UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"

CORSO DI LAUREA MAGISTRALE
IN ECONOMICS AND FINANCE

TESI DI LAUREA

"FINTECH, REGTECH AND THE ROLE OF ALTERNATIVE LENDING: AN ANALYSIS OF THE P2P PLATFORM LENDINGCLUB"

RELATORE:

CH.MO PROF. BRUNO MARIA PARIGI

LAUREANDO/A: MICHELE DI PIETRO

MATRICOLA N. 1189914

ANNO ACCADEMICO 2019 – 2020

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere. Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

Michele Di Pietro

Ringrazio il professor Parigi e la professoressa Dal Bianco per i preziosi
consigli, senza di loro tutto questo non sarebbe stato possibile.
Ringrazio i miei genitori e familiari per avermi sempre
sostenuto in tutte le scelte fin dal principio.
Ringrazio gli amici e le persone
che mi sono rimaste vicine
durante il viaggio.

TABLE OF CONTENTS

INTRODUCTION

Financial technology or FinTech has been playing an important role in reshaping the financial system by using innovative technologies to create new business models and automate the lengthy processes of the traditional financial intermediation. Although the expansion of FinTechs was for most part beneficial, it also raised important regulatory concerns. If these companies do not receive an adequate regulation they may quickly become a risk for the financial system. Countries like US are putting effort in creating a more collaborative environment, however the way by which regulators will respond to this new wave of innovation in the financial system is yet unclear. The application of technology and big data for regulatory purposes (“RegTech”) can help businesses to comply with the relevant laws and for regulators to supervise the markets and their participants. The main idea behind FinTech and RegTech solutions is that by using a large amount of data together with powerful algorithms and machine learning techniques, a company can offer a better value proposition to its customers.

In our analysis we will focus on peer-to-peer (P2P) or alternative lending activity to understand if the additional data that alternative lenders collect is helpful in providing a more accurate measure of the borrowers’ creditworthiness than what the traditional scoring system can do with the credit history. Our study will be conducted on the online platform LendingClub because the company publicly releases all the loan-level data generated each quarter through its platform, making it an attractive candidate to conduct research about.

Although many FinTech companies such as LendingClub (LC) are coming up with new innovative credit scoring mechanisms the vast majority of lenders still base its lending decisions on traditional credit scores such as FICO. FICO Scores use the information from the credit reports of a borrower to generate a 3-digit number to gauge a potential borrower's creditworthiness. The problem with this an other traditional scoring systems is that they ignore many information which are often easy to measure that could be important indicators of a borrower’s riskiness. LendingClub proposed an advanced credit scoring system based on multiple sources of data that aims to help investors in deciding how to allocate their money in the loan notes. With our study we want to address the following research questions. Is this innovative scoring system effectively better than the traditional ones? What are the most important factors used by LendingClub in determining the future loan performance? Do these additional information add a significant contribute to FICO Scores in predicting the future loan delinquencies (30-120 past due) and in explaining the pricing decisions of the company?

The study will be organized as follows. Chapter 1 will provide a broad picture about what is FinTech, its evolution and why it is innovating and disrupting the financial services industry. In Chapter 2 we will discuss the regulatory framework in which US FinTech companies operate, and see how technology and data can also help regulators in their supervising and monitoring activities. In Chapter 3 we will focus on peer-to-peer (P2P) or alternative lending. In particular we will introduce the US platform LendingClub, define its business model, and make some regulatory consideration about the same. Chapter 4 will describe the dataset on which we will perform our analysis and discuss some of the most meaningful information pertaining to loans and borrowers. In Chapter 5 we will present the final results of our study. From our findings we discovered that LC's Grades have a much higher predictive ability than FICO Scores, and the additional information effectively allowed the company to provide investor with a more accurate picture about the borrowers' creditworthiness. Among the most important features we have information from the credit reports and information that are self-reported by the borrowers. We will later discover that all these types of information bring significant improvements to FICO Scores in predicting the future loan delinquencies and in explaining the pricing decisions of the company.

CHAPTER I – THE EVOLUTION OF FINTECH

The financial system undoubtedly plays a very important role for our society, since it allows the economy to run smoothly and efficiently. We can think about the financial industry as "the oil in the engine – the engine will not turn without it but it has no role in making the engine turn" (DeMartino and McCloskey, 2016). Ultimately, financial institutions such as banks exist to facilitate a fluid exchange of funds between who wants to lend money (lender) and who wants to take it (borrower). Banks have also a very unique role in mitigating market imperfections such as asymmetry of information[1], contracting costs, duration mismatch[2] by taking part into activities like maturity transformation, payment services and information processing (Navaretti et al., 2017).

The financial industry has evolved a lot during the last decades due to frequent changes in political and geographic regimes and legislation. "FinTech" is the latest step of this evolution, and is often seen as the "new marriage" between financial services and information technology (IT). Even if the diffusion of FinTech seems a relatively new phenomenon, the linkage between Finance and Technology has its roots back in the history. The introduction of the telegraph in 1838 and the laying of the first transatlantic cable[3] in 1866 by the Atlantic Telegraph Company provided the fundamental infrastructure for the first major period of financial globalization. Subsequently, the introduction of the Automatic Teller Machine (ATM) in 1967 by Barclays Bank arguably marks the beginning of the modern evolution of FinTech (Arner et al., 2015).

One of the first appearance of the word "FinTech" in the scientific literature dates back to the 70s. Bettinger (1972) defined FinTech as "an acronym which stands for of financial technology, combining bank expertise with modern management science techniques and the computer". Prior to the 2008 financial crisis, FinTech was driven by incumbent[4] financial institutions and their investment on technology to support their risk management and internet banking operations. After 2008, a new era for financial technology ("FinTech 3.0") emerged around the globe (Arner et al., 2015). During the Financial Crisis the image of banks, especially in UK and US, took a severe hit. As a consequence the mindset of people on who has the resources and legitimacy to handle their finances shifted from banks to technology firms. This new era of FinTech is not anymore centered around the financial products or services but rather

---

[1] Asymmetry of information occurs when one party to an economic transaction possesses greater knowledge.
[2] Duration mismatch is when an intermediary borrows through short-term instruments, but lends in the long-term.
[3] A transatlantic cable is an undersea cable running under the Atlantic Ocean used for telegraph communications.
[4] Incumbent in this case refers to the leader/s in the financial industry.

on who delivers them to retail and wholesale clients through the application of new and innovative technology.

The fact that FinTech players are entering the market without a financial compliance[5] culture, poses serious challenges for regulators. To support the digital transition, both banks and FinTechs should be able to equally compete in the same environment. The regulators should hence develop a regulatory framework[6] that balances the benefits of innovation with the risks of these new approaches. Instead of introducing regulation impulsively potentially stifling innovation, regulators should step into action when the risks posed by disruptive technology become systemic and materially destabilizing. Ideally, this would progressively require a coordinated regulatory response as various types of finance providers become more correlated in the global finance domain (Anagnostopoulos, 2018).

Even if important steps have been taken, FinTech is yet an underexplored phenomenon and represents an important challenge not only for regulators but also for academia and managers in the financial industry.

1.1 – Disruption and Disintermediation

According to Christensen's (2003) theory of disruptive innovation we can distinguish between "sustaining" and "disruptive" innovations. Sustaining innovation aims to improve pre-established products, while disruptive innovation tries to create new markets and values to disrupt the already existing ones. We can consider sustaining FinTech the financial institutions that use internet and the automated processing of information to defend their market position. We can define instead as disruptive FinTech the start-ups that challenge the incumbents by offering new services and a higher degree of flexibility, security and operational efficiency (McWaters, 2015).

Even if the financial industry has always been keen to adopt and make an extensive use of IT solutions and new technologies, the wave of innovation brought by these new business models is reshaping the dynamics of this industry once again. Therefore FinTech companies were able to quickly gain market share and conquer customers in a domain which was previously dominated by well-established entities such as banks, insurance companies and credit unions[7].

[5] Financial compliance requires that a firm take steps to comply with relevant laws, policies, and regulations.
[6] The regulatory framework includes any laws, regulations, decrees and policies officially developed and approved by the government, for the purposes of regulating a specific topic.
[7] Similarly to a bank, a credit union uses the money that its members deposit to make loans to other union members.

According to Gomber et al. (2017) there are 3 main reasons that can explain the rapid adoption of FinTech products and services in the financial industry.

First, FinTech companies offer solutions that can meet the needs of customers which were not served by the traditional financial institutions. These untapped markets represent an exceptional opportunity for them to start and expand their business. For example many individuals who do not have access to traditional forms credit because they don't have a solid credit history[8] can be served by a FinTech lender with a superior ability of assessing their riskiness. Crespo et al. (2018) argued Targeted lending using big data and machine learning can bring significant benefits to lenders.

Second, FinTechs widened the spectrum of opportunities for their products by applying innovative technologies and concepts that redefined the industry standards.

Third, start-ups and IT businesses have a culture which is substantially different whit respect to traditional institutions. FinTechs have in fact agile and flexible business models, and seek to lead the change by competing with traditional financial service providers.

The long-run driver for FinTech as a disruptor is the cost of financial intermediation, which has historically been high. In his studies Philippon (2017) showed that the average cost of intermediation has been hovering around 2% of transaction amounts since 1880. Speculating about the reasons behind high cost of transaction we could think about a lack of competition on the supply side, however the large amount of money that large financial institutions have to spend to sustain their business is one of the key drivers for such high intermediation costs. Technology is now a much cheaper intermediary and a driver of competition.

If traditional financial services organizations have acted in the past as intermediaries by providing an invaluable services to clients, their functions now risk to be taken over by new business models developed by non-banking financial institutions and capital markets. Banking disintermediation refers to the disintermediation of credit and has been in place for many decades. Banks preferring to sell-off their loans instead of holding them on their balance sheets; borrowers going directly to the capital markets[9] to satisfy their needs; and savers which invest directly in securities[10] are all tangible forms of this phenomenon.

As a consequence banks moved from the low-margin business of credit provisioning to high-margin fee-based businesses such as investment banking, insurance and wealth

[8] Credit history is a record of a borrower's repayment of debts and it measures its ability and demonstrated responsibility in repaying debts.
[9] A firm can borrow money by issuing debt in the capital markets instead of going to a bank.
[10] Savers today can now invest directly on the market in various ways. For example the FinTech company Robinhood provides commission-free trading for stocks, ETFs, options, ADRs, and cryptocurrencies.

management. However, with the evolution of FinTech there is also the risk for banks of being disintermediated from their customers in their high-margin fee-based businesses.

According to PwC (2016) disintermediation is FinTech's "most powerful weapon" to shift the paradigm of traditional intermediary roles and make them obsolete. New capital raising models are one of the latest examples of how disintermediation is playing out in the financial system. Equity crowdfunding has the potential to disintermediate early stage financing, and to empower the access to capital and investment opportunities, but this is not the only place where fintech startups are growing fast. In fact other areas such as clearing and settlement, where blockchain-based startups are offering fully automated settlement[11] solutions without the need for a central counterparty, and asset management are also driving the change.

1.2 – FinTech Activities

FinTech aims to bring innovation into a wide range of financial activities. McWaters (2015) identified six core functions in which we can have FinTech activity including payments, market provisioning, investment management, insurance, deposits and lending, capital raising.

Successful FinTech applications show very similar characteristics (Das, 2019). They all have a precise definition of the problem they want to solve and the market in which they want to operate[12], and they are able to handle complex problems of data extraction, integration and analytics. Examples of innovations that are central to FinTech activities include cryptocurrencies and the blockchain, new trading systems, artificial intelligence and machine learning, peer-to-peer (P2P) lending[13], equity crowdfunding[14] and mobile payment systems (Philippon, 2019). The lending function is currently the most recurrent activity among all the FinTech companies, and our study will be finalized to understand if alternative forms of lending can effectively bring significant benefits to the financial services industry.

1.3 – The FinTech Ecosystem

As mentioned before the new wave of innovation after 2008 has been driven by a group of companies which are more agile, flexible and open minded than traditional institutions. The broader definition of FinTech covers today a range of different business models with four distinct segments (McKinsey, 2016). The companies in each segment can be divided in:

---

[11] Settlement defines a process whereby securities are delivered, usually in exchange of money.
[12] The FinTech space is wide and its participants must sharply define the niche in which they want to operate to create a unique value proposition.
[13] Peer-to-peer lending directly connects people or companies that want to loan money with other people or businesses that want to borrow money.
[14] Crowdfunding is the funding of a project by raising small amounts of money from a large number of people.

- New entrants, startups, and attackers of small dimension that seek to enter into the financial services industry by using new technologies (i.e. LendingClub);
- Incumbent financial institutions which are large entities that invest in technology to defend their position and improve their performance (i.e. Wells Fargo, Goldman Sachs);
- Large tech companies which offer financial services to improve their relationship with users and monetize their data in the meanwhile (i.e. Apple, Ant Financial)
- Infrastructure providers which are smaller tech companies that help financial institutions to digitize their business and improve their customer experience (i.e. FNX)

In the past, financial technology has often been seen as a tool related only to the financial world. Today, also Big Tech companies have entered the financial domain, and started to steal customers that traditionally have been served by incumbent financial institutions. Big Tech firms usually enter the market by offering payment services, and later expanded into lending, insurance, savings and investment products. Forst et. al. (2019) studied the drivers and implications of the growth of Big Techs in the credit market and found that in both Argentina and China, the firms that accessed credit expanded their product offerings more than those that did not.

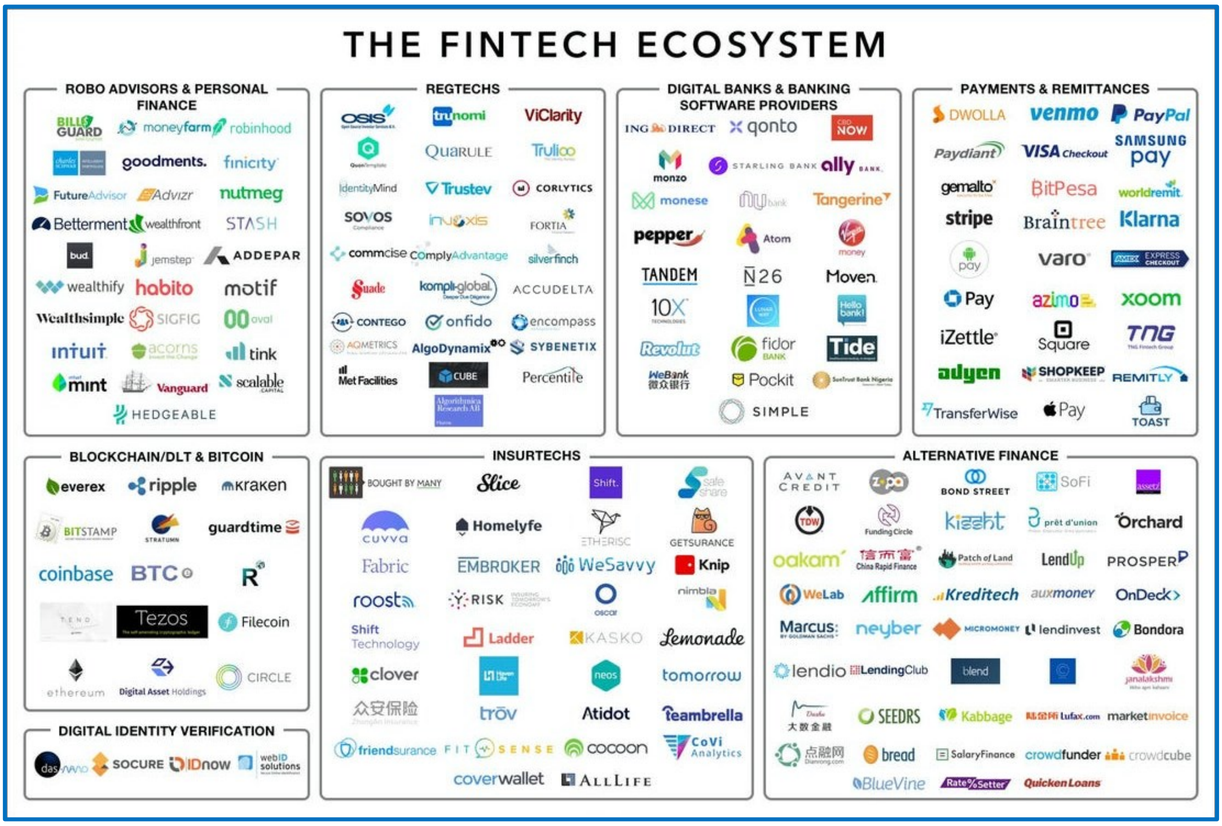
From now on we will define as “TechFin” a Big Tech firm that wants to deliver financial services on the basis of existing tech solutions. Famous examples of TechFins include Google, Amazon, Facebook and Apple in the United States, and Baidu, Alibaba, and Tencent in China.

The difference between FinTech start-ups and TechFin is in the approach by which they reach to their customers. FinTechs usually start by screening[15] the market in search for a service in which incumbents perform poorly or not perform at all. Then they try to find a solution that can be sold directly to customer or to the intermediaries themselves. TechFins start instead from a large amount of data that they have collected about their customers and then by leveraging this information advantage they move to the financial world. They can come from different fields including software, hardware, social media, e-commerce, and telecommunications, but there is one thing that they all have in common which is a privileged access to customer data from different sources which allows them to better predict user’s preferences and behavior. Both FinTechs and TechFins benefit from economies of scale from their data, however while FinTechs put financial intermediation first, TechFins are more focused towards data intermediation (Zetzsche, 2017).

[15] Screening is a process used to evaluate innovative product ideas, strategies and marketing trends.

Figure 1.1 gives a comprehensive and detailed overview on which are the companies that populate the FinTech universe. The ecosystem includes alternative forms of finance, blockchain, digital banking, digital identity, insuretechs, payments and remittances and robo-advisory.

Figure 1.1 - The FinTech ecosystem. Source: Business Insider (2020)



Among these macro-areas we can find some important names from the tech world such as Apple (Apple Pay), Google (Android Pay), Samsung (Samsung Pay), and from the financial services field like Goldman Sachs (Marcus), JP Morgan (Chase Pay).

In the last decade many FinTech start-ups also made their way in their respective markets, and here are some virtuous examples. In the payments systems industry the adoption rate of services such as Adyen, Alipay, N26, M-Pesa, PayPal and Venmo grew at impressive rates. In cross-border payments[16] and remittances[17] TransferWise and WorldRemit are competing with the well-established Western Union and MoneyGram for a share of the market. On the credit side Alibaba, LendingClub, Funding Circle, Prosper, SoFi, and Zopa are fighting with traditional banks in the unsecured consumer loan and small and medium enterprises (SME) markets with important results.

[16] Cross-border transactions are made when the payee and the transaction recipient are based in different countries.
[17] Remittance refers to a transfer of money usually made by a foreign worker to an individual in their home country.

1.4 – FinTech and Banks

To have a better idea about how the industry will evolve it's important to understand how FinTechs compare to banks in this set of activities.

The first and most important task of banks is maturity transformation which consists in using short-term funding to offer long-term loans. We have maturity transformation since there is a significantly low probability of observing a bank run[18] due to government intervention (Diamond and Dybvig, 1983). Like banks, FinTech companies can raise and pool together funds from different people, however they don't have the authorization to collect deposits which gives banks a significant advantage in providing liquidity services[19].

The ability of banks to provide liquidity services is strictly connected with their ability to offer payment services. Customer who needs liquidity would be much better off by having the possibility of paying by using their account at the same time. Now it's clear that the unique possibility of collecting deposits makes the funding for banks way cheaper with respect to any competitor. However, FinTech companies can benefit from looser regulatory requirements, lower costs thanks to technology, and better economies of scope[20].

Information processing is also a very important function of the financial sector and it's here where FinTech companies can compete the most. An increasing availability, a cheaper storage and a faster transfer of data are the key dimensions to be leveraged by FinTechs.

Differently from banks, the information used by FinTech companies is based on big data and not on long term relationships[21]. FinTechs are often built on new platforms that are cloud-based and written in modern programming languages which allow to scale up with much lower cost than traditional banking systems.

