DOI: 10.1111/fima.12295

ORIGINAL ARTICLE

Financial Management WILEY

The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform

Julapa Jagtiani[1] | Catharine Lemieux[2]

[1]Federal Reserve Bank of Philadelphia, Philadelphia, Pennsylvania

[2]Federal Reserve Bank of Chicago, Chicago, Illinois

Correspondence
Julapa Jagtiani, Federal Reserve of Philadelphia, Ten Independence Hall, Philadelphia, PA 19106.
Email: julapa.jagtiani@phil.frb.gov

Abstract

There have been concerns about the use of alternative data sources by fintech lenders. We compare loans made by LendingClub and similar loans that were originated by banks. The correlations between the rating grades (assigned by LendingClub) and the borrowers' FICO scores declined from about 80% (for loans originated in 2007) to about 35% for recent vintages (originated in 2014–2015), indicating that nontraditional data (not already accounted for in the FICO scores) have been increasingly used by fintech lenders. The rating grades perform well in predicting loan default. The use of alternative data has allowed some borrowers who would have been classified as subprime by traditional criteria to be slotted into "better" loan grades, allowing them to obtain lower priced credit.

KEYWORDS

fintech, LendingClub, marketplace lending, alternative data, shadow banking, P2P lending, peer-to-peer lending

1 | INTRODUCTION

Consumer credit has been growing steadily in recent years. As of September 2018, of the nearly $4 trillion of overall consumer credit (not secured by real estate), approximately 26% was credit card debt and only 6% was unsecured personal loans (Federal Reserve, 2018).[1] Bricker et al. (2017) find that based on the 2016 Survey of Consumer Finance, 20.8% of families felt credit constrained, and this result has been fairly consistent over recent years. Oliver Wyman (Carroll & Rehmani, 2017) estimates that as many as 60 million people may have been unable to access credit due to their thin credit files or lack of credit history. It is likely that a significant number of consumers in the subprime pool (based on the traditional measures) may not be risky borrowers, but they were subject to excessive risk premiums that reflect their low credit scores (based on inaccurate measures).

[1]The remaining 68% was student loan and auto-related debt.

Fintech lending platforms have entered the unsecured personal loan space and have the potential to fill this unmet demand for credit. Over the past decade, online alternative lenders have evolved from platforms connecting individual borrowers with individual lenders to sophisticated networks featuring institutional investors, direct lending (on their balance sheet), and securitization transactions.[2] The use of alternative data sources, big data and machine learning (ML) technology, and other complex artificial intelligence (AI) algorithms could also reduce the cost of making credit decisions and/or credit monitoring and lower operating costs for lenders. Fintech lenders could potentially pass the benefits on to borrowers.

Alternative data, when included in the credit risk analysis, could paint a fuller and more accurate picture regarding people's financial lives and their creditworthiness that could make it possible for millions of American consumers to have access to affordable credit (Cordray, 2017). Some fintech lenders have developed their own proprietary complex ML algorithms that use big data and alternative data to evaluate borrowers' credit risk. Through this new approach to credit risk evaluation, some consumers with a short credit history (i.e., one that may not satisfy a bank's traditional lending requirements) could potentially obtain a loan from an online alternative lender. Some fintech lenders specialize in making loans to those "below-prime" consumers by identifying those "invisible prime" consumers from the (traditional) subprime pool. Fintech lenders could potentially make loans to below-prime consumers at lower costs than what they would have otherwise received and without the lenders incurring any more loss (because of loan default) than the expected level of loss on loans to average consumers.

Crosman (2016) reports in *American Banker* that SoFi no longer uses FICO scores when determining loan qualifications. In addition, Kabbage claims that FICO scores are not part of its creditworthiness determination (although FICO scores are used for benchmarking and investor reporting). In the *American Banker* article, Ron Suber, former president of Prosper Marketplace, states that "Prosper gets 500 pieces of data on each borrower; the FICO score is just one data point." The company uses FICO scores to screen borrower candidates. A score of at least 640 is needed to be considered for a loan. Prosper analyzes additional data to determine its ultimate credit decision. These data sources were not normally used by traditional lenders.

We use personal installment loan-level data from LendingClub's unsecured consumer platform and compare it with similar loan-level data from traditional lenders to explore the potential consumer benefits that fintech lenders provide. Specifically, we investigate two channels: (a) whether the use of alternative data (to build internal credit rating systems such as the one designed by LendingClub) can improve consumers' ability to access credit by allowing lenders to better assess their true creditworthiness and (b) whether the use of alternative data allows fintech lenders to better risk price credit so some borrowers can get loans from fintech firms at a lower cost than they could get from traditional banks.

Our results indicate that, over the years, alternative sources of information have been increasingly used by fintech lenders to evaluate credit applications. The additional information is outside what is typically included in traditional credit ratings or the traditional credit approval criteria. Our results demonstrate that the correlation between the borrowers' FICO scores (at the time of loan application) and the rating grades assigned by LendingClub have dramatically declined over the years indicating an increased usage of alternative data in the internal rating process. We also find that credit spreads can be explained by information in LendingClub's rating grades that are not found in the FICO score or in other obvious measures of credit risk. This orthogonal component is also useful in predicting LendingClub's loan performance over the 2 years after loan origination.

Although it is not known exactly what specific set of alternative data is used by each of the specific fintech lenders, some have mentioned information drawn from bank account transactions, such as utility or rent payments, other recurring transactions, and electronic records of deposit and withdrawal transactions. Other items mentioned include insurance claims, credit card transactions, a consumer's occupation or details about their education, their use of mobile phones and related activities, Internet footprints, online shopping habits, investment choices, and so on. Concerns

[2] This is frequently referred to in prior research as peer-to-peer (P2P).

emerged that consumer privacy may be compromised in the process if information, such as insurance claims, utility bills, bank account transactions, and social network details, are used by lenders without a borrower's consent.

The rest of the paper is organized as follows. In Section 2, we present the literature review. Section 3 describes our data from various sources. Section 4 discusses the roles of alternative data and how they have been used in the credit decision process. Section 5 explores the pricing of credit (interest rate spreads) of loans originated by a fintech platform versus traditional origination. Section 6 further investigates the relationship between pricing and loan performance using regression analysis to control for other relevant risk factors, and Section 7 concludes and discusses policy implications.

2 | THE LITERATURE

Information asymmetries between lenders and borrowers have long been an important topic of banking research and, more recently, they have become a popular topic for fintech lending research. Morse (2015) reviewed the existing literature developing around fintech lending with a focus on whether the type of technologies employed by fintech firms can mitigate information frictions in lending. She posits that the process of improved capturing of soft information contained in proximity information and better profiling of loan applicants could improve the access to or price of credit. Freedman and Jin (2017) demonstrate the value of friends of the applicant committing to investing in the loan. They also find that this signal is more pronounced in lower credit grades, thus supporting the use of alternative data, such as social network, in credit decisions. Similarly, Everett (2010) determines that loans funded by investor groups perform better if someone in the group is personally connected to the borrowers. Likewise, Lin, Prabhala, and Viswanathan (2013) find that the credit quality of a borrower's friends is related to improved success in fundraising, lower interest rates, and a lower default rate. Social networks and friends may also have a negative impact on consumer credit access. Lu, Gu, Ye, and Sheng (2012) note that the reverse relationship also holds. They find a positive relationship between a friend's default and a borrower's probability of default. Research findings thus far are consistent with the argument that information drawn from social networks and friends can be useful in credit risk evaluation, especially for those with thin credit files. However, inferring credit risk from an applicant's social network and friends, rather than the consumers' own credit performance, could potentially be considered a fair lending violation. This is a topic for a separate research study.

In addition to social networks and friends, researchers have investigated the potential for other soft information to be leveraged in online loan applications. Michels (2012) finds that voluntary disclosure of hard information (other than credit scores), such as income, income source, education, and other debt, is related to interest rates that consumers are charged. Herzenstein, Sonenshein, and Dholakia (2011), through text analysis of borrower narratives, confirm limited usefulness. Gao and Lin (2012) use text mining and determine that more complex narratives are correlated with higher default rates. Yencha, Nowak, and Ross (2018) also used text mining and find that text descriptions of small businesses can predict whether a small business loan will be funded. They also determine that this information may be most useful for borrowers with low FICO scores. Duarte, Siegel, and Young (2012), Pope and Sydnor (2011), Ravina (2012), and Gonzalez and Komarova (2014) analyzed photo-based discrimination. The results are mixed. Some findings of bias lean toward attractive or trustworthy faces and against racial minorities. A central issue to the value of this line of research is that once borrowers understand that lenders are using this information, they could choose to alter the way they submit text or photo information.

Alternative data could also be derived from local economic information. For example, some fintech lenders can identify whether the loan applications are submitted from a high-crime area or in an area where factories are being shut down or relocated. Previous studies have found evidence that local economic information could serve as a possible relevant source of nontraditional information by fintech lenders (Alyakoob, Rahman, & Wei, 2017; Bertsch, Hull, & Zhang, 2016; Buchak, Matvos, Piskorski, & Seru, 2017; Chen, Hanson, & Stein, 2017; Crowe & Ramcharan, 2013; Havrylchyk, Mariotto, Rahim, & Verdier, 2018; Jagtiani & Lemieux, 2018).

Advanced technology and AI/ML algorithms have made it less costly and more effective for lenders to originate and service loans. Researchers have begun to investigate whether fintech lenders pass on the savings to consumers with lower credit costs and whether the pricing is appropriate for the risk taken. Morse (2015) explores a number of issues related to Fintech disruption and financial disintermediation and concludes that at least some cost savings seem to accrue to investors (because 80% of P2P funds come from institutional investors) and that the borrowers' social circles and local economic indicators are useful in predicting credit risk.

