THREE ESSAYS ON THE ECONOMIC IMPACT OF FINTECH LENDING

XIAORAN JIA

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN ADMINISTRATION
YORK UNIVERSITY
TORONTO, ONTARIO, CANADA

MARCH 2023

ABSTRACT

This dissertation follows the "three papers" dissertation model. It consists of three independent papers with a related theme focusing on the economic and capital market impacts of FinTech lending. While "FinTech" can be broadly described as the use of technological innovations on financial products and the provision of financial services, this dissertation will focus on the "Peer-to-peer (P2P) lending platforms" (or "FinTech lenders", "P2P lenders"). The disruptive impact of FinTech on the consumer lending sector has been increasingly acknowledged by academia in recent years (e.g., Goldstein et al. 2019). Existing studies have attributed the rapid growth of FinTech lending to regulatory arbitrage (e.g., Buchak et al. 2018), technological advantages (e.g., Fuster et al. 2019; Berg et al. 2020), and incumbent bank types (Balyuk et al. 2020). However, much less attention has been paid to the following aspects: 1) how incumbent banks react to the FinTech penetration, 2) how P2P lending changes the landscape of lending discrimination, especially with the availability of non-traditional information, and 3) whether digital inclusion, which facilities the use of internet and dissemination of information, is a potential determinant of FinTech penetration. This dissertation fills in these gaps and contributes to the literature by exploring the influence of FinTech penetration on traditional banks' risk-taking, the influence of non-traditional information on lending discrimination in P2P lending, and whether digital inclusion is a potential determinant of P2P penetration.

Chapter 1 "FinTech Competition and Bank Risk-taking" examines how FinTech-induced competition in the consumer lending market influences the changes in risk-taking of incumbent community banks and discusses potential financial stability implications. I find that banks' future change in risk-taking is positively associated with their current exposure to FinTech penetration. Path analysis shows that FinTech penetration influences bank risk-taking through the erosion of

bank charter value. Additionally, cross-sectional analysis shows that the risk-increasing effect of FinTech penetration is stronger for banks with lower ex-ante charter value, less discretionary loan loss provisions, and greater reliance on hard information. My findings suggest that 1) banks that discretionarily provision for more loan losses are better risk-disciplined, 2) weaker banks are more incentivized to increase risk-taking, and 3) banks that rely more on hard information in the loan-screening process are more incentivized to increase risk-taking in response to FinTech penetration.

In Chapter 2 "Peer-to-peer FinTech Lending, Non-traditional Information, and Racial Discrimination," I hypothesize that racial discrimination can exist in P2P lending even when race-related information is not directly observable, and that the degree of racial discrimination decreases in the precision of credit quality signals generated from both traditional and non-traditional information. I find strong evidence that loan listings in counties with a greater proportion of minority populations are associated with higher lending rates and higher loan denial rates. In cross-sectional tests, I document that racial discrimination is less pronounced when the availability of both traditional and non-traditional information is greater. Employing path analysis, I find that racial information is transmitted through the P2P platform's internal rating algorithms that utilize non-traditional information and the decision-making of platform investors.

Chapter 3 "Digital Inclusion and Financial Inclusion: Evidence from Peer-to-peer Lending" investigates the influence of digital inclusion on financial inclusion. I document that digital inclusion is positively associated with P2P lending penetration, with such relation more pronounced in county-years with more vulnerable/excluded populations. The results are robust to the use of instrumental variable (2SLS) approach, alternative measurements, weighted least squares regression, additional controls, and single-year analysis. In addition, I document that higher-risk borrowing is less likely to be denied in county-years with higher digital inclusion.

ACKNOWLEDGMENTS

First, I would like to express my deepest gratitude to my supervisor, Dr. Kiridaran (Giri) Kanagaretnam, for his valuable guidance, constructive feedback, and continuous support throughout my PhD journey. His expert knowledge, teaching, and insightful comments have been instrumental in shaping this thesis. I also owe special thanks to him for seeing my potential as an accounting researcher and admitting me into the PhD program at Schulich. I am forever grateful for his contribution to my intellectual and academic growth.

Second, I would like to thank my dissertation committee: Dr. Kiridaran (Giri) Kanagaretnam, Dr. Gregory Saxton, and Dr. Ambrus Kecskés, for their guidance, encouragement, and insightful comments. Their feedback and constructive criticism have been invaluable in improving the quality of this dissertation. I would also like to thank to all Schulich faculty members and my PhD peers for their support and encouragement during my doctoral studies. Their positive attitude and willingness to help have made this journey more enjoyable and fulfilling.

Third, I would like to express my sincere appreciation and gratitude to my two external examiners, Dr. Tiemei (Sarah) Li and Dr. Antoine Djogbenou, who have generously offered their time, expertise, and valuable feedback in reviewing my doctoral dissertation.

Finally, I would like to acknowledge the love, support, and patience of my family, especially my wife Xianqiong. Her support and encouragement have kept me motivated during challenging times. I also would like to thank my parents. My success would not have been possible without their constant love and support. I would like to give special thanks to my one-and-a-half-year-old son, who has been a source of inspiration and motivation throughout the writing of this dissertation.

TABLE OF CONTENTS

Chapter 1: FinTech Penetration and Bank Risk-taking

Abstract

Based on a sample of U.S. community banks and FinTech loans data from LendingClub and Prosper, I find a positive association between banks' future change in risk-taking and their current exposure to FinTech penetration. Path analysis reveals that FinTech penetration influences bank risk-taking through the erosion of bank charter value. Cross-sectional analysis suggests that banks with the lowest ex-ante charter values increase risk-taking the most. Additionally, I document that the risk-increasing effect of FinTech penetration is less pronounced for banks with more conservative loan loss provisioning and less reliance on hard information. My results are robust to alternative measures, propensity score matching, and a battery of sensitivity and placebo tests. My findings suggest that regulators may need to pay more attention to smaller banks with lower or deteriorating charter values.

Keywords: FinTech lending; Bank competition; Bank risk-taking; Bank loan loss accounting; Financial stability; Community banks

JEL Classifications: D14 D53 G21 G23 G32 O31

1. Introduction

The "peer-to-peer lending platforms" (or "FinTech lenders") have been growing rapidly around the world during the past decade. For example, FinTech companies held only about 0.9% of the global unsecured personal loan balances in 2010, but the percentage increased to 36.2% in 2017.[1] My data indicate a similar pattern in the U.S. —the ratio of the three-year cumulative successfully funded LendingClub and Prosper loans to community bank consumer loans increased dramatically for most U.S. states, jumping from an average of 2% in 2010 to an average of 71% in 2018 (Table 1b). Such a dramatic change indicates increasing threats from FinTech lenders to traditional banks in the retail consumer lending market.

Existing studies have attributed the rapid growth of FinTech lending to regulatory arbitrage (e.g., Buchak et al. 2018; Tang 2019; De Roure et al. 2022), technological advantages (e.g., Fuster et al. 2019; Berg et al. 2020), and market structure (Balyuk et al. 2020). However, much less attention has been paid to how banks react to FinTech-induced competition. I examine how community banks' current exposure to FinTech penetration influences their future change in risk-taking. More specifically, I study the paths through which FinTech penetration influences future bank risk-taking and explore how such a relation differs cross-sectionally.

Prior studies that examine the relationship between bank competition and bank risk-taking provide competing hypotheses. On the one hand, studies supporting a "competition-fragility" view argue that competition increases bank risk (e.g., Keeley 1990; Allen and Gale 2000; Hellmann et al. 2000; Repullo 2004), because the increase in bank competition erodes the incumbent banks' charter value, increasing banks' incentive to take more risks (Keeley 1990; Allen and Gale 2000). Under this hypothesis, FinTech penetration should be positively related to future changes in bank

[1] Source: https://www.statista.com/statistics/1019891/unsecured-personal-loans-balances/

risk-taking. Since banks' incentive to increase risk-taking is negatively associated with their charter value, I conjecture that the change in bank charter value should play a mediating role for the relation between FinTech penetration and banks' future change in risk-taking. On the other hand, a "competition-stability" view argues that competition makes banks less risky, as bank borrowers can have less risk-shifting incentives when they benefit from lower bank loan rates (Boyd and De Nicoló 2005). Under this alternative hypothesis, FinTech penetration should be negatively related to future changes in bank risk-taking.

Banks' risk-increasing incentives can also differ cross-sectionally. First, since banks with lower charter value are less concerned about avoiding bank failure (Keeley, 1990), *ex-ante* weaker banks can have stronger risk-increasing incentives when facing FinTech penetration. Second, banks that are more conservative in loan loss provisioning can have less risk-increasing incentives when exposed to FinTech penetration, as banks can be more risk-disciplined when they incorporate more information about future credit losses in their loan loss provisions (Bushman and Williams 2002). Third, as FinTech lenders are superior at processing hard information,[2] banks that rely more on hard information and less on relationship banking can be more threatened by FinTech penetration (e.g., Balyuk et al. 2020), hence they are more likely to increase risk-taking.

I use individual loan-level data from LendingClub and Prosper, two major peer-to-peer lending platforms in the U.S.[3] The primary measures of FinTech penetration are the two-year cumulative state-level successfully funded LendingClub and Prosper loans ("FinTech loans") per capita.[4] I limit my sample to community commercial banks because they operate more locally

[2] Hard information refers to conventional information such as credit bureau scores and financial statements, and unconventional "hardened" information that is transformed from non-standardized consumer data such as consumer online shopping and browsing history (e.g., Balyuk et al. 2020; Boot et al. 2020).
[3] A 2020 market research report shows that LendingClub and Prosper are the two largest P2P lenders in the U.S. (https://mangosoft.tech/blog/top-5-peer-to-peer-lending-companies-2020-full-market-research)
[4] For robustness purposes, I also use a first three digits of a Zip Code ("zip3") proxy, a FinTech loans to total bank loans proxy, and a FinTech loans to total bank consumer loans proxy.

compared to noncommunity banks and hence are more likely to be affected by the interstate variations of FinTech penetration. I proxy for the bank-level charter value using the Lerner index,[5] which measures the pricing power of banks in the market (e.g., Berger et al. 2009; Beck et al. 2013; Anginer et al. 2014).[6] Moreover, bank risk-taking is measured using the negative of the natural logarithm of the Z-score,[7] the volatility of return on assets, and the volatility of net interest margin. The change in risk-taking of a bank-quarter is measured as the level of risk-taking in the subsequent two years minus the level of risk-taking in the prior two years.

My baseline results support the "competition-fragility" hypothesis. With a final sample of 5,458 community banks in the U.S. during 2009Q1-2015Q4,[8] I find that banks' current exposure to FinTech penetration is positively related to future changes in bank risk, after controlling for both changing and static bank characteristics, macroeconomic conditions, and bank and year fixed effects. This influence is also economically significant: a one-standard-deviation-increase of a bank's current exposure to FinTech penetration is associated with a 44%-47% increase in the bank's future change in risk-taking.

Also, consistent with the "competition-fragility" hypothesis, the path analysis shows that FinTech penetration influences change in bank risk through the change in bank charter value. In the single-mediator analysis, I show that FinTech penetration is associated with a decrease in bank charter value and such a decrease in charter value is in turn related to an increase in future change in bank risk. In the multi-mediator analysis, I document that FinTech penetration is positively

[5] Lerner index is defined as (price – marginal cost) / marginal cost. The construction of the Lerner index will be introduced in detail in section 3.2. I do not use other market power indicators such as the H-statistic and Hirschman–Herfindahl index as they measure area-level market power and do not provide the bank-level variations I need.
[6] I also use net interest margin and return on assets as alternative measures of bank charter value.
[7] Since the Z-score is highly skewed, I use the natural logarithm of Z-score and the negative, so that the higher the measure, the greater the bank risk-taking.
[8] Note that I exclude the years 2016-2017 for measurement validity concerns explained in section 3. However, in sensitivity tests, I repeat all main regressions with an extended sample period (2009-2017) and get similar results.

associated with future change in bank risk through multiple channels related to bank charter value, including a decrease in bank profitability, a decrease in the level of bank capital, and a decrease in bank charter value.

In cross-sectional tests, I show that banks' risk-increasing incentives are partly driven by banks' *ex-ante* charter value, risk discipline, and vulnerability to FinTech threat. First, I document that the risk-increasing effect of FinTech penetration is stronger for *ex-ante* weaker banks. For example, for banks with the lowest charter value, a one-standard-deviation-increase in FinTech penetration is associated with a 51.67% increase in bank risk-taking, while such a relation is non-existent for banks with higher charter value. This finding implies that managers in weaker banks tend to have greater risk-shifting incentives when facing FinTech penetration. Second, I find that banks with a higher level of discretionary loan loss provisions (DLLP) in the current period (i.e., banks that are more conservative in their financial reporting) experience a lower increase in risk-taking when facing FinTech penetration. I also document that banks, when facing the FinTech threat, tend to be more risk-disciplined if their current period DLLP is positive. This finding suggests that banks that are more conservative in loan loss provisioning are also more risk-disciplined, implying that banks' forward-looking use of discretion in loan loss provisions can play a significant role in limiting their risk-increasing incentives when exposed to FinTech penetration (Bushman and Williams 2012). Last, I document that banks with a larger size, more homogeneous loans, or more residential real estate loans tend to experience a greater increase in risk-taking in response to FinTech penetration, implying that banks that rely more on hard information in the loan screening process are more likely to respond to a FinTech threat by increasing risk-taking.

To further explore how banks increase risk-taking, I examine the relationship between FinTech penetration and changes in bank loan risk. Following Bushman et al. (2016), I investigate how FinTech penetration influences the relationship between current period loan growth and the increase in net loan charge-offs (NCOs) in future periods. I find that the portion of a bank's current loans that are charged off increases with the bank's exposure to FinTech competition. Notably, such influence is nonsignificant for banks with the highest *ex-ante* charter values. In addition, I document that FinTech penetration is positively related to changes in banks' loan loss provisions (LLP) and NCOs measured in two-year windows, with such a relation negatively moderated by banks' *ex-ante* charter value. These findings suggest that banks tend to initiate riskier loans when facing the FinTech threat, with such a relation more pronounced for banks with lower charter value.

My identification strategy relies on the large and significant inter-regional variations in FinTech penetration. Although I have included a thorough list of bank and macro-level control variables and bank and year fixed effects, endogeneity concerns still exist due to unobservable bank, state, or time-varying characteristics that covary with both FinTech penetration and bank risk-taking. To further address such concerns, I divide my sample based on banks' relative exposure to FinTech penetration. I perform propensity score matching (PSM) for the "treatment effect"' of being more exposed to high FinTech penetration relative to other states. The outcome variable is banks' future change in risk-taking. Throughout different PSM specifications, the treatment effects are significant. In addition, I perform several sensitivity tests: I divide my sample into different year groups and groups with different regulatory scrutiny, and my baseline results are highly significant in each of these subsamples. To demonstrate that my results are not driven by sample period restrictions, I perform all the main regression analyses in an extended sample period from 2009Q1 to 2017Q4 and get similar results. Finally, through a battery of robustness tests, I find that the main results are robust to different measurements of FinTech penetration and several placebo tests.

This paper makes the following important contributions. First, while several studies have examined factors that contribute to the rapid growth of FinTech lending, limited attention has been paid to how such growth can influence the risk-taking behavior of incumbent banks. This paper is among the first to fill this gap by providing robust evidence that incumbent banks tend to increase their overall bank risk-taking and bank loan risk due to FinTech-induced competition. I identify bank charter value as a significant channel through which banks' risk-taking incentives are associated with the competition in the bank loan market. I document that bank charter value is both a mediator and a moderator between FinTech-induced competition and bank risk. Second, my paper contributes to the bank accounting literature by providing new empirical evidence on the relation between conservative loan loss provisioning and the risk discipline of banks in a FinTech competition setting. My finding suggests that community banks' forward-looking use of discretion in loan loss provisions may play a significant role in limiting their risk-increasing incentives when exposed to FinTech penetration. Third, compared to existing studies that use bank competition measurements such as interstate branch deregulation (Keeley 1990; Jayaratne and Strahan 1998; Goetz 2018), banking system concentration (de Nicoló et al. 2004; Beck et al. 2006), and a text-based measure of bank competition (Bushman et al. 2016), I use FinTech penetration as a novel measure of changes in competition. To the extent that incumbent banks cannot predict the market entry or expansion of FinTech lenders, FinTech penetration can be a more exogenous measure of changes in competition.

This paper is organized as follows. Section 2 provides the related literature and hypothesis development; section 3 describes the construction of measurements and discusses descriptive statistics; section 4 presents the main empirical analyses; section 5 presents additional and robustness tests, and section 6 offers concluding remarks.

2. Related Literature and Hypotheses Development

Existing studies have attributed the rapid growth of FinTech lending to regulatory arbitrage (e.g., Buchak et al. 2018; Tang 2019; De Roure et al. 2022), technological and informational advantages (Buchak et al. 2018; Hau et al. 2019; Frost et al. 2019; Fuster et al. 2019; Berg et al. 2020), and local banking market structure (Balyuk et al. 2020). Several recent studies examine the interplay between FinTech lenders and traditional banks. For example, Fuster et al. (2019) provide evidence implying the existence of a competitive relationship between FinTech lenders and traditional financial institutions. Tang (2019) finds that FinTech platforms are substitutes for bank lending regarding infra-marginal bank borrowers, while they are complements to bank lending in terms of small loans. Utilizing the 2011 European Banking Authority (EBA) capital exercise as an exogenous regulatory shock that induced a short-term bank lending cut, De Roure et al. (2022) find that FinTech lending increased in those German regions with a presence of affected banks and increased more if unaffected banks in those states had lower capital ratios.

There are somewhat contradictory theoretical predictions regarding the relationship between bank competition and bank risk-taking. On the one hand, studies supporting a "competition-fragility" view argue that bank competition increases bank risk (e.g., Marcus 1984; Keeley 1990; Boot and Thakor 1993; Allen and Gale 2000; Hellmann et al. 2000; Repullo 2004). Keeley (1990) shows that when a bank earns monopoly rents, its charter value is higher—bank owners have an incentive to avoid bank failure when bank charter value is high as they "cannot sell the charter"" in case of insolvency. Thus, banks with a higher charter value are more reluctant to increase risk. Similarly, Allen and Gale (2000) show that the equilibrium level of banks' risk shifting increases with respect to the number of competitors, arguing that the weakened profit margin for banks makes excessive risk-taking more attractive for bank managers and/or shareholders. Supporting this "competition-fragility" view, Keeley (1990) documents that the liberalization of interstate branching entry barriers diminishes banks' charter value, which increases bank risk.[9] Beck et al. (2006) document a negative relationship between banking system concentration and the likelihood of systemic crisis in a cross-country study. Bushman et al. (2016) use a text-based measure of bank competition and find that the increase in bank competition is associated with relaxed underwriting standards and higher future loan charge-offs. Based on this "competition-fragility" view, community banks' future change in risk-taking should be positively related to FinTech-induced competition.

On the other hand, Boyd and De Nicoló (2005) show that when bank competition is more intense, bank borrowers can benefit from lower borrowing rates. Therefore, bank borrowers' risk-shifting incentives are lower, which can decrease the riskiness of bank loans and hence the overall bank risk. Supporting such a "competition-stability" view, Jayaratne and Strahan (1998) find that the gradual removal of banking restrictions boosted the "natural selection" process in the U.S. banking industry, where more efficient, less-risky banks expand at the expense of inefficient and risky banks. In a cross-country study, De Nicoló et al. (2004) document that the "five-bank concentration ratio" is positively associated with bank risk.[10] In addition, Goetz (2018), using the state-specific process of banking deregulation, finds that the increase in market contestability is associated with greater bank stability, a lower share of nonperforming loans, and better profitability. Based on this alternative view, FinTech penetration should be negatively associated with the future change in bank risk. According to my discussion above, I state the first hypothesis as follows:

H1: FinTech penetration is positively related to future change in bank risk.

[9] Keeley (1990) measures bank risk by capital ratios and risk premiums on large deposits.
[10] De Nicoló et al. (2004) measure bank risk by the probability of failure of the five largest banks in a country.

Indeed, as implied by the "competition-fragility" view, competition can decrease bank charter value, which in turn increases the equilibrium level of bank risk-taking (Keeley 1990). To further explore whether bank charter value can be a significant channel of the relationship between FinTech penetration and bank risk-taking, I state the second hypothesis as follows:

H2: Bank charter value mediates the relationship between FinTech penetration and future change in bank risk.

In addition, I offer several cross-sectional predictions. First, consistent with the "competition-fragility" hypothesis, if managers of *ex-ante* stronger banks have greater motivations to avoid risky loans, then these banks should have a lower sensitivity of change in risk-taking in response to the change in FinTech-induced competition.

Second, banks that are more risk-disciplined can be less likely to increase risk-taking in response to FinTech penetration. Bushman and Williams (2002) document that banks that make more discretionary loan loss provisions for forward-looking purposes are associated with a better market discipline of bank risk-taking, as these banks are more transparent in financial and risk disclosures. In fact, the incurred credit loss estimation model used by banks in my sample period does not capture all future expected losses information (e.g., Beatty and Liao 2021). I conjecture that, to the extent that discretion in loan loss provisioning reflects forward-looking loan loss information beyond what the incurred loss model captures, banks with more loan loss discretion should be more disciplined in risk-taking.

Third, the relation between FinTech penetration and future change in bank risk-taking can differ due to banks' vulnerability to FinTech penetration. Banks can be less vulnerable to FinTech when they rely on more soft information, i.e., banks can build competitive advantages that are hard to replicate by FinTech's loan screening algorithms as such relationship-based businesses rely on soft information and repeated interactions (e.g., Boot and Thakor 2000; Berger et al. 2005; Botsch and Vanasco 2019). As smaller banks may rely more on relationship banking (e.g., Balyuk et al. 2020), they can be less vulnerable to FinTech penetration. Balyuk et al. (2020) provide empirical evidence supporting that large/out-of-market banks that rely more on hard information in the loan screening process are more challenged by FinTech lenders. Also, as bank managers tend to rely more on hard and standardized information in the loan screening process for homogeneous loans including consumer loans and consumer real estate loans (e.g., Liu and Ryan 1995, 2006), banks that have more homogeneous loans should be more vulnerable to FinTech penetration. Based on these cross-sectional predictions, I state the third hypothesis as follows:

H3: The positive relation between Fintech penetration and future change in bank risk is more pronounced for banks with (1) lower ex-ante charter value, (2) lower income-decreasing discretionary loan loss provisions, and (3) more reliance on hard information.

3. Empirical Methods

3.1 Measurements of FinTech Penetration

I measure FinTech penetration by the 8 or 12 quarters of cumulative successfully funded LendingClub and Prosper loans (or "FinTech loans") per capita at the state level. For robustness purposes, I also measure FinTech penetration as the cumulative FinTech loans per capita at the zip3 level, the cumulative FinTech loans to total bank loans at the state-level, and the cumulative FinTech loans to total bank consumer loans at the state-level. While several studies use the Metropolitan Statistical Area (MSA) as a relevant market for commercial banks (e.g., Berger and Hannan 1989; Goetz 2018), the lack of information in the combined dataset limits the use of such measures. The primary variable can be expressed as follows:

$$FinPen8Q_{s,t} \equiv \sum_{t=q-7}^{q} \frac{FinTech\ loans_{s,t}}{Population_{s,t}} \tag{1}$$

In equation (1), $FinPen8Q_{a,t}$ is the state-level FinTech penetration measure for each state s in year-quarter t over the past eight quarters. $FinTech\ loans_{s,t}$ is the dollar amount of additional FinTech loans successfully funded in year-quarter t in state s. $Population_{s,t}$ is the U.S. census population estimate in state s in year-quarter t. Note that I also use a $FinPen12Q_{a,t}$ measure of FinTech penetration, which uses 12 quarters of cumulative additional FinTech loans in the numerator.

For the zip3 measure, since a community bank can operate across multiple zip3 areas, FinTech penetration for community bank i in quarter t is weighted by the bank's proportion of branch deposits in each zip3 area in quarter t. According to the U.S. Postal Service, the first three digits of a Zip Code in the United States usually represent "a sectional center facility, the mail sorting and distribution center for an area."[11] Hence, the zip3 code represents a much smaller area than a state or MSA. Since I only consider community banks, I assume that such zip3 area can approximate the "local community" in which a community bank operates and attracts customers. The specific measurements are as follows:

$$FinPen8QZip_{i,t} \equiv \sum_{t=q-7}^{q} \sum_{a \in A_i} \frac{Deposit_{i,a,t}}{\sum_{a \in A_i} Deposit_{i,a,t}} \times \frac{FinTech\ loans_{a,t}}{Population_{a,t}} \tag{2}$$

Equation (2) represents the weighted average measure of zip3-level FinTech penetration. $FinPen8QZip_{i,t}$ is the zip3-level FinTech penetration measure for each bank i in year-quarter t. A_i denotes the set of zip3 areas in which bank i has branches with domestic deposits at time t. $FinTech\_Loans_{a,t}$ denotes the successful FinTech loans initiated in quarter t in a zip3 area a.

[11] Source: https://pe.usps.com/Archive/HTML/DMMArchive20050106/print/L002.htm

$Population_{a,t}$ denotes the population estimation in quarter t in zip3 area a, and $Deposit_{i,a,t}$ denotes bank i's total branch deposits in zip3 area a in quarter t. Note that for the zip3 measure, I only use LendingClub data, as Prosper does not provide zip3-level information.

3.2 Measurements of Bank Charter Value

I use the Lerner index to proxy for individual bank charter value in my main tests. I also use several bank profitability measures such as average net interest margin or return on assets over the past two years to proxy for bank charter value for robustness purposes. The Lerner index as a primary measure for bank charter value or market power has been used extensively in existing studies (e.g., Berger et al. 2009; Beck et al. 2013; Koetter et al. 2012). The Lerner index is constructed as follows:

$$L_{it} = (P_{it} - MC_{it})/P_{it} \tag{3}$$

where P_{it} is the price of total assets, proxied by the ratio of total operating revenue[12] to total assets, and MC_{it} is the marginal cost of total assets. Following prior studies (Mester 1987; Shaffer 1993; Berger et al. 2009; Beck et al. 2013), MC_{it} is estimated from the translog function:

$$\begin{aligned} LnC_{it} = \beta_0 + \beta_1 lnQ_{it} + \beta_2 (lnQ_{it})^2 + \sum_{j=1}^{3} \gamma_j \, lnW_{it}^{j} + \sum_{j=1}^{3}\sum_{k=1}^{3} \varphi_{j,k} \, lnW_{it}^{j} lnW_{it}^{k} \\ + \sum_{j=1}^{3} \delta_j lnQ_{it} lnW_{it}^{j} + Yearquarter_i \\ + \varepsilon_{it} \end{aligned} \tag{4}$$

Equation (4) is estimated using Call Report data, where C_{it} is the total operating costs, measured by total interest expense plus total non-interest expense; Q_{it} is the total outputs, proxied by total assets; W_1 is the price of physical capital, measured by fixed expenses and other

[12] The corresponding lines are total interest income plus total noninterest income from Call Reports.

noninterest expense over total assets; W_2 is the price of labor, measured by personnel expenses over total assets, and W_3 is the interest rate on deposits, measured by interest expense over total deposits. Consistent with Beck et al. (2013), I use time dummies to capture technological improvements. Moreover, homogeneity of degree one in the three input prices suggest the following restrictions: $\sum_{j=1}^{3}\gamma_j = 1,\ \sum_{j=1}^{3}\delta_j = 0$, and $\sum_{j=1}^{3}\varphi_{j,k} = 0\ for\ k = 1, 2, 3$.

After I run the translog function estimation with a restricted OLS regression, I obtain the estimated regression coefficients used to estimate the marginal cost:

$$MC_{it} = \frac{\partial C_{it}}{\partial Q_{it}} = \frac{C}{Q}\left(\widehat{\beta_1} + 2\widehat{\beta_2} lnQ_{it} + \sum_{j=1}^{2}\widehat{\delta_j}\frac{lnW_{it}^{j}}{lnW_{it}^{3}}\right) \tag{5}$$

Then, my main measure of existing bank market power $Lerner_{ex,it}$ is as follows:

$$LernerEx_{it} \equiv Mean\left(L_{i,t}, \dots, L_{i,t-7}\right) \tag{6}$$

In equation (6), $LernerEx_{it}$ proxies the *ex-ante* bank market power for bank i in quarter t.

3.3 Measurements of Bank Risk-taking

I use three measures for bank risk-taking, including the Z-score, the volatility of net interest margin (*σNIM*), and the volatility of return on assets (*σROA*). *σNIM* and *σROA* are measured by taking the standard deviations of return on assets and net interest margin over a past eight-quarters rolling window. They indicate risk-taking from bank operations (Laeven and Levine 2009; Jin et al. 2013). The Z-score is defined as the sum of return on assets and capital-asset ratio, divided by the standard deviation of ROA; it measures the number of standard deviations a bank's ROA has to fall before insolvency (e.g., Laeven and Levine 2009; Demirgüç-Kunt and Huizinga 2010; Beck et al. 2013; Kanagaretnam et al. 2014; Goetz 2018). Since the Z-score is highly skewed, I use the natural logarithm of Z-score. I then multiply the log of Z-score by -1, so that the higher the measure, the greater the bank risk-taking. In the main empirical analyses, I use the change in bank risk-taking (*ΔZ-Score*), which is calculated as the difference between the negative log of Z-score measured from the subsequent eight quarters and the negative log of Z-score measured from the previous eight quarters.

