Bank Misconduct and Online Lending*

Christoph Bertsch
Sveriges Riksbank

Isaiah Hull
Sveriges Riksbank

Yingjie Qi
Stockholm School of Economics

Xin Zhang
Sveriges Riksbank

First Version: November 2017
This Version: March 2020

Abstract

We introduce a high quality proxy for bank misconduct that is constructed from Consumer Financial Protection Bureau (CFPB) complaint data. We employ this proxy to measure the impact of bank misconduct on the expansion of online lending in the United States. Using nearly complete loan and application data from the online lending market, we demonstrate that bank misconduct is associated with a statistically and economically significant increase in online lending demand at the state and county levels. This result is robust to the inclusion of bank credit supply shocks and holds for both broader and more narrowly-defined bank misconduct measures. Furthermore, we show that this effect is strongest for lower rated borrowers and weakest in states with high levels of generalized trust.

Keywords: financial development, consumer loans, bank misconduct, FinTech

JEL Classification: A13, G00, G21, K00

*Christoph Bertsch, Isaiah Hull and Xin Zhang are at the Research Division of Sveriges Riksbank, 103 37 Stockholm, Sweden. Yingjie Qi is at the Department of Finance of the Stockholm School of Economics, 111 60 Stockholm, Sweden. We would like to thank Bo Becker, Giancarlo Corsetti, Julian Franks, Andreas Fuster, Mariassunta Giannetti, Luigi Guiso (discussant), Ulrike Malmendier, Christophe Pérignon (discussant), Raghavendra Rau, Vahid Saadi (discussant), Christophe Spaenjers (discussant) and Per Strömberg for comments, as well as seminar participants at the Stockholm Business School, Cleveland Fed, 2018 RIDGE December Forum in Montevideo, EFA 2018 in Warsaw, 4th IWH-FIN-FIRE Workshop in Halle, CEPR Third European Workshop on Household Finance in London, 5th EFI Workshop in Brussels, Graduate Institute Geneva, Sveriges Riksbank, and Uppsala University. All remaining errors are ours. The opinions expressed in this article are the sole responsibility of the authors and should not be interpreted as reflecting the views of Sveriges Riksbank.

1 Introduction

The Great Recession gave rise to the notion that "fraud has become a feature and not a bug" of the financial system (Zingales (2015), p.19). Such perceptions are not innocuous and may have important adverse implications for affected industries, as documented by Giannetti and Wang (2016) for corporate fraud and by Gurun et al. (2018) for fraud in the financial advisory industry. In this paper, we study misconduct in the retail banking market and link it to the expansion of online lending platforms which provide an alternative source of credit.

The online lending market has grown rapidly during the last decade, driven by new digital technologies, financial reasoning, shifts in customer preferences, and social factors.[1] By 2018, online lending had already reached around one-third of the U.S. market for unsecured personal loans (Balyuk and Davydenko 2019). Recent work suggests that online lending is a substitute for both personal loans originated by traditional banks and credit cards, and has substantive implications for small bank lending (Tang 2019; Cornaggia et al. 2018). Using nearly-complete loan and application histories from the two largest and oldest U.S. online lending platforms, LendingClub.com and Prosper.com, we study the substitution into online lending associated with bank misconduct. Specifically, we measure the impact of time and geographic variation in bank misconduct on the state and county-level expansion of online lending. Our hypothesis is that borrower perception of bank misconduct in the form of product mis-selling, unfair contractual terms, and opaque or unjustified fees, may drive existing customers away from traditional banks to online lending platforms. This is because online lenders can reduce borrower apprehension about the type of opportunistic behavior that is often associated with traditional banking by unbundling financial services and by relying on highly standardized and transparent personal loan contracts.

Controlling for local macroeconomic conditions, loan-borrower characteristics, and geographic and time fixed effects, we document a positive conditional correlation between bank misconduct complaints and the demand for online lending. Under conservative assumptions, we find that an increase in complaints by one per branch is associated with an increase

[1] For recent market studies on the broader trends reshaping retail banking see for instance the recent reports by McKinsey (`https://www.mckinsey.com/industries/financial-services/our-insights/rewriting-the-rules-in-retail-banking`) and Novantas Research (`https://www.consumerbankers.com/sites/default/files/2018-02-22%20Global%20Multi-Channel%20Survey.pdf`).

in yearly online lending demand by \$0.47M at the state level or \$23.4M at the national level.[2] This aligns closely with the estimated size of substitution estimates in the existing literature (Tang 2019; Cornaggia et al. 2018). Additionally, the impact is larger for borrowers with lower credit ratings, which indicates that bank misconduct may drive existing customers away from traditional banking to online lending platforms. These findings also hold in county level regressions with state-time fixed effects and county-level controls. We also document the robustness of our results to the inclusion of county level bank credit supply shocks. Moreover, we show that our results are robust to the use of lagged bank misconduct measures and to alternative specifications of the dependent variable that use loan originations, rather than measures of loan demand.

Beyond this, we also conduct a staggered quasi natural experiment study using a set of major bank scandals, which feature prominently in the CFPB complaint data and are identified using information from Factiva and the CFPB enforcement page. While the previous regressions can be viewed as an attempt to estimate all bank misconduct events without a difference-in-differences specification, the event study provides us with a robustness check derived from a quasi-natural experiment. Consistent with our previous results, we find that counties with a bank present that had a misconduct scandal experience increased online borrowing demand after the scandal.

Our novel measure for bank misconduct contains granular time and geographic variation and is constructed using Consumer Financial Protection Bureau (CFPB) complaint data. The CFPB consumer complaint data is often referenced in news articles and the large majority of complaints are credit related. Our complaint-based measure is indicative of perceived misconduct and tracks bank scandals closely. Roughly 25% of the consumer complaint narratives contain explicit references to acts of misconduct, criminality, cheating, theft, and deception. The remaining cases typically document alleged instances of borderline misconduct, where negligent customer service lead to material losses. In the empirical analysis, we use both a broader complaint-based measure and a more narrowly-defined misconduct measure that exclusively uses a filtered subset of comments identified by a machine learn-

[2]We assume that the share of offline lending that is substituted to the online market in response to misconduct is 5%, which is conservatively low, even for the sample period under consideration. Additionally, we use the mean amount of online lending per state in our sample (\$5.38M per month), which is considerably lower than the end-of-sample volume.

ing algorithm. The narrowly-defined bank misconduct complaints explain 80%-90% of the relationship between CFPB complaints and the online lending expansion. The remainder is potentially attributable to other non-price attributes of banks that may be reflected in CFPB complaints, such as service quality.

In addition to our main findings, we also show that the positive association between bank misconduct and the expansion of online lending is driven by the extensive margin. This is consistent with borrowers switching from traditional banks to online lending platforms, rather than choosing larger loans from online lenders. Moreover, in a separate exercise, we find evidence suggesting that the positive association between bank misconduct and the expansion of online lending is least pronounced in regions with a high level of generalized trust. It is plausible that a high level of generalized trust mutes borrowers' responsiveness to bank misconduct and also improves borrowers' access to informal credit.

There is a growing literature which studies the bank credit supply contraction after the Great Financial Crises, as a result of tightened financial regulatory rules (see, e.g., Buchak et al. (2018)) and more frequent regulatory supervision (see, e.g., Cortés et al. (2018)). Although we are mostly interested in online credit demand, it is reasonable to assume that the variation in bank credit supply at the regional level could play an important role in driving the online lending expansion. We use two measures of bank credit supply shocks from Buchak et al. (2018)–namely, the Office of Thrift Supervision's (OTS) closure and the county level bank capital requirements increase. Beyond documenting the aforementioned robustness of our findings to the inclusion of local bank credit supply shocks, we also find evidence suggestive of a complementary relationship between bank credit supply shocks and CFPB complaints. More specifically, the expansion of online lending is faster in counties where concentration of CFPB complaints coincides with bank credit supply shocks. Furthermore, in counties that suffer from the bank credit supply shocks, borrowers with high revolving credit balances appear to respond more to the CFPB complaints by borrowing from online lending platforms. Taken together, our findings suggest that banks might first cut off loans to higher risk borrowers with high revolving balances.

The finding that riskier borrower segments are more sensitive to bank misconduct is likely to be a reflection of a generally higher tendency for switching among riskier borrowers, as documented by Cornaggia et al. (2018), as well as their heightened exposure to bank misconduct. Higher risk borrowers are more likely to fail to make a payment, which may expose them to disadvantageous contractual terms. While our sample is unlikely to contain many underbanked households, this result is in line with the findings of the 2015 FDIC National Survey of Unbanked and Underbanked Households. In fact, trust in banks among underbanked households tends to be lower and can be a significant impediment for underserved borrowers seeking financial services. This survey-based evidence is reinforced by macro-level evidence, which points to a positive association between unemployment and distrust in institutions–and, in particular, banks (Stevenson and Wolfers 2011). This suggests that a low level of trust in financial institutions is likely to be particularly important for households who often lack stable employment and have low credit ratings. Taken together, riskier borrowers are likely to be most responsive to experiences of bank misconduct and news thereof.

Our analysis complements the extensive literature on financial misconduct with a focus on substitution away from the retail banking market to online platforms. It also advances the literature by contributing to our understanding of how FinTech has driven financial disintermediation and relates to the literature on the interaction between banks and shadow banks. Moreover, our paper also relates to the recent online lending platforms literature, which shows that online lending is often a substitute for bank lending (Tang 2019; Cornaggia et al. 2018).[3] Cornaggia et al. (2018) find evidence suggesting that around one quarter of the online lending volume directly substitutes personal loans by commercial banks. The overall magnitude of switching appears to be economically significant, as banks "lose 1.2% of their personal loan volume for each standard deviation increase in peer-to-peer lending" (p. 2, ibid). We also examine the personal loan substitution margin, but we consider the component associated with bank misconduct, which was a previously unexplored channel.

The rest of the paper proceeds as follows. Section 2 reviews the related literature and section 3 provides background information on the online lending market. Section 4 presents the theoretical framework and develops hypotheses. The data is described in section 5. Section 6 presents the empirical results. Thereafter, section 7 discusses identification challenges and measurement. Finally, section 8 concludes. All tables are located in the Appendix.

[3]Both lenders serve the same borrower population. When it comes to small loan sizes, online lending platforms may complement bank credit due to the lower cost of issuing loans.

2 Related literature

Our paper relates to three distinct strands of literature. First, it builds on the existing empirical research on financial misconduct and connects to the literatures on both generalized trust and trust in institutions. Second, our paper contributes to the literature on financial development and growth with a focus on financial innovation, as well as on the interaction between the incumbent banking system and the emerging non-bank competitors. And third, it complements the emerging literature on consumer credit and online lending that studies the micro- and macro-determinants of investor financing, as well as borrower behavior.

Financial Misconduct. There is a literature on financial misconduct inside and outside of the financial industry. Giannetti and Wang (2016) study corporate financial misconduct and use the Enron scandal as an exogenous shock to clients of the Arthur Anderson auditing firm. They find that states with more Arthur Andersen clients experience a larger decrease in stock market participation. Egan et al. (2017b) study misconduct in the financial advisory industry and Gurun et al. (2018) demonstrate the detrimental effect of the Madoff scandal on the investment advisors, providing suggestive evidence for the transmission of shock to investor trust originating from the Madoff scam in social networks. Regarding misconduct in the banking industry, Nguyen et al. (2016) study regulatory enforcement actions in the US and document how board monitoring and advising can reduce bank misconduct. Sakalauskait (2018) shows that misconduct appears to be positively associated with CEO bonuses during economic expansions for a sample of 30 global systemically important banks.

In light of the existing literature, the deterioration of trust in traditional banking after the Great Recession (e.g., Corsetti et al. (2010)) may have lowered the barriers to entry for new FinTech players, fueling disintermediation in some market segments targeted by online lending platforms (FSB 2017). Our measure of bank misconduct appears to be related to borrower distrust in banks and should be particularly relevant for the post-crisis sample period considered.[4] There is an extensive literature documenting the response to perceived unfair treatment in various areas (e.g., Xia et al. (2004)), which suggests that perceived

[4]In our context, distrust in banks can be defined as the subjective probability assigned to the chance of being cheated (Gambetta 2000).

unfair treatment by banks also influences borrowers' willingness to switch to online lending platforms. In related work, Guiso et al. (2013) find that lower trust in banks makes it more likely that borrowers strategically default on their mortgage debt.

To measure bank misconduct, we employ CFPB complaint data, which has also been used to study the quality of financial services. Different from the existing literature (Egan et al. 2017a and Begley and Purnanandam 2018), we do not study the quality of financial services, but focus on high frequency variation in CFPB complaints attributed to bank misconduct, which tend to be related to bank scandals and perceived fraud.

Financial innovation. Financial development has been demonstrated to facilitate economic growth (Rajan and Zingales 1998). New financial technologies can improve credit availability, for instance, by reducing asymmetric information, which enables lending to riskier borrower segments (Livshits et al. 2016). FinTech lenders can have a cost advantage, especially in screening riskier borrower segments based on hard information (Einav et al. 2013). This advantage may be related to the use of new data sources (Berg et al. 2018) or to the use of more sophisticated credit models (Fuster et al. 2018). Besides technological factors, societal factors are also important, as documented by Guiso et al. (2004) who highlight the role that trust and, more generally, social capital play in financial development.

Our paper is also related to the literature on the interaction between banks and shadow banks, including studies such as Buchak et al. (2018) and Fuster et al. (2019), which evaluate the impact of FinTech on mortgage lending. In contrast, we study the uncollateralized online consumer credit market and shed light on the role of financial misconduct in the context of financial disintermediation when new financial products become available to borrowers.

Online lending. Crowdfunding platforms have enjoyed rapid growth in recent years and have received increased attention.[5] We focus on peer-to-peer (P2P) online lending, which is the dominant worldwide form of crowdfunding (Rau 2017).[6] In the literature, the P2P online lending market has been used as a laboratory to study different micro aspects of

[5]For literature reviews on crowdfunding and crowdlending, see Belleflamme et al. (2015) and Morse (2015).
[6]For an in-depth overview refer to the Cambridge Center for Alternative Finance Benchmarking Reports.

lending, such as the role of informational frictions, using U.S. data from the Prosper.com[7] and LendingClub.com[8] consumer credit platforms. Within U.S. online lending, consumer credit is the largest market segment and tends to attract higher risk borrowers who seek small loans, a slice of the market that is underserved by traditional banks (De Roure et al. 2016). Many borrowers in this segment also use online platforms to increase their total borrowing capacity (Demyanyk et al. 2017). Online consumer credit is often uncollateralized and can be seen as a substitute for credit cards, other forms of consumer credit, or informal credit. Lending platforms operate with significantly lower costs than traditional banks and specialize in automated credit scoring. This can give FinTech lenders an advantage in screening higher risk borrowers (Einav et al. 2013), which may allow them to extend more generous loans to medium-risk borrowers and to reduce discriminatory biases (Bartlett et al. 2017). The expansion of online lending appears to be highest in regions where traditional banks are absent or under tougher regulation or capital constrained.[9]

There is also a link to the literature on informal credit, which is mostly studied mostly in the context of developing countries, since it plays a less important role when credit markets are more developed (Besley 1995; Robb and Robinson 2014). However, informal credit is still relevant in advanced economies (e.g. Zanin (2017)). While informal credit may be cheaper than formal credit, informal credit creates shadow costs that often make borrowers prefer formal finance (Lee and Persson 2016). In fact, mechanisms for third-party enforcement via crowdfunding platforms can help to mitigate social frictions and, thereby, lower the shadow costs associated with informal credit. This has been established for equity crowdfunding (Agrawal et al. 2013). An advantage of the online lending over informal lending is that it provides a standard loan contract with critical roles outsourced to the platform. In this way,

[7]Papers using data from Prosper.com study the role of soft information, such as the appearance of borrowers (Duarte et al. 2012; Pope and Sydnor 2011; Ravina 2012; Gonzales and Komarova Loureiro 2014), the importance of screening and of hard information in lending decisions (Iyer et al. 2015; Hildebrand et al. 2016; Balyuk 2018; Faia and Paiella 2017), the herding of lenders (Zhang and Liu 2012), the importance of geography-based informational frictions (Lin and Viswanathan 2016; Senney 2016), the auction pricing mechanism that existed prior to 2011 (Chen et al. 2014; Wei and Lin 2015), and the ability of marginal borrowers to substitute between financing sources (Butler et al. 2017). There are also papers studying macroeconomic developments (Crowe and Ramcharan 2013; Bertsch et al. 2017).

[8]There are papers using data from LendingClub.com to study adverse selection (Hertzberg et al. 2015), retail investor risk-aversion (Paravisini et al. 2016) and P2P as a substitute for bank lending (Tang 2019).