A few studies have attempted to compare lending rates from online alternative platforms with traditional lending channels, but these studies have been subject to significant data limitations and the results have been mixed. Mach, Carter, and Slattery (2014) report that P2P small business borrowers paid higher rates for fintech loans compared with loans obtained from traditional sources. However, they used data from LendingClub's consumer platform that were identified as small business purposes and were less likely to be comparable with small business loans made by traditional banks. The consumer loans that are marked as "small business" purposes on the LendingClub consumer platform represent less than 2% of all loans on the consumer platform and they are treated as consumer loans (LendingClub started its small business lending platform in late 2014). Demyanyk and Kolliner (2014) find that more creditworthy consumers receive preferred rates using a P2P lender over borrowing with a credit card. However, they used aggregate market rates as the comparison. In Germany, De Roure, Pelizzon, and Tasca (2016), using data from Auxmoney, a German P2P lending site, suggest that interest rates are comparable with loans made by P2P alternative lenders and those made by traditional banks. However, the interest rates used as a comparison were market rates.

In a more recent paper, De Roure, Pelizzon, and Thakor (2018) find that risk-adjusted rates on P2P loans were lower than those on bank loans in Germany and concluded that P2P lenders were bottom fishing. Jagtiani, Lambie-Hanson, and Lambie-Hanson (2019) find that, for FHA mortgage borrowers (i.e., borrowers who are more likely to be underserved than the conventional mortgage borrowers), fintech lenders offer a lower rate than traditional mortgage lenders on average. In contrast, Buchak et al. (2017), focusing on conventional mortgage loans, find evidence that fintech mortgage borrowers are among the borrowers who value fast and convenient services and that fintech lenders command an interest rate premium for their services. Another interesting study that looked at risk pricing by LendingClub found that the rates charged to higher-risk borrowers were not large enough to compensate for a higher probability of default (see Emekter, Tu, Jirasakuldech, & Lu, 2014). Our paper, using loan-level data from both LendingClub and traditional banks, is able to overcome many of the data limitations in previous studies, allowing us to compare how fintech lenders and traditional banks price consumer credit.

Taking a different approach, Hughes, Jagtiani, and Moon (2019) compare the performance of consumer loans made by large financial institutions with those made by LendingClub using a novel approach to stochastic efficiency analysis. They find that LendingClub and the financial institutions with the largest consumer portfolios were better at credit evaluation and loan management than financial institutions with smaller consumer portfolios. Although LendingClub and the largest banks did take on more credit risk, they remained closer to the efficient frontier. This is consistent with work by Serrano-Cinca, Gutierrez-Nieto, and Lopez-Palacios (2015): Using data from 2008 to 2014, they find the loan grades that LendingClub assigned were the most predictive factor of defaults suggesting that LendingClub was able to appropriately risk rank the borrowers.

Another interesting research question is whether fintech lenders have made it easier for consumers to become excessively leveraged causing loans to perform poorly in the long run. In other words, are fintech borrowers better off by being able to access this type of credit? The results have been mixed. An interesting study by Danisewicz and Elard (2018) finds that when a U.S. Court of Appeals verdict caused a decline in marketplace lending, there was a proportional and persistent rise in personal bankruptcy, particularly among low income households suggesting that fintech loans have a positive impact on consumer credit performance. However, Chava and Paradkar (2018) look at the credit profile of marketplace lending borrowers who borrowed from LendingClub to consolidate credit card debt and find that, initially, borrowers do reduce their credit card debt, but within three quarters, they received more credit from their existing banking relationships and experienced a significant increase in credit card default. They do not take into consideration important information, such as whether the borrowers also borrowed from other P2P lenders, whether the borrowers have only one loan or multiple loans with P2P lenders, and so on. They find that subprime borrowers felt the greatest impact. Di Maggio and Yao (2018) find similar results.

Looking at the funding side of fintech consumer loans, Kraussl, Kraussl, Pollet, and Rinne (2018) point out that LendingClub's portfolio generated positive abnormal returns and, therefore, could attract capital more easily to finance loan growth. In contrast, Balyuk and Davydenko (2018) note that marketplace lending platforms have evolved from trading venues into credit intermediaries. The fact that these platforms often have little skin in the game makes the fintech lending market vulnerable to large institutional investors' withdrawal from the market. The results thus far have been mixed regarding the impact on consumers, investors, or the economy overall. This is not surprising as the business models and credit evaluation techniques in the fintech space have been evolving rapidly over the past decade.

We use a unique dataset that allows us to compare online alternative lending rates with traditional credit card loans. We compare account-level credit card data that large banks submitted to the Federal Reserve for stress testing with online consumer loans that were made for credit card payoff (and debt consolidation) purposes. These data will allow us to investigate the determinants for risk pricing used by LendingClub and the performance of these loans over time, as well as serving to compare these loans with similar loans made by traditional banks. If loans are appropriately risk priced, this will provide some evidence that borrowers are not being enticed to borrow because they are offered inappropriate interest rates and that some of the efficiencies that accrue to marketplace lenders are being passed through to borrowers.

3 | THE DATA

We use four main sources of data in this paper: (a) data on loans that were originated through online alternative channels (loan-level data from the LendingClub consumer platform); (b) data on loans that were originated from traditional banking channels (loan-level data from Y-14 M reports submitted by bank holding companies with over $50 billion in total assets): (c) deposit market concentration data and bank branch information based on the Federal Deposit Insurance Corporation (FDIC) Summary of Deposits database; and (d) economic factors from the U.S. Census Bureau and Haver Analytics database.

3.1 | Fintech loans

Our research on fintech consumer lending focuses on LendingClub for two reasons. First, the company is one of the few lenders that have made their data publicly available. In addition, LendingClub is the largest fintech lender for personal loans. Therefore, the results here are likely to apply more broadly. LendingClub reports detailed information about each loan application that has been approved or denied since its inception in 2007. For each of the loans that were funded, we collect characteristics of the borrowers (e.g., FICO scores at the time of loan application, length of employment, debt-to-income (DTI) ratio, homeownership, and borrower's zip code) and characteristics of the loan (such as the loan rate, maturity of the loan [3 or 5 years], origination date, whether the verification was needed, and the loan purpose as identified in the application). We also collect the monthly payment and performance of each loan in the sample from the origination month to 24 months after origination.

Our sample includes all consumer loans that were originated from 2007 to 2015. The sample ends at 2015 (origination year) to allow us to observe loan performance over a 2-year post-origination period. In addition, we include LendingClub consumer loans in which borrowers identified as being used only to pay off credit card balances or for debt consolidation so that the sample is directly comparable with credit card loans originated by banks. At least

80% of LendingClub consumer loans in each year were used for this purpose. As of 2015, about 90% of the LendingClub consumer loans are specified as being used to pay off credit card balances or for debt consolidation (Jagtiani & Lemieux, 2018). For most of the analysis, we focus on 725,800 loans that were originated from 2010 to 2015 as data from the 2007–2009 origination vintages are less reliable and the volume was initially very small. About 76% of the sample loans were originated during the last 2 years of the sample, 2014–2015. We observe the differences between these two lending channels in terms of credit risk rating, the price of credit, and loan performance.

3.2 | Traditional loans

We use comparable loan-level (credit card loan) data from the Federal Reserve's Y-14 M reports that are reported monthly by bank holding companies with over $50 billion in assets. We use a 1% random sample of all credit card accounts reported in the Y-14 M dataset. From this dataset, we focus on the reporting period 2014–2017 and include only those accounts that were originated in 2014–2015 allowing for up to a 2-year performance period until 2017. We note that these data are constrained by the limited number of reporters and, as such, may not represent the entire population of firms that issue credit cards. However, Y-14 M reporters do represent over 80% of all credit cards issued by commercial banks. We do not include accounts that were originated prior to 2014 to avoid sample selection bias in our analysis. Accounts that were originated earlier and were closed (owing to default or other reasons) would have been dropped from the Y-14 M reports in 2014–2017. Our final sample includes 53,186 Revolver accounts (i.e., consumers who are actually borrowing, not just using, credit cards for transaction purposes) that were opened in 2014–2015. The sample includes only consumer cards (business cards and corporate cards are not included) that were issued for general purposes and private label cards. Charge cards are excluded because there is no associated credit limit for these cards. Credit card loans from the Y-14 M reports and LendingClub loans, which are used to pay off credit card loans, are the most comparable products.

It is important to note that reported credit card balances are balances as of a specific reporting date, rather than balances at the end of a statement, which varies across card accounts. The reported card balances primarily reflect spending rather than extensions of credit. To correctly compare fintech platform loans and traditional loans, we identify whether each card account is a Revolver or a Transactor. Most cardholders are Transactors, and they do not actually borrow from the bank. Because consumers report that they borrow from LendingClub to pay off their credit cards, we compare the price (interest rate spread) and performance of LendingClub loans with credit card loans controlling for loan size, origination year, and other relevant risk factors. Some credit cards have rewards (cash back or points) and/or a low rate promotion period (e.g., in the first 6 months) to encourage balance transfers from other cards. We also control for the promotion period and rewards in our analysis.

For the most part, the Y-14 M reports contain similar information on the borrowers and other risk characteristics as those reported in the LendingClub database (e.g., origination date, origination amount, location of the borrowers, and borrowers' credit scores). A few key variables (i.e., homeownership and DTI ratio) are available for LendingClub loans, but are not reported by banks in the Y-14 M report.