Farboodi et al. (2019) analyzed the expansion of big data and artificial intelligence (AI) technologies and observed that they can further increase the size of the firms that adopt them. In their study they concluded that technology driven firms have the potential to overtake traditional intermediaries also in terms of capital investment and profitability.

Although FinTech companies have been expanding at impressive rates in the financial markets, their future impact that they will have on banks and financial institutions, as of today is not completely clear. The biggest question to answer is whether and how profoundly

[18] A bank-run is when most of the depositors withdraw their funds at the same point in time, because they fear that the bank may cease to function in the near future.
[19] Liquidity services exist to provide cash or cash equivalents to meet the short-term operating needs of individuals.
[20] FinTech companies usually start by offering one innovative product or service. When their business is mature enough they can benefit from offering a broader range of products.
[21] When offering their products, banks hardly rely on the relationship with their clients. The relationship is a function of how much they trust their clients and is generally built over a long period of time.

FinTechs can replace banks and other incumbent financial institutions, and whether this replacement can trigger a virtuous cycle of healthy competition and efficient markets, or rather lead to disruption and instability in the financial system.

Like FinTechs, banks are well positioned to adopt technological innovations, and reinvent their old practices in new unexplored ways. Moreover, banks can also enjoy significant benefits from network economies and economies of scope that come from the bundling of multiple activities[22] that characterizes universal banks[23]. Even if FinTech companies can increase the competition in financial markets by providing services in a more efficient way than what traditional financial institutions can do, they will probably not completely replace banks in most of their key functions (Navaretti et al, 2017).

Many FinTech functions will probably still be carried out by traditional banks that will partially lose in terms of margins, but still keep the final interface with their customers. Intermediation will be different than today and it will be driven by internet, cloud and big data processing, but banks won't disappear. If some banks will close it won't be among the biggest ones, but mostly between small banks that don't have the resources to keep up with the constant innovation. In the long-term FinTechs will inevitably have to offer multiple services to expand their activities or enter in multiple rounds of merger & acquisition (M&A)[24] with banks. This is something that is already happening with LedingClub's acquisition of Radius Banks[25].

The business model of FinTechs, is hence slowly but gradually converging towards that of banks. This convergence of FinTech and traditional financial services will probably follow a long period of difficult political, commercial and technological challenges. As this happens, it won't be anymore clear whether FinTechs will have a neat competitive advantage over banks, besides for the costs that banks will face to reorganize their businesses (Navaretti et.al., 2017).

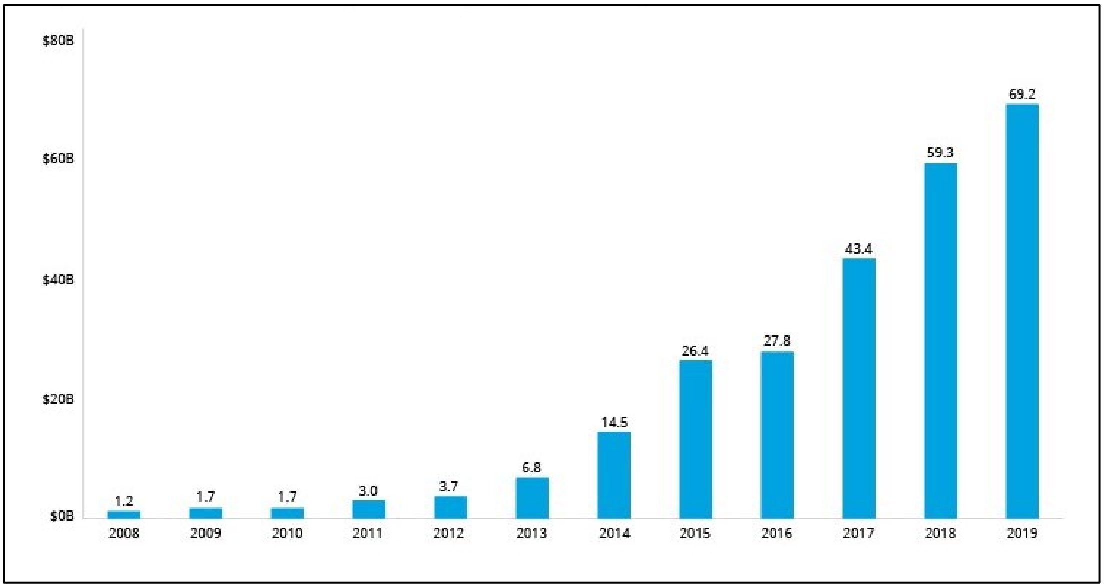
1.5 – The Future of Fintech

A recent study by Deloitte (2020) shows that FinTech start-ups funding has been sharply increasing over the years (Figure 1.2). Starting from $1.2 bn in 2008, the amount of funding grew up to $6.8 bn in 2013, and then it more than doubled in just one year up to a total of $14.5 bn. In 2015 it almost doubled again with a total of $26.4 bn, and in 2017 the cumulative funding for FinTechs passed the threshold of $100 bn.

[22] These activities may include credit, loans, deposits, asset management, investment advisory, payment processing, securities transactions, underwriting, and financial analysis.
[23] A universal bank participates in a wide range of activities and is both a commercial bank and an investment bank.
[24] Merger & Acquisition (M&A) is the consolidation of companies through various types of financial transactions.
[25] See LendingClub acquired Radius Bank in early 2020 to have access to a cheaper source of funding. https://radiusbank.com/lendingclub-announces-acquisition-of-radius-bank/

During the last two years the funding on FinTech grew at a CAGR[26] of 26.3% and this positive trend should continue in the next years since the demand for FinTech services is expected to grow with an astonishing CAGR of 25% to 30% between 2019 and 2025[27].

Figure 1.2 - FinTech funding by year (2008-2019). Source: Deloitte (2020).



In the future FinTech companies with the aid of Artificial Intelligence (AI) and Machine Learning (ML) technologies can disrupt the financial services labor market, which accounts for 6-7% of U.S. employment. As a consequence, with FinTech expansion it's realistic to expect middle-level jobs, such as financial analyst, human resource, loan officer and financial advisor to slowly disappear, and the reason is quite simple: if the job produces data, an artificial intelligence machine can be trained to replace the human on his task.

Although FinTech start-ups have the potential to disrupt and improve the financial system they must be very careful during the initial phase of their expansion. Das (2019) identified seven pitfalls to avoid when implementing FinTech:

1) Garbage-in-garbage-out (GIGO) implies that not all the data is good data. Ultimately, the success of the FinTech model depends on the quality of its input data;
2) Information overload when a firm collects too much data and uses it improperly;
3) Big is not better and having more data is not necessarily an advantage;
4) Correlation does not imply causation and models need a solid theoretical foundation;
5) FinTech infrastructure is expensive so a firm cannot lack commitment[28];

[26] CAGR is the compounded annual growth rate of an investment over a specified period of time. In other words it is the rate of return that would be required for an investment to grow from its initial balance to its ending balance.
[27] See industry trends. https://www.industryarc.com/Report/18381/fintech-market.html
[28] A FinTech company is required be fully committed in terms of capital and motivation since its inception.

6) Lack of trust can reduce the adoption rate of FinTech solutions;
7) Excessive automation can lead to lower customer satisfaction rates[29].

Overcoming these problems will prove essential for FinTechs to keep growing in the future. The player with the most accurate, detailed and valuable data about a customer is best placed to efficiently price financial services for that customer, and beat its competitors.

To sum up successful FinTech companies have the potential to disrupt the financial services industry, however they probably won't substitute the incumbent financial institutions. Banks still own better financial expertise, a well-established infrastructure, a stable customer base, and they can benefit from cheaper costs of funding. Moreover, they are now putting effort to re-specialize themselves and gain back competitiveness (De Young, 2005).

Competition will improve efficiency, bring in new winners, but also strengthen the resilient survivors. Even if intermediation will be carried out differently than today, it will keep being an essential function of financial markets. In a process that will last for about a decade we will most likely see a convergence of Big Tech, FinTech, and traditional financial services towards an integrated financial system based on more internet and more processing of big data information (Pertralia et. al., 2019).

[29] For the best customer satisfaction it's necessary to balance automation with a user-friendly interface.

CHAPTER II – REGTECH AND REGULATORY CHALLENGES

The increasing use of technology in the financial sector has put more pressure on the regulators to change their approach from regulating human behavior to controlling and supervising the algorithmic processes[30]. FinTech applications can be beneficial for the financial system, however they often amplify the burden of regulators when it comes to monitoring the market. Banks claim that if the regulatory framework for alternative lenders won't be held equally restricting for both, FinTech could become a way to access the banking system bypassing the state consumer protection laws.

Ultimately, the rapid expansion of FinTech created the need for "RegTech". The term RegTech was coined by the UK's Financial Conduct Authority (FCA)[1] in 2015 that defined it as "a subset of fintech that focuses on technologies that may facilitate the delivery of regulatory requirements more efficiently and effectively than existing capabilities". Arner et al. (2017) argued that defining RegTech as a subset of FinTech represents a narrow perspective of what it represents today as the next logical evolution of financial services regulation.

The range of application is incredibly wide and includes but not limits to financial services, regulation, social responsibility, healthcare, and cybersecurity. From a market perspective, FinTech growth was mainly driven by start-ups and IT companies, while RegTech grew as a response to the demand from financial institutions and regulators for lower compliance and regulatory costs and increasing market monitoring capabilities, respectively. In the last section of this chapter we will see how RegTech solutions can help financial institutions to comply with relevant laws and efficiently manage their risk, and regulators in monitoring their operate.

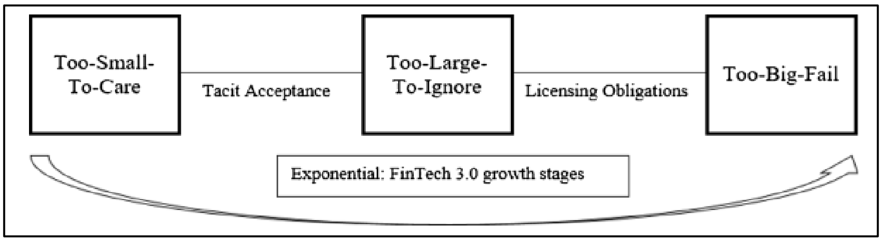
2.1 – FinTech: a regulatory challenge

The crisis of 2008 exposed significant failures in both regulation and supervision of the financial system, and made the financial market law and compliance a key topic for regulators. The financial gap and loss of credibility of financial institutions in the post-crisis led to regulatory reforms such as the Dodd Frank Act[31] and Basel III[32] which drastically increased the compliance obligation for financial institutions.

[30] Algorithm processes are mathematical processes which are commonly used to solve complex problems.
[31] The Dodd Frank Act was introduced during Obama legislation in 2010 as a response to the financial crisis of 2008. The law placed strict regulations on lenders and banks to prevent another economic recession.
[32] Basel III is a 2009 international regulatory framework on bank capital adequacy, stress testing, and market liquidity risk designed to promote stability in the international financial system.

To obtain a stable and efficient financial system, regulator have to face multiple challenges in different directions. The first challenges for regulators are to clearly identify which areas of the law should be in charge of supervising what instrument or activity, and understand when a FinTech business is becoming big enough to represent a risk for the financial system. The proliferation of small FinTechs will be a huge challenge for regulators, simply because their size, innovativeness, and numbers make them hard to find and supervise. There can be various thresholds approaches (Figure 1.3) that can be used to asses when a FinTech firm should receive a stricter regulation.

Figure 1.3 - Regulatory threshold approaches. Source: Arner et al. 2015



However, the problem that comes while trying to regulate non-traditional financial institutions is that they can move very quickly from a size that is too-small-to-care to a size which is too-big-to-fail, forcing the regulator to skip the intermediate step of too-big-to-ignore (Arner et al., 2015). If we consider the biggest players in money market funds (MMF)[33] such as Vanguard, Fidelity, and Schwab they were all established more than 40 years ago. When Alibaba launched its MMF platform Yu'E Bao it took just nine months to become the world's largest MMF. The company grew from a small number of accounts to a platform managing over $165 bn which surpassed even JP Morgan's $150 bn money market fund.

This is a clear example of how the exponential growth of FinTech platforms can create big response-timing and monitoring issues. If the entities that incorporate a higher systemic risk aren't properly stress-tested[34], the investors and the financial system may become particularly vulnerable during bad economic times. Regulators usually act in a sequential approach which consists in observing and monitoring the firms, gathering data about them, supervising and asking for compliance, and finally taking the necessary measures. Today these steps require a lengthy procedure, which makes it harder for regulators to react in time. According to Iosco (2017) FinTech providers of funds and liquidity (i.e. lenders) should receive immediate

[33] A money market fund is a mutual fund that invests only in highly liquid assets such as cash or cash equivalents.
[34] A stress test is a simulation designed to determine the resilience of a given financial instrument or institution to possible bad future market scenarios.

regulatory response as they can offer investors direct access to new and eventually riskier types of investments. Another difficult challenge will be finding a regulatory framework that promotes innovation, but at the same time is rigorous enough to maintain market confidence[35]. Anagnostopoulos, (2018) argued that after identifying which entities should be in charge of supervising, a regulator has to establish whether the current regulatory framework should be incrementally adapted to account for different FinTech realities, or radically reformed by introducing ad hoc regimes or/and exemptions for FinTechs. This is a crucial point, especially because there is still a lot of uncertainty about how the regulators around the world will respond to this latest wave of innovation in finance. The key point is finding the right trade-off between competition and stability, but unfortunately to reach a good balance more experimentation and innovation on the regulatory side is still needed. Regulators will often struggle to apply and update the existing regulatory framework to new developments. Barefoot (2015) stated that the experimentation phase will bring high levels of uncertainty and inconsistency in what will most likely become a permanent state, as soon as one new challenge are addressed, other will arise.

2.2 – FinTech regulatory framework in United States

As of today, there are no domestically or internationally agreed regulatory standards or even guidelines for FinTech intermediation. This means that each country or group of countries have very different regulatory frameworks to respond to the new challenges imposed by innovative financial services providers. Even if it's probably too early for an international harmonization, countries like United States are putting effort in creating a more collaborative environment.

In United States FinTech activities are not subject to any specific regulatory framework, but they fit in the existing financial regulation which is conducted by agencies at state or federal level. The federal government actively regulates the largest part of financial products and services[36], while individual states may establish their own statutes and regulations.

Interestingly, the Office of the Comptroller of the Currency (OCC), which regulates national banks, announced in 2018 that it will start to accept applications from financial technology companies that receive deposits, pay cheques or lend money[37]. This is part of an effort to

[35] Supervisors should consider whether these frameworks are adaptive enough to appropriately balance ensuring safety and soundness and consumer protection expectations with mitigating the risk of inadvertently raising barriers to entry for new firms or new business models (BIS, 2017).
[36] At a federal level FinTech companies may fit in the regulation of: Consumer Financial Protection Bureau (CFPB) which can enforce against the use of unfair, deceptive or abusive acts and practices (UDAAP); Federal Reserve Board of Governors which covers bank holding companies and the processing of payments; Federal Deposit Insurance Corporation (FDIC) for deposit insurance; Federal Housing Authority which covering residential mortgage loans; Security Exchange Commission (SEC) for the investment in securities; Commodity Futures Trading Commission (CFTC) for trading in commodities, and Office of the Comptroller of the Currency (OCC).
[37] See OCC Begins Accepting National Bank Charter Applications From Financial Technology Companies. https://www.occ.gov/news-issuances/news-releases/2018/nr-occ-2018-74.html

provide more choices to customers and to create greater opportunity for companies that want to offer banking services in America.

According to Tank et al. 2019 depending on its activities a FinTech company may be subjected to an extensive number of federal[38] and state[39] licensing or registration requirements. FinTechs that want to offer their services in different US states will require licensing or registration with each of them. Based on the number of states in which a FinTech firm wants to operate, it may hence become particularly burdensome to be compliant with all the different sets of rules[40]. US FinTech firms in their earlier stage often prefer to be compliant only with a limited number of states in the effort to circumvent part of the stricter federal regulation.

As US regulators are developing the regulatory framework that will govern the future FinTech space there is a high degree of uncertainty about how it will evolve, and what will be the support and collaboration from the government. With the evolution of the FinTech space, US regulators are trying to gain a deeper understanding of the dynamics that drive the FinTech industry by collaborating with its market participants. The efforts made led to the creation of innovation hubs, direct channels through which they can communicate and share their concerns with FinTech companies and vice versa. Some examples of the touchpoints between regulators and FinTech firms are:

- The CFPB's Project Catalyst[41],
- The OCC's Office of Innovation[42],
- The CFCT's LabCFTC[43].

The regulatory framework in US is yet very extensive and fragmented, and doesn't adequately establish regulatory priorities or address consumer needs. Although the regulator is trying to put more effort in connecting with these new innovative businesses, the road ahead is still long.

[38] For example federal laws include the Electronic Fund Transfer Act, the Equal Credit Opportunity Act, the Truth-in-Savings Act, the Truth-in-Lending Act, and the Securities and Exchange Act of 1934.
[39] State level regulation may include instead consumer lending, money transmission, and virtual currency licensing. The Financial Regulating Authority (FINRA)[39] also introduced the Know Your Customer (KYC) rule which requires a broker to know the essential facts regarding every customer.
[40] Among the most important entities at state level we may find: state banking agencies responsible for licensing and supervising state-chartered banks; consumer protection agencies which takes complaints about businesses that don't make good on their promises; secretaries of state appointed by the President; state securities regulators regarding the sale of securities.
[41] In 2017 it was announced a collaboration between Project Catalyst and Credit Karma, a personal finance company, to examine how engagement with educational information relates to consumers' sense of financial security and financial freedom.
[42] The office of innovation aims to implement a framework that supports responsible innovation, and enhances the safety and soundness of the federal banking system, treats customers fairly, and promotes financial inclusion.
[43] The LabCFTC seeks to promote responsible FinTech innovation and accelerate the engagement of CFCT with FinTech and RegTech driven solutions.

2.3 – The evolution of financial services regulation

As technology-driven business models are quickly gaining ground over the ones of incumbent banks, questions arise about how traditional policymaking models can deal with oversight and regulation. For example the high degree of uncertainty of the US jurisdiction may allow some firms to grow off the radars, and become a risk for the financial stability. Moreover, activities like payment services and alternative lending aren't fully covered by the regulatory authorities which should ask market participants higher transparency. If regulators won't be able to fill the legislative vacuum many firms will be able to circumvent the rules, posing serious threats to the stability of the financial system. De Roure et. al. (2018) concluded that alternative lenders can benefit from regulatory shocks that create a competitive disadvantage for banks.

For regulators it will undoubtedly be a very challenging time, however the application of technology for compliance and regulatory purposes is now playing a crucial role in assisting the decision making process of the regulatory authorities. According to Fernández De Lis (2016) RegTech provides the means to move towards a risk-based approach where that leverages big data to enable a more granular, real-time and effective differentiated supervision of markets and their participants.

Silverberg et al. (2016) identified several issues in compliance and regulatory reporting that could significantly benefit from the deployment of RegTech solutions:

- Risk data aggregation which is required for capital and liquidity reporting;
- Scenario analysis and forecasting required for stress testing and risk management;
- Monitor payment transactions to recognize money laundering and terrorism financing;
- Identification of clients as required by know-your-customers (KYC) regulations;
- Monitoring financial institutions' behavior to protect customers;
- Automating the control of financial markets;
- Identifying new regulations to apply to financial institutions.

