Federal Reserve Bank of Philadelphia

Working Papers
Research Department

WP 20-15
April 2020
https://doi.org/10.21799/frbp.wp.2020.15

Important Factors Determining Fintech Loan Default: Evidence from the LendingClub Consumer Platform

Christophe Croux
EDHEC Business School

Julapa Jagtiani
Federal Reserve Bank of Philadelphia Supervision, Regulation, and Credit Department

Tarunsai Korivi
Amazon.com

Milos Vulanovic
EDHEC Business School

Consumer Finance Institute

ISSN: 1962-5361

Important Factors Determining Fintech Loan Default: Evidence from the LendingClub Consumer Platform

Christophe Croux[1]
EDHEC Business School

Julapa Jagtiani[2]
Federal Reserve Bank of Philadelphia

Tarunsai Korivi[3]
Amazon.com

Milos Vulanovic[4]*
EDHEC Business School

April 2020

Abstract

This study examines key default determinants of fintech loans, using loan-level data from the LendingClub consumer platform during 2007–2018. We identify a robust set of contractual loan characteristics, borrower characteristics, and macroeconomic variables that are important in determining default. We find an important role of alternative data in determining loan default, even after controlling for the obvious risk characteristics and the local economic factors. The results are robust to different empirical approaches. We also find that homeownership and occupation are important factors in determining default. Lenders, however, are required to demonstrate that these factors do not result in any unfair credit decisions. In addition, we find that personal loans used for medical financing or small business financing are more risky than other personal loans, holding the same characteristics of the borrowers. Government support through various public-private programs could potentially make funding more accessible to those in need of medical services and small businesses without imposing excessive risk to small peer-to-peer (P2P) investors.

Keywords: big data, crowdfunding, financial innovation, household finance, lasso selection methods, machine learning, peer-to-peer lending, P2P/marketplace lending

JEL Codes: D10, D14, G20, G21, G 29

---

[1] Department of Data Science, Economics and Finance at EDHEC Business School, 24 Avenue Gustave Delory, CS 50411, 59057 Roubaix Cedex, France, phone: +33 (0)3 20 15 45 00; email: christophe.croux@edhec.edu.
[2] Federal Reserve Bank of Philadelphia, USA; phone: +1 (215) 574-7284; email: julapa.jagtiani@phil.frb.org.
[3] Data engineering at Amazon.com, 33 Rives de Clausen 31, 2165 Luxembourg; phone: +352 26 73 33 00; email: ttark@amazon.com.
[4] Department of Data Science, Economics and Finance at EDHEC Business School, 24 Avenue Gustave Delory, CS 50411, 59057 Roubaix Cedex 1, France; phone: +33 (0)3 20 15 45 00; email: milos.vulanovic@edhec.edu.
*Corresponding author

The authors thank Ken Benton, Mitchell Berlin, Harshali Damle, Bill Francis, Frank Fagan, Bob Hunt, Edward Lawrence, Prabesh Luitel, Snejina Panayotova, Robinson Reyes, Abdus Samad, Zvi Wiener, participants at the Financial Engineering and Business Society Conference at University of Economics in Prague, and participants at the Multinational Financial Society Conference at the Hebrew University for their helpful comments. Thanks also to Adam Lyko, Erik Dolson, and Blerta Hima for their research assistance.

1. Introduction

The market for consumer loans in peer-to-peer (P2P) or marketplace lending (MPL) settings, which started soon after the recent financial crisis, has become an important innovation that changed the entire financial landscape. Fintech lenders match lenders and borrowers, attempting to eliminate the redundant financial intermediaries.[5] Buchak et al. (2018) state that fintech lenders filled the mortgage credit gap created by the contraction of mortgage activities from traditional banks following the recent financial crisis and the Dodd–Frank Wall Street Reform and Consumer Protection Act. Tang (2019), observing regulatory change as an exogenous shock to bank credit supply, documents that P2P lending is a substitute for bank lending in terms of serving inframarginal customers. Added to their positioning as an efficient financial intermediary, fintech lenders use nontraditional data (alternative data) along with sophisticated modeling using artificial intelligence (AI) or machine learning (ML) algorithms to identify low-risk borrowers (often from the pool of borrowers with low credit scores) and to price credit more accurately, which represent a major divergence from the traditional banking; see Vallée and Zeng (2019).[6]

While most fintech lenders started as P2P lenders, they have recently supplemented funding through securitization, where fintech asset-backed securities (ABS) investors would invest in a fraction of the loan pool, rather than investing in a specific loan.[7] There have been concerns around the funding side of fintech loans — whether loans are made to borrowers who may be overleveraged and, consequently, natural candidates for bankruptcy — see Wang and Overby (2018). This would potentially impose excessive risk to fintech ABS investors. In addition, there have also been concerns related to whether this undue risk-taking from the entrants in financial intermediation warrants further inspection by regulators; see Philippon (2016) and Braggion, Manconi, and Zhu (2018). The institutional settings of P2P loan markets lead to a situation in which individual suppliers of capital bear all the risk. The risk increases when the platforms determine the funding interest rates themselves, ignoring auctions or other alternative standard supply-and-demand mechanisms; see Wei and Lin (2016). Therefore, to ensure the continuation of marketplace lending, the question of returns to investors in this market and the level of defaults are crucial ones.

[5] Funk et al. (2011) provide a review of the literature on P2P lending from its start, from 2005 until 2011, and concludes that P2P lending is becoming an essential source of funding for individuals and small businesses.

[6] Buchak et al. (2018) report that fintech lenders use a different set of information to determine interest rates compared with other lenders.

[7] According to Klafft (2008), P2P platforms have been defined as online intermediaries in which applicants place requests to obtain loans and suppliers of funds make bids to fund these loans. He dates their emergence to the year 2005. Klafft (2008) seems to have one of the earliest studies on the topic, although we were able to find only an abstract of the proceedings paper.

In this paper, we explore important factors that determine fintech loan performance, and we focus on the risk-return tradeoff on fintech lending and investments.

Research has shown that the credit decision process used by fintech lenders has been evolving rapidly over the years. A few studies that examine fintech loan defaults use data in the earliest years of the market analyzing the performance of less than 10 percent of resolved loans. Thus, their findings warrant further examination. Jagtiani and Lemieux (2019) find that the models' usage by LendingClub consumer platforms, for example, changed dramatically from 2007 to 2015. Specifically, they find that the correlation between the ratings assigned by LendingClub, and FICO scores decline from about 80 percent for loans that were issued in 2007, to only about 35 percent for loans issued in 2015. They also find that, over the years, an increasing number of consumers with low FICO scores have been able to access credit at a lower cost through the fintech lending platform. Other research studies find consistent results regarding the impact of fintech lending on consumer credit access. Danisewicz and Elard (2018) examine how financial technology affects personal bankruptcy. They document that the suppression of access to a new financial technology used by marketplace lending platforms leads to a higher incidence of personal bankruptcy filings. They conclude that fintech lending platforms have improved the screening process and the efficiency of financial intermediaries. Fintech lenders have increasingly used more and more big data and nontraditional data, in conjunction with more complex algorithms using AI/ML techniques to obtain a more complete picture of borrowers' financial lives.

Given rapid changes in fintech lending and the entire financial landscape in recent years, we include more recent loans originated in the 2015–2018 period by LendingClub consumer platforms in this study. Our samples include 1,345,549 individual personal loans that were issued during the period 2008–2018 on the LendingClub consumer platform. We contribute to the existing literature in two important ways. First, we include a more comprehensive set of risk factors than what has already been included in previous studies. Because of the change in the reporting of individual financial positions, LendingClub has provided much more detailed statistics, which enables us to observe a more comprehensive set of independent variables or default determinants than what previous studies were able to account for. Second, we have conducted a more robust analysis that includes a type of ML process (i.e., the least absolute shrinkage and selection operator (lasso) method of supervised learning). This method allows us to identify the important variables in a large set of potential determinants of loan defaults. The lasso selection method has been found to have excellent properties; see Tibshirani (1996); Meier, Van De Geer, and Bühlmann (2008); Belloni et al. (2012); Belloni and Chernozhukov (2013); Belloni, Chernozhukov, and Wei (2014); Chernozhukov, Hansen, and Spindler (2015); and Belloni et al. (2016). The method shrinks regression coefficients by penalizing their magnitude and provides a narrow set of important variables, making the results easier to interpret and resolving the problem of multicollinearity; see Meinshausen and Yu (2009). The lasso techniques have also been widely used in the financial economics literature for the prediction of expected returns; see Freyberger, Neuhierl, and Weber (2017); Chinco, Clark-Joseph, and Ye (2019); Kozak, Nagel, and Santosh (2019). The lasso techniques also seem to be the best for both the variable selection and the prediction of the corporate bankruptcy likelihood; see Tian, Yu, and Guo (2015).

We report that relevant contractual loan characteristics, borrower risk characteristics (submitted when applying to the LendingClub platform), and some relevant macroeconomic variables are essential in determining the probability of default of individual loans. Specifically, loan applicants who apply for a longer-term loan (60 months rather than 36 months) exhibit a higher likelihood of default. Similarly, loan applicants who have lower assigned credit score by LendingClub, those who are renters (not homeowners) at the moment of loan application, those who are classified as elementary or machine operators and assemblers in the standard occupation classification, and those who use the loan proceeds to finance medical expenses or small business, exhibit a higher likelihood of default. In contrast, those loan applicants who apply for loans to finance their wedding expenditures, home improvements, and car purchases, and who are classified as managers or professionals exhibit a lower probability of default. Interestingly, although fintech lenders tend to reach out to consumers with low credit scores (below prime consumers), the average default rate (unweighted) based on the LendingClub personal loan platform is found to be only 20 percent for the period 2007–2018 (including periods following the financial crisis); thus, on average, 80 percent of the borrowers did not default.

We note that, while homeownership and occupation are important in determining default risk (controlling for the risk characteristics of the borrowers), lenders cannot freely include these factors in their credit risk and pricing model. For example, lenders may include the homeownership factor in evaluating borrowers' ability to pay, but they would be required to demonstrate that using such a variable does not disadvantage individuals who are members of groups (e.g., because of their race, gender, or age) that are protected under the federal fair lending laws – i.e., the Equal Credit Opportunity Act (ECOA) and the Fair Housing Act (FHA).[8]

[8] Homeownership may be correlated with other characteristics which are prohibited bases under the federal fair lending laws.

In addition, our findings suggest that personal loans used for medical financing or small business financing are more risky than other personal loans, holding the same risk characteristics of the borrowers and economic conditions. Borrowers in need of funding for medical services and small business owners who use personal loans to fund their businesses are more likely to default than other borrowers. This implies that these loans should be segmented out for appropriate risk pricing to be fair to small P2P investors. On the other hand, it may be unfair to leave these borrowers with little access to affordable funding, since illness may not be in their control and since small businesses are so important to local economic growth. The solution to medical financing is beyond the scope of this paper. For small business owners, a similar program currently available to (more established) small businesses through current Small Business Administration (SBA) programs could potentially be extended to cover newer and smaller small businesses, which do not have sufficient business financial history.[9] These small businesses owners have turned to personal loans as their funding sources (as reflected in our personal loan data we collected from the LendingClub consumer platform) to offer nontraditional data about themselves for lenders to evaluate their true creditworthiness.[10]

The rest of the paper is organized as follows. Section 2 presents the literature review. Section 3 discusses the data sources, the data collection process, and a full description of the sample and subsamples. Section 4 presents the empirical approaches and our findings. Section 5 discusses the conclusions and policy implications.

2. The Literature Review

The majority of the fintech lending literature has focused on the impact on consumers in terms of their credit access, fair lending, consumer privacy, etc. Berger and Gleisner (2009) analyze the role of intermediaries in developing the P2P market using about 14,000 observations from the lending platform Prosper. They find that borrowers using these platforms have easier access to financing compared with the standard banking intermediaries. Duarte, Siegel, and Young (2012) use photographs of the borrowers from the Prosper lending platform, and by constructing an algorithm of perceived trustworthiness, they show that the best-perceived borrowers receive the lowest interest rates. Wei and Lin (2016) examine matching mechanisms of supply and demand in

[9] The Small Business Administration (SBA) currently provides support on small business loans, but this is not relevant for the personal loan data that we use in this paper.

[10] For people requiring medical services, the potential solution may not be strictly the financing. Medical debt has been one of the causes for millions of Americans to file for bankruptcy. Consumer credit scores are also likely to be downgraded when their medical debt gets transferred to collection agencies.

the P2P market; they study whether the obtained equilibrium interest rates are optimal ones and whether the choice of the matching mechanism determines the rates of default. They report that the likelihood of loan approval increases and the obtained interest rates are higher when fintech lenders impose a matching mechanism.

Wang and Overby (2018) exploit the timing variation in the approval by the states for LendingClub to operate within the borders and report that regulatory approval is causing higher bankruptcy filing rates in given states. Buchak et al. (2018) examine how the technological advantage of P2P platforms and the regulatory environment impacts the growth of marketplace lending. They report that P2P lenders were more active in refinancing and able to serve more creditworthy customers than traditional banks. Jagtiani and Lemieux (2018), using data from the LendingClub consumer platform, find that fintech lending has penetrated areas that are likely to be underserved by traditional lenders, such as those that have fewer bank branches per capita and in markets with highly concentrated credit card lending. They also find that the portion of LendingClub loans increases in areas where the local economy is not performing well.

Balyuk (2018), using data from the Prosper lending platform, finds that borrowing on the P2P platform eases further access to traditional banking sector products. Hertzberg et al. (2018), using LendingClub data, suggest that borrowers' choice of maturity could serve as the screening mechanism of private information. Vallee and Zeng (2019) model the behavior of the P2P platform and suggest that prescreening of borrowers' financial positions leads to the higher quality of loans offered to investors. Havrylchyk et al. (2019) examine the determinants of consumer demand for fintech loans, using data from the Prosper and the LendingClub consumer platforms. They attribute a rise in P2P lending to the deleveraging of the banks and find that marketplace lending is a substitute for traditional banking. Balyuk and Davydenko (2019) document that marketplace lending has outgrown financial intermediation function, and it is further positioning itself as a gatekeeper in the market for personal financial information.

Fintech P2P lending, which started in the personal lending space, has recently expanded into small business lending (SBL), auto refinancing, and mortgage lending. Jagtiani, Lambie-Hanson, and Lambie-Hanson (2019) examine all mortgage loan applications and originations using Home Mortgage Disclosure Act (HMDA) data and compare mortgage loans across different types of lenders. They find evidence that fintech lenders have higher market shares in areas where consumers have lower credit scores on average. More interesting, they find an increasing share of mortgage loans that are originated by fintech lenders in areas where there was a higher frequency of mortgage denial by traditional lenders in the previous period. Borrowers may have turned to fintech lenders as they had trouble getting credit through the traditional channel.

There have also been studies that explore the roles of nontraditional data used by fintech lenders and the impact on the pricing of credit. Jagtiani and Lemieux (2019) use LendingClub loans specified by the applicants that the proceeds would be used to pay off credit card balances. They compare these LendingClub loans with (loan-level) data from FR Y-14M, which contain traditional credit card loans issued by large CCAR banks. Their results indicate that alternative data and complex modeling have been increasingly used by fintech lenders to more accurately evaluate and price credit risk. Moreover, Jagtiani and Lemieux (2019) find that, after controlling for the borrowers' risk characteristics, borrowers pay significantly less on their fintech loans than what they would have had to pay on their credit card borrowing. For more background on the use of big data, alternative data, and ML by fintech lenders to make faster and better credit decisions, see Jagtiani, Vermilyea, and Wall (2018) and Goldstein, Jagtiani, and Klein (2019).[11]

A few existing studies examine fintech loan performance and default risk. Carmichael (2014) applies a discrete hazard time model to analyze a sample of LendingClub loans issues in the period 2007–2013. He reports that default is determined by the borrower's FICO score, recent credit inquiries, annual income, and loan purposes. Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios (2015), using a sample of 24,449 individual loans obtained from the LendingClub platform during the period 2008–2014, test for default determinants. They report that loan purpose, the applicant's annual income, the current housing situation, and the level of indebtedness are significant in determining loan default. Emekter et al. (2015) study the loan performance of 61,451 LendingClub loans and report that borrowers with high FICO scores and those with a low debt-to-income (DTI) ratio are less likely to default. In addition, Đurović (2017) reports that LendingClub loans with a longer maturity are riskier, while the lowest level of risk is for loan applicants who specify that they would use the loans to pay off credit card balances or for debt consolidation. In this paper, we use a significantly higher dimensional data set in modeling loan performance, and we conduct a more robust analysis using ML techniques. Our results provide deeper insights into fintech loan performance and assess the potential impact on lenders and investors participating in this innovative market.

[11] In addition, Hughes, Jagtiani, and Moon (2019) find that LendingClub became as efficient in lending as the largest U.S. banks (CCAR banks), although LendingClub belongs to a smaller size group as of 2016.