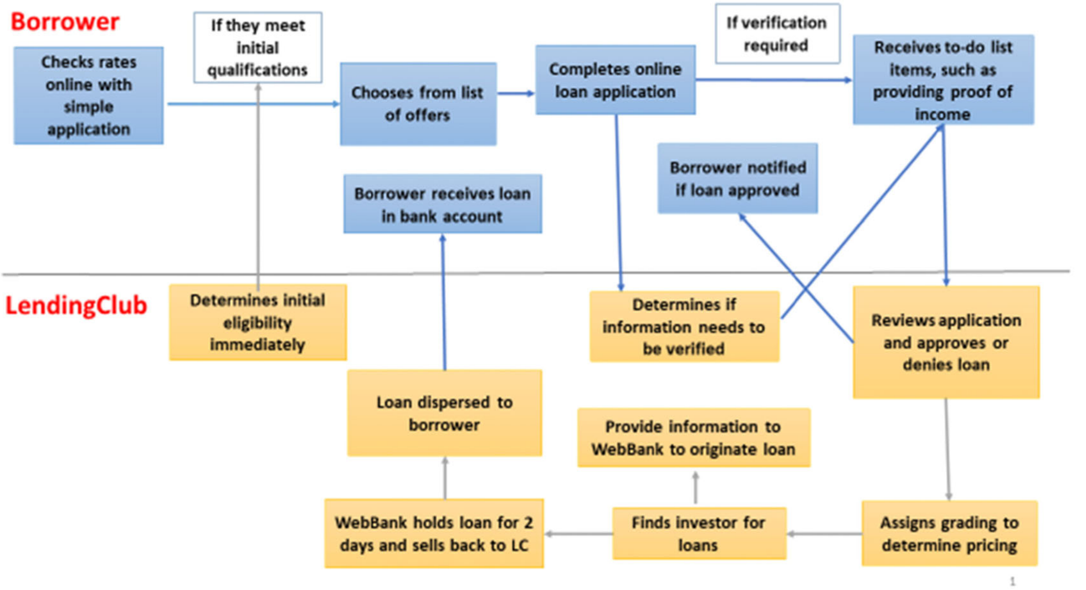
3.3 | Economic and other control factors

We collect various economic factors (e.g., local unemployment, local average household income, local home price index, and local population) from the U.S. Census Bureau database and the Haver Analytics database. We use the most appropriate and most granular level (three-digit zip code, five-digit zip code, or county) of economic factors. LendingClub loan-level data are reported at the three-digit zip code level, thus, the three-digit zip code level of economic factors is used in these cases. Using the share of outstanding credit card loans at each banking firm, we calculate local market concentration as measured by the Herfindahl-Hirschman Index (HHI).

4 | THE ROLE OF ALTERNATIVE DATA

One of the attractive features of obtaining credit from alternative lenders is how quickly lending decisions are made. An important advantage for fintech lenders is that they have access to nontraditional data sources that are not used, or not available to, by traditional bank lenders. The additional sources of information may include consumers' payment history (e.g., utility, rent, phone, alimony, and so on) history, cash flow data from bank accounts revealing recurring payments and transaction activities, such as salary and cash withdrawals, credit card transactions, medical and insurance claims, education and major, and social networks, as well as their online footprints, shopping habits, and other personal information. Some of these data are available through the various data aggregators and vendors that work directly for the lenders that provide white label services or through partnerships with the lenders. Consumers are often required to authorize lenders to access account information from their banks, credit cards, investments, or their mobile phone. Information about the timing (e.g., applying for a loan at 3:00 a.m. may not be a good signal) or location (e.g., located in a high crime area) could potentially be included in the alternative information set. These factors are not reflected in traditional credit measures, such as risk scores. Some lenders, such as PayPal, Square, and Amazon, have access to cash flow information from their own platforms allowing them to lend to small businesses and start-ups that have difficulty getting credit through traditional lending channels due to thin credit files.

In the case of LendingClub, as applications are submitted online, LendingClub's credit model grades and prices the loan, and applicants receive immediate feedback about the loan terms for which they are qualified. A verification process takes place before the loan is funded. For example, if the credit model data sources indicate the application is fraudulent, the application may be declined. If not, after an offer is presented, further income or employment verification may be requested. LendingClub has its own proprietary models that identify whether each loan application should be verified or not. As of 2015, about 70% of all loans made through the LendingClub platform were verified. Consumers are assigned a rating grade from A to G (with subgrade A1, A2, … G5) based on the full set of information after the loan has been approved. The loan application process can be summarized as shown in the diagram below.



Our analysis indicates that the use of alternative data in assigning rating grades has been increasing over the years. The correlation between LendingClub rating grades and FICO scores declined from about 80% for loans that were originated in 2007 to less than 35% for loans that were originated in 2015 (see Figure 1). We convert LendingClub's

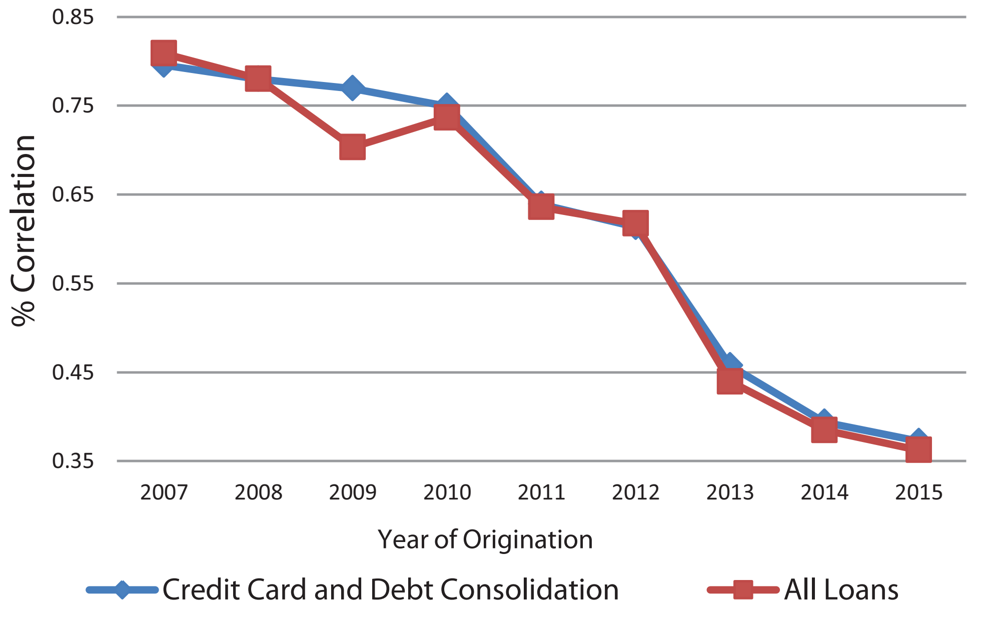


FIGURE 1 Correlation between origination FICO and rating grade assigned by LendingClub [Color figure can be viewed at wileyonlinelibrary.com]
Source: LendingClub data.

rating grades to numerical values where A is seven, B is six, …, and G is one. The declining correlation is robust. We also tried calculating the correlation when both the rating grades and the FICO scores are grouped into segments (FICO score is one if the FICO score is lower than 680; the FICO score is two, three, and four if it is 680 to 700, 700 to 750, and above 750, respectively), which also indicates that the correlation fell from 81% (for loans that were originated in 2007) to 36% (for loans that were originated in 2015).

The credit grades are increasingly defined using additional metrics beyond factors that are important in determining FICO scores. LendingClub has documented that its credit models have the Kolmogorov–Smirnov scores that outperform generic scores by identifying strong borrowers with lower FICO scores and vice versa.[3] Figures 2a, 2b, and 2c present the composition of loans for each rating grade and how that composition has evolved over the years for loans originated in 2007, 2011, and 2015, respectively. Some consumers who would be considered subprime are slotted into the "better" loan grades. For loans that originated in 2015 (see Figure 2c), about 8% that were A-rated were to borrowers with FICO scores below 680 (so-called subprime) and 28% of the B-rated borrowers had FICO scores in the subprime range. This provides evidence that the use of additional information sources could allow some borrowers with low FICO scores to access credit and potentially better pricing.

We further explore payment and default behavior of these subprime borrowers who are rated highly by LendingClub. Figure 3a illustrates the probability of loans becoming delinquent (at least 60 days past due [DPD]) within 24 months after origination for these subprime borrowers. The default probability varies significantly across rating grades with the average probability of default (PD) below 5% for A-rated versus an average PD of over 35% for G-rated even though they were all rated below 680 based on FICO scores. We also observe the same probability of becoming delinquent during a shorter performance window of 12 months. The average PD was 3% for A-rated borrowers and 19% for G-rated borrowers for this 1-year performance period. The results are

[3] See the link from the LendingClub site for more details at https://www.lendingclub.com/public/income-verification.action.

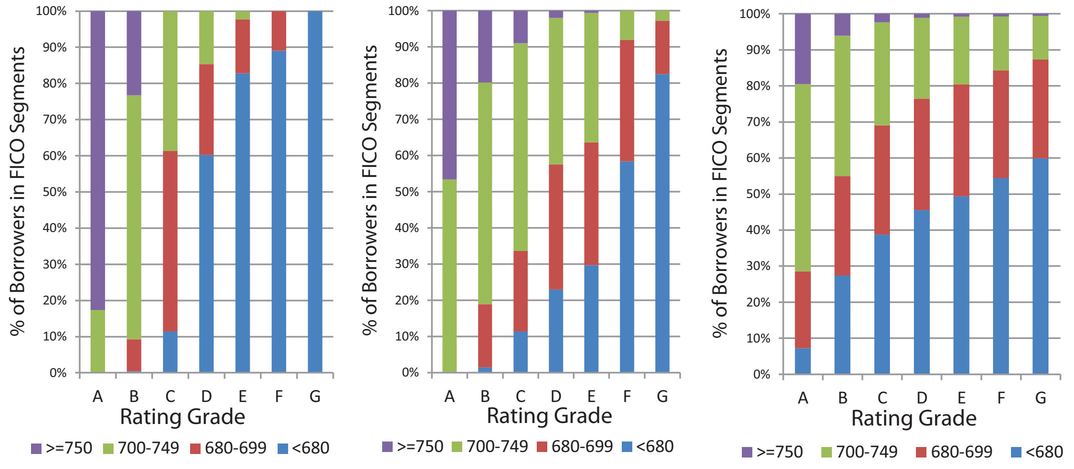


FIGURE 2 (a) FICO distribution by LendingClub rating 2007 origination; (b) FICO distribution by LendingClub rating 2011 origination; (c) FICO distribution by LendingClub rating 2015 origination [Color figure can be viewed at wileyonlinelibrary.com]
Source: LendingClub data.

robust when the performance period is expanded to 2 years. LendingClub's use of alternative data seems to enhance its ability to identify those subprime borrowers who are actually not risky—so-called the invisible prime borrowers.