3.4 Sample Selection

I use individual loan data provided by LendingClub and Prosper Marketplace, two large peer-to-peer lending companies[13] in the U.S. LendingClub provides historical loan issuance data beginning in 2007 and Prosper Marketplace provides loan issuance data beginning in 2005. My main sample period ranges from 2009Q1 to 2015Q4. I exclude the most recent years in the main analysis because of a concern about measurement validity. The U.S. market has seen exponential growth of FinTech lenders during the past decade, as numerous small FinTech companies followed the successful pioneers such as LendingClub and Prosper. As Figure 1 suggests, despite the high growth of FinTech lending throughout the U.S., the growth of LendingClub and Prosper loans per capita has slowed since 2016-2017, implying a loss of market share and hence a decline in the validity of the constructed FinTech penetration measure using data from these two companies. To ensure that the FinTech penetration measure is indeed representative of the true competitive pressure faced by incumbent community banks, I assume 2007-2015 is a period in which the FinTech penetration measure can represent the real competitive pressure faced by incumbent community banks. However, to demonstrate that my empirical results are not sensitive to the sample period selection, I provide the main empirical results with an extended sample period from 2009 to 2017 in the sensitivity tests.[14]

[13] LendingClub data: https://www.lendingclub.com/statistics/additional-statistics; Prosper data: https://www.prosper.com/investor/marketplace. Member log-in is needed to access both datasets.
[14] Note that although the raw data contains data from 2007 to 2019, lagging and forwarding requirements of my empirical specification make the earliest and latest two years unavailable for regression analyses.

In addition, I obtain bank financial data from Call Reports, branch-level data from the FDIC's Summary of Deposits (SOD) database, and community bank information from FDIC Community Banking Reference Data. State and Zip Code level population data are obtained from the United States Census Bureau, the quarterly Gross Domestic Product (GDP) by state and Zip Code is obtained from the U.S. Bureau of Economic Analysis, unemployment rate data is from U.S. Bureau of Labor Statistics, and house price index is from Federal Housing Finance Agency.

When constructing the sample, I start with a raw Call Report dataset from 2007Q1 to 2019Q4 with 361,445 bank-quarters. After merging the call data with the FinTech penetration measurement data, I exclude observations with no Zip Codes, no FDIC certificate number, and invalid state name, and I obtain a sample of 348,481 bank-quarters. Then, I exclude observations with negative or missing total assets, total interest expense, salary and employee expenses, fixed expense, non-interest expenses, loan loss provisions, total deposits, and total loans, and retain a sample of 327,365 bank-quarters. After omitting noncommunity bank observations, restricting the sample to banks in the contiguous U.S., and requiring at least 12 consecutive quarters of data, the sample size is reduced to 302,374. To mitigate concerns about mergers and acquisitions, I remove observations with a quarterly asset growth rate higher than 20%, resulting in 241,197 community bank-quarters. In addition, after generating lagged and forwarded variables, removing missing values, and restricting the sample period, I obtain a sample of 93,882 community bank-quarters for 5,560 unique community banks in the period 2009Q1-2015Q4. All continuous variables are winsorized at the 1st and 99th percentiles. The detailed sample selection process and final sample information are demonstrated in Table 3.

[Insert Table 1 and Table 2 Here]

3.5 Descriptive Statistics

There are significant cross-regional and cross-time variations of FinTech penetration, and such penetration in local bank markets is nontrivial. Summarized from the raw sample, figures 1A-1C show the scatter plots of FinTech penetration on the state-level over the years for selected states. These figures demonstrate that there are significant variations of state-level FinTech penetration both over the years and across U.S. states. Figure 2 provides two snapshots of the heatmap of the state-level FinTech penetration for 2009Q4 and 2015Q4, and shows that the variation of FinTech penetration across states at any specific point of time is large and significant. Tables 3A and 3B demonstrate that the ratios of three-year cumulative FinTech loans to the local population and to total bank consumer loans are nontrivial. For example, as Table 3B suggests, the ratio of three-year cumulative FinTech loans to state-level community bank consumer loans increases from an average of 2% in 2010 to an average of 71% in 2018.

[Insert Figure 1, Figure 2, Table 3 Here]

The descriptive statistics of the main variables are shown in Table 4, and all variables are defined in Appendix 1. As Table 4 indicates, the mean *Z-Score* is -5.04 and the median is -5.117. Bushman et al. (2016), using a much earlier sample period of 1996-2012, report a mean of -2.84 and a median of 2.46. I conjecture that the results are comparable considering that the average Z-Score for U.S. banks is trending downward (Figure 3A). In addition, the mean of the *ex-ante* Lerner index (*LERNER*) is 0.23, which is slightly higher than the measure of U.S. banks' Lerner index of 0.17 in Beck et al. (2013) but close to the international measure of 0.22 in Berger et al. (2009). The small discrepancies are possibly due to sample and sample period differences. I plot the average Lerner index over time for selected states and for the total U.S. in Figure 3B. The plots show that the Lerner index for U.S. community banks trends upward in the sample period, which partially explains why the mean Lerner index reported in Beck et al. (2013) is lower. Indeed, figures 3A and 3B univariately demonstrate that the overall U.S. community banks' charter value is negatively associated the overall level of bank risk-taking.

In addition, I validate the bank financials by comparing their means and medians to Bushman et al. (2016), who also use bank-quarter level data. As a result, summary statistics in this paper are generally comparable to the data summary of Bushman et al. (2016), with some explainable differences.[15] Finally, Table 4 Panel B reports the Pearson correlation table for the key variables used in this study; for example, the univariate analysis demonstrates that all proxies of FinTech penetration are positively correlated with changes in future risk-taking.

[Insert Table 4 and Figure 3 Here]

4. Empirical Results

4.1 Baseline Relationship between FinTech Penetration and Change in Bank Risk

The following empirical model is used to examine my main hypothesis:

$$\Delta Risk = \beta_0 + \beta_1 FinPen + \Gamma Controls + \Theta\Delta Controls + Bank\ FE + Year\ FE + \varepsilon_{i,t} \qquad (7)$$

The dependent variable is measured by *ΔZ-Score, ΔσROA*, or *ΔσNIM* to proxy for changes in bank risk-taking from the past two years to the subsequent two years. *FinPen* is a measure for FinTech penetration, which is proxied by either two-year state-level cumulative FinTech loans per capita (*FinPen8Q*) or three-year state-level cumulative FinTech loans per capita (*FinPen12Q*). *Controls* is a vector of control variables. *ΔControls* represents the first differencing of these control variables by taking the difference between the subsequent two-year average and the prior two-year average. Consistent with Laeven and Levine (2009) and Houston et al. (2010), the control variables (and their first differencing) include: the two-year rolling average of the natural logarithm of total

[15] For the sake of brevity, the detailed comparisons are not discussed in this paper.

assets (*SIZE*, *ΔSIZE*); the two-year rolling average of bank loans scaled by lagged total assets (*LOAN*, *ΔLOAN*); the two-year rolling average of LLP scaled by lagged total loans and leases, net of unearned income and allowance for losses (*LLP*, *ΔLLP*); the two-year rolling average of the natural logarithm of state GDP (*LnGDP*, *ΔLnGDP);* the two-year rolling average of the state housing price index (*HPI*, *ΔHPI);* the two-year rolling of average state unemployment rate (*UNEMP*, *ΔUNEMP*); the two-year rolling average of the Tier 1 ratio (*T1R*, *ΔT1R*), and the revenue growth rate measured in two-year rolling windows (*REVG*). In addition, I use bank fixed effects to control for any unobservable differences across banks and year fixed effects to control for any unobservable time-varying confounding effects. Standard errors are clustered by individual banks. A positive and significant β_1 is consistent with the "competition-fragility" view, indicating that, overall, FinTech penetration is positively associated with the future change in bank risk-taking.

Table 5 reports the empirical results. As expected, FinTech penetration is positively associated with the change in future bank risk. Models 1 and 2 document that FinTech penetration measured by *FinPen8Q* and *FinPen12Q*, respectively, are positively related to the increase in bank risk proxied by *ΔZ-Score* (coef. = 0.0056 and 0.0046, t = 4.91 and 4.72). More specifically, a one-standard-deviation increase in FinTech penetration is associated with a 44%-47% increase in the change in bank risk-taking. Models 3 and 4 show the FinTech penetration is positively associated with change in bank risk proxied by *ΔσNIM* (coef. = 0.0181 and 0.0147, t = 3.44 and 3.33). Models 5 and 6 document that the FinTech penetration is positively associated with the change in bank risk proxied by *ΔσROA* (coef. = 0.0762 and 0.0592, t = 3.50 and 3.20). In sum, the findings in Table 5 suggest that, overall, changes in future bank risk-taking are positively related to banks' current exposure to FinTech penetration, which is consistent with the "competition-fragility" view.

[Insert Table 5 Here]

4.2 The Mediating Role of Bank Charter Value

One of the key channels through which FinTech penetration can be related to increased bank risk-taking is bank charter value. As Keeley (1990) suggests, a bank's incentive to take risks increases as its charter value falls. In this section, I conduct several path analyses with the change in bank charter value as the main mediating variable. The source and outcome variables refer to FinTech penetration (*FinPen8Q* or *FinPen12Q*) and change in bank risk-taking. Specifically, I estimated the following structural equation models:

$$\Delta LERNER = b_0 + b_1 FinPen + CONTROLS + \varepsilon \tag{8}$$

$$\Delta ZScore = c_0 + c_1 FinPen + c_2 \Delta LERNER + CONTROLS + \varepsilon \tag{9}$$

where the set of control variables in equation (8) includes bank-level control variables and macroeconomic variables, and the set of control variables in equation (9) includes all the control variables used in my baseline regressions specified in equation (7). The path coefficient c_1 is the magnitude of the direct path and $b_1 * c_2$ is the magnitude of the indirect path. Table 6 reports the results of path analyses. In Table 5 Panel A, I use *ΔLerner* as the single mediator between FinTech penetration and change in bank risk. I document that the mediated path is significant (coef. = 0.0308 and 0.0241, t = 5.46 and 4.81) and economically meaningful (mediated path percentage = 27% and 20%). In addition, I perform multi-mediator analysis by incorporating bank profitability, bank capital, and bank charter value as multi-mediators for the relationship between FinTech penetration and change in bank risk. Table 6 Panel B shows the results of the multi-mediator analysis: by incorporating more mediators, the total mediated path becomes stronger (coef. = 0.1203 and 0.0996, t = 14.38 and 13.56) and more economically meaningful (mediated path percentage = 65% and 62%). As bank profitability and the level of bank capital are important determinants of bank charter value, the results from multi-mediator analysis corroborates my conjecture that bank charter value is one important channel connecting FinTech penetration and an increase in bank risk-taking.

[Insert Table 6 Here]

4.3 The Moderating Role of Bank Charter Value

As discussed in the hypothesis 3, banks with greater *ex-ante* market power can be more reluctant to take extra risks. I use the following model to test the role of existing bank market power on FinTech penetration and change in bank risk-taking:

$$\Delta Risk = \beta_0 + \beta_1 FinPen + \beta_2\, CharterValue + \beta_3 FinPen \times CharterValue + \Gamma Controls + \Theta\Delta Controls + Bank\ FE + Year\ FE + \varepsilon_{i,t} \quad (10)$$

where *CharterValue* stands for bank charter value and is proxied by several variables, including the *ex-ante* two-year rolling average of the Lerner index (*LERNER*), a categorical variable (*LERNERQ*) which equals the quartile of *LERNER* by year and quarter, and two alternative proxies of charter value related to *ex-ante* bank profitability, namely the *ex-ante* return on assets (*ROA_EX*) and *ex-ante* net interest margin (*NIM_EX*). All other variables are measured in the same ways as in equation (7). The main coefficient of interest in equation (10) is β_3. I expect a negative and significant β_3, which suggests that banks with higher existing charter value are less likely to increase bank risk in response to FinTech penetration.

Table 7 presents the regression results of the moderating effects of bank charter value. From Model 1 to Model 4, I use the quartile measure of Lerner index (*LERNERQ*). I find that in each of the four models, the *ex-ante* bank charter value negatively moderates the relationship between FinTech penetration and future change in bank risk-taking. I find that the coefficients of the interaction term (*FinPen* × *LERNERQ*) are increasingly significant through the *LERNER* quartiles, regardless of the choice of proxies for FinTech penetration and the choice of proxies for change in bank risk-taking. For example, in Model 1, the relationship between FinTech penetration and future change in risk-taking is significantly attenuated when the quartile of *LERNER* increases from *LERNERQ1* to *LERNERQ4*. A test of partial effects of FinTech penetration (bottom part of Table 7) suggests that when the *LERNER* quartile equals 1, the partial effect of FinTech penetration is 0.0074 (t = 5.19), while such effect attenuates and finally is insignificant (t = 1.25) when the *LERNER* quartile equals 4.

Models 5 and 6 document a similar negative moderating effect of *ex-ante* bank charter value, when a continuous measure of the Lerner index is used (coef. = -0.0113 and -0.0099, t = -2.49 and -2.48). In models 7 and 8, I use two *ex-ante* continuous bank profitability measures (*ROA_EXI* and *NIM_EX*) as a robustness check and find similar results (coef. = -1.6405 and -0.9168, t = -4.17 and -2.91), suggesting that banks with narrower profit margins are more likely to increase risk-taking when exposed to FinTech penetration. In sum, results from Table 7 imply that *ex-ante* weaker community banks tend to have much stronger risk-increasing incentives when they face FinTech-induced competition.

[Insert Table 7 here]

4.4 The Moderating Role of Discretionary Loan Loss Provisions

As discussed in the hypothesis development section, I also expect a negative moderating role of community banks' income-decreasing discretionary loan loss provisions (*DLLP*) for the relationship between FinTech penetration and change in bank risk. To the extent that *DLLP* reflects timely forward-looking credit loss information, banks with higher *DLLP* should be more risk-disciplined (Bushman and Williams, 2002). I measure *DLLP* as the residual from the regression of *LLP* using equation (11), which is a modified version of the Kanagaretnam et al. (2010) model, adding the effect of the sign of change in nonperforming loans (*DΔNPL*) following Basu et al. (2020). The residual captures a bank's magnitude of income decreasing $DLLP$.

$$LLP_{it} = \alpha_0 + \alpha_1 LLA_{it-1} + \alpha_2 NPL_{it-1} + \alpha_3 LOAN_{it} + \alpha_4 \Delta LOAN_{it} + \alpha_5 NCO_{it} + \alpha_6 \Delta NPL_{it} \times D\Delta NPL_{it} + \alpha_6 \Delta NPL_{it} + \alpha_6 D\Delta NPL_{it} + \alpha_7 \Delta lnGDP_{it} + \alpha_8 \Delta UNEMP_{it} + \alpha_9 \Delta HPI_{it} + Bank\ FE + Year\ FE + \varepsilon_{it} \quad (11)$$

where *LLP* is loan loss provision scaled by lagged total loans; *LLA* is allowance for loan losses scaled by total loans, and *NPL* is nonperforming loans scaled by lagged total loans. *DΔNPL* is a dummy variable that equals 1 if the sign of *ΔNPL* is negative, and 0 otherwise. All other variables are defined in the same way as in equation (7). *DLLP* is constructed as the residuals from the regression Model (11). In addition, I use an indicator variable *POS_DLLP* that equals 1 if the *DLLP* is positive (i.e., income-decreasing), and 0 otherwise.

Table 8 reports the moderating effects of *DLLP*. Models 1 and 2 show the results of regressing change in bank risk (proxied by *ΔZ-Score*) on the interaction between FinTech penetration (proxied by *FinPen8Q* and *FinPen12Q*). In both models, the interaction term is negative and significant (coef. = -0.5907 and -0.5181, t = -2.30 and -2.28), suggesting that banks with more income-decreasing DLLP in the current period are less likely to increase risk-taking in response to FinTech penetration. Models 3 and 4 regress change in bank risk on the interaction between FinTech penetration and the dummy variable *POS_DLLP*. These two models document that banks that have positive *DLLP* in the current period are significantly less likely to increase risk in response to FinTech penetration (coef. = -0.0015 and -0.0013, t = -2.85 and -2.83). Lastly, models 5 and 6 use the interaction between FinTech penetration and current period loan loss provision scaled by lagged total loans (*LLP*) and demonstrate that banks with more *LLP* in the current period are significantly less likely to increase risk in response to FinTech penetration (coef. = -0.8219 and -0.7291, t = -3.21 and -3.23).

[Insert Table 8 Here]

4.5 The Moderating Role of Banks' Tendency to Use Hard Versus Soft Information

In this section, I examine whether incumbent community banks' incentive to increase risk-taking depends on their reliance on hard versus soft information during banks' loan screening process. As FinTech lenders can have a competitive advantage in utilizing machine learning and big data algorithms when screening for loans, they can be good at hardening information (e.g., Balyuk et al. 2020; Boot et al. 2021). Hence, banks that rely more on hard information in the loan screening process can be more challenged by FinTech lenders. Due to such increased vulnerability to the FinTech threat, I conjecture that banks that rely more on hard information are more likely to be negatively affected by FinTech penetration, so that they are more likely to increase risk-taking when facing increased FinTech penetration.

I use three proxies for banks' likelihood of using more hard information than soft information in the loan screening process, including bank size (*SIZE*), the proportion of homogeneous bank loans (*HomoLoans%*), and the proportion of residential real estate loans (*ResReLoans%*). Balyuk et al. (2020) document that large/out-of-market banks tend to be more challenged by FinTech lenders than small/in-market banks. Hard information is easier to transmit, favoring larger banks (Boot et al. 2021). Thus, to the extent that larger banks can have more out-of-market loans that require more hard information in the loan screening process, bank size can be a plausible proxy for banks' likelihood to use more hard than soft information. In addition, relationship banking can be costly for smaller and homogeneous consumer loans. Banks with more homogeneous loans, such as consumer loans and consumer real estate loans, should be more likely to use more hard information in their loan screening process.

I present the regression results in Table 9. In models 1 and 2, I regress bank risk on the interaction between FinTech penetration and bank size. As expected, I document that larger banks are more incentivized to increase risk-taking in response to FinTech penetration (coef. = 0.0012 and 0.0011, t = 2.48 and 2.43). Models 3 and 4 use the percentage of residential real estate loans (the largest homogeneous loan type for community banks as per Table 4) as the proxy for banks' tendency to use more hard information. I find that banks with more residential real estate loans are more likely to increase risk-taking facing FinTech penetration (coef. = 0.0081 and 0.0070, t = 3.22 and 3.19). Last, models 5 and 6 document that when the proportion of homogeneous loans is used as the proxy for banks' tendency to use hard information, the moderating effect is still positive and significant (coef. = 0.0051 and 0.0044, t = 2.09 and 2.06).

[Insert Table 9 Here]

4.6 FinTech Penetration and Bank Loan Risk

The bank risk measures I use above capture overall bank risk. However, these measures are closely related to the volatility of bank earnings and can be influenced by confounding factors such as economic crisis, natural disasters, or political uncertainties. To better identify banks' own incentives to increase risk-taking, I examine the relationship between FinTech penetration and bank loan risk. In the first group of tests, I examine the effect of FinTech penetration on the relationship between current period bank loan growth and future period bank loan charge-offs. Following Bushman et al. (2016), who document that bank competition positively moderates the relationship between the bank's current period loan growth and future loan charge-offs, I estimate the following model:

$$NCO_{12m\ or\ 24m} = \beta_0 + \beta_1 LoanGrowth_t + \beta_2 FinPen_t + \beta_3 LoanGrowth_t \times FinPen_t + \delta CONTROLS + BankFE + YearFE + \varepsilon_{i,t} \quad (12)$$

where *NCO* is the total net loan charge-offs divided by lagged total loans at time *t* over the next four or eight quarters. *LoanGrowth* refers to the percentage growth of total loans in the quarter. Control variables include the interaction of *FinPen* with the proportion of each type of loans, the interaction of *LoanGrowth* with the proportion of each type of loans, current period and past period change in nonperforming loans (ΔNPL_t, ΔNPL_{t-1}, ΔNPL_{t-2}), bank size (*SIZE*), the Tier 1 capital ratio (*T1R*), return on assets (*ROA*), and bank and year fixed effects.

Table 10 Panel A reports the regression results of estimating equation (12) with four different FinTech penetration measures (two state-level measures and two zip3-level measures). Consistent with my prediction, I document that β_3 is positive and significant in each of the eight model specifications with p-values < 0.01. For example, using *FinPen8Q* as the proxy for FinTech penetration, the portion of a bank's current loans that are charged off in future 12-month periods and in future 24-month periods is significantly increasing with the bank's exposure to FinTech penetration (coef. = 0.0164 and 0.0269, t = 2.69 and 2.99). This finding implies that banks, when facing greater FinTech penetration, tend to increase the riskiness of newly originated loans.

In the second group of tests, I directly test whether FinTech penetration increases future change in bank loan risk with two proxies for change in bank loan risk, namely, the change of the two-year rolling average of *LLP* scaled by lagged total loans (*ΔLLP*) and the change of the two-year rolling average of *NCO* scaled by lagged total loans (*ΔNCO*). Loan loss provisions can reflect a bank's estimation of current and future credit losses (e.g., Beatty and Liao 2014; Khan and Ozel 2016). A higher *ΔLLP* can indicate that banks are more pessimistic about the riskiness of their current loans. However, since loan loss provisions are subject to management discretion that serves earnings or capital management purposes (e.g., Collins et al. 1995; Ahamed et al. 1999; Beatty et al. 1995, 2002; Kanagaretnam et al. 2004), I also use *ΔNCO* as a proxy for future change in loan risk, as loan charge-offs are subject to less accounting discretion.

Table 10 Panel B presents the results of regressing *ΔLLP* and *ΔNCO* on FinTech penetration with the moderating effect of bank charter value. In models 5 and 6, I use the quartile measure of the Lerner index (*LERNERQ*) and find that the coefficients of the interaction term (*FinPen* × *LERNERQ*) are increasingly significant as the Lerner index quartile increases. The partial effects of FinTech penetration are significantly positive only in the lowest two quartiles of Lerner index. In models 3 and 4, I use a dummy variable that equals 1 if the bank's charter value is above the median of bank charter value for that quarter. I find that FinTech penetration is positively associated with change in bank loan risk only when bank charter value is low. In models 1 and 2, I use a continuous measure of the Lerner index and document similar results. Finally, in models 7 and 8, I use *ΔNCO* as the dependent variable and find similar results.

[Insert Table 10 Here]

5. Additional Tests

5.1 Propensity Score Matching

In this section, I relax the assumption that FinTech penetration is linearly related to future change in bank risk and employ nonparametric methods. I conduct propensity score matching to explore the "treatment effect" of a bank-quarter being highly penetrated by FinTech lenders versus being minimally penetrated by FinTech lenders. To mitigate the confounding factor of time (i.e., FinTech penetration increases monotonically over time), I divide the sample based on banks' relative exposure to FinTech penetration for each quarter. I label a bank-quarter "treated" if it is in the states with the highest quartile of FinTech penetration in each quarter. A bank is labelled “untreated” if it is in the lowest quartile. I perform propensity score matching for the “treatment effect” of being exposed to high FinTech penetration relative to other states. The outcome variable is future change in bank risk (proxied by *ΔZ-Score*).

I report the results in Table 11. In Model 1, I implement the nearest-neighbor matching with N=1 for the observations with bank-level and macro-level control variables and Abadie and Imbens (2006) robust standard errors. The average treatment effect for the treated (ATT) for having high FinTech penetration is positive and significant (coef. = 0.1530, t = 6.07). In Model 2 I employ the Epanechnikov kernel matching with a bootstrapped standard error and find a positive and significant treatment effect of having high FinTech penetration (coef. = 0.0521, t = 3.88). In models 3 and 4, I repeat the PSM tests but with 1-on-4 matches (N=4) and find positive and significant treatment effects of having high FinTech penetration.

[Insert Table 11 Here]

5.2 Sensitivity Checks

I perform a battery of sensitivity checks with split samples to address possible endogeneity issues. Table 12 reports these results. The influence of FinTech penetration on change in bank risk can be confounded by different time periods; for example, the relationship may be different in years following the financial crisis than in later years. To address this concern, I split the sample into different groups in models 1 to 4 of Table 12 Panel A with 2011 or 2012 as the cutoff year. I find that the coefficients of FinTech penetration are positive and significant in all subsamples. In models 5 to 8, I demonstrate that the positive influence of FinTech penetration on change in bank risk is not driven by different levels of bank regulatory scrutiny. In Table 12 Panel B, I examine whether the regression results still hold when using an extended sample period from 2009 to 2017. For all 10 models, I find that the main effects are as expected.

[Insert Table 12 Here]

5.3 Alternative Measurements of FinTech Penetration

In this section, I examine the empirical results using three alternative groups of measurements of FinTech Penetration. First, I use the two-year or three-year rolling cumulative FinTech loans per capita on the zip3 level (*FinPen8QZip* and *FinPen12QZip*). The detailed construction of the measures is discussed in section 3.1. I rerun the main analyses and present the results in Table 13 Panel A, which demonstrates that all the main results are confirmed. Second, I use the two-year or three-year rolling cumulative FinTech loans divided by total bank loans on the state level (*FinPen_8qLoans* and *FinPen_12qLoans*). I repeat the base line regressions in Table 13 Panel B and get similar results. Third, I use the two-year or three-year rolling cumulative FinTech loans divided by total bank consumer loans on the state level (*FinPen_8qCSLoans* and *FinPen_12qCSLoans*). I repeat the base line regressions in Table 13 Panel B and get similar results.

[Insert Table 13 Here]

5.4 Placebo Tests

Finally, I conduct several placebo tests by checking whether banks' current exposure to FinTech penetration (*FinPen8Q* and *FinPen8Q*) is related to past changes in bank risk. I measure past changes in bank risk as the *ex-ante* two-year change in bank risk proxied by change in the natural logarithm of negative Z-score (*ΔZ-Score_p*), change in volatility of ROA (*ΔσROA_p*), and change in volatility of NIM (*ΔσNIM_p*). If FinTech penetration variables present exogenous shocks to community banks' competitive environment, then they should not be "expected" by banks *ex-ante*. As Table 13 Panel C shows, the coefficients of all FinTech penetration variables in the placebo tests are insignificant, corroborating the main empirical insights of this study.

6. Conclusions

This paper examines whether and how FinTech-induced competition influences incumbent community banks' future change in risk-taking. I address these questions by analyzing a sample of 5,458 U.S. community banks during 2009Q1 -2015Q4, with FinTech loan data from two primary FinTech lenders in the U.S.: LendingClub and Prosper Marketplace. In summary, I find that current exposure to FinTech penetration is associated with a subsequent increase in overall bank risk and bank loan risk. Path analysis shows that change in bank charter value is a significant and economically meaningful mediator between FinTech penetration and change in bank risk. Cross-sectional analysis demonstrates that the relationship between FinTech penetration and future increase in bank-risk is more pronounced when banks' *ex-ante* charter value is lower, when banks have less discretionary loan loss provisions, and when banks are more likely to use hard information in their loan screening process.

Overall, this paper supports the "competition-fragility" view: when banks face increased FinTech penetration, their charter values are negatively affected, resulting in greater equilibrium levels of bank risk-taking. Banks with higher *ex-ante* charter values are more incentivized to protect their charter values and to avoid uncertainties, so that their risk-increasing incentives are less sensitive to FinTech penetration. This paper extends the bank competition and bank risk-taking literature by introducing FinTech penetration as a unique measure, and empirically explores the moderating role of *ex-ante* bank charter value, bank accounting discretions, and the type of information used in banks' loan screening process. This paper also contributes to the bank accounting literature by providing new empirical evidence on the relation between discretionary loan loss provisioning and the risk discipline of banks in a FinTech competition setting.

This study is subject to several limitations. First, although LendingClub and Prosper are the two large FinTech lending companies in the U.S., their combined amount of loans is far from approximating the total amount of FinTech loans across the U.S. Hence, the assumption made by this paper is that FinTech loans from other lending platforms penetrate U.S. geographical areas proportionally with LendingClub and Prosper during the sample period. I addressed this concern by cutting off the sample period by the end of 2015; however, there may be unobservable concerns that other FinTech lenders' early development may collectively confound the proxy for FinTech penetration. Second, although the first differencing estimators are employed and future risk-taking is used in dependent variables, the associations between FinTech penetration and changes in future risk-taking may not result from underlying causal relations.

References

Abadie, A., and G. W. Imbens. 2006. Large sample properties of matching estimators for average treatment effects. *Econometrica* 74 (1): 235–267.

Ahmed, A. S., C. Takeda, and S. Thomas. 1999. Bank loan loss provisions: A reexamination of capital management, earnings management and signaling effects. *Journal of Accounting and Economics* 28 (1): 1–25.

Allen, F., and D. Gale. 2000. *Comparing financial systems*. Cambridge, MA: MIT Press.

Anginer, D., A. Demirgüç-Kunt, and M. Zhu. 2014. How does competition affect bank systemic risk? *Journal of Financial Intermediation* 23 (1): 1–26.

Balyuk, T., A. N. Berger, and J. Hackney. 2020. What is fueling FinTech lending? The role of banking market structure. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3633907

Basu, S., J. Vitanza, and W. Wang. 2020. Asymmetric loan loss provision models. *Journal of Accounting and Economics* 70 (2–3): 101359.

Beatty, A., S. L. Chamberlain, and J. Magliolo. 1995. Managing financial reports of commercial banks: The influence of taxes, regulatory capital, and earnings. *Journal of Accounting Research* 33 (2): 231.

Beatty, A. L., B. Ke, and K. R. Petroni. 2002. Earnings management to avoid earnings declines across publicly and privately held banks. *The Accounting Review* 77 (3): 547–570.

Beatty, A., and S. Liao. 2014. Financial accounting in the banking industry: A review of the empirical literature. *Journal of Accounting and Economics* 58 (2–3): 339–383.

Beatty, A., and S. Liao. 2021. What do analysts' provision forecasts tell us about expected credit loss recognition? *The Accounting Review* 96 (1): 1–21.