RegTech improves the efficiency of the financial system by reducing the regulatory requirements of capital, and decreasing the time it takes to investigate firms. Moreover, it offers regulators the possibility to continuously monitor firms by providing real-time insights through the use of artificial intelligence, and helps them to identify critical problems in advance instead of waiting for the facts to happen before taking enforcement action. While consumer privacy should be protected by laws and regulations, supervisors would particularly benefit from a higher and timely flow of data and improve their methods and processes. For example

Accornero and Moscatelli, 2017 argued that the big data coming from social media can be used as an effective tool to predict inflows and outflows of retail deposits when supervisors fear a potential bank run. Other examples of scientific innovations that could be applied as RegTech include cryptography for a safer and faster data sharing within institutions, biometrics to improve efficiency and security in client identification, application programming interfaces (APIs)[44] for an automated reporting of data. Thanks to technology and big data the future financial environment should see a modern and collaborative supervision system which is built on inclusive processes involving start-ups, incumbent financial institutions, regulators, and the public. For now Fintech companies and banks can retain their competitive advantages in their respective activities acknowledging that their future concerns are likely to converge.

[44] An application program interface (API) is a set of protocols which is used to build software applications and specifies how software components should interact with each other.

CHAPTER III – THE ROLE OF ALTERNATIVE LENDING: LENDINGCLUB'S CASE

The consumer credit market is one of the largest and most important credit markets, which amounts to about $4 trillion only in the United States (FED, 2019). Despite its impressive size, the this market is characterized by multiple imperfections, including excessively high rates on credit cards (Stango and Zinman, 2009) and recurrent rejections of credit applications. Bricker et al. (2017) found that a significant percent of the families surveyed felt credit constrained. Information asymmetries are also an important and very common issue in the credit market. Jaffee and Russell (1976) and Stiglitz and Weiss (1981) discovered how information asymmetries between borrowers and lenders can lead to a market equilibrium in which credit is rationed[45].

An increasing number of FinTech companies have now entered the consumer credit market, possibly because they can overcome some of these frictions and exploit profitable opportunities. For example, Frame et. al. (2001) and Einav et. al. (2013) found in their studies that the use of technology like credit scoring reduce the information asymmetry while expanding access to credit. Technological advances have facilitated the collecting and scoring of credit qualifications for prospective borrowers on online platforms, the real-time reporting of lending bids, and the online monitoring and reporting of loan performance (Morse, 2015). Philippon (2019) argued that FinTech is likely to decrease the financial intermediation costs, but also to create new regulatory issues. In his analysis on the credit market, he found that minorities that were hurt by prejudice or negative stereotyping could gain from the use of alternate data sources. On the other hand more new data can reduce the effectiveness of existing regulations.

The most common lending model among FinTechs is represented by small balance, unsecured, and fully amortizing[46] consumer loans financed by a large pool of retail investors. It is noteworthy to mention that this is not something completely new from the product perspective, because FinTech loans are similar to the personal installment loans that are offered by banks. The innovation lies instead in the lending process which allows borrowers to request loans in online marketplaces, and lenders to invest directly in consumer loans deciding whether and how much to invest. Although the alternative lending phenomenon started with retail

[45] Credit rationing is an example of market imperfection which implies that the lenders are unwilling to offer credit to borrowers at any market interest rate leaving the latter unserved.

[46] A fully amortizing loan refers to a loan with a periodic repayment according to a pre-defined schedule. When the borrower makes all the payments according to the loan's amortization schedule, the debt will be fully paid off.

investors[47], FinTech platforms rapidly attracted the interest of institutional investors. FinTech lenders are now well-established and gained the respect of the entire financial services community. TransUnion, an American consumer credit reporting agency, estimated that FinTech lending accounted for more than one third of the personal loan market in 2018[48].

3.1 – The role of alternative lending

Alternative lending, also referred as marketplace lending or peer-to-peer (P2P) lending, is a type of lending performed through online platforms that uses technology to offer a wide range of loan options to consumers and business owners outside traditional bank loans[49].

Peer-to-peer lenders bring together borrowers which were underserved by traditional lending institutions with loan investors that seek for more attractive returns. It is not rare that FinTech platforms offer up to 10% to savers which are looking for better returns than the extremely low ones that we observed in the market lately[50]. For example LendingClub offered to its investors an average return across of all the loans of about 13% between 2015 and 2020.

Started in February 2006, Prosper is the first lending platform that was launched in the U.S, and since then FinTech lending has grown rapidly, especially after 2013. The global peer-to-peer (P2P) lending market size, which was valued at $67.93 billion in 2019, is now expected to reach $558.91 billion by 2027, implying a compounded annual growth rate (CAGR) of 29.7% from 2020 to 2027[51]. Currently, the largest segment of the U.S. P2P loan market is the personal consumer loans, with over $48 billion in loans originated to 3.2 million borrowers between 2006 and 2018.

P2P platforms can set their interest rates through the posted price or reverse auctions models (Chishti, 2016). Based on the rating assigned to a borrower, with the posted price model a lending platform decides itself the interest rate to charge, while in the reverse auction the interest rate will be decided by the demand and supply. The major players in P2P lending markets, including LendingClub, set their interest rates with the posted price process, however there are still conflicting opinions on which of the two methods gives the most desirable outcome. Chen et al. (2014) concluded that unlike posted price, the reverse auction process fails

---

[47] To invest in LendingClub retail investors must be at least 18 years old, have a valid social security number, and have their identity successfully verified by LendingClub. Moreover, there is a required initial deposit of at least $1,000.

[48] FinTech continue to drive personal loans levels see https://newsroom.transunion.com/fintechs-continue-to-drive-personal-loans-to-record-levels/

[49] For the definition see https://www.morganstanley.com/im/en-us/financial-advisor/insights/investment-insights/an-introduction-to-alternative-lending.html

[50] For the investor returns see https://www.thisismoney.co.uk/money/diyinvesting/article-3062370/The-investment-trusts-backing-lending-platforms-10-return.html

[51] Peer-to-peer lending market size see https://www.alliedmarketresearch.com/peer-to-peer-lending-market#:~:text=The%20global%20peer%20to%20peer,29.7%25%20from%202020%20to%202027.

to reach the cheapest outcome for the borrower. However, they were contradicted by Wei and Lin (2015) which argued that auctions can allocate the resources in a more efficient way.

P2P lenders are often more flexible than banks in terms of repayment schedules and loan approvals, and also much faster than their traditional counterparts when providing the funds. One of the most attractive features of borrowing from alternative lenders is in fact how quickly lending decisions are made and processed by their automated systems. However, a significant concern about automated decision making is that it may create an environment which systematizes and conceals discrimination.

By analyzing the loan rates in the Swiss P2P market Cashare, the most relevant Swiss P2P platform, Dietricha and Wernli (2019) discovered some indication of discrimination by the lenders. A first indicator for discrimination is based on nationality, Swiss passport holders are in fact charged significantly lower interest rates than foreigners living in Switzerland even if the debt-to-income level for Swiss is on average higher. Another discrimination discovered by the authors comes from the gender and the presence of children living in the same household.

According to Kroll et al. (2017) correct implementation of the FinTech technology is very important to avoid false or discriminatory predictions based on race or other profiling which could result in denial of credit or higher rates for certain demographics.

Morse (2015) questioned herself about the possibility for peer-to-peer lending to effectively disintermediate and mitigate information frictions in lending such that choices and outcomes for at least some borrowers and investors are efficient. The use of internet platforms, allows P2P lenders to significantly reduce their operational costs by eliminating many of the operational expenses associated with traditional banks loans, including the cost of maintaining and staffing physical branches. It would be interesting to understand whether these cost savings are then effectively passed along to borrowers in terms of lower interest rates than the ones offered by traditional lenders.

By analyzing Auxmoney, the largest P2P loan provider in Germany, De Roure et. al. (2016) discovered that the risk-adjusted interest rate charged by the FinTech lender were in line with the interest rates charged by banks for one- to five-year loans. However, they also discovered that Auxmoney was lending relatively more to segment of borrowers that banks were unwilling or unable to supply.

One of the most important advantages to alternative lenders is that they can leverage larger amounts of traditional and non-traditional data about their customers. By increasing the amount of information on the borrowers, P2P lenders are in a good position to have great insights into the credit market conditions. The additional sources of information can include payment history, medical and insurance claims, social network data, and other different factors that are not fully reflected in the traditional credit scores. Gambacorta et al. (2019) discovered that standard credit scoring together with machine learning on other alternative data improves default prediction compared with traditional credit scoring models.

3.1.1 – Peer-to-peer (P2P) lending model

The P2P lending process is often built around four distinct entities including borrower, FinTech platform, investors, and originating bank[52], and it can be summarized as follows:

1) Before a loan is posted on a P2P platform, a borrower files the application for the loan;
2) The FinTech lender verifies the information by obtaining a credit report on the applicant and uses this and other information, to assign a risk grade and set an interest rate;
3) If accepted, a loan application is posted on the platform's website, where investors can review all the loans and commit to fund the ones with their desired risk/return;
4) When a loan is funded the partner bank originates the actual loan and becomes a creditor to the borrower;
5) The investors funds the loan by giving cash to the FinTech platform which purchases the loan note from the bank with the same money;
6) The loan note is transferred first to the lending platform and then to the investors;
7) The borrower repay its debts to the FinTech platform on due dates;
8) Since the loan note is unsecured the investors will receive their money back from the platform only if the borrower repays the loan.
9) In the meanwhile the lending platform receives a fee for originating the loan.

Differently from traditional lenders, which pool together the money of different depositors to allocate it into different loans, P2P business model allows the lenders to directly invest in each loan. Another important difference is that banks often perform other functions like risk sharing and maturity transformation in which P2P lenders are not typically involved. Although the elimination of physical branches can undoubtedly generate important cost savings, a huge advantage to FinTech lenders is that, unlike banks, they don't not carry the loans on their books which implies that they are not subject to stringent capital requirements.

The role of FinTech lenders is also different when it comes to monitor the borrowers: while banks are more focused on monitoring after they provide the loan, FinTech lenders put more effort on the screening phase before the loan origination. According to Balyuk (2019) an important feature of P2P lending is in fact the use of automated algorithms in the lending

[52] The originating bank is the entity who materially creates the loan.

process for screening, verification, funding and repayment purposes. FinTech lenders are hence in a strategic position to offer small loans which banks may be find unattractive in terms of profitability.

The innovation on the lending procedure regards the way in which information is used and the impact they have on the screening quality. Who borrows from traditional lenders is screened only once by the lender itself, while P2P borrowers are screened first by the FinTech lender and then by the investors that can decide where to allocate their money based on their preferences.

According to Morrison & Foerster (2018) peer-to-peer lending can offer a series of advantages to both borrowers and investors. Borrowers can benefit from lower average interest rates than those charged by the traditional banks on credit cards or installment loans[53], an easy to use and convenient online platform, uniform and clearly disclosed loan terms, and efficient decision-making in assigning risk grades to loan applicants. The investors' advantages include instead potentially higher risk-adjusted returns, access to high yield investment which are typically reserved to institutional investors, transparency and autonomy in deciding in which loans to invest in, and comprehensive access to the credit profile of each borrower who applied for a loan. For example LendingClub's current average adjusted NAR[54] for the safest loan notes (A) is 4.77% which in a period of extremely low treasury rates could be a good incentive for retail investors that have a more limited access to high yields debt than institutions.

3.2 – LendingClub's Case

LendingClub (LC) was founded in 2006 by Renaud Laplanche, a well-respected figure in both the technology and financial industries, and quickly become one of the most prominent U.S. start-ups to take on banks thanks to the use of innovative technology.

LC is currently the largest P2P lending platform in the U.S, and provides fixed-interest unsecured personal loans repaid monthly with 3- or 5-year maturities.

The loan amount goes from a minimum of $1,000 up to a maximum of $40,000, while interest rates charged in the period of analysis (2016-2017) ranged from 5.32% to 30.99%[55].

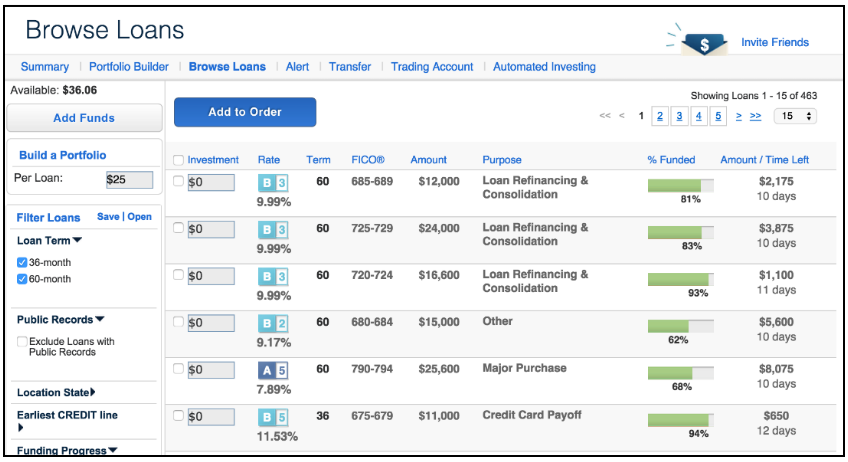
To decide the interest rate to apply to each borrower, LendingClub uses a posted price model that is based on all the information that the company collects before loan origination.

[53] Installment loans is a broad term to define to the vast majority of personal and commercial loans that involve a repayment over time with a set number of scheduled payments.
[54] The adjusted net annualized return is an annualized measure of the rate of return on the principal invested over the life of an investment.
[55] LendingClub, like other consumer lenders, used to make loans at rates that are above state interest-rate caps through arrangements with banks. But the practice was called into question by a 2016 federal appeals court ruling that could subject such lenders to state interest-rate limits. From 2016 LendingClub gave up an undisclosed amount of revenue to avoid potentially being blocked from making loans by state usury laws. Source: WSJ (2016).

LendingClub doesn't subject the borrowers that pre-pay their full loan amounts to any kind of early payment fees or penalties, however it charges them with a origination and servicing fees[56] for facilitating the origination of their loan. The origination fees are the main source of revenue for the company which occasionally retain the loan notes for itself. The focus on LendingClub is explained by the fact that the company publicly releases all the loan-level data generated each quarter through its platform, making it an attractive candidate to conduct research about.

On its platform LendingClub gives the investors two possibilities: accepting a portfolio that is automatically generated by the platform or selecting the loan notes that they want to fund. An investors has hence full control over the decision process and is helped by the company that assesses the loan riskiness on their behalf. This is where the real disintermediation takes place with lenders being able to directly decide where to put their money. By looking at Figure 3.1 we can get an idea of the large amount of flexibility that is offered through the online platform.

Figure 3.1 - LendingClub's investors interface. Source: Lendacademy[57]



On top of being able to select the amount of money to allocate on each loan note (the minimum is $25) an investor have access to all the most important information about the applicants. On the interface it's possible to see how much money an applicant needs and for what purpose, the loan term, the amount already funded by other investors and the current FICO Score of the borrower. Investors can also filter the loans based on multiple characteristics including geographical location, employment length, home ownership, credit records and many more characteristics pertaining to both loans and borrowers. Although there is undoubtedly a lot of choice on the platform the most important information provided by LendingClub is an

[56] Origination fees is the amount of money that LendingClub asks to originate the loan. Service fees are instead charged to investors for the number of borrower's payments received later than the due date.
[57] See https://www.lendacademy.com/lending-club-review/

innovative credit scoring system represented by letter grades which will be describe in the next section along their traditional counterparty, the FICO Scores.

3.2.1 – Grades vs. FICO Scores

To help the investors in the process of capital allocation, LendingClub uses algorithms and machine learning techniques to analyze the data of every applicant and generates for each loan an internal grading from A to G[58]. Grades are an indicator of a borrower's riskiness, where A indicates the lowest level of riskiness and G the highest one. Like a scoring system, grades are designed to provide investors with a comprehensive measure of the risk that they are incurring by putting their money into a specific loan note.

FICO scores are the most traditional indicators of a borrower's creditworthiness, and consist in three-digit numbers that help lenders determine how likely are borrowers to repay their loans.

New alternative credit scorings have emerged in the recent years such as the profit scoring which was studied by Serrano-Cinca and Gutiérrez-Nieto (2016). However, despite the growing number of new alternative, FICO Scores remain one of the biggest names in the marketplace and represents the industry standard. According to FICO, 90% of top lenders use FICO Scores when making lending decisions[59]. The borrower's creditworthiness affects how much the individual can borrow, the length of the loan term, and ultimately its cost (interest rate). Table 3.1 presents the division of FICO scores in different ratings, where a higher rating means higher possibilities for an applicant to get access to credit.

Table 3.1 – FICO Score ratings. Source: Myfico[60]

FICO Score	Rating	Description
<580	Poor	This FICO score is well below the average score of U.S. consumers and demonstrates to lenders that the borrower may be a risk.
580-669	Fair	This FICO score is below the average score of U.S. consumers, though some lenders may approve loans with this score.
670-739	Good	This FICO score is near or slightly above the average of U.S. consumers and most lenders consider this a good score.
740-799	Very Good	This FICO score is above the average of U.S. consumers and demonstrates to lenders that the borrower is very dependable.
800+	Exceptional	This FICO score is well above the average score of U.S. consumers and clearly demonstrates to lenders that the borrower is an exceptionally low risk.

[58] Notably, the company also uses a finer distribution of the grades by dividing each class in 5 (i.e. A1 to G5) for a total of 35 sub-grades.
[59] See ratings: https://www.myfico.com/credit-education/fico-scores-bridge#:~:text=hurt%20your%20credit-,Fact%3A,by%2090%25%20of%20top%20lenders.
[60] See www.myfico.com/credit-education/what-is-a-fico-score

FICO scores are calculated by putting together many different pieces of information from the borrower's credit report. These information are divided into five categories and weighted as follows: payment history (35%), amounts owed (30%), length of credit history (15%), new credit (10%) and credit mix (10%). Although FICO scores, consider a wide range of information from the borrower's credit report, they do not account for different types of data such as: age, salary, occupation, title, employer, employment history, living place, child/family support obligations, and different types of credit inquiries. In order to get a better knowledge about their borrowers FinTech lenders often include these or other types of alternative information in their systems. Additional sources of information can improve the risk assessment procedure and help FinTech lenders to offer a better value proposition than their traditional counterparties. In the final chapter of this study we will analyze whether the additional information collected by LendingClub can actually improve the predictive ability of the future loan performance.

3.2.2 – Past literature

In the past years several empirical studies were conducted to understand the main drivers behind the loan performance across different FinTech lenders. Table 3.2 offers a summary of the most important literature that studied the most important determinants of the default rates observed on the LendingClub's platform.

Table 3.2 – Literature on determinants of borrowers' default. Source: Polena & Regner (2016)

Study	Dataset	Method	Findings
Emekter et al. (2015)	May 2007 - June 2012 (36- & 60-month loans)	Binary logistic regression	Credit Grade, FICO score, Debt-to-Income and Revolving Credit Utilization
Serrano-Cinca et al. (2015)	January 2008 - December 2011 (36-month loans)	Univariate means test and Cox regression	Credit Grade, Annual Income, Loan Purpose, Debt-to-Income, Current Housing Situation, Credit History Length, Revolving Credit Utilization, Recent Inquiries, Delinquency in Past 2 Years, Open Credit Lines
Carmichael (2014)	June 2007 - October 2013 (36-month loans)	Dynamic logistic regression	Credit Grade, Annual Income, Loan Purpose, FICO score, Revolving Credit Utilization, Recent Inquiries, Credit History Length, Time since Last Delinquency, Loan Amount, Loan Description

What is commonly agreed is that grades are the most important predictors. However, it's not a surprise as they theoretically incorporate all the information on the dataset. Other factors that were indicated as important determinants of the borrowers' defaults are revolving line utilization, income and debt-to-income ratio.

The discrepancies observed among all of these studies can be explained by different time frames, loan terms, and methodologies. For example Carmichael (2014) and Serrano-Cinca et al. (2015) conducted the analysis only on the shorter loan terms (36-mths) while Emekter et al. (2015) used all the terms. Moreover, the variables included in the analysis were slightly different with Serrano-Cinca et al. (2015) dropping the FICO Scores from the dataset.

3.2.3 – Regulatory considerations about LendingClub

In U.S. consumer credit, whether it is originated by a bank or not, is subject to an extensive number of federal and state laws, and depending on different factors P2P lenders may fall under the authority of numerous federal and state regulators. Federal and state laws regulate all the aspects of the credit market, and prohibit any form of credit discrimination and unfair, deceptive, or abusive acts or practices (UDAAP).