Similarly, Figure 3b presents the average PD for all loans that were originated in 2014–2015 by rating grades and FICO brackets. The average PD increases as the rating grades move from A to G. A-rated borrowers, on average, have smaller PDs, whereas F-rated and G-rated borrowers have higher PDs regardless as to their FICO scores. Superprime borrowers with FICO scores above 750, who were slotted into the F- and G-rated segments by LendingClub, perform poorly with an average PD of about 40%. Again, the use of alternative data has allowed subprime borrowers who are not risky to be separated from those who are and to potentially receive a loan at a better price.

5 | PRICING OF CREDIT—FINTECH VERSUS TRADITIONAL LOANS

In this section, we explore the pricing of LendingClub loans versus similar loans from traditional lenders. Pricing is measured in terms of the credit spread between the reported interest rate and the matching Treasury rates for the same time to maturity. LendingClub uses its own loan grades to differentiate interest rates offered to borrowers. Therefore, it is not surprising that we observe a tight relationship between rating grades and interest rate spreads throughout the sample period. Figure 4 indicates that better rated borrowers receive loans at a lower price (i.e., smaller spreads). The relationship between the loan grades and spreads persists even after controlling for other relevant risk and economic factors in the regression analysis.

We demonstrate in Figure 4 that although the rating grade and spreads are consistently in rank order over the years, the spread differential between the A- and G-rated borrowers widened significantly from about 6% to more than 20% for loans originated in 2015 when more alternative data were being used in credit decisions as compared with earlier vintages. Figure 5 indicates the average PD for all of the loans in each rating grade by year of origination. The average PD and the rating grades line up better for loans that were originated in later years. Focusing on loans that were originated in 2015, the subprime borrower, with a FICO score below 680, who was slotted into a B grade would have

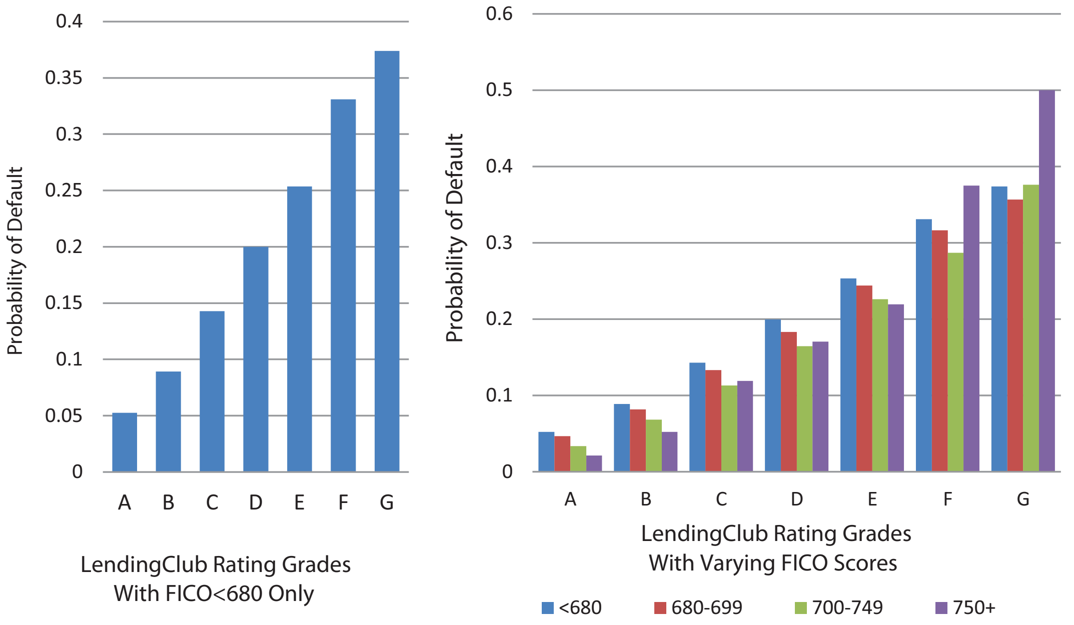


FIGURE 3 (a) Probability of being ≥60 DPD within 24 months after origination—for loans originated in 2014–2015 with FICO score <680 only. (b) Probability of being ≥60 DPD within 24 months after origination, for loans originated in 2014–2015, by credit scores and rating grades [Color figure can be viewed at wileyonlinelibrary.com]
Source: LendingClub.

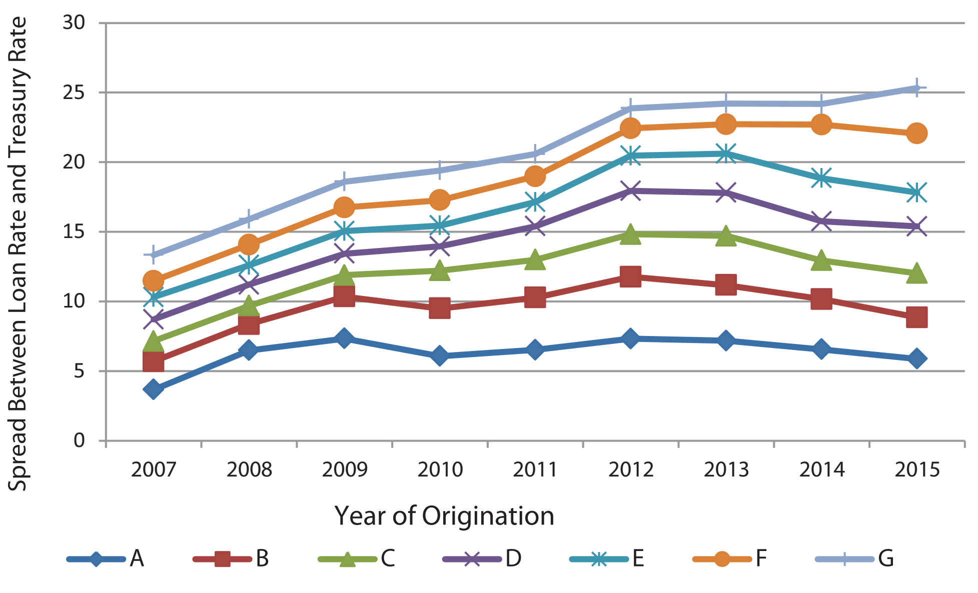


FIGURE 4 Average spread by rating grades—cards and debt consolidation (2007–2015) [Color figure can be viewed at wileyonlinelibrary.com]
Source: LendingClub data; treasury rates from the Bloomberg database.

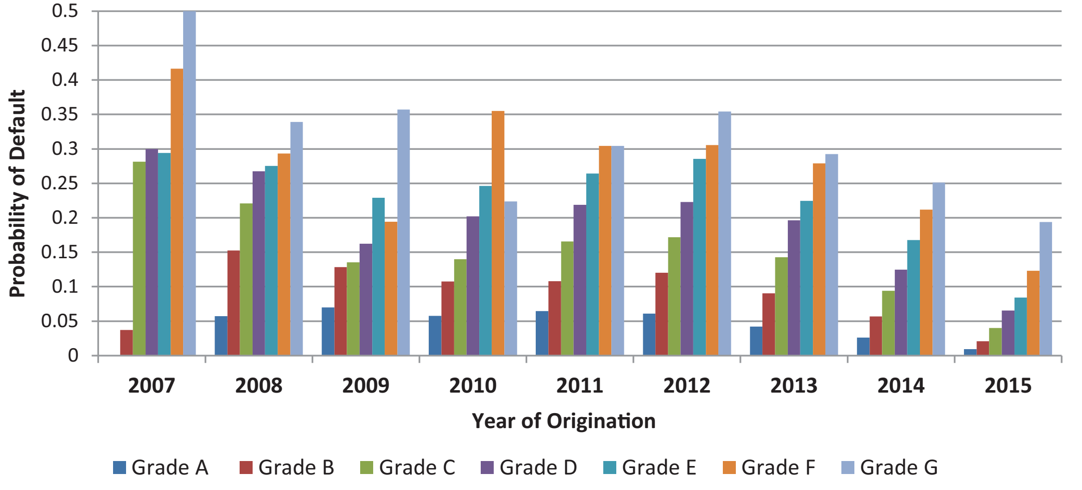


FIGURE 5 LendingClub loans ≥60 DPD within 12 months after origination—by loan grade and years [Color figure can be viewed at wileyonlinelibrary.com]
Source: LendingClub loans (cards and debt consolidation purposes only).

had to pay approximately 25% over Treasuries instead of 9% over Treasuries had he been slotted into the G-rated. In the next section, we see that a greater probability of default is observed for loans that were appropriately subject to larger credit spreads (i.e., higher price). The interest rate spreads appear to have a strong relationship with the likelihood of becoming delinquent. Again, the use of additional information allows some borrowers who would be classified as subprime by traditional criteria to be slotted into better loan grades thereby obtaining lower priced credit. More importantly, it does not appear that this credit is mispriced in terms of default risk as shown earlier in Figure 3a where the average PD for these subprime borrowers is closely related to loan grades.