[9]See, e.g., Buchak et al. (2018) and Havrylchyk et al. (2017).

the personal relationship between borrower and lender is less likely to be negatively affected. Another advantage of borrowing from a platform is that it extends the group of lenders.[10]

3 The P2P online lending market

Peer-to-peer (P2P) online lending first emerged in the U.S. in 2005 in the form of crowdfunding. *Prosper.com* was the first U.S.-based platform, followed by the current market leader, *LendingClub.com*, which was founded in 2006. According to a Federal Reserve Bank of Cleveland study, U.S. P2P lending grew by an average of 84% per quarter between 2007 and 2014 (Demyanyk 2014). In the recent years, online lenders have started to transition from lending-based crowdfunding following the classical P2P model, which entails raising small amounts of funds from multiple lenders, to a mix of P2P lending and marketplace lending, which involves securing wholesale funding from institutional investors. In 2018 online lending had already reached around one-third of the U.S. market for unsecured personal loans (Balyuk and Davydenko 2019) and the accounting firm PricewaterhouseCoopers expects it to reach 10% of revolving US consumer debt by 2025.[11]

Our primary data set is from LendingClub, which operates the largest online platform for consumer credit in the US. As of June of 2019, LendingClub had more than 3 million customers, including both investors and borrowers, and had just reached the $50 billion milestone of total loan originations. In 2018, LendingClub reached a record annual loan volume of $10.9 billion, growing by 21% year-over-year. LendingClub's base of institutional investors has grown strongly since 2014; and retail investors have become a minority, representing less than 10% of the investment volume in 2018. Since LendingClub only recently expanded to the small business loan and auto refinancing segment, we focus exclusively on personal uncollateralized loans. Loan requests range from $500 to $40,000 with a maturity of 3 to 5 years. The median borrower has a loan size of $13,000, an interest rate of 13%, a yearly income of $65,000, an employment duration of 6 years, and a low proprietary credit rating. The personal loans issued by LendingClub are used for a variety of purposes, including debt

[10] Using Prosper.com data, Lin et al. (2013) show that the chances of successful funding on the platform increase if "real friends" bid on a listing.

[11] See market study by PricewaterhouseCoopers (2015).

consolidation, large durable good purchases, and unexpected expense financing.

The loan application process relies on digital technologies. After a prospective borrower submits an application, the platform collects self-reported and publicly available information, including the borrower's credit history. LendingClub uses a credit model to decide on the borrower's qualification for the loan, to assign a credit score, and to set a fixed interest rate and repayment schedule. There is a high degree of automation and qualified borrowers can expect to receive an offer within 24 hours. The platform provides a large set of loan-borrower characteristics to investors and divides the market into two distinct segments: fractional and whole. The fractional loan market is where a crowd of investors screens posted loans and funds individual borrowers in $25 increments. The whole loan market is where individual borrowers are matched with large investors who typically purchase shares of a pass-through security. While the former market is dominated by retail investors, the latter market is dominated by institutional investors. Individual loan applicants are allocated to the fractional or whole loan market by the platform and have no influence on it. We observe whether individual loan applicants successfully obtain funding and from which market segment. Provided borrowers accept the loan, the total funding volume (net of an origination fee) is disbursed. LendingClub offloads the risk to lenders and then services the loan throughout its duration, which includes monthly installment transfers from borrowers to lenders.

The majority of loan applicants in our dataset can be considered prime borrowers. A typical LendingClub borrower has a FICO score close to 700 and the minimum score to qualify is 640. The highest rated borrowers in online lending markets may have good access to traditional sources of credit from banks and credit cards. For them, online lending could be a substitute to traditional forms of lending. They may for instance be driven to online lending platforms for reasons related to convenience benefits from the digital user experience, such as the 24/7 availability in conjunction with the fast loan application process and disbursement. In contrast, the lowest rated borrowers are likely to be underserved by traditional banks. In our primary dataset, 60% of borrowers have a FICO score below 700. This compares with 46.3% in the U.S. population in 2012. The average FICO score in the U.S. population at the end of our sample period in 2016 was hovering around the threshold of 700 with the median FICO score tracking around 20 points higher. Online lending platforms make lending decisions based on algorithms that use hard information like the FICO score and detailed borrower characteristics as an input. Investors can closely monitor platforms' lending standards and loan performance in different market segments over time.[12]

LendingClub (as well as Prosper) generates fee income that is growing in transaction volume. Specifically, LendingClub's fee structure for fractional loans comprises: 1) an origination fee of 1-6%, paid by borrowers at loan disbursement; 2) a servicing fee of 1% on the payments transferred to lenders; and 3) a set of collection fees imposed for late payment and default. The servicing fee differs for the whole loan market.

Roughly two-thirds of borrowers in our sample self-declare that they want to use the online credit to consolidate their outstanding consumer debt with other lenders. Online lending platforms advertise substantial cost advantages to borrowers switching to their platforms by taking advantage of financial technology innovations. In a recent paper, Cornaggia et al. (2018) find evidence suggesting that Prosper loan rates are 164 basis points lower than the average rates on loans issued by banks after accounting for origination fees. Adams (2018) compares the interest rates and APRs for LendingClub and Prosper with the ones of credit cards for different credit score bins and finds that LendingClub APRs are lower than credit card rates for all credit scores and that the average spreads are sizable (100-450 basis points).

4 Theoretical framework and hypothesis development

With the market entry of online lending platforms, the dominant forms of consumer credit, traditional bank loans and credit card debt, have faced growing competition, especially in the segment for uncollateralized consumer credit for higher risk borrowers. Since a large part of our empirical analysis focuses on borrower application data, we propose a stylized theoretical framework for the uncollateralized consumer credit market that emphasizes the borrower perspective and funding choice. Thereby we gain insights into how bank misconduct could affect the development of the online lending market and complement the existing literature on financial misconduct in the corporate fraud, financial advisory, and banking context.

[12]Notably, there was a LendingClub scandal in April 2016 when the founding CEO Renaud LaPlanche resigned amidst the discovery of an improper sale of loans to an institutional client that deviated from investor instructions. However, this was unrelated to borrowers and is taken care of by our regression specification.

This section provides a verbal presentation of our stylized theoretical model, which we then use for hypothesis development. A formal discussion can be found in Appendix A, which includes a formal description of the model (Appendix A.1) and derivations (Appendix A.2). The key findings are summarized in Result 1.

We consider a static model that comprises a competitive traditional banking sector and online lending platforms. There are borrowers who are potentially interested in obtaining alternative funding to consolidate their existing debt. A fraction of the borrowers receive a shock that makes them want to "shop around," that is to become interested in seeking a loan of fixed size from another lender. These borrowers enter a competitive credit market, in which a large number of risk neutral traditional banks and online lending platforms compete by offering lending rates.

Borrowers then decide whether or not to consolidate their existing debt by taking a loan from another lender, and if so, which alternative funding source to select: a term loan from a traditional bank or a loan from an online lending platform. This borrower choice is motivated by the self-reporting of loan applicants in our data set. Around three quarters of applicants for online consumer credit state debt consolidation as the loan purpose. These applicants may, for instance, wish to obtain a cheaper term loan from a traditional bank or from an online lending platform to pay off revolving consumer credit stemming from credit cards, which are frequently issued by non-bank financial firms.

A key modeling assumption is that borrowers are heterogeneous in their affinity for using traditional banks vis-à-vis online lending platforms. The idea is that for borrowers, traditional banks are more familiar and emphasize in-person customer interaction in local branches, but online lending platforms tend to offer a potentially more rewarding customer experience, especially for technology-savvy borrowers. Therefore, online lending platforms may be more appealing for some borrowers, but not for others. We can think of the heterogeneity as approximately capturing factors that affect customer convenience, such as differences in age, internet usage, and education.

We focus on a credit market equilibrium where borrowers are segmented into three groups: no debt consolidation, borrowing from another bank, and borrowing from an online lending platform. When examining the role played by bank misconduct, we consider a loss of faith in traditional banks triggered by a spike of CFPB complaints. In the model, such a shock increases the expected utility cost of borrowers from being cheated or treated unfairly by banks through, for example, unfair contractual terms or deliberate missales. From the perspective of borrowers, the shock to bank misconduct makes online lending platforms more appealing.[13] As a result, more borrowers consider taking a loan from a platform. Hypothesis 1 follows and the formal derivation can be found in Appendix A.2, equation (10).

Hypothesis 1: bank misconduct

H0: Bank misconduct is positively correlated with the expansion of online lending demand at the regional level in the US.

Under the additional assumption that higher risk borrowers are more likely to shop around than lower risk borrowers, the effect of Hypothesis 1 is stronger for higher risk borrowers, meaning that the online lending demand ratio is increasing by more in the market segment of higher risk borrowers than in the market segment of lower risk borrowers. The additional assumption is plausible and could for instance be justified with the higher cost of debt faced by higher risk borrowers, which makes them more inclined to shop around. In fact, Cornaggia et al. (2018) provide evidence pointing in this direction by showing that the negative association between online lending and commercial bank lending is particularly strong in the segment of lower rated borrowers. If the overall switching intensity is strongest in this market segment, we may expect that switching intensity following bank misconduct complaints is also strongest in this borrower segment. Hypothesis 2 follows and the formal derivations can be found in Appendix A.2, equation (11).

Hypothesis 2: heterogeneous effects of bank misconduct (across borrowers)

H0: The prediction of Hypothesis 1 is more pronounced for higher risk borrowers.

[13] While dissatisfied customers may switch to rival banks or to online lending platforms, there are at least two reasons to expect that switching to online lending is an economically significant channel. First, there is a wide-spread perception that banks' business models are, at least in part, built on the fact that customers make mistakes. As a result, the misconduct of one institution may negatively affect the perceptions of customers about other traditional banks with similar business models, while it does not worsen the image of online lending platforms, who are the newcomers to the market. Second, online lending platforms successfully advertise transparent loan terms and want to give more control to their customers. This strategy is likely to resonate particularly well with borrowers who fear getting ripped off by banks.

In relation to Hypothesis 1, we might also expect that borrowers' faith in traditional banks is less negatively affected by bank misconduct in regions with high levels of generalized trust. Generalized trust is highly correlated with trust in institutions such as banks. In regions with a high level of generalized trust, the prevalent optimism about the trustworthiness of banks will not be as easily destroyed and replaced by fears about being cheated. In addition, the availability of informal credit is also likely to be higher.[14] Hypothesis 3 follows.

Hypothesis 3: heterogeneous effects of bank misconduct (across regions)

H0: The prediction of Hypothesis 1 is less pronounced for regions with a high level of generalized trust.

5 Data and descriptive statistics

Our primary dataset consists of a panel of 1.7 million loan-borrower observations from LendingClub. In addition to our primary dataset for LendingClub, we also have a secondary dataset with application data for Prosper, which spans the 2012-2017 period.[15] Prosper's platform design is similar to LendingClub's. Our application data for both LendingClub and Prosper comes from 424B3 filings, which we retrieved from the SEC's Edgar database. Our loan data for LendingClub comes from LendingClub's loan book, retrieved from their website. The SEC filings contain all application data, including loan and borrower characteristics. The loan book contains the set of loans originated, as well as loan characteristics, borrower characteristics, and repayment status updates. See Table B.1 for the list of variable definitions and Table B.2 for the associated summary statistics of our primary dataset.

We also collect information on variables related to the core hypotheses we test in the paper: 1) Consumer Financial Protection Bureau (CFPB) Consumer Complaint data; 2) a survey-based measure of generalized trust; 3) debt origination data from the Federal Reserve Bank of New York; 4) FDIC bank branch data; 5) state-level and county-level economic and

[14] We thank our discussant, Luigi Guiso, for suggesting we analyze this link.

[15] We use LendingClub as our primary subject of study because Prosper SEC filings are incomplete over the first four years of our sample.

demographic controls; 6) regional bank concentration data from the FDIC; and 7) internet access data from the Current Population Survey (CPS). The period we have available data on P2P lending starts in 2008 and continues until 2016. Data sources (3) and (4) are used primarily to normalize our variables of interest.

The CFPB Consumer Complaint data used for our bank misconduct measure contains the name of the bank, the time of the complaint, and the location of the customer. We compute the total number of complaints per county or state. We then normalize this by the number of bank branches in the county or state, which we take from the Federal Deposit Insurance Corporation's (FDIC) summary of deposit (SoD) database. The average state has 0.20 complaints per branch and month. The CFPB complaint database covers several categories of complaints. When constructing our baseline measure for shocks to bank misconduct, we include all complaints about traditional banking service categories (e.g. bank account services, credit reporting, debt collection, etc.).

When constructing our measure of misconduct, we include complaints against the traditional banking sector, but exclude complaints against FinTech lenders, such as LendingClub and Prosper.[16] The purpose of this is to avoid the upward bias that might arise from a higher online borrowing share being associated with more complaints about FinTech lenders. We also perform additional robustness exercises, where we filter complaints into "misconduct" and "non-misconduct" categories using a deep learning model. Typical complaints in our dataset involve dissatisfaction with services, the misselling of products, as well as fees and contractual terms that are regarded as unjustified or unfair. Over 70% of complaints in the CFPB database are directly related to borrowing. By volume, the top ten categories, which account for over 60% of complaints, are listed in Table 1. Among these categories, only the deposits and withdrawals category (2.7%) is unrelated to borrowing. All other categories are either directly related to borrowing or include complaints about borrowing.

The survey-based measure of trust was obtained from the General Social Survey (GSS), which is conducted biennially by the National Opinion Research Center (NORC) at the University of Chicago. This contains a measure of generalized trust, which is available for the 1973-2016 period, and includes both the region of residence and region of residence at

[16]During the period we study, there are only few complaints against LendingClub and Prosper (17.6 complaints per year against LendingClub and 8.4 complaints per year against Prosper).

Table 1: Top ten CFPB complaint categories

Complaint category	Fraction of complaints
1) Loan modification, collection, foreclosure	13.5%
2) Incorrect information on credit report	12.3%
3) Loan servicing, payments, and escrow amount	9.3%
4) Continued attempts to collect debt not owed	7.3%
5) Account opening, closing, or management	4.6%
6) Disclosure verification of debt	3.7%
7) Communication tactics	3.0%
8) Deposits and withdrawals	2.7%
9) Dealing with my lender or servicer	2.1%
10) Application, originator, mortgage broker	2.1%

age 16 for each respondent. We focus on generalized trust for the current region of residence and average over all observations in a given state-year. This yields a measure with a scale of 0 to 1, which varies between 0.18 and 0.62 in our sample.

6 Empirical results

We test our hypotheses listed in Section 4 and present the results. We also include robustness checks to further evaluate the strength of our empirical results. All tests conducted use our primary dataset, but are robust to the inclusion of secondary data from Prosper. Section 6.1 discusses the baseline results on the state level and highlights the key role played by the extensive margin. Thereafter, section 6.2 confirms key results on the county level, which allows us to include state-time fixed effects. Finally, Section 6.3 further corroborates the robustness of our findings by studying bank credit supply shocks and takes another step towards identifying the impact of bank misconduct, including a staggered quasi natural experiment study using a set of major bank scandals.

6.1 Baseline results

To test Hypotheses 1-3, we use the following empirical model as a baseline to relate the measure of regional online lending development to our measure of shocks to bank misconduct:

$$Y_{i,t} = \beta_1 \text{CFPB complaints}_{i,t} + \gamma X_{i,t} + (A_i + B_t) + \epsilon_{i,t}. \quad (1)$$

The dependent variable, Y_{it}, is the ratio of demand for P2P debt (millions) to total debt (tens of billions) in the location i in month t. We run the regression at both the state and the county level. The P2P market is small relative to the traditional banking sector in our sample. We normalize the P2P loan demand with the household debt balance of the P2P borrowers in the state, computed using the New York Fed consumer credit panel data. As the state consumer credit report is updated annually, the main variation in our dependent variable comes from the changes of P2P loan demand. This normalization ensures that the dependent variable is stationary and bounded by one. It also provides an intuitive interpretation of the results. Importantly, however, we will show that the effect of bank misconduct on the P2P market itself is economically significant, meaning that the results continue to hold qualitatively when using P2P debt demand as left-hand side variable.[17]

The bank misconduct variable is computed as the number of consumer complaints filed to the Consumer Financial Protection Bureau in the same month, normalized by the number of bank branches in the same location. Our underlying assumption is that borrowers who file complaints are likely to disseminate their dissatisfaction with a bank branch via their local networks or via the local news media. Moreover, CFPB complaints may indicate a wide-spread customer perception of unfair treatment by institutions in a particular region.[18]

Our specification includes state fixed effects, A_i, and time fixed effects, B_t. For the county-level regressions, we use state-time fixed effects, $A_{i,t}$. State or county level controls and borrower characteristics are denoted with $X_{i,t}$. We estimate the model with OLS and report the regression results in Sections 6.1.1 for the state level and in Section 6.2 for the county level. We also use loan application level regressions to measure the impact of CFPB

[17] The results are robust to other normalization used in the literature, such as the state/county population.

[18] In Section 7, we provide supporting evidence from major bank scandals. We describe the CFPB complaints measure in more detail in Section 5.

complaints on the extensive and intensive borrowing margins in Section 6.1.2. In addition, we investigate the relationship between the CFPB complaint measure and subgroup differences in applicant credit quality in Section 6.1.3. Thereafter, Section 6.1.4 explores how the relationship between bank misconduct and the expansion of online lending is affected by regional heterogeneity in the level of generalized trust.

6.1.1 Testing Hypothesis 1

Hypothesis 1 claims that an increase in the incidence of bank misconduct drives borrowers from banks to online lenders. We use consumer complaints filed to the U.S. Consumer Financial Protection Bureau as a proxy for borrowers who are exposed to bank misconduct. As explained in the data section, we construct a measure of regional consumer complaints and normalize by the number of bank branches. We capture geographic variation through the location of the complainant. Our underlying assumption is that borrowers in states that experience a high number of complaints per bank branch will tend to perceive misconduct as more prevalent. The reasoning is that the affected consumers may disseminate their dissatisfaction with a bank branch or service to their local social network. Furthermore, instances of egregious misconduct will be more likely to be covered by local news outlets in areas where consumers are most affected.

