granted applications is higher than the overall mean variation of rejected ones; this is probably because of the increasing quality of consumers regarding creditworthiness and reliability over year.

Moreover, we can notice that the variance of loan granting is significantly higher than the variance of rejected loan applications; this is probably because of the cyclical behavior of the former and its sensitivity towards macroeconomic factors and regulatory framework.

Moving on, the second dataset, the one that includes all macroeconomic variable by State, was completely build by us taking into account all macroeconomic variables described before.

First, we downloaded all country variables for each Member State in the US. Data for the unemployment rate, total non-performing loans and number of banks by the state were downloaded from FRED (Federal Reserve Bank of St. Louis 2017) website. While data for Gross Domestic Product (together with its components), Personal Income (divided into the farm and non-farm income), Personal Consumption Expenditure (PCE Index) by State were downloaded from BEA (U.S. Department of Commerce 2017). Whereas the two different downloading sources and the great number of data files downloaded (51 data files for each variable – one for each State), we have had to deal with different issues. First, data frequencies (quarterly, monthly and annually), secondly, data dimensions (some data are expresses in US dollar, other in US thousands of dollar or percentage).

To cope with scale issue, we decide to take the logarithmic difference for variables at a level such as Gross Domestic Product, Total Non-Performing Loans, Personal Income and Personal Consumption Expenditure. This transformation would allow us to deal with estimated percentage increments for these variables mitigating the scale effect.

Regarding the frequency issue, we had to deal with three different values:

- Datasets for Gross Domestic Product, Personal Income, Number of Banks and Total Non-Performing Loans are all provided quarterly from 2007 Q1 to 2017 Q3.

- The dataset for the Unemployment rate is provided monthly from January 2007 to March 2017.
- The dataset for Personal Consumption Expenditure (PCE) is provided annually from 2007 to 2017.

Since the final goal is to create a unique dataset by merging the “Lending Club” dataset and the “Macroeconomics” dataset we had to take some assumptions.
Thus, we assumed that all quarterly values are equal for every month within the quarter and, in the same way, we assumed that all annual values are equal for every month within the year. Hence, we obtained the “Macroeconomics” dataset by merging all variables into one panel. A panel data, or longitudinal data, consists of a time series for each cross-section member in a dataset. The key feature of a panel, which distinguishes it from a pooled cross section, is that the same cross-sectional units are followed over a given period. In our case, we obtained a panel with month-by-month macroeconomic values for each State in the period under analysis.
Finally, thanks to these middle steps we had two different datasets for individuals’ characteristics and States’ macroeconomic characteristics. As a final step, we built another dataset by connecting every single applicant for a loan with his/her States’ macroeconomic values in the month in which he/her submitted the request.
This result in a pooled cross-section dataset that contains all information we need for our analysis. A pooled cross-section dataset is obtained by collecting random samples from a large population independently of each other at different points in time. The fact that the random samples are collected independently of each other implies that they need not be of equal size and will usually contain different statistical units at different points in time.
The final dataset, obtained by merging the two datasets created using the “Lending Club” dataset and the “Macroeconomics” one, has 15.648.506 observations. As already said, every single observation describes both individual’s characteristics and macroeconomic values of the state in which he/she lives. The key factor for the merge of these two datasets is the month when he/she has submitted the request for the loan.
To provide a clear overview of the data under analysis, we summarize it in a two-way table considering the variable “granted” and its frequency over each State.

Table 3.3: Granted vs. State (original dataset)

State	Granted				Total
	0		1		
AK	33,162	89.96%	3,700	10.04%	36,862
AL	254,511	93.10%	18,866	6.90%	273,377
AR	150,370	92.92%	11,461	7.08%	161,831
AZ	293,717	89.18%	35,639	10.82%	329,356
CA	1,725,282	88.94%	214,531	11.06%	1,939,813
CO	242,031	88.32%	32,000	11.68%	274,031
CT	173,028	87.93%	23,756	12.07%	196,784
DC	28,384	88.08%	3,840	11.92%	32,224
DE	46,896	91.56%	4,324	8.44%	51,220
FL	1,088,640	91.04%	107,092	8.96%	1,195,732
GA	533,754	91.46%	49,861	8.54%	583,615
HI	82,921	91.67%	7,531	8.33%	90,452
IA	146	91.25%	14	8.75%	160
ID	25,003	93.27%	1,805	6.73%	26,808
IL	517,016	89.39%	61,379	10.61%	578,395
IN	252,248	90.98%	25,009	9.02%	277,257
KS	117,704	89.96%	13,141	10.04%	130,845
KY	181,693	92.51%	14,719	7.49%	196,412
LA	218,581	92.45%	17,846	7.55%	236,427
MA	286,685	89.09%	35,100	10.91%	321,785
MD	300,664	89.29%	36,058	10.71%	336,722
ME	30,363	92.26%	2,549	7.74%	32,912
MI	383,144	90.58%	39,864	9.42%	423,008
MN	184,412	87.10%	27,316	12.90%	211,728
MO	250,279	91.13%	24,360	8.87%	274,639
MS	124,992	94.04%	7,920	5.96%	132,912
MT	36,739	89.53%	4,298	10.47%	41,037
NC	442,664	91.22%	42,585	8.78%	485,249
ND	18,897	90.52%	1,978	9.48%	20,875
NE	45,598	91.42%	4,277	8.58%	49,875
NH	60,280	89.01%	7,442	10.99%	67,722
NJ	442,643	88.67%	56,557	11.33%	499,200
NM	83,866	91.09%	8,206	8.91%	92,072
NV	171,093	88.85%	21,479	11.15%	192,572
NY	1,066,446	89.41%	126,375	10.59%	1,192,821
OH	512,914	90.90%	51,367	9.10%	564,281
OK	157,090	91.83%	13,967	8.17%	171,057
OR	142,487	88.78%	18,013	11.22%	160,500
PA	540,598	91.08%	52,925	8.92%	593,523
RI	58,021	89.63%	6,715	10.37%	64,736
SC	231,537	92.57%	18,591	7.43%	250,128
SD	29,433	90.51%	3,085	9.49%	32,518
TN	283,868	92.37%	23,461	7.63%	307,329
TX	1,251,175	90.90%	125,308	9.10%	1,376,483
UT	90,902	89.77%	10,359	10.23%	101,261
VA	390,802	89.96%	43,602	10.04%	434,404
VT	27,243	89.46%	3,211	10.54%	30,454
WA	254,729	88.83%	32,032	11.17%	286,761
WI	186,471	90.29%	20,061	9.71%	206,532

WV	46,568	90.03%	5,156	9.97%	51,724
WY	26,739	88.88%	3,346	11.12%	30,085
Total	**14,124,429**	**90.26%**	**1,524,077**	**9.74%**	**15,648,506**

Although the dataset created provides whole information about all loan applications; we considered to reduce it to have an easier-to-manage dataset. Hence, we decided to perform some manipulations on it. First, we decided to take into consideration only the last ten years of data ranging from May 2007 to March 2017 (last data available). Then, to make the dataset consistent among all variables object of analysis, we rearranged the dataset by removing residual and uncompleted observations (e.g., the ones that have no values for "state," "date" or "granted"). Finally, we created a subsample from the original dataset that will allow us to conduct our analysis more easily. The procedure has taken into considerations the final objective of this thesis and has allowed us to obtain the same framework of the original dataset. Thus, using the function -sample- in "STATA," we could create a 20%-sample that has maintained the same characteristics and proportions regarding "state" and "granted" variables.

Therefore, the resultant dataset, the one we will work with, consists of 3,129,700 observations. The overview of the subsample is presented in the following two-way table with the same structure of Table 3.3.

Table 3.4: Granted vs. State (subset overview)

State	Granted				Total
	0		1		
AK	6,632	89.96%	740	10.04%	7,372
AL	50,902	93.10%	3,773	6.90%	54,675
AR	30,074	92.92%	2,292	7.08%	32,366
AZ	58,743	89.18%	7,128	10.82%	65,871
CA	345,056	88.94%	42,906	11.06%	387,962
CO	48,406	88.32%	6,400	11.68%	54,806
CT	34,606	87.93%	4,751	12.07%	39,357
DC	5,677	88.08%	768	11.92%	6,445
DE	9,379	91.56%	865	8.44%	10,244
FL	217,728	91.04%	21,418	8.96%	239,146
GA	106,751	91.46%	9,972	8.54%	116,723
HI	16,584	91.67%	1,506	8.33%	18,090
IA	29	90.63%	3	9.38%	32
ID	5,001	93.27%	361	6.73%	5,362
IL	103,403	89.39%	12,276	10.61%	115,679
IN	50,450	90.98%	5,002	9.02%	55,452
KS	23,541	89.96%	2,628	10.04%	26,169
KY	36,339	92.51%	2,944	7.49%	39,283

LA	43,716	92.45%	3,569	7.55%	47,285
MA	57,337	89.09%	7,020	10.91%	64,357
MD	60,133	89.29%	7,212	10.71%	67,345
ME	6,073	92.25%	510	7.75%	6,583
MI	76,629	90.58%	7,973	9.42%	84,602
MN	36,882	87.10%	5,463	12.90%	42,345
MO	50,056	91.13%	4,872	8.87%	54,928
MS	24,998	94.04%	1,584	5.96%	26,582
MT	7,348	89.52%	860	10.48%	8,208
NC	88,533	91.22%	8,517	8.78%	97,050
ND	3,779	90.51%	396	9.49%	4,175
NE	9,120	91.43%	855	8.57%	9,975
NH	12,056	89.01%	1,488	10.99%	13,544
NJ	88,529	88.67%	11,311	11.33%	99,840
NM	16,773	91.09%	1,641	8.91%	18,414
NV	34,219	88.85%	4,296	11.15%	38,515
NY	213,289	89.41%	25,275	10.59%	238,564
OH	102,583	90.90%	10,273	9.10%	112,856
OK	31,418	91.84%	2,793	8.16%	34,211
OR	28,497	88.78%	3,603	11.22%	32,100
PA	108,120	91.08%	10,585	8.92%	118,705
RI	11,604	89.63%	1,343	10.37%	12,947
SC	46,307	92.57%	3,718	7.43%	50,025
SD	5,887	90.51%	617	9.49%	6,504
TN	56,774	92.37%	4,692	7.63%	61,466
TX	250,235	90.90%	25,062	9.10%	275,297
UT	18,180	89.77%	2,072	10.23%	20,252
VA	78,160	89.96%	8,720	10.04%	86,880
VT	5,449	89.46%	642	10.54%	6,091
WA	50,946	88.83%	6,406	11.17%	57,352
WI	37,294	90.29%	4,012	9.71%	41,306
WV	9,314	90.03%	1,031	9.97%	10,345
WY	5,348	88.88%	669	11.12%	6,017
Total	**2,824,887**	**90.26%**	**304,813**	**9.74%**	**3,129,700**

Table 3.4 provides, as expected, a clear indication about the composition of the subset created. In fact, observations within the new dataset maintain the same proportion of the original one regarding State and percentage of granted loan applications.