3. The Data and Descriptive Statistics

3.1 The Data Sources

We use data from several sources and merge them appropriately. In summary, all the information about loan characteristics and the borrowers' characteristics come from the LendingClub website. We then match local economic factors that are specific to the borrowers' local community to each loan observation. The most granular level we could match is to the 3-digit zip code because the borrowers' address is reported in a 3-digit zip.[12]

Fintech Loan-Level Data from LendingClub

- Fintech loan-level data are collected from the LendingClub consumer platform, a total of 1,345,549 personal unsecured installment loan observations, with two different maturities (3 years or 5 years).
- LendingClub posts its data on the public website, providing plenty of information about individual loans originated through its consumer platform since its establishment in 2007, with monthly payment updates for each loan. The variables include information on contractual loan characteristics, applicant characteristics, institutional investor characteristics, and other relevant statistics. More details on how these data are used in the analysis are provided in the Appendix.
- Our sample includes all loans originated by the LendingClub consumer platform from 2007 to 2018. The volume was quite trivial in the beginning. Most of the observations are loans originated after the year 2012. We include in the statistics only loans with clear ending resolutions. Consequently, to be included in our analysis, the loan has to be either repaid fully or charged off.
- To avoid potential misreporting of extreme values, we carefully check these variables and trim extreme values when appropriate.

Local Economic Variables from Various Data Sources

- Aside from data provided by LendingClub, we use statistics provided by the Federal Reserve Bank of St. Louis's FRED Economic Data for information on prevailing daily Treasury bill rates.
- The International Standard Classification of Occupations (ISCO-08) is used to classify applicants, based on their employment area. Based on that standard, we classify every applicant into one of the 10 base occupation categories.

[12] There are 929 3-digit zips in the United States compared with more than 6,000 5-digit zips.

- The Internal Revenue Services (IRS) is our data source for variables on the taxable income per county/zip code area of an applicant for available years. As the borrower's address (location) is reported by LendingClub in a 3-digit zip code, we calculate the average taxable income specific to each 3-digit zip.
- The Chicago Board Options Exchange (CBOE) and its Global Markets section provide the daily value of the volatility index (VIX).
- The Bureau of Economic Analysis (BEA) provides current GDP and real GDP variables at the county level. We are able to translate the county-specific GDP into the 3-digit zip level GDP. There are 3,142 counties for the 929 3-digit zips in the U.S.

General Economic Conditions and Market Sentiments

- The Policy Uncertainty website provides a few indices developed first by Baker, Bloom, and Davis (2016) that in various ways show uncertainty levels and macroeconomic environment.
- Finally, the daily level of the returns of the Russell 2000 Index is downloaded from the Policy Uncertainty website. We use this variable to proxy for the overall market performance.[13]

3.2 The Sample Summary Statistics

Table 1 reports the temporal distribution of the sample, primary applicant characteristics, and loan terms. The crucial variable for the study; namely, the default rate of individual loans, stands at 20 percent of approved loans on average of the overall 1,345,549 loans, 268,043 loans defaulted over the period 2007–2018. Jagtiani, Lambie-Hanson, and Lambie-Hanson (2019) show that fintech lenders tend to reach out to those consumers with lower credit scores and lower income (those who are likely to be underserved). On a similar note, Bhanot (2017) observes the behavior of 4,883 first-time online borrowers and concludes that consumers who failed to repay the loan do that primarily because of financial distress. Therefore, one must be careful to not compare fintech loan default rates with a traditional personal loan originated by commercial banks without appropriately controlling for the risk characteristics of the borrowers.

Figures 1 to 4 show more granular characteristics of the charged-off loans from our loan sample from the LendingClub consumer platform over the period 2007–2018 by loan purposes

[13] Korteweg (2019) surveys studies of returns in private equity investment and acknowledges wide usage of the Russell 2000 Index as a comparison benchmark (https://www.ftserussell.com/products/indices/russell-us).

(Figure 1), by the borrower's homeownership or housing situation (Figure 2), by the borrower's rating grades assigned by LendingClub (Figure 3), and by loan rates charged by LendingClub (Figure 4). The figures show that LendingClub's rating grades and loan rates are highly correlated with default risk.

Figure 5 shows the distribution of loan annual percentage rate (APR) for each of the rating grades from A to G. The least risky borrowers (A-rated) pay less than 10 percent APR, and the rate is capped at 36 percent APR for the most risky borrowers (G-rated). Finally, Figure 6 shows that the majority of loans originated by LendingClub in each year has been of the shorter maturity of 36 months rather than 60 months.

In Table 1 Panel A, the temporal distribution of P2P loans shows that fintech loan volume was growing increasingly monotonically during the period 2008–2014. It peaked at 433,872 loans originated in 2014, and then the volume started to decline. This decline in the volume of loans from 2014 primarily reflects the fact that only loans with the exact resolution of the payment are included in our analysis and are not indicating a decline in the overall volume of origination by the LendingClub platform. For example, a loan issued at the end of 2014 with five years of maturity is still not resolved and consequently is not included in our sample.

The key variables that determine the applicant's risk characteristics are reported in Panel B of Table 1. The risk premium, which is calculated as the difference between the interest APR on the loan and the matching Treasury risk-free rate, was monotonically increasing from 2007 until 2013 when it reaches its peak at 14.42 percent and then declines afterward.[14] If we consider microfinance loans as comparable and a predecessor of P2P lending, then the risk levels of fintech loans were lower compared with approximately 30 percent risk premium on microfinance loans as reported in Rosenberg, Gaul, Ford, and Tomilova (2013), lower than the interest rates of credit card mail offers extended to households as reported in Demyanyk and Kolliner (2014) and Adams (2018), and lower than risk-adjusted rates on bank loans as reported in de Roure, Pelizzon, and Thakor (2019). Overall, our data indicate that consumers could potentially benefit from the lower funding cost through fintech loans.

Panel B also reports the various statistics related to a local economic environment in which the borrowers are located. The GDP growth rate ranges from 1.65 percent to 2.59 percent, with an average over the observation period of 2.13 percent. Whether the loan applicant is a homeowner is another key variable; about 50 percent of the applicants for LendingClub loans owned a home,

[14] Figure 5 presents APR distribution by the LendingClub assigned grade.

which is lower than the national average for the general U.S population that is about 60-plus percent as reported in Shiller (2007), 63.70 percent as reported in Goodman and Mayer (2018), and 64.30 percent homeownership rate for the end of 2018 based on the Federal Reserve Bank of St. Louis's FRED Economic Data.[15] About 40 percent of the applicants reported renting. The smaller homeownership ratio for LendingClub borrowers implies that fintech loans might serve as a last resort for nonhomeowners who do not have a home as collateral, although homeownership is not one of the risk factors directly included in LendingClub's models for credit decisions.[16]

The bottom row of Panel B reports the frequency of loans that did not require a verification process to verify income sources by the LendingClub platform. The data show a rising trend of verification from a negligible number in 2007 (where most loans were not verified) to about 30 percent in 2011–2012 and has stayed flat at approximately 30 percent. In Table 1 Panel C, the purposes of the loans are reported with associated frequency. Two categories stand out (i.e., credit card repayment and debt consolidation that together consist of approximately 80 percent of all the loans originated through the platform during 2007–2018). The ratio of loans that are used to pay off credit card balances and for debt consolidation rose in more recent years, to around just under 90 percent starting in 2014–2015. As reported later in this paper, we find that these loans used for debt consolidation or to pay off credit card balances are less likely to default than loans for other purposes.

Table 2 reports descriptive statistics for all variables used in the study. Their number, mean, median, and extreme values are reported. Potential important factors that determine the successful repayment of the loan or its default and charge-off are divided into four categories: contractual loan characteristics, individual borrower risk characteristics (as of the date of loan origination), economic environment factors (which may impact P2P market and the default frequency), and those factors that describe the nature of the involvement of investors and lending institutions in the P2P loan market.

The most important contractual loan characteristics are the amount of the loan, the maturity of the loan (3 years or 5 years) and interest rate of the loan, with an average loan amount of $14,370, average loan maturity of 41.8 months, and average interest rate of 13.37 percent APR.

---

[15] See https://fred.stlouisfed.org/series/RHORUSQ156N.

[16] The homeownership variable is not one of the traditional risk factors lenders commonly use in credit decisions because they tend to be correlated with race or other prohibited bases and therefore could violate fair lending laws. The Equal Credit Opportunity Act (ECOA) and the Fair Housing Act generally prohibit lending practices that have a disproportionately negative impact on a prohibited basis (disparate impact), even though the creditor has no intent to discriminate and the practices appear neutral.

For the set of borrowers' characteristics, these variables are self-reported by the borrower and are increasingly becoming more frequently verified by LendingClub (Balyuk and Davydenko, 2019). They also include employment record, annual income, various financial positions, and credit characteristics as of the application date. On average, 62 percent of the time the applicant has less than 10 years of work experience, with an average income of $75,582, and an average DTI ratio of about 18 percent. It is interesting to note that 32 percent of the applicants were delinquent on other loans within the last two years. An average applicant has almost six years (average 70.5 months) of credit record, with 11.58 credit lines on average and a 52.8 percent credit utilization ratio. The collection of all this personal information by marketplace lending institutions is important; the financial literature recognizes that personal financial information and experience affect risk-taking levels; see Koudijs and Voth (2016). Since the LendingClub consumer platform requires that consumers have FICO scores of at least 640, those who do not have credit scores and those with thin files are not eligible to apply on the platform. The primary benefit to these consumers seems to be the use of alternative data by fintech lenders, which allow them to access credit at a lower cost.

A set of macroeconomic variables explains the environment surrounding the local market in the period under observation. Table 2 reports the risk premium of the loans, average county per capita household income, county/zip code area average income, county GDP rates and levels, volatility levels at the loan issuance date, policy uncertainty indices as well as monthly returns of the equity markets.[17] On average, personal loans originated through the LendingClub consumer platform carry a risk premium of 12.94 percent, which is much smaller than credit card rates but also could be a good investment option compared with other investment alternatives. One concern among investors has been whether the default rate on fintech loans would suddenly jump during bad times. Some borrowers are expected to be more adversely affected during a recession than others. In this paper, we explore characteristics of the borrowers who are more likely to default. Finally, Table 2 reports Institutional Investor Characteristics, indicating that 50 percent of these loans are entirely funded by institutional investors, rather than small individual investors.[18]

[17] Mollick (2014) shows that geography is an important factor in the fundraising success of marketplace lending.

[18] Balyuk and Davydenko (2019) report that about 90 percent of issued P2P loans are now funded by institutional investors. Our reporting of lower percentage is primarily due to the selection of only resolved loans by the end of 2018 in analysis. Kräussl et al. (2019) attributes an increase in the interest of institutional investors for the P2P market to high risk-adjusted performance of portfolios composed from individual loans originated on the LendingClub platform.

3.3 Characteristics of the Subsamples (Based on Loan Payment Outcome)

We divide the loan samples into two segments, based on their payment performance: default and charge-off (268,043 loans) and paid off in full (1,077,550 loans). Table 3 presents descriptive statistics for these two subsamples. We compare their means for statistical differences, where the t-statistics and p-values are reported in the last three columns, with a corresponding number of stars (one, two, and three) indicating significance level (at the 10 percent, 5 percent, and 1 percent, respectively).

On average, larger loans are associated with a higher probability of default (i.e.; loans that were charged off are larger than the loans that were paid off in full, with the average origination amount of $15,475 for defaulted loans, relative to $14,119 for good loans. As expected, charged-off loan applicants were identified essentially as being riskier, and they are required to pay a higher risk premium, with average contractual interest rates of 15.75 percent compared with 12.78 percent for those that were paid off in full. The difference between these two rates is likely to be even more significant if accounting for fees, which are usually higher for more risky borrowers.

In addition, longer-term loans are associated with a higher risk of default. One explanation is that the longer *maturity* leads to the long interval of exposure to the various shocks to individual financial position. Specifically, 40 percent of defaulted loans had a five-year maturity (60 percent with a three-year maturity) relative to only 20 percent of nondefaulted loans being five-year loans (80 percent with a three-year maturity).[19] In addition to their preference for longer maturity loans, borrowers who default on the loans also have lower self-reported *income* $69,678 (compared with $77,059), exhibit a higher *DTI* ratio of 20 percent (relative to 17.62 percent), have a higher percentage of recent *delinquencies* of 35 percent (relative to 31 percent), and had more credit inquiries in the last six months, with average of 0.81 *inquiry* (compared with 0.66). Defaulters exhibit a shorter time since the previous delinquency of 33.68 months (compared with 34.36 months), have a higher average number of open credit lines of 11.93 (compared with 11.50), have a higher average number of derogatory accounts of 0.24 (compared with 0.21 accounts), have a smaller total revolving line of $15,293 (compared with $16,377), with a higher average *credit utilization ratio* of 56 percent (compared with 52 percent), pay more late fees averaging 12 percent (compared with only 2 percent), and have more credit accounts of 5.29 on average (compared with 4.51 accounts). Those who defaulted have a statistically higher percentage of all bankcard accounts with more than 75 percent utilization ratio, at 51.89 percent of all cards (compared with 45.41 percent of all cards with at least a 75 percent utilization ratio). Finally, they also have a higher

[19] Figure 6 reports temporal distribution of loans based on the maturity choice.

number of public recorded bankruptcies of 0.15 (compared with 0.13) and a higher number of tax liens at 0.06 (compared with 0.05). These statistics could be useful in designing a loan program that would help alleviate risk to small P2P investors. For example, a combination of DTI, annual income, and credit utilization should be used in determining the loan amount and maturity of the loans that consumers are given.

Table 3 also reports how subsamples differ concerning macroeconomic indicators. A subsample of loan applicants that eventually defaulted on their loans exhibits almost a 3 percent higher risk premium than applicants who paid off the loans in full — a 15.57 percent risk premium compared with 12.59 percent for the nondefaulted segment. In addition, defaulted borrowers live in counties with lower current and real GDP levels ($66,000 versus $67,800), counties with lower GDP annual growth (2.22 percent versus 2.26 percent), and they live in a lower-income community.

4. The Empirical Approach and Findings

To set up a baseline for the further tests, we examine the likelihood of the default of individual loans in which the dependent variable is a dummy that represents the status of loan payment. We code the dependent variable as being **charged-off** equal to 1 if the applicant defaulted on the loan and the loan is consequently charged off. Otherwise, the variable takes value 0 if the loan is fully paid within the observation period. It might be that the maturity of the loan has been modified, due to early payment or extension. Since our sample only consists of loans with clear ending resolutions, the dependent variable is also well defined for the modified loans.

At first, we apply logistic regression approach in which the dependent variable charge-off is regressed on the set of independent variables that were reported in Table 3. This approach is standard in the literature examining personal or corporate defaults and enables us to determine which risk characteristics significantly impact the likelihood of default for the sample of our fintech loans; see Bastos (2010). We report the results of these logistic regressions in Table 4.

We recognize that there are limitations under the logistic regression approach because of its high dimensionality (with more than 100 independent variables) and that they potentially blur the results. To address this, we further strengthen our approach by introducing the lasso selection method, which was initially developed in Tibshirani (1996), and consequently further enhanced by Belloni and Chernozhukov (2013), Belloni et al. (2012), Belloni et al. (2014), and Belloni et al.

(2016).[20] Using the lasso selection method, we could select a set of variables that may be more important in determining default of individual fintech loans. The selected variables and coefficients are reported in Table 5. Once the lasso procedure selects a set of variables that are most likely to have an impact on defaults, we use logistic regression to reestimate the coefficients based on a smaller set of independent variables. The new set of logistic regression results obtained from the lasso selection method are reported in Table 6.

4.1 The Basic Logistic Regression Analysis

Determining the likelihood of loan default is an old and interesting question in economics. The recent financial innovations such as P2P lending enabled by highly powerful intermediary electronic platforms would allow researchers to have testing ground unseen before. Our knowledge of the factors that are important in determining the loan performance of this market is critical to its long-term success in expanding credit to those who have been underserved.

As a first step to determining the determinants of default, we use a dummy variable on the status of loan payment as the dependent variable and a set of contractual loan characteristics, individual applicant characteristics, macroeconomic variables, and institutional investor risk characteristics as independent variables. Our final sample used in the logistic regression consists of 1,064,490 loans, in which every independent variable has an observed value (not missing). Our sample observations are representative and significantly more comprehensive than in previous studies.

Table 4 presents the results of our logistic regression. We discuss variables here that show high statistically significant impact on the likelihood of default on these individual loans. When it comes to contractual loan characteristics, the most important determinant of default seems to be loan maturity. We find that people who decide to take out a longer-term loan (five-year maturity) are more likely to default, even after controlling for other risk characteristics and economic environment; these results are consistent with those found in Hertzberg et al. (2018). Again, the logistic regression results also confirm that the borrowers who were charged lower interest rates by LendingClub are less risky, and they are less likely to default, even after controlling for all the other relevant risk characteristics. This result agrees with Ryan and Zhu (2018) findings, based on data from the Prosper lending platform, noting that Prosper during the post-2013 period was

[20] Machine learning statistical techniques are offering strong additional power in analyzing behavior of economic agents; see Varian (2014), Mullainathan and Spiess (2017), Athey (2018), Björkegren and Grissen (2019).

better ex ante in judging the most appropriate interest rate to charge the borrowers that later became delinquent and defaulted.

Loan purposes could also play a role in determining default. We find that the probability of default increases if the loan is taken to fund a small business, while the likelihood of default decreases if the purpose of the loan is to fund wedding expenditures. The higher the level of reported collections in the last year (excluding medical expenses), the higher the probability of default. In addition, borrowers who applied for the loans individually have a higher chance of default than a joint loan application.