Thus far, we have observed a tight relationship between LendingClub's proprietary rating grades and the credit spreads that LendingClub charged. We also find that the correlation between LendingClub's rating grades and FICO scores has declined dramatically over the years from about 80% for loans originated in 2007 to about 35% for loans originated in 2014–2015 indicating an increasing role of alternative data sources used by LendingClub. For loans that were originated in 2015, some of the A-rated borrowers actually had FICO scores below 680 and were able to access credit at a lower rate. Following up on these borrowers two years later, we find that they did not have an increased likelihood of default. The rating grades assigned by LendingClub have been effective in identifying the invisible prime from the subprime pool of borrowers. LendingClub consumer loans only come in only two maturities: 3 or 5 years.

Table 1 reports the comparison of interest rate spreads that borrowers are charged on LendingClub loans used to pay off credit card balances versus the spreads that borrowers are charged on traditional credit card loans for borrowers with the same FICO scores. For loans that were originated from 2014 to 2015, spreads on credit card loans are significantly higher than those on LendingClub loans regardless as to the maturity of the loans (3 or 5 years). The spread differentials (i.e., the savings to consumers) range from about 8% for those with FICO scores below 680 to more than 10% for superprime borrowers (i.e., those with FICO scores of 800 or above). Because LendingClub charges origination fees, this analysis slightly overstates the difference.[4] Holding the FICO scores constant, LendingClub borrowers pay less than traditional credit card borrowers.

Further, we find that holding the probability of default fixed, LendingClub borrowers pay smaller spreads than credit card borrowers. Figures 6a and 6b compare delinquency rates across the credit spread brackets for LendingClub loans

[4]LendingClub interest rates, as reported on the LendingClub website, do not include one-time origination fees that range from 1% to 5% of the origination amount depending upon the rating grade of the borrowers. The origination fee is usually deducted from the total loan amount resulting in an approximately 1% and up to 2.5% increase in the effective annual percentage rate (APR). The interest rate from the Y-14M data is an APR.

TABLE 1 Comparing the price of credit: LendingClub loans versus Y-14 M credit card loans (revolvers only)

FICO segment at origination	% Average APR spread LendingClub (APR has incorporated the origination fees charged by LendingClub)		% Average spread bank Y-14 M (revolvers only)	Significant difference at the 1% level?	
	3-Year maturity	5-Year maturity		3-Year	5-Year
660–679	15.336 *N* = 139,337	18.113 *N* = 64,359	20.1923 *N* = 6,812	Yes	Yes
680–699	13.756 *N* = 100,033	16.764 *N* = 54,030	19.8465 *N* = 7,067	Yes	Yes
700–719	12.013 *N* = 64,271	15.351 *N* = 36,313	19.1418 *N* = 6,637	Yes	Yes
720–739	10.432 *N* = 32,512	14.033 *N* = 17,071	18.4180 *N* = 5,930	Yes	Yes
740–759	9.125 *N* = 15,403	12.818 *N* = 6,823	17.6569 *N* = 5,383	Yes	Yes
760–779	8.236 *N* = 8,081	11.972 *N* = 3,015	16.8312 *N* = 4,701	Yes	Yes
780–799	7.604 *N* = 4,458	11.338 *N* = 1,436	16.1820 *N* = 4,586	Yes	Yes
800+	6.9519 *N* = 2,509	10.699 *N* = 837	16.1668 *N* = 12,070	Yes	Yes

Note. Sample period: Loans originated in 2014–2015. Only credit spreads on credit card loans are significantly higher than consumer loans from LendingClub (regardless as to the loan maturity), even after controlling for the borrower's FICO score and accounting for the one time origination fees charged by LendingClub.

to pay off credit card balances versus credit cards loans for the period 12 months and 24 months after origination, respectively. We focus on loans that were originated during 2014 and 2015 for both the LendingClub and Y-14 M data.[5] From the Y-14 M data, we include only credit cards that carry a balance (i.e., Revolvers). Cards that involved initial promotion low interest rates are also excluded from this analysis. The analysis indicates that the average PD is consistently higher for LendingClub loans for each segment of spreads suggesting that if we hold the PD fixed, spreads are smaller for LendingClub loans than for Y-14 M credit card loans. This makes sense as otherwise consumers would have little reason to take a loan from LendingClub to pay off their credit card balances. Given the same credit risk (i.e., expected delinquency rate), consumers would be able to obtain credit at a lower rate through LendingClub than through traditional credit card loans offered by banks.

A few additional statistics are reported here related to the characteristics of LendingClub borrowers relative to traditional borrowers and their true credit risk relative to traditional borrowers. Figure 7 reports that for loans that were made to borrowers with the same FICO score brackets and originated in the same period (i.e., 2014–2015), the delinquency rate is slightly higher for LendingClub borrowers than for credit card borrowers. (Note that a small number of credit card loans reported in Y-14 M have missing FICO scores at origination and are noted in the missing FICO category in Figure 7.) All cards that involved the initial promotion low interest rates are excluded from this analysis. These results imply that for consumers with the same FICO scores, those who borrow from LendingClub tend to have a greater risk of being delinquent, on average. In addition, we find that among LendingClub borrowers with the same rating grades (A–G), homeowners are less likely to become delinquent, on average, as illustrated in Figure 8.

[5] We do not include credit card accounts from the Y-14M database that were originated prior to 2014 to avoid the sample survival bias as cards that defaulted and were closed prior to 2014 would not be included in the Y-14M reports as of 2014.

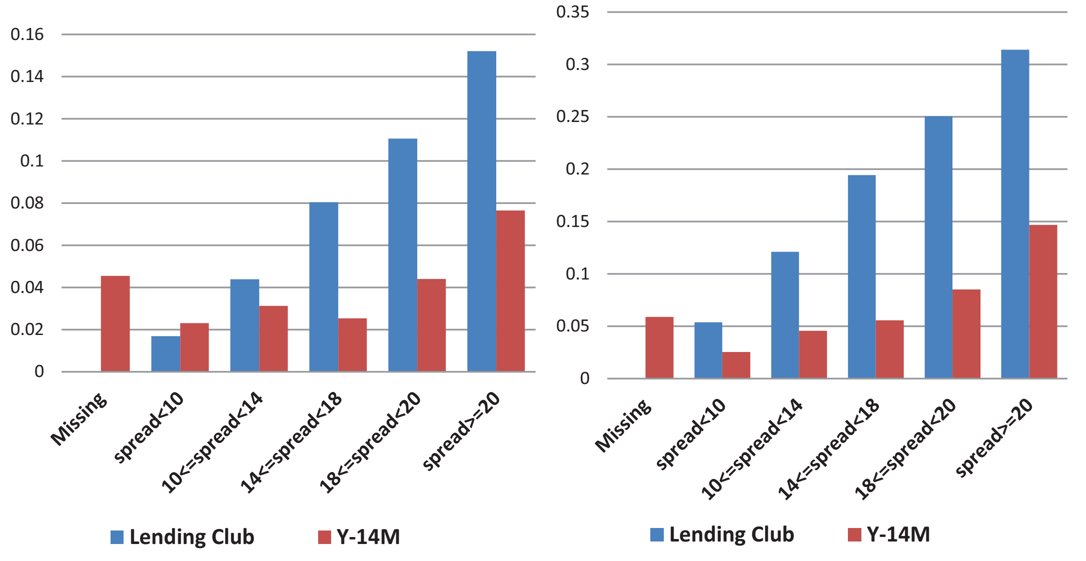


FIGURE 6 All loans were originated from January 2014 to December 2015. Delinquency status (became ≥60 DPD) is observed for the period within 12 months after loan origination. (a) LendingClub loans versus Y-14 M credit card loans (Revolvers Only)—probability of ≥60 DPD within 12 months after origination. (b) LendingClub loans versus Y-14 M credit card loans (revolvers only)—probability of ≥60 DPD within 24 months after origination [Color figure can be viewed at wileyonlinelibrary.com]
Sources: LendingClub loans (cards and debt consolidation purposes only) and Y-14 M data on credit cards.

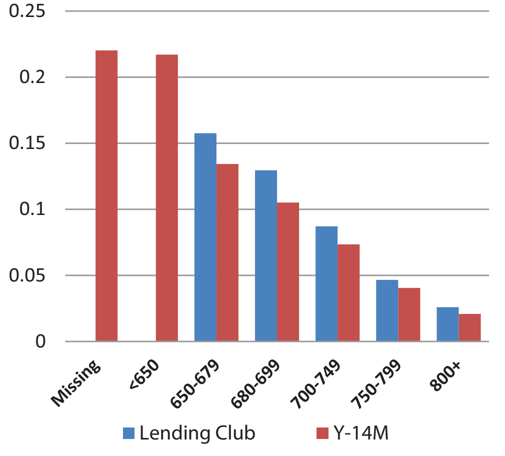


FIGURE 7 LendingClub versus Y-14 M cards (Revolvers Only): probability of ≥60 DPD within 24 months—by FICO scores [Color figure can be viewed at wileyonlinelibrary.com]
Sources: LendingClub loans (cards and debt consolidation purposes only) that were originated in 2014 and 2015 only; Y-14 M data on credit card accounts were issued to consumers from 2014 to 2015.