Chapter 4 – Loan grantsmanship over US countries

In this chapter, we will describe the main findings of our empirical analysis and outcomes we found out about the relationship between the likelihood for a P2P loan to be granted and individual's attributes (both personal and related to the State in which he/she lived).

This chapter is divided into seven sections regarding the objective of our analysis and logit regression models estimated as well as robustness check

In the first section, we summarize the objectives of this thesis, explaining the econometric technique that we have applied to the dataset. Then, we present results separately providing our interpretation.

4.1 The scope of the analysis

The primary objective of our analysis is to perceive the main determinants behind the grant or refuse of loans on the P2P lending marketplace. To realize that, we assumed that all applications submitted to Lending Club, the world's largest P2P lending platform, are representative of the whole American P2P lending marketplace. Our main purpose is to provide an additional point of view to one of the biggest economic and financial issue of our time: **the decision process**.

Overall, every day people make choices among a great number of alternatives based on a limited amount of information. The decision processes are part of our social and working lives. The choice to go to work by car rather than using public transportation on a rainy day or the compatibility of a partner is just two examples. The decision process could affect, in the same way, the potential lenders. Either due to their experience, stereotypes, and perceptions, or simply the nature of their preferences, in addition to the limited hard, verifiable information, they might base decisions on easily observable variables such as the personal characteristics of the counterpart (the potential borrower). Similarly, when assessing the creditworthiness of a potential borrower, in addition to social and financial information such as credit report, employment history, and the overall financial situation, the lenders' decision might also be influenced by personal characteristics like race, beauty, and the way borrower presents himself (Ravina 2012).

In this context, we want to put the focus on States' heterogeneity (see Chapter 3) among the United States to determine whether exists a positive or negative relationship between individuals who live in a specific State rather than in another regarding their likelihood to get a loan. In particular, we want to determine whether the lenders' perception and stereotypes might affect the decision process by based the State to which an individual belongs.

Also, we perform the same analysis on individual's likelihood to get a loan by considering all macroeconomic variables relative to his/her State. In this way, we will be able to understand whether lenders' decisions are driven by perception/stereotypes or objective macroeconomic criteria, once other personal and hard financial information is taken into account.

4.2 The econometric model

The econometric model used for estimating the sensitivity of the loan granting compared to individuals' characteristics and macroeconomic variables is the logit regression model described earlier in Chapter 2. We decided to use this model because the chosen dependent variable (granted) is a category variable built in this way:

$$\text{granted} = \begin{cases} 1 & if \text{" the loan is issued."} \\ 0 & if \text{ "the request is rejected.} \end{cases}$$

The logistic regression uses maximum likelihood estimation to provide its outcomes. By, the definition the maximum likelihood estimation (MLE) is an iterative method of estimating the parameters of a statistical model given observations, by finding the parameter values that maximize the likelihood of making the observations given the parameters. Hence, during our quantitative analysis on STATA, at each iteration, the log likelihood increases because the goal is to maximize it. When the difference between successive iterations is very small, the model is said to have converged, the iterating is stopped, and the results are displayed. (Wikipedia 2017)

To better understanding results represented in the thesis, we provide a brief description the statistics indicator of the model:

- Log likelihood: The log likelihood of the model is maximized during iterations. The maximum log likelihood found is displayed in our results. In general, its value is not significant unless for comparison between nested models.
- Number of obs.: The number of observations under analysis. Observations used in the estimation process might be smaller than the ones within the dataset. In fact, the presence of some missing values for any of the variables used in the regression could affect the number of observations included in the analysis. In fact, STATA, performing the logistic regression, remove by default any uncompleted observation concerning one of the variables taken into account.
- Wald chi2(ν): The outcome of the Wald chi-square test. Under the Wald statistical test, the maximum likelihood estimate $\hat{\theta}$ of the parameter of interest θ is compared with the proposed value θ_0. The underlying assumption is that the difference between these two estimators will be normally distributed. In our case, the square of the difference is compared to a chi-squared distribution. The number ν in the parenthesis indicates the number of degrees of freedom. This value depends on the number of predictors in the model.
- Prob > chi2: Assuming that the null hypothesis is true, in the outcome table is reported the probability of obtaining the chi-square statistic given. This is the p-value, which allows us to assess whether the estimated model is consistent or not.
- Pseudo R2: This is the pseudo R-squared. It is important to point out that, unlike the OLS regression, the logistic regression does not provide an equivalent value for R-squared. Thus, this value has to be interpreted as an estimated or approximate R-squared.
- Coef. : As already mentioned in the last section of chapter 2, the estimated values of the logistic regression are expressed in log-odds units. Hence, these estimated represents the increase in the predicted log odds of granted=1 that would be predicted by a 1 unit increase in the predictor, holding all other predictors constant.

Because these coefficients are in log-odds units, they are often difficult to interpret. However, since our final goal is to investigate whether there is a positive or negative relationship between the dependent variable and independent variables this aspect would not be considered.

- Robust Std. Err. These are the standard errors associated with the coefficients. The standard error allows us to test whether the parameter is significantly different from zero. We add the specification of robust standard errors to avoid any possible heteroscedasticity.
- z and P>|z|: These columns provide the z-value, computed by dividing the estimated coefficient by its the standard error, and the relative p-value. In our analysis, we decide to set the significance level at 1% and, therefore, p-values above this threshold will be considered not statistically significant.

4.3 Variables and Transformations

Our analysis was conducted with a pooled cross-section dataset that provides us much information regarding individuals' characteristics and the State to which they belong.

However, to better understanding the outcomes provided, by our models, we had to deal some transformations on variables of interest because the raw data were downloaded in differente scales and, for this reason, our estimates could have been affected regarding magnitude. All variables have been transformad so that we obtaind an estimation of their percentage increments over time to aviod this problem (see Chapter 3 for more details). Since the estimated coefficients should tell us the amount of increase in the predicted log odds of granted =1 expected by 1 unit increase in predictors (our variables), the data tranformation should allow us to provide a better interpretation on the estimations.

As earlier mentioned in the previous chapter, all macroeconomic variables are taken in logarithmic difference, except for the "number of banks" that is maintained at a level and the unemployment rate that is already provided in percentage. Moreover, we take the logarithmic transformation of the loan amount requested and debt-to-income ratio to consider their estimated percentage increments.

Finally, to provide another point of view to our analysis, we decde to add the estimated percentage increment for the number of applications per month forwarded to Lending Club. This variable could allow us to investigate how the P2P platform has run the outstanding growth of applications throughout years.

After these adjustements, we decided to divide our analysis into five different models. The main characteristics of our models are that all of them included at least four rapresentative variables for invidivudals' characteristics that we consider traditional drivers of the loan decision process. Then, we will include, from time to time, different and more detailed information about States and their macroeconomic structure.

4.4 The decision process driven by individual's characteristics and macroeconomic variables

The first model we present is performed by considering that the decision process could be affected by the macroeconomic structure of the State to which belong the potential borrower. In particular, we want to investigate how the likelihood of getting a loan could be affected by different features of the application. Thus, we include into our model one variables striclty linked to the loan application (loan amount), one that belongs to hard financial information of the individual (debt-to-income ratio), one that indicates the employment status (employment length) together with six macroeconomic and banking variables. Finally, we also include the estimated percentage increment of the number of applications submitted to the platform.

Therefore, the model we estimate has the following structure:

$$
\begin{aligned}
granted = & \propto_0 + \propto_1 Log(Loan\ Amount) + \propto_2 Log(Debt\ to\ income) \\
& + \propto_3 Emp.Length + \propto_4 D(Log(applications\ per\ month) \\
& + \propto_5 Unemp.Rate + \propto_6 D(Log(Real\ GDP) + \propto_7 D\big(Log(PI)\big) \\
& + \propto_8 D(Log(Tot.NPL) + \propto_9 Number\ of\ banks \\
& + \propto_{10} D\big(Log(PCE)\big) + u_t
\end{aligned}
$$

The results of the estimations, including the standardized coefficients and the option of robust errors to heteroscedasticity, are summarized in the following table:

Table 4.1: Estimation outcomes of the decision model. The effect of macroeconomic variables

Logistic regression						Number of obs=	2,230,157
						Wald chi2(10)=	378036.97
						Prob>chi2=	0.000
Log pseudolikelihood=	-578599.11					Pseudo R2=	0.2873
Granted	Coef.	Robust Std. Err.	Z	P>z	[95% Conf. Interval]		
Log_Loan_Amount	**0.310**	0.002	127.240	0.000	0.306	0.315	
Log_Debt_to_Income	**-0.035**	0.002	-21.930	0.000	-0.039	-0.032	
Emp_Length	**0.365**	0.001	598.100	0.000	0.364	0.366	
DLog_applications_per_month	**0.085**	0.013	6.340	0.000	0.059	0.111	
Unemp_Rate	**0.004**	0.000	29.370	0.000	0.003	0.004	
DLog_Real_GDP	**-3.106**	0.343	-9.040	0.000	-3.779	-2.432	
DLog_Personal_Income	**7.763**	0.318	24.380	0.000	7.139	8.388	
DLog_Total_NPL	**-0.415**	0.024	-17.580	0.000	-0.461	-0.369	
Number_of_Banks	**0.000**	0.000	3.780	0.000	0.000	0.000	
DLog_Pers_Cons_Exp	**2.668**	0.322	8.280	0.000	2.036	3.299	
_cons	**-6.563**	**0.029**	**-225.580**	**0.000**	**-6.620**	**-6.506**	

In table 4.1, we can note that all estimated coefficients are significantly different from zero. Their p-values are all equal zero and allow us to reject the null hypothesis that coefficients may be zero at the 1% significance level. Focusing on estimated coefficients, the first surprising finding is that the loan amount requested has a positive effect on the dependent variable. This finding could be interpreted as a peculiar aspect connected with the P2P lending market, where major investments rather than small-scale ones attract investors. This interpretation could be strengthened by the fact that, in recent years, an increasing number of institutional investors have entered into this business.