In Table B.5, we test Hypothesis 1 by regressing the ratio of online lending demand (m$) to total household debt (10b$) on the average number of consumer complaints per bank branch at the state level. In total, we have 2839 state-month observations spanning from 2012 to 2016. In the even-numbered columns, we use a specification that controls for the following variables at the state-level: population density, GDP, the unemployment rate, and the population size. We also control for borrower characteristics by including the state-level averages for the interest rates charged on loans, the gross income levels of borrowers, and the number of years employed. In columns 3 and 4, we include fixed effects to control for time-invariant state level characteristics. Columns 5 and 6 include both state and year fixed effects to control for time-invariant, state-level characteristics, and other common sources of time series variation that may have contributed to growth. In columns 7 and 8, we replace year fixed effects with year-month fixed effects, along with state fixed effects. This imposes more rigorous control on the time trend and the business cycle component to capture any shocks that occurred in a particular month.

Our estimate of the impact of bank misconduct complaints on the expansion of online lending is positive and statistically significant at the 1% level. Column 8 contains our most conservative regression specification. We find that increasing the number of complaints by one per bank branch is associated with a 0.9ppt increase in the online debt ratio. This amounts to an increase in online lending demand by at least \$0.47M at the state level or \$23.4M at the national level.[19] Furthermore, it aligns closely with the order of magnitude of substitution estimates in the existing literature (Tang 2019; Cornaggia et al. 2018) and suggests that bank misconduct has an economically significant impact on the growth of online lending, which we further corroborate in Section 6.2. We also find that these results are robust to the inclusion of additional county-level controls or the use of lagged CFPB complaint values. We include the following as additional controls: a measure for banking competition,[20] educational attainment, and internet penetration. The results are reported in Table C.1 in the Online Appendix. The coefficients of the CFPB complaint measure change only marginally and remain statistically significant. The county-level regressions with state-time fixed effects and county-level controls are shown in our robustness checks in Section 6.2. We are able to explore the variation of online lending and CFPB complaint at the county level with a stricter empirical specification. All the results are consistent qualitatively. Furthermore, all key results are robust to the replacement of the dependent variable with the level of online lending demand, the replacement of loan origination data with loan application data, and the inclusion of additional controls.

6.1.2 Extensive and intensive margin

Next, we analyze the impact of CFPB complaints on the extensive and intensive margins using loan application-level information. We measure the effect on the extensive margin by

[19] We assume that the share of offline lending that is substituted to the online market in response to misconduct is 5%, which is conservatively low, even for the sample period under consideration. Additionally, we use the mean amount of online lending per state in our sample (\$5.38M per month), which is considerably lower than the end-of-sample volume.

[20] We use the HHI index to measure bank competition. Recall that the construction of the CFPB complaints measure incorporates the number of bank branches.

regressing the fraction of P2P borrowers in the state's population on the average borrower pool quality variables and state level controls. The intensive margin is measured using a regression that links the loan application size to borrower-loan characteristics. From Table B.6 and B.7, we find that the results for the CFPB are mostly driven by the extensive margin, rather than the intensive margin. Equivalently, CFPB complaints at the state level motivate higher online lending growth, but do not affect borrowers' decisions about the loan amount.

These results align well with intuition. Namely, a positive shock to bank misconduct is associated with a higher number of borrowers switching to online lending. Conversely, there is no compelling reason to expect that the typical borrower who switches to request more funds. Thus, it seems plausible that state-level gains would be primarily driven by the extensive margin. This is also consistent with our theoretical framework, which makes predictions about the extensive margin and reflects the finding in Tang (2019) and Cornaggia et al. (2018) that online lending is often a substitute for bank lending.

6.1.3 Testing Hypothesis 2

To identify which borrower quality group is more sensitive to bank misconduct and other trust measures, we repeat the exercises from Table B.5, but for borrowers with high and low credit ratings. Forty percent of the borrowers have a FICO score above 700. We consider these borrowers to be of high quality and assign a value of one to them for the dummy variable "highrating." The rest of the borrowers, roughly corresponding to sixty percent of the sample, get a highrating dummy equal to zero.

We include the highrating dummy, the CFPB consumer complaint measure, and their interaction term in the regression presented in Table B.8. Following the same specification as in Table B.5, we find that the positive relationship between bank misconduct and P2P borrowing is driven by the low rated borrowers. In other words, lower quality borrowers are more likely to be picked up by online lending platforms ("bottom fishing") when they are exposed to misconduct or are unsatisfied with the service they get from traditional banking. This result is consistent with Cornaggia et al. (2018) who find that there is a higher tendency for switching among riskier borrowers. Again, we further corroborate the finding in county-level regressions in Section 6.2.

6.1.4 Testing Hypothesis 3

We next examine the heterogeneous effect of bank misconduct across regions. We repeat the exercise from Table B.5, but group the states according to their level of generalized trust. The cross-sectional generalized trust measure is computed as the average of the GSS's generalized trust question over the period 1973-2006, which pre-dates our sample.

The results are presented in Table B.9, where we include the CFPB measure, the generalized trust measure, the full set of controls from Table B.5, and year-month fixed effects. In columns 1, 2, and 3, we split the sample into terciles along the generalized trust dimension: bottom, middle and top. We find that the positive relationship between bank misconduct and P2P borrowing is concentrated in the middle tercile, for which the CFPB coefficient remains both positive and highly significant. Next, we split the sample in two generalized trust subgroups: low and high. The results are presented in columns 4 and 5, respectively. Now the positive relationship between bank misconduct and the P2P borrowing is concentrated in the low generalized trust subgroup.

Taken together, we find suggestive evidence that the positive association between bank misconduct and the P2P borrowing is less pronounced in regions with a high level of generalized trust. Hence, we cannot reject Hypothesis 3. It appears that borrowers in regions with an intermediate level of generalized trust are most responsive to bank misconduct. This result is plausible, since the occurrence of bank misconduct is likely to confirm the prevalence of pessimistic trust beliefs when the level of generalized trust is very low. Instead, a high level of generalized trust is likely to mute borrowers' responsiveness to bank misconduct. In addition, a high level of generalized also improves borrowers' access to informal credit.

6.2 County level results

Above, we discussed the empirical tests for the hypotheses presented in Section 4. We found support for the hypothesis that higher bank misconduct increases participation in the online lending market, especially by higher risk borrowers. In the following section, we extend these results to the county level, which allows us to include state-time fixed effects to control for unobserved characteristics at the state-time level.

In the CFPB database, the complainant's location can be identified using ZIP code information recorded in the data from December 2011. For the online lending applications filed by customers before October 2014, we can locate the borrower at the city-county level. We use a matched sample between December 2011 and October 2014 to repeat the baseline test in Table B.5, but at the county level. We supplement the data with county-level information from the Bureau of Economic Analysis, which contains total area, population, total income, and number of jobs in each county. In total, we have 53,658 county-month observations (see Table B.3 for summary statistics).

We regress the ratio of online lending demand (m$) to total household debt (10b$) on the average number of consumer complaints per bank branch at the county level. The results are presented in Table B.10. In the even-numbered columns, we use a specification that controls for county-level variables: population density, the logarithm of total income, the number of jobs, the population size, banking competition, educational attainment, and internet penetration. We also control for borrower characteristics by including the county-level averages for the debt-to-income ratio, the interest rates charged on loans, the gross income levels of borrowers, and the number of years employed. In columns 1 and 2, state-year fixed effects are included to absorb changes that occurred in a specific state and year. In columns 3 and 4, we include state-month fixed effects to control for potential seasonality within a year. Columns 5 and 6 include the strictest fixed effects permissible–namely, state-by-year-month fixed effects–to absorb any shocks that might have happened in a given state and month in our sample. This conservative econometric specification helps to control for observable and unobservable variation at the state level over time.

Our estimate of the impact of bank misconduct on the expansion of online lending is positive and statistically significant at the 1% level. Column 6 contains our most conservative regression specification. We find that increasing the number of complaints by 1 per bank branch at the county level is associated with a 2.8ppt increase in the online debt ratio. Stated differently, a one standard deviation increase of complaints per bank branch is associated with an 10.8% of standard deviation increase in the ratio of online debt to total debt. The impact is larger than for the state-level regression result because there is a high degree of variation in CFPB complaints per branch at the county level. This suggests that bank misconduct has an economically significant impact on the online lending sector development at the county level.

We also conduct the same exercise for borrower subgroups at county level as we did in Table B.8. The corresponding county level regression results are presented in Table B.11. We find consistently negative and significant coefficients for the high credit rating group dummy and its interaction with CFPB complaints per branch. This implies that the relationship between bank misconduct and the development of online lending is mainly driven by borrowers with lower credit scores. It confirms the conjecture that the online lending platforms are "bottom fishing"–that is, serving lower credit quality borrowers.

One of the concerns regarding our bank misconduct measure is that the complaints against bank misconduct could happen before switching is observed. We replace the CFPB complaint number per branch with 1 month lagged values in the aforementioned regressions. The results, shown in Tables C.2 in the Online Appendix, indicate that the qualitative results still hold with statistical significance. The magnitude, however, is smaller than the regression results using the contemporaneous CFPB complaints variable.[21]

6.3 Robustness

6.3.1 Bank credit supply shocks

A growing literature has studied the contraction of bank credit supply after the Financial crisis as a result of tightened financial regulatory rules (see, e.g., Buchak et al. (2018)) and more frequent regulatory supervision (see, e.g., Cortés et al. (2018) on the effect of stress testing on corporate credit). Although we are mostly interested in the online lending credit demand, it is reasonable to assume that the variation of credit supply at the regional level could play an important role in driving the online lending expansion.

We run a number of regressions to verify the relevance of bank credit supply in our study. In particular, we borrow a few measures of bank credit supply shocks from Buchak et al. (2018)–namely, the Office of Thrift Supervision's (OTS) closure and the county level bank

[21]Beyond the reported robustness checks, we also repeat our main exercises on trust in traditional banks, but in an instrumental variables regression setting. We use lags of the CFPB measure. The coefficient estimates are quantitatively similar and statistically significant at the 1% level.

capital increases. We expand the baseline regression as follows:

$$\begin{aligned} Y_{i,j,t} &= \beta_1 \text{CFPB complaints}_{i,j,t} + \beta_2 \text{Supply shocks}_{i,j} \\ &+ \beta_3 \text{CFPB complaints}_{i,j,t} \text{ x Supply shocks}_{i,j} + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t}, \end{aligned} \quad (2)$$

which includes *Supply shocks* that can be specified as a dummy or as a continuous treatment variable.

The closure of the OTS in 2011 is widely considered to be a tightening of regulation in the traditional banking industry.[22] The Dodd-Frank ACT terminated the OTS supervisory body and reallocated the financial institutions formally regulated by OTS to the OCC, FDIC, Federal Reserve, and CFPB. The literature documents that the OTS's closure generated a credit contraction in locations with many OTS banks. We explore the difference between counties with formally OTS supervised banks and counties not affected by the institutional change. To do so, we generate a dummy variable to indicate whether a county contains treated banks, and run the regression in separate samples. Columns 1 and 2 in Table B.12 report the baseline regression results for counties not treated by the OTS and for counties treated by the supply shock. Column 3 shows the expanded regression that contains the interaction between the supply shock dummy and the CFPB complaint measure. It appears that the positive relationship between CFPB complaints and the expansion of online lending does not disappear once we include the OTS's closure variable as the supply shock.

Next, we look at two variants. First, we use a continuous variable to measure the fraction of the bank deposit market affected by the shock using the Summary of Deposit dataset prior to the OTS's closure in 2011. Column 4 reports the regression results for the expanded specification with the share of affected banks at the county level and its interaction with CFPB. Second, we use the county level bank capital changes within our sample period as an index to see the potential credit contraction due to higher capital requirement after the Great Financial Crisis. Column 5 in the table shows that the counties with a higher capital requirement increase are likely to experience a faster expansion of online lending if there are more bank misconduct complaints filed to the CFPB. The interaction term is also significant, meaning that the counties with higher captial requirement increase will experience even faster

[22]See, e.g., Agarwal et al. (2014) for a detailed study of the quality of regulatory enforcement in the U.S.

P2P lending market growth if there are more consumer complaints.

We further the analysis by taking subgroups of borrowers who are more likely to be affected by the credit crunch. In our sample, almost all borrowers have a non-zero and usually high revolving credit balance. This rules out the possibility that a lack of access to credit is driving our results. We define *high revolvers* as individuals holding more than the median value of revolving debt in a given county. The regressions specification is:

$$\begin{aligned} Y_{i,j,t} &= \beta_1 \text{CFPB complaints}_{i,j,t} + \beta_2 \text{High revolvers}_{i,j} \\ &+ \beta_3 \text{CFPB complaints}_{i,j,t} \text{ x High revolvers}_{i,j} + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t}. \end{aligned} \quad (3)$$

High revolvers$_{i,j}$ is a dummy that indicates if borrowers in county j have revolving credit higher than the state i median borrower in the sample. We accumulate the loan sizes for the high/low revolvers and normalize them with the aggregate state outstanding credit as before. Moreover, we interact the CFPB complaints in the county with the dummy variable for high revolvers. Table B.13 presents the regression results. The columns 1–3 document the results using various fixed effects. We consistently find that high revolvers in a county with a higher number of CFPB complaints borrow more with the online lending platform.

Next, we introduce credit supply shocks. The last two columns contain results from counties not exposed to (column 4) and exposed to (column 5) the OTS closure credit supply shock. It appears that the CFPB complaint and the online lending activities are correlated in both treated and non-treated counties. However, the reactions are comparable for high revolvers and low revolvers. In the counties that suffer from credit supply shocks, high revolvers are the ones responding more to the CFPB complaints by borrowing online. Our findings also suggest that banks might cut off loans to risky borrowers with high revolving balances and are reminiscent of the finding that banks undergoing stress testing cut risky small business loans (Cortés et al. 2018).

6.3.2 New measure of bank misconduct

There are few existing papers using CFPB complaint data. One example is Begley and Purnanandam (2018), who study bank service quality. We extend the existing literature by using CFPB complaint data to construct a high frequency measure that serves as a high-quality proxy for bank misconduct. Typical CFPB complaints include fees and contractual terms that are regarded as unjustified or unfair, dissatisfaction with services, or misselling of products. Often, the severity of such issues was allegedly sufficient to generate financial losses for the complainant.

As explained in section 5, the large majority of CFPB complaints are related to credit. Overall, CFPB complaints can be seen as allegations of either misconduct or borderline misconduct in the form of negligence and deception. Despite containing some noise, our CFPB complaints-based measure is indicative of wide-spread customer perception of unfair treatment by banks that can be traced to specific regions with a high concentration of customer dissatisfaction.[23] Moreover, egregious misconduct is likely to be covered in local news outlets in areas with a high concentration of customer dissatisfaction; and major bank scandals are arguably key factors in the deterioration of trust. The relationship between bank scandals and shocks to CFPB complaints suggests that our measure is likely to serve well as a proxy for bank misconduct. Spikes in CFPB complaints about a particular institution tend to be associated with a critical media coverage that also sows distrust in the same institution. On the other hand, media reports regularly refer to the number of CFPB complaints to gauge the magnitude of bank scandals, which increases the publicity of CFPB complaints.

The 2013 Bank of America surge in customer complaints serves as an example of this phenomenon. Bank of America experienced a strong spike in CFPB complaints in January 2013, which took half a year to ebb off. Customer complaints centered around mortgage loan modifications and loan servicing, which was especially pronounced for loans in delinquency or foreclosure–a business area that has been contentious to BoA before, reflected in a 2.8bn penalty by the Office of the Comptroller of the Currency related to earlier mortgage abuses.[24] Notably, the surge in CFPB complaints acted as a catalyst for wide-spread news coverage of Bank of America's financial misconduct.[25]

Another good example is the Wells Fargo scandal in 2016, which involved the creation of fake accounts and severely damaged the bank's reputation among retail customers. Unlike

[23]Attenuation bias due the noisy CFPB complaint measure suggests that the effects we measure are a lower bound for the true effects.

[24]See `https://violationtracker.goodjobsfirst.org/violation-tracker/-bank-of-america-0`.

[25]See, e.g., `https://www.wsj.com/articles/SB10001424127887323361804578388791087793804`.

many other banks, Wells Fargo appeared to be a stable financial institution with fewer risky business practices during the years following the Great Financial Crisis. However, this positive perception was overturned when Wells Fargo pursued an overly aggressive approach to cross-sell products to their retail customers, providing strong incentives for their sales personnel to open up new accounts for customers in order to boost fee-based revenue. The twisted incentive scheme and the lack of supervision eventually lead to a large nationwide scandal. Over a five-year period, Wells Fargo employees opened 2 million accounts without customer authorization. The Wells Fargo scandal was widely reported in late 2016 when the CFPB fined Wells Fargo 100m$ for the illegal practice of secretly opening deposit and credit card accounts for customers.[26] In addition to this, the scandal also generated the largest surge of CFPB complaints for a single bank in the CFPB database.

6.3.3 Quasi-natural experiment: major bank scandals

The regressions in the main specification can be treated as attempts to estimate all bank misconduct events without a difference-in-differences (DID) specification. To further examine the robustness of the channel, we will look at a list of prominent bank scandal events as a quasi-natural experiment. The CFPB database covers a broad range of bank misconduct events and scandals, some of which are well-known and have attracted national attention. We identify major bank misconduct events using Factiva and CFPB enforcement actions. Table 2 provides a summary of bank scandals we investigate, including the institution involved, the broad category of the scandal, and the date at which it became public knowledge.

For each bank scandal, we included a county in the treatment group if it has a bank with a deposit market share exceeding 20% (the sample median) that is fined for misconduct by the CFPB or OCC, and control otherwise. The sample period is the same as the main analysis, namely from January 2011 to September 2014. We investigate whether such events increased online lending demand in the counties with branches of the bank where the scandals occurred relative to counties without branches of banks that had scandals. We use the following

[26] See CFPB Consent Order: `https://files.consumerfinance.gov/f/documents/092016_cfpb_WFBconsentorder.pdf`.