Borrower characteristics at the moment of loan origination have implications on loan performance. Borrowers who belong to the category of professionals have a lower likelihood of default. Although with the lower degree of statistical significance, similar results are reported for applicants classified as associate professionals. Workers belonging to classifications as elementary occupations, machine operators and assemblers, service and sales workers, and craft and trade workers all exhibit a higher likelihood of default, even after controlling for other risk characteristics and economic environment.

Housing status impacts the likelihood of default as well. Borrowers who are homeowners (or have an outstanding mortgage) have a lower likelihood of default, while renters are associated with an increased likelihood of the default. Borrowers with higher DTI ratio or with a higher number of credit inquiries in the last two years are more likely to default. A variable with the highest coefficient of all in the analysis (1.80) is one documenting whether an applicant used to pay late fees. Apparently, the likelihood of default is significantly higher for applicants who in the recent past used to pay late fees on their credit accounts.

Macroeconomic conditions also impact the likelihood of default on behalf of applicants. The macroeconomic variable with a high coefficient is the return on the Russell 2000 Index, implying that the positive return on the index is associated with an increased likelihood of default overall, after controlling for all the risk characteristics of loans and borrowers. We also find a higher likelihood of default for loans financed entirely by institutional investors.

Most important, our results also show that LendingClub's models used to predict the borrower's likelihood of default are accurately reflected in the credit rating assigned by LendingClub: grade A (best) to G (worst). We find that applicants who were classified in the top score grade A, B, and C exhibit a lower probability of default. The size of the coefficient on these three subscores increase monotonically, indicating that borrowers with an A-rated score have a lower probability of default than the borrowers with a B-rated score, and consequently, the borrowers with a B-rated score are less likely to default compared with those who are C-rated. Our findings confirm that LendingClub has the right risk-assessment tools when evaluating borrowers and their likelihood of default. These results are also consistent with Jagtiani and Lemieux (2019), who also find that, while the rating grades assigned by LendingClub are accurate in predicting loan default, these rating grades have minimal correlation with borrowers' FICO scores, which have been primarily used for a credit decision, especially for credit card loan applications.

4.2 The Lasso Selection Method and Post-Lasso Logistic Regression Analysis

To further improve the predictive accuracy and to improve the variable selection accuracy (with about 100 independent variables), we use the lasso selection method to help us streamline the variable set. Of the initial set of 99 independent variables (used earlier in Table 4), the lasso selection method selects only 58 variables to be included. These variables are grouped in an identical way as in the summary statistics tables and presented in Table 5.

We then apply the logistic regression that includes only the selected 58 independent variables. The final regression has 1,095,012 individual loans, in which the dependent variable **charged-off** is regressed against 58 lasso method selected independent variables that describe contractual loan characteristics, borrower characteristics, macroeconomic conditions, and involvement of institutional investors. The results are reported in Table 6.

Most of the variables that are significant in the baseline logistic regression (Table 4) continue to be significant (Table 6), but there are notable improvements. Reported results on the contractual loan characteristics show similar coefficients and directions as previously reported, while there is an increase in the level of information provided on the impact of the loan purpose variables on the likelihood of default. Namely, almost all loan purposes are statistically significant at the 1 percent level. The default likelihood increases if the loans are taken with the purpose of financing home improvement, major purchases, medical-related expenses, small business-related costs, and moving expenses. On the contrary, the likelihood of default decreases when the loans are used to finance car purchases, repay credit card debt, house purchases, and wedding expenditures. Similar to the previous results reported in Table 4, when the applicant exhibits an increasing number of collections in the last year, excluding medical expenses, and when the loan is applied for individually, the likelihood of default increases.

The borrower's occupation and homeownership continue to be important.[21] Unlike renters, homeowners are less likely to default. Most of the variables that are seen as having negative implications on creditworthiness — such as the DTI ratio (prior to application), number of delinquencies two years prior to the issuance of the loan, the number of credit inquiries prior to the loan, and the number of derogatory public records —are associated with an increased likelihood of default. As in the original regression, the indicator of whether an applicant used to pay late fees before obtaining loans shows a high economic significance and increases the likelihood of default. Similarly, an increase in the level of reported public bankruptcies increases the likelihood of default.

In post-lasso selection regression, macroeconomic variables exhibit similar results as before. Risk premium and the returns on the Russell 2000 Index (in the month of the loan application) increase the likelihood of default. Although the magnitude is not strong, higher volatility of the equity options in the month of loan origination seems to be associated with a lower probability of default. The improvement in the growth of GDP is related to a lower likelihood of loan default. Finally, the likelihood of default decreases if the higher amount of the loan is financed by institutional investors.

Once again, we find that the ratings assigned by LendingClub remain important factors in determining loan performance and default likelihood. The results are robust in supporting the use of alternative data and appropriate ML analysis in credit decisions. Using alternative data in credit decisions has become a new trend in the financial landscape and regulations.

On the other hand, there are concerns about consumer privacy and fair lending associated with the use of these data and ML algorithms in credit decisions, and this has become a popular topic of debate. Regulators attempt to strike the right balance in encouraging fintech innovations while providing consumer privacy and fair lending protections. Several important questions have remained unanswered. What data about consumers could be shared? Who owns the consortium data? Who is responsible when information about consumers is shared with outside parties and causes damages to consumers? Which alternative data could be used to those who have been denied credit and which could be used to expand credit to the underserved? Are consumers

[21] The likelihood of default decreases when an applicant is classified as belonging to the following groups: managers, professionals, technicians, and associate professionals. Opposite results are reported for the following employment occupation classifications: elementary occupations, machine operators, services and sales workers, and craft and related trade workers as they exhibit a higher likelihood of default and consequent loan charge-offs.

becoming too leveraged because of additional funding access provided through the fintech platform?

4.3 Robustness Checks for Sample Selection Bias

Our observed sample for the analysis includes only loans that have a clear ending resolution and consequently were either charged off or fully paid off loans. This potentially creates an issue of the sample selection as the loans included in the period after 2013 for five-year loans and loans included after 2015 for three-year loans may be overrepresenting defaulted loans. To address the issue of the sample restriction, we conduct two procedures. The results are in Table 7.

First, we create a subsample including loans only in which we have ending outcome as the only possibility for all of them. This subsample contains 651,555 loans that are free of possible selection bias, or about 59.51 percent of the total loans used to the full-sample analysis, as reported in Table 6. Second, we apply a Heckman-type correction for sample selection – see Heckman (1979) and Marchenko and Genton (2012). This method is very popular for linear models, and it has been extended to binary choice models (Van de Ven and Van Pragg (1981)), and the application is available via standard software (De Luca and Perotti (2011)). The binary regression model is complemented with a selection equation, having a binary dependent variable equal to 1, if the loan has "clear ending resolution" and 0 otherwise. The selection equation is estimated from a much larger sample, including the loans that were still unresolved at the end of the observation period, yielding a total of 2,023,934 observations. The regressors used in the selection equation are the duration (between issuing the loan and the end of the observation period), the time to maturity, and the interest rate. It turns out that the results are pretty robust to the specification of the regressors in the selection equation.

In the first four columns of Table 7, we present the original results (taken from Table 6); in the next four columns, the results based on the subsample of 651,555 loans as described above, and in the last four columns, the results with the Heckman correction. The only significant difference is for the dummy variable "LC credit rating F (Y/N)." This variable is positively related to the likelihood of default estimating the model from the full sample and negatively related to the likelihood of default when using the subsample. In the subsample analysis, our loan observations stop at the 2013 origination year (for five-year loans) and stop at the 2015 origination year (for three-year loans). Thus, the subsample analysis is based on older loans, which could be driving the differences in the findings. In the early years, LendingClub was comparatively closer in assigning a credit score to standard scoring systems as with FICO than in later years (after 2015 origination);

see Jagtiani and Lemieux (2019). For the model with the Heckman correction, all estimation results except one are very similar to those reported in Table 6 (in the first four columns). We conclude that the detected important drivers for fintech loan default are not significantly affected by the selecting mechanism to construct our sample.

5. The Conclusions and Policy Implications

This paper examines the default determinants of loans in P2P lending settings using loan-level data from the LendingClub consumer platform during the period 2007–2018. We examine a number of factors that may potentially be important in determining fintech loan defaults. We started with a comprehensive set of contractual loan characteristics, borrower characteristics, and macroeconomic variables as independent variables (total about 100 independent variables), using logistic regression analysis, and find interesting results indicating potential important roles of nontraditional data in credit decisions.

To further validate results and explore whether an alternative set of explanatory variables would become significant in determining fintech defaults, we also conduct an alternative analysis using a more robust methodology, using lasso selection methods to narrow our initial comprehensive set of 100 explanatory variables before applying the logistic regression analysis. The process reduced the number of independent variables from 100 to 58 variables. The logistic regression analysis was applied on these selected variables (post-lasso selection), and the results are very similar (but with additional insights around the impact of loan purposes on default) to the initial results from the previous step.

Overall, we find that borrowers who choose to take loans with a longer maturity (five-year loans), those with lower assigned credit scores, nonhomeowners, and those belonging to elementary or machine operators and assemblers (not a manager or executive) are more likely to default. In addition, after controlling for borrowers' risk characteristics, we find that loan purposes also play a role; for example, loans that are used to finance medical expenses or small business costs (rather than paying off credit card balances or funding wedding expenses) exhibit a higher likelihood of default. Borrowers who use loan proceeds to finance a wedding, house purchase-related, and car purchases experience lower likelihood of default. More important, the results show that LendingClub's own rating scores are highly accurate in predicting defaults, and it is significant even after controlling for the obvious risk characteristics of the borrowers, loan characteristics, and the local economic factors. The results are also robust to alternative empirical approaches, with and without the lasso selection process.

Our findings are consistent with an argument that the use of alternative data in credit decision could result in creditworthy "subprime" borrowers being able to access credit at a much lower cost than they otherwise would. Jagtiani and Lemieux (2019) find that most of the "invisible prime" borrowers, who have been rated poorly by the traditional credit scoring process, have a very low default probability that is similar to the default probability of (traditional) super-prime borrowers, suggesting that regulators could consider allowing lenders to use certain alternative data to identify good borrowers from the traditionally subprime pool as a way to expand credit access to low-score borrowers.

Our results overall suggest that homeownership and occupation are important in determining default, controlling for credit ratings, and other risk factors. However, we note that such variables cannot be incorporated into underwriting or pricing models without careful examination by the lender to demonstrate that they do not result in disparate treatment that would adversely affect members of groups protected under the nation's federal fair lending laws — i.e., the Equal Credit Opportunity Act (ECOA) and the Fair Housing Act (FHA).[22]

More broadly, it has become more common now that lenders would subscribe to data analytic services from outside vendors. Potential violations of privacy and fair lending laws might potentially lie inside the "black box" provided by third-party vendors. While third-party vendor risk has always been a concern among bank regulators, the nature of the risk has changed significantly in the new financial landscape, where many of the credit decisions for loans that are on banks' balance sheet may be largely determined by the models used by data aggregators and AI vendors. The uncertainty may be greater for those nonbank lenders that are not subject to regular in-depth banking examinations (by federal and/or state banking regulators) and may be too small to fall under the supervision of the Consumer Financial Protection Bureau.

In addition, our overall results indicate that certain loan purposes (controlling for risk characteristics of the borrowers and economic conditions) such as medical financing and small business financing are riskier than other loan purposes. This would imply that they should be segmented out for more appropriate risk evaluation and fair pricing for P2P investors. It is interesting to observe that, in late 2014, LendingClub also established a separate lending platform that deals with small business loans only, aiming to serve those small firms that cannot access business loans through the SBA program but requiring larger loans than what they could get on the consumer platform. That small business platform later became part of the Opportunity Funds in

[22] The concern is that such variables may be correlated with race, age, color, national origin, religion, or gender.

2019.[23] In addition, a separate lending platform was later established to serve those with specific medical needs (financing through a doctor's office), subject to a different process and credit risk models.

As for policy implications, we note that borrowers in need of funding for medical services and for small businesses are more risky than other borrowers, and they may have difficulties getting access to affordable funding sources. While the SBA currently provides support for more established small businesses that have some track records, newer and smaller small business owners (without sufficient business credit history) have had to turn to personal loans as their funding sources, as reflected in our personal loan data from the LendingClub consumer platform. Some government support has begun and could be expanded to support small businesses. For example, more public–private partnerships with fintech firms could be expanded, such as the partnership in 2019 between LendingClub, the (nonprofit) Community Development Financial Institution (CDFI), and Funding Circle (another fintech small business lending platform). More programs like this one would help to enhance access to affordable credit for small businesses without imposing excessive credit risk to small P2P investors.

In closing, we note that there remain uncertainties around fintech credit decisions, given the rapid advance in technology. Some small community banks find themselves in fierce competition with fintech lenders in their own local community. Others have benefited greatly through the various partnership programs with fintech platforms, as a way to digitize their credit decisions without a large investment in their own in-house technology. Investors are interested in understanding whether fintech lenders would replace traditional banks or become part of the traditional bank holding companies. The recent announcement of LendingClub to acquire Radius Bancorp is consistent with a belief that fintech lending and retail banking would likely converge over time. This is apparently an opportune time for researchers to further explore the impact of fintech on consumers, lenders, fintech investors, and the financial system overall.

[23] See more details at Bloomberg News (April 23, 2019); https://www.bloomberg.com/press-releases/2019-04-23/lendingclub-partners-with-opportunity-fund-and-funding-circle-increasing-financial-inclusion-and-small-businesses-access-to-c.

References

Adams, R. M. (2018). "Do Marketplace Lending Platforms Offer Lower Rates to Consumers?" (2018-10-22). Board of Governors of the Federal Reserve System (U.S.). https://doi.org/10.17016/2380-7172.2268.

Athey, S. (2018). "The Impact of Machine Learning on Economics," in *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Baker, S. R., N. Bloom, and S.J. Davis. (2016). "Measuring Economic Policy Uncertainty." The Quarterly *Journal of Economics* 131(4), 1593–1636. https://doi.org/10.1093/qje/qjw024.

Balyuk, T. (2018). "Financial Innovation and Borrowers: Evidence from Peer-to-Peer Lending." Rotman School of Management Working Paper (2802220). http://dx.doi.org/10.2139/ssrn.2802220.

Balyuk, T., and S. A. Davydenko. (2019). "Reintermediation in FinTech: Evidence from Online Lending. http://dx.doi.org/10.2139/ssrn.3189236.

Bastos, J. A. (2010). "Forecasting Bank Loans L-Given-Default." *Journal of Banking & Finance*, 34(10), 2510–2517. https://doi.org/10.1016/j.jbankfin.2010.04.011.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. (2012). "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica* 80(6), 2369–2429. http://dx.doi.org/10.3982/ECTA9626.

Belloni, A., and V. Chernozhukov. (2013). "Least Squares After Model Selection in High-Dimensional Sparse Models." *Bernoulli*, 19(2), 521–547. http://dx.doi.org/10.3150/11-BEJ410.

Belloni, A., V. Chernozhukov, and C. Hansen. (2014). "High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives* 28(2), 29–50. https://doi.org/10.1257/jep.28.2.29.

Belloni, A., V. Chernozhukov, and Y. Wei. (2016). "Post-Selection Inference for Generalized Linear Models with Many Controls." *Journal of Business & Economic Statistics* 34(4), 606–619. https://doi.org/10.1080/07350015.2016.1166116.

Berger, S.C., and F. Gleisner. (2009). "Emergence of Financial Intermediaries in Electronic Markets: The Case of Online P2P Lending." *BuR Business Research Journal* 2(1): May 2009. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1568679.

Bhanot, S. P. (2017). "Cheap Promises: Evidence from Loan Repayment Pledges in an Online Experiment." *Journal of Economic Behavior & Organization* 140, 246–266. https://doi.org/10.1016/j.jebo.2017.04.007.

Bjorkegren, D., and D. Grissen. (2019). "Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment." The World Bank, Policy Research Working Paper WPS9074. December 6, 2019.

Bloomberg News. (2019). “LendingClub Partners with Opportunity Fund and Funding Circle, Increasing Financial Inclusion and Small Businesses’ Access to Credit.” April 23. https://www.bloomberg.com/press-releases/2019-04-23/lendingclub-partners-with-opportunity-fund-and-funding-circle-increasing-financial-inclusion-and-small-businesses-access-to-c.

Braggion, F., A. Manconi, A., and H. Zhu. (2018). “Can Technology Undermine Macroprudential Regulation? Evidence from Peer-to-Peer Credit in China.” http://dx.doi.org/10.2139/ssrn.2957411.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru. (2018). “Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks.” *Journal of Financial Economics* 130(3), 453–483. https://doi.org/10.1016/j.jfineco.2018.03.011.

Carmichael, D. (2014). “Modeling Default for Peer-to-Peer Loans.” http://dx.doi.org/10.2139/ssrn.2529240.

Chernozhukov, V., C. Hansen, and M. Spindler. (2015). “Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach.” *Annual Review of Economics* 7(1), 649–688.

Chinco, A., A. D. Clark-Joseph, and M. Ye. (2019). “Sparse Signals in the Cross-Section of Returns.” *The Journal of Finance* 74(1), 449–492. https://doi.org/10.1111/jofi.12733.