The estimated coefficient of the Debt-to-Income ratio provides the predicted result. This evidence allows us to confirm that an increase in debt-to-income ratio lowered the individual's likelihood to get the loan requested. Even the estimate connected with the employment length validate our previous prediction. In fact, the outcome of the model confirms that the higher the number of years a person has been employed in the current service, the better is his/her chance of a favorable ruling.

Moving on through the analysis we can observe that the likelihood to get a loan increasing as the number of total applications per month increase. The interpretation of this finding might be that as the number of applications rises, the quality of applicants rises too. Conseguently, this effect could indicate that the decision process within P2P lending platform is driven by a predetermined threshold that allows rejecting requests above a certain level of risk.

Turning to the interpretation of the estimated coefficients connected to the macroeconomic structure of the State from which the application is connected, we can find that the coefficient of the unemployment rate has a positive effect on the dependent variable chosen. To provide an interpretation of this value, we can follow traditional macroeconomics theory. Hence, given the inverse relationship between the unemployment rate and inflation rate (Phillip Curve), we can argue that, in the short run, an increase in the unemployment rate results in a decrease in the inflation rate and, consequently, in a decrease in the interest rate. This conjuncture could be considered an incentive for the traditional credit market since one can borrow money at a lower price. Conversely, on lenders side, a decrease in the interest rate discourage their willingness to lend money. In this context, the P2P lending market could offer a better investments opportunity and the possibility to invest in assets generating higher returns compared to traditional financial instruments.

Focusing on the estimated coefficient of the real gross domestic product, we find that an increment in the real GDP is connected with a decrease regarding loan granting. In this case, we could argue that, following the equilibrium in the money market (LM curve), an increase in the production leads to an increase in the money demand. This effect could lead to a new market equilibrium at a higher interest rate level. Consequently, lenders are more willing to grant a loan through traditional channels rather than on P2P lending platforms.

Proceeding with our analysis, we find that, as expected, the percentage increment in the mean of personal incomes in a State is positively correlated with the loan granting. The interpretation of this value might seem easy and predictable. Nevertheless, this result brings up an interesting topic about deep economic diversity between States.

Moving on the interpretation of the estimated coefficient of the total amount of NPL registered, we can note that it points out a positive effect on the dependent variable. This finding follows and confirms our prediction that the higher is the amount of NPL retained within a financial institution in a State; the lower is their willingness to grant credit. Moreover, according to our results, this relationship could be considered valid for both traditional credit channels and P2P lending channels.

Another variable taken into consideration is the number of banks per State. Its introduction, without any prediction, is because we decide to analyze whether exist an effective relationship between this variable and the loan granting. Our results allow us to state that there is a slight positive relation between the two. This occurence could be interpreted in different ways. For example, the decision for a bank group to establish a new subsidiary is usually subject to a deep market analysis. The final goal is to reach the highest number of potential customers in the selected area to improve the business. For this reason, a high number of banks in a specific area should be connected with a high number of potential customers for the overall financial market. Furthermore, in this context, the P2P lending market could find a breeding ground and reach a higher number of potential clients.

The last variable included in the model is the estimated percentage increase of the PCE index. As already said, the PCE allows us to reveal changes in expenditures that fall within a pre-established fixed bucket. The estimated coefficient, in this case, underlines a positive effect this indicator has towards the loan granting. To provide an interpretation, we can state that, in a specific State, growth in consumption expenditure could be a signal of a healthy economy and a good indicator for people capability to repay their debts.

4.5 The decision process through years

To enhance the precision of the model, we decide to add a time dummy variable to the regression. This operation could allow us to control for time-specific fixed effect, i.e., a shock which impact is restricted to a given period and is not controlled by other explanatory variables.

Thus, the model we estimate has the following structure:

$$
\begin{aligned}
granted =& \propto_0 + \propto_1 Log(Loan\ Amount) + \propto_2 Log(Debt\ to\ income) \\
& + \propto_3 Emp.Length + \propto_4 D(Log(applications\ per\ month) \\
& + \propto_5 Unemp.Rate + \propto_6 D(Log(Real\ GDP) + \propto_7 D\big(Log(PI)\big) \\
& + \propto_8 D(Log(Tot.NPL) + \propto_9 Number\ of\ banks \\
& + \propto_{10} D\big(Log(PCE)\big) + \propto_i Years + u_t
\end{aligned}
$$

$for\ i = 11, \ldots 19$

The results of the estimations, including the standardized coefficients and the option of robust errors to heteroscedasticity, are summarized as follows:

Table 4.2: Estimation outcomes of the decision model. The effect of macroeconomic variables through years

Logistic regression					*Number of obs=*	*2,230,157*
					Wald chi2(19)=	*366931*
					Prob>chi2=	*0.000*
Log pseudolikelihood=	*-572240.17*				*Pseudo R2=*	*0.2951*
Granted	**Coef.**	**Robust Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
Year						
2008	-0.333	0.168	-1.990	0.047	-0.662	-0.005
2009	0.027	0.163	0.170	0.866	-0.291	0.346
2010	**0.568**	0.158	3.600	0.000	0.259	0.876
2011	**0.861**	0.156	5.530	0.000	0.556	1.166
2012	**1.329**	0.155	8.570	0.000	1.025	1.632
2013	**1.367**	0.155	8.830	0.000	1.063	1.670
2014	**1.161**	0.154	7.520	0.000	0.858	1.463
2015	**1.314**	0.154	8.520	0.000	1.011	1.616
2016	**0.672**	0.154	4.360	0.000	0.370	0.974
Log_Loan_Amount	**0.286**	0.002	116.460	0.000	0.281	0.291
Log_Debt_to_Income	**-0.033**	0.002	-20.590	0.000	-0.037	-0.030
Emp_Length	**0.369**	0.001	584.170	0.000	0.368	0.371
DLog_applications_per_month	0.001	0.014	0.110	0.915	-0.025	0.028
Unemp_Rate	0.000	0.000	-0.280	0.781	0.000	0.000
DLog_Real_GDP	**-1.060**	0.350	-3.030	0.002	-1.745	-0.374
DLog_Personal_Income	-0.830	0.355	-2.340	0.019	-1.526	-0.135
DLog_Total_NPL	0.047	0.027	1.750	0.081	-0.006	0.099
Number_of_Banks	0.000	0.000	-1.280	0.202	0.000	0.000
DLog_Pers_Cons_Exp	**4.279**	0.412	10.390	0.000	3.471	5.086
_cons	**-6.910**	**0.157**	**-43.960**	**0.000**	**-7.218**	**-6.602**

In table 4.2, we can observe that, including the time specification, the estimated coefficients for the year 2008 and 2009 are not significant as well as some explanatory variables. In fact, their p-values are above the significance level (1%). However, the estimated coefficients for other variables maintain their effect on the dependent one. The outcome could be interpreted because the regression model, by introducing the time-specific fixed effect, filter and select only the most relevant predictors. Hence, as seen before, the estimated coefficient of the increase of the loan ticket requested in one application has a positive effect, while the effect of an increasing debt-to-income ratio is negative on the likelihood to get the loan. At the same time, the estimated coefficient on the employment length seems to confirm one of the previous models. Thus, we can state that, overall, the higher the number of years employed in the current services, the higher the probability that a loan would be issued. Moving on macroeconomic variables connected the State in which the applicant lives, the negative effect of an increase in the real gross domestic product has been confirmed as well as the positive effect connected to the behavior of the PCE index.

4.6 The decision process driven by individuals' characteristics by State

To provide another point of view to our analysis, in this section, we decide to remove the macroeconomic variables from the model replacing them with a dummy variable for States. This procedure allows us to investigate and analyze the effects on the decision process taking into consideration only individuals' features provided from loan applications. For this reason, the model specification is the following:

$$\begin{aligned} granted =& \propto_0 + \propto_1 Log(Loan\ Amount) + \propto_2 Log(Debt\ to\ income) \\ & + \propto_3 Emp.Length + \propto_4 D(Log(applications\ per\ month)) \\ & + \propto_i States + u_t \end{aligned}$$

$$for\ i = 5, \dots 54$$

Similarly to the other models, the one above includes three different variables that describe three relevant aspects included in a loan application:

- A specific feature of the single loan application: the estimated increase in the loan amount requested (the logarithmic transformation of the loan size an individual apply for)
- Information directly connected to the potential borrower's financial state: the estimated increase in the debt-to-income ratio that represents a measure of the total debt affordable.
- A social characteristic of the individual: the number of years employed in the current service. This feature is usually considered to provide information about the working (and thus income) stability of the individual.

Moreover, we decide to add the estimated increase in the total amount of applications submitted to Lending Club and to assess the effect that could have in every single State.

The results of the estimations are summarized in Table 4.3. Here again, we include the standardized coefficients and the option of robust errors to heteroscedasticity:

Table 4.3 Estimation outcomes of the decision model. The effect of individual's characteristics

Logistic regression					Number of obs=	2,817,305
					Wald chi2(54)=	456686.08
					Prob>chi2=	0.000
Log pseudolikelihood=	-691623.71				Pseudo R2=	0.2835
Granted	Coef.	Robust Std. Err.	z	P>z	[95% Conf. Interval]	
Log_Loan_Amount	**0.324**	0.002	146.590	0.000	0.319	0.328
Log_Debt_to_Income	**-0.033**	0.001	-22.650	0.000	-0.036	-0.030
Emp_Length	**0.368**	0.001	657.790	0.000	0.367	0.369
DLog_applications_per_month	**0.113**	0.013	8.500	0.000	0.087	0.139
States						
AL	**-0.302**	0.051	-5.920	0.000	-0.401	-0.202
AR	**-0.279**	0.053	-5.220	0.000	-0.384	-0.175
AZ	**0.208**	0.049	4.210	0.000	0.111	0.305
CA	**0.126**	0.047	2.650	0.008	0.033	0.219
CO	**0.251**	0.050	5.050	0.000	0.154	0.349
CT	**0.168**	0.051	3.310	0.001	0.068	0.268
DC	**0.550**	0.067	8.230	0.000	0.419	0.681
DE	-0.120	0.063	-1.920	0.055	-0.243	0.003
FL	-0.035	0.048	-0.740	0.460	-0.129	0.058
GA	-0.069	0.049	-1.430	0.154	-0.165	0.026
HI	**-0.291**	0.057	-5.090	0.000	-0.402	-0.179
IA	-0.684	1.021	-0.670	0.503	-2.686	1.318