Finally, we explore whether LendingClub is more willing to make larger loans to more creditworthy borrowers and smaller loans to those who may have trouble getting credit through traditional channels. The data we have do not allow us to see whether the origination amount is less than the initial amount requested, but we explore the size of loans originated by LendingClub controlling for different credit risk variables. Figure 9a presents the share of LendingClub loans in the various brackets of origination balances controlling for FICO scores. The distribution of loan balances across the

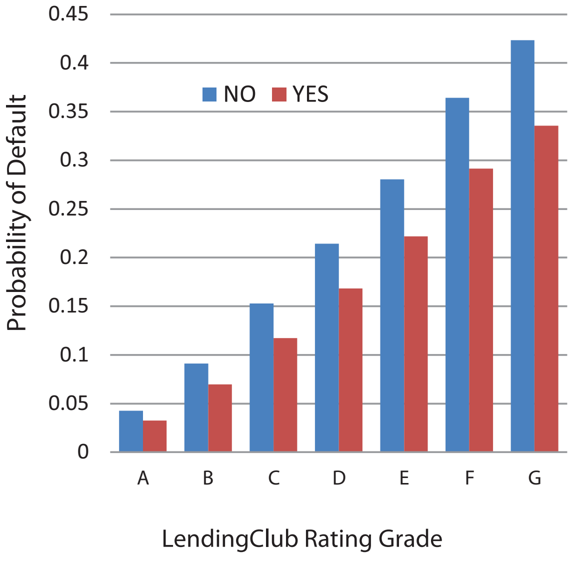


FIGURE 8 LendingClub loans probability of ≥60 DPD within 24 months by loan grades and homeownership [Color figure can be viewed at wileyonlinelibrary.com]

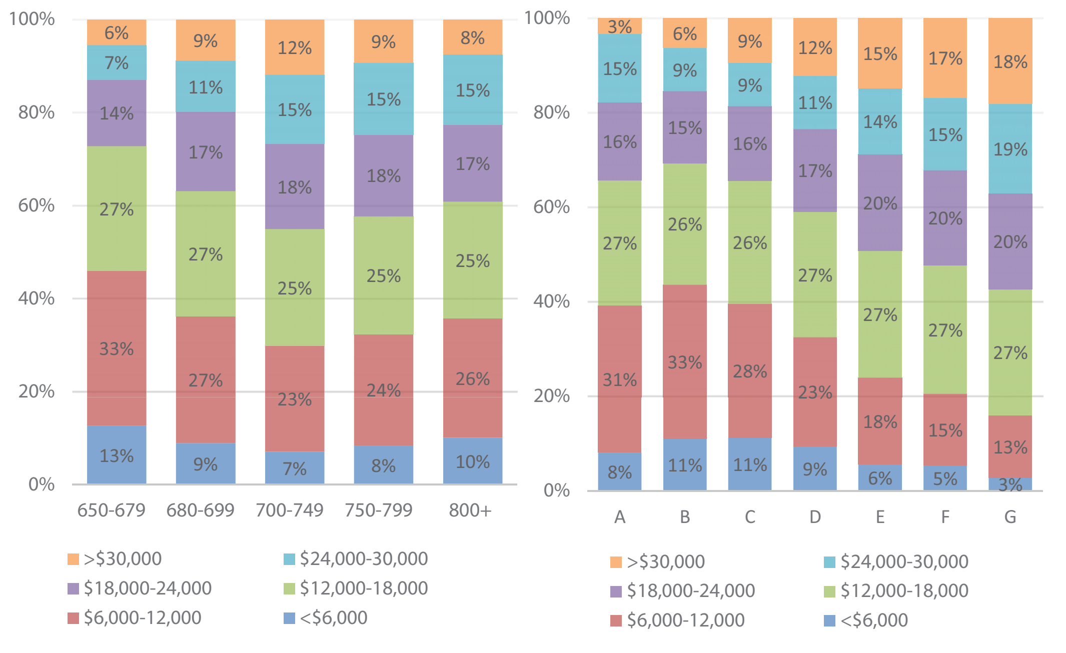


FIGURE 9 (a) Number of accounts of LendingClub loans (originated in 2014–2015) by origination balance, controlling for FICO scores. (b) Number of accounts of LendingClub loans (originated in 2014–2015) by origination balance, controlling for rating grades [Color figure can be viewed at wileyonlinelibrary.com]
Source: LendingClub loans (cards and debt consolidation purposes only).

FICO scores does not vary significantly for all borrowers with FICO scores above 680. Average loan balances seem to be smaller for borrowers with FICO scores below 680. Less than 10% of loan accounts originating in 2014–2015 were made with an origination balance above $30,000, and less than 25% of loan accounts were originated with a balance above $24,000.

Unlike with FICO scores, when we control for the rating grades A–G, Figure 9b suggests a relationship between the loan origination amount and the rating grade, but the relationship is the opposite of what one would expect under a credit rationing scenario. Lower loan grades (F-rated and G-rated) are associated with larger origination amounts. About 18% of G-rated borrowers received a loan with an origination amount larger than $30,000 compared with only 3% and 6% for A-rated and B-rated borrowers, respectively. Almost 40% of G-rated borrowers received a loan from LendingClub with an origination amount larger than $24,000. Of these G-rated borrowers who receive a large loan (i.e., at least a $30,000 origination amount) from LendingClub, 90% of them have FICO scores below 700 and about 65% of them have FICO scores below 680. These results suggest that LendingClub does not try to reduce its risk taking when making loans to poorly rated borrowers by giving them smaller loans.

6 | REGRESSION ANALYSIS

Our analysis thus far indicates that for loans that originated in 2014 and 2015, LendingClub's rating grades A–G are based on information that is not highly correlated with the borrowers' FICO score and seem to do a good job of identifying the invisible prime (i.e., those who are less risky subprime borrowers). The rating grades are highly related to the borrowers' probability of becoming delinquent on their loans within 2 years of loan origination. For robustness testing, the regression analysis in this section will demonstrate that the rating grades, which contain alternative data, are superior to FICO scores in predicting default and for accurate risk pricing even after controlling for a set of other risk factors.

First, based on our logistic regression analysis (coefficients are not reported here) of default probability (e.g., being at least 60 DPD within 2 years after origination), we present the receiver operating characteristic (ROC) curves in Figure 10. We plot the ROC curves for four different default probability model specifications based on the following sets of explanatory variables: (a) FICO scores only, (b) rating grades A–G only, (c) FICO scores and other control factors, such as borrower's income, debt-to-income ratio, length of employment, number of credit inquiries prior to loan application, homeownership, and local economic environment (e.g., home price index [HPI] and unemployment rate), and (d) rating grades A–G and the same set of other control factors as in Model (3). The results in Figure 10 indicate that Model (1), which uses FICO scores to predict delinquency, does not perform as well as the other three models. Its ROC lies closest to the 45 degree line. In addition, we find that Model (2), which uses both FICO scores and the relevant set of risk factors, does not perform as well as Model (3), which uses only rating grades A–G without other risk factors, to predict delinquency over the 24 months after origination. Model (4), using both rating grades A–G and the same set of risk factors used in Model (2), performs slightly better than using the rating grades alone. These results are consistent with our earlier findings that the rating grades assigned by LendingClub are more powerful in predicting a borrower's default probability than a set of FICO scores, other traditional risk variables, and economic factors combined.

Furthermore, the regression results presented in Table 2 demonstrate that the rating grades, which are highly correlated, as expected, with interest rates that the borrowers are charged, are better at predicting the borrower's default probability than FICO scores, other borrower risk characteristics, and the economic conditions combined. The dependent variable is the interest rate spread (on LendingClub loans).

First, we focus on Columns 1 and 3 of Table 2. The key independent variables are the various rating grades in Column 1 and the FICO score segments in Column 3. The results indicate that there is a strong relationship between rating grades and credit spreads with an adjusted R-square of almost 90% as indicated in Column 1. The coefficients for rating grades are all statistically significantly positive and in rank order in which the coefficients are positive for B-rated loans and the coefficients are largest (positive) for G-rated loans. Unlike Column 1, the relationship between credit spreads and FICO scores at origination (in Column 3) is not as tight with an adjusted R-square of only about 18%. The coefficients for FICO scores are, as expected, positive, statistically significant, and in rank order. These results confirm that although FICO scores have been used by fintech lenders as an initial broad measure of credit risk, FICO scores alone are not granular enough to sufficiently predict each consumer's default probability or to be used for risk pricing.