Danisewicz, P., and I. Elard. (2018). “The Real Effects of Financial Technology: Marketplace Lending and Personal Bankruptcy,” working paper on SSRN, July 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3208908.

De Luca, G. (2008). “SNP and SML Estimation of Univariate and Bivariate Binary-Choice Models. “*The Stata Journal 8*(2), 190–220. https://doi.org/10.1177/1536867X0800800203.

De Luca, G., and V. Perotti. (2011). “Estimation of Ordered Response Models with Sample Selection.” *The Stata Journal* 11(2), 213–239. https://doi.org/10.1177/1536867X1101100204.

Demyanyk, Y., and D. Kolliner. (2014). “Peer-to-Peer Lending Is Poised to Grow.” Federal Reserve Bank of Cleveland *Economic Trends*. August 14, 2014.

De Roure, C., L. Pelizzon, and A. V. Thakor. (2019). “P2P Lenders versus Banks: Cream Skimming or Bottom Fishing?” http://dx.doi.org/10.2139/ssrn.3174632.

Duarte, J., S. Siegel, and L. Young. (2012). “Trust and Credit: The Role of Appearance in Peer-to-Peer Lending.” *The Review of Financial Studies* 25(8), 2455–2484, https://doi.org/10.1093/rfs/hhs071.

Đurović, A. (2017). “Estimating Probability of Default on Peer to Peer Market–Survival Analysis Approach.” *Journal of Central Banking Theory and Practice* 6(2), 149–167. http://dx.doi.org/10.1515/jcbtp-2017-0017.

Emekter, R., Y. Tu, B. Jirasakuldech, and M. Lu. (2015). “Evaluating Credit Risk and Loan Performance in Online Peer-to-Peer (P2P) Lending.” *Applied Economics* 47(1), 54–70. https://doi.org/10.1080/00036846.2014.962222.

Freyberger, J., A. Neuhierl, and M. Weber. (2017). "Dissecting Characteristics Nonparametrically" (w23227). National Bureau of Economic Research. https://doi.org/10.3386/w23227.

Funk, B., D. Buerckner, M. Hilker, F. Kock, M. Lehmann, and P. Tiburtius. (2011). "Online Peer-to-Peer Lending — A Literature Review." *The Journal of Internet Banking and Commerce* 16(2), 1–18. https://www.researchgate.net/publication/236735575_Online_Peer-to-Peer_Lending--A_Literature#fullTextFileContent.

Goldstein, I., J. Jagtiani, and A. Klein. (2019) "Fintech and the New Financial Landscape" Bank Policy Institute (BPI): Banking Perspectives, Q1 Volume 7, March 2019. https://www.bankingperspectives.com/fintech-and-the-new-financial-landscape/.

Goodman, L. S., and C. Mayer. (2018). "Homeownership and the American Dream." *Journal of Economic Perspectives* 32(1), 31–58 http://dx.doi.org/10.1257/jep.32.1.31.

Havrylchyk, O., and C. Mariotto, T. Rahim, and M. Verdier. (2019). "What Has Driven the Expansion of the Peer-to-Peer Lending?" (August 26). http://dx.doi.org/10.2139/ssrn.2841316.

Heckman, J. J. (1979). "Sample Selection Bias as a Specification Error." Econometrica: *Journal of the Econometric Society*, 153–161. https://www.jstor.org/stable/1912352.

Hertzberg, A., A. Liberman, and D. Paravisini. (2018). "Screening on Loan Terms: Evidence from Maturity Choice in Consumer Credit." *The Review of Financial Studies* 31(9), 3532–3567. https://doi.org/10.1093/rfs/hhy024.

Hughes, J.P., J. Jagtiani, and C. Moon. (2019). "Consumer Lending Efficiency: Commercial Banks Versus a Fintech Lender." FRB Philadelphia Working Paper 19-22 (April).

Jagtiani, J., and C. Lemieux. (2019). "The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform." *Financial Management* Winter 2019, 48(4), 1009–1029. http://dx.doi.org/10.21799/frbp.wp.2018.15.

Jagtiani, J., and C. Lemieux. (2018). "Do Fintech Lenders Penetrate Areas That Are Underserved by Traditional Banks?" *Journal of Economics and Business*, June 2018. https://www.sciencedirect.com/science/article/pii/S0148619518300390.

Jagtiani, J., L. Lambie-Hanson, and T. Lambie-Hanson. (2019). "Fintech Lending and Mortgage Credit Access, Federal Reserve Bank of Philadelphia Working Paper 19-47 (November). https://www.philadelphiafed.org/-/media/research-and-data/publications/working-papers/2019/wp19-47.pdf.

Jagtiani, J., T. Vermilyea, and L. Wall. (2018). "The Roles of Big Data and Machine Learning in Bank Supervision." The Clearing House: *Banking Perspectives* Quarter 1.

Klafft, M. (2008). "Peer to Peer Lending: Auctioning Microcredits over the Internet." Proceedings of the International Conference on Information Systems, Technology and Management, A. Agarwal, R. Khurana, editors, IMT, Dubai. https://ssrn.com/abstract=1352383.

Korteweg, A. (2019). "Risk Adjustment in Private Equity Returns." *Annual Review of Financial Economics* 11, 131–152.

Koudijs, P., and H. J. Voth. (2016). "Leverage and Beliefs: Personal Experience and Risk-Taking in Margin Lending." *American Economic Review* 106(11), 3367–3400. http://dx.doi.org/10.1257/aer.20140259.

Kozak, S., S. Nagel, and S. Santosh. (2019). "Shrinking the Cross-Section." *Journal of Financial Economics.* https://doi.org/10.1016/j.jfineco.2019.06.008.

Kräussl, R., Z. Kräussl, J. M. Pollet, and K. Rinne. (2018). "The Performance of Marketplace Lenders: Evidence from LendingClub Payment Data." http://dx.doi.org/10.2139/ssrn.3240020.

Marchenko, Y. V., and M. G. Genton. (2012). "A Heckman Selection-T Model." *Journal of the American Statistical Association* 107(497), 304–317. https://doi.org/10.1080/01621459.2012.656011.

Meier, L., S. Van De Geer, and P. Bühlmann. (2008). "The Group Lasso for Logistic Regression." *Journal of the Royal Statistical Society: Series B* (Statistical Methodology) 70(1), 53–71. https://doi.org/10.1111/j.1467-9868.2007.00627.x.

Meinshausen, N., and B. Yu. (2009). "Lasso-Type Recovery of Sparse Representations for High-Dimensional Data." *The Annals of Statistics* 37(1), 246–270. https://projecteuclid.org/euclid.aos/1232115934.

Mollick, E. (2014). "The Dynamics of Crowdfunding: An Exploratory Study." *Journal of Business Venturing* 29(1), 1–16. https://doi.org/10.1016/j.jbusvent.2013.06.005.

Mullainathan, S., and J. Spiess. (2017). "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31(2), 87–106. http://dx.doi.org/10.1257/jep.31.2.87.

Philippon, T. (2016). "The Fintech Opportunity" (w22476). National Bureau of Economic Research. https://doi.org/10.3386/w22476.

Rosenberg, R., S. Gaul, W. Ford, and O. Tomilova. (2013). "Microcredit Interest Rates and Their Determinants: 2004–2011." In *Microfinance 3.0* (69–104). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41704-7_4.

Ryan, S. G., and C. Zhu. (2018). "Fintech Isn't So Different from Traditional Banking: Trading Off Aggregation of Soft Information for Transaction Processing Efficiency," http://dx.doi.org/10.2139/ssrn.3212902.

Serrano-Cinca, C., B. Gutierrez-Nieto, and L. López-Palacios. (2015). "Determinants of Default in P2P Lending." *PloS one* 10(10), e0139427. https://doi.org/10.1371/journal.pone.0139427.

Shiller, R. J. (2007). "Understanding Recent Trends in House Prices and Home Ownership (w13553). National Bureau of Economic Research. http://dx.doi.org/10.3386/w13553.

Tang, H. (2019). "Peer-to-Peer Lenders versus Banks: Substitutes or Complements?" *Review of Financial Studies*. Forthcoming. https://doi.org/10.1093/rfs/hhy137.

Tian, S., Y. Yu, and H. Guo. (2015). "Variable Selection and Corporate Bankruptcy Forecasts." *Journal of Banking & Finance* 52, 89–100. https://doi.org/10.1016/j.jbankfin.2014.12.003.

Tibshirani, R. (1996). “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society*. Series B (Methodological), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Vallee, B., and Y. Zeng. (2019). “Marketplace Lending: A New Banking Paradigm?” *The Review of Financial Studies* 32(5), 1939–1982. https://doi.org/10.1093/rfs/hhy100.

Van de Ven, W. P. M. M., and B. M. S. Van Pragg. (1981). “The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection.” *Journal of Econometrics* 17: 229–252. https://doi.org/10.1016/0304-4076(81)90028-2.

Varian, H. R. (2014). “Big Data: New Tricks for Econometrics.” *Journal of Economic Perspectives* 28(2), 3–28. http://dx.doi.org/10.1257/jep.28.2.3.

Wang, H., and E. M. Overby. (2018). “How Does Online Lending Influence Bankruptcy Filings?” Georgia Tech Scheller College of Business Research Paper (17-20). http://dx.doi.org/10.2139/ssrn.2958916.

Wei, Z., and M. Lin. (2016). “Market Mechanisms in Online Peer-to-Peer Lending.” *Management Science* 63(12), 4236–4257. https://doi.org/10.1287/mnsc.2016.2531.

Table 1: Loan Temporal Distribution

This table describes the sample that consists of 1,345,593 loan applications from LendingClub consumer platform during the period 2007–2018. Panel A presents statistics per year on the total number of approved loans and on the outcomes of these loans. Panel B presents a summary of the main variables for each year. The *risk premium* is defined as the difference between the rate of interest obtained at the Data Lending and T-bill rate. Economic policy disagreement is derived from the Political Uncertainty Index. *GDP growth rate* represents the average annual growth rate of the U.S. economy. The rest of the variables represents major applicant characteristics in terms of income, debt-to-income ratio, and house ownership status. Panel C summarizes loan purposes as classified by Data Lending.

Panel A: Sample Temporal Distribution

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
Number of Loans	603	2393	5281	12537	21721	53367	132602	433872	333475	221932	112176	15634	1345593
Percent	0.04	0.18	0.39	0.93	1.61	3.97	9.85	32.24	24.78	16.49	8.34	1.16	100
Loans charged off	158	496	723	1757	3297	8644	20939	80656	71377	57014	21978	1004	268043
Percent	0.06	0.19	0.27	0.66	1.23	3.22	7.81	30.09	26.63	21.27	8.20	0.37	100.00
Loans paid off	445	1897	4558	10780	18424	44723	111663	353216	262098	164918	90198	14630	1077550
Percent	0.04	0.18	0.42	1.00	1.71	4.15	10.36	32.78	24.32	15.30	8.37	1.36	100.00
Loan default rate	0.26	0.21	0.14	0.14	0.15	0.16	0.16	0.19	0.21	0.26	0.20	0.06	0.20

Panel B: Major Applicant Characteristics

Risk premium	8.14	10.66	12.30	11.85	12.18	13.55	14.42	13.55	12.43	13.13	13.23	11.54	12.25
Economic policy disagreement among local economic forecasters	54.97	64.93	144.86	106.95	125.70	118.86	92.44	89.98	69.25	54.35	63.30	61.18	87.23
GDP Growth rate	1.65	2.07	2.05	2.03	2.02	2.07	2.02	2.04	2.59	2.33	2.35	2.37	2.13
Average income of applicant in $ thousand	64.74	65.24	69.19	69.51	69.46	69.72	73.16	74.39	76.06	78.77	79.83	80.96	72.58
Average borrower's debt to income ratio prior to loan application	10.71	13.20	12.47	13.10	13.85	16.66	17.19	17.90	18.91	18.77	18.59	18.09	15.79
Number of applicants with homeownership rent	338	1266	2676	5954	9947	24129	50859	173326	135549	85284	41224	5412	535964
Percent	0.56	0.53	0.51	0.47	0.46	0.45	0.38	0.40	0.41	0.38	0.37	0.35	0.40
Number of applicants with homeownership mortgage	206	931	2048	5617	10157	24947	70716	217968	161759	109711	57058	8034	669152
Percent	0.34	0.39	0.39	0.45	0.47	0.47	0.53	0.50	0.49	0.49	0.51	0.51	0.50
Number of loans with income status not verified	603	2040	3141	6025	6949	19869	38959	132070	91377	61956	3725	5648	372362
Percent	1	0.85	0.59	0.48	0.32	0.37	0.29	0.30	0.27	0.28	0.03	0.36	0.28

Panel C: Loan Purposes

Loan Purposes	Count	Percent	Loan Purposes	Count	Percent
Car related	14043	1.04	Medical related	14873	1.11
Credit card	296170	22.01	Moving	9271	0.69
Debt consolidation	791347	58.81	Small business	15419	1.15
Home improvement	84302	6.27	Vacation	8598	0.64
House related	6562	0.49	Wedding related	2357	0.18
Major purchase related	27983	2.08	Education	423	0.03
Renewable energy related	923	0.07	Other	73322	5.45
			Total	1345593	100.00

Table 2: Descriptive Statistics

This table describes the number of observations, the mean, median, standard deviation, minimum value, and maximum values for all the variables used in the study (see also Appendix 1). Variables are divided into four categories. The Loan Characteristics set of variables describes loans from the point of view of the lending platform — loan amount, maturity, interest rates, and other contractual characteristics of the loan. The Borrower Characteristics set of variables consists of all available variables that were provided by the applicant. Macroeconomic Variables describe various variables that explain the economic environment. The Institutional Investors characteristics explain the involvement of institutional investors in this market.

	Number of observations	Mean	Median	Std.	Min	Max
Loan Characteristics						
Percentage of requested loan funded by Lending Club	1345593	1	1	0.01	0.10	1.00
Loan amount	1345593	14378.51	12000	8586.09	500.00	40000.00
Funded amount	1345593	14369.69	12000	8582.11	500.00	40000.00
Loan maturity	1345593	41.81	36	10.28	36.00	60.00
Interest rate on loan	1345593	13.37	12.99	4.66	5.31	30.99
Monthly payment	1345593	437.51	375.91	257.60	4.93	1719.83
Number of collections in the last year excluding medical	1345448	0.02	0	0.14	0.00	20.00
Application type individual (Y/N)	1345593	0.99	1	0.1012579	0	1
The number of delinquent accounts	1345564	0.01	0	0.08	0.00	14.00
Total current balance	1275317	140424.90	80465	155888.10	0.00	8000078.00
Borrower Characteristics						
Employment length under 10 years (Y/N)	1345593	0.62	1	0.49	0	1
Employment length over 10 years (Y/N)	1345593	0.33	0	0.47	0	1
Annual income self-reported	1345589	75582.23	65000	65844.42	0	9550000
Borrower's debt to income ratio prior to loan application	1345419	18.09	17.53	9.76	-1	999
Delinquency in 2 years prior to the loan	1345564	0.32	0	0.88	0	39
Credit inquiries in past 6 months prior to the loan	1345563	0.70	0	0.99	0	33
The number of open credit lines	1345564	11.58	11	5.40	0	90
Number of derogatory public records	1345564	0.22	0	0.60	0	86
Total credit revolving balance	1345593	16204.32	11186	22136.46	0	2904836
Percentage of revolving line utilized	1344770	52.80	53	24.24	0	892.30
The total number of credit lines	1345564	25.30	24	11.99	1	176
Late fees paid (Y/N)	1345593	0.04	0	0.19	0	1.00
Total revolving high credit/credit limit	1275317	32022.09	23600	36941.99	0	9999999

Table 2. Descriptive Statistics (continued)

	Number of observations	Mean	Median	Std.	Min	Max
Number of trade accounts opened in last 2 years.	1295563	4.67	4	3.13	0	64
Average current balance of all accounts	1275293	13449.05	7425	16198.18	0	958084
Number of months since last revolving account	1275316	12.93	8	16.02	0	372
Number of months since last account	1275317	7.81	5	8.63	0	314
Number of mortgage accounts	1295563	1.73	1	2.05	0	51
Months since most recent bankcard account opened	1282919	23.71	13	30.25	0	639
Months since most recent inquiry	1173066	6.70	5	5.84	0	25
Number of accounts ever 120 or more days past due	1275317	0.51	0	1.31	0	51
Number of currently active bankcard accounts	1275317	3.64	3	2.21	0	33
Number of currently active revolving trades	1275317	5.67	5	3.25	0	63
Number of satisfactory bankcard accounts	1287003	4.71	4	2.89	0	63
Number of bankcard accounts	1275317	8.27	7	4.83	0	70
Number of installment accounts	1275317	8.59	7	7.37	0	150
Number of open revolving accounts	1275317	8.28	7	4.48	0	83
Number of revolving accounts	1275316	14.87	13	8.12	0	128
Number of revolving trades with balance >0	1275317	5.62	5	3.18	0	45
Number of satisfactory accounts	1287003	11.63	11	5.41	0	90
Number of accounts currently 120 days past due	1228102	0.00	0	0.03	0	6
Number of accounts currently 30 days past due	1275317	0	0	0.06	0	4
Number of accounts 90 or more days past due in last 2 years	1275317	0	0	0.50	0	39
Number of accounts opened in last year	1275317	2.16	2	1.80	0	32
Percent of trades never delinquent	1275170	94.22	97.90	8.60	0	100
Percentage of all bankcard accounts > 75% of limit	1281496	46.71	50	35.76	0	100
Number of public record bankruptcies	1344228	0.14	0	0.38	0	12
Number of tax liens	1345488	0.05	0	0.39	0	85
Total high credit/credit limit	1275317	172640.40	111620	176197.60	0	9999999
Total credit balance excluding mortgage	1295563	49093.19	37008	46970.54	0	3408095
Total bankcard high credit/credit limit	1295563	21035.50	14600	21008.05	0	1105500
Total installment high credit/credit limit	1275317	41358.94	31112	42546.08	0	2101913
Total installment high credit/credit limit		41358.94		42546.08	0.00	2101913.00
Macroeconomic Variables						
Risk premium	1345593	12.94	13.19	4.66	2.30	30.72
Average income on tax-based data	1345058	164058.10	185021.20	133049.30	2915.38	2261346
VIX average	1345593	18.97	19.97	3.59	12.64	36.37
Current county GDP in thousand dollars based on BEA	1342685	28200000	70300000	127000000	97815	692000000
Real county GDP in thousand dollars based on BEA	1342685	27300000	67400000	122000000	95335	656000000
GDP Growth in percentage	1342685	2.10	2.25	2.66	-23.60	24.63
Policy-related economic uncertainty based on NEWS	1345593	108.06	120.67	38.37	44.78	283.67