ID	**-0.379**	0.078	-4.890	0.000	-0.531	-0.228
IL	0.092	0.048	1.900	0.057	-0.003	0.187
IN	-0.033	0.050	-0.660	0.512	-0.131	0.065
KS	0.030	0.053	0.570	0.567	-0.074	0.135
KY	**-0.214**	0.052	-4.120	0.000	-0.316	-0.112
LA	**-0.169**	0.051	-3.290	0.001	-0.269	-0.068
MA	0.113	0.050	2.280	0.023	0.016	0.210
MD	0.106	0.049	2.140	0.032	0.009	0.203
ME	**-0.313**	0.071	-4.390	0.000	-0.453	-0.173
MI	-0.021	0.049	-0.420	0.676	-0.117	0.076
MN	**0.317**	0.050	6.300	0.000	0.218	0.416
MO	-0.024	0.050	-0.470	0.636	-0.122	0.075
MS	**-0.424**	0.055	-7.680	0.000	-0.532	-0.316
MT	0.068	0.065	1.040	0.296	-0.059	0.195
NC	-0.008	0.049	-0.150	0.878	-0.103	0.088
ND	-0.069	0.078	-0.880	0.379	-0.223	0.085
NE	**-0.180**	0.062	-2.880	0.004	-0.302	-0.057
NH	0.029	0.058	0.490	0.621	-0.085	0.143
NJ	0.100	0.049	2.060	0.039	0.005	0.196
NM	-0.081	0.056	-1.430	0.151	-0.191	0.030
NV	**0.234**	0.051	4.580	0.000	0.134	0.334
NY	0.079	0.048	1.660	0.096	-0.014	0.173
OH	-0.038	0.049	-0.770	0.440	-0.133	0.058
OK	**-0.150**	0.052	-2.860	0.004	-0.253	-0.047
OR	**0.209**	0.052	4.030	0.000	0.108	0.311
PA	-0.092	0.049	-1.890	0.058	-0.187	0.003
RI	0.014	0.059	0.240	0.813	-0.101	0.129
SC	**-0.173**	0.051	-3.380	0.001	-0.273	-0.072
SD	-0.099	0.070	-1.430	0.154	-0.235	0.037
TN	**-0.170**	0.050	-3.370	0.001	-0.268	-0.071
TX	-0.005	0.048	-0.100	0.919	-0.098	0.089
UT	0.136	0.055	2.480	0.013	0.028	0.243
VA	0.056	0.049	1.150	0.249	-0.039	0.152
VT	-0.009	0.070	-0.130	0.898	-0.146	0.128
WA	**0.199**	0.050	4.010	0.000	0.102	0.297
WI	0.024	0.051	0.470	0.641	-0.076	0.124
WV	0.056	0.062	0.900	0.369	-0.066	0.177
WY	**0.183**	0.068	2.680	0.007	0.049	0.316
_cons	**-6.274**	**0.052**	**-120.910**	**0.000**	**-6.376**	**-6.173**

In the above table 4.3, we note that all estimated coefficients, associated with the four categories described before, are significantly different from zero. Their p-values allow rejecting the null hypothesis that these estimated coefficients may be zero, even at the 1% significance level. Although the absence of control for time-specific fixed effects, the effects induced by the estimated coefficients on loan granting seems to be in line with interpretations given on the estimates of the first model. Indeed, signs of the coefficients estimated in this case are the same for each explanatory variable and, therefore, leads to the same interpretation. It is important to notice that this should not consider for granted, since the model specification, in this case, incorporates more information, i.e., the State specification.

In fact, the most relevant aspect of the model under analysis can be found in the estimated coefficients for every single State. Precisely, the model allows us to capture different effects derived from the State to which the potential borrower belong. At a glance, the table above gives us many indications about States' effects on loan granting since twenty-two estimates are significant at the 1% level. To provide an overview of the estimated coefficients and their impact on the dependent variables, we represent them in the table below. States highlighted in green are associated with a positive effect, whereas States highlighted in the red depicts a negative effect. The table included the clusterization defined in Chapter 3 and the relative percentage contribution given by each State to the total US economy.

Table 4.4: The decision process driven by individual's characteristics. States' estimated coefficients

Far West		% of Tot Economy	Rocky Mountain		% of Tot Economy	Plains		% of Tot Economy	SouthWest		% of Tot Economy
California	CA	13,3%	Colorado	CO	1,8%	Minnesota	MN	1,8%	Texas	TX	9,5%
Washington	WA	2,5%	Utah	UT	0,8%	Missouri	MO	1,6%	Arizona	AZ	1,6%
Oregon	OR	1,2%	Idaho	ID	0,4%	Iowa	IA	1,0%	Oklahoma	OK	1,1%
Nevada	NV	0,8%	Wyoming	WY	0,3%	Kansas	KS	0,8%	New Mexico	NM	0,5%
Hawaii	HI	0,4%	Montana	MT	0,3%	Nebraska	NE	0,6%			
Alaska	AK	0,3%				North Dakota	ND	0,3%			
						South Dakota	SD	0,3%			

SouthEast		% of Tot Economy	Great Lakes		% of Tot Economy	MidEast		% of Tot Economy	New England		% of Tot Economy
Florida	FL	4,9%	Illinois	IL	4,3%	New York	NY	8,1%	Massachusetts	MA	2,7%
Georgia	GA	2,8%	Ohio	OH	3,4%	Pennsylvania	PA	3,8%	Connecticut	CT	1,5%
North Carolina	NC	2,8%	Michigan	MI	2,6%	New Jersey	NJ	3,2%	New Hempshire	NH	0,4%
Virginia	VA	2,7%	Indiana	IN	1,8%	Maryland	MD	2,0%	Maine	ME	0,3%
Tennessee	TN	1,7%	Wisconsin	WI	1,7%	District of Columbia	DC	0,7%	Rhode Island	RI	0,3%
Louisiana	LA	1,5%				Delaware	DE	0,4%	Vermont	VT	0,2%
Alabama	AL	1,2%									
Kentucky	KY	1,1%									
South Carolina	SC	1,1%									
Arkansas	AR	0,7%									
Mississippi	MS	0,6%									
West Virginia	WV	0,4%									

Interpretation of Table 4.4 could be that applications submitted from a State located in Far West region are boosting the likelihood to get a loan, while applications forwarded from many States in the Southeast are more likely to be rejected. The above statement is valid except for some isolated cases (e.g., Connecticut, District of Columbia, Arizona,..) that belong to other regions. The interpretation of those cases might be found in their economic characteristics (e.g., tax policies to incentive the market) as well as their regulatory framework. The table also depicts some "negative" exceptions such as Hawaii and Oklahoma. In the former, the rationale behind the negative effect on loan granting is probably connected to a geographic issue, in the latter; tax policies and the regulatory framework that characterize the State could influence this effect. (Perry 2012)

We can note that, in general, negative effects seem to be mainly located in poorer States, whereas richer States is tending to present a positive effect on loan granting.

4.7 The decision process driven by individual's characteristics by State, through years

To increase the precision of the model we decide to repeat the same procedure described above by adding a time dummy variable. As already said this could allow us to control for the time-specific fixed effect that is not explained by independent variables within the regression. Hence, the model we estimate has the following structure:

$$\begin{aligned} granted = &\propto_0 + \propto_1 Log(Loan\ Amount) + \propto_2 Log(Debt\ to\ income) \\ &+ \propto_3 Emp.Length + \propto_4 D(Log(applications\ per\ month) \\ &+ \propto_i States + \propto_j Years + u_t \end{aligned}$$

$$for\ i = 5, \dots 54$$

$$for\ j = 55, \dots 64$$

The results of the estimations, including the standardized coefficients and the option of robust errors to heteroscedasticity, are summarized in the following table:

Table 4.5: Estimation outcomes of the decision model. The effect of individual's characteristics through years

Logistic regression					Number of obs=	2,817,305
					Wald chi2(64)=	433396.19
					Prob>chi2=	0.000
Log pseudolikelihood=	-679141.21				Pseudo R2=	0.2964
Granted	**Coef.**	**Robust Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
Year						
2008	-0.417	0.168	-2.480	0.013	-0.746	-0.087
2009	-0.238	0.161	-1.470	0.141	-0.554	0.078
2010	**0.500**	0.158	3.170	0.002	0.191	0.810
2011	**0.853**	0.156	5.470	0.000	0.547	1.158
2012	**1.260**	0.155	8.120	0.000	0.956	1.563
2013	**1.287**	0.155	8.330	0.000	0.984	1.590
2014	**1.144**	0.154	7.410	0.000	0.841	1.447
2015	**1.288**	0.154	8.340	0.000	0.985	1.590
2016	**0.655**	0.154	4.240	0.000	0.352	0.957
2017	0.383	0.154	2.480	0.013	0.081	0.686
States						
AL	**-0.302**	0.052	-5.830	0.000	-0.403	-0.200
AR	**-0.273**	0.054	-5.020	0.000	-0.380	-0.166
AZ	**0.217**	0.050	4.320	0.000	0.118	0.315
CA	**0.125**	0.048	2.580	0.010	0.030	0.219
CO	**0.267**	0.051	5.270	0.000	0.167	0.366
CT	**0.165**	0.052	3.200	0.001	0.064	0.267
DC	**0.562**	0.067	8.340	0.000	0.430	0.694
DE	-0.108	0.064	-1.700	0.089	-0.233	0.016
FL	-0.022	0.049	-0.450	0.650	-0.117	0.073
GA	-0.051	0.049	-1.030	0.302	-0.148	0.046
HI	**-0.297**	0.058	-5.110	0.000	-0.410	-0.183
IA	0.527	1.002	0.530	0.599	-1.438	2.491
ID	-0.056	0.078	-0.720	0.472	-0.210	0.097
IL	0.104	0.049	2.110	0.035	0.008	0.201
IN	-0.019	0.051	-0.370	0.710	-0.119	0.081
KS	0.028	0.054	0.510	0.610	-0.078	0.133
KY	**-0.207**	0.053	-3.930	0.000	-0.311	-0.104
LA	**-0.167**	0.052	-3.200	0.001	-0.269	-0.065
MA	0.111	0.050	2.210	0.027	0.013	0.210
MD	0.120	0.050	2.380	0.017	0.021	0.218
ME	-0.095	0.073	-1.310	0.189	-0.238	0.047
MI	-0.014	0.050	-0.290	0.776	-0.112	0.084
MN	**0.327**	0.051	6.390	0.000	0.227	0.427
MO	-0.016	0.051	-0.310	0.760	-0.116	0.084
MS	**-0.359**	0.056	-6.410	0.000	-0.469	-0.249
MT	0.071	0.066	1.080	0.278	-0.058	0.200
NC	-0.005	0.050	-0.110	0.912	-0.103	0.092
ND	0.123	0.079	1.540	0.123	-0.033	0.278
NE	-0.010	0.063	-0.150	0.878	-0.134	0.114
NH	0.039	0.059	0.650	0.515	-0.077	0.154
NJ	0.103	0.049	2.080	0.037	0.006	0.200