Beck, T., O. De Jonghe, and G. Schepens. 2013. Bank competition and stability: Cross-country heterogeneity. *Journal of Financial Intermediation* 22 (2): 218–244.

Beck, T., A. Demirgüç-Kunt, and R. Levine. 2006. Bank concentration, competition, and crises: First results. *Journal of Banking & Finance* 30 (5): 1581–1603.

Berg, T., V. Burg, A. Gombović, and M. Puri. 2020. On the rise of FinTechs: Credit scoring using digital footprints. A. Karolyi, ed. *The Review of Financial Studies* 33 (7): 2845–2897.

Berger, A. N., and T. H. Hannan. 1989. The price-concentration relationship in banking. *The Review of Economics and Statistics* 71 (2): 291-299.

Berger, A. N., L. F. Klapper, and R. Turk-Ariss. 2009. Bank competition and financial stability. *Journal of Financial Services Research* 35: 99–118.

Berger, A. N., N. H. Miller, M. A. Petersen, R. G. Rajan, and J. C. Stein. 2005. Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial Economics* 76 (2): 237–269.

Boot, A., P. Hoffmann, L. Laeven, and L. Ratnovski. 2021. Fintech: What's old, what's new? *Journal of Financial Stability* 53: 100836.

Boot, A. W. A., and A. V. Thakor. 1993. Self-interested bank regulation. *American Economic Review* 83 (2): 206–212.

———. 2000. Can relationship banking survive competition? *The Journal of Finance* 55 (2): 679–713.

Botsch, M., and V. Vanasco. 2019. Learning by lending. *Journal of Financial Intermediation* 37: 1–14.

Boyd, J. H., and G. De Nicoló. 2005. The theory of bank risk taking and competition revisited. *The Journal of Finance* 60 (3): 1329–1343.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru. 2018. Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics* 130 (3): 453–483.

Bushman, R. M., B. E. Hendricks, and C. D. Williams. 2016. Bank competition: Measurement, decision-making, and risk-taking. *Journal of Accounting Research* 54 (3): 777–826.

Bushman, R. M., and C. D. Williams. 2012. Accounting discretion, loan loss provisioning, and discipline of banks' risk-taking. *Journal of Accounting and Economics* 54 (1): 1–18.

Claessens, S., and N. van Horen. 2014. The impact of the global financial crisis on banking globalization. *IMF Economic Review* 63(4): 868-918.

Claessens, S., and L. Laeven. 2004. Competition in the financial sector and growth: A cross-country perspective. In *Financial Development and Economic Growth*, C. A. E. Goodhart, ed., 66–105. London: Palgrave Macmillan UK.

Collins, J. H., D. A. Shackelford, and J. M. Wahlen. 1995. Bank differences in the coordination of regulatory capital, earnings, and taxes. *Journal of Accounting Research* 33 (2): 263-291.

De Roure, C., L. Pelizzon, and A. Thakor. 2022. P2P lenders versus banks: Cream skimming or bottom fishing? A. Ellul, ed. *The Review of Corporate Finance Studies* 11 (2): 213–262.

Demirgüç-Kunt, A., and H. Huizinga. 2010. Bank activity and funding strategies: The impact on risk and returns. *Journal of Financial Economics* 98 (3): 626–650.

Frost, J., L. Gambacorta, Y. Huang, H. S. Shin, and P. Zbinden. 2019. BigTech and the changing structure of financial intermediation. *Economic Policy* 34 (100): 761–799.

Fuster, A., M. Plosser, P. Schnabl, and J. Vickery. 2019. The role of technology in mortgage lending. *The Review of Financial Studies* 32 (5): 1854–1899.

Goetz, M. R. 2018. Competition and bank stability. *Journal of Financial Intermediation* 35: 57–69.

Hau, H., Y. Huang, H. Shan, and Z. Sheng. 2019. How FinTech enters China's credit market. *AEA Papers and Proceedings* 109: 60–64.

Hellmann, T. F., K. C. Murdock, and J. E. Stiglitz. 2000. Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review* 90 (1): 147–165.

Houston, J. F., C. Lin, P. Lin, and Y. Ma. 2010. Creditor rights, information sharing, and bank risk taking. *Journal of Financial Economics* 96 (3): 485–512.

Jayaratne, J., and P. E. Strahan. 1998. Entry restrictions, industry evolution, and dynamic efficiency: Evidence from commercial banking. *The Journal of Law and Economics* 41 (1): 239–274.

Kanagaretnam, K., G. V. Krishnan, and G. J. Lobo. 2010. An empirical analysis of auditor independence in the banking industry. *The Accounting Review* 85 (6): 2011–2046.

Kanagaretnam, K., C. Y. Lim, and G. J. Lobo. 2014. Influence of national culture on accounting conservatism and risk-taking in the banking industry. *The Accounting Review* 89 (3): 1115–1149.

Kanagaretnam, K., G. J. Lobo, and D. H. Yang. 2004. Joint tests of signaling and income smoothing through bank loan loss provisions. *Contemporary Accounting Research* 21 (4): 843–884.

Keeley, M. C. 1990. Deposit insurance, risk, and market power in banking. *American Economic Review* 80 (5): 1183–1200.

Laeven, L., and R. Levine. 2009. Bank governance, regulation and risk taking. *Journal of Financial Economics* 93 (2): 259–275.

Liu, C. C., and S. G. Ryan. 1995. The effect of bank loan portfolio composition on the market reaction to and anticipation of loan loss provisions. *Journal of Accounting Research* 33 (1): 77.

———. 2006. Income smoothing over the business cycle: Changes in banks' coordinated management of provisions for loan losses and loan charge-offs from the pre-1990 bust to the 1990s boom. *The Accounting Review* 81 (2): 421–441.

Marcus, A. J. 1984. Deregulation and bank financial policy. *Journal of Banking & Finance* 8 (4): 557–565.

Mester, L. J. 1987. A multiproduct cost study of savings and loans. *The Journal of Finance* 42 (2): 423–445.

Repullo, R. 2004. Capital requirements, market power, and risk-taking in banking. *Journal of Financial Intermediation* 13 (2): 156–182.

Shaffer, S. 1983. Non-structural measures of competition. *Economics Letters* 12 (3–4): 349–353.

Tang, H. 2019. Peer-to-peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies* 32 (5): 1900–1938.

Appendix 1: Key Variable Definitions

Variable Name		Variable Description and Data Sources
Bank Risk-taking Related Variables: ***(Source: Call Report Data)***		
σROA	=	Standard deviation of bank-quarter return on assets (ROA) calculated in two-year windows.
ΔσROA	=	The change of *σROA* from prior two years to subsequent two years.
σNIM	=	Standard deviation of bank-quarter net interest margin (NIM) calculated in two-year windows.
ΔσNIM	=	The change of *σNIM* from prior two years to subsequent two years.
Z-Score	=	$ZScore_{it} = (-1) * \log\left[\frac{Mean(ROA)_{i,t+1\ to\ t+8} + Mean(CAR)_{i,t+1\ to\ t+8}}{\sigma ROA_{i,t+1\ to\ t+8}}\right]$ where: - $Mean(ROA)_{i,t+1\ to\ t+8}$: mean of bank i's *ROA* - $Mean(CAR)_{i,t+1\ to\ t+8}$: mean of bank i's capital to assets ratio - $RoaStdPost_{i,t+1\ to\ t+8}$: Standard deviation of bank i's *ROA* - *t+1 to t+8*: the time window over which *Z-Score* is calculated
ΔZ-Score	=	Change in bank *Z-Score* from prior two years to subsequent two years.
FinTech Penetration (*FinPen*) Measurements: ***(Source: LendingClub, Prosper)***		
FinPen8Q	=	The primary measure of *FinPen*: The two-year cumulative state-quarter level successfully funded FinTech loans over state-quarter population.
FinPen12Q	=	The primary measure of *FinPen*: The three-year cumulative state-quarter level successfully funded FinTech loans over state-quarter population. (Source: LendingClub, Prosper).
FinPen8QZip	=	The two-year cumulative zip3-quarter level successfully funded FinTech loans over zip3-quarter population. For banks operating across zip3 areas, the measure is weighted by banks' zip3-level branch deposits.
FinPen12QZip	=	The three-year cumulative zip3-quarter level successfully funded FinTech loans over zip3-quarter population. For banks operating across zip3 areas, the measure is weighted by banks' zip3-level branch deposits.
FinPenCSLoans	=	The two or three-year cumulative state-quarter level successfully funded FinTech loans over state-quarter community bank consumer loans.
FinPenLoans	=	The two or three-year cumulative state-quarter level successfully funded FinTech loans over state-quarter total community bank loans.
Bank-level Variables[16]: ***(Source: Call Report Data)***		
LERNER	=	The measure of two-year rolling average of *ex-ante* bank charter value. The construction of the Lerner index is discussed in detail in section 3.2.
LERNER_Q	=	A categorical variable generated from the quartiles of *LERNER* in each quarter.
LERNER_HIGH	=	A dummy variable that equals 1 if *LERNER* is above the median of *LERNER* in each quarter

[16] Unless explicitly stated in some regression models, the bank-level variables will be measured as the average of the past eight quarters. The changes of bank-level variables stand for the first differencing of the bank-level variables to match the change in bank risk-taking from past eight quarters to subsequent eight quarters. For example, *SIZE* is the past two-year average of the natural logarithm of total assets for bank i in quarter t, and *ΔSIZE* is the change of *SIZE* from past two years to the forward two years.

DLLP	=	Discretionary loan loss provision, calculated as the residual from the models following Kanagaretnam et al. (2010) and Basu et al. (2020): $LLP_{it} = \alpha_0 + \alpha_1 LLA_{it-1} + \alpha_2 NPL_{it-1} + \alpha_3 LOAN_{it} + \alpha_4 \Delta LOAN_{it} + \alpha_5 NCO_{it} + \alpha_6 \Delta NPL_{it} \times D\Delta NPL_{it} + \alpha_6 \Delta NPL_{it} + \alpha_6 D\Delta NPL_{it} + \alpha_7 \Delta lnGDP_{it} + \alpha_8 \Delta UNEMP_{it} + \alpha_9 \Delta HPI_{it} + Bank\ FE + Year\ FE + \varepsilon_{it}$ where *ΔDNPL* is a dummy variable that equals 1 if the sign of *ΔNPL* is negative, and 0 otherwise; other variables are defined in this Appendix.
POS_DLLP	=	A dummy variable that equals 1 if *DLLP* is positive, and 0 otherwise.
SIZE	=	Natural logarithm of total assets.
EBTP	=	Earnings before tax and provisions scaled by lagged total assets.
REVG	=	The growth rate of *EBTP* from past two years to forward two years.
LLP	=	Loan loss provisions scaled by lagged total loans net of unearned revenue and allowances.
LOAN	=	Total loans net of unearned revenue and allowances scaled by lagged total assets.
NPL	=	Nonperforming loans scaled by lagged total loans net of unearned revenue and allowances.
NCO	=	Net charge-offs (= charge-offs – recoveries) scaled by lagged total loans net of unearned revenue and allowances.
T1R	=	Tier 1 capital ratio (=Tier 1 capital / total risk-weighted Assets)
LLA	=	Allowance for loan losses scaled by lagged total loans net of unearned revenue and allowances.
EQTY	=	Amount of shareholders' equity scaled by lagged total loans net of unearned revenue and allowances.
Csloans%	=	Consumer loans scaled by total loans net of unearned revenue and allowances.
Ciloans%	=	Commercial and industrial loans scaled by total loans net of unearned revenue and allowances.
Reloans%	=	Real estate loans scaled by total loans net of unearned revenue and allowances.
ResReloans%	=	Residential real estate loans scaled by total loans net of unearned revenue and allowances.
Homo%	=	Homogeneous loans (i.e., consumer loans and consumer real estate loans) scaled by total loans net of unearned revenue and allowances.
Heter%	=	Heterogenous loans (i.e., commercial loans and commercial real estate loans) scaled by total loans net of unearned revenue and allowances.
LoanGrowth	=	The growth rate of total loans in the current period over lagged total assets.
$NCO_{12m\ or\ 24m}$	=	The future 12-month or 24-month net loan chargeoffs.
ΔRISK_p	=	Past eight period change in bank risk (*RISK* is proxied by Z-Score, *σROA*, and *σNIM.*
Macroeconomic Variables[17]: (*Source: Various, See Below*)		
LnGDP	=	Natural Logarithm of per capita GDP of each state-quarter. (Source: BEA).
HPI	=	House price index of each state-quarter. (Source: Federal Housing Finance Agency).
UNEMP	=	Unemployment rate of each state-quarter. (Source: U.S. Bureau of Labor Statistics).

[17] To match the main dependent and explanatory variables, the macro variables are measured as the average of the past eight quarters. The changes of macro variables stand for the first differencing of the macro variables to match the change in bank risk-taking from past eight quarters to subsequent eight quarters. For example, *LnGDP* is the past two-year average of state level GDP, and *ΔLnGDP* is the change of *LnGDP* from past two years to the forward two years.

Figure 1: Scatter Plots of FinTech Penetration Measures by Year for Selected States

Figure 1A:

3-Year FinTech Loans Per State Capita for Selected States

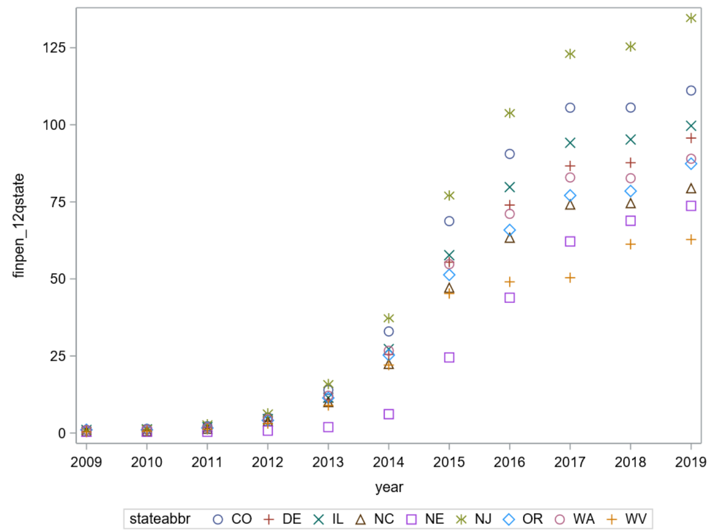


Figure 1B:

3-Year FinTech Loans to State-level Bank Consumer Loans for Selected States

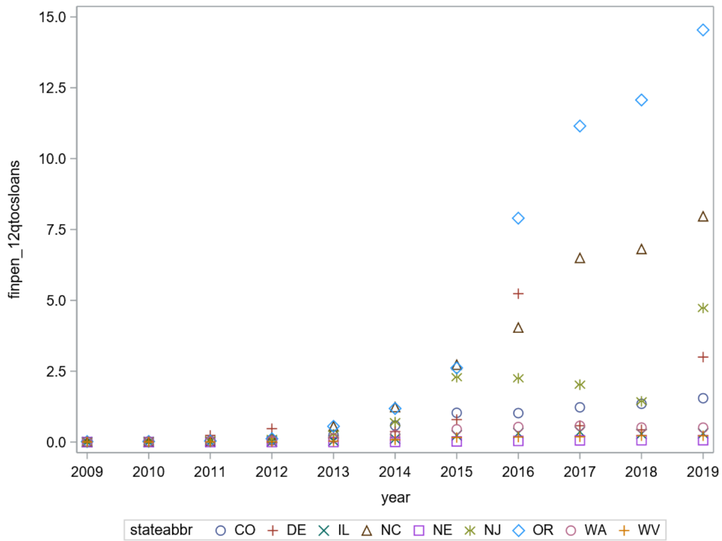


Figure 1C:

3-year FinTech Loans to State-level Total Bank Loans for Selected States

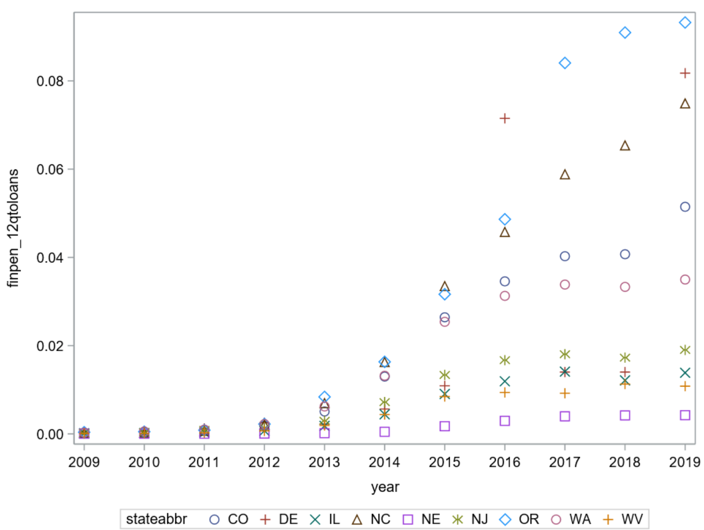


Figure 1A shows scatter plots of the three-year cumulative successfully funded LendingClub and Prosper Marketplace loans scaled by state population in each year from 2009 to 2019 (the y-axis) against year (the x-axis) for 10 selected U.S. states. Each of the 10 symbols represents a different state. Figure 1B shows scatter plots of the three-year cumulative successfully funded LendingClub and Prosper Marketplace loans scaled by state-level community bank consumer loans in each year from 2009 to 2019 (the y-axis) against year (the x-axis) for 10 selected U.S. states. Each of the 10 symbols represents a different state. Figure 1C shows scatter plots of the three-year cumulative successfully funded LendingClub and Prosper Marketplace loans scaled by state-level total community bank loans in each year from 2009 to 2019 (the y-axis) against year (the x-axis) for 10 selected U.S. states. Each of the 10 symbols represents a different state.

Figure 2: Two Snapshots of Fintech Penetration Heatmap as of 2009Q4 and 2015Q4

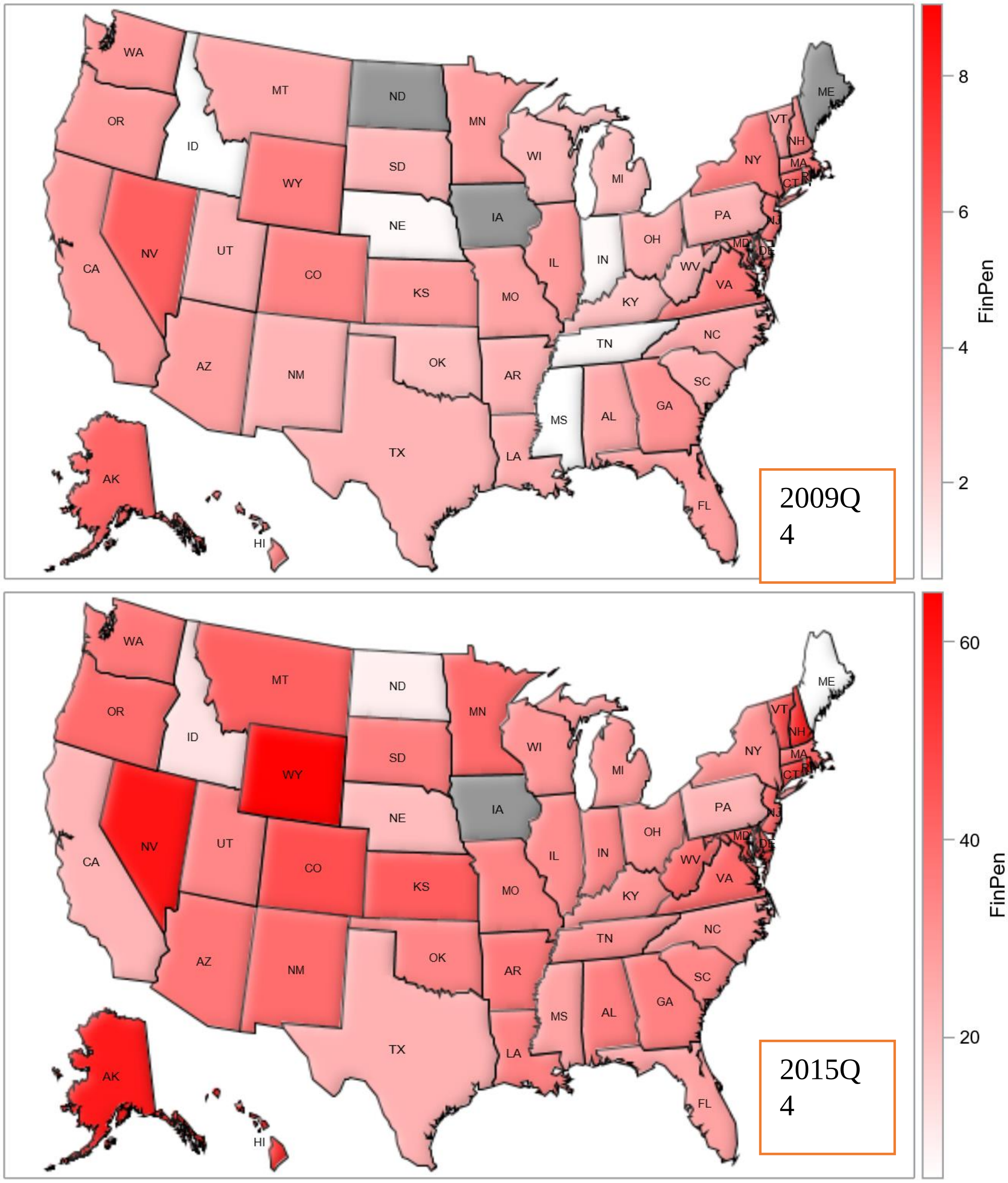


Figure 2 shows two snapshots of the heatmap of FinTech penetration (measured as three-year cumulative successfully funded LendingClub and Prosper Marketplace loans per capita) across all U.S. states. The first heatmap depicts a snapshot of FinTech penetration across U.S. states as of 2009Q4 (which is close to the beginning of the sample period). The second heatmap shows a snapshot of FinTech penetration across U.S. states as of 2015Q4 (which is close to the end of the sample period). Darker red suggests greater FinTech penetration. States with grey color are the ones with no FinTech loans data. Also, note that the scales are different for the two graphs, i.e., in the first heatmap, FinTech penetration ranges from about \$2 to \$8 per capita; in the second heatmap, FinTech penetration ranges from about \$20 to \$60 per capita.

Figure 3: Time Series Scatter Plot of Forward-looking Community Bank Risk-taking (Left) and Charter Value (Right) Over 2007-2016 for Selected U.S. States and for U.S. in Total.

Figure 3A

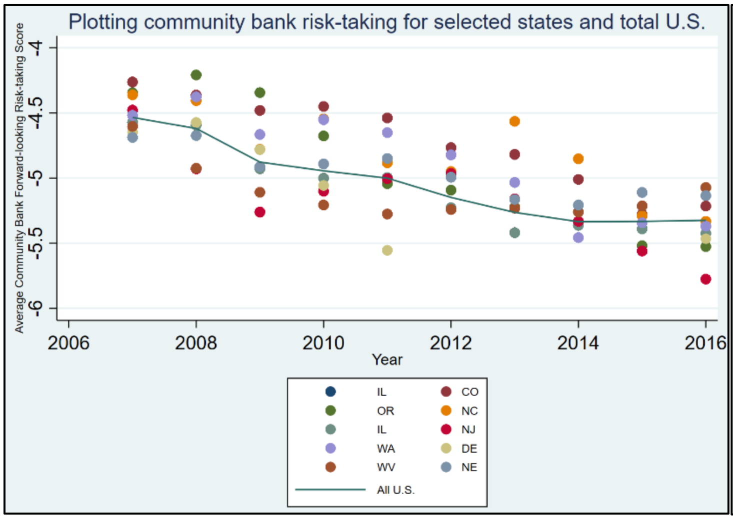


Figure 3B

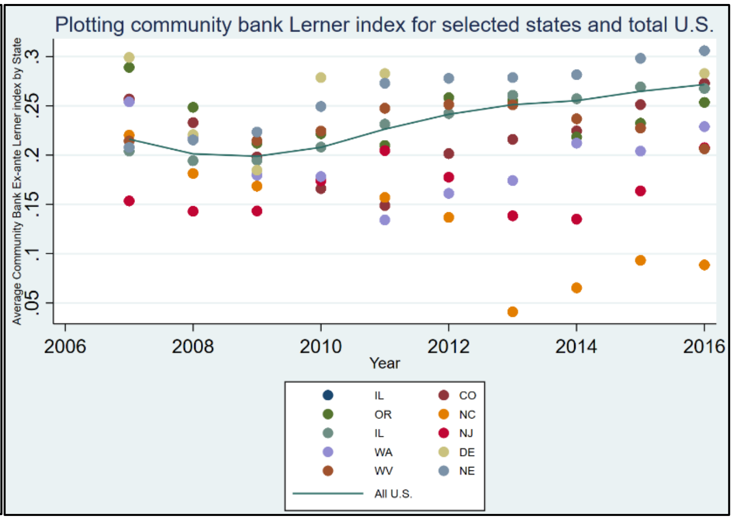


Figure 3A shows the time series variation for the forward-looking community bank risk-taking measured as the negative of natural logarithm of Z-score (i.e., it is measured based on the subsequent two years of bank financials data), for 10 selected U.S. states (the circles with different colors) and the entire U.S. (the solid line) from 2006-2016 (approx. the sample period). In this graph, a higher measure of negative log of Z-score suggests higher risk; therefore, the graph shows an overall downward trending of community bank risk-taking for the selected states and the U.S in total during the sample period.

Figure 3B depicts the times series variation for the state-level average *ex-ante* bank charter value proxied by Lerner index (i.e., the mark-up of price over marginal cost over price as discussed in section 3.2) for 10 selected U.S. states (the circles with different colors) and the entire U.S. (the solid line) from 2006 to 2016 (approx. the sample period). Overall, there is a upward trending of bank charter value over the sample period. Notably, such trend is the opposite of the trending of bank risk-taking. Hence, Figure 3 and Figure 4 depict an overall negative univariate relationship between bank market power and bank risk-taking thoughout the sample years.

Table 1: Main Sample Selection

Sample Exclusion Criteria	Resultant Sample Size
Raw Call Report data queried from WRDS database (2007q1-2019q4)	361,445
• Exclude: obs. with no Zip Codes, no FDIC certificate No., invalid state name	348,481
• Exclude: obs. with missing/negative total assets, total interest expense, salary and employee expense, fixed expenses, non-interest expenses, loan loss provisions, total deposits, and total loans.	327,365
• Exclude: noncommunity bank observations	303,113
• Exclude: U.S. states in list ("AK," "AS," "GU," "PW," "PR," "VI," "HI")	302,374
• Exclude: obs. with quarterly total assets growth rate higher than 20%	298,924
• Exclude: missing values generated from lagging and forwarding for eight periods and missing values from all main and control variables, require sample period to end before 2016	93,882

Final Sample:

Year-quarter range	2009Q1 to 2015Q4[18]
Number of unique community banks	5,560

[18] Note that I also present empirical results using an extended sample period from 2009Q1 to 2017Q4 in Table 12 Panel B.

Table 2: Loan Amount and Loan Terms Distributions in The FinTech Loans Sample

Loan Amount Groups	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Less than $10k	1,630,417	41.93%	1,630,417	41.93%
$10k to $20k	1,414,338	36.38%	3,044,755	78.31%
$20k to $30k	568,592	14.62%	3,613,347	92.93%
$30k to $40k	274,841	7.07%	3,888,188	100%

Loan Terms Category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
12	1,613	0.04%	1,613	0.04%
36	2,733,545	70.3%	2,735,158	70.35%
60	1,153,030	29.65%	3,888,188	100%

The first part of Table 2 tabulates the frequency distribution of the individual successfully funded FinTech loans by four loan amount groups. As the table suggests, most LendingClub and Prosper loans are smaller loans under $20k. This implies that FinTech lenders mainly compete with local community banks in terms of smaller sized loans. The second part of Table 2 shows the frequency distribution of the loan terms. Most loans have a term of three years.

Table 3A: 3-year Cumulative FinTech Loans Per Capita by State at Year End 2010, 2014, and 2018.

State	2010	2014	2018	State	2010	2014	2018	State	2010	2014	2018
NV	1.07	39.17	138.98	GA	1.44	27.13	92.50	TN	0.29	18.27	68.05
NJ	1.38	37.24	125.42	DE	1.15	25.50	87.71	AL	0.74	21.67	67.98
MD	1.56	35.07	125.14	MN	1.16	26.14	86.99	LA	0.53	20.74	67.68
RI	1.14	34.49	119.36	TX	0.72	24.79	84.30	AR	0.55	20.88	67.51
NH	1.12	31.00	107.49	WA	1.14	26.71	82.68	OK	0.59	19.03	66.87
CO	1.36	32.97	105.59	KS	0.62	25.77	78.55	ND	N/A	N/A	65.76
HI	0.74	31.18	102.38	OR	1.06	25.29	78.52	WI	0.69	18.93	64.48
AK	0.79	33.06	101.76	OH	0.86	23.34	76.34	MS	0.26	11.30	62.93
CT	1.47	36.48	100.53	NC	0.58	22.44	74.61	PA	0.84	21.12	62.91
WY	1.05	34.48	100.24	VT	0.58	25.69	72.99	SD	N/A	18.46	61.29
MA	1.49	30.18	99.09	MO	0.99	21.70	71.24	WV	0.62	22.06	61.26
VA	1.41	32.09	98.65	MI	0.76	21.07	71.04	UT	0.91	19.64	60.16
IL	1.20	27.27	95.23	SC	0.69	20.45	70.55	KY	0.48	17.72	59.13
NY	1.30	33.86	95.03	IN	0.43	20.10	70.43	ID	0.71	5.43	53.49
CA	1.44	30.06	94.62	MT	0.62	22.45	69.61	ME	N/A	0.40	52.59
FL	1.11	25.93	93.61	NM	0.73	21.62	69.06	IA	N/A	0.36	N/A
AZ	1.05	26.18	92.68	NE	0.41	6.12	68.87				

Table 3A provides more information to show the variations of FinTech penetration across states and over time—for example, in 2010 FinTech penetration is only about $1 per capita, and increases to about $50-$140 in 2018.