Disagreement among local economic forecasters	1345593	85.44	78.68	20.44	44.17	166.72
CPI disagreement measure	1345593	77.03	77.37	15.27	40.81	190.43
Forecast of 10-year dollar-weighted sum of expiring tax	1345593	83.94	316.03	396.56	18.92	1597.29
Policy-related uncertainty index	1345593	100.32	107.89	26.38	59.32	245.13
Russell 2000 Index monthly return	1345593	0.01	0.01	0.05	-0.21	0.22
Institutional Investors Characteristics						
Loan financing by investor in whole (Y/N)	1345593	1.00	1	0.50	0	1
The loan amount funded by the investors	1345593	12000.00	14342	8588.15	0	40000
Fraction of the loan funded by the investors	1345593	1	1	0.04	0	1

Table 3: Descriptive Statistics for Subsamples

This table describes the mean, median, standard deviation, minimum value, and maximum values for all variables used in the study for subsamples in the period 2007–2018. The first subsample consists of 268,043 charged-off loans. The second subsample consists of 1,077,550 paid-off loans. The Descriptive Statistics on the loan data is provided by Lending Club. The last three columns test the statistical difference of the means of variables in the subsamples, and symbols *, **, and *** represent statistical significance at the 10%, 5%, and 1% level.

	Loans paid off				Loans charged off						
	Mean	Std.	Min	Max	Mean	Std.	Min	Max	T stat	Sig.	P-value
Loan Characteristics											
Percentage of requested loan funded by LendingClub	0.9995	0.0133	0.1013	1	0.9996	0.0119	0.1547	1	-3.5741		0.000
Loan amount	14106.5400	8546.2820	500	40000	15475.7400	8658.5580	900	40000	-73.373	***	0.000
Funded amount	14097.5200	8542.1920	500	40000	15467.7000	8654.8200	900	40000	-73.458	***	0.000
Loan maturity 36 months (Y/N)	0.7961	0.4029	0	1	0.6020	0.4895	0	1	189.887	***	0.000
Interest rate on loan	12.7802	4.4445	5.31	30.99	15.7500	4.7410	5.31	30.99	-290.000	***	0.000
Monthly payment	431.5371	257.4272	4.93	1719.83	461.6152	256.9119	21.62	1714.54	-54.162	***	0.000
Number of collections in the last year excluding medical collections	0.0151	0.1402	0	20	0.0205	0.1556	0	6	-16.543	***	0.000
Application type individual (Y/N)	0.9898	0.1003	0	1	0.9889	0.1049	0	1	4.215	***	0.000
The number of delinquent accounts	0.0050	0.0767	0	14	0.0059	0.0836	0	6	-5.206	***	0.000
Total current balance	145974.80	160665.70	0	8000078	118430.60	133048.30	0	3437283	89.727	***	0.000
Borrower Characteristics											
Employment length under 10 years (Y/N)	0.6174	0.4860	0	1	0.6155	0.4865	0	1	1.835	*	0.067
Employment length over 10 years (Y/N)	0.3325	0.4711	0	1	0.3107	0.4628	0	1	21.722	***	0.000
Annual income self-reported	77045.5900	66916.87	0.00	9550000.0	69678.5400	60972.13	0	8900060	54.817	***	0.000
Borrower's debt to income ratio prior to loan application	17.6091	9.7820	-1.00	999	20.0321	9.4230	0	999	-120.00	***	0.000
Delinquency in 2 years prior to the loan	0.3110	0.8592	0.00	39	0.3538	0.9362	0	27	-21.479	***	0.000
Credit inquiries in past 6 months prior to the loan	0.6713	0.9777	0.00	33	0.8129	1.0417	0	8	-63.662	***	0.000
The number of open credit lines	11.5001	5.3527	0.00	90	11.9254	5.5656	0	76	-35.631	***	0.000
Number of derogatory public records	0.2091	0.5866	0.00	63	0.2446	0.6461	0	86	-25.884	***	0.000
Total credit revolving balance	16405.2600	22912.6400	0.00	2904836	15393.6700	18658.4300	0	1746716	23.914	***	0.000
Percentage of revolving line utilized	52.0232	24.3611	0.00	892.30	55.9187	23.5056	0.00	366.60	-76.119	***	0.000
The total number of credit lines	25.3583	11.9618	1.00	162	25.0804	12.1092	2.00	176.00	10.647	***	0.000
Late fees paid (Y/N)	0.0201	0.1404	0.00	1	0.1174	0.3219	0.00	1.00	-150.00	***	0.000
Total revolving high credit/credit limit	32940.81	38950.44	0.00	9999999	28381.1700	27277.24	0.00	1998700.0	68.854	***	0.000
Number of trade accounts opened in last 2 years	4.5101	3.0536	0.00	53	5.2905	3.3554	0.00	64.00	-110.00	***	0.000
Average current balance of all accounts	14095.4900	16827.2600	0.00	958084	10887.1400	13108.2800	0.00	355824.00	104.271	***	0.000
Total open to buy on revolving bankcards	10250.9000	15412.9600	0.00	559912	7208.7260	11422.8600	0.00	293031.00	111.940	***	0.000
Credit limit usage ratio	60.2401	27.9987	0.00	339.60	65.1542	27.0764	0.00	255.20	-81.737	***	0.000
Charge-offs in last year	0.0091	0.1097	0.00	10	0.0098	0.1134	0.00	8.00	-2.592	***	0.000
Delinquent amount	12.8304	698.7417	0.00	94521	17.3718	832.1442	0.00	76929.00	-2.603	***	0.009

Table 3: Descriptive Statistics for Subsamples (continued)

	Loans paid off				Loans charged off						
	Mean	Std.	Min	Max	Mean	Std.	Min	Max	T stat	Sig.	P-value
Number of months since first loan	126.9022	51.3033	0.00	999	123.7011	53.9026	0.00	720.00	26.737	***	0.000
Number of months since first revolving account	184.1336	93.2851	2.00	851	172.8825	93.7927	2.00	842.00	54.397	***	0.000
Number of months since last revolving account	13.3630	16.4148	0.00	334	11.2091	14.1918	0.00	372.00	66.522	***	0.000
Number of months since last account	8.0533	8.8278	0.00	314	6.8263	7.7035	0.00	289.00	69.969	***	0.000
Number of mortgage accounts	1.8003	2.0866	0.00	51	1.4289	1.8812	0.00	29.00	88.046	***	0.000
Months since most recent bankcard account opened	24.4908	30.9064	0.00	639	20.6211	27.2692	0.00	611.00	62.635	***	0.000
Months since most recent inquiry	6.8603	5.8893	0.00	25	6.0785	5.6238	0.00	25.00	60.339	***	0.000
Number of accounts ever 120 or more days past due	0.5011	1.3020	0.00	51	0.5331	1.3244	0.00	34.00	-10.995	***	0.000
Number of currently active bankcard accounts	3.5967	2.1789	0.00	33	3.8302	2.3336	0.00	27.00	-45.920	***	0.000
Number of currently active revolving trades	5.5532	3.1735	0.00	63	6.1447	3.4900	0.00	47.00	-78.156	***	0.000
Number of satisfactory bankcard accounts	4.6875	2.8678	0.00	63	4.7911	2.9591	0.00	44.00	-16.016	***	0.000
Number of bankcard accounts	8.3143	4.8320	0.00	70	8.1164	4.8259	0.00	57.00	18.568	***	0.000
Number of installment accounts	8.5656	7.3023	0.00	150	8.7020	7.6393	0.00	128.00	-8.162	***	0.000
Number of open revolving accounts	8.2026	4.4438	0.00	83	8.5969	4.6240	0.00	72.00	-38.931	***	0.000
Number of revolving accounts	14.8808	8.1089	1.00	128	14.8095	8.1790	0.00	101.00	3.954	***	0.000
Number of revolving trades with balance >0	5.5086	3.1049	0.00	45	6.0839	3.4078	0.00	44.00	-77.815	***	0.000
Number of satisfactory accounts	11.5505	5.3627	0.00	90	11.9564	5.5619	0.00	76.00	-33.432	***	0.000
Number of accounts currently 120 days past due (updated in past 2 months)	0.0009	0.0305	0.00	3	0.0009	0.0337	0.00	6.00	-0.920		0.358
Number of accounts currently 30 days past due (updated in past 2 months)	0.0034	0.0620	0.00	4	0.0039	0.0668	0.00	4.00	-3.884	***	0.000
Number of accounts 90 or more days past due in last 2 years	0.0877	0.4934	0.00	39	0.0989	0.5177	0.00	26.00	-9.933	***	0.000
Number of accounts opened in last year	2.0872	1.7621	0.00	30	2.4727	1.9316	0.00	32.00	-91.979	***	0.000
Percent of trades never delinquent	94.2613	8.5952	0.00	100	94.0805	8.6248	10.50	100.00	9.504	***	0.000
Percentage of all bankcard accounts > 75% of limit	45.4072	35.7480	0.00	100	51.8940	35.3597	0.00	100.00	-83.022	***	0.000
Number of public record bankruptcies	0.1313	0.3730	0.00	12	0.1509	0.4023	0.00	11.00	-22.862	***	0.000
Number of tax liens	0.0495	0.3817	0.00	63	0.0590	0.4339	0.00	85.00	-10.428	***	0.000
Total high credit/credit limit	179512.30	181647.90	0.00	9999999	145406.6000	149624.50	0.00	9999999	98.650	***	0.000
Total credit balance excluding mortgage	49148.29	47668.29	0.00	3408095	48874.1500	44087.06	0.00	1896461	2.789	**	0.005
Total bankcard high credit/credit limit	21782.63	21639.40	0.00	1105500	18065.4000	17979.41	0.00	560800	90.322	***	0.000
Total installment high credit/credit limit	41367.15	42942.86	0.00	2101913	41326.4000	40935.89	0.00	954503	0.447		0.655
Macroeconomic Variables											
Risk premium	12.5961	4.4470	2.30	30.72	15.5725	4.7307	3.55	30.72	-290.00	***	0.000
Average income on tax-based data	186423.40	134636.10	2915.38	2261346.00	179363.40	126286.6000	2915.38	2261346.0	25.523	***	0.000
VIX average	19.9788	3.6165	12.64	36.37	19.9245	3.4655	12.75	36.37	7.194	***	0.000
Current county GDP in thousand dollars based on BEA	70700000	127000000	97815	692000000	68900000	126000000	104995	692000000	6.466	***	0.000

Real county GDP in thousand dollars based on BEA	67800000	122000000	95335	656000000	66000000	120000000	98276	656000000	6.995	***	0.000
GDP Growth in percentage	2.2606	2.6596	-23.60	24.63	2.2165	2.6759	-23.60	24.63	7.622	***	0.000
Policy-related economic uncertainty based on NEWS	120.6312	38.3449	44.78	283.67	120.8209	38.4625	63.88	283.67	-2.284	***	0.022
Disagreement among local economic forecasters	79.3254	20.4524	44.17	166.72	76.0987	20.1961	44.17	166.72	73.755	***	0.000
CPI disagreement measure	77.5923	15.4570	40.81	190.43	76.4692	14.4625	40.81	190.43	35.444	***	0.000
Forecast of 10-year dollar-weighted sum of expiring tax	319.8890	402.0730	18.92	1597.29	300.4859	373.0918	18.92	1597.29	23.692	***	0.000
Policy-related uncertainty index	108.0969	26.6482	59.32	245.13	107.0503	25.2414	71.26	245.13	18.975	***	0.000
Russell 2000 Index monthly return	0.0082	0.0509	-0.21	0.22	0.0086	0.0505	-0.21	0.22	-3.857	***	0.000
Institutional Investors Characteristics											
Loan financing by investor in whole (Y/N)	0.5371	0.4986	0	1	0.5548	0.4970	0	1	-16.479	***	0.000
The loan amount funded by the investors	14069.4800	8548.7990	0	40000	15441.8500	8658.3940	0	40000	-73.540	***	0.000
Fraction of the loan funded by the investors	0.9974	0.0371	0	1	0.9981	0.0304	0	1	-10.598	***	0.000

Table 4: Logit Regression Analysis

This table report results from the logit regression analysis. The sample consists of 1,064,990 loans that were issued in the period 2007–2018. The dependent variable for regression equals 1 if the loan is charged off, and 0 otherwise. Aside from macroeconomic variables, the statistics on loan characteristics are collected from LendingClub. The symbols *, **, and *** represent statistical significance of coefficients at the 10%, 5%, and 1% level.

Variables	Logit regression analysis results					
	Coef.	Std.	z		p	Marginal
Loan Characteristics						
Percentage of requested loan funded by LendingClub	0.4931	6.5296	0.08		0.9400	0.069
Loan amount	0.0003	0.0001	1.86	*	0.0630	0.000
Loan maturity 36 months	-0.7780	0.0147	-52.90	***	0.0000	-0.109
Interest rate on loan	0.0091	0.0020	4.57	***	0.0000	0.001
Monthly payment	0.0009	0.0001	12.76	***	0.0000	0.000
LC credit rating A (Y/N)	-1.0768	0.0460	-23.40	***	0.0000	-0.151
LC credit rating B (Y/N)	-0.6439	0.0397	-16.24	***	0.0000	-0.090
LC credit rating C (Y/N)	-0.3506	0.0347	-10.10	***	0.0000	-0.049
LC credit rating D (Y/N)	-0.1735	0.0304	-5.71	***	0.0000	-0.024
LC credit rating E (Y/N)	-0.0534	0.0273	-1.95	*	0.0510	-0.007
LC credit rating F (Y/N)	-0.0219	0.0268	-0.82		0.4150	-0.003
Income status not verified (Y/N)	-0.1444	0.0065	-22.35	***	0.0000	-0.020
Loan purpose car (Y/N)	-0.2115	0.1026	-2.06	**	0.0390	-0.030
Loan purpose credit card	-0.1105	0.0983	-1.12		0.2610	-0.015
Loan purpose debt consolidation	-0.0845	0.0982	-0.86		0.3890	-0.012
Loan purpose home improvement	-0.0133	0.0987	-0.13		0.8930	-0.002
Loan purpose house related	-0.2208	0.1047	-2.11	**	0.0350	-0.031
Loan purpose major purchase	-0.0110	0.0999	-0.11		0.9130	-0.002
Loan purpose medical	0.1247	0.1009	1.24		0.2170	0.017
Loan purpose moving	0.0897	0.1025	0.88		0.3810	0.013
Loan purpose small business	0.3307	0.1007	3.28	***	0.0010	0.046
Loan purpose vacation	-0.0577	0.1035	-0.56		0.5770	-0.008
Loan purpose wedding related	-0.3746	0.1468	-2.55	***	0.0110	-0.053
Loan purpose other	-0.0459	0.0987	-0.46		0.6420	-0.006
Number of collections in the last year excluding medical collections	0.1481	0.0169	8.75	***	0.0000	0.021
Application type individual (Y/N)	0.4956	0.0255	19.45	***	0.0000	0.070