NM	-0.081	0.057	-1.420	0.155	-0.193	0.031
NV	**0.230**	0.052	4.440	0.000	0.129	0.332
NY	0.075	0.049	1.540	0.123	-0.020	0.170
OH	-0.028	0.049	-0.570	0.572	-0.125	0.069
OK	**-0.144**	0.053	-2.700	0.007	-0.248	-0.039
OR	**0.216**	0.053	4.100	0.000	0.113	0.319
PA	-0.087	0.049	-1.770	0.077	-0.184	0.010
RI	0.016	0.060	0.270	0.787	-0.101	0.133
SC	**-0.157**	0.052	-3.030	0.002	-0.259	-0.055
SD	-0.108	0.071	-1.530	0.126	-0.246	0.030
TN	**-0.161**	0.051	-3.150	0.002	-0.261	-0.061
TX	0.008	0.049	0.160	0.874	-0.087	0.103
UT	0.131	0.055	2.370	0.018	0.023	0.240
VA	0.057	0.050	1.150	0.251	-0.040	0.155
VT	-0.011	0.071	-0.150	0.877	-0.150	0.128
WA	**0.202**	0.051	4.000	0.000	0.103	0.301
WI	0.028	0.052	0.540	0.588	-0.074	0.130
WV	**-0.162**	0.063	-2.590	0.010	-0.284	-0.039
WY	0.173	0.069	2.510	0.012	0.038	0.308
Log_Loan_Amount	**0.293**	0.002	130.490	0.000	0.289	0.298
Log_Debt_to_Income	**-0.027**	0.001	-18.350	0.000	-0.030	-0.024
Emp_Length	**0.372**	0.001	633.700	0.000	0.371	0.373
DLog_applications_per_month	0.017	0.013	1.300	0.193	-0.009	0.043
_cons	**-6.869**	**0.163**	**-42.130**	**0.000**	**-7.189**	**-6.550**

Looking at table 4.5, we observe that the estimated coefficients for the year 2008, 2009 and 2017 are not significant at 1% level. Moreover, the estimated coefficient associated to the percentage increase in the number of applications per month is no longer significant. This outcome is probably due to the introduction of the time-specific fixed effect and the consequent exclusion of variables considered less relevant in explaining the dependent variable.

Focusing on estimated coefficients for States, thanks to the clusterization provided by the Bureau of Economic Analysis (BEA), we follow the same procedure described before to look for any difference compared to the outcomes of the previous model.

Table 4.6: The decision process driven by individual's characteristics through years. States' estimated coefficients

Far West		% of Tot Economy	Rocky Mountain		% of Tot Economy	Plains		% of Tot Economy	SouthWest		% of Tot Economy
California	CA	13,3%	Colorado	CO	1,8%	Minnesota	MN	1,8%	Texas	TX	9,5%
Washington	WA	2,5%	Utah	UT	0,8%	Missouri	MO	1,6%	Arizona	AZ	1,6%
Oregon	OR	1,2%	Idaho	ID	0,4%	Iowa	IA	1,0%	Oklahoma	OK	1,1%
Nevada	NV	0,8%	Wyoming	WY	0,3%	Kansas	KS	0,8%	New Mexico	NM	0,5%
Hawaii	HI	0,4%	Montana	MT	0,3%	Nebraska	NE	0,6%			
Alaska	AK	0,3%				North Dakota	ND	0,3%			
						South Dakota	SD	0,3%			

SouthEast		% of Tot Economy	Great Lakes		% of Tot Economy	MidEast		% of Tot Economy	New England		% of Tot Economy
Florida	FL	4,9%	Illinois	IL	4,3%	New York	NY	8,1%	Massachusetts	MA	2,7%
Georgia	GA	2,8%	Ohio	OH	3,4%	Pennsylvania	PA	3,8%	Connecticut	CT	1,5%
North Carolina	NC	2,8%	Michigan	MI	2,6%	New Jersey	NJ	3,2%	New Hempshire	NH	0,4%
Virginia	VA	2,7%	Indiana	IN	1,8%	Maryland	MD	2,0%	Maine	ME	0,3%
Tennessee	TN	1,7%	Wisconsin	WI	1,7%	District of Columbia	DC	0,7%	Rhode Island	RI	0,3%
Louisiana	LA	1,5%				Delaware	DE	0,4%	Vermont	VT	0,2%
Alabama	AL	1,2%									
Kentucky	KY	1,1%									
South Carolina	SC	1,1%									
Arkansas	AR	0,7%									
Mississippi	MS	0,6%									
West Virginia	WV	0,4%									

In Table 4.6, we observe that our earlier interpretation is confirmed. In fact, applications submitted from States located in Far West region present again a higher probability to be accepted. On the contrary, potential borrowers who live in the Southeast seem to have a fewer probability of success in their requests.

The table also confirms the same exceptions discussed in the previous model.

4.8 The general decision process model

The final step of our analysis is to provide a model that includes all variables taken into consideration during the study. Thus, the model incorporates all individuals' characteristics together with macroeconomic variables relate to their State, as well as States' and years specification. Hence, the model we estimate has the following structure:

$$
\begin{aligned}
granted =& \propto_0 + \propto_1 \; Log(Loan\ Amount) + \propto_2 \; Log(Debt\ to\ income) \\
& + \propto_3 \; Emp.Length + \propto_4 \; D(Log(applications\ per\ month)) \\
& + \propto_5 \; Unemp.Rate + \propto_6 \; D(Log(Real\ GDP) + \propto_7 \; D\big(Log(PI)\big) \\
& + \propto_8 \; D(Log(Tot.NPL) + \propto_9 \; Number\ of\ banks \\
& + \propto_{10} \; D\big(Log(PCE)\big) + \propto_i \; States + \propto_j \; Years + u_t
\end{aligned}
$$

$$for\ i = 5, \dots 54$$

$$for\ j = 55, \dots 64$$

The results of the estimations, including the standardized coefficients and the option of robust errors to heteroscedasticity, are summarized in Table 4.7.

Table 4.7 Estimation outcomes of the general decision process model (States and years specification included)

Logistic regression					*Number of obs=*	*2,230,157*
					Wald chi2(69)=	*366251.86*
					Prob>chi2=	*0.000*
Log pseudolikelihood=	*-571013.87*				*Pseudo R2=*	*0.2966*
Granted	**Coef.**	**Robust Std. Err.**	**Z**	**P>z**	**[95% Conf. Interval]**	
Year						
2008	**-0.513**	0.168	-3.050	0.002	-0.842	-0.184
2009	**-0.432**	0.165	-2.610	0.009	-0.756	-0.108
2010	**0.491**	0.158	3.110	0.002	0.182	0.800
2011	**0.876**	0.156	5.620	0.000	0.570	1.181
2012	**1.268**	0.155	8.160	0.000	0.964	1.572
2013	**1.270**	0.155	8.170	0.000	0.965	1.574
2014	**1.204**	0.155	7.770	0.000	0.901	1.508
2015	**1.337**	0.155	8.620	0.000	1.033	1.641
2016	**0.705**	0.155	4.540	0.000	0.400	1.009
States						
AL	**-0.454**	0.066	-6.920	0.000	-0.582	-0.325
AR	**-0.398**	0.065	-6.090	0.000	\	-0.270
AZ	**0.207**	0.055	3.750	0.000	0.099	0.315
CA	-0.049	0.072	-0.680	0.494	-0.190	0.092
CO	**0.212**	0.060	3.560	0.000	0.095	0.328
CT	0.136	0.056	2.410	0.016	0.025	0.246
DC	**0.629**	0.073	8.580	0.000	0.486	0.773
DE	-0.114	0.070	-1.630	0.103	-0.250	0.023
FL	**-0.173**	0.067	-2.570	0.010	-0.305	-0.041
GA	**-0.239**	0.074	-3.230	0.001	-0.383	-0.094
HI	**-0.251**	0.064	-3.910	0.000	-0.376	-0.125
IA	0.224	0.998	0.220	0.823	-1.733	2.181
ID	0.079	0.091	0.860	0.388	-0.100	0.258
IL	**-0.442**	0.132	-3.360	0.001	-0.700	-0.184
IN	-0.109	0.061	-1.800	0.072	-0.227	0.010
KS	**-0.309**	0.093	-3.330	0.001	-0.491	-0.127
KY	**-0.388**	0.072	-5.410	0.000	-0.528	-0.247
LA	**-0.318**	0.064	-4.930	0.000	-0.444	-0.191