Table 2: Traditional banking scandals included in difference-in-differences exercise

Institution	Category	Date
Wells Fargo	Lending	October 2012
JP Morgan Chase	Credit cards	August 2013
Bank of America	Loan servicing and foreclosure processing	January 2013
Regions Bank	Accounting	January 2012
Capital One	Add-on products	July 2012
PNC Bank	Racial discrimination	December 2013

Notes: The table shows the list of traditional banking scandals used in the difference-in-differences exercises. The columns provide the institution involved in the scandal, the broad category of the scandal, and the date at which the scandal became public knowledge.

modified regression specification:

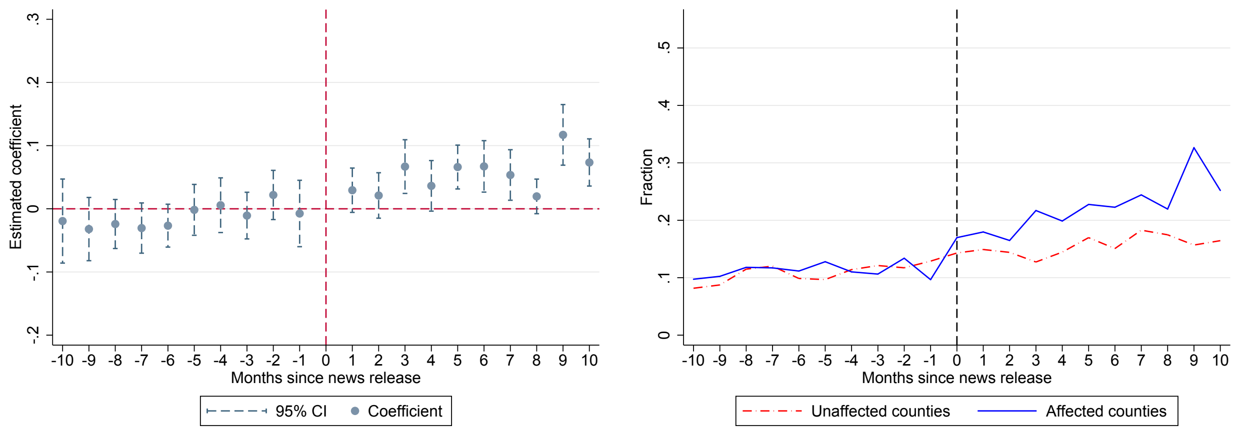
$$\begin{aligned} Y_{i,j,t} &= \beta_1 \text{Misconduct}_{i,j,t} + \beta_2 \text{Post}_t \\ &+ \beta_3 \text{Misconduct}_{i,j,t} \text{ x } \text{Post}_t + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t}, \end{aligned} \tag{4}$$

where $\text{Misconduct}_{i,j,t}$ is the treatment dummy variable that indicates whether the misconduct bank in county j of state i had a market share higher than 20% in the year before the misconduct disclosure. Post_t is an event dummy that is equal to 1 after the bank scandal is reported in the news media.[27] The left sub-figure of Figure 1 shows the relative difference in online lending demand from counties in the treatment and control group before and after the bank scandal hits. We see a significant difference in the trend after the scandal hits the counties in the treatment group. In particular, the counties in the treatment group see a faster increase in their online lending demand. In addition to this, we plot the estimated levels for the two groups separately in the right sub-figure of Figure 1 and find that they were similar prior to the scandals. Recent work by Kahn-Lang and Lang (2018) suggests that a DID specification will typically be more plausible if the treatment and control groups not only have parallel trends prior to the event, but also have similar levels.

This result is further confirmed in a set of DID regressions, which are reported in Tables

[27] If a county is affected by several scandals, we define the post period with the earliest bank misconduct scandal in the sample.

Figure 1: Difference-in-differences exercise



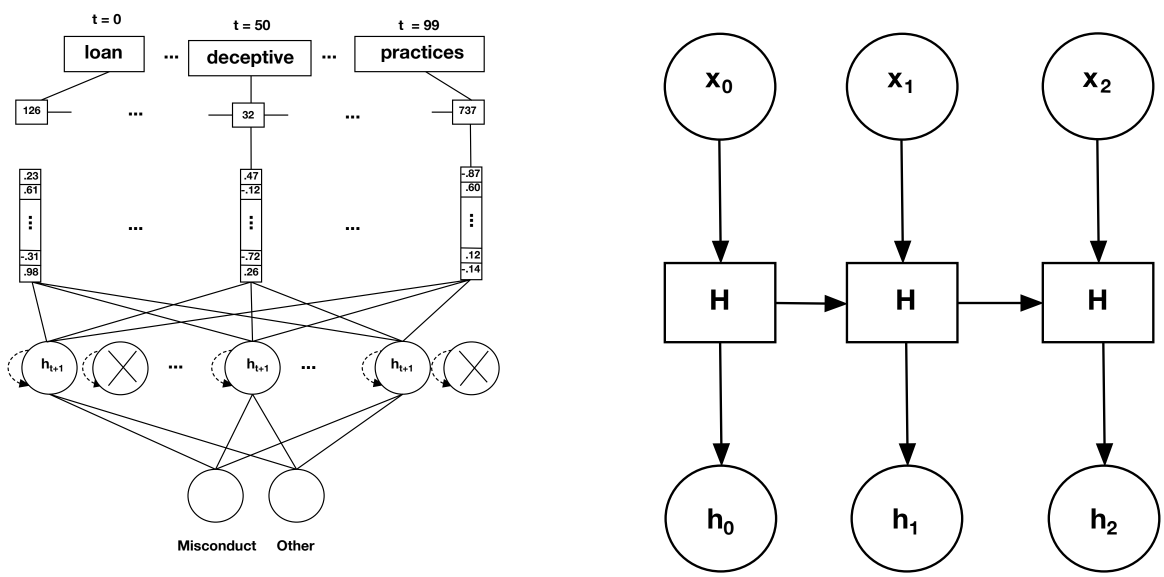
Notes: The sub-figure on the left shows the estimated difference in the P2P's share of total debt between treated and control counties. The horizontal axis shows the number of months that have elapsed since a major banking scandal occurred in the treatment counties. The vertical axis shows the difference in the P2P's share of total debt. We identify the date of bank scandals through the use of newspaper articles drawn from Factiva and CFPB enforcement actions. These events are also associated with sharp increases in the number of reported CFPB complaints. The subfigure on the right shows the levels of the estimates for each group separately.

B.14 and B.15. We compare the characteristics of counties in the treatment and control group in the year prior to the news release in Table B.4, and control for the variables that were significantly different in the two groups in the Difference-in-Differences regressions. We find that counties affected by bank scandals experience a larger online lending expansion after the publication of the news about the scandal. In fact, the DID interaction term coefficient, β_3, is significantly positive. After the scandal, counties where a bank that engaged in misconduct is present also experienced a stronger online borrowing demand. Table B.15 uses a continuous measure as the bank misconduct variable, namely the bank's deposit market share prior to the misconduct event, to measure how much the county is affected by a bank misconduct shock. It appears that the counties with larger deposit shares in banks with scandals experience increased online lending demand even more after misconduct events.

6.3.4 Narrowly-defined bank misconduct

In this section, we demonstrate the relevance of bank misconduct complaints in the CFPB data. The Consumer Financial Protection Bureau has been created under the mission to protect consumers in the financial sector. Given the potential risk of predatory lending and unfair practices in the financial service industry, the CFPB opens up the possibility for consumers to file against bank misconduct and fraudulent activities. We clean the complaint database and explore the textual information provided by consumers using machine learning. In particular, we first hand-classify a subsample of 10,000 complaints as alleging misconduct or not. We then train a recurrent neural network (RNN) to predict the class of each complaint in the subsample. Finally, we use the model to assign a probability of fraudulence to all complaint narratives in our sample.

Figure 2: Recurrent Neural Network Architecture



Notes: The figure on the left depicts our model's architecture. We use a recurrent neural network with word embeddings. The figure on the right depicts an unrolled recurrent neural network. The inputs, which are dense representations of word vectors, are processed in sequence, allowing for the memory of earlier inputs to affect the model's interpretation of words vectors that appear later in the sequence.

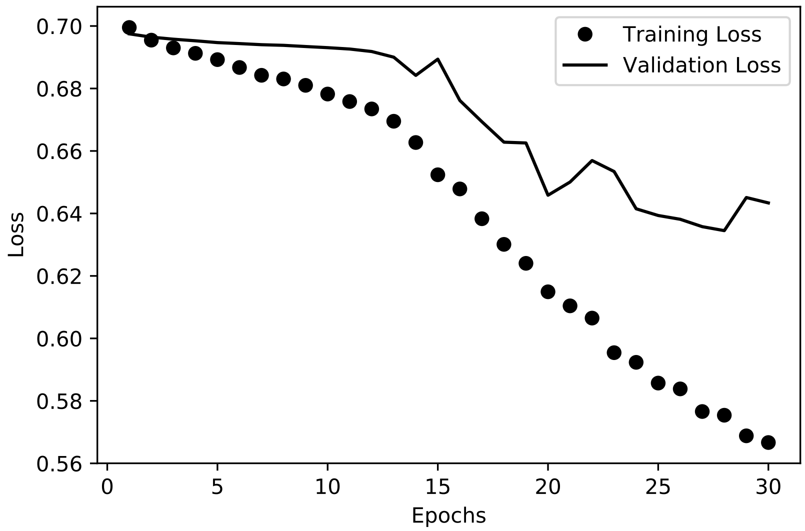
Our choice of RNN architecture is shown in Figure 2. Note that the input layer takes sequences of words, each of which is associated with a unique integer. We allow for the inclusion of 10,000 words and for the use of up to 100 words in a sequence. These inputs are then passed to a word embeddings layer, which constructs dense representations of words, each consisting of 64 elements. The embeddings layer allows the model to learn representations of word vectors that contain information about meaning and similarity. The layer accepts integer inputs and outputs 64-element word embeddings. We then apply dropout, which randomly drops all incoming edges for a random subset of the nodes in the following layer. This forces the model to develop robust representations of the input data that do not rely heavily on any particular feature, which prevents overfitting. After the dropout layer, we apply a layer that consists of 32 recurrent neural network cells. Figure 2 shows the "unrolled" version of an RNN cell. Here, x_0, ..., x_{99}, refer to 100 dense vector outputs from the embedding layer. These outputs are fed into a recurrent layer, which makes use of the sequence in which they appear. Note that the vector in the j^{th} position accepts an input from the vector in the $\{j+1\}^{th}$ position, allowing the model to retain memory of earlier words in the sequence. Finally, the model applies a fully-connected layer, which connects all of the RNN cell output nodes to a model output node.

We trained the model on 80% of the sample, retaining the remaining 20% for validation purposes.[28] We terminated the training process when the model failed to improve further in the validation sample. Performance in both the training and validation samples is given in Figure 3. Finally, we used the trained model to assign probabilities to complaints. In particular, we tried to identify the degree to which each category and subcategory of complaint tended to contain complaints that were about misconduct with a high probability. Based on this classification, we computed the number of CFPB complaints for each county or state at each month using the probabilities of the complaint categories classified to be bank misconduct related.

The regression is the same as the baseline in equation (1), but the CFPB measures are decomposed into two groups "bank misconduct related" and "others." We count the number of complaints that fall into each category and normalize them by the number of branches in the county. Table B.16 provides the results of the regression at the county level. Columns

[28]We use a binary crossentropy loss function and train the model with the adaptive moment (adam) optimizer with a learning rate of 10^{-4}. Word sequences that contain fewer than 100 elements are padded with zeros.

Figure 3: Training and Validation Losses



Notes: The figure above plots the training and validation sample losses by training epoch. We use a binary crossentropy loss function and train using the adaptive moment (adam) optimizer. The training and validation samples are divided into batches of 512 observations. Each epoch consists of a complete pass over all batches in the sample. The model parameters are never trained on the validation sample, which allows us to use the validation sample to terminate the training process before the model begins to overfit.

1 and 2 use the issue reported in the CFPB to identify the probability that the complaint is related to a potential bank misconduct. Columns 3 and 4 rely on the sub-category in the reported issue from the CFPB database. We show the regression results for the bank misconduct related complaints in columns 1 and 3, and for the other CFPB complaints in columns 2 and 4. It appears that both the bank misconduct and the non-misconduct related complaints can explain a substantial fraction of the online loan expansion at the county level. However, the bank misconduct related complaints contribute to 80%–90% of the relationship between CFPB complaints and the online lending expansion. The rest of the relationship may be attributed to other non-price attributes of banks, such as service quality.

6.3.5 Placebo test: the case of captive auto lenders

The online lending market has expanded to cover an increasing number of credit market segments served by traditional financial institutions. For our sample period, the auto loan market is one of the few exceptions with little to no P2P lending activity. We can identify a number of cases in the LendingClub loan application dataset where individual borrowers seek credit to finance a vehicle purchase; however, it remains uncommon over our sample period. This allows us to conduct a placebo test, where we investigate whether CFPB complaints against auto lenders explains the expansion of online lending. Since direct substitution into online lending should limited for automobile financing, we expect to find little to no effects from misconduct, especially for captive auto lenders, where the reputational effects of misconduct are arguably confined to a narrow segment of the credit market.[29]

To conduct this exercise, we clean the CFPB database and extract auto loan related products complaints. We run the county level baseline regression when exclusively auto loan related complaints are included. We find weaker, less significant results for the CFPB complaints in the baseline regression with the full set of controls and fixed effects. Additionally, the effect becomes insignificant when we look at captive automotive lenders, which do not lend outside of automobile loans.[30] Our finding suggests that the online lending expansion cannot be explained by consumer complaints against traditional lenders in a market segment where substitution into online lending is difficult. Furthermore, this finding is even stronger if we limit the sample to non-bank auto loan providers. The regression results are available upon request.

7 Discussion

Our paper adopts a number of identification strategies and includes a few robustness checks. The baseline regression and the quasi-natural experiment are employed to take a broad and narrow perspective to disentangle the effect of bank misconduct on the expansion of online lending in the United States. From both whole sample, and the event study, we find consistent results that the bank misconduct plays a role in explaining the fast expansion of online lending across different counties and states. We also take advantage of the rich

[29]We thank the referee for this suggestion.

[30]We use the CFPB complaints filed against the top ten captive auto lenders: BMW Financial Services, Toyota Financial Services, GM Financial, Honda Financial Services, Subaru Motors Finance, Mercedez-Benz Financial Services, VW Credit, Nissan Motor Acceptance Corporation, Lexus Financial Services, and Hyundai Motor Finance.

information in the CFPB complaint database to run a machine learning exercise and look at a few subcategories of the products. All these tests point to the potential channel of bank misconduct and frauds that drives borrowers to alternative credit providers. There might be a few remaining challenges which will be discussed below.

We encounter two common identification challenges. The first is the potential endogeneity of bank misconduct: rather than bank misconduct affecting online lending, online lending may instead affect bank misconduct. The second is omitted variable bias: a confounding variable may jointly determine both bank misconduct and online lending. Reverse causality can be trivially ruled out, since the participants in online lending are a small fraction of the population and cannot plausibly influence bank misconduct on the county or state level. Nevertheless, we explain the identification strategy and robustness checks in the remainder of the section.

An advantage of our approach over related research is that our measure of misconduct contains granular time and geographic variation. It is derived from complaints filed to the CFPB about the local office or branch of a particular financial institution. We assume that an increase in complaints in a given county-month indicates a deterioration in how consumers were treated or perceived themselves to be treated by financial institutions. A reduction in this metric would indicate that consumers perceive an increase in the likelihood of being cheated or treated unfairly.

Using a measure with county-time variation enables us to use state-month-year fixed effects. This eliminates the possibility that a variable with state-month-year variation could be jointly determining bank misconduct and the county-level online lending share. Our identification strategy relies on comparing two counties with similar characteristics, but different amounts of consumer complaints filed against traditional banks. In addition to this, we attempt to control for all plausible confounders with county-time variation, including average loan characteristics, income, employment, population size, population density, educational attainment, internet penetration, and competition in the banking sector. We also run these regressions at the state level and find qualitatively similar results. The difference-in-differences regressions using a quasi-natural experiment further confirm that our main results are related to bank misconduct news. The robustness of our results indicates that they are unlikely to be driven by local economic conditions differences or other region specific factors.

We further demonstrate that the relationship between online lending and bank misconduct comes primarily through the extensive margin, which is measured as the number of online borrowing applicants divided by the population size. Given the small number of applications per state and the correspondingly small number of complaints about online lending in the CFPB database, the number of applications cannot plausibly impact our bank misconduct measure in any substantial way. Furthermore, Section 6.3.1 presents additional robustness checks using supply-side shocks to bank credit at the county level and subgroups of borrowers that are likely to be differentially affected by a bank credit contraction. These findings jointly alleviate the concern of endogeneity for the measure of bank misconduct due to not controlling for bank supply shocks. They also suggest that alternative theories based on supply-side effects and credit rationing are unlikely to dominate the relationship between misconduct and online lending expansions suggested in our paper. An example of such a supply-side channel could be a "retailer effect," since traditional banks and other institutional investors became active on online lending platforms in 2014. While we do not have information about the investors on online lending platforms, traditional banks investing on online lending platforms arguably strive for geographic diversification. As a result, any excess bank funds that accumulate as a consequence of local bank misconduct are unlikely to be targeted to the same locations via bank investment in online lending platforms.