Some of the most important federal laws to which LendingClub is subject include the Truth in Lending Act which prescribes uniform methods for computing the cost of credit, credit terms disclosure, and errors resolution, the Equal Credit Opportunity Act that prohibits lenders to conduct discriminating practices against credit applicants, and the Fair Credit Reporting Act that requires creditors to accurately report information to reporting agencies. Other laws that regulate the relationships between financial institutions and consumers include privacy and data security and anti-money laundering laws.

Initially P2P loan notes were not considered as securities, however the SEC ruled out that the loan notes are at all effects securities[61]. The reasons of such decision is because lenders are motivated by an expected return on their funds, the loans are offered to the general public, a reasonable investor is likely to expect that the loans are an investment, and there is no alternate regulatory scheme that reduces the risks to investors presented by the P2P lending platform (Morrison & Foerster, 2018). Since LendingClub's loan notes have all the characteristics to be considered as securities, the company had to register with the SEC and be subjected to its regulation along with the consumer credit one.

Because of this complex web of regulations, many FinTech lenders decided to partner with banks as they have access to cheaper funding plus an already existing compliance infrastructure. However, when a FinTech platform is partnering with a bank to originate its loans it would be considered as a service provider with respect to its origination, therefore it may still be subjected to regulatory oversight and examination. In LendingClub's case the Consumer Financial

[61] The SEC applied to the P2P platform Prosper the same analysis that was used in the case of Reves v. Ernst & Young, 494 U.S. 56 (1990)

Protection Bureau (CFPB) has the unfair, deceptive, or abusive acts or practices (UDAAP) enforcement authority over both the originating bank and the platform being involved in P2P lending. Another important regulatory concern with P2P lenders such as LendingClub is that they do not share with investors the potential losses coming from the borrower defaults. As we have seen the investors are paid back only if LendingClub receives the money, and if a borrower misses its payments the investors will also be charged with an additional servicing fee for the recollection. Although FinTech lenders offer very interesting opportunities to investors, this remains a clear advantage to them which may opportunistically behave to reach different type of goals, which is exactly what happened during LendingClub's scandal of 2016.

3.2.4 – The scandal of 2016

In 2016 LendingClub's management forced Laplanche, the CEO of the company at that time, to resign from its position, after an internal investigation discovered that $22 million of subprime sold between March and April of 2016 were conflicting with the investor's expressed terms. According to the S.E.C., a division of the company led by Laplanche, manipulated the way in which the funds were managed without disclosing it to the investors, in order to boost the demand for some of the loans that were sold through the lending platform.

After the scandal came out, the company saw its share price falling dramatically with a massive disruption for the shareholders value. Since then the new management that was put in charge struggled to gain back credibility, and the company had to purchase part of the loan notes that remained unsold. What happened with LendingClub demonstrated that the regulatory authorities in U.S. were not fully prepared to monitor a company that grew so much so quickly, and that more regulatory scrutiny is required in the future to avoid similar situations.

In this scenario it becomes particularly important the role that RegTech is assuming in reducing the information gap between regulators and businesses. For example a real-time monitoring of the loans that are originated through the online platform can prevent situations in which the platform's data is altered by the management, or incorrect as a consequence of reporting errors. In particular having a mechanism that automatically verifies whether the loans that are about to be originated comply with all the relevant regulations and, meet all the investors criteria would have certainly helped in avoiding LendingClub's scandal.

CHAPTER IV – DATASET

Differently from the vast majority of traditional and non-traditional lenders, LendingClub offers the unique opportunity to access and analyze the loan level data collected from its platform. Moreover, the fact that LendingClub is the largest FinTech lender for personal loans, makes most of the research results more likely to apply broadly.

LC collects data on all its applicant before and after loan origination. All these information will be later used to study their impact in the loan pricing mechanism and on the predictive ability of the future loan performance. In particular we are interested in the loans that went delinquent in the two years after origination because we will use to use them to assess the predictive ability of the most important dataset variables.

Our sample of observations includes all the loans that were originated through the platform between 2016 and 2017. The full dataset available on the P2P platform contains a total of 150 variables of different types including continuous, binary or factors. On its inside we can find several information about the loans such as interest rate charged, amount, term, and the grade assigned by the platform, and about the borrowers including employment length and type, annual income, debt-to-income ratio (DTI), home ownership, FICO Score, and many other qualitative and quantitative indicators that the company uses to assess their riskiness.

The total amount of information contained in the dataset is very extensive, however only a smaller portion of it will be relevant to our analysis. An essential part of this research was in fact selecting all the variables that were most suitable of being introduced in the models at a later stage. To this purpose a primary step was performing an initial selection to drop all the uninformative variables, including the ones that were empty[62] or couldn't be introduced into the models. All the variables relative to the joint applications, which are a very small subset of the total applications, were also dropped to focus only on the individual ones.

The initial round of screening, which was in most part discretionary, was followed by a second one meant to identify those variables that would significantly reduce the population without adding any material benefits to the analysis. For this reason the variables that presented a significant amount of missing values (> 20%) were dropped from the dataset.

Another important step was to identify which variables contained the same or a very similar information, because depending on their correlation such variables could lead to a problem of perfect or imperfect multicollinearity[63]. Since the problem of multicollinearity can make the

[62] Empty variables are those one that did not contain any type of information.
[63] We have multicollinearity when two or more independent variables in a multiple regression model are highly linearly related.

models estimates more unstable and difficult to interpret by increasing the variance of the regression coefficients, the variables that presented a high correlation were also dropped from the dataset. After the initial selection, the remaining variables were cleaned[64] and re-adapted to obtain a more uniform dataset, which made possible to implement the models at a later stage.

In addition to collecting the data before loan origination, LendingClub keeps track of the loan performance throughout time. This information was essential to test the ability of LC in predicting the future delinquency rates, therefore it was merged with the initial dataset. In the next part of the chapter we will go through some of the most interesting information collected by LendingClub to better understand its customer base.

4.1 – Loans and borrowers characteristics

In the following section we will analyze some of the main characteristics related to loans and borrowers. The grade and the interest rates are automatically assigned by LendingClub whenever a new application is received, while the other characteristics are the result of the application process.

4.1.1 Data cleaning

Before analyzing the main borrowers' characteristics it was necessary to identify and eliminate all the potential outliers[65] that could bias the statistical estimates at a later stage.

For example in the borrower income there were some entries which weren't in line with the rest of the observations. The range for the annual income reached values of up to $11 million of dollars while the typical borrower income was in reality much lower for most of the individuals. It's hence reasonable to believe that borrowers with a very high income are unlikely to use the platform to obtain small personal loans, therefore extreme income values are probably the result of reporting errors. To avoid including data that could bias our estimates the distribution of the annual income was trimmed up to the 99^{th} percentile.

The debt-to-income ratio (DTI) was another variable which required to be adjusted. The DTI is a very important measure of the borrowers' ability to repay their debts on time, and it is constructed by dividing a borrower's monthly debt payment to his or her monthly gross income. Lenders prefer borrowers that aren't overextended[66] (low DTI) since they want to make sure that they will be able collect the full loan principal plus the interest rate. The debt-to-income

---

[64] Data cleaning is the process of detecting and correcting errors or inaccurate records from a dataset. The cleaning process starts with the identification of incomplete, incorrect or irrelevant information, and ends with the replacement, modification, or deletion of it from the dataset.
[65] Outliers are extreme or abnormal values that lie outside the overall pattern of a distribution of variables.
[66] A borrower is overextended when is carrying and excessive amount of debts with respect of their income

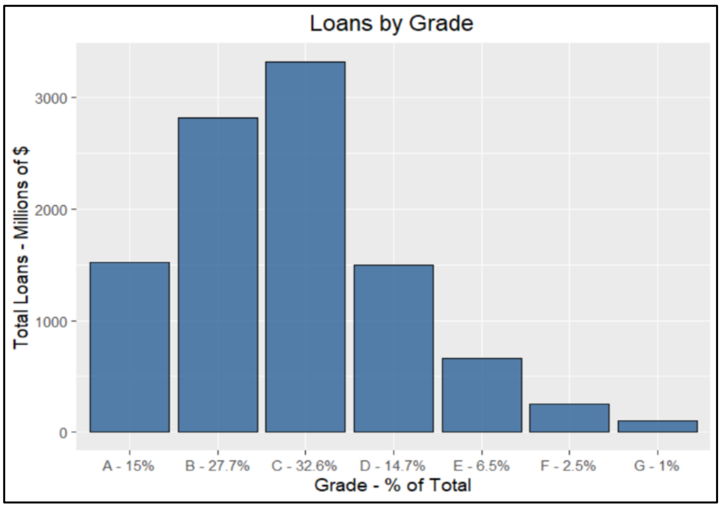
ratio variable presented some anomalies among its observations. The variable included observations that were negative or too high to make any financial sense (i.e. one person cannot have negative or disproportionate amounts debt). To exclude potential outliers in debt-to-income ratio variable was limited in the positive domain and trimmed up to 99th percentile.

Also revolving credit balances, numbers of credit lines and revolving trades presented some extreme values at the right of the distribution that were also trimmed up the 99th percentile.

4.1.2 - Total loans by grade

As we know LendingClub assigns to each of them a different grade that represents a measure of the borrower's riskiness. From the chart on Figure 4.1 we can see that the majority of loan applications had an assigned grade of C (32.6%), B (27.7%) and A (15.0%), and D (14.7%). The lowest grades E (6.5%), F (2.5%) and G (1.0%) instead represented a minority of the total originations in the period 2016-2017.

Figure 4.1 – Loan originations by grade. Sources: LendingClub's data, Author.

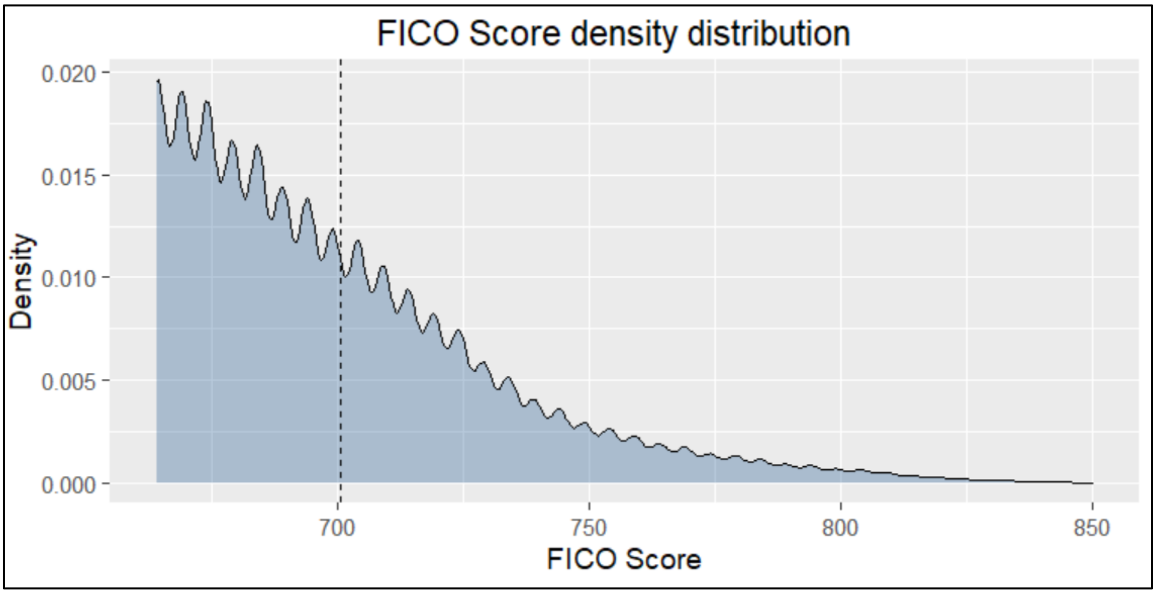


What's interesting about these numbers is that most of the funded loans are not from the safest class (A), but from more remunerative ones such as B or C, which may indicate that the investors effectively came to the company in seek of higher profits than what the market can offer, even if this can translate in higher risks for them. It's important to remind that the loan notes are unsecured which means that when a loan defaults or is charged off the investors will suffer the entire loss.

4.1.3 - FICO Scores

The data on the FICO Scores shows that in the period between 2016 and 2017 the company accepted only borrowers that had a FICO score higher than 660 (Fair rating). To limit a potentially excessive risk LendingClub sets in fact a minimum threshold for FICO Scores below which it doesn't accept any credit application. Figure 4.2 shows the kernel density plot[67] of the FICO Score variable.

Figure 4.2 – Loan originations by grade. Sources: LendingClub's data, Author.



Note: In the vertical axis we have the probability of the density function derived with the kernel density estimation[68], and the black dotted line represents the average FICO Score.

The average borrower had a FICO score of about 700 points (Good Rating). The largest part of the distribution includes borrowers with fair/good (660-739) ratings with fewer individuals having an exceptional rating (>800). While borrowing from LendingClub may be a necessity for some applicants who do not have access to traditional forms of lending, we cannot say the same for borrowers with a high creditworthiness. The most intuitive reason why a borrower with a high or exceptional FICO score would apply for credit from LendingClub is that they are offered significantly lower rates on their loans. Other reasons may include an easier application process and a faster decision-making procedure.

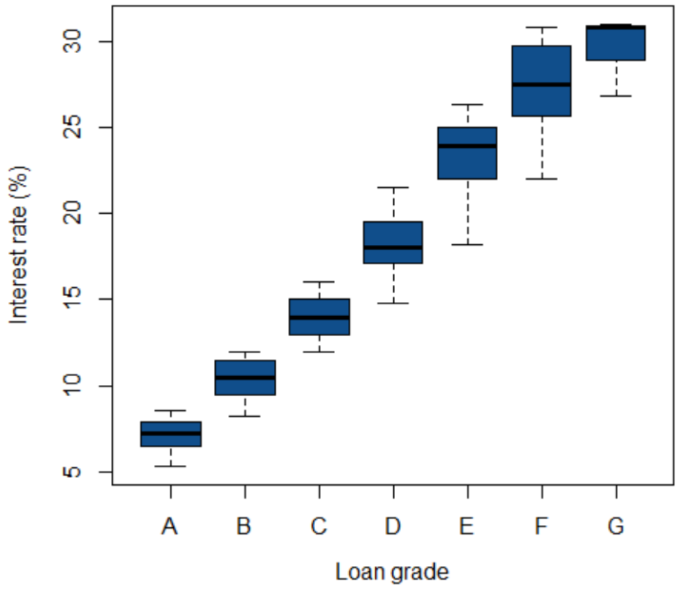
4.1.4 - Interest rates

The interest rates on loans are a function of all the information that LendingClub collects to assess the riskiness of the applicants. The boxplots[69] in Figure 4.3 show the interest rates that

[67] A density plot visualizes visualize the distribution of data over a continuous interval. This is a useful alternative to the histogram for continuous data because it is not affected by the number of bins used.
[68] Kernel density estimators are non-parametric density estimators. The kernel density estimation creates a smoothed a curve of the distribution by weighting the distance of all the points in each specific location along the distribution.
[69] Boxplots are a standardized way of displaying the dataset through five summary statistics minimum, maximum, median, and first and third quartiles.

the company charges for each grade. Higher grades are associated with lower interest rates. The difference between the interest rates charged on the loans with the highest grade and loans with the lowest one can reach up to 2,567 basis points (25.67%).

Figure 4.3 - Loan grades and related interest rate (%). Sources: LendingClub's data, Author.



Note: The top and the bottom of the whiskers indicate max and min, while the boxes show the interquartile range including 1st quartile, median, and 3rd quartile.

A difference that is this large shows how the company lends to very different types of borrowers (i.e. prime to sub-prime), providing the investors with a wide range of opportunities in terms of risk-return. During the period of analysis the average interest among all the loans grades was 13.29%, and as we would expect this value lies in the range between B and C classes, the most frequent ones. The division of loans in small notes (min. $25 each), and the possibility of obtaining higher returns while achieving a high level of diversification are determinant factor in attracting a larger pool of both institutional and retail investors.

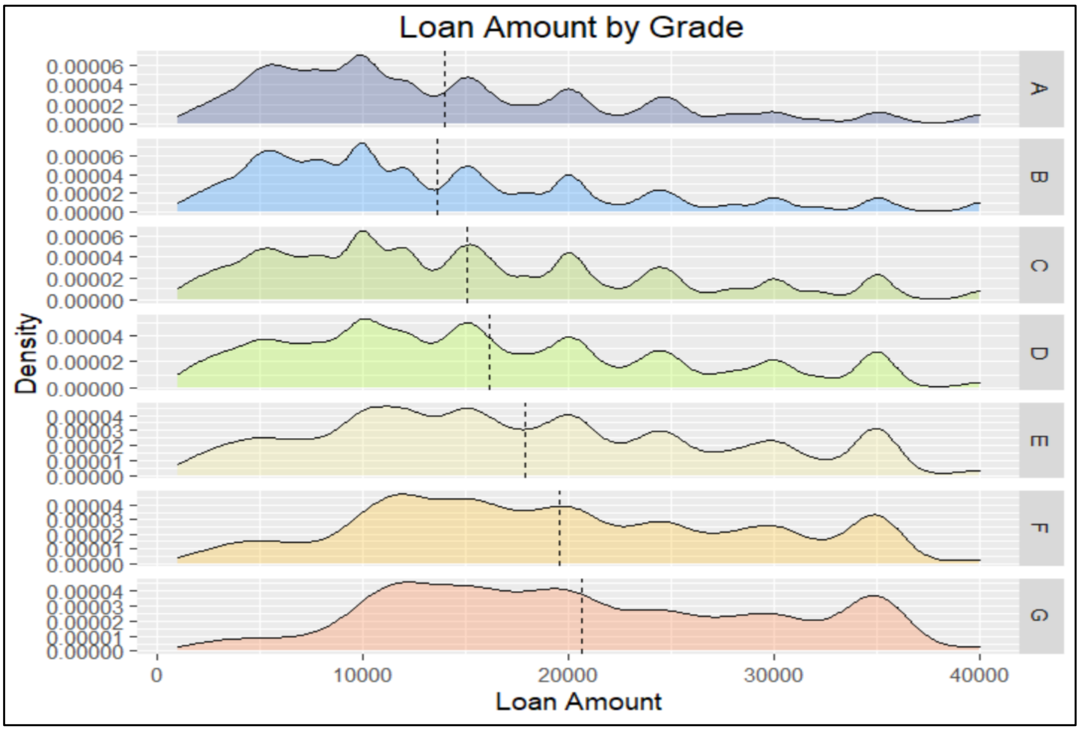
4.1.5 - Loan amount

By breaking down the density function of the loan amount by each grade we can understand how much credit applicants of different classes demanded and received[70]. Figure 4.4 shows the density plot of the loan amount divided by each grade. As we can see the loan amount varied significantly among each class of grades. From the density distribution by grade we have that

[70] Note that all the loans that were approved by LendingClub were fully funded in the period of analysis. This is because each approved loan note is not issued until it's completely funded.

the loan demand by the borrowers progressively increases as we move from the highest to the lowest grade.

Figure 4.4 - Loan amount by borrower's grade. Sources: LendingClub's data, Author.



Note: On the vertical axis we have the probability of the density function, while the dotted lines indicate the average loan amounts for each grade.

The fact that applicants with higher creditworthiness borrowed less on average is most likely a consequence of their better financial situation or social environment that granted them an easier access to traditional sources of credit. Sub-prime borrowers are usually forced to increase their demand for credit to make up for the impossibility of borrowing elsewhere.

For individuals with lower creditworthiness and that have already accumulated a significant amount of debt, FinTech lenders often represent the last resort where they can go to satisfy their credit needs. Lower loan grades (F and G) were associated with larger origination amounts which means that applicants with the worst financial situation on average were able to borrow more that their wealthier counterparties. This finding discovers a relationship that is the opposite of what we would expect under a credit rationing scenario. The same results were observed by Balyuk (2019) by analyzing the loans from the online platform of Prosper. In her study she found that P2P borrowers not only expand their credit access through P2P lending platforms, but banks subsequently increase credit supply to these customers thereafter. The key variable was the quantity of revolving credit (i.e. credit cards, lines of credit) provided by banks. In particular P2P lending was associated with an average increase in revolver limits of $1,020.