MA	0.116	0.055	2.100	0.036	0.008	0.225
MD	0.111	0.056	2.000	0.046	0.002	0.220
ME	-0.091	0.084	-1.090	0.277	-0.255	0.073
MI	-0.095	0.061	-1.560	0.120	-0.214	0.025
MN	-0.054	0.103	-0.530	0.596	-0.256	0.147
MO	**-0.357**	0.093	-3.820	0.000	-0.540	-0.174
MS	**-0.438**	0.065	-6.770	0.000	-0.565	-0.311
MT	0.065	0.073	0.890	0.374	-0.079	0.209
NC	-0.020	0.055	-0.360	0.720	-0.129	0.089
ND	-0.011	0.094	-0.110	0.909	-0.195	0.174
NE	-0.217	0.087	-2.510	0.012	-0.387	-0.048
NH	0.014	0.065	0.220	0.829	-0.114	0.142
NJ	0.064	0.055	1.150	0.249	-0.045	0.172
NM	-0.099	0.063	-1.580	0.114	-0.223	0.024
NV	**0.239**	0.057	4.220	0.000	0.128	0.350
NY	-0.031	0.060	-0.520	0.602	-0.148	0.086
OH	**-0.176**	0.064	-2.750	0.006	-0.301	-0.051
OK	**-0.422**	0.082	-5.130	0.000	-0.583	-0.260
OR	**0.246**	0.058	4.250	0.000	0.132	0.359
PA	**-0.209**	0.062	-3.380	0.001	-0.330	-0.088
RI	0.035	0.065	0.540	0.591	-0.092	0.162
SC	**-0.169**	0.058	-2.910	0.004	-0.282	-0.055
SD	-0.173	0.078	-2.200	0.028	-0.327	-0.019
TN	**-0.332**	0.070	-4.740	0.000	-0.470	-0.195
TX	**-0.576**	0.136	-4.220	0.000	-0.843	-0.309
UT	**0.160**	0.061	2.600	0.009	0.039	0.280
VA	-0.034	0.059	-0.570	0.566	-0.149	0.082
VT	-0.004	0.077	-0.050	0.962	-0.155	0.148
WA	**0.207**	0.056	3.690	0.000	0.097	0.318
WI	**-0.234**	0.081	-2.910	0.004	-0.392	-0.076
WV	**-0.217**	0.067	-3.250	0.001	-0.348	-0.086
WY	0.146	0.075	1.940	0.053	-0.002	0.293
Log_Loan_Amount	**0.283**	0.002	114.640	0.000	0.278	0.288
Log_Debt_to_Income	**-0.028**	0.002	-17.500	0.000	-0.032	-0.025
Emp_Length	**0.370**	0.001	583.200	0.000	0.368	0.371
DLog_applications_per_month	-0.001	0.014	-0.040	0.967	-0.027	0.026
Unemp_Rate	0.000	0.000	0.730	0.465	0.000	0.000
DLog_Real_GDP	**-1.464**	0.361	-4.060	0.000	-2.171	-0.757
DLog_Personal_Income	**-2.022**	0.358	-5.650	0.000	-2.724	-1.320
DLog_Total_NPL	-0.009	0.027	-0.330	0.738	-0.062	0.044
Number_of_Banks	**0.001**	0.000	4.790	0.000	0.001	0.002
DLog_Pers_Cons_Exp	**-2.498**	0.613	-4.070	0.000	-3.700	-1.295
_cons	**-6.760**	**0.168**	**-40.300**	**0.000**	**-7.089**	**-6.432**

Looking at table 4.7, we note that all years estimated coefficients are significantly different from zero. Their p-values allow rejecting the null hypothesis that these estimated coefficients may be zero, even at the 1% significance level. About estimated coefficients for individuals and States' macroeconomic characteristics, the introduction of the dummy variables for years and States leads to the exclusion of many predictors such as the estimated growth number of applications per month, the unemployment rate and the estimated growth of the total amount of non-performing loans. Meanwhile, the estimated coefficients for the number of banks in a specific state is re-included in the model. This occurrence could be interpreted considering that, in this case, we have taken into consideration the State specification, which was not included in the previous model (see 4.5 the decision process through years). Furthermore, for the same reason, we notice that the estimated coefficients for percentage increments of personal income and PCE index now present a negative relationship concerning the loan granting. The interpretation of these changes could be directly linked to the difference between models specification. This fact suggests that controlling for application's origin effectively affect the loan decision process.

In fact, while in the model presented in section 4.5, in which there is no specification regarding the state of origin of the loan application, the percentage increase in personal income (as well as the PCE) was interpreted as an incentive signal to the totality of the credit market given by the increase consumers' ability to repay their debts. The current model, which considers the local dimension of the credit market, this increase has the opposite effect.

This effect can be connected to the fact that an increase in personal income at the local level goes to stimulate consumption and lowers the demand for money for investments (on both traditional and P2P channels). In any case, further investigations would be needed on the specific topic that is outside the scope of our study.

Focusing on estimated coefficients for States, we repeat the same process described previously highlighting in green all States that present a positive relationship concerning the acceptance of a loan application together with States that present a negative relationship, highlighted in red.

Table 4.8 The general decision process model. States' estimated coefficients

Far West		% of Tot Economy	Rocky Mountain		% of Tot Economy	Plains		% of Tot Economy	SouthWest		% of Tot Economy
California	CA	13,3%	Colorado	CO	1,8%	Minnesota	MN	1,8%	Texas	TX	9,5%
Washington	WA	2,5%	Utah	UT	0,8%	Missouri	MO	1,6%	Arizona	AZ	1,6%
Oregon	OR	1,2%	Idaho	ID	0,4%	Iowa	IA	1,0%	Oklahoma	OK	1,1%
Nevada	NV	0,8%	Wyoming	WY	0,3%	Kansas	KS	0,8%	New Mexico	NM	0,5%
Hawaii	HI	0,4%	Montana	MT	0,3%	Nebraska	NE	0,6%			
Alaska	AK	0,3%				North Dakota	ND	0,3%			
						South Dakota	SD	0,3%			

SouthEast		% of Tot Economy	Great Lakes		% of Tot Economy	MidEast		% of Tot Economy	New England		% of Tot Economy
Florida	FL	4,9%	Illinois	IL	4,3%	New York	NY	8,1%	Massachusetts	MA	2,7%
Georgia	GA	2,8%	Ohio	OH	3,4%	Pennsylvania	PA	3,8%	Connecticut	CT	1,5%
North Carolina	NC	2,8%	Michigan	MI	2,6%	New Jersey	NJ	3,2%	New Hempshire	NH	0,4%
Virginia	VA	2,7%	Indiana	IN	1,8%	Maryland	MD	2,0%	Maine	ME	0,3%
Tennessee	TN	1,7%	Wisconsin	WI	1,7%	District of Columbia	DC	0,7%	Rhode Island	RI	0,3%
Louisiana	LA	1,5%				Delaware	DE	0,4%	Vermont	VT	0,2%
Alabama	AL	1,2%									
Kentucky	KY	1,1%									
South Carolina	SC	1,1%									
Arkansas	AR	0,7%									
Mississippi	MS	0,6%									
West Virginia	WV	0,4%									

In Table 4.8, we notice that previous interpretations are confirmed and enhanced by more information. For example, according to the outcomes provided by the model, applications submitted from regions such ad South East (all except North Carolina and Virginia), Great Lakes (Illinois, Ohio, and Wisconsin), South West (Texas and Oklahoma) and Plains (Missouri and Kansas) have a lower probability to be accepted. On the other hand, applications forwarded from regions like Far West or Rocky Mountain seem to have a higher probability to be accepted compared to the others.

Considering the increasing model specification through the last three models, these evidences allow us to capture the trend that characterizes the loans granted all over the US country. Moreover, as table 4.8 points out, people who live in the Fast West (excluding the singular "Hawaii case") seem to be facilitated in the loan decision process comparing to residents in others area. However, this evidence leave room for discussion about other drivers of the decision process to assess in the future.

4.9 Robustness Check

To fully assess the robustness of our empirical results we re-estimate the above relations using our dataset. It is important to underline that, since Lending Club is the world's largest P2P lending platform in the US, we expect almost same results from this procedure. Reports are presented from Appendix A to Appendix E. As already said; we assume that Lending Club is representative for the total P2P lending market in the US and shares many similarities in the loan application process with other P2P lending platforms.

The robustness check was conducted performing the bootstrap resampling with fifty replications of the dataset. Results provided for the first three models (Appendix A, B, and C) are consistent and confirm our previous analysis. Considering the other models presented (Appendix D and E), although some exceptions, the outcome of the resampling shares many similarities with original models. For this reason, evidence and interpretations concerning the States' effect on the loan granting can be considered valid. However, further investigations may be undertaken by substituting Lending Club data with data obtained from another P2P lending platform (e.g., Prosper, SoFi, Upstart).

Chapter 5 - Conclusions and Final Remarks

5.1 Findings suggest

Remembering the original purpose of the study and trying to summarize all findings, it is unquestionable that other factors, different from those considered, affect the loan decision process and we have not been able to capture them all. However, the analysis of the last three models gives a precise overview of the local dimension of the P2P lending marketplace and its decision process. Following the BEA clusterization by regions, we were able to find that, when assessing the creditworthiness, in addition to social and financial information such as credit history, employment status, and the overall financial situation, the lenders' decision might also be influenced by the perception they have with respect to people who lives in a certain geographic area. Overall, we found that people who live in the Far West have a higher probability of getting a loan, once other hard financial information is taken into account. Conversely, people who live in the Southeast are less likely to get loans. The origins of this geo-related discrimination could be found in the original location of Lending Club, headquartered in California, or can be traced back to the American Civil War (1861-1865).

5.2 Remarks and hints

The research and the relative findings suggest that considering the complexity of the loan decision process, the debate is far from a definitive conclusion. Rather, our analysis is just the beginning and might be further expanded in the future.

The interpretation provided and exceptions discovered in our models could be more deeply analyzed even considering many other aspects related to social and demographic diversity. The topic departs substantially from the economic context and introduces to sociological and cultural aspects related to the population under analysis, which was not part of our quest.

Our study adds one more piece to the analysis conducted about the loan decision process driven by ethical diversity and taste-based discrimination (Ravina 2012), opening a new discussion around some drivers conditioning of our decisions. In fact, the overall perceptions and stereotypes we have against someone we consider “different” from “us” are a meaningful part of our decision-making process, and not just only in the economic field.

5.3 Potential developments of the quest

Starting from our point of view and the results of our quest, to increase the precision of the estimates by including in the dataset further variables that describe the social and cultural characteristics of borrowers and lenders.

This further research could extend the context of the study by combining economic and social factors allowing refining an optimal model of the decision process, a unique model, which does contain the elements of our research and that of Ravina, would increase the reliability of the process accuracy.

Furthermore, other investigations might be conducted utilizing data substituting lending Club data with other platforms such as Propser, SoFi, and Upstarte, to consolidate the results and outcomes of our study.

References

de Roure, Calebe, Loriana Pelizzon, and Anjan Thakor. 2018. *P2P Lenders versus Banks: Cream Skimming or Bottom Fishing?* Working Paper, Frankfurt: Mimeo.

de Roure, Calebre, Loriana Pelizzon, and Paolo Tasca. 30/2016. *How does P2P lending fit into the consumer credit market?* Frankfurt: Deutsche Bundesbank.

Faia, Ester, and Monica Paiella. 2008. *P2P Lending, Information Externetalities, Social Network, and Loans' Substitution.* Frankfurt: Goethe University.

FDIC. 2018. *Federal Deposit Insurance Corporation.* 01 01. https://www.fdic.gov/.

Federal Reserve Bank of St. Louis. 2017. *Federal Reserve Economic Data | FRED | St. Louis Fed.* 2 12. https://fred.stlouisfed.org/.

Franks, Julian, Nicolas Serran-Velarde, and Oren Sussman. 2016. *Marketplace Lending, Information Aggregation, and Liquidity.* London: London Business School.

How much cost information website. 2015. *Howmuch.net - Understanding Money.* 2 14. https://howmuch.net/.

Lending Club. 2016. *Annual Report.* Report, San Francisco, USA: Lending Club.