Table 3B: 3-Year Cumulative Fintech Loans Scaled by Community Bank Consumer Loans by State at Year End 2010, 2014, and 2018.

State	2010	2014	2018	State	2010	2014	2018	State	2010	2014	2018
OR	0.02	1.18	12.07	CT	0.04	0.28	0.55	IN	0.00	0.07	0.24
AZ	0.16	5.24	11.30	OH	0.01	0.08	0.54	MS	0.00	0.04	0.23
NC	0.01	1.23	6.81	TX	0.00	0.13	0.53	WV	0.00	0.08	0.23
CA	0.06	1.02	2.79	WA	0.01	0.25	0.51	ME	N/A	0.00	0.23
FL	0.03	0.44	1.64	NY	0.01	0.23	0.48	PA	0.00	0.11	0.22
MD	0.02	0.41	1.53	VA	0.01	0.20	0.48	AR	0.00	0.06	0.21
NJ	0.03	0.70	1.44	SC	0.01	0.16	0.46	OK	0.00	0.05	0.20
AK	0.02	0.65	1.36	DE	0.05	0.37	0.43	HI	0.01	0.09	0.19
CO	0.02	0.58	1.35	VT	0.00	0.17	0.35	KS	0.00	0.06	0.19
GA	0.01	0.09	0.91	MT	0.00	0.10	0.34	LA	0.00	0.05	0.18
NM	0.01	0.28	0.79	AL	0.00	0.10	0.32	TN	0.00	0.05	0.18
ID	0.01	0.09	0.74	MA	0.01	0.09	0.29	KY	0.00	0.04	0.12
NV	0.17	0.21	0.73	MO	0.00	0.08	0.29	NE	0.00	0.01	0.07
NH	0.01	0.36	0.67	WI	0.00	0.08	0.29	SD	N/A	0.01	0.04
RI	0.02	0.26	0.67	IL	0.01	0.12	0.29	ND	N/A	N/A	0.03
UT	0.03	0.62	0.61	MN	0.00	0.10	0.28	IA	N/A	0.00	N/A
MI	0.01	0.25	0.56	WY	0.00	0.11	0.27				

Table 3B provides more information to show the variations of FinTech penetration across states and over time with the alternative measure of FinTech penetration, i.e., FinTech loans as a percentage of bank consumer loans. This measure also shows large variations across states and years. For example, in 2010 the measure is only about 1%-5%, and the percentage increases to more than 100% for some states in 2018, suggesting that loans from LendingClub and Prosper alone have surpassed the total local community bank consumer loans.

Table 4: Descriptive Statistics and Correlations Table

Panel A: Descriptive Statistics

	N	Mean	Std dev	p25	p50	p75
LERNER	93,882	0.228	0.128	0.158	0.237	0.311
Z-Score	93,881	-5.044	0.750	-5.556	-5.117	-4.627
ΔZ-Score	93,881	-0.142	0.685	-0.545	-0.132	0.262
σROA	93,882	0.001	0.001	0.000	0.001	0.001
ΔσROA	93,882	0.000	0.001	0.000	0.000	0.000
σNIM	93,882	0.000	0.000	0.000	0.000	0.001
ΔσNIM	93,882	0.000	0.000	0.000	0.000	0.000
FinPen8Q	86,574	7.958	11.838	0.607	1.820	10.469
FinPen12Q	85,713	9.096	13.473	0.846	2.139	11.675
FinPen8QZip	52,616	7.178	7.915	1.103	3.905	11.076
FinPen12QZip	39,614	10.252	9.698	2.491	7.192	15.278
DLLP	74,752	0.000	0.002	-0.001	0.000	0.000
SIZE	93,882	11.963	0.996	11.275	11.922	12.605
EBTP	93,882	0.003	0.002	0.002	0.003	0.004
EQTY	93,882	0.108	0.028	0.090	0.102	0.119
REVG	93,882	-0.026	0.833	-0.177	-0.027	0.144
LOAN	93,882	0.621	0.153	0.524	0.637	0.732
LLP	93,882	0.001	0.002	0.000	0.001	0.001
NPL	93,882	0.020	0.023	0.005	0.013	0.027
NCO	93,882	0.001	0.002	0.000	0.000	0.001
T1R	93,882	0.165	0.070	0.121	0.145	0.185
Csloans%	93,882	0.064	0.065	0.020	0.044	0.085
Ciloans%	93,882	0.131	0.087	0.070	0.115	0.175
Reloans%	93,882	0.727	0.180	0.621	0.759	0.860
ResReloans%	93,882	0.333	0.190	0.200	0.306	0.429
Homo%	93,882	0.399	0.199	0.255	0.371	0.518
Heter%	93,882	0.601	0.199	0.482	0.629	0.745
ΔLnGDP	93,302	0.067	0.041	0.049	0.067	0.083
ΔHPI	93,302	6.128	15.203	-3.750	8.161	15.336
ΔUNEMP	93,302	-0.756	1.337	-1.600	-1.025	-0.400

Table 4 (continued)

Panel B: Correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	*LERNER*	1.000												
(2)	*ΔZ-Score*	0.047	1.000											
(3)	*ΔσROA*	0.115	0.661	1.000										
(4)	*ΔσNIM*	0.036	0.224	0.149	1.000									
(5)	*FinPen8Q*	0.110	**0.022**	**0.036**	**0.060**	1.000								
(6)	*FinPen12Q*	0.108	**0.023**	**0.036**	**0.061**	1.000	1.000							
(7)	*FinPen8QZip*	0.099	**0.039**	**0.043**	**0.058**	0.888	0.887	1.000						
(8)	*FinPen12QZip*	0.081	**0.054**	**0.047**	**0.056**	0.872	0.872	0.998	1.000					
(9)	*DLLP*	0.060	0.019	0.009	-0.007	0.003	0.002	-0.025	-0.028	1.000				
(10)	*SIZE*	0.175	-0.045	-0.014	-0.022	0.029	0.029	-0.076	-0.092	0.045	1.000			
(11)	*EBTP*	0.742	0.004	0.047	0.039	-0.029	-0.028	-0.017	-0.012	0.078	0.182	1.000		
(12)	*EQTY*	0.230	0.019	0.022	0.025	0.082	0.081	0.093	0.088	-0.001	-0.173	0.142	1.000	
(13)	*REVG*	0.036	-0.024	0.048	0.034	0.051	0.051	0.065	0.074	-0.010	0.014	0.068	0.001	1.000
(14)	*LOAN*	0.036	0.023	0.022	-0.015	-0.008	-0.004	-0.017	-0.009	-0.015	0.240	0.203	-0.245	0.047
(15)	*LLP*	-0.128	0.050	0.004	-0.064	-0.196	-0.193	-0.216	-0.191	0.567	0.124	-0.061	-0.126	-0.066
(16)	*NPL*	-0.329	0.054	0.001	-0.048	-0.177	-0.177	-0.211	-0.206	-0.017	0.081	-0.262	-0.083	-0.111
(17)	*NCO*	-0.172	0.029	-0.008	-0.065	-0.160	-0.159	-0.189	-0.173	-0.066	0.097	-0.127	-0.109	-0.073
(18)	*T1R*	0.109	0.001	0.005	0.023	0.109	0.106	0.120	0.108	-0.009	-0.271	-0.041	0.775	-0.022
(19)	*Csloans%*	0.077	0.005	-0.001	0.014	-0.040	-0.042	0.009	0.019	0.018	-0.282	0.072	0.086	-0.012
(20)	*Ciloans%*	0.069	0.007	0.007	-0.007	-0.024	-0.024	-0.015	-0.019	0.062	0.007	0.074	-0.079	0.024
(21)	*Reloans%*	-0.211	-0.019	-0.014	-0.014	0.047	0.046	-0.032	-0.048	-0.018	0.341	-0.132	-0.069	-0.006
(22)	*ResReloans%*	-0.155	-0.014	-0.007	0.019	0.086	0.084	0.049	0.042	-0.078	0.081	-0.140	0.043	-0.012
(23)	*Homo%*	-0.119	-0.012	-0.007	0.021	0.070	0.068	0.050	0.046	-0.064	-0.017	-0.103	0.077	-0.013
(24)	*Heteo%*	0.119	0.012	0.007	-0.021	-0.070	-0.068	-0.050	-0.046	0.064	0.017	0.103	-0.077	0.013
(25)	*ΔLnGDP*	0.000	-0.029	-0.016	0.047	-0.149	-0.153	-0.189	-0.211	-0.005	-0.084	0.019	0.018	0.010
(26)	*ΔHPI*	0.161	-0.024	0.019	0.077	0.557	0.548	0.484	0.390	0.027	-0.048	0.027	0.088	0.045
(27)	*ΔUNEMP*	-0.073	0.046	0.001	-0.088	-0.233	-0.221	-0.017	0.201	-0.025	0.059	0.001	-0.112	-0.015

Table 4 (continued)

		(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)
(14)	*LOAN*	1.00													
(15)	*LLP*	0.11	1.00												
(16)	*NPL*	0.04	0.40	1.00											
(17)	*NCO*	0.06	0.70	0.40	1.00										
(18)	*T1R*	-0.59	-0.17	-0.11	-0.14	1.00									
(19)	*Csloans%*	-0.31	-0.04	-0.11	-0.03	0.22	1.00								
(20)	*Ciloans%*	0.01	0.03	-0.04	0.01	-0.15	0.02	1.00							
(21)	*Reloans%*	0.25	0.07	0.19	0.07	-0.07	-0.44	-0.51	1.00						
(22)	*ResReloans%*	0.04	-0.06	0.00	-0.05	0.22	-0.06	-0.47	0.63	1.00					
(23)	*Homo%*	-0.06	-0.06	-0.04	-0.06	0.29	0.30	-0.45	0.44	0.93	1.00				
(24)	*Heteo%*	0.06	0.06	0.04	0.06	-0.29	-0.30	0.45	-0.44	-0.93	-1.00	1.00			
(25)	*ΔLnGDP*	-0.05	-0.05	-0.03	-0.03	0.02	0.03	0.04	-0.13	-0.08	-0.07	0.07	1.00		
(26)	*ΔHPI*	-0.12	-0.30	-0.22	-0.23	0.13	0.02	0.05	-0.09	-0.06	-0.05	0.05	0.19	1.00	
(27)	*ΔUNEMP*	0.14	0.19	0.05	0.10	-0.15	0.00	0.04	-0.04	-0.09	-0.09	0.09	-0.32	-0.58	1.00

This table provides the descriptive statistics (Panel A) and Pearson's correlation (Panel B) of the main variables used in this study. The detailed definitions of the variables are provided in the Appendix. All correlations with absolute values greater than 0.02 are statistically significant at the 0.01 level or better (two-tailed).

Table 5: Baseline Regressions of Future Change of Bank Risk on Current Exposure to FinTech Penetration

	(1)		(2)		(3)		(4)		(5)		(6)	
Dependent Var.	*ΔZ-Score*		*ΔZ-Score*		*ΔσNIM*		*ΔσNIM*		*ΔσROA*		*ΔσROA*	
FinPen8Q	**0.0056*** **	**(4.91)**			**0.0181*** **	**(3.44)**			**0.0762*** **	**(3.50)**		
FinPen12Q			**0.0046*** **	**(4.72)**			**0.0147*** **	**(3.33)**			**0.0592*** **	**(3.20)**
SIZE	-0.1571*	(-1.77)	-0.1572*	(-1.75)	-2.180***	(-4.06)	-2.265***	(-4.17)	-7.147***	(-2.67)	-7.434***	(-2.73)
ΔSIZE	-0.0301	(-0.24)	-0.0436	(-0.35)	0.3532	(0.36)	0.3497	(0.36)	-4.8702	(-1.53)	-4.9951	(-1.55)
REVG	0.0094	(1.26)	0.0093	(1.24)	0.070*	(1.90)	0.0694*	(1.86)	1.0255***	(3.88)	1.022***	(3.83)
LOAN	-0.5260**	(-2.38)	-0.5276**	(-2.36)	-0.5000	(-0.28)	-0.5735	(-0.31)	-16.94***	(-2.67)	-16.87***	(-2.61)
ΔLOAN	-0.959***	(-5.06)	-0.950***	(-4.97)	1.1841	(1.10)	1.1387	(1.07)	-14.12***	(-3.02)	-13.95***	(-2.94)
LLP	72.651***	(6.97)	73.52***	(6.98)	-166.122*	(-1.85)	-167.19*	(-1.91)	829.83***	(2.99)	853.88***	(3.02)
ΔLLP	102.8***	(13.86)	103.33***	(13.72)	41.2979	(0.71)	30.7894	(0.53)	1533.8***	(7.09)	1553.9***	(6.93)
T1R	-0.5572	(-1.10)	-0.5914	(-1.15)	-5.3132	(-1.56)	-5.9738*	(-1.74)	-23.1110	(-1.57)	-23.4917	(-1.57)
ΔT1R	-3.703***	(-6.81)	-3.71***	(-6.77)	5.6825*	(1.91)	5.5494*	(1.85)	-30.538**	(-2.07)	-30.203**	(-2.03)
LnGDP	-0.2912	(-1.19)	-0.2172	(-0.86)	3.8685***	(2.64)	4.2871***	(2.83)	2.4987	(0.55)	4.6307	(0.94)
ΔLnGDP	-0.3100	(-1.11)	-0.2266	(-0.76)	5.1077***	(3.14)	5.8152***	(3.32)	0.2672	(0.06)	2.4347	(0.47)
HPI	0.0000	(0.04)	0.0002	(0.20)	0.0067	(1.03)	0.0073	(1.13)	-0.0196	(-0.89)	-0.0180	(-0.81)
ΔHPI	-0.0002	(-0.14)	-0.0002	(-0.22)	0.0311***	(3.58)	0.0297***	(3.39)	-0.0372	(-1.27)	-0.0388	(-1.30)
UNEMP	0.0070	(0.46)	0.0101	(0.66)	0.3374***	(3.67)	0.3497***	(3.74)	0.3993	(1.37)	0.4790	(1.62)
ΔUNEMP	-0.0010	(-0.08)	-0.0011	(-0.08)	0.1473*	(1.73)	0.1491*	(1.66)	-0.1014	(-0.40)	-0.0553	(-0.22)
LERNER	1.868***	(10.60)	1.89***	(10.58)	-0.4283	(-0.38)	-0.2434	(-0.22)	52.062***	(6.65)	52.534***	(6.58)
Constant	5.269*	(1.71)	4.281	(1.34)	-26.437	(-1.38)	-30.890	(-1.58)	55.078	(0.89)	30.432	(0.46)
Bank FE	Yes		Yes		Yes		Yes		Yes		Yes	
Year FE	Yes		Yes		Yes		Yes		Yes		Yes	
Clustered SE	Bank		Bank		Bank		Bank		Bank		Bank	
Adj. R-squared	0.27		0.27		0.24		0.24		0.34		0.34	
N	86,241		85,404		86,242		85,405		86,242		85,405	

Table 5 reports the results of tests examining the effect of FinTech penetration on future change in bank risk-taking. The sample consists of bank-quarter observations from 2009Q1 to 2005Q4. All variables are defined in Appendix A. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. All continuous variables are winsorized at 1% and 99%. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Path Analysis of the Mediating Role of Bank Charter Value

Panel A: Single-mediator Analysis:

	FinTech Penetration Proxied By			
	FinPen8Q	T-stat	*FinPen12Q*	T-stat
Direct Path				
ρ (ΔZ-Score, FinPen)	0.0850***	(3.21)	0.0973***	(3.63)
Indirect Path				
ρ (ΔZ-Score, ΔLerner)	-2.2716***	(-51.54)	-2.2932***	(-51.29)
ρ (ΔLerner, FinPen)	-0.0136***	(-5.50)	-0.0105***	(-4.84)
Total mediated path	**0.0308*** **	**(5.46)**	**0.0241*** **	**(4.81)**
Mediated path as a percentage of total path	**27%**		**20%**	
Bank-level controls	Yes		Yes	
Macro-level controls	Yes		Yes	
N	86,366		85,510	

Panel B: Multi-mediator Analysis:

	FinTech Penetration Proxied By			
	FinPen8Q	T-stat	*FinPen12Q*	T-stat
Direct Path				
ρ (ΔZ-Score, FinPen)	0.0646**	(2.41)	0.0622***	(2.65)
Indirect Paths				
ρ (ΔZ-Score, ΔLerner)	-2.2820***	(-51.55)	-2.2819***	(-51.29)
ρ (ΔLerner, FinPen)	-0.0136***	(-5.50)	-0.0105***	(-4.84)
ρ (ΔZ-Score, NIMQ)	-9.6202***	(-5.11)	-9.1047***	(-4.82)
ρ (NIMQ, FinPen)	-0.0028***	(-58.11)	-0.0025***	(-58.14)
ρ (ΔZ-Score, ΔT1R)	-3.2746***	(-21.86)	-3.2712***	(-21.76)
ρ (ΔT1R, FinPen)	-0.0190***	(-26.94)	-0.0162***	(-26.10)
Total mediated path	**0.1203*** **	**(14.38)**	**0.0996*** **	**(13.56)**
Mediated path as a percentage of total path	**65%**		**62%**	
Bank-level controls	Yes		Yes	
Macro-level controls	Yes		Yes	
N	86,366		85,510	

Table 6 reports result from path analysis that examines the direct effect and the indirect effect of FinTech penetration on future change in bank risk through the mediator of bank charter value. Panel A uses *ΔLerner* as a single mediator. Panel B uses *NIMQ*, *ΔT1R*, and *ΔLerner* as multi-mediators. A recursive path model with observable variables is used. The standard error is heteroskedasticity-robust. All continuous variables are winsorized at 1% and 99%. The constant term is not presented. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Coefficients on *FinPen* are scaled by 100 for display purposes.

Table 7: Moderating Effect of Bank Charter Value

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	*ΔZ-Score*	*ΔZ-Score*	*ΔσROA*	*ΔσNIM*	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*
FINPEN Proxy	*FinPen8Q*	*FinPen12Q*	*FinPen8Q*	*FinPen8Q*	*FinPen8Q*	*FinPen12Q*	*FinPen8Q*	*FinPen8Q*
FinPen	0.0074***	0.0061***	0.1426***	0.0254***	0.0083***	0.0069***	0.0100***	0.0127***
	(5.19)	(5.01)	(5.08)	(3.99)	(5.02)	(4.86)	(6.19)	(4.38)
LERNERQ2 × FinPen	**-0.0007**	**-0.0006**	**-0.0771*****	**-0.0070**				
	(-0.54)	**(-0.54)**	**(-3.23)**	**(-1.05)**				
LERNERQ3 × FinPen	**-0.0031****	**-0.0027****	**-0.1112*****	**-0.0125***				
	(-2.38)	**(-2.38)**	**(-4.42)**	**(-1.79)**				
LERNERQ4 × FinPen	**-0.0056*****	**-0.0049*****	**-0.1401*****	**-0.0167****				
	(-4.10)	**(-4.06)**	**(-4.96)**	**(-2.23)**				
FinPen × LERNER					**-0.0113****	**-0.0099****		
					(-2.49)	**(-2.48)**		
FinPen × ROA_EX							**-1.6405*****	
							(-4.17)	
FinPen × NIM_EX								**-0.9168*****
								(-2.91)
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank
Adj. R-squared	0.27	0.27	0.36	0.24	0.27	0.27	0.22	0.22
N	86,241	85,404	86,242	86,242	86,241	85,404	86,241	86,241

Tests of Coefficients of Linear Combinations (Marginal Effect of FINPEN in Each LERNER_EX Quartile):

	(1)	(2)	(3)	(4)
FinPen × (1 + LERNERQ1)	**0.0074*****	**0.0061*****	**0.1426*****	**0.0253*****
	(5.19)	**(5.01)**	**(5.08)**	**(3.99)**
FinPen × (1 + LERNERQ2)	**0.0067*****	**0.0056*****	**0.0655*****	**0.0184*****
	(5.11)	**(1.92)**	**(2.89)**	**(2.84)**
FinPen × (1 + LERNERQ3)	**0.0043*****	**0.0035*****	**0.0314**	**0.0129***
	(3.26)	**(3.03)**	**(1.25)**	**(1.94)**
FinPen × (1 + LERNERQ4)	**0.0018**	**0.0013**	**0.0024**	**0.0086**
	(1.25)	**(1.05)**	**(0.09)**	**(1.28)**

Table 7 reports the results of examining the moderating role bank charter value on the relation between FinTech penetration on future change in bank risk-taking. The sample consists of bank-quarter observations from 2009Q1 to 2005Q4. All variables are defined in Appendix A. Columns 1-4 use the quartile of Lerner index by time as the moderator. Columns 5-6 use the continuous measure of Lerner index as the moderator. Columns 7-8 use alternative proxies of bank charter value. Also shown are partial effects of FinTech penetration for each Lerner index quarter. Continuous variables are winsorized at 1% and 99%. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. * p < .1, ** p < .05, *** p < .01

Table 8: Moderating Effect of DLLP

FinPen Proxy	(5) *ΔZ-Score* *FinPen8Q*	(6) *ΔZ-Score* *FinPen12Q*	(3) *ΔZ-Score* *FinPen8Q*	(4) *ΔZ-Score* *FinPen12Q*	(1) *ΔZ-Score* *FinPen8Q*	(2) *ΔZ-Score* *FinPen12Q*
FinPen	0.0069***	0.0055***	0.0067***	0.0054***	0.0062***	0.0050***
	(2.95)	(2.69)	(3.00)	(2.80)	(2.84)	(2.64)
FinPen × DLLP	**-0.5907****	**-0.5181****				
	(-2.30)	**(-2.28)**				
FinPen × POS_DLLP			**-0.0015*****	**-0.0013*****		
			(-2.85)	**(-2.83)**		
FinPen × LLP					**-0.8219*****	**-0.7291*****
					(-3.21)	**(-3.23)**
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Bank	Bank	Bank	Bank	Bank	Bank
Adj. R-squared	0.29	0.29	0.27	0.27	0.27	0.27
N	68,856	68,386	86,241	85,404	86,241	85,404

Table 8 reports the results of examining the moderating role of discretionary loan loss provisions on the relation between FinTech penetration on future change in bank risk-taking. The sample consists of bank-quarter observations from 2009Q1 to 2005Q4. All variables are defined in Appendix A. Columns 1-2 use the continuous measure of DLLP as the moderator. Columns 3-4 use the dummy variable *POS_DLLP* as the moderator. Columns 5-6 use the LLP as the moderator. All continuous variables are winsorized at 1% and 99%. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. * p < .1, ** p < .05, *** p < .01

Table 9: Moderating Effect of Information Type

FinPen Proxy	(1) *ΔZ-Score* *FinPen8Q*	(2) *ΔZ-Score* *FinPen12Q*	(3) *ΔZ-Score* *FinPen8Q*	(4) *ΔZ-Score* *FinPen12Q*	(5) *ΔZ-Score* *FinPen8Q*	(6) *ΔZ-Score* *FinPen12Q*
FinPen × SIZE	**0.0012****	**0.0011****				
	(2.48)	**(2.43)**				
FinPen × ResReLoans%			**0.0081*****	**0.0070*****		
			(3.22)	**(3.19)**		
FinPen × HomoLoans%					**0.0051****	**0.0044****
					(2.09)	**(2.06)**
FinPen	-0.0095	-0.0086	0.0020	0.0013	0.0032	0.0024
	(-1.49)	(-1.54)	(0.81)	(0.62)	(1.28)	(1.10)
SIZE	-0.1827**	-0.1800**				
	(-2.16)	(-2.10)				
ResReLoans%			-0.0695	-0.0907		
			(-0.33)	(-0.43)		
HomoLoans%					-0.1302	-0.1370
					(-0.69)	(-0.72)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Bank	Bank	Bank	Bank	Bank	Bank
Adj. R-squared	0.27	0.27	0.27	0.27	0.27	0.27
N	86,241	85,404	86,241	85,404	86,241	85,404

Table 9 reports the results of examining the moderating role of banks' tendency to use hard vs. soft information on the relation between FinTech penetration on future change in bank risk-taking. The sample consists of bank-quarter observations from 2009Q1 to 2005Q4. All variables are defined in Appendix A. Columns 1-2 use the bank size as the moderator. Columns 3-4 use banks' proportion of residential real estate loans as the moderator. Columns 5-6 use banks' proportion of homogeneous loans as the moderator. All continuous variables are winsorized at 1% and 99%. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. * p < .1, ** p < .05, *** p < .01

Table 10: FinTech Penetration and Bank Loan Risk

Panel A: FinTech Penetration and Individual Bank Loan Risk: Future Charge-offs.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Var.	NCO_{12m}	NCO_{24m}	NCO_{12m}	NCO_{24m}	NCO_{12m}	NCO_{24m}	NCO_{12m}	NCO_{24m}
FINPEN Proxy	*FinPen8Q*	*FinPen8Q*	*FinPen12Q*	*FinPen12Q*	*FinPen8QZip*	*FinPen8QZip*	*FinPen12QZip*	*FinPen12QZip*
FinPen × *LoanGrowth*	**0.0164***** **(2.69)**	**0.0269***** **(2.99)**	**0.0143***** **(2.66)**	**0.0233***** **(2.92)**	**0.0431***** **(3.91)**	**0.0813***** **(5.42)**	**0.0409***** **(4.87)**	**0.0648***** **(5.13)**
FinPen	0.0002*** (7.75)	0.0005*** (9.08)	0.0002*** (8.05)	0.0004*** (9.33)	0.0005*** (7.90)	0.0008*** (8.30)	0.0003*** (6.30)	0.0006*** (7.05)
LoanGrowth	-0.0106*** (-7.93)	-0.0141*** (-6.36)	-0.0106*** (-7.84)	-0.0140*** (-6.25)	-0.0122*** (-5.40)	-0.0161*** (-4.75)	-0.0123*** (-4.58)	-0.0167*** (-4.30)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank
Adj. R-squared	0.63	0.76	0.64	0.76	0.66	0.78	0.66	0.78
N	86,448	86,448	85,608	85,608	52,452	52,452	39,423	39,423

Table 10 Panel A reports the results of examining the effect of FinTech penetration on the relationship between current period loan growth and future period loan charge-offs. The sample consists of bank-quarter observations from 2009Q1 to 2005Q4. All variables are defined in Appendix A. Columns 1-2 use the *FinPen8Q* as proxy for FinTech penetration. Columns 3-4 use the *FinPen12Q* as proxy for FinTech penetration. Columns 5-6 use the *FinPen8QZip* as proxy for FinTech penetration. Columns 7-8 use the *FinPen12QZip* as proxy for FinTech penetration. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. * p < .1, ** p < .05, *** p < .01

Panel B: FinTech Penetration and Individual Bank Loan Risk: Future Change in Loan Loss Provisions and Charge-offs.

Dependent Variable	(1) *ΔLLP*	(2) *ΔLLP*	(3) *ΔLLP*	(4) *ΔLLP*	(5) *ΔLLP*	(6) *ΔLLP*	(7) *ΔNCO*	(8) *ΔNCO*
FINPEN Proxy	*FinPen8Q*	*FinPen12Q*	*FinPen8Q*	*FinPen12Q*	*FinPen8Q*	*FinPen12Q*	*FinPen8Q*	*FinPen12Q*
FinPen × LERNER	-0.0324***	-0.0290***					-0.0348***	-0.0306***
	(-4.14)	(-4.28)						
LERNER_H × FinPen			-0.0083***	-0.0073***				
			(-5.63)	(-5.67)				
LERNERQ2 × FinPen					-0.0049***	-0.0041**		
					(-2.61)	(-2.50)		
LERNERQ3 × FinPen					-0.0110***	-0.0095***		
					(-5.35)	(-5.30)		
LERNERQ4 × FinPen					-0.0109***	-0.0095***		
					(-4.85)	(-4.87)		
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank
Adj. R-squared	0.77	0.78	0.77	0.77	0.77	0.77	0.61	0.61
N	86,242	85,405	86,242	85,405	86,242	85,405	86,242	85,405
Tests of Coefficients of Linear Combinations (Marginal Effect of FINPEN in Each LERNER_EX Quartile)								
FinPen + LERNER_LOW × FinPen			0.0067***	0.0061***				
			(3.00)	(3.26)				
FinPen + LERNER_HIGH × FinPen			-0.0015	-0.0011				
			(-0.68)	(-0.56)				
FinPen + LERNERQ1 × FinPen					0.0093***	0.0083***		
					(3.98)	(4.18)		
FinPen + LERNERQ2 × FinPen					0.0044*	0.0043**		
					(1.78)	(2.03)		
FinPen + LERNERQ3 × FinPen					-0.0017	-0.0011		
					(-0.68)	(-0.55)		
FinPen + LERNERQ4 × FinPen					-0.0014	-0.0010		
					(-0.58)	(-0.49)		

Table 10 Panel B reports the results of examining the relationship between FinTech penetration and change in bank loan risk. The sample consists of bank-quarter observations from 2009Q1 to 2005Q4. All variables are defined in Appendix A. Columns 1-2 use the continuous measure of the Lerner index to interact with FinTech penetration. Columns 3-4 use a dummy variable *LERNER_H* as the moderator. Columns 5-6 Lerner index quartiles by time as the moderator. Columns 7-8 use *ΔNCO* as the dependent variable. Also shown are partial effects of FinTech penetration for subsample groups. All continuous variables are winsorized at 1% and 99%. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. * p < .1, ** p < .05, *** p < .01

Table 11: Testing the Effect of FinTech Penetration with Propensity Score Matching

	(1) Nearest-Neighbor Matching (N=1)	(2) Epanechnikov Kernel	(3) Nearest-Neighbor Matching (N=1)	(4) Epanechnikov Kernel
Outcome Var.	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*
Average Treatment Effect for the Treated (ATT):				
FinPen_8QHigh	**0.1530****** (t = 6.07)**	**0.0521****** (t = 3.88)**		
FinPen_12QHigh			**0.1706****** (t = 3.59)**	**0.0694****** (t = 5.03)**
Control Var.	Yes	Yes	Yes	Yes
Bank FE	No	No	No	No
Year FE	No	No	No	No
Standard Error	AI (2006)	Bootstrapped	AI (2006)	Bootstrapped
Adj. R-squared	0.1172	0.2680	0.1622	0.2680
N	7,006	13,999	1,945	13,999

Table 11 reports results using various setups of propensity score matching. *FinPen_8QHigh* or *FinPen_12QHigh* is the status of receiving "treatment" of high FinTech threat for community bank-quarter i at time t. It equals 1 when the FinTech penetration for the bank-quarter is in the highest quartile of FinTech penetration measure by holding the year-quarter fixed, and 0 otherwise. In all the propensity score matching models, the dependent variable is future change in bank risk-taking (*ΔZ-Score*). In Model 1, the treated and control groups are 1-on-1 matched based on firm-level characteristics (i.e., firm size, total loans scaled by lagged total assets, loan loss provisions, earnings before tax and provisions scaled by lagged total assets, and Tier 1 ratio) and macroeconomic controls (i.e., natural logarithm of GDP, housing price index, and unemployment rate). Model 1 uses the standard errors that are Abadie and Imbens (2006) robust. In Model 2, I employ the Epanechnikov kernel matching for the treated and control observations. The standard errors are bootstrapped. In models 3-4, I repeat the first two model estimations but with *FinPen_12QHigh* as the treatment variable. The average treatment effects for the treated (ATT) are reported in this table. For all models, I impose common support by dropping the 10 percent of the treatment observations at which the p-score density of the control observations is the lowest. Propensity scores are calculated using the logit function. I use a caliper of 0.25 for the nearest-neighbor matchings. All firm level continuous variables are winsorized at 1% and 99%. * p < .1, ** p < .05, *** p < .01. All coefficients are multiplied by 100 for display purposes.