The number of delinquent accounts	0.0256	0.0655	0.39		0.6960	0.004
Total current balance	0.0000	0.0000	-0.43		0.6640	0.000
Borrower Characteristics						
Employment title managers (Y/N)	-0.0106	0.0072	-1.48		0.1400	-0.001
Employment title professionals (Y/N)	-0.1811	0.0088	-20.54	***	0.0000	-0.025
Employment title associate professionals (Y/N)	-0.0176	0.0091	-1.94	**	0.0520	-0.002
Employment title clerical support workers (Y/N)	0.0033	0.0321	0.10		0.9190	0.000
Employment title elementary occupations (Y/N)	0.2610	0.0478	5.46	***	0.0000	0.037
Employment title machine operators assemblers (Y/N)	0.2738	0.0128	21.44	***	0.0000	0.038
Employment title service and sales workers (Y/N)	0.1846	0.0291	6.34	***	0.0000	0.026
Employment title agri forest fishery workers (Y/N)	-0.0768	0.0836	-0.92		0.3580	-0.011
Employment title craft trades workers (Y/N)	0.0802	0.0180	4.46	***	0.0000	0.011
Employment length below 10 years (Y/N)	-0.0239	0.0057	-4.20	***	0.0000	-0.003
Homeownership mortgage (Y/N)	-0.1118	0.0093	-12.08	***	0.0000	0.001
Homeownership rent (Y/N)	0.1584	0.0092	17.20	***	0.0000	0.022
Annual income self-reported	0.0000	0.0000	-11.50	***	0.0000	0.000
Borrower's debt to income ratio prior to loan application	0.0184	0.0004	45.25	***	0.0000	0.003
Delinquency in 2 years prior to the loan	0.0785	0.0043	18.46	***	0.0000	0.011
Credit inquiries in past 6 months prior to the loan	0.0304	0.0033	9.24	***	0.0000	0.004
The number of open credit lines	-0.0402	0.0134	-3.00	***	0.0030	-0.006
Number of derogatory public records	0.0382	0.0113	3.38	***	0.0010	0.005
Total credit revolving balance	0.0000	0.0000	-2.28	***	0.0230	0.000
Percentage of revolving line utilized	0.0010	0.0002	3.87	***	0.0000	0.000
The total number of credit lines	-0.0070	0.0042	-1.68	*	0.0930	-0.001
Late fees paid (Y/N)	1.8070	0.0111	163.05	***	0.0000	0.253
Total revolving high credit/credit limit	0.0000	0.0000	-8.18	***	0.0000	0.000
Number of trade accounts opened in last 2 years	0.0391	0.0013	29.64	***	0.0000	0.005
Average current balance of all accounts	0.0000	0.0000	-8.61	***	0.0000	0.000
Total open to buy on revolving bankcards	0.0000	0.0000	4.61	***	0.0000	0.000
Credit limit usage ratio	-0.0006	0.0002	-2.37	***	0.0180	0.000
Charge-offs in last year	0.0024	0.0234	0.10		0.9170	0.000
Delinquent amount	0.0000	0.0000	3.12	***	0.0020	0.000

	Logit regression analysis results					
	Coef.	Std.	z		p	Marginal
Number of months since first loan	0.0002	0.0001	3.49	***	0.0000	0.000
Number of months since first revolving account	-0.0002	0.0000	-4.70	***	0.0000	0.000
Number of months since last revolving account	0.0002	0.0003	0.60		0.5480	0.000
Number of months since last account	-0.0020	0.0006	-3.59	***	0.0000	0.000
Number of mortgage accounts	-0.0258	0.0045	-5.67	***	0.0000	-0.004
Months since most recent bankcard account opened	-0.0018	0.0001	-13.92	***	0.0000	0.000
Months since most recent inquiry	-0.0063	0.0006	-10.93	***	0.0000	-0.001
Number of accounts ever 120 or more days past due	0.0090	0.0026	3.45	***	0.0010	0.001
Number of currently active bankcard accounts	0.0046	0.0039	1.19		0.2320	0.001
Number of currently active revolving trades	-0.0892	0.0166	-5.39	***	0.0000	-0.013
Number of satisfactory bankcard accounts	-0.0052	0.0037	-1.38		0.1670	-0.001
Number of bankcard accounts	0.0117	0.0016	7.09	***	0.0000	0.002
Number of installment accounts	0.0007	0.0042	0.16		0.8730	0.000
Number of open revolving accounts	0.0035	0.0025	1.37		0.1700	0.000
Number of revolving accounts	-0.0066	0.0043	-1.51		0.1300	-0.001
Number of revolving trades with balance >0	0.1232	0.0165	7.46	***	0.0000	0.017
Number of satisfactory accounts	0.0412	0.0134	3.08	***	0.0020	0.006
Number of accounts currently 120 days past due (updated in past 2 months)	-0.0763	0.1078	-0.71		0.4790	-0.011
Number of accounts currently 30 days past due (updated in past 2 months)	0.1072	0.0767	1.40		0.1620	0.015
Number of accounts 90 or more days past due in last 2 years	-0.0353	0.0073	-4.81	***	0.0000	-0.005
Number of accounts opened in last year	-0.0002	0.0022	-0.11		0.9120	0.000
Percent of trades never delinquent	0.0020	0.0005	4.46	***	0.0000	0.000
Percentage of all bankcard accounts > 75% of limit	0.0020	0.0001	14.69	***	0.0000	0.000
Number of public record bankruptcies	-0.0089	0.0130	-0.68		0.4940	-0.001
Number of tax liens	-0.0235	0.0129	-1.82	*	0.0690	-0.003
Total high credit/credit limit	0.0000	0.0000	-1.51		0.1320	0.000
Total credit balance excluding mortgage	0.0000	0.0000	19.21	***	0.0000	0.000
Total bankcard high credit/credit limit	0.0000	0.0000	-11.23	***	0.0000	0.000
Total installment high credit/credit limit	0.0000	0.0000	-21.25	***	0.0000	0.000
Macroeconomic Variables						
Average income on tax-based data	0.0000	0.0000	-10.76	***	0.0000	0.000
VIX average	0.0028	0.0009	3.13	***	0.0020	0.000
Current GDP in thousand dollars based on BEA	0.0000	0.0000	4.52	***	0.0000	0.000

Real GDP in thousand dollars based on BEA	0.0000	0.0000	-4.48	***	0.0000	0.000
GDP Growth in percentage	-0.0063	0.0010	-6.39	***	0.0000	-0.001
Policy-related economic uncertainty based on NEWS	-0.0008	0.0001	-9.06	***	0.0000	0.000
Disagreement among local economic forecasters	-0.0089	0.0002	-52.37	***	0.0000	-0.001
CPI disagreement measure	-0.0018	0.0003	-6.08	***	0.0000	0.000
Forecast of 10-year dollar-weighted sum of expiring tax	0.0000	0.0000	3.60	***	0.0000	0.000
Russell 2000 Index monthly return	0.3666	0.0556	6.60	***	0.0000	0.051
Institutional Investors Characteristics						
Loan financing by investor in whole (Y/N)	0.0188	0.0057	3.29	***	0.0010	0.003
The loan amount funded by the investors	-0.0003	0.0001	-1.95	*	0.0520	0.000
Fraction of the loan funded by the investors	-3.4407	1.6998	-2.02	**	0.0430	-0.483
Mc Fadden R square	12.71%					
LR Ratio	-470515.93					
Number of observations	1064990					

Table 5: Lasso Selection

This table reports results from the lasso regression. The sample consists of 1,095,012 loans that merged in period 2007–2018. The dependent variable for regression equals 1 if the loan is charged off, and 0 otherwise. The statistics on loan characteristics are collected from the LendingClub website.

Variables	**Lasso results**	
	Lasso	**Post-est OLS**
Loan Characteristics		
Loan maturity 36 months	-0.1095	-0.1049
Interest rate on loan	0.0059	0.0058
LC credit rating A (Y/N)	-0.0527	-0.0734
LC credit rating B (Y/N)	-0.0541	-0.0662
LC credit rating C (Y/N)	-0.0281	-0.0360
LC credit rating E (Y/N)	0.0279	0.0307
LC credit rating F (Y/N)	0.0374	0.0454
Income status not verified (Y/N)	-0.0184	-0.0217
Loan purpose car	-0.0088	-0.0189
Loan purpose credit card	-0.0021	-0.0022
Loan purpose home improvement	0.0059	0.0044
Loan purpose house related	-0.0201	-0.0293
Loan purpose major purchase related	0.0053	0.0047
Loan purpose medical related	0.0191	0.0138
Loan purpose moving expenses related	0.0104	0.0033
Loan purpose small business related	0.0546	0.0542
Loan purpose wedding related	-0.0356	-0.0596
Number of collections in the last year excluding medical collections	0.0197	0.0225
Application type individual (Y/N)	0.0586	0.0572
Borrower Characteristics		
Employment title managers (Y/N)	-0.0017	-0.0043
Employment title professionals (Y/N)	-0.0231	-0.0248
Employment title elementary occupations (Y/N)	0.0404	0.0490
Employment title machine operators assemblers (Y/N)	0.0452	0.0509
Employment title services and sales workers (Y/N)	0.0281	0.0325
Employment title craft and related trades workers (Y/N)	0.0107	0.0164

Employment length below 10 years (Y/N)	-0.0026	-0.0052
Homeownership mortgage (Y/N)	-0.0168	-0.0202
Homeownership rent (Y/N)	0.0263	0.0279
Borrower's debt to income ratio prior to loan application	0.0022	0.0021
Delinquency in 2 years prior to the loan	0.0090	0.0097
Credit inquiries in past 6 months prior to the loan	0.0034	0.0024
Number of derogatory public records	0.0026	0.0023
The total number of credit lines	-0.0007	-0.0010
Late fees paid (Y/N)	0.3675	0.3679
Number of trade accounts opened in last 2 years	0.0061	0.0061
Number of mortgage accounts	-0.0042	-0.0065
Months since most recent bankcard account opened	-0.0002	-0.0002
Months since most recent inquiry	-0.0007	-0.0007
Number of accounts ever 120 or more days past due	0.0009	0.0025
Number of bankcard accounts	0.0009	0.0001
Number of revolving accounts	-0.0009	-0.0005
Number of revolving trades with balance >0	0.0058	0.0053
Number of accounts currently 30 days past due (updated in past 2 months)	-0.0011	-0.0093
Percentage of all bankcard accounts > 75% of limit	0.0059	0.0010
Number of public record bankruptcies	-0.0016	-0.0036
Macroeconomic Variables		
Average of VIX index on the month of loan issued	0.0003	0.0009
GDP Growth in percentage	-0.0008	-0.0014
Policy-related economic uncertainty based on NEWS	-0.0001	-0.0001
Disagreement among local economic forecasters	-0.0013	-0.0013
CPI disagreement measure	-0.0001	-0.0001
Russell 2000 Index monthly return	0.0344	0.0529
Institutional Investors Characteristics		
Loan financing by investor in whole (Y/N)	0.0014	0.0021
Fraction of the loan funded by the investors	-0.9947	-1.3188

Table 6: Lasso Selected Variables Logit Regression

This table report results from multinomial logit regression analysis post lasso selection. The sample consists of 1,095,012 loans that merged in the period 2007–2018. The dependent variable for regression equals 1 if the loan is charged off, and 0 if the loan status is the other. The statistics on loan characteristics are collected from Lending Club. The symbols *, **, and *** represent statistical significance of coefficients at the 10%, 5%, and 1% level.

	Logistic regression results - Post Lasso Selection				
	Coef.	**Std.**	**z**		**p**
Loan Characteristics					
Loan maturity 36 months	-0.6363	0.0062	-102.25	***	0.00
Interest rate on the loan	0.0291	0.0014	20.25	***	0.00
LC credit rating A (Y/N)	-0.9153	0.0193	-47.50	***	0.00
LC credit rating B (Y/N)	-0.4555	0.0128	-35.47	***	0.00
LC credit rating C (Y/N)	-0.1747	0.0089	-19.71	***	0.00
LC credit rating E (Y/N)	0.0973	0.0098	9.94	***	0.00
LC credit rating F (Y/N)	0.1084	0.0162	6.68	***	0.00
Income status not verified (Y/N)	-0.1700	0.0060	-28.17	***	0.00
Loan purpose car	-0.1689	0.0290	-5.82	***	0.00
Loan purpose credit card	-0.0189	0.0065	-2.92	***	0.00
Loan purpose home improvement	0.0301	0.0109	2.76	***	0.01
Loan purpose house purchase related	-0.1614	0.0355	-4.55	***	0.00
Loan purpose major purchase related	0.0310	0.0185	1.68	*	0.09
Loan purpose medical related	0.1040	0.0234	4.45	***	0.00
Loan purpose moving expenses related	0.0464	0.0289	1.60		0.11
Loan purpose small business related	0.3405	0.0223	15.29	***	0.00
Loan purpose wedding related	-0.3742	0.1044	-3.58	***	0.00
Number of collections in the last year excluding medical collections	0.1323	0.0152	8.70	***	0.00
Application type individual (Y/N)	0.4286	0.0248	17.31	***	0.00
Borrower Characteristics					
Employment title managers (Y/N)	-0.0335	0.0068	-4.93	***	0.00
Employment title professionals (Y/N)	-0.1898	0.0084	-22.60	***	0.00
Employment title elementary occupations (Y/N)	0.2930	0.0453	6.47	***	0.00

Employment title machine operators assemblers (Y/N)	0.3018	0.0121	24.95	***	0.00
Employment title services and sales workers (Y/N)	0.1967	0.0274	7.17	***	0.00
Employment title craft and related trades workers (Y/N)	0.1053	0.0172	6.13	***	0.00
Employment length below 10 years (Y/N)	-0.0346	0.0053	-6.54	***	0.00
Homeownership mortgage (Y/N)	-0.1527	0.0087	-17.61	***	0.00
Homeownership rent (Y/N)	0.1728	0.0086	20.03	***	0.00
Borrower’s debt to income ratio prior to loan application	0.0165	0.0003	53.09	***	0.00
Delinquency in 2 years prior to the loan	0.0557	0.0028	19.83	***	0.00
Credit inquiries in past 6 months prior to the loan	0.0299	0.0030	9.85	***	0.00
Number of derogatory public records	0.0162	0.0049	3.30	***	0.00
The total number of credit lines	-0.0080	0.0004	-21.69	***	0.00
Late fees paid (Y/N)	1.7906	0.0106	168.69	***	0.00
Number of trade accounts opened in last 2 years	0.0363	0.0010	37.84	***	0.00
Number of mortgage accounts	-0.0510	0.0016	-31.76	***	0.00
Months since most recent bankcard account opened	-0.0020	0.0001	-19.45	***	0.00
Months since most recent inquiry	-0.0055	0.0005	-9.99	***	0.00
Number of accounts ever 120 or more days past due	0.0104	0.0019	5.42	***	0.00
Number of bankcard accounts	0.0013	0.0010	1.36		0.17
Number of revolving accounts	-0.0030	0.0007	-4.16	***	0.00
Number of revolving trades with balance >0	0.0356	0.0010	36.67	***	0.00
Number of accounts currently 30 days past due (updated in past 2 months)	0.0463	0.0378	1.22		0.22
Percentage of all bankcard accounts > 75% of limit	0.0023	0.0001	28.52	***	0.00
Number of public record bankruptcies	0.0416	0.0080	5.23	***	0.00
Macroeconomic Variables					
Average of VIX index on the month of loan issued	0.0050	0.0008	6.01	***	0.00
GDP Growth in percentage	-0.0099	0.0009	-10.70	***	0.00
Policy-related economic uncertainty based on NEWS	-0.0009	0.0001	-10.63	***	0.00
Disagreement among local economic forecasters	-0.0089	0.0002	-56.10	***	0.00
CPI disagreement measure	-0.0016	0.0003	-5.99	***	0.00
Russell 2000 Index monthly return	0.3729	0.0531	7.02	***	0.00
Institutional Investors Characteristics					
Loan financing by investor in whole (Y/N)	0.0149	0.0054	2.75	***	0.01
Fraction of the loan funded by the investors	-5.9372	1.0257	-5.79	***	0.00

Table 7: Robustness Checks for Sample Selection

This table addresses sample restrictions in the process of the selection of loan outcomes. We report three sets of results from logit regression analysis post lasso selection. The four columns labeled as "(A) Full sample" report final paper results. The four columns "(B) Subsample" report results — includes loans originated before the end of 2015 and with those with maturity of 36 months. The last four columns report results after the application of the Heckman selection model, where the selection in the sample is instrumentalized with time variable and interest rate variable. The symbols *, **, and *** represent statistical significance of coefficients at the 10%, 5%, and 1% level.