Milne, Alister, and Paul Parboteeah. 2016. *The Business Model and Economics of Peer-to-Peer Lending.* 1000 Brussels, Belgium: Euopean Credit research Institute.

Perry, Gene. 2012. *Oklahoma Policy Institute.* 04 18. https://okpolicy.org/in-the-know-oklahomas-creditworthiness-imperiled-by-tax-cuts/.

PWC LLP. 2015. *Peer pressure, how peer-to-peer lending platforms are transforming the consumer lending industry.* Delaware, USA: PricewaterhouseCooper LLP.

Ravina, Enrichetta. 2012. "Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets." *SSRN website.* Edited by Elsevier Ltd. 11 22. doi: http://dx.doi.org/10.2139/ssrn.1107307 .

Rodríguez, Germàn. 2007. "Lecture Notes on Generalized Linear Models." In *Generalized Linear Models*, by Priceton University, Appendix B. Priceton: http://data.princeton.edu/wws509/notes/.

Tang, Huan. 2018. *Peer-to-Peer Lenders versus Banks, Sobstitues or Complements.* Paris: HEC Paris.

U.S. Department of Commerce. 2017. *U.S. Bureau of Economic Analysis (BEA).* 2 12. https://www.bea.gov.

Ziegler, Tania, EJ Reedy, Annie Le, Bryan Zhang, Randall S. Kroszner, and Kieran Garvey. 2017. *The americas alternative finance industry report.* Working report, Cambridge: University of Cambridge.

Wikipedia. 2017. *Categorical variable.* 12 31. Accessed 02 01, 2018. https://en.wikipedia.org/wiki/Categorical_variable.

Wikipedia. 2017. *Consumer Financial Protection Bureau.* 12 31. Accessed 02 01, 2018. https://en.wikipedia.org/wiki/Consumer_Financial_Protection_Bureau.

Wikipedia. 2017. *Maximum likelihood estimation.* 12 31. Accessed 02 01, 2018. https://en.wikipedia.org/wiki/Maximum_likelihood_estimation.

Wikipedia. 2017. *U.S. Securities and Exchange Commission.* 12 31. Accessed 02 01, 2018. https://it.wikipedia.org/wiki/Securities_and_Exchange_Commission.

Appendix A

Logistic regression			Number of obs=	2,230,157		
			Replications=	50		
			Wald chi2(10)=	611893.94		
			Prob>chi2=	0.000		
Log pseudolikelihood=	-578599.11		Pseudo R2=	0.2873		
Granted	**Observed Coef.**	**Bootstrap Normal-based Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
Log_Loan_Amount	**0.310**	0.003	115.770	0.000	0.305	0.316
Log_Debt_to_Income	**-0.035**	0.002	-22.650	0.000	-0.038	-0.032
Emp_Length	**0.365**	0.001	651.650	0.000	0.364	0.366
DLog_applications_per_month	**0.085**	0.011	7.540	0.000	0.063	0.107
Unemp_Rate	**0.004**	0.000	35.340	0.000	0.003	0.004
DLog_Real_GDP	**-3.106**	0.346	-8.980	0.000	-3.784	-2.428
DLog_Personal_Income	**7.763**	0.302	25.740	0.000	7.172	8.355
DLog_Total_NPL	**-0.415**	0.025	-16.660	0.000	-0.464	-0.366
Number_of_Banks	**0.000**	0.000	3.130	0.002	0.000	0.000
DLog_Pers_Cons_Exp	**2.668**	0.312	8.560	0.000	2.057	3.279
_cons	**-6.563**	0.032	-207.170	0.000	-6.625	-6.501

Appendix B

Logistic regression				Number of obs=	2,230,157		
				Replications=	50		
				Wald chi2(19)=	636313.7		
				Prob>chi2=	0.000		
Log pseudolikelihood=		-572240.17		Pseudo R2=	0.2951		
Granted		**Observed Coef.**	**Bootstrap Normal-based Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
Year							
	2008	-0.333	0.168	-1.980	0.048	-0.664	-0.003
	2009	0.027	0.181	0.150	0.880	-0.328	0.383
	2010	**0.568**	0.166	3.420	0.001	0.243	0.893
	2011	**0.861**	0.157	5.480	0.000	0.553	1.168
	2012	**1.329**	0.159	8.370	0.000	1.017	1.640
	2013	**1.367**	0.163	8.400	0.000	1.048	1.686
	2014	**1.161**	0.158	7.320	0.000	0.850	1.471
	2015	**1.314**	0.159	8.250	0.000	1.002	1.625
	2016	**0.672**	0.159	4.240	0.000	0.361	0.983
Log_Loan_Amount		**0.286**	0.003	109.380	0.000	0.281	0.291
Log_Debt_to_Income		**-0.033**	0.002	-19.930	0.000	-0.037	-0.030
Emp_Length		**0.369**	0.001	631.090	0.000	0.368	0.371
DLog_applications_per_month		0.001	0.012	0.120	0.907	-0.023	0.026
Unemp_Rate		0.000	0.000	-0.300	0.765	0.000	0.000
DLog_Real_GDP		**-1.060**	0.352	-3.010	0.003	-1.750	-0.369
DLog_Personal_Income		-0.830	0.356	-2.330	0.020	-1.527	-0.133
DLog_Total_NPL		0.047	0.026	1.830	0.067	-0.003	0.097
Number_of_Banks		0.000	0.000	-1.270	0.202	0.000	0.000
DLog_Pers_Cons_Exp		**4.279**	0.447	9.580	0.000	3.404	5.154
_cons		**-6.910**	0.166	-41.700	0.000	-7.234	-6.585

Appendix C

Logistic regression			Number of obs=	2,817,305	
			Replications=	48	
			Wald chi2(47)= .		
			Prob>chi2= .		
Log pseudolikelihood=	-691623.71		Pseudo R2=	0.2835	

Granted	Observed Coef.	Bootstrap Normal-based Std. Err.	z	P>z	[95% Conf. Interval]	
Log_Loan_Amount	**0.324**	0.002	158.590	0.000	0.320	0.328
Log_Debt_to_Income	**-0.033**	0.002	-21.340	0.000	-0.036	-0.030
Emp_Length	**0.368**	0.001	645.750	0.000	0.367	0.369
DLog_applications_per_month	**0.113**	0.012	9.070	0.000	0.089	0.138
States						
AL	**-0.302**	0.053	-5.710	0.000	-0.405	-0.198
AR	**-0.279**	0.052	-5.330	0.000	-0.382	-0.177
AZ	**0.208**	0.050	4.180	0.000	0.111	0.306
CA	**0.126**	0.046	2.740	0.006	0.036	0.216
CO	**0.251**	0.048	5.230	0.000	0.157	0.346
CT	**0.168**	0.051	3.330	0.001	0.069	0.267
DC	**0.550**	0.056	9.860	0.000	0.441	0.660
DE	-0.120	0.061	-1.970	0.049	-0.240	0.000
FL	-0.035	0.046	-0.770	0.440	-0.125	0.054
GA	-0.069	0.048	-1.460	0.144	-0.162	0.024
HI	**-0.291**	0.063	-4.600	0.000	-0.414	-0.167
IA	-0.684	0.977	-0.700	0.484	-2.599	1.231
ID	**-0.379**	0.079	-4.800	0.000	-0.534	-0.225
IL	0.092	0.048	1.920	0.055	-0.002	0.186
IN	-0.033	0.048	-0.680	0.498	-0.128	0.062
KS	0.030	0.051	0.600	0.548	-0.069	0.130
KY	**-0.214**	0.050	-4.260	0.000	-0.313	-0.116
LA	**-0.169**	0.053	-3.210	0.001	-0.272	-0.066
MA	0.113	0.050	2.280	0.023	0.016	0.210
MD	0.106	0.049	2.140	0.032	0.009	0.203
ME	**-0.313**	0.066	-4.750	0.000	-0.443	-0.184
MI	-0.021	0.049	-0.420	0.673	-0.116	0.075
MN	**0.317**	0.054	5.840	0.000	0.211	0.424
MO	-0.024	0.050	-0.480	0.632	-0.121	0.073
MS	**-0.424**	0.057	-7.440	0.000	-0.535	-0.312
MT	0.068	0.062	1.090	0.275	-0.054	0.189
NC	-0.008	0.047	-0.160	0.872	-0.099	0.084
ND	-0.069	0.091	-0.760	0.450	-0.248	0.110
NE	-0.180	0.072	-2.490	0.013	-0.321	-0.038
NH	0.029	0.065	0.440	0.659	-0.099	0.156
NJ	0.100	0.045	2.240	0.025	0.013	0.188
NM	-0.081	0.053	-1.510	0.130	-0.185	0.024
NV	**0.234**	0.051	4.620	0.000	0.135	0.333
NY	0.079	0.047	1.690	0.091	-0.013	0.172
OH	-0.038	0.050	-0.750	0.455	-0.136	0.061
OK	**-0.150**	0.053	-2.850	0.004	-0.253	-0.047
OR	**0.209**	0.050	4.160	0.000	0.111	0.308
PA	-0.092	0.045	-2.060	0.040	-0.180	-0.004
RI	0.014	0.054	0.260	0.797	-0.092	0.120
SC	**-0.173**	0.048	-3.630	0.000	-0.266	-0.079
SD	-0.099	0.064	-1.550	0.121	-0.224	0.026
TN	**-0.170**	0.056	-3.040	0.002	-0.279	-0.060
TX	-0.005	0.048	-0.100	0.920	-0.099	0.089
UT	0.136	0.054	2.530	0.011	0.031	0.241
VA	0.056	0.048	1.180	0.238	-0.037	0.150
VT	-0.009	0.079	-0.110	0.910	-0.165	0.147
WA	**0.199**	0.051	3.910	0.000	0.099	0.299
WI	0.024	0.053	0.450	0.650	-0.079	0.127
WV	0.056	0.054	1.040	0.300	-0.049	0.160
WY	**0.183**	0.071	2.570	0.010	0.043	0.322
_cons	**-6.274**	0.050	-124.590	0.000	-6.373	-6.176