Although for different reasons FinTech adoption is not the same all over the world, the results from the American market show the potential to create a more inclusive financial system. Frost (2020) showed that unmet demand (i.e. financial inclusion) together with high costs of finance, and high banking sector mark-ups are a strong driver in emerging countries and in underserved market segments.

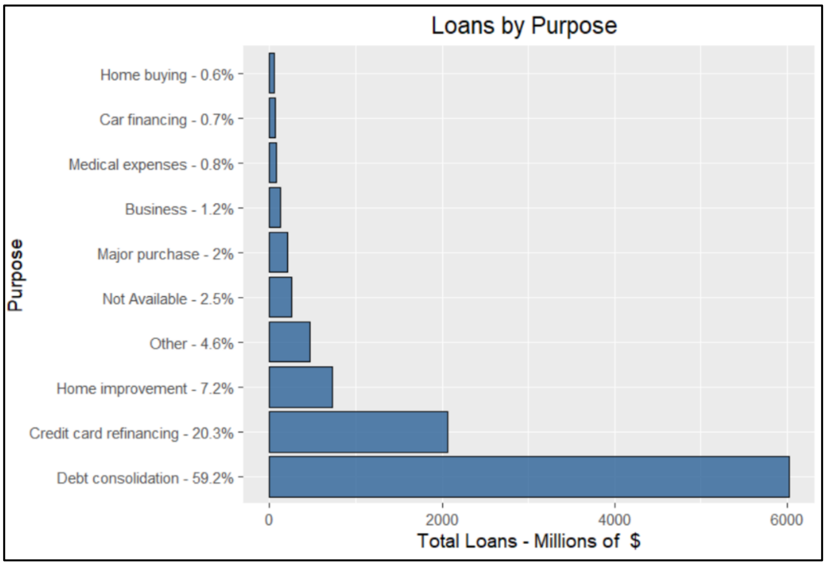
4.1.6 – Loan terms

Another interesting finding from the data analysis is that the most frequent loan term is 36 months (61.6%) with 60 months being slightly less common (38.4%). According to LendingClub loans with 60-month terms have historically delivered higher returns compared to those with 36-month terms, however investing in longer term notes adds both risk and return exposure[71] which could explain the preference for shorter terms. Different maturities allow the investors to diversify their exposures even further achieving higher risk-adjusted returns.

4.1.7 – Total loans by purpose

Figure 4.5 presents the amount of loans divided by the purpose that led to their origination. Among the most common purposes declared by credit applicants on the platform we can find debt consolidation[72] (58.6%), and Credit card refinancing (20.6%). Less frequent purposes include home improvement (7.2%), major purchase (2.0%) and business activities (1.2%). Minor reasons include medical expenses (0.9%), car financing (0.7%) and house buying (0.6%).

Figure 4.5 - Total loans by purpose. Sources: LendingClub's data, Author.



[71] For the investing performance see https://www.lendingclub.com/investing/investment-performance
[72] Debt consolidation generally refers to the process of combining multiple debts from credit cards, high-interest loans, and other bills into a single monthly payment.

The majority of the applicants hence borrowed from LendingClub to consolidate or refinance the debt that they owed to other institutions. There are many factors that can explain why borrowers apply to LC to satisfy these specific needs. Usually the debt consolidation and credit card refinancing loans offered by traditional institutions are secured by a collateral that a borrower has to put up to reduce the credit risk. By offering unsecured personal loans, LendingClub can decrease the burden of a borrower that cannot afford to offer a collateral or take on more debt. Gambacorta et. al. (2020) found that a greater use of credit that was granted on the basis of machine learning and big data could reduce the importance of collateral in credit markets and potentially weaken the financial accelerator[73] mechanism.

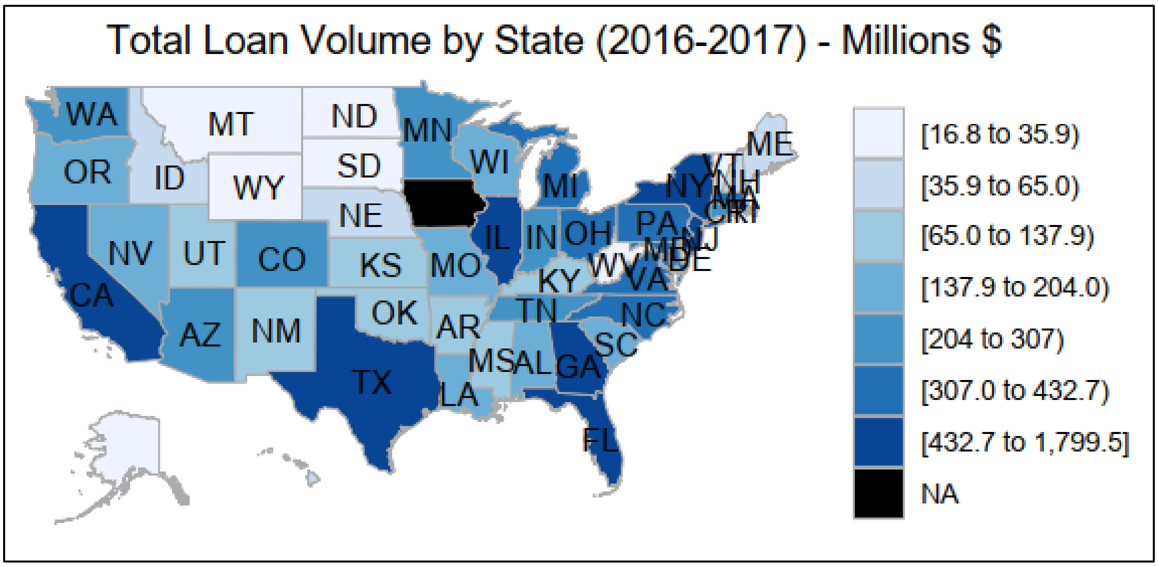
Moreover, the advantages of debt consolidation often translate in lower interest rates that can help the borrowers in saving money with lower monthly payments and a quicker debt repayment. WalletHub (2020) reported that average APR[74] is currently 17.89% for new credit card accounts and 14.52% for existing ones. The average APR for existing LendingClub accounts is instead at 12.74% across all the loans and 11.50% if we consider only 36-months term which explains why LendingClub could be a good opportunity for debt consolidation. However, we cannot say the same for every single loan purpose. For example, Mach, Carter, and Slattery (2014) used LendingClub's loan data to make an explorative analysis on the interest rates of small business loans, and found out that rates on loans with small business purposes are subject to a higher rate even after they controlled for the quality of applicants.

4.1.8 – Total loans by geographical region

By analyzing the geographical distribution of loan originations we can understand in which states the company conducts most of its business. From Figure 4.6 we can see that in the period between 2016 and 2017 among the states with the highest volume of originations we had New York, New Jersey, Florida, California, and Texas. There are multiple factors that can explain the levels of loan activity that we observed including local GDP, unemployment and different state-level regulations which could go in favor or against FinTech lending. States with higher GDP per capita and lower unemployment tend to have a larger presence of the company, and vice versa. The regulatory framework of each state is also an extremely important factor to explain FinTech activity. As we have seen in Chapter 2 every U.S. state has its own set of rules governing the financial activity.

[73] The financial accelerator refers to the process by which adverse shocks to the economy may be amplified by worsening credit-market conditions.
[74] The annual percentage rate (APR) corresponds an effective APR, is the interest rate for a whole year, rather than just a monthly fee/rate, as applied on a loan, mortgage loan, credit card, etc.

Figure 4.6 - Total loans originations by State. Sources: LendingClub's data, Author.



While different states promote FinTech activity as a driver of innovation, other ones are more concerned to guarantee a common regulatory field for both traditional and non-tradition financial institutions. The most extreme example of how the regulatory framework can impact FinTech activity is in Iowa which makes financial intermediation a difficult task for any lender due to a state law that rebates origination fees in the event of loan prepayment. Given that LendingClub mostly profits from the fees earned through loans origination it decided to forego operating in the state.

CHAPTER 5 – EMPIRICAL ANALYSIS

Lenders are constantly trying to find new safe ways to loan money to the growing number of people who demand for credit. Credit scores exist to evaluate the potential risk of lending money to consumers and to mitigate the losses due to bad debt[75]. The companies that design credit scores are in business to help lenders in predicting credit risk, and who creates the best, most predictive scoring system will be able to stand out from the competition.

Although there is a growing number of innovative scoring systems such as the one proposed by LendingClub, the vast majority of traditional lenders base their lending decisions on FICO Scores which are derived from the information collected from the credit bureaus[76]. FICO Scores are designed by the Fair Isaac Company which claims that "scoring solutions are easily understood and provide automated decision-making capability that can be integrated into a company's operations as the basis for taking action". The main problem with FICO Scores is that they ignore important pieces of information that are outside the credit report of an applicant and that could be leveraged to improve the decision-making process of a lender.

The intuition behind alternative lending is that by using a larger set of information on the credit applicants a P2P lender can provide investors with better insights into the habits of consumers than what the sole credit history can do. The introduction of multiple sources of data in the credit risk analysis, can help FinTech lenders in expanding credit access to more individuals improving inclusion and efficiency in the financial system.

Non only LendingClub's scoring system includes an extensive amount of information from the credit reports, but it also collect many information that are not considered FICO Scores such as annual income, employment, geographical provenience, home ownership and many more.

Steve Allocca, the president of LendingClub, stated in 2019 that traditional credit scoring such as FICO Score misrepresent the creditworthiness of individuals. Allocca believes that using credit bureau and historical loan performance data as the primary way for the underwriting and provisioning of credit is increasingly inadequate in a world where financial life is digitized[77]. Instead of relying only on historical data, Allocca said that "two C's" of credit for unsecured lending are capacity[78] and capital, two information that are relatively easy to measure today and provide a better picture of a borrower's creditworthiness.

---

[75] We refer to bad debt as loans that for a variety of reasons cannot be collected from the borrower.

[76] Credit bureaus gather account information from various creditors and sell them for a fee to lenders so they can make a decision on granting loans.

[77] See https://www.pymnts.com/consumer-finance/2020/lendingclub-president-fico-model-wont-save-credit-scoring-dinosaur/

[78] Capacity mainly refers to the individuals' income, and their ability to handle discretionary expenses. A good capacity translates into recurring net savings, which translates into an accumulation of capital

His comment on credit bureaus and traditional scoring systems was: “It becomes obvious that if they don’t move dramatically, the world is going to quickly pass them by”.

Many FinTech companies already took the distance from FICO Scores. For example the online lender SoFi announced in 2016 that the company no longer factors them into its loan-qualification process. Moreover, Kabbage claims that FICO Scores are not anymore part of its creditworthiness determination. A similar position was also taken by Prosper which uses FICO Scores as one data entry out of hundreds on top of using them to screen the applicants. Understanding how LendingClub defines its internal grades is very complicated, given that the company doesn’t disclose what are the exact variables used to generate them. However, it’s acknowledged that the proprietary grades increasingly rely on different additional metrics beyond FICO Scores (Jagtiani and Lemieux, 2017).

An essential task for a lender is finding the best ways to accurately predict the riskiness of its borrowers. Alternative lenders such as LendingClub provide new innovative solutions with the goal of overcoming the limitations of traditional credit scoring.

Our analysis aims to directly compare the performance of LC’s scoring system to the one of FICO Scores to understand if alternative lending really brings significant improvements in assessing the riskiness of borrowers.

In particular we will address the following research questions. We know that FinTech lenders collect an extensive number of datapoints in addition to FICO Scores about their customers, but what are the factors with the highest ability in explaining the loan delinquencies observed on the online platform? Can LendingClub’s grades really outperform the FICO Scores in terms of predictive ability? If yes what is the impact of the most important information that the company collects in explaining the delinquency rates? How much do these factors actually weight in the price decisioning. Not only we are interested in understanding which is the most performant scoring solution among LC’s grades and FICO Scores, but we also want to get an insight about the importance of the additional information in explaining the delinquencies and the pricing decisions that are observed on the platform.

5.1 – Methodologies

In the following sections of the chapter we will introduce all the methodologies that will be used in our analysis. For every methodology we will define its most important theoretical foundations and the way in which we will build our models.

5.1.1 – Random Forest

To identify the most important features will use a Random Forest model because it offers the possibility of ranking them during the classification mechanism. The use of this machine learning technique is justified by the fact that it has one of the highest accuracy among the existing algorithms, it runs efficiently on large datasets, it gives an insight of what are the most important variables in the estimation, and generates an unbiased internal estimate of the prediction error as the forest is built.

Random Forest is a supervised learning algorithm[79] that was first proposed by Ho (1995) by using a random subspace method[80]. The algorithm was then expanded by Breiman (1996, 2000, 2001, 2004) in a series of papers that demonstrated how significant gains in classification accuracy can be achieved by using ensembles of decision trees. Random Forest builds multiple decision trees and puts them together to obtain a more accurate prediction without the risk of overfitting[81] the data. The objective of this machine learning model is to generalize well to new data that the model has never seen before.

The name "Random Forest" derives from the fact that the ensemble is constituted by different decision trees that are grown in a random fashion.

Decision Trees are a type of machine learning classifying technique that were proposed by Breiman et. al. (1984). The decision trees split a dataset into smaller and smaller subsets to draw conclusions about a target value. The tree is built on two parts, namely decision nodes and leaves[82]. At each node the three splits the dataset into 2 buckets, each of them including the observations that are similar among each other and different from the ones in the other bucket.

Breiman (2001) formalized the main concepts of Random Forest and described a method of building forests by using decisional trees combined with specific randomization techniques. We will now will go through some of the most important theoretical foundations of the model.

A random forest is a collection of tree-structured classifiers $h_k(X) = \{h(\mathbf{x}, \Theta_k), k = 1, \ldots\}$ where $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class.

[79] We define as supervised learning the process where computers group data together based on predetermined characteristics.
[80] The random subspace method is an ensemble learning method that seeks to reduce the correlation between estimators by training them on random samples from the full dataset.
[81] Overfitting refers to a model that was trained "too well" so that it learned the noise in the training data to the extent it that reduces the performance of the model when generalizing to new different data.
[82] The nodes indicate the point where the data is split, while the leaves are the final outcome.

Given an ensemble of tree classifiers h1(x), h2(x), . . . , hK(x), with a training[83] set drawn at random from the distribution of the random vector X, Y, which contains all the model's features[84] and the variable of interest, we can define define the margin[85] function as:

$$mg(\boldsymbol{X}, Y) = avg_k \boldsymbol{I}\ (h_k(\boldsymbol{X}) = Y) - \max_{j \neq Y} avg_k \boldsymbol{I}\ (h_k(\boldsymbol{X}) = j)$$

where I (·) is the indicator function[86] and k the number of tree classifiers. The margin function measures the extent to which the average number of votes at **X**, *Y* for the right class (i.e. "Loan delinquent") exceeds the average votes for the other class j (i.e. "Loan not delinquent"). If *mg*(**X**, Y) > 0 the set of classifiers votes for the correct classification. If *mg*(**X**, Y) < 0 the set of classifiers votes for the incorrect one. The generalization or prediction error[87] is given by:

$$PE^* = P_{\boldsymbol{X},Y}\,(mg(\boldsymbol{X}, Y) < 0)$$

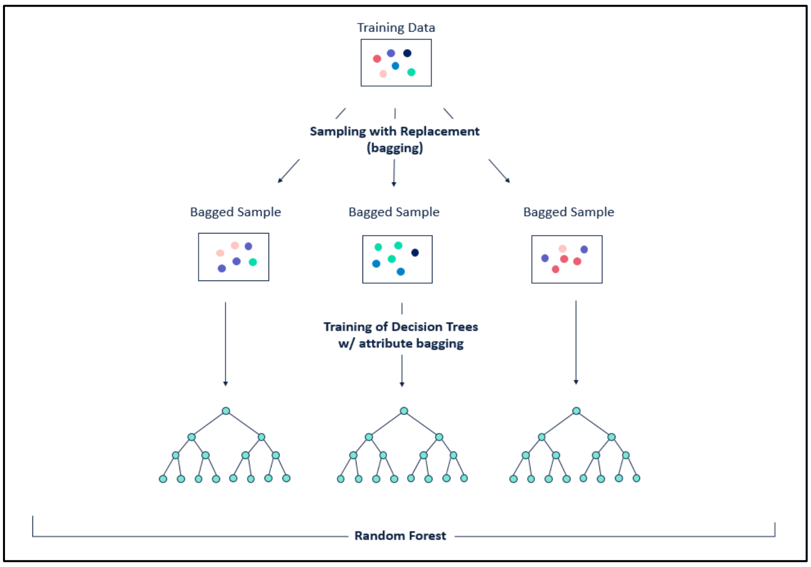
where the subscripts X, Y indicate that the probability is over the X, Y space. From the Law of Large Numbers[88] it can be demonstrated that as the number of trees k grows to infinity the random forest model does not overfit, but produces instead a limiting value of the generalization error. In other words the more trees we add to the forest the better the result, however the marginal improvement in performance will be lower for each tree that we add to the forest.

By expanding the analysis of Amit and Geman (1997), Breiman (2001) also derived an upper bound for the generalization error in terms of two parameters that measure the strength of the individual trees in the forest and the correlation between them. The strength of a tree classifier indicates its accuracy in classifying the outcomes, while the correlation measures the dependence across different trees. Ideally we would like to have the most accurate tree classifiers while keeping the correlation among them as low as possible.

To reduce the correlation among the classifiers ensemble of trees is often generated with the bootstrap aggregating or "bagging" method (Figure 5.1), which is based on the idea that a combination of independent learning models increases the overall result. In other words by growing multiple classifiers and aggregating their results we have a more accurate classifier.

[83] The training set contains a known output on which a model learns to generalize the results to other new data. Note that the training set will be different for each tree classifier.
[84] Feature is a concept closely related to that of explanatory variable used in statistical techniques.
[85] We define the margin of a data point as the proportion of votes for the correct class minus the maximum proportion of votes for the other classes.
[86] An indicator function is a random variable that takes value "1" when the event happens and value "0" when the event doesn't happen.
[87] The generalization error measures how an algorithm is able to predict the values from unseen data.
[88] In statistics the law of large numbers states that as the size of a population sample grows, its mean gets closer to the average of the whole population. $\lim_{n \to \infty} P(|X - \mu| \geq \in) = 0$

Figure 5.1 – Bootstrap aggregation/bagging method. Source: Alteryx (2019).



The bootstrap aggregating method creates several subsets of data that are chosen at random and with replacement[89] from the training sample. This randomization technique guarantees that every tree is trained[90] on a set that is different from each other but comes from the same distribution. The subset of observations that is selected by the bootstrap aggregation is called in-bag set, while what remains out forms the out-of-bag set (OOB). This type of partitioning of the data is meant to ensure that the trees built on the in-bag-set have no bias towards what falls into the OOB set. The classification error can then be tested in the OOB set so that we have a fair estimate of the true error rate of the classifier.

Random Forest – Mean decreasing impurity

The first part of our analysis aims to select the most important features in determining the loan delinquencies. Random forest is particularly useful because it allows us to rank by importance the variables in the dataset through the mean decrease impurity which is internally computed during the model run. The criteria of mean decrease impurity derives from the concept of "Gini Impurity" that we will now define. Assume C the number of total classes and p(i) the probability of picking a datapoint with the class i, we have that the Gini Impurity or Gini Index is:

$$G = \sum_{i=1}^{C} p(i) * (1 - p(i))$$

[89] Each sample unit of the population can be selected multiple times in the same sample.

[90] The process of training an machine learning model involves providing an algorithm with training data from which it can learn from.