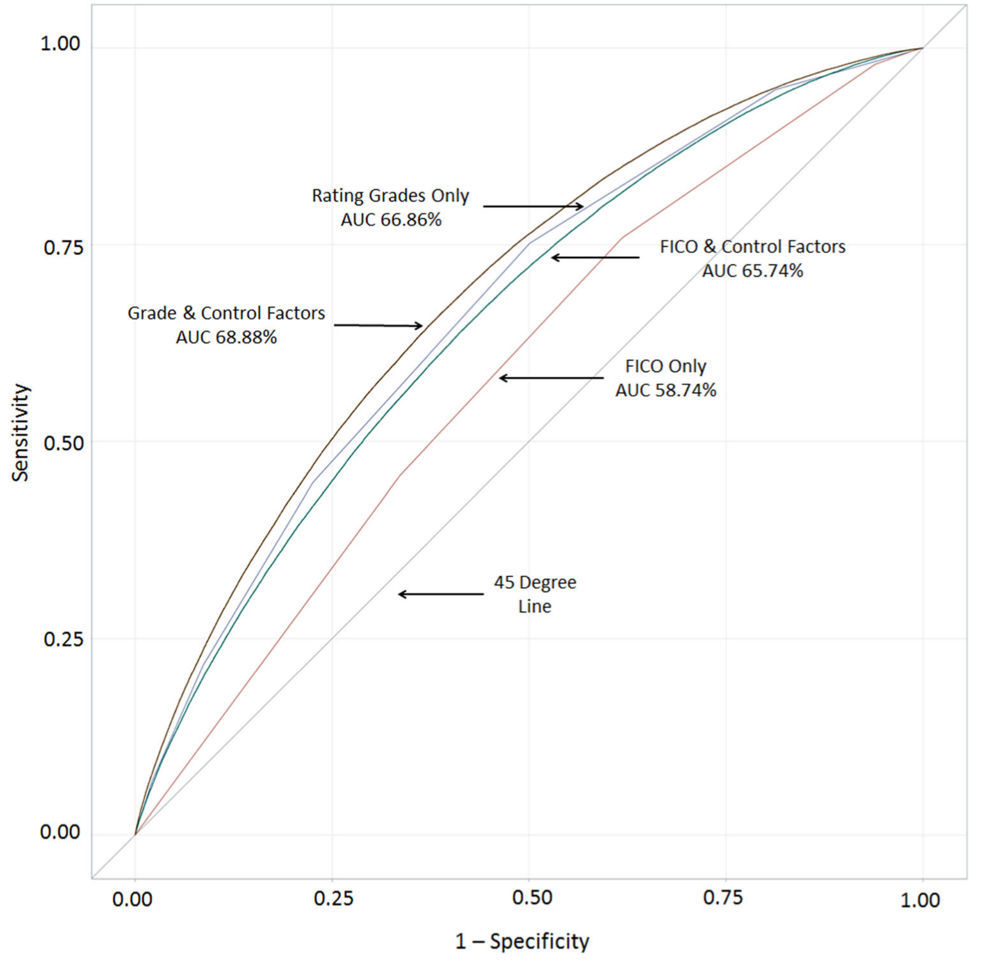


FIGURE 10 This figure illustrates the discriminatory power of four different models of default probability specifications by providing the receiver operating characteristics curve (ROC curve) and the area under curve (AUC). The ROC curves are estimated using a logit regression of the default dummy (at least 60 DPD within two years after origination) on: (a) FICO scores only, (b) rating grades only, (c) FICO scores and other control factors, and (d) rating grades and the same set of other control factors [Color figure can be viewed at wileyonlinelibrary.com]
Source: LendingClub data and economic factors from the U.S. Census Bureau and Haver Analytics database.

Next, we focus on Columns 2 and 4. We include additional control factors that are intended to capture borrowers' risk characteristics and the local economic environment. Borrower risk characteristics include the DTI ratio at origination, homeownership, length of employment, income, loan amount, and the number of consumer's credit inquiries during the period before loan origination. Economic factors included in the analysis are the local unemployment rate, the local HPI, year dummies, and the HHI measure of the local credit market concentration in the borrower's zip code. Most important, in Columns 2 and 4, we also include a dummy *D(default within 24 months after origination)* indicating whether the loan defaulted (i.e., at least 60 DPD within 24 months after loan origination) and another dummy *D(2014-15)* × *D(default within 24 months after origination)* indicating whether the loan originated in 2014–2015 and defaulted within 24 months after origination date.

The coefficient of the default indicator, *D(default within 24 months after origination)*, is positive and significant in both Columns 2 and 4 indicating the positive relationship between credit spreads and the actual default probability is not fully captured by the other risk measures. We note that the coefficient is, however, much larger in Column 4 than in

TABLE 2 Regression results—LendingClub consumer loans important factors that determine credit spreads

	LendingClub rating grades A–G		Origination FICO scores	
Independent variable	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	6.3339***	9.7106***	6.4811***	8.1175***
	(.0001)	(.0001)	(.0001)	(.0001)
D(Default within 24 months after origination)		0.2197***		1.3499***
		(.0001)		(.0001)
D(2014-15) × D(default within 24 months after origination)		−0.0979***		0.3387***
		(.0001)		(.0001)
D(rating grade B)	3.5325***	3.2572***		
	(.0001)	(.0001)		
D(rating grade C)	6.5685***	6.3532***		
	(.0001)	(.0001)		
D(rating grade D)	9.6647***	9.4415***		
	(.0001)	(.0001)		
D(rating grade E)	12.2472***	12.1488***		
	(.0001)	(.0001)		
D(rating grade F)	15.9334***	15.8618***		
	(.0001)	(.0001)		
D(rating grade G)	18.0956***	18.1991***		
	(.0001)	(.0001)		
D(650< FICO at origination <680)			7.2882***	6.0992***
			(.0001)	(.0001)
D(680< FICO at origination <700)			6.0733***	4.8362
			(.0001)	(.0001)
D(700< FICO at origination <750)			3.9252***	0.6505***
			(.0001)	(.0001)
D(750< FICO at origination <800)			1.0541***	0.6505***
			(.0001)	(.0001)
D(homeownership)		−0.0819***		−0.2051***
		(.0001)		(.0001)
D(employment >10 years)		0.0082***		0.1540***
		(.0044)		(.0001)
Debt-to-income ratio at origination		0.0037***		0.0647***
		(.0001)		(.0001)
Log(borrower's income)		−0.1292***		−1.597***
		(.0001)		(.0001)
Log(origination loan amount)		−0.0493***		1.7489***
		(.0001)		(.0001)

(Continues)

TABLE 2 (Continued)

Independent variable	LendingClub rating grades A–G (1)	(2)	Origination FICO scores (3)	(4)
Number of credit inquiries 6 months before	−2.27	0.1076***		0.8983***
		(.0001)		(.0001)
Home price index (3-digit zip)?		0.0001***		−0.0000
		(.0001)		(.8678)
Unemployment rate (3-digit zip)?		0.0064***		0.0280***
		(.0001)		(.0001)
D(origination year 2014)		−1.3258***		−1.2583***
		(.0001)		(.0001)
D(origination year 2015)		−2.1937***		−2.2765***
		(.0001)		(.0001)
D(Y-14M card loans HHI >2,500)		0.0029		−0.0749***
		(.5825)		(.00001)
Adjusted R^2	88.63%	93.42%	17.62%	34.25%
Observation number (*N*)	725,800	663,576	725,800	663,576

Note. The sample period begins in 2013 in columns 2 and 4 owing to the unavailability of reliable Y-14 M data prior to 2013. The data are used to calculate the HHI market concentration measure. Also note that all loans were originated up to 2015 to allow 24 months of performance to observe the loans' default behavior.
Sample period: 2010—2017.
Data are at the loan level from LendingClub's consumer platform (for credit cards or debt consolidation only). All loans were originated from 2007 to 2015 with a 2-year performance period ending in 2017 or earlier. The dependent variables are interest rate spreads calculated as the difference between the interest rates charged on the loans and the equivalent risk-free loans (the U.S. Treasury rate of securities with the same time to maturity). The variables *Rating Grade A* and *FICO at Origination Greater Than 800* serve as the base case. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Column 2 (i.e., 1.3499 in Column 4 and 0.2197 in Column 2) implying that the default dummy picks up some of the risk factors specific to the loan and the borrower that are not captured by the FICO scores in Column 4 even after controlling for a set of other relevant risk factors.

The second dummy indicator that identifies loans made in 2014–2015 that defaulted, *D(2014-15)* × *D(default within 24 months after origination)*, is significantly negative in Column 2 but significantly positive in Column 4. In Column 2, when the rating grades and other control factors are included in the analysis, a combination of these two coefficients is a very small number (i.e., 0.2197 – 0.0979) compared to the equivalent number in Column 4 (i.e., 1.3499 + 0.3387) where FICO scores are included in the analysis instead of the rating grades. For loans that were originated in later years (2014–2015), when more alternative data were used in assigning rating grades and in credit pricing, the rating grades and the set of risk factors in Column 2 capture much of the risk for specific loans and borrowers. In contrast, when FICO scores are used instead of rating grades in Column 4, much of the default risk was not fully captured by the model resulting in significantly large positive coefficients for the two default dummy variables that indicate the borrower's actual default within 2 years after loan origination. Again, these results indicate that the interest rate spread charged by LendingClub, based primarily on the rating grades A–G and additional risk factors, is more predictive of default risk consistent with our earlier results presented in Figure 10.

Our control variables are mostly significant with the expected signs across all columns in Table 2. For example, we observe a significantly positive relationship between interest rate spreads that LendingClub charges and loan amounts and the number of credit inquiries by the borrowers within 6 months prior to loan origination (i.e., measuring how desperately the borrowers need additional credit). In addition, we find that LendingClub charges smaller credit spreads to borrowers who own a home, have been employed for more than 10 years, and have a higher income. The market concentration variable, *D(Y-14M card loans HHI >2,500)*, is either negative or insignificant implying that LendingClub is likely to offer loans at a lower rate to consumers who live in the zip codes with a high consumer loan market concentration (i.e., areas that would benefit from more lenders including fintech alternative lenders).

Overall, our results suggest that the use of alternative data could provide increased access to credit at a lower cost to those creditworthy individuals who have a thin credit history or have poor FICO scores—see Jagtiani, Vermilyea, and Wall (2018) and Goldstein, Jagtiani, and Klein (2019) for further discussion regarding the use of alternative data, big data, and machine learning in credit decisions. Fintech lenders should be cautious about which alternative data they use and keep in mind that some set of alternative data that may work well for some groups of consumers may not be representative and stable enough to be used for others depending upon how the data were collected.

7 | CONCLUSIONS

Fintech has been playing an increasing role in shaping the financial and banking landscape. Technology has allowed both banks and fintech lenders to serve small businesses and consumers without brick-and-mortar investments. In this paper, we explored the impact of fintech lending on consumers' ability to access credit and the price of credit. In addition, we examined the role of alternative information sources potentially used by these nonbank alternative lenders. Although the alternative data sources and algorithms that online alternative lenders use have allowed for faster and lower cost credit assessments, these innovations could potentially carry a risk of disparate treatment and fair lending violations.