Table 12: Sensitivity Tests

Panel A: Regression of Future Change in Bank Risk on Fintech Penetration for Different Subsamples

	(1) Year <= 2012	(2) Year > 2012	(3) Year <= 2011	(4) Year > 2011	(5) Banks with more regulatory scrutiny (TA > $500M)	(6) Banks with less regulatory scrutiny (TA < $500M)	(7) Banks with more regulatory scrutiny (high T1R)	(8) Banks with less regulatory scrutiny (low T1R)
FinPen 8Q	**0.0528*****	**0.0054*****	**0.1846*****	**0.0046*****	**0.0095*****	**0.0050*****	**0.0064*****	**0.0057*****
	(3.61)	**(3.33)**	**(5.17)**	**(3.06)**	**(3.20)**	**(4.04)**	**(4.10)**	**(3.30)**
*SIZE_*	-0.5778***	0.1398	-0.7181***	-0.0422	-0.3548	-0.1433	-0.3989**	-0.0850
	(-3.19)	(0.56)	(-3.13)	(-0.24)	(-1.49)	(-1.42)	(-2.38)	(-0.73)
ΔSIZE	-0.3916**	0.0183	-0.2520	0.1997	-0.1533	-0.0084	0.1160	-0.1708
	(-2.09)	(0.07)	(-1.13)	(0.96)	(-0.38)	(-0.06)	(0.58)	(-1.04)
REVG	0.0068	0.0324**	0.0079	0.0246**	0.0054	0.0093	0.0316***	0.0003
	(0.88)	(2.41)	(1.00)	(2.09)	(0.21)	(1.19)	(2.67)	(0.03)
LOAN	-1.0659***	-0.6275	-2.2676***	-0.7954*	-0.9515	-0.5104**	-0.2466	-0.7320**
	(-2.77)	(-1.17)	(-4.51)	(-1.94)	(-1.35)	(-2.18)	(-0.80)	(-2.03)
ΔLOAN	-1.6083***	-0.0709	-2.4325***	-0.4349	-0.6945	-0.9764***	-0.4376	-1.3319***
	(-5.70)	(-0.18)	(-6.92)	(-1.36)	(-1.02)	(-4.94)	(-1.60)	(-4.82)
LLP	66.6556***	28.9250	68.0667***	75.1171***	66.8296	72.1644***	57.3839***	53.9161***
	(4.88)	(1.15)	(4.49)	(3.96)	(1.57)	(6.84)	(3.72)	(3.60)
ΔLLP	92.63***	95.52***	91.10***	116.89***	87.41***	104.78***	95.11***	84.35***
	(10.73)	(5.25)	(9.86)	(7.84)	(3.05)	(13.89)	(9.04)	(7.98)
T1R	-1.3990	-0.9411	-2.2945**	-1.1326	-2.4352	-0.5344	-0.5511	-0.8233
	(-1.60)	(-0.76)	(-2.03)	(-1.15)	(-1.26)	(-1.01)	(-0.91)	(-0.55)
ΔT1R	-5.2753***	-2.0630**	-5.9898***	-2.4469***	-5.6067**	-3.4043***	-1.3419**	-8.4885***
	(-6.36)	(-2.22)	(-5.95)	(-3.11)	(-2.36)	(-6.13)	(-1.98)	(-8.63)
Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.41	0.42	0.53	0.33	0.32	0.27	0.30	0.33
N	51,077	34,796	38,307	47,540	11,365	74,832	43,729	42,291

Table 12 reports the results of sensitivity tests. Table 12 Panel A presents the regression results of the baseline model with different year, total assets, and bank capital subsample cutoffs. Models 1-4 show subsample regressions based on different year groups. Models 5-6 use the FDICIA internal control total assets threshold of $500 million as the total assets' cutoff points. Models 7-8 use Tier 1 capital cutoffs. All variables are defined in Appendix A. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. All continuous variables are winsorized at 1% and 99%. * p < .1, ** p < .05, *** p < .01.

Panel B: Regression of Future Change in Bank Risk on Fintech Penetration Using an Extended Sample From 2009-2017

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Var.	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*	*ΔZ-Score*	NCO_{12m}	NCO_{24m}	*ΔNCO*
FinPen Proxy	*FinPen8Q*	*FinPen12Q*	*FinPen8Q*	*FinPen8Q*	*FinPen8Q*	*FinPen8Q*	*FinPen8Q*	*FinPen8Q*	*FinPen8Q*	*FinPen8Q*
FinPen	**0.0032***** **(3.34)**	**0.0020***** **(2.61)**	0.0039*** (3.60)	0.0030*** (2.80)	-0.0106*** (-3.24)	0.0020* (1.85)	0.0016 (1.50)	0.0002*** (9.64)	0.0004*** (10.35)	0.0078*** (3.98)
LERNERQ2 × FinPen			**-0.0005** **(-0.69)**							
LERNERQ3 × FinPen			**-0.0013*** **(-1.78)**							
LERNERQ4 × FinPen			**-0.0017**** **(-2.27)**							
DLLP × FinPen				**-0.5705***** **(-4.84)**						
SIZE × FinPen					**0.0011***** **(4.35)**					
HomoLoans% × FinPen						**0.0026**** **(1.99)**				
ResReLoans% × FinPen							**0.0040***** **(2.93)**			
LoanGrowth × FinPen *(* 100)*								**0.0002***** **(6.25)**	**0.0281***** **(5.83)**	
LERNER × FinPen										**-0.0267***** **(-5.09)**
Bank, Macro Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank	Bank
Adj. R-squared	0.22	0.22	0.22	0.23	0.22	0.22	0.22	0.60	0.71	0.58
N	108,148	107,352	108,148	87,726	108,148	108,148	108,148	108,377	108,377	108,149

Table 12 Panel B reports the regression results of the baseline model with an extended sample period from 2009Q1 to 2017Q4 to demonstrate that my main results are not driven by restricted sample periods. All variables are defined in Appendix A. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. All continuous variables are winsorized at 1% and 99%. * p < .1, ** p < .05, *** p < .01.

Table 13: Regression Results with Alternative Measures and Placebo Tests

Panel A: Regression Results with Three-digit Zip Code Measure of Fintech Penetration						
	(1)	(2)	(3)	(4)	(5)	(6)
	ΔZ-Score	ΔZ-Score	ΔZ-Score	ΔZ-Score	ΔZ-Score	ΔZ-Score
FinPenZip Proxy	*8QZip*	*12QZip*	*8QZip*	*8QZip*	*8QZip*	*8QZip*
FinPenZip	**0.0058***** **(2.80)**	**0.0051***** **(2.59)**	0.0070*** (2.65)	0.0045* (1.94)	0.0002 (0.06)	-0.0013 (-0.43)
LERNERQ2 × FinPenZip			0.0020 (0.77)			
LERNERQ3 × FinPenZip			-0.0027 (-1.02)			
LERNERQ4 × FinPenZip			**-0.0072**** **(-2.50)**			
FINPEN_8QZIP × DLLP				**-1.0941**** **(-2.34)**		
FinPenZip × HomoLoans%					**0.0117**** **(2.40)**	
FinPenZip × ResReLoans%						**0.0165***** **(3.35)**
DLLP				5.8318** (2.04)		
HomoLoans%					-0.7290** (-2.57)	
ResReLoans%						-0.7449** (-2.38)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustered SE	Bank	Bank	Bank	Bank	Bank	Bank
Adj. R-squared	0.32	0.36	0.32	0.34	0.32	0.32
N	52,109	39,171	52,109	42,370	52,109	52,109

Table 13 Panel A reports the main regression results with three-digit Zip Code level of FinTech penetration measures (*FINPEN_8QZIP*, *FINPEN_12QZIP*). All variables are defined in Appendix A. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. All continuous variables are winsorized at 1% and 99%. * p < .1, ** p < .05, *** p < .01.

Panel B: Regression Results with FinTech Loans to Bank Loans/Bank Consumer Loans as Measures of FinTech Penetration

	(1) *ΔZ-Score*	(2) *ΔZ-Score*	(3) *ΔZ-Score*	(4) *ΔZ-Score*	(5) *ΔσNIM*	(6) *ΔσROA*	(7) *ΔσNIM*	(8) *ΔσROA*
FinPen_8qCSLoans	0.3077***						1.8789**	5.7235**
	(2.58)						(2.12)	(2.42)
FinPen_8qLoans		11.8165**			91.52***	173.13*		
		(2.46)			(3.15)	(1.94)		
FinPen_12qCSLoans			0.2696**					
			(2.55)					
FinPen_12qLoans				10.1431**				
				(2.43)				
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank and Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-squared	0.27	0.27	0.27	0.27	0.24	0.36	0.24	0.36
N	86241.00	86241.00	85404.00	85404.00	86242.00	86242.00	86242.00	86242.00

Table 13 Panel B reports the baseline regression results with alternative FinTech penetration measures that are based on cumulative FinTech loans to state-level bank loans or state-level bank consumer loans (*FinPen_8qCSLoans, FinPen_8qLoans, FinPen_12qCSLoans, FinPen_12qLoans*). All variables are defined in Appendix A. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. All continuous variables are winsorized at 1% and 99%. * p < .1, ** p < .05, *** p < .01. Standard errors are clustered at bank level.

Panel C: Placebo Tests with Past Changes of Bank Risk as Dependent Variables

	(1) *ΔZ-Score_p*	(2) *ΔσROA_p*	(3) *ΔσNIM_p*	(4) *ΔZ-Score_p*	(5) *ΔσROA_p*	(6) *ΔσNIM_p*
FinPen8Q	0.0001	-0.0000	0.0000			
	(0.05)	(-0.72)	(0.27)			
FinPen12Q				0.0002	-0.0000	0.0000
				(0.09)	(-0.72)	(0.28)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Bank and Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.33	0.25	0.29	0.32	0.25	0.29
N	51,813	81,384	81,384	51,750	80,841	80,841

Table 13 Panel C reports the baseline regression results with past change in bank risk-taking as the dependent variables for placebo testing purposes. All variables are defined in Appendix A. All regressions include bank and year fixed effects. Standard errors are clustered by bank. t statistics in parentheses. All continuous variables are winsorized at 1% and 99%. * p < .1, ** p < .05, *** p < .01. Standard errors are clustered at bank level.

Chapter 2: Peer-to-peer FinTech Lending, Non-traditional Information, and Racial Discrimination

Abstract*

We hypothesize that racial discrimination can exist in peer-to-peer (P2P) lending even when racial information is not directly observable, and that the degree of racial discrimination decreases in the precision of credit quality signals generated from both traditional and non-traditional information. Using a large sample of loan listings from a sizeable P2P lender in the U.S., we find strong evidence that loan listings in counties with a greater proportion of minority population are associated with higher lending rates and higher loan denial rates. In cross-sectional tests, we document that racial discrimination is less pronounced when the availability of both traditional and non-traditional information is greater. Employing path analysis, we find that racial information is transmitted through the P2P platform's internal rating algorithms that utilize non-traditional information and the decision-making of platform investors.

Keywords: Discrimination; FinTech; Peer-to-peer lending; Non-traditional information; Statistical discrimination

JEL codes: G23 G28 J14 J15 K22

*This paper is coauthored with my supervisor, Kiridaran (Giri) Kanagaretnam. However, I took the lead role in the implementation and writing the paper. .

1. Introduction

For decades, the unequal and unfair treatment of minority borrowers by traditional financial institutions has been a significant concern for policymakers. The U.S. government has enacted a series of consumer protection regulations in pursuit of equal access to credit. Such regulations include, for example, the Fair Housing Act (FHA) in 1968, the Equal Credit Opportunity Act (ECOA) in 1974, the Home Mortgage Disclosure Act (HMDA) in 1975, and the Community Reinvestment Act (CRA) in 1977. However, despite the efforts by policymakers, anecdotal evidence of discrimination against minorities is still prevalent—for instance, in 2016, federal regulators filed a complaint against BancorpSouth, a Mississippi bank with $13.9 billion in total assets, for mortgage lending discrimination against African Americans and other minorities; in 2017, Wells Fargo was sued by the city of Philadelphia for imposing riskier loans on African American and Hispanic borrowers. [19]

Peer-to-peer (or P2P, FinTech) lenders match potential loan borrowers and lenders through a platform. Unlike traditional banks, P2P platforms pre-screen the loan listings by applying machine learning and big data techniques and assigning loan ratings before presenting the listings to all platform investors (e.g., Philippon 2016; Fuster et al. 2019, 2022). Investors then bid for the loan listings. P2P lending has been increasing rapidly over the past decade. For example, the share of personal loans granted by P2P lenders in the U.S. was only 5% in 2013—the ratio grew to 38% in 2018.[20] While there has been abundant evidence of racial discrimination in traditional bank lending, which documents significant differences between minority and majority borrowers in loan acceptance and loan interest rates (Black et al. 1978; Schafer and Ladd 1981; Munnell et al. 1996;

[19] Sources for the two press stories: https://www.wsj.com/articles/mississippi-bank-accused-of-mortgage-redlining-1467242094 and https://www.washingtonpost.com/news/get-there/wp/2017/05/15/philadelphia-sues-wells-fargo-for-allegedly-discriminating-against-minority-borrowers/

[20] Source: https://www.statista.com/statistics/935629/distribution-personal-loans-by-source-usa/

Ladd 1998; Begley and Purnanandam 2021), much less attention has been paid to potential racial discrimination in P2P lending.

Considering these, we examine the prevalence of racial discrimination in a P2P lending setting and explore how credit signals for lenders can influence the severity of racial discrimination. Some recent studies argue that, to the extent that P2P lenders can acquire more precise signals of borrowers' credit quality than traditional banks through the utilization of non-traditional information, P2P lending can effectively reduce the discrimination bias shown in traditional bank lending (e.g., Philippon 2019; Howell et al. 2021). However, Fuster et al. (2022) posit that minority mortgage borrowers are less likely to gain from machine learning algorithms. Indeed, Bartlett et al. (2022) document significant rate disparities between minority and majority borrowers for mortgages issued by FinTech lenders. While Bartlett et al. (2022) provide important evidence of racial discrimination in large single-family mortgage loans, we focus on much smaller unsecured personal loans that dominate P2P lending with much higher interest rates.

Moreover, our study extends Pope and Sydnor (2011) (PS), who rely on racial information identified from self-disclosed unverifiable profile pictures and find that listings with blacks in the picture are less likely to be funded and pay higher interest rates. Similar to PS, we use a sample of P2P loan listings from Prosper Marketplace (Prosper). Our study is different from PS in at least three important aspects. First, in the PS study, the racial status of borrowers is derived from pictures manually assessed by research assistants. As per Prosper, the self-reported information is not fully verified and can be "incomplete, inaccurate, or intentionally false," and thus Prosper does not encourage its use in investment decisions.[21] Our study uses the proportion of minority population at the county level as a proxy for the possible racial information potentially utilized by

[21] See, for example, Prosper's 10-K filings: https://www.prosper.com/prospectus

the platform and investors. In fact, since about 2015, Prosper loan listings no longer contain profile pictures.[22] Because borrowers still disclose their city and state of residence, which are verified by Prosper, we argue that the county-level proportion of minority population is a reasonable proxy for the platform's and investors' set of racial information about borrowers. Second, if investors do use the racial information derived from profile pictures as in the PS study, we would expect the extent of racial discrimination to be alleviated in our sample during which profile pictures are not supplied in most of the listings. In this sense, any evidence of racial discrimination in our study would suggest that racial discrimination could exist indirectly via area-level minority proportions. Third, we investigate the role of the availability and precision of non-traditional and traditional information, which is not examined in PS.

Moreover, our study extends Pope and Sydnor (2011) (PS), who uses racial information identified from self-disclosed unverified profile pictures and document that P2P loan listings with blacks in the picture are less likely to be funded and pay higher interest rates. Our study is different from PS in the following aspects. First, Prosper has experienced rapid growth between PS's sample period (from 2006 to 2007) and our sample period (from 2013 to 2019). For example, the annual listing amount of Prosper increased from $274 million in 2006 to $5.27 billion in 2015, and the number of loan applications increased from 47K listings in 2006 to 372K listings in 2015. Second, Prosper's business model has changed between the two sample periods. For instance, Prosper no longer allows for the bidding for interest rates by investors, suggesting a fundamental change in how racial information can be incorporated into interest rate determination. Also, in the PS study, the racial status of borrowers is derived from pictures manually assessed by research assistants. As

[22] We obtain this information by examining each year's Prosper 10-k filing. We find that the loan listing format with a profile picture no longer exists since the 2015 10-k filing. We also checked that (in 2022) all the loan listings on Prosper's website do not contain profile pictures..

per Prosper, the self-reported information is not fully verified and can be "incomplete, inaccurate, or intentionally false".[23] Our study uses the proportion of minority population at the county level as a proxy for racial information potentially utilized by the platform and investors. In fact, Prosper loan listings no longer contain profile pictures in most years of our sample period.[24] Because borrowers still disclose their city and state of residence, which Prosper verifies, we argue that the county-level proportion of the minority population is a reasonable proxy for the platform and investors' perception of racial background of borrowers. Third, if investors do use the racial information derived from profile pictures as in the PS study, we would expect the extent of racial discrimination to be alleviated in our sample, during which profile pictures are not supplied in most of the listings. In this sense, any evidence of racial discrimination in our study would suggest that racial discrimination could exist indirectly via county-level minority proportions. Fourth, we additionally investigate the role of the availability and precision of non-traditional and traditional information on racial discrimination, which is not examined in PS.

We examine racial discrimination in P2P lending using a large sample of 1,093,797 P2P loan listings from Prosper over the 2013 to 2019 period. We measure race at the county level with the proportion of the Latinx and/or Black population (or 'minority proportion').[25] Because the Prosper platform and its investors can only learn borrowers' racial information via their city and state (which can be well-matched to counties), we reason that our county-level measure of minority proportion can proxy for the platforms' and investors' use of racial information. Although we use a county-level proxy, we conduct our analyses at both county and individual loan listing levels.

[23] See, for example, Prosper's 10-K filings: https://www.prosper.com/prospectus

[24] We obtain this information by examining each year's Prosper 10-k filing. We find that the loan listing format with a profile picture no longer exists since the 2015 10-k filing. We also checked that (in 2022) all the loan listings on Prosper's website do not contain profile pictures.

[25] We follow Bartlett et al. (2022) and use Black or Latinx as the proxy for minority population.

For county-level analysis, we collapse the characteristics of loan listings into county averages and construct a sample of 12,600 county-year observations. We regress county-level loan denial and interest rates on county-level minority proportion and an array of loan, demographic, and macroeconomic controls. We find robust evidence that county-level minority proportion is positively associated with county-level average P2P loan denial rate and loan interest rate. This effect is also economically significant. A one standard deviation increase in county-level minority proportion is associated with a 3.3% increase in loan denial rate and a 1.1% increase in loan interest rate, respectively. The loan rate effect translates to a 17-basis-points (bps) increase in the county-level mean loan interest rate. Similarly, for loan-level analysis, we find that P2P loan listings from counties with larger minority proportions are associated with a higher likelihood of denial and higher loan interest rates.

Next, we summarize Prosper's credit pricing and funding process into three phases and use three path analyses to examine how racial discrimination transmits through the three phases. In the first phase of path analysis, we find that only 30% of the transmission of racial information to Prosper's internal credit rating is through the FICO score, suggesting that racial information is largely incorporated in the platform's "custom risk model" that is used in tandem with credit bureau scores in determining the internal ratings. In the second phase of path analysis, we find that 74% of the transmission of racial information to the final quoted interest rate is through Prosper's internal ratings, suggesting that racial information is incorporated in the "additional factors" that are used together with the internal rating in determining the final interest rate. In the final path analysis, we find that a substantial portion of the transmission of racial information to loan denial decisions is related to investors' own judgments (86%) in addition to Prosper's recommendations.

We perform several cross-sectional analyses to test our hypothesis that racial discrimination in P2P lending is less pronounced when the availability of non-traditional information is greater. First, we measure the availability of non-traditional information with a state-year-level proxy provided by the National Telecommunications and Information Administration (NTIA). This proxy measures the percentage of 15+ people who use online shopping, online financial services, and social networks at the state level each year.[26] At the county level, we find that the influence of race on the P2P loan denial rate and loan interest rate is weaker in states with more available non-traditional information. At the loan level, we find that the relationship between race and P2P loan interest rate is less pronounced for loans in states with more people engaged in online shopping, more people using online financial services, and more people using social networks. Moreover, we find that the influence of race on loan interest rate is weaker for loan listings with prior interactions with the platform, more prior on-time loan payments, and higher credit bureau scores, consistent with our prediction that racial discrimination is less pronounced for borrowers with more precise traditional credit information. In sum, the cross-sectional findings support our hypothesis that the manifested racial discrimination in loan pricing and loan acceptance decisions is less pronounced to the extent that lenders can effectively utilize the ubiquitous traditional and non-traditional information.

In addition, we perform quantile regressions on both the baseline and the cross-sectional analysis at the county level. We find that non-traditional information tends to alleviate racial discrimination the most for higher-risk borrowers, implying that higher-risk borrowers, who may not have sufficient information collected by the credit bureau, benefit the most from the utilization of non-traditional information. Last, we perform several robustness tests using alternatively constructed information proxies and the proportion of the population that is Black as the alternative

[26] The data related to state online shopping and online financial services users are available for years 2013, 2015, 2017, and 2019, while the data related to state social network users are available for years 2015, 2017, and 2019.

proxy for minority proportions. We find similar results using these alternative proxies.

We contribute to the literature in several important ways. First, our study contributes to the lending discrimination literature by being one of the first to empirically document the effect of the precision of credit quality signals acquired by P2P lenders on racial discrimination. Hence, our study is an empirical implementation of the theoretical insights of Aigner and Cain (1977) and Philippon (2019). Second, building on abundant evidence that lenders do discriminate (e.g., Schafer and Ladd 1981; Munnell et al. 1996; Ladd 1998; Fairlie et al. 2021), our paper provides new empirical insights on lending discrimination by documenting evidence of racial discrimination in P2P lending. In particular, it extends Pope and Sydnor (2011), who examine evidence of racial discrimination in P2P lending with racial information extracted from profile pictures, and Bartlett et al. (2022), who focus on single-family mortgage loans. Unlike Pope and Sydnor (2011), our paper utilizes verified borrowers' location information and uses the minority proportion at the county level as the proxy for racial information. Third, we exploit the unique business model of a P2P platform, which enables us to take a closer look at the credit pricing and funding process of the P2P platform and explore where racial discrimination occurs and how racial information is transmitted through the credit pricing and funding process. Last, we contribute to the growing FinTech lending and information economics literature by documenting the role of FinTech lenders in decreasing the information friction in the consumer credit market.

Regarding policy implications, our study corresponds to the Federal Deposit Insurance Corporation's (FDIC) vision of using big and alternative data in finance and banking to address the challenges of financial inclusion.[27] Our study implies that governments can promote the use of

---

[27] For example, during the FDIC's *Banking on Data: Great Possibilities, Great Responsibilities* webinar in 2021, Jelena McWilliams, the former Chairman of FDIC, said that FDIC is "actively engaged with banks, FinTechs, and other stakeholders on the full spectrum of issues including data analytics, artificial intelligence…in pursuit of sound

big data for the entire banking sector to minimize racial and economic discrimination.

The rest of this study is organized as follows. We synthesize relevant literature in section 2, present a simple framework and develop testable hypotheses in section 3, describe the institutional background, data, and empirical design in section 4, discuss the empirical results in section 5, and make concluding remarks in section 6.

2. Literature Review

2.1. Economic Theories of Discrimination

Theories of lending discrimination are developed based on well-established literature on economic theories of discrimination, which can be classified into "taste"-based and "statistical" discrimination models.[28] Becker (1957) attributes racial discrimination to a "taste" for discrimination, which refers to non-minorities' prejudice against certain minority groups for merely "taste" reasons. Nevertheless, more recent studies depart from Becker (1957) because they find it difficult to use "taste" to explain differences in certain economic outcomes such as wages and earnings (e.g., Arrow 1972, 1973; Phelps 1972; McCall 1973; Cain 1986;). These studies emphasize "statistical" discrimination theories and argue that the manifested discrimination is caused by information imperfections, with the assumption that the observable signals for minorities are noisier than the observable signals for non-minorities (Aigner and Cain 1977; Lundberg and Startz 1983; Cornell and Welch 1996). Under such "statistical" discrimination, decision-makers put less weight on the observable signals of minorities (e.g., Oettinger 1996; Altonji and Pierret 2001). As implied by the "statistical" discrimination theory, lenders may put less weight on the observable information of minority borrowers, and they may find it less costly

public policy," and that big data "offers the promise to provide for a more inclusive financial system that offers affordable access to financial services and products for our most vulnerable populations."

[28] A more detailed review of economics of discrimination literature is provided by Darity and Mason (1998) and Bertrand and Mullainathan (2004).

to use group characteristics, such as race, as a proxy for creditworthiness (e.g., Carr and Megbolugbe 1993; Ladd 1998).

2.2. Lending Discrimination in Traditional Financial Institutions and FinTech Lenders

Several studies, building on theories of the economics of discrimination, empirically examine discrimination in lending. Most of the early studies focus on mortgage lending. For example, using HMDA data, Black et al. (1978) find that minority borrowers are rejected more often than non-minority borrowers. Schafer and Ladd (1981) use mortgage application data available under the state law in California and New York, and document that Black applicants had significantly higher chances of loan denial than Whites. However, both studies were criticized due to the limitation of control variables available in the data. In a much more comprehensive study conducted by the Federal Reserve Bank of Boston, the researchers collected 38 more variables from lenders in Boston in addition to the HMDA dataset (Munnell et al. 1996). They document that, holding all other financial and property characteristics constant, the probability of mortgage application denial was 8.3 percentage points higher for a minority applicant.

Realizing the significant disparities in how financial institutions treat minority and non-minority borrowers, U.S. policymakers have enacted several fair lending regulations since the 1960s, forbidding the use of variables that do not determine creditworthiness.[29] However, recent studies still find evidence of discrimination in traditional financial institutions.[30] For example, studies focusing on the pre-2017-2018 financial crisis period document that Black borrowers get significantly higher mortgage rates (e.g., Ghent et al. 2014; Reid et al. 2017; Bayer et al. 2018)

[29] As Ladd (1998) and Bartlett et al. (2021, 2022) summarize, the ECOA (Section 701, March 1976) states that it is "unlawful for any creditor to discriminate against … on the basis of race, color, religion, national origin, sex or marital status, age …"; FHA (1968) puts forward similar statements except for the omission of marital status.

[30] Besides the studies discussed in this section, other studies, such as Courchane and Nickerson (1997), Cavalluzzo et al. (2002), Black et al. (2003), Blanchflower et al. (2003), Blanchard et al. (2008), and Bone et al. (2019) also document significant evidence of racial discrimination in lending with respect to mortgages, credit cards, or small business loans originated in traditional financial institutions.

and pay higher fees when the broker is White (Ambrose et al. 2021). Among more recent studies, using confidential survey data, Fairlie et al. (2021) document that it is more difficult for Black start-ups to raise external debt and that Black entrepreneurs apply for loans less often due to the greater expected chance of credit denial. Focusing on the auto loan market, Butler et al. (2022) find that the approval rates for minority applicants are 1.5 percentage points lower than for non-minority borrowers. Indeed, discrimination also exists in other aspects of the lending process (Ladd, 1998). For instance, Begley and Purnanandam (2021) find that the dilution in financial service quality induced by the CRA regulation is significantly higher in areas with more minorities.