8 Conclusion

Since the Great Recession, perceptions of bank misconduct have become widespread. Such perceptions can be harmful to the traditional banking sector, as they have been shown to be in other sectors. The focus of this paper is on credit-related misconduct in the retail banking market and how it contributed to the expansion of U.S. online lending. Our findings suggest that bank misconduct may have played an economically significant role in facilitating the expansion of FinTech products, services, and instruments. Further analysis reveals that the effect is driven by the extensive margin, meaning that borrowers switch from traditional banks to online lending platforms, rather than choosing larger loans from online lenders.

We find that the positive association between bank misconduct and the expansion of online lending is robust to the inclusion of county-level bank credit supply shocks. Moreover, there appears to be an additional effect in the form of a complementary relationship between credit supply shocks and CFPB complaints: undercapitalized banks appear to cut off riskier borrowers first. And these borrowers are likely to have the highest levels of dissatisfaction and perceive fraudulent and unfair practices. We also find suggestive evidence that the positive association between bank misconduct and the expansion of online lending is less pronounced in regions with a high level of generalized trust, which is known to foster informal lending and to be highly correlated with trust in banks and institutions more generally.

One novel contribution of this paper is way in which we measure of bank misconduct using CFPB complaints. This yields a high frequency proxy for credit-related bank misconduct. We employ both broad and narrow definitions of bank misconduct. With the help of a machine learning algorithm that classifies the complaints as either being related to misconduct or a different topic, we show that a large part of the relationship between CFPB complaints and the online lending expansion can be plausibly attributed to bank misconduct. The rest of the relationship is attributable to other non-price attributes, such as bank service quality.

Our findings contribute to the understanding of the interaction between the incumbent banking system and its emerging non-bank competitors. The impact of bank misconduct goes beyond the traditional banking system. Bank regulation and consumer credit policy should be mindful of potential effects, especially when the shape of the modern credit market is vastly different from the traditional banking sector.

References

Adams, Robert M., "Do Marketplace Lending Platforms Offer Lower Rates to Consumers?," *FEDS Notes*, 2018.

Agarwal, Sumit, David Lucca, Amit Seru, and Francesco Trebbi, "Inconsistent Regulators: Evidence from Banking," *The Quarterly Journal of Economics*, 2014, *129* (2), 889–938.

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb, "Some Simple Econonomics of Crowdfunding," *NBER Working Paper 19133*, 2013.

Balyuk, Tetyana, "Financial Innovation and Borrowers: Evidence from Peer-to-Peer Lending," *Rotman School of Management Working Paper No. 2802220*, 2018.

— **and Sergei Davydenko**, "Reintermediation in Fintech: Evidence from Online Lending," *Mimeo*, 2019.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace, "Consumer Lending Discrimination in the FinTech Era," *mimeo*, 2017.

Begley, Taylor A. and Amiyatosh K. Purnanandam, "Color and Credit: Race, Regulation, and the Quality of Financial Services," *Mimeo*, 2018.

Belleflamme, Paul, Nessrine Omrani, and Martin Peitz, "The Economics of Crowdfunding Platforms," *Mimeo*, 2015.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri, "On the Rise of FinTechs - Credit Scoring Using Digital Footprints," *Mimeo*, 2018.

Bertsch, Christoph, Isaiah Hull, and Xin Zhang, "Monetary Normalizations and Consumer Credit: Evidence from Fed Liftoff and Online Lending," *Mimeo*, 2017.

Besley, Timothy, "Nonmarket Institutions for Credit and Risk Sharing in Low-Income Countries," *Journal of Economic Perspectives*, 1995, *9* (3), 115–127.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru, "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks," *Journal of Financial Economics*, 2018, *130* (3), 453–483.

Butler, Alexander W., Jess Cornaggia, and Umit G. Gurun, "Do Local Capital Market Conditions Affect Consumers' Borrowing Decisions?," *Management Science*, 2017, *63* (12), 4175–4187.

Chen, Ning, Arpita Ghosh, and Nicolas S. Lambert, "Auctions for social lending: A theoretical analysis," *Games and Economic Behavior*, 2014, *86*, 367–391.

Cornaggia, Jess, Brian Wolfe, and Woongsun Yoo, "Crowding Out Banks: Credit Substitution by Peer-to-Peer Lending," *Mimeo*, 2018.

Corsetti, Giancarlo, Michael P. Devereux, Luigi Guiso, John Hassler, Gilles Saint-Paul, Hans-Werner Sinn, Jan-Egbert Sturm, and Xavier Vives, "A Trust-Driven Financial Crisis," *EEAG Report on the European Economy*, 2010.

Cortés, Kristle, Yuliya Demyanyk, Lei Li, Elena Loutskina, and Philip E. Strahan, "Stress Tests and Small Business Lending," *NBER Working Paper 24365*, 2018.

Crowe, Christopher and Rodney Ramcharan, "House Prices and Household Credit Access: Evidence from an Internet Bank," *Journal of Money, Credit & Banking*, 2013, *45* (6), 1085–1115.

De Roure, Calebe, Loriana Pelizzon, and Paolo Tasca, "How Does P2P Lending Fit into the Consumer Credit Market?," *Deutsche Bundesbank Discussion Paper No. 30/2016*, 2016.

Demyanyk, Yuliya, "Peer-to-Peer Lending Is Poised to Grow," *Federal Reserve Bank of Cleveland*, 2014.

__ , Elena Loutskina, and Daniel Kolliner, "The Taste of Peer-to-Peer Loans," *Mimeo*, 2017.

Duarte, Jefferson, Stephan Siegel, and Lance Young, "Trust and Credit: The Role of Appearance in Peer-to-peer Lending," *Review of Financial Studies*, 2012, *25* (8), 2455–2484.

Egan, Mark, Ali Hortaçsu, and Gregor Matvos, "Deposit Competition and Financial Fragility: Evidence from the US Banking Sector," *American Economic Review*, 2017, *107* (1), 169–216.

__ , Gregor Matvos, and Amit Seru, "The Market for Financial Adviser Misconduct," *Mimeo*, 2017.

Einav, Liran, Mark Jenkins, and Jonathan Levin, "The Impact of Information Technology on Consumer Lending," *RAND Journal of Economics*, 2013, *44* (2), 249–274.

Faia, Ester and Monica Paiella, "P2P Lending: Information Externalities, Social Networks and Loans' Substitution," *Mimeo*, 2017.

FDIC, "Federal Deposit Insurance Corporation National Survey of Unbanked and Underbanked Households," 2015.

FSB, *FinTech Credit Market Structure, Business Models and Financial Stability Implications* 2017.

Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery, "The Role of Technology in Mortgage Lending," *Review of Financial Studies*, 2019, *32* (5), 1854–1899.

__ , Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther, "Predictably Unequal? The Effects of Machine Learning on Credit Markets," *Mimeo*, 2018.

Gambetta, Diego, "Can We Trust Trust?," in Diego Gambetta, ed., *Trust: Making and Breaking Cooperative Relations*, 2000, pp. 213–237.

Giannetti, Mariassunta and Tracy Y. Wang, "Corporate Scandals and Household Stock Market Participation," *Journal of Finance*, 2016, *71* (6), 2591–2636.

Gonzales, Laura and Yuliya Komarova Loureiro, "When Can a Photo Increase Credit?: The Impact of Lender and Borrower Profiles on Online P2P Loans," *Journal of Experimental and Behavorial Finance*, 2014, *2*, 44–58.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "The Role of Social Capital in Financial Development," *American Economic Review*, 2004, *94* (3), 526–556.

—, —, and —, "The Determinants of Attitudes towards Strategic Default on Mortgages," *Journal of Finance*, 2013, *68* (4), 1473–1515.

Gurun, Umit G, Noah Stoffman, and Scott E Yonker, "Trust Busting: The Effect of Fraud on Investor Behavior," *The Review of Financial Studies*, 2018, *31* (4), 1341–1376.

Havrylchyk, Olena, Carlotta Mariotto, Talal Rahim, and Marianne Verdier, "What Drives the Expansion of the Peer-to-Peer Lending?," *Mimeo*, 2017.

Hertzberg, Andrew, Andres Liberman, and Daniel Paravisini, "Adverse Selection on Maturity: Evidence From on-Line Consumer Credit," *Mimeo*, 2015.

Hildebrand, Thomas, Manju Puri, and Jörg Rocholl, "Adverse Incentives in Crowdfunding," *Management Science*, 2016, *63* (3), 587 – 608.

Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo F.P. Luttmer, and Kelly Shue, "Screening Peers Softly: Inferring the Quality of Small Borrowers," *Management Science*, 2015, *62* (6), 1554 – 1577.

Kahn-Lang, Ariella and Kevin Lang, "The Promise and Pitfalls of Differences-in-Differences," *NBER Working Paper 24857*, 2018.

Lee, Samuel and Petra Persson, "Financing from Family and Friends," *The Review of Financial Studies*, 2016, *29* (9), 2341–2386.

Lin, Mingfeng and Siva Viswanathan, "Home Bias in Online Investments: An Empirical Study of an Online Crowdfunding Market," *Management Science*, 2016, *62* (5), 1393 – 1414.

—, Nagpurnanand R. Prabhala, and Siva Viswanathan, "Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to- Peer Lending," *Management Science*, 2013, *59* (January 2013), 17–35.

Livshits, Igor, James C. Mac Gee, and Michèle Tertilt, "The Democratization of Credit and the Rise in Consumer Bankruptcies," *Review of Economic Studies*, 2016, *83* (4), 1673–1710.

Morse, Adair, "Peer-to-Peer Crowdfunding: Information and the Potential for Disruption in Consumer Lending," *NBER Working Paper No. 20899*, 2015, pp. 1–30.

Nguyen, Duc Duy, Jens Hagendorff, and Arman Eshraghi, "Can Bank Boards Prevent Misconduct?," *Review of Finance*, 2016, *20* (1), 1–36.

Paravisini, Daniel, Veronica Rappoport, and Enrichetta Ravina, "Risk Aversion and Wealth: Evidence from Person-to-Person Lending Portfolios," *Management Science*, 2016, *63* (2), 279 – 297.

Parlour, Christine A., Uday Rajan, and Haoxiang Zhu, "FinTech Disruption, Payment Data, and Bank Information," *Mimeo*, 2019.

Pope, Devin G. and Justin R. Sydnor, "What's in a Picture? Evidence of Discrimination from Prosper.com," *The Journal of Human Resources*, 2011, *46* (1), 53–92.

PricewaterhouseCoopers, "Peer Pressure: How Peer-to-Peer Lending Platforms are Transforming the Consumer Lending Industry," 2015, (February).

Rajan, Raghuram G. and Luigi Zingales, "Financial Dependence and Growth," *The American Economic Review*, 1998, *88* (3), 559–586.

Rau, Raghavendra, "Law, Trust, and the Development of Crowdfunding," *Mimeo*, 2017.

Ravina, Enrichetta, "Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets," *Mimeo*, 2012, pp. 1–79.

Robb, Alicia M. and David T. Robinson, "The Capital Structure Decisions of New Firms," *Review of Financial Studies*, 2014, *27* (1), 153–179.

Sakalauskait, Ieva, "Bank Risk-Taking and Misconduct," *Bank of Lithuania Working Paper Series No 53 / 2018*, 2018.

Senney, Garrett T., "The Geography of Bidder Behavior in Peer-to-Peer Lending Markets," *Mimeo*, 2016.

Stevenson, Betsey and Justin Wolfers, "Trust in Public Institutions over the Business Cycle," *The American Economic Review*, 2011, *101* (3), 281–287.

Tang, Huan, "Peer-to-Peer Lenders Versus Banks: Substitutes or Complements?," *Review of Financial Studies*, 2019, *32* (5), 1900–1938.

Wei, Zaiyan and Mingfeng Lin, "Market Mechanisms in Online Crowdfunding," *Mimeo*, 2015.

Xia, Lan, Kent B. Monroe, and Jennifer L. Cox, "The Price Is Unfair! A Conceptual Framework of Price Fairness Perceptions," *Journal of Marketing*, 2004, *68* (4), 1–15.

Zanin, Luca, "Determinants of the Conditional Probability that a Household has Informal Loans given Liquidity Constraints Regarding Access to Credit Banking Channels," *Journal of Behavioral and Experimental Finance*, 2017, *13*, 16–24.

Zhang, J. and P. Liu, "Rational Herding in Microloan Markets," *Management Science*, 2012, *58* (5), 892–912.

Zingales, Luigi, "Presidential address: Does Finance Benefit Society?," *The Journal of Finance*, 2015, *70* (4), 1327–1363.

A Formal model

A.1 Setup and results

Consider a stylized simultaneous-move, one-shot game with a continuum of risk-neutral borrowers of mass $m > 0$, indexed by $i \in [0, m]$, who are potentially interested in consolidating their revolving consumer debt (credit card or bank debt) by obtaining a long-term, uncollateralized loan of fixed size, $L = 1$, from an another lender. There are two *observable* types of borrowers indexed by the superscript $j = \ell, h$: Lower risk and higher risk borrowers. A fraction $0 < \alpha < 1$ of borrowers is of the lower risk type $j = \ell$ and a fraction $(1 - \alpha)$ of borrowers is of the higher risk type $j = h$. The respective *exogenous* repayment probabilities of lower and higher risk borrowers, p_ℓ and p_h, are common knowledge with $0 < p_h < p_\ell \leq 1$. Since we are in our baseline regression using a FICO score cutoff of 700, we can consider the case with approximately equally sized borrower sizes, i.e. we henceforth set $\alpha = 1/2$.

Suppose not every borrower wants to consolidate their revolving credit card or bank debt by obtaining a loan from another lender. Instead, only borrowers who receive a shock that makes them want to "shop around" and become interested in seeking a credit offer from another lender. The probability to receive this shock is exogenous and depends on the borrower type: $q_j \in (0, 1)$. We assume $q_h > q_\ell$, meaning that for higher risk borrowers there is a higher probability that they are interested in obtaining a loan from another lender.

Borrowers hit by the shock enter a competitive uncollateralized credit market. In the credit market borrowers can choose from a large number of risk neutral traditional banks and online lending platforms that compete by setting lending rates. Similar to Parlour et al. (2019) we assume that borrowers are heterogeneous in their affinity for using a traditional bank (which applies for both, an alternative bank lender and for their existing lender) vis-à-vis an online lending platform. This heterogeneity is meant to capture in a crude way factors that affect customer convenience, such as differences in age, internet usage and education. More specifically, we consider a simple linear model where the utility convenience benefit of borrower i from obtaining an alternative credit from an online lending platform is given by $u(i) = b(i - c)$, with $b > 0$ and $c \in (0, bm)$, where $u(i)$ is independent of the borrower type j. Notably, $u(i)$ is negative for the segment of borrowers with a small i who have an affinity for traditional banks and it is positive for the segment of borrowers with a large i who have an affinity for online lending.

Let $r_j > 1$ be the uniform reservation interest rate of borrowers, which captures the borrowing cost incurred by borrowers when deciding not to consolidate their debt. We assume that borrowers are wary of being negatively affected by financial misconduct, since they believe they may be cheated or treated unfairly with a certain probability (i.e. borrowers distrust lenders). For simplicity, we assume that the expected utility cost of being cheated by banks is uniform across borrowers and given by $\tau \geq 0$ and applies equally to the borrowers' exiting creditors and to other banks. The utility cost is uniform across borrowers and captures the disutility from facing unfair fees or contractual terms in a stylized way.[31]

Let $f_{B,h} > 1$ and $f_{B,\ell} > 1$ be the constant marginal cost to issue one loan for banks to borrowers of type h and type ℓ, respectively, where $f_{B,h} > f_{B,\ell}$. Moreover, let $f_O > 1$ be the cost to issue one loan for online lenders to borrowers of either type. While the cost of issuing a loan to a higher risk borrower is higher for banks, the cost of issuing loans for online lenders is invariant in the borrower type. This assumption reflects the reliance of online lenders on fully automated credit risk assessment. Moreover, it allows us to generate a relative cost advantage of screening based on hard information especially in riskier borrower segments (Einav et al. 2013) if we assume that $f_O \in (f_{B,\ell}, f_{B,h})$. In this plausible scenario online lending platforms naturally target the higher risk borrower segments first.

The timing of the game is as follows. First, the shock to "shop around" realizes and the affected borrowers enter the credit market. Second, lenders simultaneously offer lending rates to borrowers who then decide whether or not to consolidate their debt and which offer for alternative funding to accept. We next describe the credit market equilibrium and summarize the key comparative statics.

Result 1 *An individual borrower i of observable risk type $j = \ell, h$ who receives the shock and shops around optimally decides to take a loan from another lender if and only if:*

$$r_j + \tau > \min\{f_{B,j}/p_j + \tau, f_O/p_j - u(i)\} \equiv \xi_j(i). \tag{5}$$

[31] Evidently, borrowers may also be wary of being cheated by online lending and τ can also be understood as the differential expected utility cost of being cheated by banks vis-à-vis online lending platforms.

She prefers traditional banks over online lending platforms if:

$$(f_{B,j} - f_O)/p_j + \tau < -u(i). \tag{6}$$

Instead, if inequality (6) is violated, she chooses to take a loan from an online lender. If:

$$r_\ell > f_{B,\ell}/p_\ell \tag{7}$$

$$r_h > f_O/p_h - \tau - b(m - c) \tag{8}$$

$$-b(m - c) < (f_{B,h} - f_O)/p_h + \tau < bc. \tag{9}$$

There exists a credit market equilibrium characterized by a segmentation into borrowers who shop around and (a) do not consolidate their debt, (b) seek a loan from another bank, and (c) seek a loan from online lending platforms. The corresponding lending volumes are derived in Appendix A.2. The online lending demand ratio is increasing in borrowers' utility cost of being cheated, τ, and the increase is stronger for higher credit risk borrowers of type $j = h$.