Since our results are derived based on loans originated on the LendingClub platform, the largest personal unsecured installment lenders, one should be cautious in extrapolating the interpretation of our findings to all loans originated through other online alternative platforms. In addition, our data on traditional lenders are based on Y-14 M data, which are constrained by the limited number of reporters (i.e., only the largest U.S. banks that are subject to CCAR stress testing) excluding bank holding companies under $50 billion in total assets and credit unions.

We find that the use of nontraditional information from alternative data sources has allowed consumers with fewer or inaccurate credit records (i.e., based on FICO scores) to have access to credit. Some creditworthy consumers that have poor FICO scores have been identified using additional information and have been rated as low-risk borrowers by LendingClub. The correlation between rating grades and FICO scores declined steadily from over 80% for loans that were originated in 2007 to about 35% for loans originated in 2015. Interestingly, these rating grades (with only a 35% correlation with FICO) continued to serve as a good predictor of future loan delinquency over the next 2 years. There is additional information in LendingClub's own internal rating grades that is not already incorporated in the obvious traditional risk factors. This has enabled some borrowers to be assigned better loan ratings and receive lower priced credit.

Our previous research in Jagtiani and Lemieux (2018) presented evidence that fintech lenders fill credit gaps in areas where bank offices may be less available and provide credit to creditworthy borrowers that banks may not be serving. Our further research in this paper finds that loans from fintech lenders seem to be "appropriately" risk priced. Banks are responding to these innovations by partnering with fintech firms. This relationship is evolving quickly.

Our results provide important implications for data sharing, such as in the open banking environment, and consumer protection. Although consumers' information and privacy should be protected by laws and regulations, certain alternative information could play a key role in allowing lenders to fully understand the credit quality of potential borrowers and must be willing to grant certain consumers' access to credit that would not have been granted otherwise. Banks could potentially benefit from the alternative data sources and big data by partnering with online fintech lenders and other AI vendors. Credit rating agencies could consider incorporating nontraditional data in their credit scoring models while protecting consumer privacy. Further research remains to be done to fully explore other aspects of risk to borrowers presented by these new innovations, such as inherent bias in some alternative data, and whether these fintech lending innovations have allowed consumers to become excessively leveraged.

ACKNOWLEDGMENTS

The authors thank Leigh-Ann Wilkins, Erik Dolson, Raman Quinn Maingi, and John Nguyen for their research assistance. They also thank Onesime Epouhe for his assistance with the stress test data. Helpful comments and suggestions from Tracy Basinger, Robin Prager, Joe Hughes, Bob Hunt, Robert Wardrop, Raghu Rau, Paul Calem, Chris Cumming, Kathleen Hanley, Giuliana Borello, and participants at the following conferences are appreciated: the Annual FDIC Conference, the European Financial Management Association (EFMA), the American Economic Association (AEA), the annual NYU Fintech, the Central Bank Research Association (CEBRA), and the Financial Stability Board FIN meeting. The opinions expressed in this paper are the authors' own views and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve Bank of Chicago, or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

REFERENCES

Alyakoob, M., Rahman, M., & Wei, Z. (2017). *Where you live matters: The impact of local financial market competition in managing online peer-to-peer loans*. Purdue University Working Paper.

Balyuk, T., & Davydenko, S. (2018). *Reintermediation in fintech: Evidence from ONLINE Lending*. Emory University and University of Toronto Working Paper.

Bertsch, C., Hull, I., & Zhang, X. (2016). *Monetary normalizations and consumer credit: Evidence from Fed liftoff and online lending*. Sveriges Riksbank Working Paper 319.

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

Buchak, G., Matvos, G., Piskorski, T., & Seru, A. (2017). *Fintech, regulatory arbitrage, and the rise of shadow banks*. NBER Working Paper 23288.

Carroll, P., & Rehmani, S. (2017). *Alternative data and the unbanked*. Oliver Wyman Insights.

Chava, S., & Paradkar, N. (2018). *Winners and losers of marketplace lending: Evidence from borrower credit dynamics*. Georgia Tech Working Paper 18-16.

Chen, B. S., Hanson, S. G., & Stein, J. C. (2017). *The decline of big bank lending to small banks: Dynamic impacts on local credit and labor markets*. NBER Working Paper 23843.

Cordray, R. (2017). *Prepared remarks of CFPB Director Richard Cordray at the LendIt USA Conference*. Retrieved from https://www.consumerfinance.gov/about-us/newsroom/prepared-remarks-cfpb-director-richard-cordray-lendit-usa-conference/

Crosman, P. (2016, June 14). Will fintechs kill the FICO score? *American Banker*.

Crowe, C., & Ramcharan, R. (2013). House prices and household credit access: Evidence from prosper.com. *Journal of Money, Credit and Banking*, *45*, 1085–1115.

Danisewicz, P., & Elard, I. (2018). *The real effects of financial technology: Marketplace lending and personal bankruptcy*. University of Bristol and Shanghai University of International Business and Economics Working Paper.

Demyanyk, Y., & Kolliner, D. (2014). *Peer-to-peer lending is poised to grow*. Cleveland, OH: Federal Reserve Bank of Cleveland. Retrieved from https://www.clevelandfed.org/newsroom-and-events/publications/economic-trends/2014-economic-trends/et-20140814-peer-to-peer-lending-is-poised-to-grow.aspx

De Roure, C., Pelizzon, L., & Tasca, P. (2016). *How does P2P lending fit into the consumer credit market*. Deutsche Bundesbank Discussion Paper 30.

De Roure, C., Pelizzon, L., & Thakor, A. (2018). *P2P lenders versus banks: Cream skimming or bottom fishing?* SAFE Working Paper 206.

Di Maggio, M., & Yao, V. (2018). *Fintech borrowers: Lax-screening or cream skimming?* Harvard Business School and Georgia State University Working Paper.

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

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

Everett, C. R. (2010). Group membership, relationship banking and loan default risk: The case of online social lending. *Banking and Finance Review*, *7*, 15–54.

Federal Reserve. (2018). *Consumer credit, G.19*. Retrieved from https://www.federalreserve.gov/releases/g19/current/

Freedman, S., & Jin, G. (2017). The information value of online social networks: Lessons from peer-to-peer lending. *International Journal of Industrial Organization*, *51*, 185–222.

Gao, Q., & Lin, M. (2012). *Linguistic features and peer-to-peer loan quality: A machine learning approach*. City University of New York, Georgia Institute of Technology, and University of Arizona Working Paper.
Goldstein, I., Jagtiani, J., & Klein, A. (2019). *Fintech and the new financial landscape*. Washington, DC: Economic Perspectives.
Gonzalez, L., & Komarova, Y. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance*, *2*, 44–58.
Havrylchyk, O., Mariotto, C., Rahim, T., & Verdier, M. (2018). *What has driven the expansion of the peer-to-peer lending?* Université Paris, Boston University, and the European Commission Working Paper.
Herzenstein, M., Sonenshein, S., & Dholakia, U. M. (2011). Tell me a good story and I may lend you my money: The role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research*, *48*, S138–S149.
Hughes, J., Jagtiani, J., & Moon, C. (2019). *Consumer lending efficiency: Commercial banks versus a fintech lender*. Federal Reserve Bank of Philadelphia, Working Paper.
Jagtiani, J., Lambie-Hanson, L., & Lambie-Hanson, T. (2019). *Fintech mortgage lending: Impact on credit access and price*. Federal Reserve Bank of Philadelphia Working Paper.
Jagtiani, J., & Lemieux, C. (2018). Do fintech lenders penetrate areas that are underserved by traditional banks? *Journal of Economics and Business*, *100*, 43–54.
Jagtiani, J., Vermilyea, T., & Wall, L. (2018). The roles of big data and machine learning in bank supervision. *Banking Perspectives* (Vol. 6, pp. 44–51) Washington, DC: The Clearing House.
Kraussl, R., Kraussl, Z., Pollet, J., & Rinne, K. (2018). *Performance of market place lenders: Evidence from lending club payments data*. Center for Financial Studies Working Paper Series 598.
Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, *59*, 17–35.
Lu, Y., Gu, B., Ye, Q., & Sheng, Z. (2012). *Social influence and defaults in peer-to-peer lending networks*. Conference on Information Systems and Technology (CIST 2012) White Paper.
Mach, T., Carter, C., & Slattery, C. (2014). *Peer-to-peer lending to small businesses*. Federal Reserve Board of Governors, Finance and Economics Discussion Series 2014-10.
Michels, J. (2012). Do unverifiable disclosures matter? Evidence from peer-to-peer lending. *Accounting Review*, *87*, 1385–1413.
Morse, A. (2015). *Peer-to-peer crowdfunding: Information and the potential for disruption in consumer lending*. NBER Working Paper.
Pope, D., & Sydnor, J. (2011). What's in a picture? Evidence of discrimination from prosper.com. *Journal of Human Resources*, *46*, 53–92.
Ravina, E. (2012). *Love & loans: The effect of beauty and personal characteristics in credit markets*. Northwestern University Working Paper.
Serrano-Cinca, C., Gutierrez-Nieto, B., & Lopez-Palacios, L. (2015). Determinants of default in P2P lending. *PLoS ONE*, *10*(10), e0139427. https://doi.org/10.1371/journal.pone.0139427
Yencha, C., Nowak, A., & Ross, A. (2018). Small business borrowing and peer-to-peer lending: Evidence from lending club. *Contemporary Economic Policy*, *36*, 318–336.

How to cite this article: Jagtiani J, Lemieux C. The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*. 2019;48:1009–1029. https://doi.org/10.1111/fima.12295