FinTech broadly refers to the use of technological innovations in financial products and the provision of financial services (e.g., Thakor, 2020). P2P lending, considered one of the key elements of global FinTech innovations, intends to match potential loan borrowers and lenders through a platform that applies big data and machine learning techniques in the loan screening process (e.g., Philippon 2016, 2019; Fuster et al. 2019, 2022). With the rapid growth of FinTech lending in the past decade, researchers started to examine lending discrimination in the FinTech era. Philippon (2019) argues that the use of non-traditional data (such as phone bills, shopping histories, subscriptions, or browsing histories) and machine learning in consumer credit can reduce racial discrimination against minorities. However, Fuster et al. (2022) apply machine learning models to U.S. mortgages and find that minority borrowers are less likely to gain from machine learning. The empirical evidence on discrimination in P2P lending, however, is scarce. Among the pioneers, Pope and Sydnor (2011) extract borrowers' racial information from unverified profile photos of Prosper listings and document that black applicants are less likely to be funded and pay higher loan rates. However, profile photos are not verified and only exist in less than half of the listings, hence Prosper cautions against their use in investment decisions. Our study relies on verified borrowers' location information and uses the minority proportion at the county level as the proxy for racial information. We also examine the role of traditional and non-traditional information, which is not investigated in Pope and Sydnor (2011). Moreover, Bartlett et al. (2022) provide evidence of racial discrimination in the single-family mortgage sector. They document that minority borrowers pay higher risk-adjusted interest rates on GSE-securitized and FHA-insured loans. Unlike Bartlett et al. (2022), we focus on much smaller unsecured personal loans that dominate P2P lending and are subject to much less fair lending scrutiny than is focused on mortgages. We also derive our hypotheses with a framework that allows us to decompose the factors that might influence the degree of racial discrimination in lending.

3. A Simple Framework and Testable Hypotheses

We use a simple model adapted from Aigner and Cain (1977) and Philippon (2019) to develop testable hypotheses about P2P lending and racial discrimination. Based on the classical "statistical" discrimination model of Aigner and Cain (1977), Philippon (2019) proposes several models to compare the discrimination between traditional banks and P2P lenders. While we focus only on the discrimination on P2P platforms, we adopt the Philippon (2019) framework partially with our own derivations. Following Philippon (2019), we denote z as the information related to the racial status of the borrowers. We define $z = A$ for the majority group and $z = B$ for minority groups. Unlike Philippon (2019), we assume that the minority and the majority borrowers have the same credit quality for simplicity and, more importantly, *ceteris paribus* purposes.[31]

We define q as the unobservable credit quality of an individual of any racial group with a mean of $\bar{q}$ and variance $var(q)$. The precision of q is defined as $p \equiv 1/var(q)$. Our goal is to examine how discrimination-related factors can affect the prediction deviation from $\bar{q}$ (i.e., $\bar{q} -$

[31] So that the model automatically controls the true credit quality of the minority and the majority borrowers.

$E[q|z=B]$). For each borrower, we define signal $y_1 = q + \varepsilon_1$, which can be observed by all lenders, including traditional banks and P2P lenders. ε_1 represents the noise of the signal, which is normally distributed with a mean of zero and a precision of p_1. An example of y_1 is an individual's standard credit bureau score. In addition, we define another noisy signal $y_2 = q + \varepsilon_2$, which can only be observed by P2P lenders via their utilization of non-traditional data. Similarly, ε_2 is the noise and is normally distributed with a mean of zero and a precision of p_2.

Finally, following Philippon (2019), we assume that, to the extent that P2P platforms and/or investors show prior prejudice against minorities, they will perceive the average credit quality of the borrowers with a discount of δ: i.e., for biased lenders, the perceived average loan quality for minority borrowers (i.e., when $z = B$) would be $(\bar{q} - \delta)$ instead of $\bar{q}$. We argue that δ is likely positive, since in our sample the borrowers disclose city and state information, which can be used by platforms and/or platform investors to deduce some racial information.[32]

Putting all these together, the conditional expectation of the credit quality of a potential borrower given the two noisy signals and the racial group B can be expressed as:[33]

$$E[q|y_1, y_2, z=B] = (1 - \hat{\beta}_1 - \hat{\beta}_2)(\bar{q} - \delta) + \hat{\beta}_1 y_1 + \hat{\beta}_2 y_2 \qquad (1)$$

where $\hat{\beta}_1$, $\hat{\beta}_2$ are OLS estimators of β_1, β_2 in the regression model $q = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + u$. In equation (1), we follow Phelps (1972) and Aigner and Cain (1977), and model $E[q|y_1, y_2, z=B]$ with an operational regression. Detailed explanations and derivations are shown in Appendix B.

The average bias is then computed by applying the law of iterated expectations by taking averages along the dimensions of y_1 and y_2 (we still condition on B):

---

[32] We note that this prejudice, δ, can be either "taste"-based or "statistical"—i.e., lenders may apply this prejudice based on the average historical performance of minority vs. non-minority lenders. Thus, δ reflects lenders' prior beliefs.

[33] See Appendix B for the detailed derivation for the conditional and unconditional expectations of q.

$$E[q|B] = E[E[q|y_1, y_2, B]] = \bar{q} - \delta(1 - \frac{p_1}{p+p_1} - \frac{p_2}{p+p_2}) \quad (2)$$

Therefore, the racial discrimination bias:

$$b = \bar{q} - E[q|B] = \delta(1 - \frac{p_1}{p+p_1} - \frac{p_2}{p+p_2}) \quad (3)$$

In equation (3), we are particularly interested in three parameters: δ, the prior prejudice of the platform and/or investors against minorities; p_1, the precision of traditional credit quality signals, and p_2, the precision of credit quality predictions via non-traditional data used by P2P platforms through their machine learning algorithms. From this setup, we formulate the following propositions:

P1. A necessary condition for the existence of P2P platform racial discrimination is that P2P platforms and/or investors show some prior prejudice against minorities—i.e., $\delta > 0$.[34] Assuming that all signals are noisy, $\delta > 0$ becomes both a necessary and a sufficient condition for the existence of discrimination on P2P platforms.

P2. P2P platform racial discrimination is negatively related to the precision level of the standard credit score signal, i.e., $\partial b/\partial p_1 < 0$ when $\delta > 0$.

P3. P2P platform racial discrimination is negatively related to the precision level of the credit quality predictions from the use of non-traditional data, i.e., $\partial b/\partial p_2 < 0$ when $\delta > 0$.

We use the above propositions to generate two main empirical predictions. The first is about racial discrimination in P2P lending and the second is about factors that can influence racial discrimination in P2P lending via changing the precision level of signals.

First, we assume that, given that borrowers disclose some information about their race, the

[34] We do not consider the counter-intuitive case when δ is below zero—i.e., we assume that lenders do not discriminate against non-minorities.

platform and/or investors will use some of that information when making loan approval and loan pricing decisions.[35] Then, according to Proposition 1, P2P lenders will show racial discrimination against minorities. In our context, the P2P platform data we use contain information about borrowers' city and state, which can be well matched to counties and hence show some information about the borrowers' racial status.[36] Therefore, we have the following *ceteris paribus* predictions:

H1a: County-level loan interest rate and loan denial rate are positively related to the county-level proportion of the minority population.

H1b: Loan-level interest rate and the probability of loan denial are higher for loans applied in counties with higher proportions of the minority population.

Second, implied by Proposition 3, the degree of racial discrimination should be smaller when the signals acquired by P2P lenders through the utilization of non-traditional data with machine learning techniques are more precise. We reason that in the FinTech era, such signal precision can depend on three factors. The first factor is the availability of non-traditional data, because the availability of large data sets is necessary to effectively "feed" the machine learning models. (e.g., Fuster et al. 2022). The second factor is the quality of algorithms, which can be properly controlled in this study because we use data only from one P2P platform. The third factor is time, or technology development, for machine learning performance depends substantially on evolving computing power (e.g., Al-Jarrah et al. 2015). In this study, we are interested in the ubiquity of non-traditional data, and hence we have the following *ceteris paribus* cross-sectional prediction:

H2a: The racial discrimination in P2P lending is less pronounced when the availability of non-traditional data is greater.

---

[35] As discussed, such prejudice reflects lenders' prior beliefs based on historical information, or "taste," if any.

[36] As illustrated in the descriptive data, there are significant variations in minority proportions across counties.

In addition, implied by Proposition 2, the degree of racial discrimination should be smaller when the signals from traditional credit measurements are more precise. In the context of P2P lending, traditional credit measurements can include, for instance, borrowers' credit scores, employment and income information, and prior interactions with the platform, such as the existence of prior loans and information about prior payments. We expect that, like non-traditional credit signal precision, traditional credit signal precision can be greater when the available information set is larger. Therefore, we posit that, when borrowers have prior loans with the platform and make more on-time payments, lenders will perceive the traditional credit measurements precisely due to a larger information set. Similarly, we expect that when borrowers have a higher credit score, lenders will perceive the higher credit score as a positive sign for information precision because a higher credit bureau score suggests less uncertain loan payments. Therefore:

H2b: The racial discrimination in P2P lending is less pronounced when borrowers 1) have prior loans on the P2P platform, 2) have more prior on-time payments on the P2P platform, and 3) have higher credit bureau scores.

4. Institutional Background, Data, and Empirical Design

4.1. The Business Model and Credit Pricing System of Prosper[37]

Reported as one of the largest P2P lenders in the U.S.,[38] Prosper adopts a business model in which all the listed loan interest rates are pre-determined by Prosper. Since 2010, Prosper no longer allows the Dutch-auction-like bidding for interest rates. Hence, investors only bid for loan amounts, not the loan rates which are pre-set by the platform. Therefore, an investor's objective is

[37] All information about loan pricing, funding, and general business model is obtained from Prosper's Prospectus report: https://www.sec.gov/Archives/edgar/data/1416265/000141626510000555/prosperposam310d22d10.htm

[38] A 2020 market research report shows that LendingClub and Prosper are the two largest P2P lenders in the U.S. (https://mangosoft.tech/blog/top-5-peer-to-peer-lending-companies-2020-full-market-research)

to select a loan listing(s) and bid for X% of the requested loan amount (X ranges from 0 to 100) to maximize her/his utility. In addition, if a listing does not receive the bid(s) of the minimum amount required to fund, the listing will terminate. Under this business model, Prosper services the entire life-cycle of a loan and makes profits through collecting fees from borrowers.[39] Up to mid-2022, Prosper offered three financing options: personal loans, credit cards, and home equity line of credit. Because Prosper launched the credit card product in 2022 and the home equity line of credit at the end of 2019, our sample contains only personal loan listings.

The credit pricing system of Prosper works as follows. A borrower submits her/his loan listing requests to Prosper. Prosper then submits the listing to a third-party bank named WebBank, which verifies the minimum credit score requirements and performs identity and anti-fraud checks.[40] After the loan listing passes the initial screening, the borrower's credit risk will be assessed by Prosper in a 10 × 10 matrix. The first dimension of the matrix is the Prosper score, which is a score of 1 to 10 determined by a custom risk model based on possible data related to the borrowers.[41] The second dimension of the matrix is credit bureau scores divided into 10 groups. After a "base rating" is assigned to a listing via the 10 × 10 matrix, it is adjusted by the loan term and information about previous Prosper loan(s). The adjusted rating, an internal rating developed by Prosper, is called the "Prosper Rating." Finally, Prosper sets the borrower loan rates based on the Prosper Rating and additional factors such as "group affiliations, the general economic environment, and competitive conditions."[42]

Importantly, borrowers on the platform do not disclose their racial information directly—

[39] This includes a one-time loan listing fee, late-payment fees, and prepayment fees, as per the Prospectus report.
[40] Prosper initially requires a minimum credit score of 640, which is different from LendingClub that requires a debt-to-income ratio above 0.35 and a credit score above 660 (e.g., Tang 2019).
[41] Prosper claims that the "custom risk model" uses historical Prosper data and is built on the Prosper borrower population, i.e., the model inputs all historical Prosper loan records and makes predictive analysis.
[42] A more illustrative version of the Prosper credit rating system is shown in Table 1, Panel A and Panel B.

however, they disclose their city and state of residence, which are the only source of data that the platform and investors can use to deduce the possible racial status of a borrower. Hence, we use the county-level proportion of minority population as the proxy for the platform's and investors' set of racial information. Under Prosper's pricing model, we expect that such racial information is incorporated in the custom risk model and the processing of additional factors.

4.2. Sample

We retrieve detailed P2P listings data from Prosper's website, including all loan applications from 2013 to 2019. A total of 1,538,451 loan listings (including both funded and unfunded loans) are downloaded from the Prosper website. The dataset contains loan listings from 49 states (including the District of Columbia and excluding Alaska and Iowa) and 2,393 counties in the U.S. The dataset includes variables such as the listing creation/funded date, loan amount, loan term, FICO credit scores, self-disclosed income, self-disclosed employment length, city and state of residence, and information about prior interactions with the platform. Following Tang (2019), we match the city and state of each loan listing to U.S. counties and delete unmatched observations. Then, we match the county of each observation to county-level racial information, including the proportion of Black and/or Latinx population and the proportion of Black population. We also match the county of each listing to county-level personal income, per capita GDP, Housing Price Index (HPI), unemployment rate, percentage of the population that is married, and percentage of adults with post-secondary education. To reduce the influence of extreme values, we winsorize the top and bottom 1% of the variables *DECTY*, *RATECTY*, and *RATE*. Last, we drop observations with missing values on any loan-level characteristics. These data screening steps result in a final sample of 1,093,797 loan listings, of which 825,975 are successfully funded loans, 259,731 are cancelled or expired listings, and 8,091 are listings under other status (e.g., pending or withdrawn). For county-level analysis, we also collapse the loan-level data into county-level data and obtain 12,600 county-year observations after deleting county-level missing values.

[Insert TABLE 1 here]

4.3. Measure for the Availability of Non-traditional Information

We obtain data for the proxy for the availability of non-traditional information from the National Telecommunications and Information Administration (NTIA) Internet Use Survey.[43] Considered as an important source for policymakers and researchers on internet use and digital divide, the survey includes 50 questions to 50,000 households in the U.S. at the state level (*NTIA* 2020). The NTIA survey data has been used in studies such as Van Dijk and Hacker (2003) and Van Dijk (2006), and is widely acknowledged in studies examining issues such as the digital divide (e.g., Ono and Zvodny 2002; Robinson et al. 2003) and online shopping behaviors (e.g., Klopping and McKinney 2004; Porter and Donthu 2006; Hausman and Siekpe 2009).

For our study, the variables of interest are related to various online activities, such as the percentage of 15+ persons doing online shopping (*ECOM%*), the percentage of 15+ persons using online financial services such as banking, investing, and paying bills (*OFS%*), and the percentage of 15+ persons using online social networks (*SNW%*). The data is shown in Table 2. For both county- and loan-level analyses, we construct an index variable (*INFO%*) as the sum of *ECOM%*, *OFS%*, and *SNW%* to proxy for the availability of non-traditional information for a state-year. In regression analysis, we use an indicator variable (*INFO*), which equals 1 if *INFO%* is above the median value of all state-years, and 0 otherwise. For loan-level analyses, we also use the indicator variables *ECOM*, *OFS* and *SNW*, which are defined in detail in Appendix A. Since *ECOM%* and *OFS%* are available only for the years 2013, 2015, 2017, and 2019, and *SNW%* is only available

[43] Link to the data source: https://www.ntia.doc.gov/data/explorer#sel=socialNetworkUser&disp=map

for years 2015, 2017, and 2019, we only use observations in the corresponding years in the cross-sectional tests.

[Insert TABLE 2 here]

4.4. Empirical Design

To test H1 (county-level and loan-level tests for presence of racial discrimination in lending), we estimate the following cross-sectional regressions as our baseline models, with standard errors of the estimates clustered by year and county:

$$DECTY \text{ or } RATECTY = \alpha_0 + \alpha_1 MP + \alpha_2 VC + \alpha_3 W + Year\_FE + State\_FE + \varepsilon \quad (4)$$

$$DE \text{ or } RATE = \beta_0 + \beta_1 MP + \beta_2 VL + \beta_3 W + Year\_FE + State\_FE \text{ or } County\_FE + \varepsilon \quad (5)$$

The dependent variables for equation (4) are the county-level loan denial rate (*DECTY*) or the county-level average loan interest rate (*RATECTY*). The dependent variables for equation (5) are the loan-level indicator variable *DE*[44] or the loan-level interest rates (*RATE*). When *DE* is used as the dependent variable, a logit regression model is used to estimate equation (5). *VC* is a vector of the average value of loan characteristics at the county level, and *W* is a vector of county-level demographic and economic characteristics. *VL* is a vector of individual loan-level loan characteristics. *Year_FE* is year fixed effects. For the baseline regressions, we include state fixed effects (*State_FE*) so that the effects of state-level financial regulations do not affect our results. We also apply county fixed effects (*County_FE*) for the loan-level regressions in equation (5). H1a hypothesizes that the county-level minority proportion is positively associated with county-level average loan denial rate and average loan rate; hence we expect α_1 to be positive. H1b hypothesizes that loan listings in counties with a higher minority proportion tend to have a higher loan rate and are more likely to be denied; hence we expect β_1 to be positive.

[44] The variable equals 1 when the loan listing is cancelled by the platform or expired, and 0 otherwise.

We select control variables for loan characteristics based on the credit pricing system of Prosper. As discussed in section 4.1, after the initial screening for credit scores and identity and anti-fraud checks, Prosper uses a 10 × 10 matrix to determine the baseline interest rates. We use the FICO score (*SCORE*) to control for the credit score dimension of the matrix, and we use the borrower's self-disclosed income range (*INCSELF*), self-disclosed employment tenure (*EMPLEN*), and several variables related to prior interactions with the platforms (*PRIOR*, *PRIORA*, *PRIORP*, *PRIORO*, *PRIORL*) to control for the custom risk model dimension of the matrix. Because Prosper adjusts the baseline interest rates based on the loan amount and loan term, we control for the requested loan amount (*AMT*) and loan term (*TERM*).

In addition, since Prosper determines the final interest rates by incorporating some additional factors such as "group affiliations, the general economic environment, and competitive conditions," we also include several county-level control variables. We control for county-level personal income per capita (*INCOME*) because it is associated with discrimination in lending (e.g., Ferguson and Peters 1995; Tootell 1996). We control for the county-level proportion of the population with post-secondary education (*EDU*) because racial discrimination can be more salient for borrowers with low education (Cheng et al. 2015). We control for the county-level proportion of the population that is male (*MALE*) because gender is documented as one significant factor in P2P lending racial discrimination (e.g., Chen et al. 2017) and bank lending racial discrimination (e.g., Cozarenco and Szafarz 2018). We also control for the county-level proportion of the population that is married (*MARRIED*) because Tootell (1996) finds that being married significantly increases the probability of being approved for a loan. Finally, to control for the macroeconomic conditions, we control for county-level unemployment (*UNEMP*), county-level housing price index (*HPI*), and the natural logarithm of county-level GDP per capita (*GDP*).

5. Results

5.1. Descriptive Statistics

Table 3 presents the descriptive statistics and correlations. Table 3 Panel A reports the descriptive statistics for regression variables at the county level. The mean (median) county-level percentage of African American and/or Latinx population is 19.1% (12.3%), consistent with Bartlett et al. (2022) who report a minority borrower proportion of 10% for GSE loans and 24.6% for FHA loans; the mean (median) county-level percentage of African American population is 10.1% (3.5%); the mean (median) county-level P2P loan denial rate is 28.3% (25%)[45], and the mean (median) county-level P2P loan interest rate is 15.68% (15.33%). In addition, the mean (median) county-level percentage of P2P loans that have prior borrowing records on the platform is 22.6% (20.4%), suggesting that on average, 22.6% of the loans are from repeat borrowers. Table 3 Panel B reports the descriptive statistics for regression variables at the loan level. The mean probability that a loan is denied is 23.7%, which is calculated by taking the mean of the indicator variable *DE* across all loan listings; the mean (median) individual loan interest rate is 14.94% (13.54%); the mean (median) individual loan amount is $13,630 ($12,000) U.S. dollars; the mean (median) loan term is 43.2 (36) months, and the mean (median) FICO score is 704 (690). Tang (2019), who uses data from LendingClub loans from 2009 to 2012, reports a mean loan amount of $13,104 and mean FICO score of 652 for all loan listings, and a mean interest rate of 13.3% for funded loans, suggesting that applicants' characteristics of these two platforms are comparable despite different sample periods. In addition, Table 3 Panel C reports Spearman correlations between the variables in our analyses at the county level. As predicted by H1a, we observe a significant and negative correlation of between loan interest rates and the minority proportion.

[45] The mean (median) of loan denial rate is comparable to Munnell et al. (1996) who report a loan rejection rate of 20% for Whites and 28% for minorities.

[Insert TABLE 3 here]

Table 4 presents the univariate analysis at the loan level. We divide the loan listing sample into quartiles based on the minority proportion (*MP*) and compare loan characteristics between the highest and the lowest quartiles. As reported in Table 4, compared to those with the lowest quartile of *MP*, the loan listings located in counties with the highest quartile of *MP* have a higher probability of loan denial and a higher loan interest rate, which is consistent with our baseline predictions in H1b. Because tables 3 and 4 only present pairwise univariate correlations, we defer inferences to the multivariate tests reported in the following sections.

[Insert TABLE 4 here]

5.2. Main Empirical Results

5.2.1. Evidence of Racial Discrimination in Lending

In this section, we report the results for the test of H1a, which examine the association between the county-level minority proportion and county-level loan denial/interest rates. In all models, we cluster the standard errors at the year and county level. In Table 5 Panel A Column (1), we regress county-level loan denial rates on the minority proportion without state fixed effects, and in Column (2), we report the results including both year and state fixed effects. In both columns, we report a positive and statistically significant coefficient on *MP*, which indicates that the county-level average loan denial rates (i.e., the percentage of listings that are denied) are positively associated with the county-level minority proportion. This relation is also economically significant. Using Column 1 as an illustration, a one standard deviation increase in *MP* is associated with a 3.3% increase in *DECTY*.[46] In Table 5 Panel A Column (3), we regress county-level average loan

[46] The impact of a one standard deviation increase in *MP* on *DECTY* is computed as 0.052 (the coefficient of *MP*) × 0.179 (the sample standard deviation of *MP*) ÷ 0.283 (the sample mean of *DECTY*) × 100% = 3.3%. Analogously, the economic significance in Column 2 is 2.6%.

interest rates on the minority proportion without state fixed effects, and in Column (4) we report the results including both year and state fixed effects. In both columns, we report a positive and statistically significant coefficient on *MP*, which indicates that the county-level average loan interest rates increase with county-level minority proportion. This association is also economically significant. As Column (3) illustrates, a one standard deviation increase in *MP* is associated with a 1.1%, or 17-bps increase in the county-level mean loan interest rate.[47]

The signs of the coefficients of the control variables are generally consistent with the credit pricing criteria of Prosper and prior studies. We document that the listing amount is negatively associated with loan interest rates, consistent with Prosper permitting more funding for borrowers with better credit ratings. We also report a positive relation between loan interest rates and loan term, consistent with a normal term structure of interest rates. Likewise, as expected, loan rates are negatively related to FICO scores and negatively related to borrowers' self-disclosed income range. Moreover, we find that interest rates are positively related to the number of borrowers' currently active Prosper loans, indicating that Prosper may be concerned about the borrowing capacity of borrowers who actively hold more Prosper loans. In addition, consistent with Cheng et al. (2005), we find that county-level education is negatively associated with both loan denial rates and loan interest rates with large effect magnitudes. We also find that the proportion of the population that is male is positively associated with loan denial rates, which is consistent with Chen et al. (2017). However, unlike Chen et al. (2017) who find that females tend to pay higher interest rates, we document that the county-level male population is positively related to county-level interest rates.

[Insert TABLE 5 Panel A here]

[47] The impact of a one standard deviation increase in *MP* on *RATECTY* is computed as 0.954 (the coefficient of *MP*) × 0.179 (the sample standard deviation of *MP*) ÷ 15.680 (the sample mean of *RATECTY*) × 100% = 1.1%, which is equivalent to a 1.1% × 15.68 = 17-bps increase in mean county-level interest rate. Analogously, the economic significance in Column (4) is 0.5%, or 8-bps increase in mean county-level interest rate.

Moreover, we report the results for the test of H1b, which examine the association between the county-level minority proportion and individual loan level interest rates/probability of loan denial. In Table 5 Panel B Column (1), we regress loan-level quoted interest rates on the county-level minority proportion with loan-level controls and year and state fixed effects, and in Column (2), we report the results with full controls. In both columns, we report a positive and statistically significant coefficient on *MP*, which indicates that loans that originated in counties with greater minority proportion tend to have higher quoted interest rates. Column (1) illustrates that a one standard deviation increase in *MP* is associated with a 0.7% increase in quoted loan interest rates, or about 10-bps increase in mean interest rates.[48] In Table 5 Panel B Column (3), we use a logit model and regress the indicator variable of loan denial on the county-level minority proportion with loan-level controls and year and state fixed effects, and in Column (4), we report the results with full controls. In both columns, we report a positive and statistically significant coefficient on *MP*, suggesting that loans that originated in counties with greater minority proportion are more likely to be denied. Last, as a robustness check, in columns (5) and (6), we re-estimate columns (1) and (2) with year and county fixed effects and find statistically significant (though with lower t stats) results.[49]

[Insert TABLE 5 Panel B here]

5.2.2. Path Analyses – Where Does Discrimination Come From?

As discussed in detail in section 4.1, there are three phases during the funding process of Prosper loan listings. In the first phase, an internal Prosper rating is generated by a matrix

---

[48] The impact of a one standard deviation increase in *MP* on *RATE* is computed as 0.55 (the coefficient of *MP*) × 0.179 (the sample standard deviation of *MP*) ÷ 14.935 (the sample mean of *RATE*) × 100% = 0.6%, which is equivalent to a 0.6% × 14.935 = 10-bps increase in mean loan-level interest rate. Analogously, the economic significance in Column (4) is 0.58%, or 8.7-bps increase in mean loan-level interest rate.

[49] We do not perform re-estimations of columns (3) and (4) with county fixed effects because the large number of counties led to a non-convergence problem of coefficient estimation in the logit model.

consisting of two dimensions: the credit bureau scores and a custom risk model that incorporates all borrowers' historical data on the platform. In the second phase, the final quoted interest rates are determined by the internal Prosper rating and other factors such as group affiliations, the general economic environment, and competitive conditions. In the last phase, investors bid for the loan amount. Therefore, in this section, we aim to examine the potential source of discrimination in loan approval decisions and pricing. We employ path analyses to examine the sources of bias in each of the three phases summarized above.

5.2.2.1 Path Analysis 1

In the first path analysis, we examine whether racial discrimination is present in the determination process of the internal Prosper ratings. We test how much of the relation between race and the internal Prosper rating is attributable to the credit bureau score. We estimate the following structural equation model to test the mediating role of the credit bureau score:

$$SCORE = b_0 + b_1 MP + b_2 SCONTROLS + \varepsilon \qquad (6)$$

$$PRATE = c_0 + c_1 MP + c_2 SCORE + c_3 PCONTROLS + \varepsilon \qquad (7)$$

In equation (6), we regress the credit bureau score (*SCORE*) proxied by FICO scores on minority proportion (*MP*) and control variables that can be potentially related to *SCORE*, including *INCSELF*, *EMPLEN*, *INCOME*, *EDU*, *MALE*, *MARRIED*, *UNEMP*, *HPI*, *GDP*, and year dummies. In equation (7), we regress the internal Prosper rating (*PRATE*) on *MP*, *SCORE*, and all the control variables that are included in the baseline model in equation (5) except state fixed effects. The path coefficient c_1 is the magnitude of the direct path and $b_1 * c_2$ is the magnitude of the indirect path. The indirect path suggests the portion of internal Prosper rating discrepancy between minorities and non-minorities that can be attributable to credit bureau ratings. We are interested in the direct path—a significant and economically meaningful direct path implies that, *ceteris paribus*, racial discrimination may exist in the custom risk model that the platform uses to derive the Prosper score, the second dimension (credit score is the first dimension) when calculating the internal Prosper rating.

Table 6 Panel A reports the direct and indirect path coefficients. The indirect relation between *MP* and *PRATE* through *SCORE* is -0.072 (t = -12.60) and is about 30% of the total effect, suggesting that a significant portion of Prosper rating differences comes from credit bureau scores. The direct relation between *MP* and *PRATE* is -0.171 (t = -20.49) and is about 70% of the total effect, indicating that racial information is incorporated in the custom risk model in addition to all the other control variables included in the model.

5.2.2.2 Path Analysis 2

In the second path analysis, we examine whether racial discrimination is present in the second phase of the Prosper funding process, in which the internal Prosper ratings are considered together with additional factors in determining the final loan interest rates. We estimate the following structural equation model to test the mediating role of Prosper ratings:

$$PRATE = b_0 + b_1 MP + b_2 CONTROLS + \varepsilon \qquad (8)$$

$$RATE = c_0 + c_1 MP + c_2 PRATE + c_3 CONTROLS + \varepsilon \qquad (9)$$

In equation (8), we regress the Prosper rating (*PRATE*) on minority proportion (*MP*) and all the control variables that are included in the baseline model in equation (5) except state fixed effects. In equation (9), we regress the quoted loan interest rate (*RATE*) on *MP*, *PRATE,* and all the control variables that are included in the baseline model in equation (5) except state fixed effects. The indirect path (b_1 * c_2) suggests the proportion of race-related loan interest rate discrepancy attributable to the internal Prosper rating. In this test, we are interested in the direct path (c_1)—a significant and economically meaningful direct path implies that, *ceteris paribus*, racial discrimination may exist in the "additional factors" that the platform uses to calculate the final interest rates.