In our case the Gini Impurity or Index[91] is a measure of the probability that a particular element (i.e. loan) is incorrectly classified if it is randomly classified according to the distribution of the features in the set. The mean decrease in impurity (MDI) leverages the concept of impurity to measure the importance of a feature in estimating the value of the variable of interest across all of the trees that build the forest. At each node of a decision tree the sample is split based on the feature that decreases the node impurity the most. If for a feature we make a cumulative sum of the decrease in impurity at each node and average it across all the trees in the forest we obtain the mean decrease impurity[92]. A higher mean decrease impurity simply indicates that a variable has a higher importance in classifying a loan as delinquent.

5.1.2 – Logistic Regression

To understand the predictive ability of LC's Grades, FICO Scores and the most important feature we will use a logit model. The logistic regression is a binary response model where the dependent variable of interest (delinquency) can take only two values:

$$\hat{y}_i = \begin{cases} 1 & \text{loan delinquent with probability } \mathrm{p}_i \\ 0 & \text{loan not delinquent with probability } 1 - \mathrm{p}_i \end{cases}$$

With logit we want to assign to parametrize the probability p_i with the function:

$$p_i = Pr(\hat{y}_i = 1|x_i) = F(x_i\boldsymbol{\beta})$$

where x_i is a vector of the most predictive variables, β is a vector of parameters and F(.) the logistic function. Given that $y_i|x_i$ follows a Bernoulli distribution[93], we have that:

$$E\ (\hat{y}_i|x_i) = 1 \times Pr(\hat{y}_i = 1|x_i) + 0 \times Pr(\hat{y}_i = 0|x_i) = Pr(\hat{y}_i = 1|x_i)$$

Logit assigns to the regressors and parameters the function:

$$F(x_i\boldsymbol{\beta}) = \Lambda(x_i\beta) = \frac{\exp(x_i\boldsymbol{\beta})}{1 + \exp(x_i\boldsymbol{\beta})}, \qquad \Lambda(x_i\beta) = \int_{-\infty}^{z} \lambda(v)\, dv$$

[91] Gini Index varies between 0 and 1, where 0 denotes that all elements belong to a single class, and 1 denotes that they are randomly distributed the classes. An index of 0.5 denotes equally distributed elements into some classes.
[92] Note that the decrease in impurity at each node is weighted by the proportion of samples from the initial distribution that reaches that node. For example, if only a limited number samples end up in the left node after the first split, the importance of a feature may be lower because the decrease in impurity on the left node only affects very few samples.
[93] The Bernoulli distribution of a random variable takes the value 1 with prob. *p* and the value 0 with prob. (*1-p*).

where Λ(z) is the probability density function[94] (pdf) of the standard logistic random variable and $\lambda(z) = \Lambda(z)\,[1 - \Lambda(z)]$ is its density function. If the function F(.) is a cumulative distribution[95] we ensure that the estimated response probabilities are strictly between 0 and 1.

To estimate the parameters of a logit model it's necessary to use a maximum likelihood estimation (MLE) which is indispensable for limited dependent variable models (Woolridge, 2003). The maximum likelihood estimation aims to estimate the parameters of a probability distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. Since there is no explicit solution for the parameters to solve the maximization problem the software uses numerical iterative procedures[96].

Logistic Regression – ROC Curves

In our analysis we will use receiver operating characteristics (ROC) curves to visualize the predictive ability of the logistic regression models. With binary response models the predictions are calculated as follows:

$$\hat{y} = \begin{cases} 1 & \text{if } \hat{p} = F\big(\boldsymbol{x}\widehat{\boldsymbol{\beta}}\big) > 0.5 \\ 0 & \text{if } \hat{p} = F\big(\boldsymbol{x}\widehat{\boldsymbol{\beta}}\big) \leq 0.5 \end{cases}$$

The problem with this type of prediction is that when most of the sample has y = 1 it will deliver $\hat{y} = 1$ for all observations, and vice versa. A possible solution to this problem is to consider a cutoff point, c, such that:

$$\hat{y} = \begin{cases} 1 & \text{if } \hat{p} = F\big(\boldsymbol{x}\widehat{\boldsymbol{\beta}}\big) > c \\ 0 & \text{if } \hat{p} = F\big(\boldsymbol{x}\widehat{\boldsymbol{\beta}}\big) \leq c \end{cases}$$

This different type of prediction allows us to build the receiver operating characteristics (ROC) curve. The ROC curve is used to show the diagnostic ability of binary classifiers and plots the true positive rate (TPR) against the false positive rate (FPR). The true positive rate or "Sensitivity" is the fraction of observations that were correctly classified to be positive among all the positive observations, while the false positive rate is the fraction of observations that are incorrectly predicted to be positive out of all negative observations. The false positive rate can be also described as "1 - Specificity" where the specificity is the proportion of negatives that

[94] A probability density function of a continuous random variable provides the probability that the value of the random variable would equal that of the sample.
[95] The cumulative distribution gives the probability that the random variable X is less than or equal to x. The cumulative distribution is defined as: $F(x)=P[X \leq x]$
[96] An iterative method is a mathematical process that starting from an initial guess builds up a sequence of improving approximate solutions.

are correctly classified. The higher the area under the curve (AUC), the better the predictive ability of a model and, in the extreme case of a model that doesn't have any predictive ability (random guessing) the curve simply becomes a straight line.

5.1.3 – Multiple linear regression

Linear models are among the most famous and commonly used statistical techniques. Multiple linear regression (MLR), also known as multiple regression, is a statistical model that uses several explanatory or independent variables (determinants of delinquencies) to predict the outcome of a response or dependent variable (loan prices[97]). In our case we want to analyze whether the most important predictors have a significant impact on the pricing decisions of the company. The linear multiple regression is widely used in the practice because it's easy to implement and allows for a direct interpretation of the statistical estimates.

5.2 – Empirical Results

In this section we will show the results of the different empirical models and discuss their main findings. To understand what are the most important drivers of the future loan performance we will use a Random Forest model and we will present all the predictors in order of importance based on the mean decreasing impurity concept.

Later on we will show the results of different logistic regression specifications and compare the actual ability of LendingClub grades, FICO Scores and the most important features in predicting the future loan performance. To visualize the predictive ability of the features in the most meaningful way we will use the receiving operating characteristics (ROC) curves.

Finally, we will use a linear model to understand if the determinants of the delinquency rates are also important drivers for LendingClub's pricing decisions.

5.2.1 –Features Importance

We will now discuss our classification problem, which aims to identify to which of a set of different characteristics a delinquent loan (30-120 days past due) belongs. In particular we will use a machine learning technique to understand which are the most important features in the classification of a loan as delinquent.

Our Random Forest model included all the variables that remained after the initial screening phase excluding the grades[98], and the variables that were updated after loan origination (i.e.

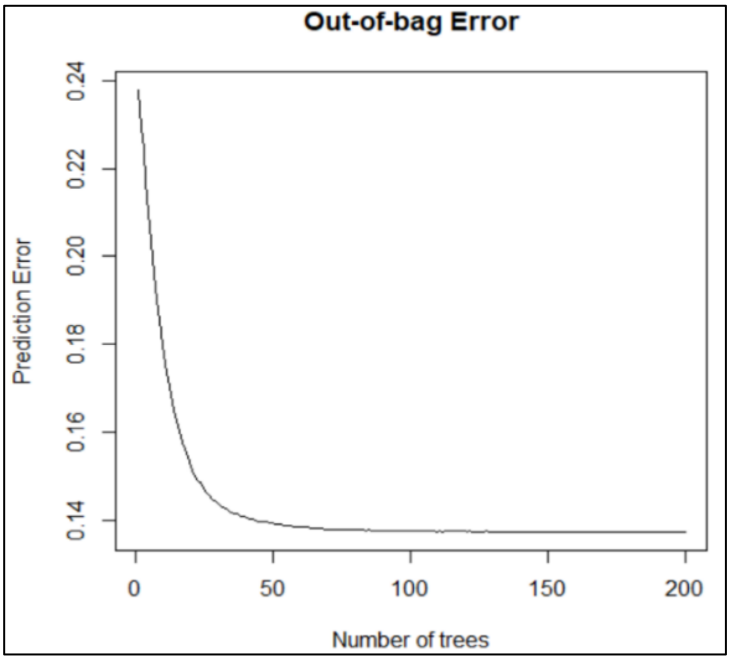
[97] The loan prices are measured in terms of credit spreads which are calculated by subtracting the treasury yields for the same period and maturity to the interest rates charged on the platform.
[98] Grades theoretically account for all the information collected on the dataset.

recent FICO Scores). To achieve a good compromise between accuracy and computational speed the forest was grown out of 200 different decisional trees and from a sample of 200,000 observations that were drawn at random from the initial distribution of 683,448 observations.

Each decisional tree in the forest was constructed by using a different bootstrap sample[99] that was randomly extracted from a total of 200,000 observations, and what was left out of the bootstrap sample became the out-of-bag (OOB) set which we used as a test set for the classification error. Since with Random Forests the classification error is already estimated internally during the model run there was no need for cross-validation[100].

To get an unbiased estimate of the classification error as more trees are added to the forest we will show the out-of-bag (OOB) error (Figure 5.2) which is the prediction error (i.e. loans that are incorrectly classified) that the Random Forest internally estimated from the information that did not enter in the bootstrap sample (i.e. out-of-bag set).

Figure 5.2 – Random Forest out-of-bag error.



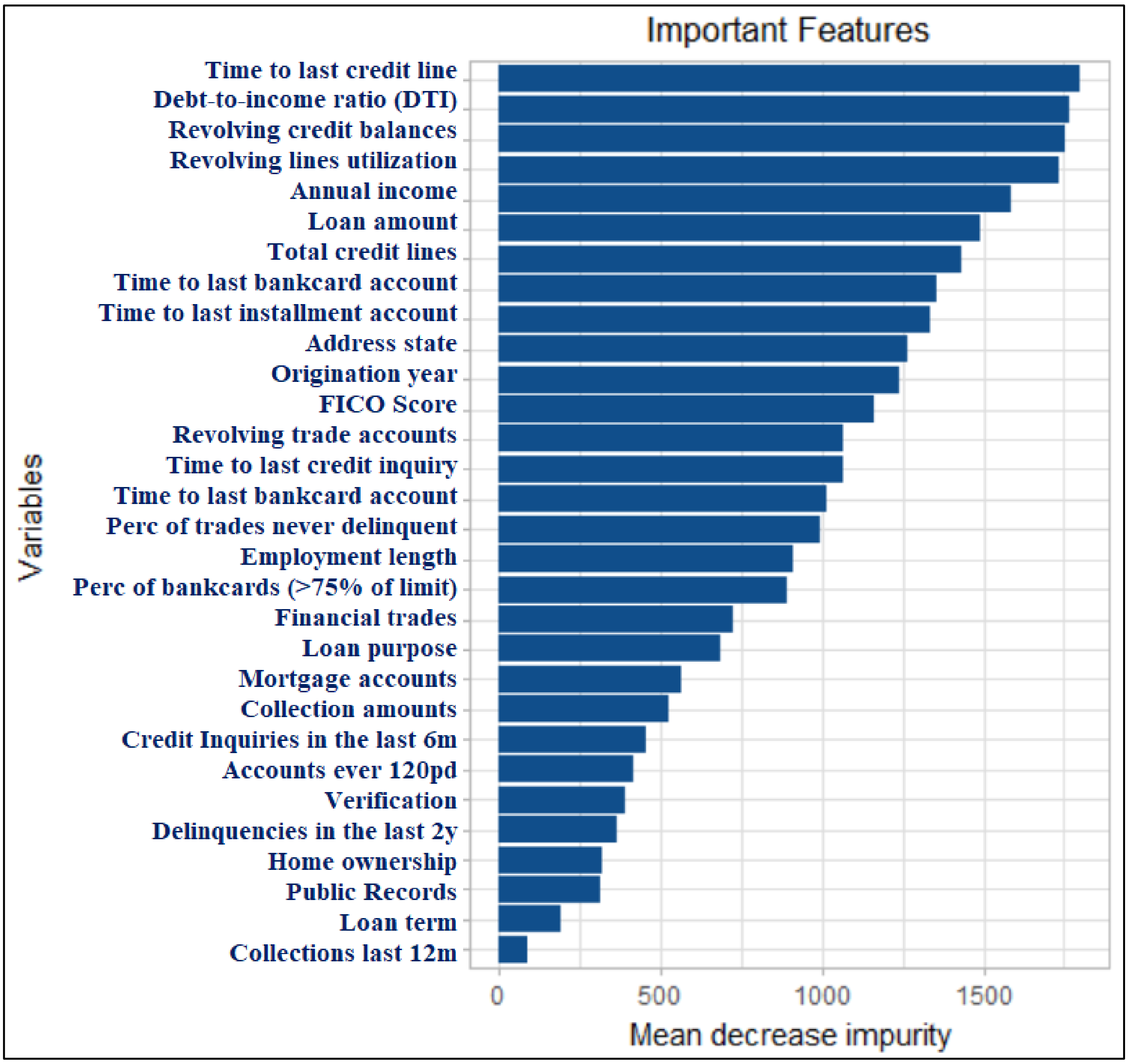
Note: The prediction error rate indicates the probability that the loans are incorrectly classified by the random forest model.

When the number of trees in the forest grows, the prediction error gradually decreases and approaches to its limiting value. Notably, the prediction error sees its biggest improvement in the first 100 simulations and stabilizes thereafter at about 14%. With 200 simulations we hence had a good prediction improvement without sacrificing computational speed.

[99] A sample that is randomly generated by bootstrap aggregation.
[100] Cross-validation is a model validation technique that is used to asses how the results of a statistical analysis will generalize to new observations.

The determinants of the delinquency rates are presented in order of importance[101] in Figure 5.3. The feature importance was obtained through the concept of mean decreasing impurity that we defined before, therefore the most important features in the classification model are the ones that cause the highest total decrease in impurity averaged across all the trees.

Figure 5.3 – Delinquency determinants ranked by importance.



Note: A higher mean decrease in impurity means that a feature has more importance in classifying the loans as delinquent.

According to the random forest model among the most important factors in determining the delinquencies we find recent credit lines, debt-to-income ratio, revolving credit balances, revolving credit utilization rate, annual income, loan amount, total credit lines, address state and origination year. Notably, FICO Scores are not the most important feature across the dataset. A complete list of the features and their description is presented in Table 5.1.

[101] The importance refers to the features with the most important role in predicting the future delinquencies rates observed on the platform

Table 5.1 – Description of the delinquency determinants.

	VARIABLES	DESCRIPTION
1	**Time to last credit line**	Months since the most recent credit line was opened.
2	**Debt-to-income ratio (DTI)**	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
3	**Annual income**	The self-reported annual income provided by the borrower during registration.
4	**Loan amount**	The listed amount of the loan applied for by the borrower.
5	**Revolving credit balances**	Total revolving credit balance.
6	**Revolving lines utilization**	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
7	**Total credit lines**	The total number of credit lines currently in the borrower's credit file.
8	**Time to last bankcard account**	Months since most recent bankcard account opened.
9	**Time to last installment account**	Months since most recent installment accounts opened.
10	**Address state**	The state provided by the borrower in the loan application.
11	**Origination year**	Indicates when the loan was issued.
12	**FICO Score**	The borrower's FICO Score before origination.
13	**Revolving trade accounts**	Number of active revolving trade accounts.
14	**Time to last credit inquiry**	The number of months since the borrower's last credit inquiry.
15	**Time to last bankcard account**	Months since most recent account opened.
16	**Perc of trades never delinquent**	Percent of revolving trades that never went delinquent.
17	**Employment length**	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
18	**Perc of bankcards (>75% of limit)**	Percentage of all bankcard accounts > 75% of limit.
19	**Financial trades**	Total number of financial trades.
20	**Loan purpose**	The loan purpose provided by the borrower.
21	**Mortgage accounts**	Total number of mortgage accounts.
22	**Collection amounts**	Total collection amounts ever owed by an applicant.
23	**Credit Inquiries in the last 6m**	The number of credit inquiries in the 6 months before origination.
24	**Accounts ever 120pd**	Total number of accounts ever 120 or more days past due.
25	**Verification**	Indicates if income was verified by LC, not verified, or if the income source was verified.
26	**Delinquencies in the last 2y**	The number of 30+ days past-due incidences of delinquency in the borrower's credit file in the 2 years before origination.
27	**Home ownership**	The home ownership status provided by the borrower during registration or obtained from the credit report. (i.e. Rent, Own, Mortgage, Other).
28	**Public Records**	Number of derogatory public records.
29	**Loan term**	The number of payments on the loan. Values are in months and can be either 36 or 60.
30	**Collections last 12m**	Number of credit collections in the 12 months before origination.

Although this is a slightly different type of research our findings don't differ too much from what was discovered in the past literature. For example Emekter et al. (2015) found that the debt-to-income ratio, and the revolving line utilization are both important determinants of the borrowers' defaults. Moreover, Carmichael (2014) and Serrano-Cinca et al. (2015) discovered that annual income and credit history length are also very important predictors.

As we can see there are many different pieces of information about the customers. For example information about credit line history, new credit lines, credit mix, and credit balances can all be found on the credit report of an applicant and should be already factored inside FICO Scores. Annual income, DTI, address states, employment length, loan purpose are instead all examples of information that are completely ignored by FICO Scores.

Feature Importance – Most important features

Now that we have ranked the features by importance we want to assess their effective ability of predicting the delinquency rates observed on the platform along with LC's Grades, FICO Scores. To avoid creating an excessively complex model and risk to overfit the data we will focus only on the 20 most important features in addition to FICO Scores. The selected features are reported and described in Table 5.2 where they are divided between information that can be found on the credit reports[102] and information that are self-reported by the applicants on the online platform. We will later use both these groups of information as explanatory variables in our models. First we will introduce only the credit report information and then the other features collected by LC. The idea is to understand if collecting all these different types of information adds a significant benefit to FICO Scores in predicting the delinquencies.

Table 5.2 – important information from credit reports and self-reported by the borrowers.

TYPE OF INFORMATION	VARIABLE
Credit reports	Time to earliest credit line, Revolving credit balances, Revolving lines utilization, Total credit lines, Time to last bankcard account, Time to last installment account, Revolving trade accounts, Time to last credit inquiry, Time to last bankcard account, Perc of trades never delinquent, Perc of bankcards (>75% of limit), Financial trades, Mortgage accounts
Self-reported by borrowers	Debt-to-income ratio (DTI), Annual income, Loan amount, Address state, Origination year, Employment length, Loan purpose

[102] Which should theoretically be factored in the FICO Scores.

The summary statistics of the most important continuous variables are presented in Table 5.3 and include Minimum, 1st. Quartile, Median, Mean, 3rd Quartile and Maximum. We will now briefly go through some of these statistics to get a better insight about LC's customers.

Table 5.3 – Most important continuous features summary statistics.

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Time to last credit line	3.1	11.2	14.7	16.1	19.9	83.1
Revolving credit balances	0	5,978	11,134	15,097	19,483	102,595
Revolving lines utilization	0.0%	30.4%	48.0%	48.6%	66.7%	100.0%
Total credit lines	3.0	16.0	23.0	24.3	31.0	61.0
Time to last bankcard account	0.0	6.0	13.0	22.2	26.0	562.0
Time to last installment account	0.0	7.0	12.0	19.5	22.0	503.0
Revolving trades	0.0	3.0	5.0	5.59	7.0	17.0
Time to last credit inquiry	0.0	2.0	5.0	6.90	10.0	25.0
Time to last bankcard account	0.0	3.0	5.0	7.0	9.0	175.0
Perc of acc never delinquent	5.0%	90.5%	97.1%	93.5 %	100.0%	100.0%
Perc of bankcards (>75% of limit)	0.0%	0.0%	33.3%	40.2%	66.7%	100.0%
Financial trades	0.0	0.0	0.0	1.37	2.0	12.0
Mortgage accounts	0.0	0.0	1.0	1.45	2.0	7.0
Debt-to-income ratio (DTI)	0%	12.6%	18.3%	18.8%	24.7%	47.0%
Annual income	6,000	50,000	70,000	78,969	98,000	280,500
Loan amount	1,000	8,000	12,375	14,876	20,000	40,000

Note: The time variables are expressed in months.