The results guide the hypothesis development and the derivations can be found in Appendix A.2. As expected, a higher level of bank misconduct is, at the margin, associated with an increase in the online lending demand and, hence, volume (Hypothesis 1). This effect is larger for higher risk borrowers since higher risk borrowers are more likely to shop around and, therefore, potentially benefit from seeking a loan from another lender (Hypothesis 2).[32]

Intuitively, the conditions in (7), (8), and (9) help us to focus on the most relevant case. First, the condition in inequality (7) together with the second inequality in (9) assures that alternative bank lenders can obtain a positive market share in equilibrium since they are *(a)* not dominated by the existing lender in the market for low risk borrowers (albeit they may be dominated by the existing lender in the market for high risk borrowers) and *(b)* not driven out of the market by online lending platforms. To see this, notice that the second inequality in (9) implies that $(f_{B,\ell} - f_O)/p_\ell + \tau < bc$, because of the assumption that $f_O \in (f_{B,\ell}, f_{B,h})$.

Second, the condition in inequality (8) together with the first inequality in (9) assures

[32]Notice that all formal results are stated as "weak" increases or decreases and not "strict" increases or decreases. We have strict increases or decreases for cases when the solution is interior. Since there are a many cases to consider, we relegate the details to Appendix A.2.

that online lending platforms can obtain a positive market share since they are *(a)* not driven out of the market by the existing lender and *(b)* not dominated by other banks in the market for high risk borrowers. Note that online lending platforms may still be driven out of the market for lower risk borrowers.

Taken together, the conditions ensure that there exists a credit market equilibrium where a positive mass of borrowers stays with their existing lenders, while a positive mass of borrowers switch to each of the alternatives: banks and online lenders.

A.2 Derivations

The volumes for no debt consolidation, bank lending and online lending ($\mathcal{N}$, $\mathcal{B}$ and $\mathcal{O}$) are:

$$
\begin{aligned}
\mathcal{N} = \mathcal{N}_\ell + \mathcal{N}_h &\equiv \frac{1}{2}\left((1-q_\ell)m + q_\ell \int_{r_\ell+\tau}^{\xi(i)} di\right) + \frac{1}{2}\left((1-q_h)m + q_h \int_{r_h+\tau}^{\xi(i)} di\right) \\
\mathcal{B} = \mathcal{B}_\ell + \mathcal{B}_h &\equiv \frac{1}{2}\, q_\ell \int_{f_{B,\ell}/p_\ell}^{r_\ell} \Phi_{NB,\ell}\Phi_{B,\ell}(i)di + \frac{1}{2}\, q_h \int_{f_{B,h}/p_h}^{r_h} \Phi_{NB,h}\Phi_{B,h}(i)di \\
\mathcal{O} = \mathcal{O}_\ell + \mathcal{O}_h &\equiv \frac{1}{2}\, q_\ell \int_{f_O/p_\ell - u(m)}^{r_\ell+\tau} \Phi_{NO,\ell}(i)\Phi_{O,\ell}(i)di + \frac{1}{2}\, q_h \int_{f_O/p_h - u(m)}^{r_h+\tau} \Phi_{NO,h}(i)\Phi_{O,h}(i)di
\end{aligned}
$$

where the integration bounds relate to inequalities (5) and (6), with:

$$
\begin{aligned}
\Phi_{NB,j} &\equiv \begin{cases} 1 & if\ f_{B,j}/p_j < r_j \\ 0 & otherwise \end{cases} \qquad , \forall j = \ell, h \\
\Phi_{NO,j}(i) &\equiv \begin{cases} 1 & if\ f_O/p_j - u(i) < r_j + \tau \\ 0 & otherwise \end{cases} \qquad , \forall j = \ell, h \\
\Phi_{B,j}(i) &\equiv \begin{cases} 1 & if\ inequality\ (6)\ holds \\ 0 & otherwise \end{cases} \qquad \Phi_{O,j}(i) \equiv 1 - \Phi_{B,j}(i), \forall j = \ell, h.
\end{aligned}
$$

Recall that borrower types are observed and that both traditional banks and online lenders behave competitively by setting rates equal to their risk adjusted funding costs, which are given by $f_{B,j}/p_j$ and f_O/p_j, respectively. If inequality (5) is violated for all borrowers then none is willing to seek alternative funding. We have that all borrowers of type $j = \ell, h$ are unwilling to borrow from another bank if $\Phi_{NB,j} = 0$ and they are unwilling to borrow from an online lending platform if $\Phi_{NO,j}(i) = 0, \forall i$. A set of necessary and sufficient conditions for $\mathcal{N} < m$ and $\mathcal{O}, \mathcal{B} > 0$ is given by inequalities (7), (8) and (9).

The marginal borrower $\widehat{i}_{B,j}$ who is indifferent between seeking a loan from an alternative bank or from an online lending platform can be derived as:

$$\widehat{i}_{B,j} = \frac{(f_O - f_{B,j})/p_j - \tau}{b} + c$$

and the marginal borrower $\widehat{i}_{N,j}$ who is indifferent between the exiting lender and a loan from an online lending platform can be derived as:

$$\widehat{i}_{N,j} = \frac{f_O/p_j - r_j - \tau}{b} + c.$$

Provided there exists an interior solution, the mass of borrowers preferring a loan from online lending platforms strictly increases in bank misconduct since $\frac{d\widehat{i}_{B,j}}{d\tau} < 0$ and $\frac{d\widehat{i}_{N,j}}{d\tau} < 0$. In the market segment of higher risk borrowers interiority is guaranteed either if online lending platforms compete with the existing lender (i.e. $r_h < f_{B,h}/p_h$ and $r_h < f_O/p_h - \tau + bc$) or if online lending platforms compete with other banks (i.e. $r_h > f_{B,h}/p_h$). Instead, in the market segment of lower risk borrowers interiority is guaranteed if online lending platforms compete with other banks (i.e. $-b(m - c) < (f_{B,\ell} - f_O)/p_\ell + \tau$).

Next, we define the online lending demand ratio as $\mathcal{O}/m$. Provided inequalities (7), (8) and (9) hold and there exists an interior solution, we can derive the following results:

$$\frac{d\mathcal{O}/m}{d\tau} > 0 \tag{10}$$

$$\frac{d\mathcal{O}_\ell/m}{d\tau} < \frac{d\mathcal{O}_h/m}{d\tau}, \tag{11}$$

where we used the fact that for the market segment of higher risk borrowers there either exits the marginal borrower $\widehat{i}_{B,h}$ or the marginal borrower $\widehat{i}_{N,h}$ (or neither of them if we have a corner solution). Instead, for the market segment of lower risk borrowers there exists the marginal borrower $\widehat{i}_{B,h}$ (or not if we have a corner solution).

The first result in inequality (10) follows from $\frac{d\widehat{i}_{B,j}}{d\tau}, \frac{d\widehat{i}_{N,j}}{d\tau} \leq 0$, $\forall j$, where at least one of the derivatives holds with strict equality whenever there is an interior solution in at least one of the markets. The second result in inequality (11) is derived by examining the different cases that can arise. If online lending platforms are only active in the market for higher risk borrowers, the second result follows trivially since the left-hand side of inequality (11) is zero. Instead, if online lending platforms are active in both markets, they must compete in the market for lower risk borrowers with other banks, while they compete in the market for higher risk borrowers either with other banks or with the existing lender.

In both cases inequality (11) holds because:

$$\frac{d\mathcal{O}_\ell}{d\tau} = \frac{q_\ell}{2b} \quad < \quad \frac{d\mathcal{O}_h}{d\tau} = \frac{q_h}{2b}.$$

This concludes the derivations for Result 1.

B Main tables

Table B.1: Variable definitions

Variable	Description	Source
State level		
CFPB complaints	Total number of consumer complaints regarding banking services	Consumer Financial Protection Bureau (CFPB)
Generalized trust	A survey-based measure of general trust in other people	General Social Survey (GSS)
Total debt	Total dollar amount of household debt; state level	NY Fed Consumer Panel
P2P debt	Total dollar amount of P2P debt; state level	Platforms, computed by authors
GDP	Gross Domestic production in the past 12 months	US Census Bureau
Population	Total number of population registered in the state in the past 12 months	Bureau of Labor Statistics
Total Area	Total land area of the state	US Census Bureau
Branch	Total number of bank branches in the state	FDIC SoD data
College attainment rate	Percentage of population that have a college (or higher) degree	US Census Bureau
Bank competition	Herfindahl-Hirschman Index (HHI) calculated with each bank's market share in the state	Federal Deposite Insurance Corporation (FDIC)
Internet access	Percentage of population that has internet access from some location	Current Population Survey 2009
County level		
CFPB complaints	Total number of consumer complaints regarding banking services	CFPB
Total debt	Total dollar amount of household debt; county level	NY Fed Consumer Panel
GDP	Gross Domestic production in the past 12 months	US Census Bureau
Population	Total number of population registered in the state in the past 12 months	Bureau of Labor Statistics
Total Area	Total land area of the county	US Census Bureau
Branch	Total number of bank branches in the county	FDIC SoD data
College attainment rate	Percentage of population that have a college (or higher) degree	US Census Bureau
Bank competition	HHI calculated with each bank's market share in the county	FDIC
Internet access	Residential Fixed Connections over 200 kbps in at least one direction per 1,000 households	Current Population Survey
Loan-Borrower level		
Maturity	Maturity that is recorded under the loan identification number	Platforms
Loan size	Dollar Amount of loan that is applied for	Platforms
Loan interest rate	Interest rate that is assigned to the loan	Platforms
Credit rating	Credit rating	Platforms
Loan purpose	The purpose for the online lending loans	Borrowers
Income	Annual income	Borrowers reported, verified by platforms
Employment	Length of employment history	Borrowers reported, verified by platforms

Table B.2: Summary statistics

	Mean	SD	P25	Median	P75	N
State-level variables (by state)						
Generalized trust	0.40	0.11	0.33	0.39	0.46	49
State-level variables (by state and year)						
Credit card debt (10b$)	1.53	1.83	0.38	0.91	1.94	459
Total debt (10b$)	24.35	33.20	6.04	15.19	32.06	459
Population density(1000/km2)	0.11	0.43	0.02	0.03	0.07	459
Log GDP	12.21	1.01	11.43	12.29	12.94	459
Log population	14.94	1.03	14.20	15.08	15.64	459
Unemployment rate	6.76	2.12	5.10	6.60	8.10	459
Number of bank branches (k)	1.82	1.73	0.47	1.46	2.38	459
HHI	0.11	0.09	0.06	0.09	0.12	459
College+ attainment (in %)	27.15	5.44	23.60	26.00	29.80	459
State-level variables (by state and month)						
CFPB complaints per branch	0.20	0.66	0.06	0.10	0.16	2839
P2P debt (m$)	5.83	11.41	0.31	1.59	6.32	4653
P2P debt (m$)/Bank debt (10 bn$)	0.24	0.25	0.02	0.15	0.41	4653
Average DTI ratio	0.17	0.06	0.13	0.17	0.19	4653
Average interest rate	0.13	0.01	0.13	0.13	0.14	4653
Average annual income (k$)	75.08	26.41	65.78	73.16	81.56	4653
Average employment duration	5.30	1.11	4.95	5.50	5.82	4653
Loan-borrower level variables						
Credit card debt (per cap)	2947.62	566.29	2515.00	2880.00	3350.00	459
Total debt (per cap)	46358.24	12645.88	36595.00	43295.00	54255.00	459
Loan size	14961.15	8791.76	8000.00	13000.00	20000.00	1745948
Interest rate	0.14	0.05	0.10	0.13	0.16	1745948
High rating	0.37	0.48	0.00	0.00	1.00	1745948
Annual income (k$)	79.72	180.73	45.00	65.00	90.00	1745948
Employment	5.56	3.79	2.00	5.00	10.00	1745948

Notes: This table shows the summary statistics for all variables used in the empirical analysis. The sample covers the largest P2P platform, LendingClub, between 2008 and 2016. Where possible, we use state-level variables with monthly frequencies, while the loan-borrower level variables contain individual-specific information. The CFPB data extends back to 2012. Thus, the observation number is smaller than for other variables. It also means that the loan level regression with the CFPB complaints measure will use a fraction of the whole application sample. The variable *generalized trust* is computed as the state-level average of positive responses to the General Social Survey's (GSS) question about generalized trust over the 1973-2006 period, however there are only 49 states in our sample.

Table B.3: Summary statistics: county level

	Mean	SD	P25	Median	P75	N
County-level variables (by county and year)						
Pop. density (1000/km2)	0.111	0.750	0.008	0.020	0.052	7723
Log county income	14.070	1.495	13.024	13.877	14.917	7723
Log county population	10.503	1.402	9.567	10.365	11.313	7723
Log county jobs	9.739	1.491	8.665	9.567	10.600	7723
College	0.186	0.087	0.125	0.164	0.224	7723
Internet	0.003	0.001	0.003	0.003	0.004	7723
HHI	0.207	0.197	0.072	0.149	0.273	7723
County-level variables (by county and month)						
P2P debt (m$)/Bank debt (10 bn$)	0.195	0.298	0.041	0.100	0.221	53658
CFPB complaints per branch	0.187	1.149	0.000	0.000	0.074	53658
Average DTI ratio	0.182	0.058	0.148	0.179	0.215	53658
Average interest rate	0.145	0.031	0.128	0.144	0.161	53658
Average annual income (k$)	71.029	104.640	50.004	63.322	78.498	53658
Average employment duration	5.824	2.716	4.250	5.776	7.667	53658

Notes: This table shows the summary statistics for all variables used in the county level regressions. The sample covers the largest P2P platform, LendingClub, between 2012 and 2016. Where possible, we use county-level variables with monthly frequencies, while the loan-borrower level variables contain individual-specific information. The CFPB data doesn't cover all counties in different months.

Table B.4: Summary statistics: Difference-in-Differences test

	Treat				Control				Difference in mean
	Mean	SD	Median	N	Mean	SD	Median	N	
Average DTI ratio	0.18	0	0.06	240	0.18	0	0.06	320	-0.00
Average interest rate	0.14	0	0.03	240	0.15	0	0.03	320	0.01**
Average annual income (k$)	61.19	58	21.20	240	65.68	61	29.85	320	4.49*
Average employment duration	5.83	6	2.48	240	5.96	6	2.82	320	0.13
Pop. density (1000/km2)	0.29	0	0.62	240	0.09	0	0.22	320	-0.20***
Log county income	15.52	16	1.37	240	14.67	14	1.15	320	-0.85***
Log county population	11.83	12	1.25	240	11.04	11	1.07	320	-0.78***
Log county jobs	11.17	11	1.33	240	10.37	10	1.15	320	-0.81***
College	0.24	0	0.10	240	0.21	0	0.09	320	-0.04***
Internet	0.00	0	0.00	240	0.00	0	0.00	320	-0.00**
HHI	0.11	0	0.14	240	0.13	0	0.12	320	0.02*

Notes: This table compares characteristics of counties that are in the treatment and control group in the Difference-in-Differences regressions. A county is in the treatment group if it has a bank with a market share exceeding 20% that is fined for misconduct, and in the control group otherwise. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.5: Online lending and CFPB consumer complaints

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CFPB	0.017**	0.021***	0.023***	0.012***	0.009***	0.008***	0.009***	0.009***
	(0.007)	(0.005)	(0.007)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
DTI		0.669*		0.408*		0.284*		0.113***
		(0.366)		(0.216)		(0.150)		(0.038)
interest rate		-4.054***		-2.493**		-0.814		1.300*
		(1.413)		(1.091)		(1.078)		(0.712)
income		0.003***		0.001**		0.000		-0.000
		(0.001)		(0.001)		(0.000)		(0.000)
employment		-0.000		-0.001		0.004		0.013*
		(0.013)		(0.009)		(0.009)		(0.007)
pop. density		0.007		-1.299***		-1.195***		-1.141***
		(0.012)		(0.185)		(0.164)		(0.163)
log GDP		-0.049**		-0.546***		-0.404***		-0.442***
		(0.020)		(0.140)		(0.138)		(0.139)
log population		0.076***		1.746***		-0.293*		-0.213
		(0.018)		(0.564)		(0.156)		(0.145)
unemployment rate		-0.052***		-0.129***		-0.024***		-0.024***
		(0.007)		(0.010)		(0.003)		(0.003)
Year FE	NO	NO	NO	NO	YES	YES	NO	NO
State FE	NO	NO	YES	YES	YES	YES	YES	YES
Year-Month FE	NO	NO	NO	NO	NO	NO	YES	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
No. of observations	2839	2839	2839	2839	2839	2839	2839	2839
Adj. R-squared	0.002	0.270	0.111	0.567	0.684	0.692	0.874	0.881

Notes: This table reports the results for the period 2012-2016 using regression equation:

$$Y_{i,t} = \beta_1 \text{CFPB complaints}_{i,t} + \gamma X_{i,t} + (A_i + B_t) + \epsilon_{i,t}.$$

The dependent variable is the ratio of online debt (m$) to total household debt (10b$) in state i at a monthly frequency. *CFPB complaints* is the number of consumer complaints per branch in state i in month t. In the even-numbered columns, we include a number of independent variables to control for the average quality of loans. *DTI* is the simple average debt-to-income ratio of all the loans from state i at month t. *Interest rate* is the average value of the interest rates for loans originated in the state in month t. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in state i in month t. We also control for state level variables, such as population density ($1000/km^2$), the logarithm of GDP, the logarithm of population, and the state unemployment rate. Columns 5 and 6 include year fixed effects. Columns 3-8 include state fixed effects. And columns 7 and 8 include year-month fixed effects. Standard errors, in parentheses, are corrected for the clustering of observations by year-month. The results also hold with clustering at the state level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.6: Extensive margin: number of applicants/population and CFPB complaints