Appendix D

Logistic regression			Number of obs=	2,817,305		
			Replications=	46		
			Wald chi2(45)= .			
			Prob>chi2= .			
Log pseudolikelihood=	-679141.21		Pseudo R2=	0.2964		
Granted	**Observed Coef.**	**Bootstrap Normal-based Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
Year						
2008	**-0.417**	0.161	-2.580	0.010	-0.733	-0.100
2009	-0.238	0.151	-1.570	0.115	-0.533	0.058
2010	**0.500**	0.149	3.360	0.001	0.209	0.792
2011	**0.853**	0.142	5.990	0.000	0.574	1.132
2012	**1.260**	0.142	8.870	0.000	0.981	1.538
2013	**1.287**	0.143	8.990	0.000	1.006	1.568
2014	**1.144**	0.143	8.000	0.000	0.864	1.425
2015	**1.288**	0.142	9.050	0.000	1.009	1.566
2016	**0.655**	0.143	4.580	0.000	0.374	0.935
2017	**0.383**	0.142	2.700	0.007	0.105	0.662
States						
AL	**-0.302**	0.059	-5.140	0.000	-0.417	-0.187
AR	**-0.273**	0.062	-4.410	0.000	-0.395	-0.152
AZ	**0.217**	0.058	3.720	0.000	0.103	0.331
CA	0.125	0.060	2.080	0.038	0.007	0.242
CO	**0.267**	0.056	4.740	0.000	0.156	0.377
CT	**0.165**	0.058	2.840	0.005	0.051	0.280
DC	**0.562**	0.067	8.330	0.000	0.430	0.694
DE	-0.108	0.079	-1.370	0.172	-0.264	0.047
FL	-0.022	0.060	-0.370	0.712	-0.139	0.095
GA	-0.051	0.062	-0.830	0.409	-0.172	0.070
HI	**-0.297**	0.060	-4.940	0.000	-0.414	-0.179
IA	0.527	1.106	0.480	0.634	-1.641	2.695
ID	-0.056	0.089	-0.630	0.529	-0.232	0.119
IL	0.104	0.061	1.720	0.086	-0.015	0.223
IN	-0.019	0.059	-0.320	0.748	-0.134	0.097
KS	0.028	0.065	0.420	0.673	-0.100	0.155
KY	**-0.207**	0.068	-3.070	0.002	-0.340	-0.075
LA	-0.167	0.067	-2.480	0.013	-0.299	-0.035
MA	0.111	0.054	2.050	0.041	0.005	0.218
MD	0.120	0.062	1.920	0.054	-0.002	0.242
ME	-0.095	0.084	-1.140	0.256	-0.260	0.069
MI	-0.014	0.062	-0.230	0.819	-0.136	0.108
MN	**0.327**	0.060	5.440	0.000	0.209	0.445
MO	-0.016	0.062	-0.250	0.802	-0.137	0.106
MS	**-0.359**	0.068	-5.270	0.000	-0.493	-0.226
MT	0.071	0.062	1.160	0.247	-0.049	0.192
NC	-0.005	0.057	-0.100	0.923	-0.117	0.106
ND	0.123	0.087	1.410	0.158	-0.047	0.292
NE	-0.010	0.066	-0.150	0.882	-0.138	0.119
NH	0.039	0.070	0.550	0.580	-0.098	0.175
NJ	0.103	0.057	1.790	0.073	-0.010	0.216
NM	-0.081	0.065	-1.260	0.208	-0.208	0.045
NV	**0.230**	0.057	4.060	0.000	0.119	0.341
NY	0.075	0.061	1.220	0.223	-0.045	0.195
OH	-0.028	0.063	-0.450	0.655	-0.151	0.095
OK	-0.144	0.056	-2.540	0.011	-0.255	-0.033
OR	**0.216**	0.067	3.220	0.001	0.085	0.347
PA	-0.087	0.059	-1.470	0.142	-0.204	0.029
RI	0.016	0.059	0.270	0.784	-0.099	0.131
SC	**-0.157**	0.059	-2.680	0.007	-0.272	-0.042
SD	-0.108	0.063	-1.730	0.084	-0.231	0.014
TN	-0.161	0.064	-2.520	0.012	-0.286	-0.036
TX	0.008	0.060	0.130	0.898	-0.109	0.125
UT	0.131	0.070	1.880	0.060	-0.005	0.268
VA	0.057	0.060	0.950	0.341	-0.061	0.175
VT	-0.011	0.070	-0.160	0.875	-0.148	0.126
WA	**0.202**	0.055	3.690	0.000	0.095	0.310
WI	0.028	0.063	0.440	0.657	-0.096	0.152
WV	-0.162	0.068	-2.400	0.017	-0.294	-0.029
WY	0.173	0.085	2.050	0.041	0.007	0.339
Log_Loan_Amount	**0.293**	0.003	112.920	0.000	0.288	0.298
Log_Debt_to_Income	**-0.027**	0.002	-17.340	0.000	-0.030	-0.024
Emp_Length	**0.372**	0.001	598.300	0.000	0.371	0.373
DLog_applications_per_month	0.017	0.013	1.300	0.193	-0.009	0.043
_cons	**-6.869**	0.148	-46.530	0.000	-7.158	-6.580

Appendix E

Logistic regression		Number of obs=	2,230,157
		Replications=	47
		Wald chi2(45)=	.
		Prob>chi2=	.
Log pseudolikelihood=	-571013.87	Pseudo R2=	0.2966

Granted		Observed Coef.	Bootstrap Normal-based	z	P>z	[95% Conf. Interval]	
Year							
	2008	**-0.513**	0.176	-2.910	0.004	-0.858	-0.168
	2009	**-0.432**	0.168	-2.570	0.010	-0.761	-0.103
	2010	**0.491**	0.155	3.160	0.002	0.186	0.795
	2011	**0.876**	0.149	5.890	0.000	0.584	1.167
	2012	**1.268**	0.151	8.420	0.000	0.973	1.563
	2013	**1.270**	0.152	8.330	0.000	0.971	1.568
	2014	**1.204**	0.151	7.950	0.000	0.908	1.501
	2015	**1.337**	0.153	8.740	0.000	1.037	1.637
	2016	**0.705**	0.151	4.650	0.000	0.408	1.001
States							
AL		**-0.454**	0.057	-7.950	0.000	-0.565	-0.342
AR		**-0.398**	0.067	-5.970	0.000	-0.529	-0.268
AZ		**0.207**	0.052	3.960	0.000	0.104	0.310
CA		-0.049	0.063	-0.790	0.432	-0.172	0.074
CO		**0.212**	0.056	3.750	0.000	0.101	0.322
CT		**0.136**	0.053	2.560	0.010	0.032	0.239
DC		**0.629**	0.076	8.280	0.000	0.480	0.778
DE		-0.114	0.085	-1.340	0.181	-0.280	0.053
FL		**-0.173**	0.059	-2.910	0.004	-0.290	-0.056
GA		**-0.239**	0.059	-4.040	0.000	-0.355	-0.123
HI		**-0.251**	0.069	-3.620	0.000	-0.386	-0.115
IA		0.224	0.971	0.230	0.818	-1.679	2.127
ID		0.079	0.108	0.730	0.467	-0.134	0.291
IL		**-0.442**	0.104	-4.240	0.000	-0.646	-0.238
IN		-0.109	0.061	-1.800	0.072	-0.227	0.010
KS		**-0.309**	0.078	-3.980	0.000	-0.461	-0.157
KY		**-0.388**	0.064	-6.080	0.000	-0.513	-0.263
LA		**-0.318**	0.059	-5.390	0.000	-0.433	-0.202
MA		0.116	0.060	1.940	0.052	-0.001	0.233
MD		0.111	0.053	2.080	0.038	0.006	0.216
ME		-0.091	0.071	-1.270	0.202	-0.230	0.049
MI		-0.095	0.058	-1.640	0.102	-0.208	0.019
MN		-0.054	0.083	-0.650	0.513	-0.218	0.109
MO		**-0.357**	0.077	-4.650	0.000	-0.508	-0.207
MS		**-0.438**	0.057	-7.720	0.000	-0.549	-0.327
MT		0.065	0.077	0.840	0.399	-0.086	0.217
NC		-0.020	0.054	-0.370	0.715	-0.127	0.087
ND		-0.011	0.084	-0.130	0.898	-0.176	0.155
NE		**-0.217**	0.083	-2.620	0.009	-0.380	-0.055
NH		0.014	0.063	0.220	0.823	-0.109	0.137
NJ		0.064	0.052	1.230	0.217	-0.037	0.165
NM		-0.099	0.061	-1.620	0.106	-0.220	0.021
NV		**0.239**	0.054	4.440	0.000	0.133	0.344
NY		-0.031	0.055	-0.560	0.574	-0.140	0.077
OH		**-0.176**	0.057	-3.100	0.002	-0.287	-0.065
OK		**-0.422**	0.068	-6.210	0.000	-0.555	-0.289
OR		**0.246**	0.059	4.190	0.000	0.131	0.360
PA		**-0.209**	0.057	-3.660	0.000	-0.321	-0.097
RI		0.035	0.076	0.460	0.645	-0.114	0.183
SC		**-0.169**	0.056	-3.030	0.002	-0.278	-0.059
SD		-0.173	0.068	-2.530	0.011	-0.307	-0.039
TN		**-0.332**	0.065	-5.140	0.000	-0.459	-0.206
TX		**-0.576**	0.108	-5.310	0.000	-0.788	-0.363
UT		**0.160**	0.061	2.610	0.009	0.040	0.280
VA		-0.034	0.053	-0.640	0.524	-0.138	0.070
VT		-0.004	0.076	-0.050	0.961	-0.152	0.145
WA		**0.207**	0.055	3.780	0.000	0.100	0.315
WI		**-0.234**	0.064	-3.670	0.000	-0.359	-0.109
WV		**-0.217**	0.065	-3.340	0.001	-0.345	-0.090
WY		0.146	0.063	2.330	0.020	0.023	0.268
Log_Loan_Amount		**0.283**	0.002	128.260	0.000	0.278	0.287
Log_Debt_to_Income		**-0.028**	0.002	-18.970	0.000	-0.031	-0.026
Emp_Length		**0.370**	0.001	601.680	0.000	0.369	0.371
DLog_applications_per_month		-0.001	0.012	-0.050	0.963	-0.025	0.024
Unemp_Rate		0.000	0.000	0.760	0.446	0.000	0.000
DLog_Real_GDP		**-1.464**	0.325	-4.500	0.000	-2.102	-0.826
DLog_Personal_Income		**-2.022**	0.342	-5.910	0.000	-2.692	-1.352
DLog_Total_NPL		-0.009	0.025	-0.370	0.714	-0.058	0.039
Number_of_Banks		**0.001**	0.000	5.800	0.000	0.001	0.002
DLog_Pers_Cons_Exp		**-2.498**	0.667	-3.740	0.000	-3.805	-1.190
_cons		**-6.760**	0.167	-40.410	0.000	-7.088	-6.433