From the summary statistics we can see that on average borrowers obtained new credit line in the 16 months before application with about 25% of them being able to get access to credit in the same year on which they applied.

Revolving line balances and revolving line utilization refers to what is the maximum credit line that the borrower has access to and how much it was used. Average applicants had access up to about $15,000 of credit and used almost half of their credit line. A recommendation that is generally given to borrowers is to keep their revolving debt utilization below 30% of the total available credit line, so that the utilization rate doesn't hurt their FICO score. However, from the dataset statistics we can see that most of the applicants had an utilization rate which was above that percentage. The fact that the average LendingClub's customer carries larger amounts of debt than what is normally recommended can explain why most frequent loan purposes are debt consolidation and credit card refinancing.

The statistics for the debt-to-income indicate that the average applicant had a DTI of 18.8%, with 75% of the total applicants having a DTI lower than 24.7%. Usually a DTI lower than 35% is the most attractive to lenders, with higher values making it progressively harder for borrowers to get access to credit. With traditional lenders the 43% percent debt-to-income ratio is an important threshold for who wants to apply for credit, as it often represents the highest ratio a borrower can have to get access to qualified mortgages[103]. Larger lenders may still lend to borrowers with a DTI higher than 43%, however the loan will assume a different connotation which means that the borrower may face abusive interest rates and conditions. LendingClub's unsecured personal loans offered a certain degree of flexibility with borrowers being accepted with a DTI of up to 47%. Other non-numerical factors that are ignored by FICO Scores include information about the employment, the address state, the loan purpose, and the origination date.

The annual income variable represents one of the most important piece of information because it's an immediate indicator of a borrower's ability to repay its debt. From the summary statistics we found out that the average applicant of LendingClub earns about $79,000/y with three quarter of the borrowers earning no more than $100,000/y. About 25% of the remaining borrowers declared instead at least six figures income to the platform.

On average applicants borrowed $14,876 from LendingClub but we know that the amount varies among the different classes of borrowers. Less creditworthy individuals were in fact able to borrow more than the prime borrowers.

The non-continuous variables will be taken as dummies: the employment length will take the value of 1 when the borrower worked for ten years or more, and 0 otherwise. The state address will take the value 1 for the states with the highest loan originations[104] and 0 otherwise. The loan purpose will value 1 when the purpose is debt consolidation or credit card refinancing and 0 otherwise. The year dummy will value 1 when the year is 2016 and 0 otherwise.

5.2.2 – Predictive Ability

Now that we've selected our group of important features we will run four different specifications of logistic regression to understand whether grades have a better predictive power than FICO Scores and what is the contribute of the most predictive features in explaining the future loan performance. Unlike before, with Logit we had to split our data in training set and

[103] Qualified mortgages offer several benefits to the borrowers prohibiting excessive upfront costs and fees, and toxic loan features such as interest-only loans, negative amortization loans, or terms longer than 30 years.
[104] States with higher originations include New York, New Jersey, Florida, California and Texas.

testing set to see how the models perform outside the sample on which they are built. Therefore, the total sample of 683,448 observations was split in training set and testing set, where 70% of the data was used for training, and the remaining 30% for testing. In the training set we estimated all the parameters of the logistic regression, and we used the testing set to evaluate how well the models predict the loan delinquencies through the ROC Curves. It's very important to randomly select the data so that testing and training sets come from the same distribution. Moreover, to guarantee comparability among all the model specifications they used the same training and testing sets.

The outcomes of the logistic regressions are presented in Table 5.4 and have the following model specifications:

(1) LendingClub grades

(2) FICO Scores

(3) FICO Scores + Credit report information

(4) FICO Scores + Credit report and borrower's self-reported Information

To test for exclusion restrictions (i.e. incremental models parameters jointly equal to zero) we used the likelihood ratio (LR) test[105] that is based on the difference in the log-likelihood functions for the unrestricted and restricted models. Under the null hypothesis the test statistic LR is distributed as a chi-squared with q degrees of freedom, one for each exclusion restriction.

The test was run twice, first between model (3) and model (2) and second between model (4) and model (3). In both the cases the p-values for the likelihood ratio test rejected the null hypothesis indicating that all the coefficients of the incremental models are jointly significant. In other words if we exclude these variables from the specification the fall in the log-likelihood is large enough to conclude that they are important.

Since the parameters of the logistic regression are in log-odds they are not directly interpretable. This is why they were transformed in the odds ratio (OR) by taking the exponential (Table 5.5).

When the odds ratio is > 1 we have that the variable is associated with an increase in the probability of delinquency, while an odds ratio < 1 indicates the opposite. An OR = 1 indicates instead no association between the predictor and the probability of observing a delinquency.

[105] The likelihood ratio test is a statistical test of the goodness-of-fit between two models (restricted and unrestricted).

Table 5.4 – Logistic Regression Outputs.

$\hat{Y}$ = Loan Delinquency	*Model (1) LendingClub Grades (A-G)*		*Model (2) FICO Scores*		*Model (3) FICO Scores + Credit report information*		*Model (4) FICO Scores + All important information*	
	β	**(SE)**	**β**	**(SE)**	**β**	**(SE)**	**β**	**(SE)**
Intercept	-3.1916^{***}	(0.0186)	6.5998^{***}	(0.1120)	4.872^{***}	(0.1377)	5.6718^{***}	(0.1408)
D(*Grade B*)	0.8787^{***}	(0.0208)						
D(*Grade C*)	1.4429^{***}	(0.0200)						
D(*Grade D*)	1.9314^{***}	(0.0210)						
D(*Grade E*)	2.2675^{***}	(0.0232)						
D(*Grade F*)	2.6251^{***}	(0.0289)						
D(*Grade G*)	2.7589^{***}	(0.0396)						
FICO Scores			-0.0121^{***}	(0.0002)	-0.0093^{***}	(0.0002)	-0.0108^{***}	(0.0002)
Time to last credit line					-0.0078^{***}	(0.0007)	-0.0069^{***}	(0.0007)
Revolving credit balances					$-6\cdot e^{-06***}$	$(4\cdot e^{-07})$	$-1.4\cdot e^{-05***}$	$(5\cdot e^{-07})$
Revolving lines utilization					0.0017^{***}	(0.0003)	0.0007^{*}	(0.0003)
Total credit lines					-0.0003^{***}	(0.0005)	-0.0035^{***}	(0.0003)
Time to last bankcard account					-0.0033^{***}	(0.0002)	-0.0037^{***}	(0.0002)
Time to last installment acct					-0.0048^{***}	(0.0002)	-0.0035^{***}	(0.0002)
Revolving trade acounts					0.0226^{***}	(0.0015)	0.0197^{***}	(0.0016)
Time to last credit inquiry					-0.0251^{***}	(0.0008)	-0.0255^{***}	(0.0008)
Time to last bankcard account					-0.0111^{***}	(0.0009)	-0.0133^{***}	(0.0009)
Perc of trades never delinquent					0.0023^{***}	(0.0005)	0.0010^{*}	(0.0005)
Perc of bankcards (>75% of limit)					0.0015^{***}	(0.0002)	0.0016^{***}	(0.0002)
Financial trades					0.0116^{***}	(0.0016)	0.0097^{***}	(0.0021)
Mortgage accounts					-0.0907^{***}	(0.0030)	-0.0851^{***}	(0.0031)
Debt-to-income ratio (DTI)							0.0187^{***}	(0.0008)
Annual Income							$-3\cdot e^{-06***}$	$(5\cdot e^{-07})$
Loan Amount							0.00004^{***}	(0.0000)
D(*Address state*)							0.1096^{***}	(0.0088)
D(*Origination year*)							-0.0649^{***}	(0.0087)
D(*Employment length*)							-0.0556^{***}	(0.0095)
D(*Loan purpose*)							-0.2132^{***}	(0.0104)
Observations	**478,413**		**478,413**		**478,413**		**478,413**	

*Note: The symbols ***, **, and * represent the statistical significance of the coefficients estimates at the 0.1%, 1%, and 5% levels, respectively.*

Table 5.5 – Logistic Regression Odds Ratios.

$\hat{Y}$ = *Loan Delinquency*	Model (1) *LendingClub Grades (A-G)*	Model (2) *FICO Scores*	Model (3) *FICO Scores + Credit report information*	Model (4) *FICO Scores + All important information*
	Exp(β)	**Exp(β)**	**Exp(β)**	**Exp(β)**
Intercept	0.0411	734.987	130.5935	290.5810
D(Grade B)	2.4078			
D(Grade C)	4.2332			
D(Grade D)	6.8991			
D(Grade E)	9.6557			
D(Grade F)	13.8071			
D(Grade G)	15.7833			
FICO Scores		0.9879	0.9907	0.9891
Time to last credit line			0.9921	0.9930
Revolving credit balances			0.9999	0.9999
Revolving lines utilization			1.0017	1.0006
Total credit lines			0.9997	0.9964
Time to last bankcard account			0.9967	0.9963
Time to last installment account			0.9951	0.9964
Revolving trade accounts			1.0229	1.0199
Time to last credit inquiry			0.9751	0.9747
Time to last bank account opening			0.9889	0.9867
Perc of trades never delinquent			1.0023	1.0010
Perc of bankcards (>75% of limit)			1.0015	1.0016
Financial trades			1.0117	1.0097
Mortgage accounts			0.9132	0.9183
Debt-to-income ratio (DTI)				1.0188
Annual Income				0.9999
Loan Amount				1.00004
D(*Address state*)				1.1158
D(*Origination year*)				1.0670
D(*Employment Length*)				0.9458
D(*Loan Purpose*)				0.8079
Observations	**478,413**	**478,413**	**478,413**	**478,413**

Note: The coefficients represent the odds ratio (OR) of the regression. An OR > 1 means that the variable is associated with an increase in the probability of delinquency, while an OR < 1 indicates the opposite. An OR = 1 indicates instead no association.

From model (1) on Table 5.5 we understand that higher grades are generally associated with high probabilities for a loan to become delinquent. The same result was also observed by Emekter et. al (2014) with the default rates on the loans. As we would expect we also observed that higher FICO credit scores are associated with lower delinquency probabilities (model 2).

If we look instead at the most important features from the credit reports (model 3) we have that all the time variables have a negative association with the delinquencies. For example the more time that have passed since the first account has opened, or from the last credit inquiry the lower will be the probability of observing a bad loan. Higher revolving credit balances are associated with lower delinquencies, but a higher revolving line utilization means that a loan is more likely to become bad. In a similar way a higher percentage of bankcard accounts close to the limit increases the delinquency probability.

From the self-reported information on model (4) we have that higher annual incomes decrease the probability of observing delinquencies while higher loan amounts have the opposite effect. Intuitively when applicants had a stable job for more than 10 years they also had a lower probability of becoming delinquent. The states with the highest level of activity were also the ones with the highest probability of delinquency. A possible explanation for this result could be that the states with the highest operativity are also the ones in which the company expanded the credit access to less creditworthy borrowers the most. Finally, when the purpose of a loan was debt consolidation we observed a lower probability for the loan to go bad. This last result suggests that borrowers that go to LendingClub to consolidate a more expensive debt are in general a safer choice for the company.

Predictive ability – FICO Scores vs. Grades

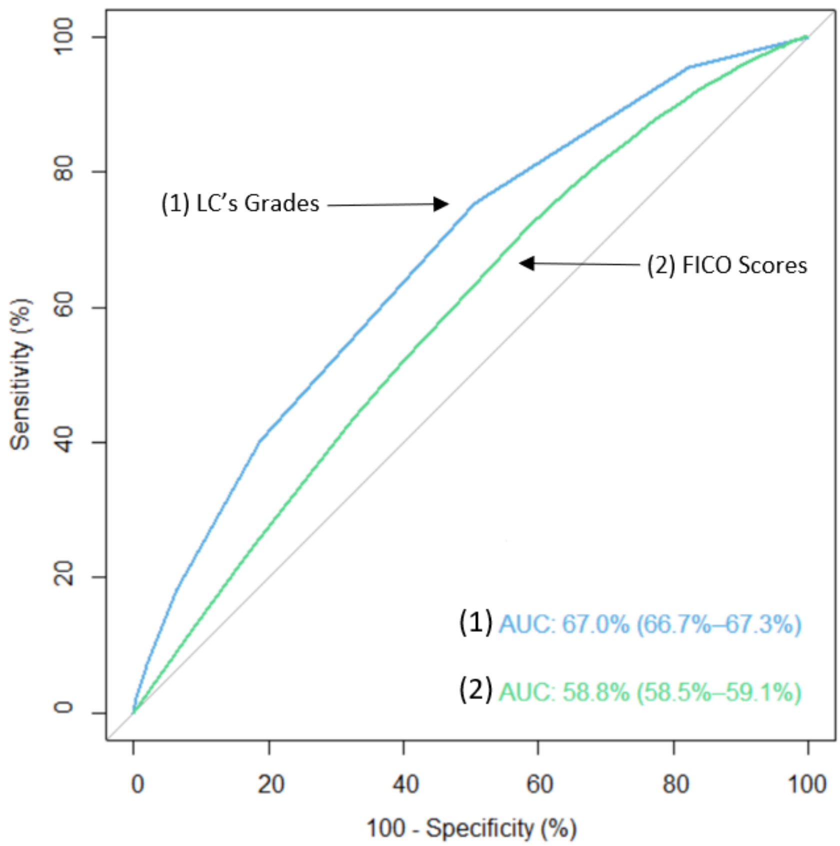
Now that we presented our models we will visualize their actual predictive performance through the receiving operative curves. As we know from the previous sections, the higher the area under the curve the better is the predictive ability of a model. Sensitivity is the fraction of observations that were correctly classified to be positive among all the positive observations, while the specificity is the proportion of negatives that are correctly classified. The confidence intervals[106] at 95% confidence level (CI) of the ROC curves are represented inside the parenthesis, and we will use them to compare the predictive ability of the different models. If

[106] The confidence interval is a range of values in which we are sure that the true value lies with a specified degree of confidence (i.e. 99%, 95%, 90%). For example with a 95% confidence interval we are 95% certain that the true value is inside the range.

the confidence interval around one ROC curve overlap that of a second ROC curve, we can say that there is no statistical difference between the outcomes (Kerekes, 2008).

From Figure 5.4 which includes LC's grades and FICO Scores we can immediately see that the grades have a predictive ability which is way above FICO Scores with an AUC of 67.0% and 59.1%, respectively.

Figure 5.4 – Predictive ability of LC Grades and FICO Scores.



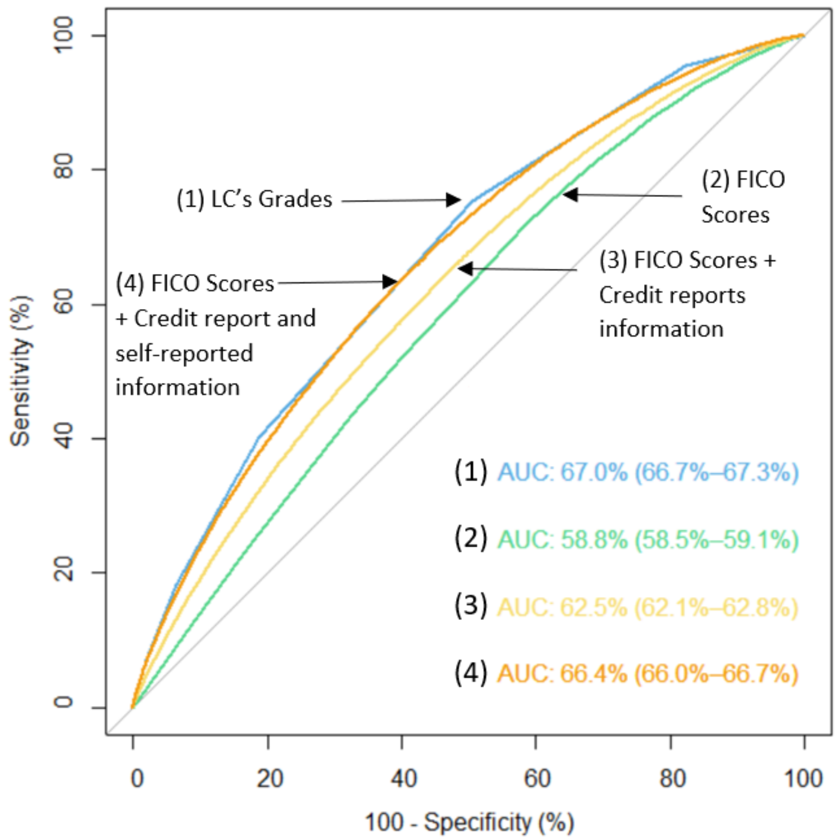
Note: The area under the curve (AUC) indicates the predictive ability of a model. The higher the area the better the model. The CI of the curves at 95% confidence level are shown in parenthesis.

Moreover, we have that the confidence intervals for grades and FICO Score do not overlap with each other confirming the superiority of LendingClub's grades in predicting the future loan performance. Therefore, we can conclude that the higher amount of data that is used to generate grades adds important benefits in assessing the riskiness of the applicants.

Predictive ability – Most important information

Although the previous result is very important we also want to understand what is the contribution of the most important features in predicting the delinquency rates observed on the platform. From Figure 5.5 we can see how the ROC curves of the models with the additional information compare to the ones of LC's grades and FICO Scores.

Figure 5.5 – Predictive ability LC Grades and Fico Scores with important features.



Note: The area under the curve (AUC) indicates the predictive ability of a model. The higher the area the better the model. The CI of the curves at 95% confidence level are shown in parenthesis.

Comparing model (3) to model (2) we can notice that model (3) has a higher predictive ability with an AUC of 62.5% and 58.8%, respectively. Moreover, the confidence intervals do not overlap with each other indicating that the set of information contained in model (2) add a significant benefit in predicting the future delinquencies. Interestingly, the FICO Scores performance can still be improved by just adding the information from the credit files which can explain why LendingClub still includes these types of information on its dataset.

Looking instead at model (4) and model (3) we have that the self-reported information also add a significant benefit in terms of predictive ability (AUC: 66.4% vs. 62.5%), with the confidence intervals confirming the result. The self-reported information are hence all important elements that should be factored to improve the scoring system.

Notably, if we compare model (4) to model (1) we can see that they have the AUC (66.4% vs. 67.0%) for these two models is now much closer than before (also from the graphical representation it's easy to see how close are the two curves). Moreover, the upper bound of the confidence interval of model (4) crosses the lower bound of model (1) at 66.7% indicating that the two models have substantially the same predictive ability. Not only the model (4), which includes all the most important features, was able to outperform model (3) with FICO Scores and credit information, but it was also able to reach the predictive ability of LC's grades confirming the importance of including such information in the scoring system.

A superior ability of assessing the borrowers' creditworthiness can translate in the possibility of expand credit access. Recent literature also found significant benefits coming from additional data sources outside the traditional credit records. For example Berg et al. (2019) analyzed the information content of the "digital footprint"[107] for predicting consumer default finding that it equals or exceeds the predictive power of traditional credit bureau scores. Jagtiani and Lemieux (2019) found instead that most of the "invisible prime" borrowers, who have been rated poorly by the traditional credit scoring mechanism, had a very low default probability that was very close to the default probability of prime borrowers, suggesting that collecting alternative data is a good way to expand credit access to less creditworthy borrowers.

5.2.3 – Pricing analysis

Throughout different multiple linear regression models we now want to understand whether the most important delinquency determinants have a significant impact in explaining the pricing decisions of LendingClub. The analysis will be based on four model specifications which are similar to the one that we have used before:

(1) LendingClub grades

(2) FICO Scores

(3) FICO Scores + Credit report information

(4) FICO Scores + Credit report and borrower's self-reported information

The dependent variable (loan prices) is now represented by the interest rate spreads, which are calculated as the difference between the interest rates charged by LendingClub and the equivalent U.S. Treasury interest rates of securities with the same maturity (i.e. 3 or 5 years).