Variables	(A)Full sample					(B)Subsample					(C)Heckman correction				
	Coef.	Std.	z		p	Coef.	Std.	z			Coef.	Std.	z		
Loan Characteristics															
Loan maturity 36 months	-0.6363	0.0062	-102.25	***	0.00	-0.4823	0.0160	-30.06	***	0.00	-0.3015	0.0028	-106.16	***	0.00
Interest rate on loan	0.0291	0.0014	20.25	***	0.00	0.0620	0.0030	20.75	***	0.00	0.0181	0.0007	27.20	***	0.00
LC credit rating A (Y/N)	-0.9153	0.0193	-47.50	***	0.00	-0.5859	0.0329	-17.82	***	0.00	-0.4115	0.0087	-47.39	***	0.00
LC credit rating B (Y/N)	-0.4555	0.0128	-35.47	***	0.00	-0.2353	0.0216	-10.88	***	0.00	-0.2162	0.0059	-36.61	***	0.00
LC credit rating C (Y/N)	-0.1747	0.0089	-19.71	***	0.00	-0.0453	0.0141	-3.20	***	0.00	-0.0860	0.0041	-20.93	***	0.00
LC credit rating E (Y/N)	0.0973	0.0098	9.94	***	0.00	0.0057	0.0173	0.33		0.74	0.0492	0.0045	10.85	***	0.00
LC credit rating F (Y/N)	0.1084	0.0162	6.68	***	0.00	-0.0930	0.0298	-3.12	***	0.00	0.0623	0.0075	8.34	***	0.00
Income status not verified (Y/N)	-0.1700	0.0060	-28.17	***	0.00	-0.1209	0.0084	-14.33	***	0.00	-0.0809	0.0028	-29.33	***	0.00
Loan purpose car	-0.1689	0.0290	-5.82	***	0.00	-0.1151	0.0404	-2.85	**	0.00	-0.0745	0.0133	-5.61	***	0.00
Loan purpose credit card	-0.0189	0.0065	-2.92	***	0.00	-0.0156	0.0092	-1.69	*	0.09	-0.0073	0.0030	-2.46	***	0.01
Loan purpose home improvement	0.0301	0.0109	2.76	***	0.01	0.0000	0.0167	0.00		1.00	0.0171	0.0050	3.40	***	0.00
Loan purpose house related	-0.1614	0.0355	-4.55	***	0.00	-0.0346	0.0530	-0.65		0.51	-0.0735	0.0164	-4.49	***	0.00
Loan purpose major purchase related	0.0310	0.0185	1.68	*	0.09	-0.0289	0.0278	-1.04		0.30	0.0129	0.0085	1.52		0.13
Loan purpose medical related	0.1040	0.0234	4.45	***	0.00	0.0303	0.0332	0.91		0.36	0.0351	0.0108	3.24	***	0.00
Loan purpose moving expenses related	0.0464	0.0289	1.60		0.11	0.0008	0.0393	0.02		0.99	0.0205	0.0134	1.53		0.13
Loan purpose small business related	0.3405	0.0223	15.29	***	0.00	0.2487	0.0312	7.98	***	0.00	0.1671	0.0102	16.35	***	0.00
Loan purpose wedding related	-0.3742	0.1044	-3.58	***	0.00	-0.3301	0.1042	-3.17	***	0.00					
Number of collections in the last year excluding medical	0.1323	0.0152	8.70	***	0.00	0.1046	0.0224	4.67	***	0.00	0.0606	0.0066	9.15	***	0.00
Application type individual (Y/N)	0.4286	0.0248	17.31	***	0.00	0.0315	0.2139	0.15		0.88	0.1922	0.0114	16.89	***	0.00
Borrower Characteristics															
Employment title managers (Y/N)	-0.0335	0.0068	-4.93	***	0.00	-0.0455	0.0105	-4.33	***	0.00	0.1336	0.0152	8.81	***	0.00
Employment title professionals (Y/N)	-0.1898	0.0084	-22.60	***	0.00	-0.2311	0.0126	-18.29	***	0.00	0.0020	0.0087	0.24		0.81
Employment title elementary occupations (Y/N)	0.2930	0.0453	6.47	***	0.00	0.2559	0.0657	3.90	***	0.00	0.1886	0.0321	5.87	***	0.00

Employment title machine operators Assemblers (Y/N)	0.3018	0.0121	24.95	***	0.00	0.2490	0.0185	13.46	***	0.00	0.1545	0.0080	19.36	***	0.00
Employment title services and sales workers (Y/N)	0.1967	0.0274	7.17	***	0.00	0.0470	0.0258	1.82	*	0.07	0.1003	0.0147	6.81	***	0.00
Employment title craft and related trades workers (Y/N)	0.1053	0.0172	6.13	***	0.00	0.0858	0.0262	3.27	***	0.00	0.0902	0.0113	7.97	***	0.00
Employment length below 10 years (Y/N)	-0.0346	0.0053	-6.54	***	0.00	-0.0356	0.0077	-4.62	***	0.00	-0.0203	0.0023	-8.79	***	0.00
Homeownership mortgage (Y/N)	-0.1527	0.0087	-17.61	***	0.00	-0.1360	0.0128	-10.59	***	0.00	-0.0646	0.0040	-16.27	***	0.00
Homeownership rent (Y/N)	0.1728	0.0086	20.03	***	0.00	0.1246	0.0125	9.98	***	0.00	0.0852	0.0040	21.57	***	0.00
Borrower’s debt to income ratio prior to loan application	0.0165	0.0003	53.09	***	0.00	0.0179	0.0005	38.31	***	0.00	0.0100	0.0001	68.13	***	0.00
Delinquency in 2 years prior to the loan	0.0557	0.0028	19.83	***	0.00	0.0401	0.0041	9.83	***	0.00	0.0257	0.0013	20.20	***	0.00
Credit inquiries in past 6 months prior to the loan	0.0299	0.0030	9.85	***	0.00	0.0212	0.0044	4.86	***	0.00	0.0150	0.0014	10.68	***	0.00
Number of derogatory public records	0.0162	0.0049	3.30	***	0.00	-0.0044	0.0072	-0.62		0.54	0.0123	0.0022	5.57	***	0.00
The total number of credit lines	-0.0080	0.0004	-21.69	***	0.00	-0.0085	0.0004	-20.77	***	0.00	-0.0050	0.0001	-39.40	***	0.00
Late fees paid (Y/N)	1.7906	0.0106	168.69	***	0.00	1.6878	0.0145	116.50	***	0.00	0.8218	0.0050	165.92	***	0.00
Number of trade accounts opened in last 2 years	0.0363	0.0010	37.84	***	0.00	0.0499	0.0014	34.69	***	0.00	0.0150	0.0004	34.20	***	0.00
Number of mortgage accounts	-0.0510	0.0016	-31.76	***	0.00	-0.0329	0.0023	-14.18	***	0.00	-0.0218	0.0007	-29.90	***	0.00
Months since most recent bankcard account opened	-0.0020	0.0001	-19.45	***	0.00	-0.0025	0.0002	-15.78	***	0.00	-0.0010	0.0000	-21.06	***	0.00
Months since most recent inquiry	-0.0055	0.0005	-9.99	***	0.00	-0.0062	0.0008	-7.69	***	0.00	-0.0022	0.0003	-8.94	***	0.00
Number of accounts ever 120 or more days past due	0.0104	0.0019	5.42	***	0.00	0.0127	0.0029	4.42	***	0.00	0.0065	0.0009	7.35	***	0.00
Number of bankcard accounts	0.0013	0.0010	1.36		0.17	-0.0025	0.0027	-0.91		0.37	-0.0050	0.0009	-5.86	***	0.00
Number of revolving accounts	-0.0030	0.0007	-4.16	***	0.00	-0.0277	0.0066	-4.17	***	0.00	0.0031	0.0017	1.75	*	0.08
Number of revolving trades with balance >0	0.0356	0.0010	36.67	***	0.00	0.0509	0.0067	7.60	***	0.00	0.0164	0.0018	9.30	**	0.00
Number of accounts currently 30 days past due	0.0463	0.0378	1.22		0.22	-0.0022	0.0544	-0.04		0.97	0.0216	0.0173	1.25		0.21
Percentage of all bankcard accounts > 75% of limit	0.0023	0.0001	28.52	***	0.00	0.0016	0.0001	13.60	***	0.00	0.0011	0.0000	30.74	***	0.00
Number of public record bankruptcies	0.0416	0.0080	5.23	***	0.00	0.0624	0.0116	5.40	***	0.00	0.0177	0.0036	4.89	***	0.00
Macroeconomic Variables															
Average of VIX index on the month of loan issued	0.0050	0.0008	6.01	***	0.00	0.0017	0.0013	1.33		0.19	0.0040	0.0004	10.62	***	0.00
GDP Growth in percentage	-0.0099	0.0009	-10.70	***	0.00	-0.0067	0.0013	-5.23	***	0.00	-0.0050	0.0004	-11.56	***	0.00
Policy-related economic uncertainty based on NEWS	-0.0009	0.0001	-10.63	***	0.00	-0.0008	0.0001	-6.54	***	0.00	-0.0005	0.0000	-11.97	***	0.00
Disagreement among local economic forecasters	-0.0089	0.0002	-56.10	***	0.00	-0.0094	0.0003	-31.52	***	0.00	-0.0030	0.0001	-42.14	***	0.00
CPI disagreement measure	-0.0016	0.0003	-5.99	***	0.00	0.0003	0.0005	0.51		0.61	-0.0003	0.0001	-2.42	**	0.02

Russell 2000 Index monthly return	0.3729	0.0531	7.02	***	0.00	0.3153	0.0755	4.18	***	0.00	0.1768	0.0245	7.23	***	0.00
Institutional Investors Characteristics															
Loan financing by investor in whole (Y/N)	0.0149	0.0054	2.75	***	0.01	-3.7584	4.6910	-0.80		0.42	-1.5392	2.2383	-0.69		0.49
Fraction of the loan funded by the investors	-5.9372	1.0257	-5.79	***	0.00	-4.6325	1.6144	-2.87	***	0.00	-2.5305	0.4523	-5.59	***	0.00
Mc Fadden R square	12.11%					8.99%					8.99%				
LR Ratio	-511053					-255698					-1878745				
Number of observations	1095012					651555					2023934				

Figure 1
The figure shows the temporal distribution of the charged-off loans by loan purpose. Borrowers while requesting a loan report the purpose of the loan and then LendingClub classifies it within the categories: car, home improvement, moving expenses, vacation expenses, credit card refinancing, house purchase-related, other, debt consolidation, major purchase, renewable energy, educational expense, medical expense, and small business expense.

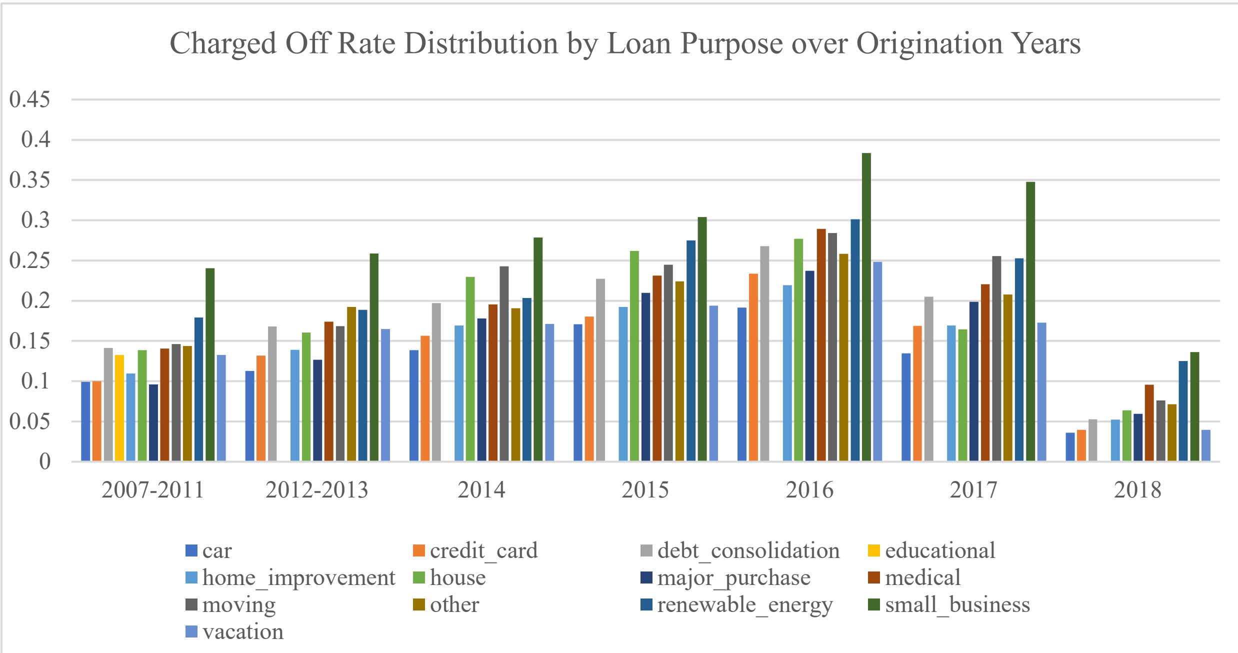


Figure 2

The figure shows the temporal distribution of the charged-off loans by homeownership. While applying for a loan, borrowers state their homeownership status, and LendingClub classifies it within the categories: mortgage, none, other, rent, and own.

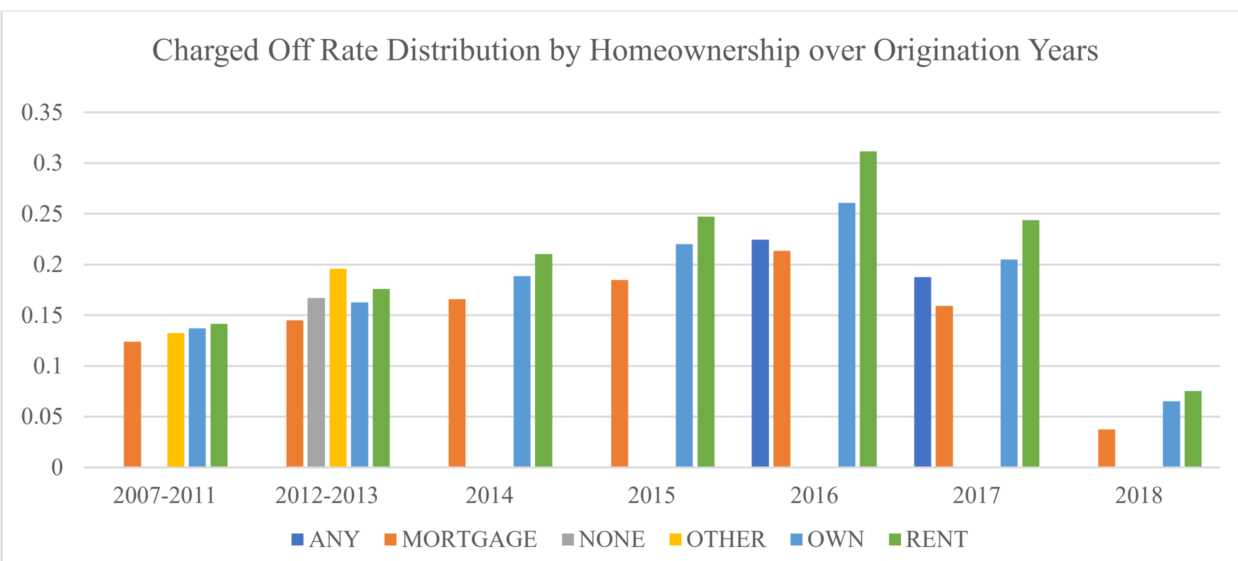


Figure 3
The figure shows the temporal charge-off rate based on LendingClub's assigned credit score at the loan origination.

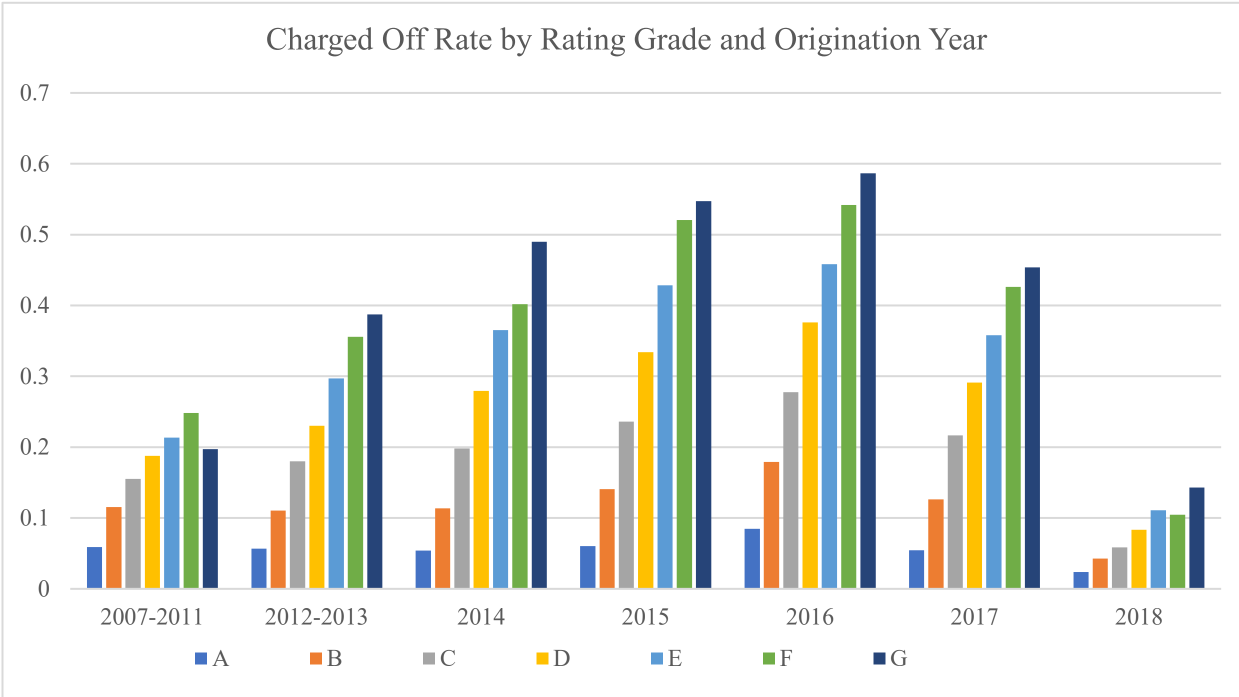


Figure 4
The figure shows the temporal distribution of the charge-offs based on the issuance interest rate.