Table 6 Panel B reports the direct and indirect path coefficients. The indirect relation between *MP* and *RATE* through *PRATE* is 1.870 (t = 32.82) and is about 74% of the total effect, suggesting that a significant portion of loan interest rate differences comes from the internal Prosper ratings. The direct relation between *MP* and *RATE* is 0.666 (t = 44.56) and is about 26% of the total effect, indicating that racial information may be incorporated in the "additional factors" in the second phase, above and beyond all the other control variables included in the model.

5.2.2.3 Path Analysis 3

In the third path analysis, we examine whether racial discrimination is present in the final phase of the Prosper funding process, in which investors decide to bid for the listing amounts. We assume that investors rely on two general factors when deciding whether (and by how much) to bid for a loan listing, namely, the Prosper recommendations (i.e., the Prosper rating) and their own judgments. We estimate the following structural equation model to test the mediating role of Prosper ratings:

$$PRATE = b_0 + b_1 MP + b_2 CONTROLS + \varepsilon \quad (10)$$

$$DE = c_0 + c_1 MP + c_2 PRATE + c_3 CONTROLS + \varepsilon \quad (11)$$

In equation (10), we regress the Prosper rating (*PRATE*) on minority proportion (*MP*) and all of the control variables that are included in the baseline model in equation (5) except state fixed effects. In equation (11), we use a linear probability model and regress the loan denial indicator variable (*DE=1*) on *MP*, *PRATE*, and all of the control variables that are included in the baseline model in equation (5) except state fixed effects. The indirect path ($b_1 * c_2$) suggests the proportion of race-related loan denial probability discrepancy attributable to the Prosper recommendation.

The direct path (c_1) indicates the proportion of race-related loan denial probability discrepancy due to investors' own judgments.

Table 6 Panel C reports the direct and indirect path coefficients. The indirect relation between *MP* and *DE* through *PRATE* is 0.005 (t = 21.87) and is about 14% of the total effect, suggesting that only a small portion of loan denial probability discrepancy comes from the internal Prosper ratings. The direct relation between *MP* and *DE* is 0.031 (t = 7.83) and is about 86% of the total effect, indicating that racial information may be significantly incorporated in investors' own judgments. This implies that investors may refer to the racial-related information of loan listings beyond Prosper's loan ratings when making the final funding decisions.

A caveat for the third path analysis is that it does not incorporate Prosper's use of "additional factors" in phase two. Hence, we replace the path variable (*PRATE*) in equation (10) with the quoted interest rate (*RATE*), assuming that the *RATE* should include all the information about Prosper's recommendation of loan listings. Since *RATE* contains more information, we expect that the indirect path should be stronger in this alternative specification. Table 6 Panel D reports the expected results: The indirect relation between *MP* and *DE* through *RATE* is 0.006 (t = 26.40) and is about 20% of the total effect. The direct relation is 0.024 (t = 7.81) and is about 80% of the total effect. As expected, the indirect relation is stronger in this alternative specification of path analysis 3.

[Insert TABLE 6 here]

5.2.3. Cross-sectional Analysis: The Effects of Non-traditional Information

In this section, we report the results for the test of H2a, which examines the cross-sectional differences in racial discrimination between loan listings in areas with more available non-traditional information and in areas with less available non-traditional information. In Table 7

Column (1), we regress the county-level loan denial rate (*DECTY*) on the interaction between our state-level proxy for the availability of non-traditional information (*INFO*) and the minority proportion (*MP*) with all control variables and year and state fixed effects, and in Column (3), we repeat the regression in Column (1) by replacing *INFO* with the alternative proxy (*INFO_A*). In both columns, we find negative and statistically significant coefficients on *INFO* × *MP* and *INFO_A* × *MP*, suggesting that the relation between county minority proportion and county average loan denial rate is less pronounced in states with more available non-traditional information. In addition, in Table 7 Column (2), we regress the county-level average loan interest rates on *INFO* × *MP*, and in Column (4), we repeat the regression in Column (2) by replacing *INFO* with *INFO_A*. In both columns, we report negative and statistically significant coefficients on *INFO* × *MP* and *INFO_A* × *MP*, indicating that the relation between county minority proportion and county average loan interest rate is less pronounced in states with more available non-traditional information. In Table 7 columns (5) to (8), we repeat the regressions in columns (1) to (4) by replacing state fixed effects with county fixed effects, and we find similar results. Overall, the results reported in Table 7 corroborate Proposition 3 of our framework and H2a, indicating that non-traditional information plays an important role in alleviating racial discrimination on P2P platforms.

[Insert TABLE 7 here]

To gain additional insights into how different types of non-traditional data affect racial discrimination, we examine how the three components of the proxy for the availability of non-traditional information can affect the relation between P2P loan interest rate and race on the loan level. In Table 8 Column (1), we regress the loan level interest rate (*RATE*) on the interaction term between county minority proportion (*MP*) and our proxy for the availability of non-traditional information (*INFO*). As expected, we find a negative and statistically significant coefficient on *INFO* × *MP*. In Column (2), we regress *RATE* on the interaction term between *MP* and *ECOM*, the state-level percentage of 15+ persons who do online shopping. We find a negative and statistically significant coefficient on *ECOM* × *MP*, suggesting that lending racial discrimination is less pronounced in states with more people doing online shopping. In Column (3), we regress *RATE* on the interaction term between *MP* and *OFS*, the state-level percentage of 15+ persons who use online financial services. We find a negative and statistically significant coefficient on *OFS* × *MP*, suggesting that lending racial discrimination is less pronounced in states with more people using online financial services. In Column (4), we regress *RATE* on the interaction term between *MP* and *SNW*, the state-level percentage of 15+ persons who use social networks. We find a negative and statistically significant coefficient on *SNW* × *MP*, suggesting that lending racial discrimination is less pronounced in states with more people using social networks. In Table 8 columns (5) to (8), we repeat the regressions in columns (1) to (4) by replacing state fixed effects with county fixed effects, and we find similar results.[50] Overall, the results reported in Table 8 suggest that each of the components of our proxy for the availability of non-traditional data can play a significant role in affecting the relationship between loan interest rates and race.

[Insert TABLE 8 here]

5.2.4. Additional Cross-sectional Analysis

In this section, we report the results for the test of H2b, in which we predict that racial discrimination in P2P lending is less pronounced when lenders perceive that the traditional credit measurements of borrowers are more precise. We test whether racial discrimination in P2P lending

[50] As Table 8 reports, when we replace state dummies with county dummies the significance level of coefficients of the interaction terms decreases. In fact, the coefficient on *SNW* × *MP* is insignificant with county fixed effects. The other two (*ECOM* × *MP* and *OFS* × *MP*) are significant at the α=0.1 and α=0.05 level, respectively.

is less pronounced when borrowers have more prior loans, make more prior on-time payments, and have higher credit scores. In Table 9 Column (1), we regress the loan-level interest rate on the interaction between minority proportion (*MP*) and an indicator variable that equals 1 if the borrower has prior loans with the platform (*PRIOR*). We find that the coefficient on the interaction term is statistically significant (t = -5.44), suggesting that when borrowers have prior loans on the platform, the relationship between *MP* and *RATE* is less pronounced. Similarly, Column (2) reports a negative and statistically significant coefficient on the interaction term *PRIORO_High* × *MP* (t = -3.32), indicating that when borrowers have more prior on-time payments, the relationship between *MP* and *RATE* is less pronounced. In addition, Column (3) documents a negative and statistically significant interaction term *SCORE_High* × *MP,* suggesting that for those borrowers who have higher credit bureau scores, the relationship between *MP* and *RATE* is less pronounced. While untabulated, we also perform all the tests in Table 9 with county fixed effects and find similar results. Overall, the results reported in Table 9 imply that when borrowers' credit information is perceived as more precise by the platform, the need to use race as a proxy for credit worthiness is reduced, corroborating H2b and Proposition 2 of our framework.

[Insert TABLE 9 here]

5.3. Additional Insights from Quantile Regression Results

Economists often use quantile regression to examine how an explanatory variable is associated with the entire distribution, rather than the mean, of the response variable (e.g., Angrist et al. 2006; Lemieux 2008). In this section, we explore the following two issues: a) is county-level minority proportion positively associated with the interest rate and denial rate gap between high-risk and low-risk borrowers?, and b) does the relation in a) change with the increasing availability of non-traditional information?

To address question a), we test how conditional quantiles of county-level loan interest rates and loan denial rates are associated with county-level minority proportion. Table 10 Model (1) reports how *MP* is associated with the 25th, median, and 75th quantiles of *RATECTY*, and Model (3) reports how *MP* is associated with the 25th, median, and 75th quantiles of *DECTY*. We also report the corresponding OLS estimates for each model with the same control variables. In both models, we find that county-level minority proportion is positively associated with all quantiles of *RATECTY* and *DECTY* to an extent similar to that in the OLS estimates. Our results indicate that the interest rate and denial rate gaps do not widen in counties with greater minority proportion.

To address question b), we check how the interaction between the proxy for non-traditional information availability and minority proportion (*INFO* × *MP*) is related to the conditional quantiles of *RATECTY* and *DECTY*. Table 10 Model (2) reports how *INFO* × *MP* is associated with the 25th, median, and 75th quantiles of *RATECTY*, and Model (4) reports how *INFO* × *MP* is associated with the quantiles of *DECTY*. In these two models, we find that the estimated coefficients of *INFO* × *MP* are much larger for higher quantiles of *RATECTY* and *DECTY*. These results indicate that non-traditional information tends to alleviate lending racial discrimination the most for higher risk borrowers. We consider this finding insightful because it implies that higher risk borrowers, who may not have sufficient information collected by standard credit bureau, may benefit the most from P2P platforms' utilization of non-traditional information.

[Insert TABLE 10 here]

5.4. Additional Robustness Checks

We perform two more robustness checks for our main hypothesis (H1) and our cross-sectional predictions involving non-traditional information (H2a). Table 11 reports the loan-level cross-sectional analysis of the effect of non-traditional information on the relation between the minority proportion and loan-level interest rate, with the alternative construction of the proxies for non-traditional information and components of the non-traditional information. We report similar results as in Table 8 columns (1) to (4). Because the alternative proxies of non-traditional information are constructed as the rank within years, the results reported in Table 9 suggest that our main cross-sectional findings are not driven by different years. In addition, Table 12 columns (1) and (2) repeat the baseline county-level regressions in Table 5 columns (2) and (4), but use a different proxy for minority population (*AA*, or the proportion of African Americans population, instead of *MP*). We find statistically significant and economically meaningful coefficients in both models. In Table 12 columns (3) and (4), we repeat the cross-sectional analysis at the county-level using *AA* as the alternative proxy for *MP*, and we find similar statistically significant results. Finally, we confirm that with *AA* as an alternative proxy for *MP*, we find similar results in loan-level regressions in Table 12 columns (5) and (6).

[Insert TABLE 11 and TABLE 12 here]

6. Concluding Remarks

Realizing the abundant anecdotal and empirical evidence of racial discrimination in traditional bank lending, policymakers have enacted an array of fair lending regulations since the 1960s, forbidding the use of variables that do not determine creditworthiness. However, recent anecdotal and empirical evidence suggests that racial discrimination still exists. For example, a well-known study by the Federal Reserve Bank of Boston collects 38 more variables in addition to the HMDA dataset and documents that the probability of loan denial for a minority applicant was 8.3 percentage points higher than that for a non-minority applicant (Munnell et al. 1996).

Considering the rapid development of P2P lending and the lack of empirical evidence of racial discrimination in a P2P lending setting, especially regarding small and unsecured personal loans, we examine racial discrimination in a P2P lending setting. With the help of a model that outlines the conditions under which the degree of racial discrimination in P2P lending can vary, we posit that racial discrimination can exit in P2P lending and that the degree of racial discrimination decreases in the precision of credit quality signals generated from both traditional and non-traditional information sources. Using a large sample of 1,093,797 loan listings from a sizeable P2P lender in the U.S. from 2013 to 2019, we find strong evidence that loan listings in counties with more minorities have higher interest rates and higher probabilities of loan denial. In cross-sectional tests, we find that racial discrimination is less pronounced when the availability of both traditional and non-traditional information is higher.

Notably, Pope and Sydnor (2011) document the first empirical evidence of racial discrimination in P2P lending based on information extracted from unverified profile pictures. Bartlett et al. (2022) provide the first empirical evidence of racial discrimination in FinTech lending focusing on large single-family mortgage loans. Our study is different from these two studies and extends the literature in three aspects. First, unlike Pope and Sydnor (2011), we rely on verified borrowers' location information and use the county-level proportion of minority population as the proxy for racial information. Considering that Prosper listings no longer contain profile photos in most of our sample years, we argue that our proxy for racial information mimics the racial information set of the platform and investors. Second, we derive our hypotheses from a simple framework adopted from Aigner and Cain (1977) and Philippon (2019) that allows us to decompose the factors that could influence the degree of racial discrimination in P2P lending. In this sense, we examine the role of the availability and precision of both traditional and non-traditional credit information in affecting the degree of racial discrimination. Third, we take a close look at the credit pricing and funding process of the P2P platform and find that racial information transmits through the platform's internal rating algorithms that utilize non-traditional information and that loan denial decisions are largely due to platform investors' own judgments.

Our study contributes to the economics of discrimination and FinTech lending literature by being one of the first to document the effects of the precision of different credit quality signals on P2P lending racial discrimination. Also, our study exploits the unique features of the P2P platform we use and examines how racial discrimination is transmitted through the credit pricing and funding process. Our study has important policy implications for policymakers who can promote the use of big data for the entire banking sector to reduce racial and economic discrimination.

References

Aigner, D. J., and G. G. Cain. 1977. Statistical theories of discrimination in labor markets. *ILR Review* 30 (2): 175–187.

Al-Jarrah, O. Y., P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, and K. Taha. 2015. Efficient machine learning for big data: A review. *Big Data Research* 2 (3): 87–93.

Altonji, J. G., and C. R. Pierret. 2001. Employer learning and statistical discrimination. *Quarterly Journal of Economics* 116 (1): 313–350.

Ambrose, B. W., J. N. Conklin, and L. A. Lopez. 2021. Does borrower and broker race affect the cost of mortgage credit? *The Review of Financial Studies* 34 (2): 790–826.

Angrist, J., V. Chernozhukov, and I. Fernández-Val. 2006. Quantile regression under misspecification, with an application to the US wage structure. *Econometrica, 74*(2): 539-563.

Arrow, K. J. 1972. Models of job discrimination. In A. H. Pascal, ed., *Racial Discrimination in Economic Life*, 83–102. Lexington: D.C. Heath.

Arrow, K. J. 1973. The theory of discrimination. In *Discrimination in Labor Markets, edited by O. Anshenfelter and A. Rees*, 3–33.

Bartlett, R., A. Morse, N. Wallace, and R. Stanton. 2021. Algorithmic discrimination and input accountability under the Civil Rights Acts. *Berkeley Technology Law Journal*, 36: 675-736.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace. 2022. Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics* 143 (1): 30–56.

Bayer, P., F. Ferreira, and S. L. Ross. 2018. What drives racial and ethnic differences in high-cost mortgages? The role of high-risk lenders. *The Review of Financial Studies* 31 (1): 175–205.

Becker, G. S. 1957. *The economics of discrimination*. Chicago: The University of Chicago Press.

Begley, T. A., and A. Purnanandam. 2021. Color and credit: Race, regulation, and the quality of financial services. *Journal of Financial Economics* 141 (1): 48–65.

Bertrand, M., and S. Mullainathan. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review* 94 (4): 991–1013.

Black, H. A., R. L. Schweitzer, and L. Mandell. 1978. Discrimination in mortgage lending. *The American Economic Review* 68 (2): 186–191.

Black, H. A., T. P. Boehm, and R. P. DeGennaro. 2003. Is there discrimination in mortgage pricing? The case of overages. *Journal of Banking and Finance* 27 (6): 1139–1165.

Blanchard, L., B. Zhao, and J. Yinger. 2008. Do lenders discriminate against minority and woman entrepreneurs? *Journal of Urban Economics* 63 (2): 467–497.

Blanchflower, D. G., P. B. Levine, and D. J. Zimmerman. 2003. Discrimination in the small-business credit market. *Review of Economics and Statistics* 85 (4): 930–943.

Bone, S. A., G. L. Christensen, J. D. Williams, S. Adams, A. Lederer, and P. C. Lubin. 2019. Shaping small business lending policy through matched-pair mystery shopping. *Journal of Public Policy & Marketing* 38 (3): 391–399.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru. 2018. Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics* 130 (3): 453–483.

Butler, A. W., E. J. Mayer, and J. P. Weston. 2022. Racial disparities in the auto loan market. *The Review of Financial Studies (Forthcoming)*

Cain, G. G. 1986. The economic analysis of labor market discrimination: A survey. In O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, 1: 693–785, New York: Elsevier Sci.

Carr, J. H., and I. F. Megbolugbe. 1993. The Federal Reserve bank of Boston study on mortgage lending revisited. *Journal of Housing Research* 4 (2): 277–313.

Cavalluzzo, K. S., L. C. Cavalluzzo, and J. D. Wolken. 2002. Competition, small business financing, and discrimination: Evidence from a new survey. *The Journal of Business* 75 (4): 641–679.

Chen, D., X. Li, and F. Lai. 2017. Gender discrimination in online peer-to-peer credit lending: Evidence from a lending platform in China. *Electronic Commerce Research*, 17(4): 553-583.

Cheng, P., Z. Lin, and Y. Liu. 2015. Racial discrepancy in mortgage interest rates. *The Journal of Real Estate Finance and Economics*, 51(1): 101-120.

Cornell, B., and I. Welch. 1996. Culture, information, and screening discrimination. *Journal of Political Economy* 104 (3): 542–571.

Courchane, M., and D. Nickerson. 1997. Discrimination resulting from overage practices. In G. J. Benston, W. C. Hunter, and G. G. Kaufman, eds., *Discrimination in Financial Services*, 133–151. Boston, MA: Springer US.

Cozarenco, A., and A. Szafarz. 2018. Gender biases in bank lending: Lessons from microcredit in France. *Journal of Business Ethics*, 147(3): 631-650.

Darity, W. A., and P. L. Mason. 1998. Evidence on discrimination in employment: Codes of color, codes of gender. *Journal of Economic Perspectives* 12 (2): 63–90.

Fairlie, R., A. Robb, and D. T. Robinson. 2021. Black and white: Access to capital among minority-owned start-ups. *Management Science* 68 (4): 2377–3174.

Ferguson, M. F., and S. R. Peters. 1995. What constitutes evidence of discrimination in lending? *The Journal of Finance*, 50(2): 739-748.

Fuster, A., M. Plosser, P. Schnabl, and J. Vickery. 2019. The role of technology in mortgage lending. *The Review of Financial Studies* 32 (5): 1854–1899.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2022. Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance* 77 (1): 5–47.

Ghent, A. C., R. Hernández-Murillo, and M. T. Owyang. 2014. Differences in subprime loan pricing across races and neighborhoods. *Regional Science and Urban Economics* 48: 199–215.

Hausman, A. V., and J. S., Siekpe. 2009. The effect of web interface features on consumer online purchase intentions. *Journal of Business Research*, 62(1): 5-13.

Howell, S., T. Kuchler, D. Snitkof, J. Stroebel, and J. Wong. 2021. *Automation and racial disparities in small business lending: Evidence from the paycheck protection program*. Cambridge, MA: National Bureau of Economic Research.

Klopping, I. M., and E., McKinney. 2004. Extending the technology acceptance model and the task-technology fit model to consumer e-commerce. *Information Technology, Learning & Performance Journal*, 22(1): 35-48

Ladd, H. F. 1998. Evidence on discrimination in mortgage lending. *Journal of Economic*

Perspectives 12 (2): 41–62.

Lemieux, T. 2008. The changing nature of wage inequality. *Journal of Population Economics, 21*(1): 21-48.

Lundberg, S. J., and R. Startz. 1983. Private discrimination and social intervention in competitive labor markets. *The American Economic Review* 73 (3): 340–347.

McCall, J. J. 1973. *Income mobility, racial discrimination, and economic growth*. Lexington, MA: Lexington Books.

Munnell, A. H., G. M. B. Tootell, L. E. Browne, and J. McEneaney. 1996. Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review* 86 (1): 25–53.

Oettinger, G. S. 1996. Statistical discrimination and the early career evolution of the Black-White wage gap. *Journal of Labor Economics* 14 (1): 52–78.

Ono, H., and M. Zavodny. 2002. Race, Internet usage, and e-commerce. T*he Review of Black Political Economy*, 30(3): 7-22.

Phelps, E. S. 1972. The statistical theory of racism and sexism. *The American Economic Review* 62 (4): 659–661.

Philippon, T. 2016. *The FinTech opportunity*. Cambridge, MA: National Bureau of Economic Research.

Philippon, T. 2019. *On Fintech and financial inclusion*. Cambridge, MA: National Bureau of Economic Research.

Porter, C. E., and N. Donthu. 2006. Using the technology acceptance model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics. *Journal of Business Research*, 59(9): 999-1007.

Reid, C. K., D. Bocian, W. Li, and R. G. Quercia. 2017. Revisiting the subprime crisis: The dual mortgage market and mortgage defaults by race and ethnicity. *Journal of Urban Affairs* 39 (4): 469–487.

Robinson, J. P., P. DiMaggio, and E. Hargittai. 2003. New social survey perspectives on the digital divide. *IT & Society*, 1(5): 1-22.

Schafer, R., and H. F. Ladd. 1981. *Discrimination in mortgage lending*. MIT-Harvard Joint Center for Urban Studies. Cambridge, MA: MIT Press.

Tang, H. 2019. Peer-to-peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies*, 32(5): 1900-1938.

Thakor, A. V. 2020. Fintech and banking: What do we know? *Journal of Financial Intermediation* 41: 100833

Tootell, G. M. B. 1996. Can studies of application denials and mortgage defaults uncover taste-based discrimination? Working Papers 96-10, *Federal Reserve Bank of Boston*.

Van Dijk, J. A. 2006. Digital divide research, achievements and shortcomings. *Poetics*, 34(4-5): 221-235.

Van Dijk, J. A., and K. Hacker. 2003. The digital divide as a complex and dynamic phenomenon. *The Information Society*, 19(4): 315-326.

Appendix A: Variable Definitions

MP	=	Our main proxy for the minority proportion, which is the county-year-level percentage of African American and/or Latinx population. (Note: all demographic data are obtained from the U.S. Census Bureau)
AA		The county-year-level percentage of African American population. We use this proxy in robustness tests.
DECTY	=	The county-year-level percentage of Prosper loans that are "Expired" or "Cancelled" by Prosper, as indicated by the Prosper loan listing dataset. (Note: all loan level data are obtained from Prosper's developer platform)
RATECTY	=	The average county-year-level quoted Prosper loan borrower rate (i.e., a value of 15 suggests 15%, we multiply the rate by 100 for display purposes).
DE	=	An indicator variable on the individual loan level, which equals 1 if the loan is denied (i.e., "Expired" or "Cancelled" by the platform, not by the applicants), and 0 otherwise.
RATE	=	The individual loan level quoted borrower interest rate.
PRATE	=	Rating of loan listings developed internally by Prosper. The rating result serves as a reference for the determination of loan interest rate for both borrowers and investors.
INFO	=	A state-year-level indicator variable for the availability of alternative information that can be potentially used by Prosper or other FinTech lenders. This proxy is constructed by taking the sum of three percentages provided by National Telecommunications and Information Administration (NTIA), including the percentage of 15+ persons using online social networks, the percentage of 15+ persons using online financial services, and the percentage of 15+ persons using online shopping. The variable *INFO* is then constructed as an indicator variable that equals 1 if the sum of percentages is above the median value of all states across all available years, and 0 otherwise. The data is available only at the state-year level for years 2013, 2015, 2017, and 2019 regarding Ecommerce and online financial services users, and for years 2015, 2017, and 2019 regarding social network users. *Source: NTIA*, United States Department of Commerce. https://www.ntia.doc.gov/data/digital-nation-data-explorer#sel=socialNetworkUser&disp=map
INFO_A	=	We construct *INFO_A* in a way similar to *INFO*, except that we convert the percentages into ranks across the states by year. Hence, the alternative proxy intends to capture within-year cross-state variations. The states that have a rank over 25 (i.e., there are 51 states including DC in our sample) will be assigned a value of 1 to indicate having a higher level of relative availability of alternative information compared to other states for a given year.
ECOM	=	A state-year-level indicator variable for the percentage of 15+ persons using online shopping. It equals 1 if the value (*ECOM %*) is above the median of all states across all years, and 0 otherwise. Alternatively, we construct a variable *ECOM_A* in a way similar to *INFO_A* to measure the relative availability of online shopping information in a state compared to other states for a given year.
OFS	=	A state-year-level indicator variable for the percentage of 15+ persons using online financial services. It equals 1 if the value (*OFS %*) is above the median of all states across all years, and 0 otherwise. Alternatively, we construct a variable *OFS_A* in a way similar to *INFO_A* to measure the relative availability of online financial services information in a state compared to other states for a given year.
SNW	=	A state-year-level indicator variable for the percentage of 15+ persons using online social networks. It equals 1 if the value (*SNW %*) is above the median of all states

		across all years, and 0 otherwise. Alternatively, we construct a variable *SNW_A* in a way similar to *INFO_A* to measure the relative availability of online social networks information in a state compared to other states for a given year.
AMTCTY	=	The average county-year-level loan amount requested.
TERMCTY	=	The average county-year-level loan term.
SCORECTY	=	The average county-year-level borrower's credit bureau score.
INCSELFCTY	=	The average county-year-level self-disclosed income range.
EMPLENCTY	=	The average county-year-level self-disclosed employment length.
DPRIORCTY	=	The county-year-level proportion of loan borrowers that have prior Prosper loans.
PRIORCTY	=	The average county-year-level number of prior loans across all borrowers.
PRIORACTY	=	The average county-year-level number of prior loans active across all borrowers.
PRIORPCTY	=	The average county-year-level prior loan principal outstanding in thousands.
PRIOROCTY	=	The average county-year-level number of prior on-time payments.
PRIORLCTY	=	The average county-year-level number of prior late payments over 1 month.
AMT	=	The individual loan level loan amount requested.
TERM	=	The individual loan level loan term.
SCORE	=	The individual loan level borrower's credit bureau (FICO) score.
INCSELF	=	The individual loan level self-disclosed income range.
EMPLEN	=	The individual loan level self-disclosed employment length.
PRIOR	=	The individual loan level number of prior loans with Prosper.
PRIORA	=	The individual loan level number of prior loans that are currently active.
PRIORP	=	The individual loan level prior loan principal outstanding in thousands.
PRIORO	=	The individual loan level number of prior on-time payments.
PRIORL	=	The individual loan level number of prior late payments over 1 month.
PEN	=	The county-year-level total completed Prosper loans per capita (penetration).
INCOME	=	County-year-level per capita personal income in thousands *Source:* U.S. Bureau of Economic Analysis (BEA).
EDU	=	County-year-level proportion of adult population receiving post-secondary education.
MALE	=	County-year-level proportion of population that is male.
MARRIED	=	County-year-level proportion of adult population that is married.
UNEMP	=	County-year-level unemployment rate. (*Source:* U.S. Bureau of Labor Statistics)
HPI	=	County-year-level housing price index. (*Source:* Federal Housing Finance Agency)
GDP	=	The natural logarithm of county-year-level GDP per capita. (*Source:* U.S. Bureau of Economic Analysis)

Appendix B: Derivations of the Framework

We define q as the credit quality of a borrower, where $q \sim N(\bar{q}, var(q))$. $y_1 = q + \varepsilon_1$, where y_1 is a traditional credit signal for q with $\varepsilon_1 \sim (0, var(\varepsilon_1))$. y_1 can be more than one measures, for simplicity, we assume y_1 is a single credit score. $y_2 = q + \varepsilon_2$, where y_2 is a non-traditional credit signal for q with $\varepsilon_2 \sim (0, var(\varepsilon_2))$.

$z = A$ if the borrower belongs to the non-minority group, which is observable to lenders. $z = B$ if the borrower belongs to the minority group, which is observable to lenders.

We assume that to the extent that P2P platforms and/or investors show prior prejudice against minorities when making lending decisions, they will perceive the average credit quality of the borrowers with a discount of δ: i.e., for biased lenders, the perceived average loan quality for minority borrowers would be $(\bar{q} - \delta)$ instead of $\bar{q}$.

Following Philippon (2019), we denote the precisions $p \equiv 1/var(q)$, $p_1 \equiv 1/var(\varepsilon_1)$, and $p_2 \equiv 1/var(\varepsilon_2)$, for convenience purposes.

Lenders can observe y_1 and y_2, where $z = A \text{ or } B$, and they expect these two signals can provide information about the unobservable variable q. Hence, we also assume that $\bar{y}_1 = \bar{y}_2 = \bar{q}$ and that $var(q) = cov(q, y_1) = cov(q, y_2)$.[51] For simplicity purposes, we also assume that $cov(y_1, y_2) = 0$, which indicates that the two signals contain two unrelated sets of credit information about individuals.