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CFPB	0.071***	0.079***	0.063***	0.030***	0.020***	0.020***	0.020***	0.019***
	(0.022)	(0.016)	(0.019)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)
DTI		1.527*		1.055*		0.720*		0.218**
		(0.903)		(0.561)		(0.390)		(0.105)
interest rate		-12.974***		-7.828**		-3.360		1.267
		(4.053)		(3.081)		(3.070)		(1.863)
income		0.008***		0.002		-0.000		-0.001
		(0.002)		(0.001)		(0.001)		(0.001)
employment		0.002		0.003		0.015		0.038**
		(0.039)		(0.024)		(0.022)		(0.017)
pop. density		-0.058		-1.516***		-1.235***		-1.155***
		(0.041)		(0.471)		(0.404)		(0.378)
log GDP		0.517***		-1.508***		-1.168***		-1.273***
		(0.064)		(0.380)		(0.364)		(0.368)
log population		-0.420***		7.300***		1.869***		2.078***
		(0.062)		(1.687)		(0.416)		(0.437)
unemployment rate		-0.150***		-0.371***		-0.100***		-0.100***
		(0.018)		(0.028)		(0.009)		(0.009)
Year FE	NO	NO	NO	NO	YES	YES	NO	NO
State FE	NO	NO	YES	YES	YES	YES	YES	YES
Year-Month FE	NO	NO	NO	NO	NO	NO	YES	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
No. of observations	2839	2839	2839	2839	2839	2839	2839	2839
Adj. R-squared	0.005	0.324	0.119	0.627	0.722	0.730	0.891	0.896

Notes: This table reports the extensive margin regression for the period 2012-2016:

$$Y_{i,t} = \beta_1 \text{CFPB complaints}_{i,t} + \gamma X_{i,t} + (A_i + B_t) + \epsilon_{i,t}.$$

The dependent variable is the fraction of state i's residents (in basis points) who are online loan applicants in month t. *CFPB complaints* is the number of consumer complaints per branch in state i in month t. In the even-numbered columns, we include a number of independent variables to control for the average quality of loans. *DTI* is the simple average debt-to-income ratio of all the loans from state i at month t. *Interest rate* is the average value of the interest rates for loans originated in the state in month t. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in state i in month t. We also control for state level variables, such as population density ($1000/km^2$), the logarithm of GDP, the logarithm of population, and the state unemployment rate. Columns 5 and 6 include year fixed effects. Columns 3-8 include state fixed effects. And columns 7 and 8 include year-month fixed effects. Standard errors, in parentheses, are corrected for clustering of observations by year-month. The results hold with the clustering at the state level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.7: Intensive margin: size of loan request and CFPB complaints

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CFPB	0.003	0.005	0.000	0.000	-0.001*	-0.000	-0.000	-0.000
	(0.007)	(0.007)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
homeowner		0.047***		0.019***		0.019***		0.019***
		(0.005)		(0.003)		(0.003)		(0.003)
has mortgage		0.034***		0.037***		0.037***		0.037***
		(0.003)		(0.002)		(0.002)		(0.002)
employment		0.003***		0.004***		0.004***		0.004***
		(0.000)		(0.000)		(0.000)		(0.000)
income		0.000***		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)		(0.000)
interest rate		0.316***		0.280***		0.290***		0.304***
		(0.025)		(0.020)		(0.017)		(0.017)
maturity		0.074***		0.074***		0.074***		0.073***
		(0.004)		(0.004)		(0.004)		(0.004)
Year FE	NO	NO	NO	NO	YES	YES	NO	NO
State FE	NO	NO	YES	YES	YES	YES	YES	YES
Year-Month FE	NO	NO	NO	NO	NO	NO	YES	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
No. of observations	1680778	1680778	1680778	1680778	1680778	1680778	1680778	1680778
Adj. R-squared	0.000	0.236	0.138	0.306	0.140	0.306	0.142	0.309

Notes: This table reports the intensive margin regression for the period 2012-2016:

$$Y_{i,d,t} = \beta_1 \text{CFPB complaints}_{i,t} + \gamma X_{i,d,t} + (A_i + B_t) + \epsilon_{i,d,t}.$$

The dependent variable is the loan size from borrower d in state i in month t. *CFPB complaints* is the number of consumer complaints per branch in state i in month t. The even-numbered columns include a number of independent variables to control for individual loan and borrower characteristics. Note that *homeowner* and *has mortgage* are dummy variables that capture an applicant's homeownership status and whether or not they have a mortgage. The variable *employment* measures the applicant's employment duration in years and the variable *income* measures the applicant's annual income. We also control for each applicant's FICO score, as well as two loan characteristics: the interest rate and the maturity in months. Columns 5 and 6 include year fixed effects. Columns 3-8 contain state fixed effects. And columns 7 and 8 contain year-month fixed effects. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.8: Online lending and CFPB complaints: by borrower quality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CFPB	0.010**	0.012***	0.014***	0.008***	0.007***	0.007***	0.007***	0.007***
	(0.004)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
highrating=1	-0.073***	-0.078***	-0.073***	-0.078***	-0.073***	-0.075***	-0.073***	-0.075***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
highrating=1 x CFPB	-0.005***	-0.004**	-0.005***	-0.004**	-0.005***	-0.005***	-0.005***	-0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
interest rate		-0.144***		-0.122***		-0.061**		-0.035*
		(0.041)		(0.033)		(0.026)		(0.018)
income		0.000		-0.000		-0.000		0.000
		(0.000)		(0.000)		(0.000)		(0.000)
employment		-0.000		0.000		0.000		-0.000
		(0.000)		(0.000)		(0.000)		(0.000)
pop. density		0.014***		-0.605***		-0.587***		-0.587***
		(0.004)		(0.090)		(0.081)		(0.081)
log GDP		-0.019***		-0.272***		-0.195***		-0.199***
		(0.006)		(0.070)		(0.066)		(0.066)
log population		0.036***		1.026***		-0.150**		-0.144**
		(0.007)		(0.308)		(0.071)		(0.070)
unemployment rate		-0.035***		-0.070***		-0.012***		-0.012***
		(0.004)		(0.005)		(0.001)		(0.001)
Year FE	NO	NO	NO	NO	YES	YES	NO	NO
State FE	NO	NO	YES	YES	YES	YES	YES	YES
Year-Month FE	NO	NO	NO	NO	NO	NO	YES	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
No. of observations	5644	5644	5644	5644	5644	5644	5644	5644
Adj. R-squared	0.085	0.272	0.176	0.558	0.678	0.682	0.843	0.847

Notes: This table repeats the specifications from table B.5, but includes additional measures for borrower quality. It regresses the ratio of online debt (m$) to total household debt (10b$) in state i at month t on a high credit rating dummy, the number of CFPB consumer complaints per bank branch, and their interactions. *highrating=1* means that the borrower has a FICO score higher than 700. *CFPB complaints* is the number of consumer complaints per branch in state i in month t. We include a number of independent variables to control for the average quality of loans. *DTI* is the simple average debt-to-income ratio of all loans originated in state i in month t. *Interest rate* is the average value of interest rates for P2P loans in state i in month t. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in state i in month t. At the state level, we control for population density ($1000/km^2$), the logarithm of GDP, the logarithm of population, and the state unemployment rate. All columns include the *highrating=1* dummy and its interaction with the CFPB complaint variable. All even-numbered columns contain the full set of controls. Columns 5 and 6 include year fixed effects. Columns 3-8 include state fixed effects. And columns 7 and 8 include year-month fixed effects. Standard errors, in parentheses, are corrected for clustering of observations by year-month. The results hold with the clustering at the state level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.9: Online lending and CFPB complaints: by generalized trust

	(1)	(2)	(3)	(4)	(5)
CFPB	0.023	0.029***	-0.001	0.021***	-0.002
	(0.015)	(0.006)	(0.004)	(0.006)	(0.004)
generalized trust	-0.394***	-0.709***	0.019	-0.406***	-0.079
	(0.114)	(0.141)	(0.134)	(0.072)	(0.066)
DTI	0.221**	-0.019	0.051	0.134*	0.042
	(0.092)	(0.063)	(0.065)	(0.079)	(0.050)
interest rate	1.540***	-1.584**	1.287**	1.652***	0.799*
	(0.596)	(0.707)	(0.518)	(0.508)	(0.426)
income	0.000	-0.002***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
employment	0.007	0.032***	0.012*	0.009	0.018***
	(0.007)	(0.006)	(0.007)	(0.006)	(0.005)
population density	-0.045***	0.156***	-0.033	-0.051***	-0.069
	(0.009)	(0.049)	(0.097)	(0.008)	(0.073)
log GDP	0.020	0.046**	0.134***	0.048***	0.134***
	(0.020)	(0.022)	(0.033)	(0.016)	(0.024)
log population	0.002	-0.027	-0.176***	-0.016	-0.161***
	(0.018)	(0.023)	(0.034)	(0.015)	(0.026)
unemployment rate	-0.003	0.005	0.014***	0.005	0.018***
	(0.004)	(0.004)	(0.005)	(0.003)	(0.004)
Subsample	Bottom	Middle	Top	Low	High
Year-Month FE	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
No. of observations	874	986	891	1301	1450
Adj. R-squared	0.008	0.346	0.580	0.593	0.729

Notes: This table reports the results of a subsample regression over the period 2012-2016:

$$Y_{i,t} = \beta_1 \text{CFPB complaints}_{i,t} + \beta_2 \text{Generalized trust}_i + \gamma X_{i,t} + (B_t) + \epsilon_{i,t}.$$

The dependent variable is the ratio of online debt (m$) to total household debt (10b$) in state i and at month t. Our key explanatory variable, *CFPB*, is the number of consumer complaints per branch in state i in month t. *Generalized trust* is a measure of social trust taken from the General Social Survey (GSS) during the 1973-2006 period. We consider two exhaustive partitions of states by generalized trust levels. The first divides states into bottom, middle, and top groups (columns (1)–(3)). And the second divides states into high and low groups (columns (4)–(5)). *DTI* is the simple average debt-to-income ratio of all loans originated in state i in month t. *Interest rate* is the average value of interest rates for P2P loans in state i in month t. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in state i in month t. At the state level, we control for population density ($1000/km^2$), the logarithm of GDP, the logarithm of population, and the state unemployment rate. All specifications include year-month fixed effects and the full set of controls. Standard errors, in parentheses, are corrected for clustering of observations by year-month. The results hold with the clustering at the state level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.10: Online lending and CFPB consumer complaints at the county level

	(1)	(2)	(3)	(4)	(5)	(6)
CFPB	0.021***	0.028***	0.022***	0.029***	0.021***	0.028***
	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)
DTI		-0.045*		-0.012		-0.048*
		(0.023)		(0.024)		(0.026)
interest rate		0.504***		0.554***		0.510***
		(0.042)		(0.042)		(0.045)
income		0.000***		0.000***		0.000***
		(0.000)		(0.000)		(0.000)
employment		0.003***		0.003***		0.003***
		(0.000)		(0.000)		(0.000)
pop. density (1000/km2)		0.025***		0.026***		0.025***
		(0.005)		(0.005)		(0.005)
log county income		-0.011		-0.004		-0.012
		(0.041)		(0.042)		(0.041)
log county population		-0.086**		-0.091**		-0.085**
		(0.034)		(0.036)		(0.034)
log county jobs		0.006		0.002		0.006
		(0.027)		(0.027)		(0.028)
college		0.218**		0.208**		0.219**
		(0.097)		(0.098)		(0.099)
internet		37.578***		36.735***		37.467***
		(12.434)		(12.391)		(12.638)
HHI		0.440***		0.440***		0.439***
		(0.097)		(0.098)		(0.099)
State-Year FE	YES	YES	NO	NO	NO	NO
State-Month FE	NO	NO	YES	YES	NO	NO
State-Year-Month FE	NO	NO	NO	NO	YES	YES
Controls	NO	YES	NO	YES	NO	YES
No. of observationservations	53658	53658	53658	53658	53658	53658
Adj. R-squared	0.159	0.368	0.133	0.356	0.152	0.363

Notes: This table reports the results of the following county-level regression equation:

$$Y_{i,j,t} = \beta_1 \text{CFPB complaints}_{i,j,t} + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t}.$$

The dependent variable measures the ratio of online debt (m$) to total household debt (10b$) in county j, state i, and month t. The key explanatory variable, *CFPB*, is the number of consumer complaints per branch in county j, state i, and month t. We include a number of independent variables to control for the average quality of loans. *DTI* is the simple average debt-to-income ratio of all the loans from county j at month t. *Interest rate* is the average value of interest rates for P2P loans in the county j at month t. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in county j and month t. County-level controls include population density ($1000/km^2$), the logarithm of county average income, the logarithm of population, the logarithm of number of jobs, the HHI of bank market shares (to proxy for bank competition), the percentage of the population with a college or above degree (to proxy for financial literacy), and internet usage. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.11: Online lending and CFPB complaints: by borrower quality at the county level

	(1)	(2)	(3)	(4)	(5)	(6)
CFPB	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
highrating=1	-0.010***	-0.005***	-0.009***	-0.005***	-0.010***	-0.005***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
highrating=1 x CFPB	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
DTI		-0.367***		-0.369***		-0.371***
		(0.044)		(0.045)		(0.045)
interest rate		-0.754***		-0.757***		-0.762***
		(0.124)		(0.121)		(0.128)
income		0.000		0.000		0.000
		(0.000)		(0.000)		(0.000)
employment		-0.003***		-0.003***		-0.003***
		(0.001)		(0.001)		(0.001)
pop. density (1000/km2)		0.012***		0.012***		0.012***
		(0.002)		(0.002)		(0.002)
log county income		0.002		0.003		0.002
		(0.025)		(0.025)		(0.026)
log county population		-0.106***		-0.107***		-0.106***
		(0.024)		(0.024)		(0.024)
log county jobs		-0.004		-0.004		-0.005
		(0.018)		(0.018)		(0.018)
college		0.069		0.067		0.069
		(0.068)		(0.069)		(0.068)
internet		10.633		10.211		10.645
		(7.085)		(7.073)		(7.150)
HHI		0.334***		0.331***		0.333***
		(0.066)		(0.067)		(0.067)
State-Year FE	YES	YES	NO	NO	NO	NO
State-Month FE	NO	NO	YES	YES	NO	NO
State-Year-Month FE	NO	NO	NO	NO	YES	YES
Controls	NO	YES	NO	YES	NO	YES
No. of observations	82950	82950	82950	82950	82950	82950
R-squared	0.156	0.468	0.141	0.465	0.166	0.474

Notes: This regression examines the impact of CFPB complaints on the county-level online lending share by borrower credit rating type. The dependent variable is the ratio of online debt (m$) to total household debt (10b$) in county j, state i, and month t. We regress the dependent variable on a dummy for high credit rating, CFPB consumer complaints, and their interactions. *Highrating=1* means that a borrower has a FICO score above 700. *CFPB* is the number of consumer complaints per branch in county j of state i in month t. We include a number of independent variables to control for the average quality of loans. *DTI* is the simple average debt-to-income ratio of all the loans from county j in month t. The variable *Interest rate* is computed as the average value of interest rates for P2P loans in county j in month t. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in county j in month t. County-level controls include population density ($1000/km^2$), the logarithm of income, the logarithm of population, the logarithm of number of jobs, the bank market share HHI (to proxy for bank competition), the percentage of the population with a college or above degree (to proxy for financial literacy), and internet usage. Even-numbered columns include the full set of controls. Additionally, columns 1 and 2 include state-year fixed effects, columns 3 and 4 include state-month fixed effects, and columns 5 and 6 include state-year-month fixed effects. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.12: Online lending and CFPB complaints with supply shocks

	(1)	(2)	(3)	(4)	(5)
CFPB	0.024***	0.018***	0.030***	0.030***	0.027***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)
supply			0.028***	0.012	-0.001
			(0.009)	(0.016)	(0.004)
CFPB x supply			-0.014	-0.015	0.002***
			(0.009)	(0.012)	(0.000)
DTI	-0.040	-0.017	-0.046*	-0.048*	-0.049*
	(0.031)	(0.038)	(0.026)	(0.026)	(0.026)
interest rate	0.547***	0.249***	0.511***	0.511***	0.511***
	(0.051)	(0.054)	(0.046)	(0.045)	(0.045)
income	0.000***	0.000	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
employment	0.004***	0.001	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
pop. density (1000/km2)	0.233**	0.013***	0.025***	0.025***	0.025***
	(0.099)	(0.001)	(0.005)	(0.005)	(0.005)
log county income	-0.031	0.068**	-0.009	-0.010	-0.014
	(0.039)	(0.031)	(0.041)	(0.042)	(0.041)
log county population	-0.105***	-0.088	-0.087**	-0.086**	-0.085**
	(0.030)	(0.060)	(0.034)	(0.035)	(0.034)
log county jobs	0.004	-0.015	0.002	0.005	0.008
	(0.032)	(0.041)	(0.028)	(0.028)	(0.028)
college	0.235*	0.053	0.223**	0.226**	0.221**
	(0.127)	(0.102)	(0.099)	(0.099)	(0.099)
internet	50.924***	-14.565	37.323***	37.064***	37.635***
	(13.181)	(8.851)	(12.645)	(12.652)	(12.668)
HHI	0.374***	0.111**	0.434***	0.439***	0.437***
	(0.100)	(0.042)	(0.098)	(0.099)	(0.101)
State-Year FE	NO	NO	NO	NO	NO
State-Month FE	NO	NO	NO	NO	NO
State-Year-Month FE	YES	YES	YES	YES	YES
Controls	COUNTY	COUNTY	COUNTY	COUNTY	COUNTY
Cluster	STATE	STATE	STATE	STATE	STATE
No. of observations	41472	12186	53658	53658	53658
Adj. R-squared	0.378	0.310	0.364	0.363	0.363