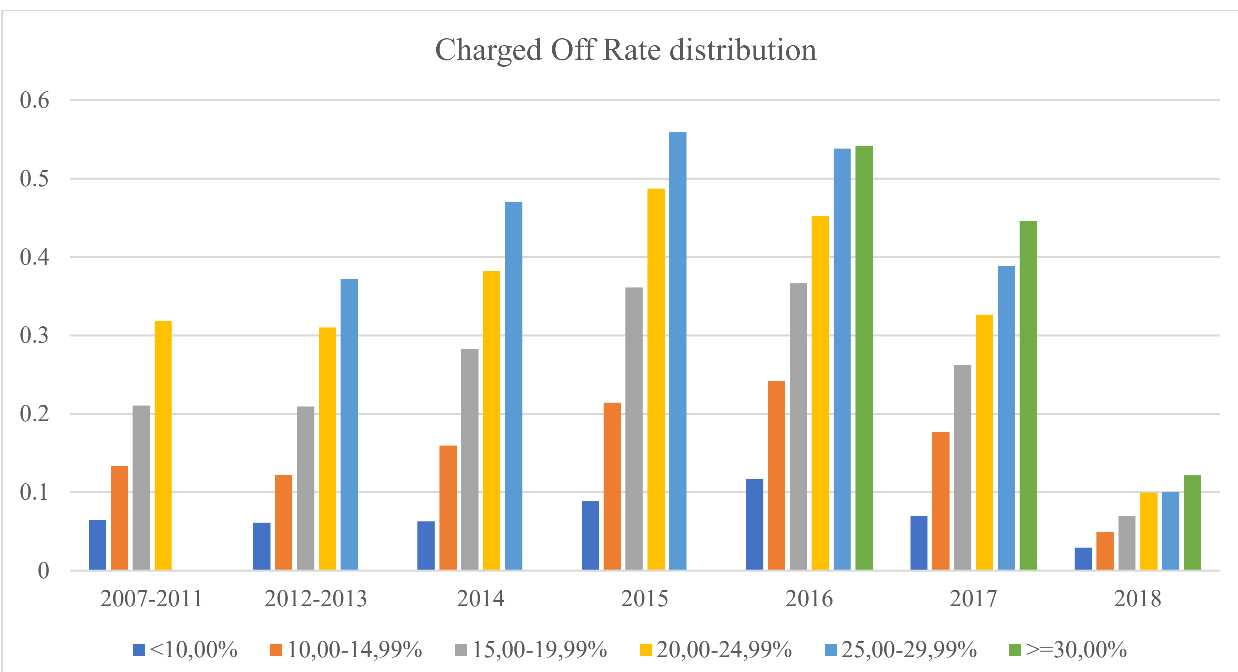


Figure 5
The figure shows the annual percentage rate (APR) obtained by borrowers as a function of the credit score assigned by Lending Club.

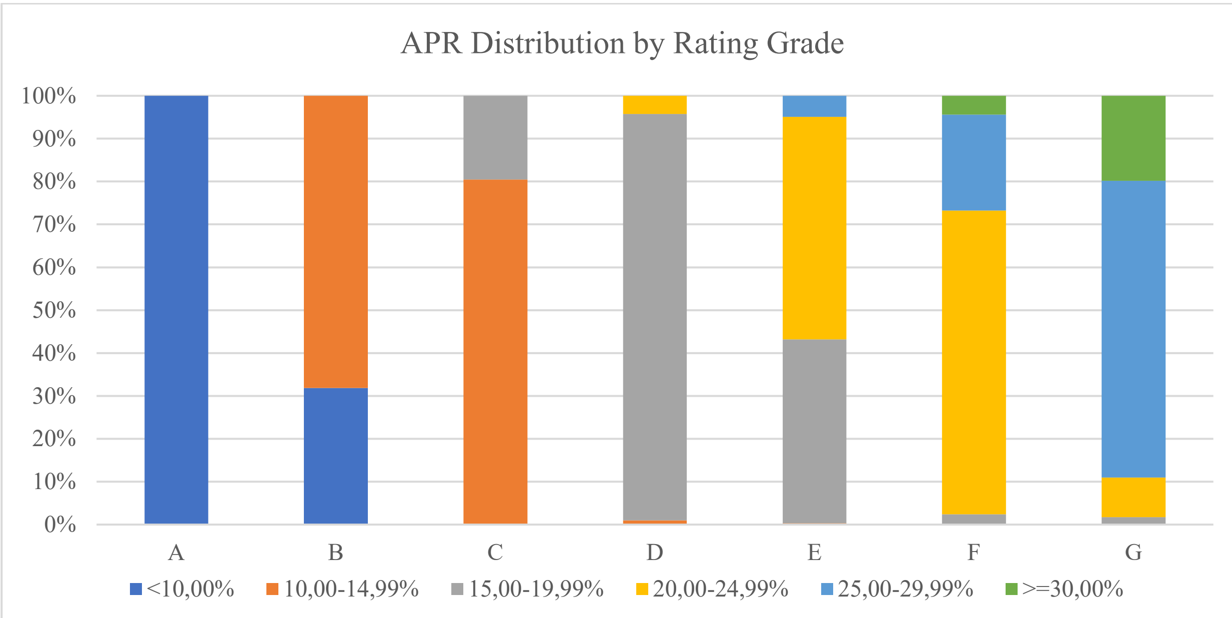
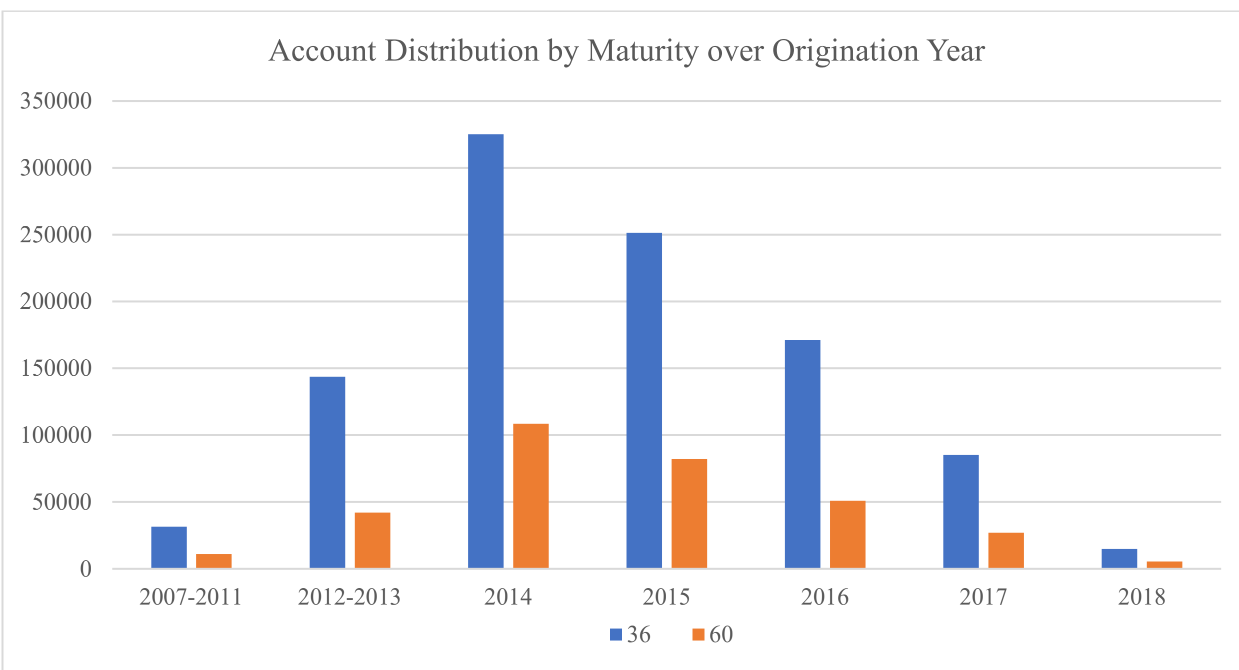


Figure 6
The figure shows the temporal distribution of loans based on maturity. LendingClub originates loans with only two maturities: 36 months and 60 months.



Appendix 1: Variables Definition

The Appendix 1 presents in detail how every independent variable is defined.

Code	Variable	Description	Variable type
1	Loan payment status	Current status of the loan. We create dependent variable charged off, based on the payment status.	Dummy variables based on the payment status.
100	**Loan characteristics**		
101	Percentage of requested loan funded by Lending Club	The ratio of funded loan amount to requested loan amount by borrower.	Ratio
102	Loan amount	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.	$ amount
103	Funded amount	The total amount committed to that loan at that point in time.	$
104	Loan maturity	Maturity of the loan. Values are in months and can be either 36 or 60. (Y/N)	Dummy variables based on the term months.
105	Interest rate on loan	Interest rate	Percentage
106	Monthly payment	The monthly payment owed by the borrower if the loan originates.	$
107	LC credit rating	Lending Club assigned loan grade (Y/N) at issuance	Dummy variable
108	LC subcredit rating	LC assigned loan subgrade (Y/N)	Dummy variables based on the subgrade issued.
109	Income verification status	Indicates if income was verified by Lending Club, not verified, or if the income source was verified. We use this information to create dummy variables. (Y/N)	Dummy variables based on whether verification was conducted
111	Loan purpose description	Loan description category provided by the borrower. Purposes are: car purchase, credit card consolidation, debt consolidation, home improvement, house purchase related, major purchase, medical expense, moving expense, other, small business related, vacation financing and wedding related expenditure, educational related, renewable energy related. (Y/N)	Dummy variables based on the purpose of loan.
112	Number of collections in the last year excluding medical collections	Number of collections in 12 months excluding medical collections	Integer
113	Months since most recent 90-day or worse rating	Months since most recent 90-day or worse rating	Integer
114	Application type individual (Y/N)	Indicates whether the loan is an individual application or a joint application with two co-borrowers (Y/N)	Dummy variables based on the type of application.
115	The number of delinquent accounts	The number of accounts on which the borrower is now delinquent	Integer
116	Total current balance	Total current balance of all accounts	$
200	**Borrower Characteristics**		
201	Employment title	The job title supplied by the Borrower when applying for the loan. The International Standard Classification of Occupations (ISCO-08) is used to create dummy variables.	The International Standard Classification of Occupations (ISCO-08) is used to create dummy variables.
202	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. (Y/N)	Dummy variables based on the employment length.

203	Homeownership	The homeownership status provided by the borrower during registration or obtained from the credit report. Our values are: Rent, Own, Mortgage and Other. We use this information to create dummy variables. (Y/N)	Dummy variables based on the homeownership
204	Annual income self-reported	The self-reported annual income provided by the borrower during registration.	$
205	Borrower zip code	The first 3 numbers of the zip code provided by the borrower in the loan application.	Integer
206	Borrower's debt-to-income ratio prior to loan application	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.	Ratio
207	Delinquency in 2 years prior to the loan	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years	Integer
208	Earliest credit line opened	The month the borrower's earliest reported credit line was opened	Integer
209	Credit inquiries in past 6 months prior to the loan	The number of inquiries in past 6 months (excluding auto and mortgage inquiries)	Integer
210	Months since the last delinquency	The number of months since the borrower's last delinquency.	Integer
211	Months since the last public record	The number of months since the last public record.	Integer
212	The number of open credit lines	The number of open credit lines in the borrower's credit file.	Integer
213	Number of derogatory public records	Number of derogatory public records	Integer
214	Total credit revolving balance	Total credit revolving balance	$
215	Percentage of revolving line utilized	Percentage of revolving line utilized.	$
216	The total number of credit lines	The total number of credit lines currently in the borrower's credit file	Integer
217	Late fees paid (Y/N)	Late fees received to date.	$
218	Total revolving high credit/credit limit	Total revolving high credit/credit limit	$
219	Number of trade accounts opened in last 2 years	Number of trade accounts opened in past 24 months.	Integer
220	Average current balance of all accounts	Average current balance of all accounts	$
221	Total open to buy on revolving bankcards	Total open to buy on revolving bankcards.	Ratio
222	Credit limit usage ratio	Ratio of total current balance to high credit/credit limit for all bankcard accounts.	Ratio
223	Charge-offs in last year	Number of charge-offs within 12 months	Integer
224	Delinquent amount	The past-due amount owed for the accounts on which the borrower is now delinquent.	$
225	Number of months since first loan	Months since oldest bank installment account opened	Integer
226	Number of months since first revolving account	Months since oldest revolving account opened	Integer
227	Number of months since last revolving account	Months since most recent revolving account opened	Integer
228	Number of months since last account	Months since most recent account opened	Integer
229	Number of mortgage accounts	Number of mortgage accounts	Integer

230	Months since most recent bankcard account opened	Months since most recent bankcard account opened.	Integer
231	Months since most recent bankcard delinquency	Months since most recent bankcard delinquency	Integer
232	Months since most recent inquiry	Months since most recent inquiry.	Integer
233	Months since most recent revolving delinquency	Months since most recent revolving delinquency.	Integer
234	Number of accounts ever 120 or more days past due	Number of accounts ever 120 or more days past due	Integer
235	Number of currently active bankcard accounts	Number of currently active bankcard accounts	Integer
236	Number of currently active revolving trades	Number of currently active revolving trades	Integer
237	Number of satisfactory bankcard accounts	Number of satisfactory bankcard accounts	Integer
238	Number of bankcard accounts	Number of bankcard accounts	Integer
239	Number of installment accounts	Number of installment accounts	Integer
240	Number of open revolving accounts	Number of open revolving accounts	Integer
241	Number of revolving accounts	Number of revolving accounts	Integer
242	Number of revolving trades with balance >0	Number of revolving trades with balance greater than zero	Integer
243	Number of satisfactory accounts	Number of satisfactory accounts	Integer
244	Number of accounts currently 120 days past due (updated in past 2 months)	Number of accounts currently 120 days past due (updated in past 2 months)	Integer
245	Number of accounts currently 30 days past due (updated in past 2 months)	Number of accounts currently 30 days past due (updated in past 2 months)	Integer
246	Number of accounts 90 or more days past due in last 2 years	Number of accounts 90 or more days past due in last 24 months	Integer
247	Number of accounts opened in last year	Number of accounts opened in past 12 months	Integer
248	Percent of trades never delinquent	Percent of trades never delinquent	Percentage
249	Percentage of all bankcard accounts > 75% of limit	Percentage of all bankcard accounts > 75% of limit.	Percentage
250	Number of public record bankruptcies	Number of public record bankruptcies	Integer
251	Number of tax liens	Number of tax liens	Integer
252	Total high credit/credit limit	Total high credit/credit limit	Ratio
253	Total credit balance excluding mortgage	Total credit balance excluding mortgage	Ratio
254	Total bankcard high credit/credit limit	Total bankcard high credit/credit limit	Ratio
255	Total installment high credit/credit limit	Total installment high credit/credit limit	Ratio

300	**Institutional Investors Characteristics**		
301	Loan financing by investor (whole or fractional)	The initial listing status of the loan from the perspective of the investor. Possible values are – Whole and Fractional. Created dummy variables based on these values. (Y/N)	Dummy variables based on the listing status.
302	The loan amount funded by the investors	The loan amount funded by the investors.	
303	Fraction of the loan funded by the investors	The ratio of the loan amount financed by investors to total funded amount.	Ratio
400	**Macroeconomic Variables**		
401	Risk premium	Difference of interest rate of loan to T-bill rate of loan month issual	Percentage
402	Average income on tax-based data	Average income on tax data based on zip code provided by IRS	$. Average income from tax data based on zip code collected from Internal Revenue Service.
403	VIX average	Average of VIX index on the month of loan issued	Integer. Average difference of open and close index values on month of loan issual. Collected from CBOE Global Markets, Inc.
404	Current GDP in thousand dollars based on BEA	Current GDP in thousand dollars based on Bureau of Economic Analysis December 12th announcement.	$. Based on data provided by Bureau of Economic Analysis December 12th announcement. GDP data is merged based on county-zip code.
405	Real GDP in thousand dollars based on BEA	Real GDP in thousand dollars	$. Based on data provided by Bureau of Economic Analysis December 12th announcement. GDP data is merged based on county-zip code.
406	GDP Growth in percentage	GDP Growth in percentage	Percentage. Based on data provided by Bureau of Economic Analysis December 12th announcement. GDP data are merged based on county-zip code.
407	Policy-related economic uncertainty based on NEWS	Policy related economic uncertainty based on NEWS	Data are collected from http://www.policyuncertainty.com and merged based on loan issue date.
408	Disagreement among local economic forecasters	Disagreement among local economic forecasters	Data are collected from http://www.policyuncertainty.com and merged based on loan issue date.
409	CPI disagreement measure	CPI disagreement measure	Data are collected from http://www.policyuncertainty.com and merged based on loan issue date.
410	Forecast of 10-year dollar-weighted sum of expiring tax	Each year's forecast is a 10-year horizon dollar-weighted sum of expiring tax.	
411	Policy-related uncertainty index	Policy-related uncertainty index draws on the Federal Reserve Bank of Philadelphia's *Survey of Professional Forecasters*.	Data are collected from http://www.policyuncertainty.com and merged based on loan issue date.
412	Russell 2000 Index level	Russell 2000 monthly data added based on month of loan issued.	Russell 2000 Index data included based on loan issue date.
701	Post charge-off recovery (Y/N)	Post charge off gross recovery. Borrowers paid some amount after charged off.	