FICO Scores are now divided in different dummies representing intervals which are closely related to their corresponding ratings: fair/good (660-679), good (680-739), very good (740-800), exceptional (>800). Moreover, to allow for a non-linear relationship we took the logarithm of the variables revolving balances, annual income and loan amount. All the coefficient estimates from the different multiple linear regression models are showed in Table 5.6. The dummies Grade "A" and FICO Scores higher than 800 are used as the base case.

[107] Digital footprint or digital shadow refers to your unique set of traceable digital activities, actions, contributions and communications manifested on the Internet or on digital devices.

Table 5.6 – Multiple Linear Regression Outputs.

$\hat{Y}$ = *Credit Spreads*	*Model (1) LendingClub Grades (A-G)*		*Model (2) FICO Scores*		*Model (3) FICO Scores + Credit report information*		*Model (4) FICO Scores + All important information*	
	β	**(SE)**	**β**	**(SE)**	**β**	**(SE)**	**β**	**(SE)**
Intercept	5.3190***	(0.0038)	6.2828***	(0.0545)	14.3158***	(0.0945)	14.5287***	(0.1864)
D(Grade B)	3.4760***	(0.0047)						
D(Grade C)	7.0861***	(0.0047)						
D(Grade D)	11.5782***	(0.0056)						
D(Grade E)	16.5162***	(0.0076)						
D(Grade F)	20.5215***	(0.0118)						
D(Grade G)	23.0291***	(0.0183)						
D(FICO 741-800)			1.51450***	(0.0576)	0.6529***	(0.0553)	0.7419***	(0.0524)
D(FICO 681-740)			5.49953***	(0.0550)	3.0438***	(0.0543)	3.1879***	(0.0515)
D(FICO 660-670)			7.41905***	(0.0563)	4.1160***	(0.0565)	4.4631***	(0.0536)
Time to last credit line					-0.0463***	(0.0008)	-0.0436***	(0.0008)
Revolving credit balances					-0.6186***	(0.0078)	-0.9126***	(0.0080)
Revolving lines utilization					0.0361***	(0.0004)	0.0346***	(0.0005)
Total credit lines					-0.0136***	(0.0006)	-0.0218***	(0.0006)
Time to last bankcard account					-0.0094***	(0.0002)	-0.0105***	(0.0002)
Time to last installment account					-0.0147***	(0.0002)	-0.0105***	(0.0002)
Revolving trade accounts					0.1224***	(0.0021)	0.1027***	(0.0020)
Time to last credit inquiry					-0.1036***	(0.0010)	-0.1064***	(0.0009)
Time to last bank account					-0.0588***	(0.0010)	-0.0681***	(0.0009)
Perc of trades never delinquent					-0.0005	(0.0006)	-0.0095***	(0.0006)
Perc of bankcards (>75% of limit)					0.0148***	(0.0002)	0.0150***	(0.0002)
Financial trades					0.0417***	(0.0022)	0.0283***	(0.0021)
Mortgage accounts					-0.1795***	(0.0035)	-0.1349***	(0.0034)
Debt-to-income ratio (DTI)							0.0826***	(0.0008)
ln(Annual Income)							-1.5207***	(0.0147)
ln(Loan Amount)							2.0947***	(0.0087)
D(*Address state*)							0.3741***	(0.0106)
D(*Origination year*)							-0.0179*	(0.0090)
D(*Employment length*)							0.0967***	(0.0114)
D(*Loan purpose*)							-1.1638***	(0.0127)
Adj. R-squared	**93.82%**		**9.93%**		**18.33%**		**27.12%**	
Observations	**683,448**		**683,448**		**683,448**		**683,448**	

*Note: The symbols ***, **, and * represent the statistical significance of the coefficients estimates at the 0.1%, 1%, and 5% levels, respectively.*

From model (1) we can see that the estimates for LC grades coefficients are all positive and in rank order. The regression shows that loans with the highest grade ("A") are associated with a price of 5.32% (i.e. the intercept coefficient), with the other grades having an incrementally higher price. For example, loans with "B" grade have a predicted credit spread of 8.80%[108], while loans with the lowest grade ("G") have a credit spread of 28.35%. Similar results are found with FICO Scores in model (2), where lower credit ratings are associated with lower spreads and vice versa.

By looking at model (1) we can see that the model including only LC Grades fits well the data with an R^2 of 93.82% indicating that there is a very strong relationship between LendingClub's proprietary grades and the credit spreads. If we compare instead model (1) with model (2) we observe very different R-squared (93.82% vs. 9.93%).

FICO Scores alone explained only a smaller part of the credit spreads which demonstrates how LendingClub relies on a much larger set of information when it comes to price its loans, but this hasn't always been the same. Initially FICO Score weighted much more on the pricing decisioning of LendingClub than what they do now. In their study on LC's loans, Jagtiani & Lemieux (2018a) found that in 2007 correlation between the LendingClub's grades and FICO scores was about 80 percent before decreasing to about 35 percent for loans originated in 2014-2015. As LendingClub expanded its customer base and consequently the amount of additional information it could collect, alternative data have assumed a growing importance in guiding the pricing decisions.

Looking at model (4) we can see that all the coefficients for the additional information are significant at a 5% level giving a clear indication of a strong relationship between them and the rates charged by LendingClub. Even after accounting for heteroskedasticity by computing White standard errors[109] the parameters remained significant.

Focusing on the most important predictors from the credit reports we can see that higher revolving credit balances are associated with lower credit spreads, but when the revolving line utilization increases the effect is the opposite. In other words individuals who have access to larger lines of credit are charged lower rates but if they are not able to manage them appropriately they will be charged more. When the time since new applicants opened new credit lines increases the borrowers benefit from better rates as they don't show a high need for credit.

Annual income is one of the most direct indicators of a borrower's ability to repay, and as we would expect borrowers with a higher income are charged less by the company. The debt-

[108] A credit spread of 8.80% is the result of 5.32% (Intercept) + 3.48 (Coefficient for grade B)
[109] White standard errors are heteroskedasticity robust standard errors. In their calculation it is not imposed any assumptions on the structure of heteroskedasticity.

to-income measures instead the adequateness of the borrower income to repay the debt. When this index grows we have a clear indication that the sources of income are becoming less adequate to cover the debt and therefore higher spreads are applied.

Borrowers who demanded for more credit are also charged more on average, but we also know that the less creditworthy borrowers are the ones who are more in need for funds.

In general the parameters are consistent with what we observed with the delinquency rates and variables that increase the probability of a borrower to become delinquent are also the ones that increase the rates that they will be charged.

Comparing model (3) with model (2) we can see that the inclusion of the additional features from the credit bureaus improves R^2 up to 18.33% confirming that FICO Scores are just one out of many determinants for the price. When we also include the self-reported information (model 4) we can see that the R^2 increases even more up to 27.12% which indicates that also these factors are important drivers of LendingClub's decisions.

Interestingly, by including all the self-reported information it was still not possible to get close to the R-squared observed with LendingClub grades which suggests that in reality the pricing mechanism is more complex and includes also other factors that weren't disclosed on the dataset. However, we still found a strong relationship between the use of the additional information collected by the company and the credit spreads confirming that LendingClub price decision account for many more factors beyond FICO Scores.

Given the significant benefits that additional data can bring to lenders it's not surprising that an increasing number of FinTech lenders are abandoning FICO Scores by implementing more advanced scoring systems. It's important to mention that Banks can also benefit from alternative data sources and "soft information" which is why we are likely to see a convergence of banks and FinTechs towards an integrated financial system that leverages alternative information and big data to offer a better value proposition to all the customers.

CONCLUSIONS

FinTech has the potential reshape the way by which intermediation is carried out in the financial system. Technology has allowed new firms to serve small businesses and consumers without the expensive investments of the traditional financial institutions. FinTech can be applied to a wide spectrum of activities, and in our study we focused on the lending activity which is also the most recurrent one in the FinTech space.

From our analysis on the platform LendingClub we discovered that LC's innovative credit scoring system outperforms FICO Scores in predicting the future performance of the loans. Our findings also highlighted that information such as credit history length, debt-to-income ratio, revolving credit balances, revolving credit utilization rate, and annual income are among the most important determinants of a borrower's delinquency.

Information about credit history length, revolving credit balances and utilization rates should all enter in the generation of FICO Scores. While annual income, DTI, and other non-numerical factors such as employment length, geographical area, loan purpose, loan amount and issuing date are all important information that FICO Scores do not account for. By testing the predictive abilities of the most important features we saw that the information from both the credit reports and that are self-reported by the borrowers add a significant benefit to FICO Score in explaining the future loans performances. FICO Scores have a predictive ability which is way below the one of LendingClub's grades, however with a model that included the 20 most important features we were able to replicate the same predictive ability of the broader grades categories (A-G) confirming the importance of collecting these information. In the final part of our study we discovered that all these features also played a significant role in explaining the rates that the borrowers were charged on the platform. Interestingly, we were not able to explain all the variability that we observed in the credit spreads which suggests that the pricing mechanism is more complex and includes even more information beyond FICO Scores.

The fact that many FinTech lenders are gradually abandoning traditional scoring systems is just one consequence of the transformation that is happening in the credit markets. The more alternative lenders grow and further increase their availability of data, the better it will be their ability to reach new customers that are underserved by traditional lenders.

While an access to a large amount of datapoints about customers may cause regulatory and privacy concerns, it undoubtedly plays a key role in allowing lenders to better understand the credit quality of the potential borrowers.

To avoid scandals like the one of LendingClub and promote a more inclusive financial system it is very important for regulators to develop a regulatory framework that balances the benefits of innovation with the risks of these new approaches. While consumers' privacy should be protected by laws the use of technology and alternative data can also play a fundamental role to help regulators in supervising and monitoring the markets participants.

BIBLIOGRAPHY

Accornero, M. & Moscatelli, M., (2018). *"Listening to the buzz: social media sentiment and retail depositors' trust"*

Amit Y., Geman D., (1997). *"Shape quantization and recognition with randomized trees"*. Neural Computation. 9 (7): 1545–1588.

Anagnostopoulos, I. (2018). *Fintech and Regtech: Impact on regulators and banks.* Journal of Economics and Business, Volume 100.

Alteryx, (2020). *Seeing the Forest for the Trees: An Introduction to Random Forest.*

Arner, D. W., Barberis, J. N. and Buckley, R. P. (2015). *The Evolution of Fintech: A New Post-Crisis Paradigm?* University of Hong Kong Faculty of Law Research Paper No. 2015/047; UNSW Law Research Paper No. 2016-62.

Arner, D. W., Barberis, J. N. and Buckley, R. P. (2017). *FinTech, RegTech, and the reconceptualization of financial regulation.* Northwestern Journal of International Law and Business. 37. 373-415.

Balyuk, T., (2019). *Financial Innovation and Borrowers: Evidence from Peer-to-Peer Lending.*

Bank for International Settlements (BIS) (2017). *Sound practices: Implications for Fintech developments for banks supervisors.* Consultative document.

Barefoot, J. A. (2015). *Disrupting fintech law.* Fintech Law Reporter, 18(2), 1–18.

Berg, T., Burg, V., Gombović, A., and Puri, M., (2019). *On the rise of fintechs – credit scoring using digital footprints.* FDIC, Working paper.

Breiman, L., (1996). *Bagging predictors.* Machine Learning, 24:123–140.

Breiman, L., (2000). *Some infinity theory for predictor ensembles.* Technical Report 577, UC Berkeley.

Breiman, L., (2001). *Random forests.* Machine Learning, 45:5–32.

Breiman, L., (2004). *Consistency for a simple model of random forests.* Technical Report 670, UC Berkeley.

Breiman, L., Friedman J. H., Olshen, R. A., and Stone C. J., (1984). *Classification and Regression Trees.* Chapman & Hall, New York.

Bricker J., Dettling L.J., Henriques A., Hsu J.W., Jacobs L., Moore K.B., (2017). *Changes in U.S. family finances from 2013 to 2016: Evidence from the survey of consumer finances.* Federal Reserve Bulletin, 103 (3), pp. 1-42

Business Insider, (2020a). *Overview of the fintech industry: Latest trends, market research and analysis from our ecosystem report.* https://www.businessinsider.com/fintech-ecosystem-report

Carmichael, D. (2014). *Modeling default for peer-to-peer loans.* University of Houston - C.T. Bauer College of Business

Chen, N., Ghosh, A., Lambert, N.S., (2014). *Auctions for social lending: A theoretical analysis.* Games Econ. Behav. 86, 367–391.

Chishti, S. (2016). *How peer to peer lending and crowdfunding drive the fintech revolution in the UK.* New Economic Windows.

Christensen, C. M. (2003). *The innovator's dilemma: The revolutionary book that will change the way you do business.* The American Political Science Review.

Crespo, I., Naveira, C. F., & Kwon, H. (2018). *Using data to unlock the potential of an SME and mid-corporate franchise* (McKinsey & Company).

Das, S.R. (2019). *The future of fintech.* Financial Management; 48: 981– 1007.

Deloitte, (2020). *Fintech investors: Enthusiastic yet strategically picking their spots.* Deloitte Center for Financial Services.

DeMartino, G. F. and McCloskey, D. N. (2016). *The Oxford Handbook of Professional Economic Ethics.*

Demyanyk, Y., and Kolliner, D., (2014). *"Peer-to-Peer Lending Is Poised to Grow".* Federal Reserve Bank of Cleveland.

DeYoung, R. (2005). *The Performance of Internet-Based Business Models: Evidence from the Banking Industry.* The Journal of Business, 78(3), 893-948.

De Roure, C., Pelizzon, L., and Tasca, P., (2016)De Roure. *"How does P2P lending fit into the consumer credit market?"* Discussion Papers 30/2016, Deutsche Bundesbank.

De Roure, C., Pelizzon, L. and Thakor, A. (2018). *P2P lenders versus banks: Cream skimming or bottom fishing?* Working Paper.

Diamond, D. W. and Dybvig, P. H. (1983). *"Bank Runs, Deposit Insurance, and Liquidity"*, Journal of Political Economy, University of Chicago Press, vol. 91(3), pages 401-419.

Dietricha, A., Wernli, R., (2019). *Determinants of Interest Rates in the P2P Consumer Lending Market: How Rational are Investors?*

Einav, L., Jenkins, M., and Levin, J., (2013). *The impact of credit scoring on consumer lending.* RAND Journal of Economics, 44(2): 249–274.

Emekter, R., Tu, Y., Jirasakuldech, B., (2014). *Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending.* Applied Economics, 47:1, 54-70.

FED, 2019, Consumer credit – G.19, Board of Governors of the Federal Reserve System.

Fernández De Lis, S. (2016). *RegTech, the new magic word in FinTech, banking outlook*. MarchBBVA.

FSB (2017). *Financial Stability Implications from FinTech.*

Frame W.S., Srinivasan A., Woosley L., (2001). *The effect of credit scoring on small-business lending.* Journal of Money, Credit and Banking, 33 (3), pp. 813-825

Frost, J., (2020). *The Economic Forces Driving Fintech Adoption Across Countries.* BIS Working Paper No. 838.

Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., Zbinden, P., (2019). *BigTech and the changing structure of financial intermediation.* BIS Working Papers, No. 779.

Gambacorta, L., Huang, Y., Qiu, H., (2019), *How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm.* BIS Working Papers, No. 834.

Gambacorta, L., Huang, Y., Li, Z., Qiu, H., Chen, S., (2020). *Data vs collateral.* BIS Working Papers 881.

Gomber, P., Koch, J. A. and Siering, M. (2017). *Digital Finance and FinTech: Current research and future research directions.* Journal of Business Economics 87, 537–580.

Gulamhuseinwala, I., Subas, R. and Viljoen, A., (2015). *Innovating with regtech - Turning regulatory compliance into a competitive advantage.* Ernst and Young.

Ho, T. K., (1995). *Random Decision Forests*. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16. pp. 278–282.

Jaffee D. M., Russell T., (1976). *Imperfect information, uncertainty, and credit rationing.* Quarterly Journal of Economics, 90 (4) pp. 651-666

Kerekes, J., (2008). *Receiver Operating Characteristic Curve Confidence Intervals and Regions.* IEEE Geoscience and Remote Sensing Letters, vol. 5, no. 2, pp. 251-255.

Kroll J. A., Huey J., Barocas S., Felten E. W., Reidenberg J. R., Robinson D. G., & Yu H. (2017). *Accountable Algorithms*, 165 U. Pa. L. Rev. 633.

Iosco, (2017). *Research report on financial technologies.*

Jagtiani, J. and Lemieux, C. (2017). *Fintech Lending: Financial Inclusion, Risk Pricing, and Alternative Information.* FRB of Philadelphia Working Paper No. 17-17.

Jagtiani, J., & Lemieux, C., (2018). *Do fintech lenders penetrate areas that are underserved by traditional banks?* Journal of Economics and Business, Volume 100, Pages 43-54.

Jagtiani, J., & Lemieux, C., (2019). *The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform.* Federal Reserve Bank of Philadelphia, United States. Working Paper.

Mach T. L., Carter C. M., and Slattery C. R., (2014). *Peer-to-peer lending to small businesses.* Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C.

McWaters, J. F. (2015). The Future of Financial Services – *How disruptive innovations are reshaping the way financial services are structured, provisioned and consumed.* World Economic Forum 1–178.

Morrison & Foerster, (2018). *Practice Pointers on P2P Lending: How It Works, Current Regulations and Considerations.*

Morse, A. (2015). *"Peer-to-Peer Crowdfunding: Information and the Potential for Disruption in Consumer Lending"* Annual Review of Financial Economics, vol 7(1), pages 463-482.

Navaretti, G., G. Calzonari and A. Pozzolo (2017). *"FinTech and banking: Friends or foes?"*, European Economy: Banks, Regulation and the Real Sector 2017(2).

Petralia, K., Philippon, T., Rice, T., Véron, N. (2019). *Banking Disrupted? Financial Intermediation in an Era of Transformational Technology.* International center for monetary and banking studies.

Philippon, T. (2017). *The fintech opportunity.* BIS Working Papers, No. 655.

Philippon, T., (2019). *On Fintech and Financial Inclusion.* BIS Working Papers, No. 841.

Polena, M., Regner, T., (2016). *Determinants of borrowers' default in P2P lending under consideration of the loan risk class.* Jena Economic Research Papers, No.2016-023, Friedrich Schiller University Jena, Jena.

PwC (2016). *Global FinTech Report 2016.*

Serrano-Cinca, C., Gutiérrez-Nieto, B. (2016). *The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending.* Department of Accounting and Finance, University of Zaragoza, Zaragoza, Spain.

Serrano-Cinca, C., Gutierrez-Nieto, B., Lopez-Palacios, L. (2015). *Determinants of Default in P2P Lending*. PLOS ONE, 10(10):1–22.

Silverberg, K., Portilla, A., French, K., van Liebergen, B., and Van Den Berg, S. (2016). *RegTech financial services: Technology solutions for compliance and reporting.* Institute of International Finance, March 2016.

Stango, V., and Jonathan Z., (2009). *What do consumers really pay on their checking and credit card accounts? Explicit, implicit, and avoidable costs.* American Economic Review 99, 424–429.

Stiglitz, J. E., and Andrew W., (1981). *Credit rationing in markets with imperfect information.* American Economic Review 71, 393–410.

Tank, M. H. K., Whitaker D., Grant, A., Johnsen E., Hare J. L., Lucente K., Lee, V., Luce D. D., (2019). *Fintech regulation in the USA.* DLA Piper.

WalletHub, (2020). *Credit Card Landscape Report*. *What Is the Average Credit Card Interest Rate?* https://wallethub.com/edu/cc/average-credit-card-interest-rate/50841/

Wei, Z., Lin, M., 2015. *Market mechanisms in online crowdfunding.*

Woolridge, J. M., (2003). *Introductory econometrics: A modern approach.* South-Western Publishing Co., 2ed.

WSJ, (2016). *LendingClub to Change Its Fee Model*. The Wall Street Journal.

Zetzsche, D. A., Buckley, R. P., Arner, D. W., Barberis, J. N., (2017). *From FinTech to TechFin: The Regulatory Challenges of Data*-Driven Finance. EBI Working Paper Series, 2017, no. 6.