Notes: This table reports the results of regression equation:

$$Y_{i,j,t} = \beta_1 \text{CFPB complaints}_{i,j,t} + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t},$$

or the interaction with the supply shock:

$$Y_{i,j,t} = \beta_1 \text{CFPB complaints}_{i,j,t} + \beta_2 \text{Supply shocks}_{i,j} + \beta_3 \text{CFPB complaints}_{i,j,t} \text{ x Supply shocks}_{i,j} + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t}.$$

We use CFPB consumer complaints, supply shocks from the closure of the Office of Thrift Supervision (OTS), and bank capital changes at the county level as the key explanatory variables. The dependent variable is the ratio of online debt (m$) to total household debt (10b$) in county j, state i, and month t. *CFPB* is the number of consumer complaints per branch in county j of state i in month t. We include a number of independent variables to control for the average quality of loans. *DTI* is the simple average debt-to-income ratio of all the loans from county j at month t. *Interest rate* is the average value of interest rates for P2P loans in the county j at month t. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in county j and month t. We also control for county level variables such as population density ($1000/km^2$), the logarithm of total income, the logarithm of population, the logarithm of number of jobs, bank market share HHI (to proxy for bank competition), the percentage of the population with a college or above degree (to proxy for financial literacy), and internet usage. All columns contain state-year-month fixed effects, as well as the full set of state and county-level controls. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.13: Online lending and CFPB complaints: by credit balances

	(1)	(2)	(3)	(4)	(5)
CFPB	0.025***	0.025***	0.025***	0.028***	0.018***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
high revolver	0.100***	0.100***	0.100***	0.102***	0.087***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
CFPB x high revolver	0.007***	0.007***	0.007***	0.007*	0.008***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.001)
DTI	-0.289***	-0.294***	-0.289***	-0.280***	-0.291***
	(0.019)	(0.019)	(0.019)	(0.020)	(0.030)
interest rate	-0.100***	-0.115***	-0.103***	-0.053	-0.260***
	(0.029)	(0.027)	(0.030)	(0.036)	(0.056)
income	0.000***	0.000***	0.000***	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
employment	0.000	-0.001**	0.000	0.000	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
pop. density (1000/km2)	-0.006***	-0.006***	-0.006***	0.086	-0.006***
	(0.002)	(0.002)	(0.002)	(0.054)	(0.002)
log county income	0.072**	0.079**	0.072**	0.070*	-0.028
	(0.031)	(0.030)	(0.031)	(0.039)	(0.029)
log county population	-0.245***	-0.254***	-0.245***	-0.246***	-0.177***
	(0.038)	(0.038)	(0.039)	(0.044)	(0.041)
log county jobs	0.115***	0.115***	0.115***	0.108***	0.165***
	(0.017)	(0.017)	(0.018)	(0.025)	(0.030)
college	-0.011	-0.027	-0.013	-0.049	0.033
	(0.058)	(0.058)	(0.059)	(0.066)	(0.083)
internet	1.562	1.181	1.607	9.750	-10.162
	(6.842)	(6.838)	(6.940)	(10.421)	(8.929)
HHI	0.324***	0.321***	0.323***	0.323***	0.032
	(0.049)	(0.049)	(0.050)	(0.075)	(0.039)
State-Year FE	YES	NO	NO	NO	NO
State-Month FE	NO	YES	NO	NO	NO
State-Year-Month FE	NO	NO	YES	YES	YES
Controls	LOAN	LOAN	LOAN	COUNTY	COUNTY
Cluster	STATE	STATE	STATE	STATE	STATE
No. of observations	76898	76898	76898	46497	20070
Adj. R-squared	0.384	0.382	0.394	0.390	0.439

Notes: This table reports the regression of the ratio of online debt (m$) to total household debt (10b$) in county j in month t on a dummy for a higher borrower credit balance, CFPB consumer complaints, and their interaction:

$$Y_{i,j,t} = \beta_1 \text{CFPB complaints}_{i,j,t} + \beta_2 \text{High revolvers}_{i,j} + \beta_3 \text{CFPB complaints}_{i,j,t} \text{ x High revolvers}_{i,j} + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t}.$$

Column 1 includes state-year fixed effects, column 2 includes state-month fixed effects, and column 3 includes state-year-month fixed effects. Column 4 contains results from counties not exposed to the OTS closure supply shock to and column 5 includes results from those that were exposed. The variable *high revolver=1* indicates that the borrower has a revolving credit balance higher than the median borrower from the same state over the sample period. *CFPB* is the number of consumer complaints per branch in county j of state i in month t. We include a number of independent variables to control for the average quality of loans, including the average borrower's *DTI*, *Interest rate*. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in county j in month t. At the county-level, we also control for population density ($1000/km^2$), the logarithm of income, the logarithm of population, and the logarithm of number of jobs, the bank market share HHI (to proxy for bank competition), the percentage of population with a college or above degree (to proxy for financial literacy), and internet usage. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table B.14: Difference-in-Differences test

	(1)	(2)	(3)	(4)
Misconduct	0.098**	0.253***	0.045**	0.203***
	(0.037)	(0.035)	(0.022)	(0.029)
Post	-0.020	-0.203***	-0.008	-0.209***
	(0.033)	(0.027)	(0.035)	(0.029)
Misconduct x Post	0.241***	0.338***	0.304***	0.396***
	(0.040)	(0.035)	(0.048)	(0.043)
DTI		0.063		0.074
		(0.051)		(0.050)
interest rate		0.532***		0.515***
		(0.136)		(0.132)
income		0.001***		0.001***
		(0.000)		(0.000)
employment		-0.000		0.000
		(0.001)		(0.001)
pop. density (1000/km2)		0.167***		0.165***
		(0.048)		(0.049)
log county income		0.166		0.160
		(0.100)		(0.098)
log county population		-0.372***		-0.365***
		(0.118)		(0.114)
log county jobs		-0.055		-0.054
		(0.050)		(0.050)
college		0.147		0.173
		(0.279)		(0.271)
internet		-30.760		-32.007
		(21.882)		(21.638)
HHI		0.834***		0.836***
		(0.142)		(0.144)
State FE	YES	YES	NO	NO
Year-Month FE	YES	YES	NO	NO
State-Year-Month FE	NO	NO	YES	YES
Controls	NO	YES	NO	YES
Cluster	STATE	STATE	STATE	STATE
No. of observations	52,444	52,444	52,363	52,363
Adj. R-squared	0.173	0.444	0.172	0.443

Notes: This table contains the results of the equation:

$$Y_{i,j,t} = \beta_1 \text{Misconduct}_{i,j,t} + \beta_2 \text{Post}_t + \beta_3 \text{Misconduct}_{i,j,t} \text{ x Post}_t + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t},$$

where $\text{Misconduct}_{i,j,t}$ is the treatment dummy variable that indicates whether county j of state i has a bank with a scandal that had a market share higher than 20% in the year before the misconduct disclosure. Post_t is an event dummy that is equal to 1 after the bank scandal is reported in the news media. We include a number of controls for the average quality of loans, including the average borrower's *DTI* and the *interest rate*. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in county j in month t. At the county-level, we also control for population density ($1000/km^2$), the logarithm of income, the logarithm of population, the logarithm of number of jobs, the bank market share HHI (to proxy for bank competition), the percentage of the population with a college or above degree (to proxy for financial literacy), and internet usage. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

Table B.15: Difference-in-Differences test with a continuous measure of treatment

	(1)	(2)	(3)	(4)
Misconduct	0.730**	1.998***	0.349**	1.652***
	(0.284)	(0.275)	(0.168)	(0.229)
Post	0.038	-0.125***	0.013	-0.155***
	(0.045)	(0.043)	(0.041)	(0.045)
Misconduct x Post	0.715***	1.228***	1.144***	1.601***
	(0.229)	(0.238)	(0.237)	(0.269)
DTI		0.059		0.067
		(0.052)		(0.051)
interest rate		0.540***		0.519***
		(0.136)		(0.133)
income		0.001***		0.001***
		(0.000)		(0.000)
employment		-0.000		-0.000
		(0.001)		(0.001)
pop. density (1000/km2)		0.180***		0.177***
		(0.053)		(0.055)
log county income		0.177*		0.171*
		(0.091)		(0.090)
log county population		-0.361***		-0.356***
		(0.107)		(0.104)
log county jobs		-0.086		-0.085
		(0.053)		(0.054)
college		0.218		0.240
		(0.219)		(0.213)
internet		-31.970		-33.142
		(24.262)		(24.023)
HHI		0.914***		0.913***
		(0.160)		(0.162)
State FE	YES	YES	NO	NO
Year-Month FE	YES	YES	NO	NO
State-Year-Month FE	NO	NO	YES	YES
Controls	NO	YES	NO	YES
Cluster	STATE	STATE	STATE	STATE
No. of observations	52,444	52,444	52,363	52,363
Adj. R-squared	0.156	0.440	0.155	0.440

Notes: We repeat the analysis of Table B.14 with one difference: *Misconduct* is now a continuous measure of treatment. It is defined as the pre-scandal deposit market share of the banks with a major scandal, divided by total county deposits in the year before the misconduct disclosure. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

Table B.16: Online lending and CFPB complaints: classified bank misconduct

	(1)	(2)	(3)	(4)
CFPB	0.141***	0.030***	0.230***	0.028***
	(0.022)	(0.005)	(0.049)	(0.005)
DTI	-0.045	-0.045*	-0.045	-0.044*
	(0.027)	(0.027)	(0.027)	(0.026)
interest rate	0.519***	0.514***	0.519***	0.516***
	(0.048)	(0.049)	(0.048)	(0.049)
income	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
employment	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
pop. density (1000/km2)	0.025***	0.025***	0.026***	0.025***
	(0.005)	(0.005)	(0.006)	(0.005)
log county income	-0.010	-0.009	-0.014	-0.009
	(0.041)	(0.041)	(0.043)	(0.041)
log county population	-0.086**	-0.085**	-0.086**	-0.085**
	(0.033)	(0.034)	(0.032)	(0.034)
log county jobs	0.005	0.003	0.008	0.003
	(0.026)	(0.026)	(0.026)	(0.026)
college	0.225**	0.230**	0.215**	0.230**
	(0.103)	(0.103)	(0.103)	(0.103)
internet	37.938***	37.199***	40.496***	37.171***
	(13.028)	(12.589)	(14.724)	(12.538)
HHI	0.442***	0.444***	0.439***	0.444***
	(0.101)	(0.100)	(0.103)	(0.100)
State-Year FE	NO	NO	NO	NO
State-Month FE	NO	NO	NO	NO
State-Year-Month FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Cluster	STATE	STATE	STATE	STATE
No. of observations	53658	53658	53658	53658
Adj. R-squared	0.363	0.364	0.359	0.364

Notes: This table reports the results of the following county-level regression:

$$Y_{i,j,t} = \beta_1 \text{CFPB complaints}_{i,j,t} + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t}$$

The dependent variable is the ratio of online debt (m$) to total household debt (10b$) in county j, state i, and month t. The CFPB complaints variable is exclusively constructed using complaints that were classified either as "misconduct" (columns 1 and 3) and "not misconduct" (columns 2 and 4). Columns 1 and 2 classify misconduct and non-misconduct by using CFPB narratives to assign a class to each issue category. Columns 3 and 4 instead perform classification at the sub-issue level. We include a number of independent variables to control for the average quality of loans, including the average borrower's *DTI* and *Interest rate*. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in county j in month t. At the county-level, we also control for population density ($1000/km^2$), the logarithm of income, the logarithm of population, the logarithm of number of jobs, the bank market share HHI (to proxy for bank competition), the percentage of the population with a college or above degree (to proxy for financial literacy), and internet usage. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

C Online Appendix

Table C.1: Online lending and CFPB consumer complaints with additional controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CFPB	0.017**	0.023***	0.023***	0.012***	0.009***	0.008***	0.009***	0.009***
	(0.007)	(0.005)	(0.007)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
DTI		0.537*		0.408*		0.280*		0.108***
		(0.319)		(0.216)		(0.149)		(0.038)
interest rate		-4.225***		-2.487**		-0.796		1.324*
		(1.397)		(1.088)		(1.083)		(0.715)
income		0.003***		0.001**		0.000		-0.000
		(0.001)		(0.001)		(0.000)		(0.000)
employment		0.002		-0.001		0.003		0.013*
		(0.013)		(0.009)		(0.009)		(0.007)
population density		0.041**		-1.298***		-1.196***		-1.141***
		(0.019)		(0.184)		(0.166)		(0.164)
log GDP		0.106***		-0.552***		-0.381***		-0.415***
		(0.024)		(0.142)		(0.135)		(0.135)
log population		-0.066***		1.757***		-0.374**		-0.305**
		(0.024)		(0.585)		(0.157)		(0.147)
unemployment rate		-0.056***		-0.129***		-0.021***		-0.021***
		(0.006)		(0.011)		(0.003)		(0.002)
college		-1.210***		73.285**		-39.587**		-35.111**
		(0.095)		(35.339)		(16.130)		(15.174)
internet		-0.210**		-31.681*		18.940**		16.587**
		(0.086)		(16.440)		(8.185)		(7.658)
HHI		0.215***		-0.120		0.453***		0.515***
		(0.035)		(0.262)		(0.158)		(0.137)
Year FE	NO	NO	NO	NO	YES	YES	NO	NO
State FE	NO	NO	YES	YES	YES	YES	YES	YES
Year-Month FE	NO	NO	NO	NO	NO	NO	YES	YES
Controls	NO	YES	NO	YES	NO	YES	NO	YES
No. of observations	2839	2839	2839	2839	2839	2839	2839	2839
Adj R-squared	0.002	0.326	0.111	0.567	0.684	0.693	0.874	0.882

Notes: We repeat the regression in Table B.5. The dependent variable is the ratio of online debt (m$) to total household debt (10b$) in state i and in month t. CFPB complaints is the number of consumer complaints per branch in state i in month t. In the even-numbered columns, we include a number of independent variables to control for the average quality of loans. DTI is the simple average debt-to-income ratio of all the loans from state i at month t. Interest rate is the average value of the interest rates for loans originated in the state in month t. Income and Employment measure the average annual income and years of employment for the borrowers in state i in month t. We also control for state level variables, such as population density (1000/km2), the logarithm of GDP, the logarithm of population, and the state unemployment rate. Columns 5 and 6 include year fixed effects. Columns 3-8 include state fixed effects. And columns 7 and 8 include year-month fixed effects. In addition, we also control for the percentage of the population with a college degree, internet usage and the HHI Index of bank competition. Standard errors, in parentheses, are corrected for clustering of observations by year-month. The results hold with the clustering at the state level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.2: Online lending and lagged CFPB consumer complaints at the county level

	(1)	(2)	(3)	(4)	(5)	(6)
CFPB	0.001***	0.002***	0.001***	0.002***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
DTI		-0.092**		-0.061*		-0.099**
		(0.035)		(0.034)		(0.038)
interest rate		0.408***		0.442***		0.410***
		(0.049)		(0.046)		(0.051)
income		0.000**		0.000**		0.000**
		(0.000)		(0.000)		(0.000)
employment		0.003***		0.003***		0.003***
		(0.000)		(0.000)		(0.000)
pop. density (1000/km2)		0.022***		0.022***		0.021***
		(0.004)		(0.004)		(0.004)
log county income		-0.023		-0.017		-0.024
		(0.037)		(0.038)		(0.038)
log county population		-0.050***		-0.055***		-0.049***
		(0.014)		(0.016)		(0.014)
log county jobs		0.002		-0.002		0.002
		(0.032)		(0.031)		(0.032)
college		0.178*		0.170*		0.180*
		(0.092)		(0.091)		(0.094)
internet		26.993*		25.209*		26.955*
		(15.239)		(14.922)		(15.663)
HHI		0.421***		0.418***		0.420***
		(0.124)		(0.124)		(0.127)
State-Year FE	YES	YES	NO	NO	NO	NO
State-Month FE	NO	NO	YES	YES	NO	NO
State-Year-Month FE	NO	NO	NO	NO	YES	YES
Controls	NO	YES	NO	YES	NO	YES
No. of observations	41177	41177	41177	41177	41177	41177
Adj. R-squared	0.136	0.293	0.108	0.280	0.126	0.284

Notes: This table reports the results of the following regression:

$$Y_{i,j,t} = \beta_1 \text{CFPB complaints}_{i,j,t-1} + \gamma X_{i,j,t} + (A_{i,t}) + \epsilon_{i,j,t}.$$

The lagged CFPB consumer complaints is the key explanatory variable. The dependent variable is the ratio of online debt (m$) to total household debt (10b$) in county j in state i in month t. *CFPB* is the number of consumer complaints per branch in county j of state i in month $t-1$. We include a number of independent variables to control for the average quality of loans. *DTI* is the simple average debt-to-income ratio of all the loans originated in county j at month t. *Interest rate* is the average value of interest rates for P2P loans in the county j at month t. *Income* and *Employment* measure the average annual income and years of employment for the borrowers in county j and month t. County-level controls include population density ($1000/km^2$), the logarithm of county average income, the logarithm of population, the logarithm of the number of jobs, bank market share HHI (to proxy for bank competition), the percentage of the population with a college or above degree (to proxy for financial literacy), and internet usage. Standard errors, in parentheses, are corrected for clustering of observations by state. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.