Therefore, our goal is to derive $E[q|y_1, y_2, z]$. Following Phelps (1972) and Aigner and Cain (1977), we model $E[q|y_1, y_2]$ with an operational regression equation (A1), because lenders could measure the actual q of a potential borrower on the basis of a *post hoc* assessment of the borrowers' loan payment performance.

$$q = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + u \tag{A1}$$

Suppose that $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ are OLS estimators of β_0, β_1, β_2:

$$E[q|y_1, y_2] = \hat{\beta}_0 + \hat{\beta}_1 y_1 + \hat{\beta}_2 y_2 \tag{A2}$$

Since $\hat{\beta}_0 = \bar{q} - \hat{\beta}_1 \bar{y}_1 - \hat{\beta}_2 \bar{y}_2$, and we assumed that $\bar{y}_1 = \bar{y}_2 = \bar{q}$, we have:

$$E[q|y_1, y_2] = (1 - \hat{\beta}_1 - \hat{\beta}_2)\bar{q} + \hat{\beta}_1 y_1 + \hat{\beta}_2 y_2 \tag{A3}$$

Because we assumed that lenders may prejudice against minorities by the amount of δ, we have:

$$E[q|y_1, y_2, z = B] = (1 - \hat{\beta}_1 - \hat{\beta}_2)(\bar{q} - \delta) + \hat{\beta}_1 y_1 + \hat{\beta}_2 y_2 \tag{A4}$$

By the law of iterated expectations, we have the unconditional expectation of q by taking the averages along the dimensions of y_1 and y_2 in equation (A4):

$$E[q|z = B] = \bar{q} - \delta(1 - \hat{\beta}_1 - \hat{\beta}_2) \tag{A5}$$

Now, we apply the OLS solutions formula to obtain $\hat{\beta}_1$ and $\hat{\beta}_2$, with all our assumptions described above:

$$\hat{\beta}_1 = \frac{cov(y_1,q)var(y_2) - cov(y_1,y_2)cov(y_2,q)}{var(y_1)var(y_2) - cov(y_1,y_2)^2} = \frac{cov(y_1,q)var(y_2)}{var(y_1)var(y_2)} = \frac{(\frac{1}{p}+\frac{1}{p_2})\frac{1}{p}}{(\frac{1}{p}+\frac{1}{p_2})(\frac{1}{p}+\frac{1}{p_1})} = \frac{p_1}{p+p_1} \tag{A6}$$

Similarly, $\hat{\beta}_2 = \frac{p_2}{p+p_2}$. So, we have:

$$E[q|z = B] = \bar{q} - \delta(1 - \frac{p_1}{p+p_1} - \frac{p_2}{p+p_2}) \tag{A7}$$

The downward Bias:

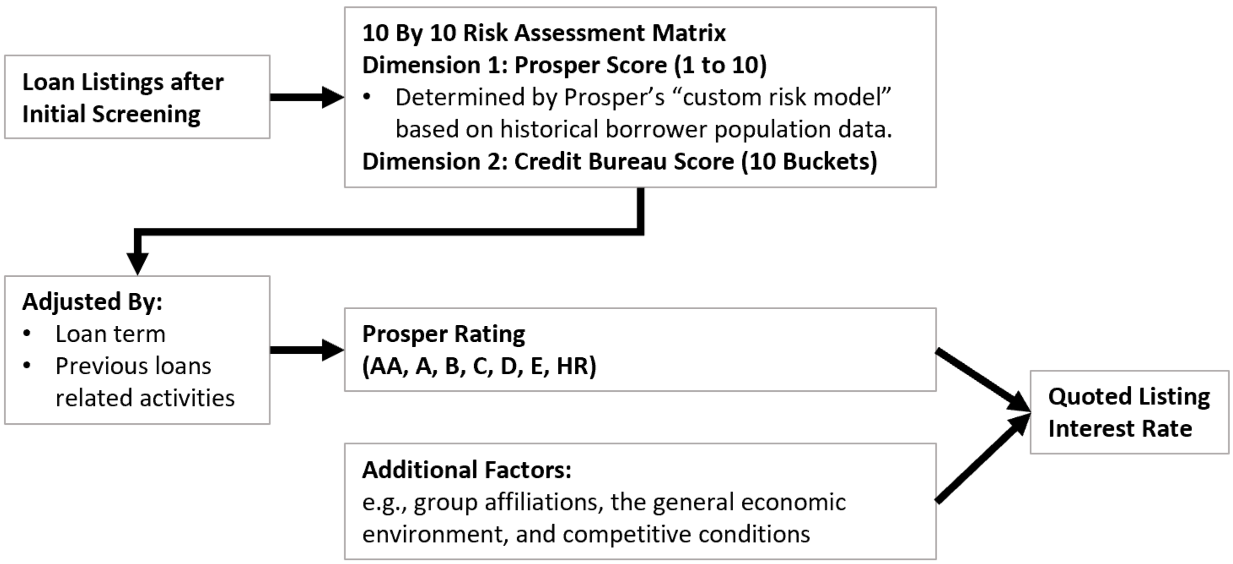
$$b = \bar{q} - E[q|z = B] = \delta(1 - \frac{p_1}{p+p_1} - \frac{p_2}{p+p_2}) \tag{A8}$$

Therefore, as long as $\delta > 0$, we have $\partial b / \partial p_1 < 0$ and $\partial b / \partial p_2 < 0$. That is, when δ>0, the discriminative bias will be lower when the precisions level of credit quality signals are higher.

[51] These assumptions are consistent with Aigner and Cain (1977)

Table 1: Prosper's Credit Pricing System and Sample Construction

Panel A: Prosper's credit pricing system (summarized from Prosper's prospectus report)



Prosper Rating	Corresponding estimated mean annual loss rate (%)
AA	0 – 1.99
A	2 – 3.99
B	4 – 5.99
C	6 – 8.99
D	9 – 11.99
E	12 – 14.99
HR	>= 15

This information is summarized from Prosper's 2010 prospectus report, post-effective amendment No. 3. Prosper uses a 10 by 10 matrix to determine the base probability of default rate using Prosper Score (1 to 10) and Credit Bureau Score (10 buckets). Then, borrower interest rate is adjusted by the loan term and whether a borrower has previous Prosper loan(s), as well as additional considerations. The quoted listing rate is NOT open for bidding. Investors only bid for whether to invest in the loan listed and the loan amount.

Panel B: Sample construction

Original loan listings data from 2013 to 2019	1,538,451
After matching the original data to U.S. counties and deleting unmatched observations	1,317,474
After matching with county level race, personal income, GDP, HPI, unemployment rate, marital status, and education variables, and deleting unmatched observations	1,300,886
After deleting missing values of loan-level characteristic variables	**1,093,797**
• Final sample with successfully funded loans	825,975
• Final sample with loan listings that are cancelled or expired	259,731
• Final sample with loan listings that are under other status	8,091
After collapsing into county-level data	13,867
After deleting county-level missing values of loan amount, loan term, loan status, loan rate, FICO scores, self-reported income range, and self-reported employment length	**12,600**

Table 2: Proxies for P2P Loan Applicants' Information Environment at State Level

	E-Commerce User %				E-Financial Services User %				Social Network User %		
State\Year	2013	2015	2017	2019	2013	2015	2017	2019	2015	2017	2019
AK	58%	75%	71%	73%	53%	68%	65%	69%	75%	74%	71%
AL	39%	62%	60%	67%	39%	54%	60%	61%	74%	77%	76%
AR	49%	62%	66%	68%	43%	57%	64%	66%	70%	80%	79%
AZ	50%	69%	66%	77%	49%	65%	66%	78%	66%	75%	72%
CA	55%	70%	69%	74%	51%	67%	68%	70%	69%	74%	73%
CO	59%	77%	76%	79%	60%	73%	72%	73%	70%	77%	75%
CT	56%	74%	74%	81%	46%	64%	67%	73%	69%	73%	74%
DC	72%	80%	81%	85%	67%	75%	76%	87%	79%	78%	78%
DE	50%	71%	70%	76%	45%	63%	65%	68%	70%	71%	68%
FL	47%	65%	66%	70%	47%	66%	66%	73%	71%	75%	73%
GA	48%	66%	64%	68%	48%	64%	65%	72%	73%	77%	70%
HI	57%	71%	67%	75%	48%	58%	59%	63%	70%	69%	70%
IA	55%	65%	72%	76%	52%	57%	65%	69%	71%	80%	82%
ID	53%	75%	73%	69%	53%	67%	68%	70%	74%	78%	75%
IL	53%	73%	69%	75%	47%	66%	64%	71%	72%	73%	73%
IN	54%	64%	72%	73%	51%	60%	67%	70%	71%	78%	78%
KS	53%	63%	69%	72%	52%	62%	67%	69%	73%	81%	77%
KY	49%	69%	63%	65%	46%	61%	62%	66%	75%	79%	77%
LA	48%	63%	60%	64%	46%	61%	60%	64%	70%	72%	75%
MA	55%	79%	76%	80%	49%	67%	72%	72%	70%	70%	73%
MD	67%	74%	77%	77%	58%	65%	70%	79%	69%	77%	73%
ME	54%	67%	72%	76%	50%	61%	70%	70%	74%	79%	78%
MI	54%	68%	69%	74%	51%	64%	66%	70%	70%	71%	72%
MN	64%	72%	77%	78%	61%	67%	67%	73%	71%	73%	75%
MO	53%	69%	69%	72%	50%	61%	63%	66%	70%	77%	80%
MS	43%	59%	57%	61%	37%	53%	50%	56%	74%	78%	77%
MT	54%	72%	76%	74%	44%	63%	67%	67%	70%	75%	74%
NC	51%	67%	68%	61%	45%	60%	67%	59%	70%	77%	72%
ND	55%	74%	76%	75%	51%	63%	69%	69%	69%	78%	79%
NE	56%	66%	71%	77%	56%	59%	71%	75%	73%	77%	79%
NH	63%	74%	74%	80%	54%	65%	73%	73%	68%	72%	74%
NJ	52%	71%	65%	73%	47%	68%	61%	67%	67%	71%	68%
NM	48%	62%	60%	72%	43%	61%	60%	70%	66%	69%	69%
NV	52%	72%	65%	71%	48%	73%	66%	70%	71%	70%	71%
NY	50%	65%	65%	69%	41%	62%	61%	64%	68%	70%	72%
OH	54%	70%	68%	73%	51%	64%	65%	69%	69%	76%	79%
OK	43%	56%	63%	65%	40%	54%	62%	66%	67%	74%	78%
OR	60%	73%	76%	76%	56%	67%	73%	72%	71%	73%	74%
PA	56%	72%	73%	76%	47%	64%	65%	71%	68%	72%	73%
RI	56%	70%	74%	75%	49%	63%	64%	69%	68%	71%	72%
SC	49%	66%	69%	72%	47%	60%	65%	65%	66%	74%	72%
SD	51%	64%	66%	78%	54%	60%	63%	72%	71%	77%	81%
TN	49%	68%	66%	67%	43%	60%	65%	63%	73%	76%	74%
TX	45%	64%	64%	66%	48%	68%	66%	72%	75%	76%	75%
UT	61%	75%	78%	81%	61%	77%	76%	79%	76%	82%	76%
VA	61%	75%	74%	78%	54%	70%	68%	74%	68%	74%	72%
VT	58%	75%	71%	78%	45%	64%	64%	69%	68%	71%	69%
WA	63%	78%	74%	79%	64%	71%	71%	79%	75%	72%	70%
WI	59%	65%	73%	74%	54%	62%	68%	71%	70%	72%	76%
WV	44%	60%	56%	67%	36%	54%	56%	61%	72%	69%	74%
WY	51%	74%	73%	76%	48%	64%	66%	67%	68%	78%	73%

This table provides the sample composition of the state-year level percentages of age 15+ persons using E-commerce, online financial services, and social network. Data is obtained from NTIA, United States Department of Commerce. https://www.ntia.doc.gov/data/digital-nation-data-explorer#sel=socialNetworkUser&disp=map. Also, note that our sample of Prosper loan listings does not contain loan listings in AK and IA.

Table 3: Descriptive Statistics and Correlations

Panel A: Descriptive statistics for variables at county level

Variable	N	Mean	Q1	p50	Q3	Std dev
MP	12,600	0.191	0.050	0.123	0.294	0.179
AA	12,600	0.101	0.010	0.035	0.131	0.143
DECTY	12,600	0.283	0.167	0.250	0.346	0.165
RATECTY	12,600	15.680	14.081	15.327	17.222	2.561
AMTCTY	12,600	12.664	11.157	12.746	14.174	2.693
TERMCTY	12,600	43.564	41.684	43.368	45.231	3.851
SCORECTY	12,600	706.128	697.756	704.955	713.187	15.191
INCSELFCTY	12,600	4.081	3.810	4.067	4.347	0.452
EMPLENCTY	12,600	121.046	98.129	115.903	137.323	45.546
DPRIORCTY	12,600	0.226	0.111	0.204	0.313	0.141
PRIORCTY	12,600	0.272	0.065	0.200	0.393	0.329
PRIORACTY	12,600	0.137	0.000	0.100	0.190	0.210
PRIORPCTY	12,600	1.130	0.000	0.653	1.410	2.828
PRIOROCTY	12,600	1.635	0.063	0.815	2.000	2.994
PRIORLCTY	12,600	0.009	0.000	0.000	0.000	0.132
PEN	12,600	5.125	2.520	4.600	7.032	3.428
INCOME	12,600	41.804	34.720	39.400	45.539	12.132
EDU	12,600	0.131	0.092	0.120	0.161	0.054
MALE	12,600	0.498	0.488	0.494	0.502	0.019
MARRIED	12,600	0.430	0.402	0.435	0.463	0.047
ECOM%[52]	7,670	0.652	0.614	0.662	0.716	0.086
OFS%	7,670	0.620	0.592	0.642	0.674	0.085
SNW%	6,091	0.736	0.712	0.732	0.762	0.033
UNEMP	12,600	5.392	3.900	5.000	6.500	2.057
HPI	12,600	143.643	125.080	138.670	156.930	28.251
GDP	12,600	10.582	10.289	10.571	10.836	0.444

Panel B: Descriptive statistics for variables at loan level

Variable	N	Mean	Q1	p50	Q3	Std dev
DE	1,093,797	0.237	0.000	0.000	0.000	0.426
RATE	1,093,797	14.935	10.280	13.540	18.550	6.237
AMT	1,093,797	13.630	7.500	12.000	19.000	8.132
TERM	1,093,797	43.206	36.000	36.000	60.000	11.001
SCORE	1,093,797	704.42	669.50	689.50	729.50	39.65
INCSELF	1,093,797	4.315	3.000	4.000	5.000	1.161
EMPLEN	1,093,797	109.55	28.00	76.00	161.00	106.36
PRIOR	1,093,797	0.290	0.000	0.000	0.000	1.077
PRIORA	1,093,797	0.143	0.000	0.000	0.000	0.732
PRIORP	1,093,797	0.939	0.000	0.000	0.000	3.104
PRIORO	1,093,797	1.684	0.000	0.000	0.000	10.654
PRIORL	1,093,797	0.011	0.000	0.000	0.000	0.601

[52] Data of *ECOM %* and *OFS %* is available for years 2013, 2015, 2017 and 2019; data of *SNW %* is only available for years 2015, 2017 and 2019.

Panel C: Correlations for main variables on the county level

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1)	*DECTY*	1.00																	
(2)	*RATECTY*	0.05	1.00																
(3)	*MP*	-0.01	**0.05**	1.00															
(4)	*AA*	0.01	**0.08**	0.72	1.00														
(5)	*AMTCTY*	-0.05	-0.33	0.06	0.00	1.00													
(6)	*TERMCTY*	0.08	0.13	-0.08	-0.02	0.18	1.00												
(7)	*SCORECTY*	-0.11	-0.46	-0.02	-0.02	0.17	0.00	1.00											
(8)	*INCSELFCTY*	-0.14	-0.14	0.11	0.01	0.49	0.07	0.13	1.00										
(9)	*EMPLENCTY*	0.02	-0.05	0.00	0.07	0.00	0.02	0.06	-0.02	1.00									
(10)	*DPRIORCTY*	-0.29	0.04	-0.05	-0.01	-0.06	-0.10	0.30	0.16	0.04	1.00								
(11)	*PEN*	-0.38	-0.18	0.08	-0.00	0.39	-0.04	0.00	0.21	-0.06	-0.07	1.00							
(12)	*INCOME*	-0.21	-0.10	-0.06	-0.15	0.19	-0.04	0.08	0.35	-0.12	0.09	0.28	1.00						
(13)	*EDU*	-0.21	-0.08	-0.07	-0.11	0.16	-0.01	0.04	0.33	-0.15	0.01	0.27	0.74	1.00					
(14)	*MALE*	0.06	-0.02	-0.05	-0.19	0.00	-0.03	0.02	-0.02	-0.02	0.04	-0.02	-0.11	-0.16	1.00				
(15)	*MARRIED*	0.07	-0.02	-0.61	-0.59	0.01	0.05	0.00	-0.03	0.01	0.05	0.02	0.13	0.05	0.03	1.00			
(16)	*UNEMP*	0.36	0.10	0.22	0.24	-0.14	0.11	-0.14	-0.16	0.08	-0.27	-0.32	-0.40	-0.40	-0.04	-0.19	1.00		
(17)	*HPI*	-0.21	-0.10	0.14	-0.11	0.16	-0.13	0.15	0.28	-0.14	0.24	0.22	0.43	0.31	0.12	-0.04	-0.35	1.00	
(18)	*GDP*	-0.15	-0.04	0.13	-0.06	0.11	-0.03	0.04	0.25	-0.12	0.02	0.20	0.60	0.50	-0.05	-0.13	-0.32	0.34	1.00

This table provides the descriptive statistics (panels A and B) and Spearman's correlations (Panel C) of the main variables used in this study. The detailed definitions of the variables are provided in Appendix A. All correlations with absolute values greater than 0.02 are statistically significant at the 0.01 level or better (two-tailed)

Table 4: Univariate Analysis
Loan-level Characteristics Mean Differences in Counties with Highest/lowest Minority Proportion

		(1)	(2)	(2) – (1)		
	N in each subsample	Counties with the lowest quartile *MP*	Counties with the highest quartile *MP*	Differences	Test of Difference T-stat H0: diff ≠ 0	P-Value
DE	545,761	0.23	0.25	0.02***	19.50	< 0.001
RATE	545,761	14.92	15.09	0.17***	9.98	< 0.001
AMT	545,761	13.25	13.67	0.42***	19.25	< 0.001
TERM	545,761	43.64	42.77	-0.88***	29.47	< 0.001
SCORE	545,761	705.21	703.52	-1.69***	-15.78	< 0.001
INCSELF	545,761	4.19	4.33	0.14***	44.59	< 0.001
EMPLEN	545,761	114.30	107.83	-6.47***	-22.35	< 0.001
PRIOR	545,761	0.29	0.27	-0.02***	-6.11	< 0.001
PRIORP	545,761	0.94	0.90	-0.05***	-5.87	< 0.001
PRIORO	545,761	1.76	1.55	-0.21***	-7.58	< 0.001
PRIORL	545,761	0.009	0.011	0.001*	-0.86	0.392

Table 3 compares the differences in the mean values of some key loan characteristics variables of interest between counties with the lowest quartile of Black/Latinx population percentage and counties with the highest quartile of Black/Latinx population percentage. Continuous variables are winsorized at top and bottom 1%. All variables are defined in the Appendix A . *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively, based on a two-tailed test.

Table 5: Tests of Existence of Racial Discrimination

Panel A: County level P2P loan failure rate/loan interest rate and race

	(1) *DECTY*		(2) *DECTY*		(3) *RATECTY*		(4) *RATECTY*	
MP	**0.052*****	**(7.44)**	**0.041*****	**(4.62)**	**0.954*****	**(8.72)**	**0.437*****	**(3.17)**
AMTCTY	0.006***	(8.25)	0.006***	(8.69)	-0.127***	(-4.41)	-0.128***	(-9.88)
TERMCTY	-0.001*	(-2.06)	-0.001	(-1.41)	0.101***	(5.22)	0.099***	(14.81)
SCORECTY	-0.000	(-0.06)	0.000	(0.03)	-0.078***	(-8.93)	-0.076***	(-30.34)
INCSELFCTY	-0.016***	(-4.67)	-0.014***	(-2.86)	-0.140	(-1.34)	-0.149**	(-2.10)
EMPLENCTY	-0.000	(-0.77)	-0.000	(-0.55)	-0.000	(-0.60)	-0.000	(-0.02)
PRIORCTY	-0.014	(-1.42)	-0.017*	(-1.73)	-0.280	(-1.85)	-0.165	(-0.99)
PRIORACTY	-0.029*	(-2.19)	-0.028	(-1.49)	0.717*	(2.29)	0.684**	(2.12)
PRIORPCTY	0.002**	(3.09)	0.002***	(2.72)	-0.026	(-1.35)	-0.028	(-1.57)
PRIOROCTY	-0.002*	(-2.04)	-0.001**	(-2.57)	-0.000	(-0.04)	-0.002	(-0.22)
PRIORLCTY	0.016	(1.69)	0.016**	(2.05)	-0.061	(-0.48)	-0.108	(-1.02)
INCOME	0.000	(0.93)	0.000***	(3.23)	-0.005	(-1.86)	-0.001	(-0.65)
EDU	-0.165**	(-3.22)	-0.169***	(-5.35)	-2.862***	(-5.76)	-1.873***	(-4.36)
MALE	0.157**	(2.71)	0.068	(1.03)	-0.586	(-0.40)	2.289**	(2.24)
MARRIED	0.262***	(15.78)	0.232***	(8.25)	0.462	(1.09)	-0.185	(-0.45)
PEN	-0.012***	(-6.07)	-0.014***	(-21.58)	0.021	(1.41)	0.014	(1.58)
UNEMP	0.001	(1.44)	0.003***	(3.56)	-0.058***	(-3.91)	-0.014	(-1.10)
HPI	0.000***	(4.92)	0.000***	(5.39)	-0.001	(-1.25)	-0.001	(-0.97)
GDP	-0.001	(-0.20)	-0.000	(-0.13)	0.018	(0.25)	0.104**	(2.06)
Constant	0.179	(1.30)	0.187**	(2.04)	76.356***	(12.08)	72.742***	(34.08)
Year Fixed Effects	Yes		Yes		Yes		Yes	
State Fixed Effects	No		Yes		No		Yes	
SE Cluster	Year, County		Year, County		Year, County		Year, County	
R-squared	0.46		0.59		0.60		0.62	
N	12,600		12,600		12,600		12,600	

This table reports the regression of county-level Prosper loan failure rate and loan interest rate on race and an array of county-level loan-related, macroeconomic, and demographic control variables. The main independent variable is *Black/Latinx %_CTY*. Columns 1 and 2 show the results containing county-level macroeconomic and demographic controls only: Column 1 uses the county-level percentage of failed Prosper loans as the dependent variable and Column 2 uses the county-level average Prosper borrowers loan interest rate as the dependent variable. Columns 3 and 4 show the results containing all control variables: Column 3 uses the county-level percentage of failed Prosper loans as the dependent variable and Column 4 uses the county-level average Prosper borrowers loan interest rate as the dependent variable. The detailed definitions of all variables are provided in Appendix A. All columns include state and year fixed effects. Coefficients on the year and state indicator variables are not tabulated for brevity. The continuous variables are winsorized at top and bottom 1% to eliminate the confounding effects of outliers. The t-statistics reported in parentheses are based on robust standard errors clustered by year and state. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

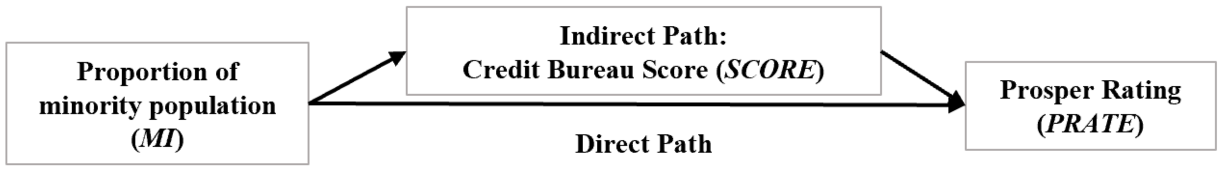
Panel B: Loan level FinTech loan failure event/loan interest rate and race

	(1)		(2)		(3)		(4)		(5)		(6)	
	RATE		*RATE*		*DE = 1*		*DE = 1*		*RATE*		*RATE*	
MP	**0.550*****	**(6.74)**	**0.486*****	**(4.72)**	**0.210*****	**(6.00)**	**0.231*****	**(4.06)**	**6.72****	**(4.87)**	**6.045***	**(2.28)**
AMT	-0.081***	(-25.04)	-0.067***	(-67.06)	0.017***	(22.93)	0.017***	(22.96)	-0.081***	(-63.19)	-0.081***	(-63.19)
TERM	0.130***	(61.64)	0.129***	(61.64)	-0.001***	(-4.17)	-0.001***	(-4.35)	0.129***	(159.17)	0.129***	(159.14)
SCORE	-0.076***	(-93.58)	-0.082***	(-283.6)	0.002***	(7.63)	0.002***	(7.64)	-0.076***	(-266.4)	-0.082***	(-282.4)
INCSELF	-0.331***	(-25.60)	-0.380***	(-25.75)	-0.043***	(-7.02)	-0.039***	(-6.10)	-0.326***	(-41.75)	-0.326***	(-41.75)
EMPLEN	-0.000	(-0.21)	-0.000	(-0.65)	-0.000***	(-7.72)	-0.000***	(-8.29)	-0.000	(-1.05)	-0.000	(-1.06)
PRIOR	-0.493***	(-9.61)	-0.491***	(-9.60)	-0.633***	(-23.93)	-0.631***	(-23.87)	-0.493***	(-13.52)	-0.492***	(-13.51)
PRIORA	0.044	(0.56)	0.043	(0.55)	0.073**	(2.29)	0.073**	(2.27)	0.050	(0.90)	0.049	(0.88)
PRIORP	0.068***	(12.55)	0.068***	(12.54)	-0.039***	(-13.20)	-0.039***	(-13.21)	0.068***	(14.74)	0.068***	(14.81)
PRIORO	-0.018***	(-7.51)	-0.018***	(-7.51)	-0.010***	(-5.47)	-0.010***	(-5.45)	-0.018***	(-8.23)	-0.018***	(-8.26)
PRIORL	0.030*	(1.85)	0.030*	(1.84)	0.030***	(5.43)	0.030***	(5.33)	0.031*	(1.67)	0.031*	(1.68)
INCOME			-0.0007	(-0.57)			0.002***	(3.20)			-0.034***	(-3.57)
EDU			-1.3178***	(-5.39)			-1.307***	(-12.75)			0.9009	(0.29)
MALE			-0.2310	(-0.19)			0.310	(0.61)			-13.745*	(-1.73)
MARRIED			0.1454	(0.45)			0.736***	(4.17)			0.980	(0.42)
UNEMP			-0.0132	(-1.33)			0.008*	(1.78)			-0.099***	(-4.40)
HPI			0.0014	(1.64)			0.002***	(6.86)			-0.0004	(-0.21)
GDP			0.0000	(0.36)			0.000	(0.16)			0.0000	(1.11)
Constant	69.416***	(60.37)	69.779***	(53.19)	-0.555	(-1.38)	-1.209**	(-2.38)	66.961***	(43.79)	75.66***	(17.76)
Year FE	Yes		Yes		Yes		Yes		Yes		Yes	
State FE	Yes		Yes		Yes		Yes		No		No	
County FE	No		No		No		No		Yes		Yes	
SE Cluster	Year, County		Year, County		Year, County		Year, County		Year, County		Year, County	
Observations	1,093,797		1,093,797		1,093,797		1,093,797		1,093,797		1,093,797	
R^2	0.36		0.36		n/a		n/a		0.36		0.36	
ROC	n/a		n/a		0.66		0.67		n/a		n/a	

Columns 1, 2, 5, and 6 of this table report results of OLS regressions of individual Prosper loan level borrower interest rate on race and an array of loan-level and county-level control variables. Columns 3 and 4 report results of logistic regressions of individual Prosper loan denial event on race and an array of loan-level and county-level control variables. Columns 1-4 include year and state fixed effects, and columns 5 and 6 include year and county fixed effects. Columns 1, 3, and 5 include only county-level loan-related control variables while columns 2, 4, and 6 include all control variables. The detailed definitions of all variables are provided in Appendix A. Continuous variables are winsorized at top and bottom 1%. Coefficients on the year, state and county indicator variables are not tabulated for brevity. The t-statistics reported in parentheses are based on robust standard errors clustered by year and state for columns 1-4, and by year and county for columns 3 and 4. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 6: Path Analyses—Where Does the Discrimination on P2P Platforms Come From?

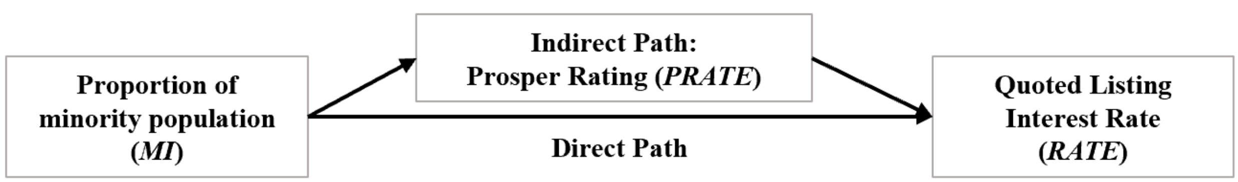
Panel A: Path analysis of platform ratings on race through credit bureau scores



		Coefficients	T-stat
Direct Path	p (*MP, PRATE*)	-0.171***	(-20.49)
Indirect Path	p (*MP, SCORE*)	-3.584***	(-12.60)
	p (*SCORE, RATE*)	0.020***	(780.93)
Total magnitude of indirect effect		-0.072***	(-12.60)
Percentage of direct effect to the total effect		70%	
Percentage of indirect effect to the total effect		30%	
Control variables		Yes	
N		1,093,797	

This table reports result from a path analysis that examines the direct effect and the indirect effect through the credit bureau ratings (*SCORE*). The second state dependent variable is *PRATE*. A recursive path model with observable variables is used. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The standard error is heteroskedasticity-robust.

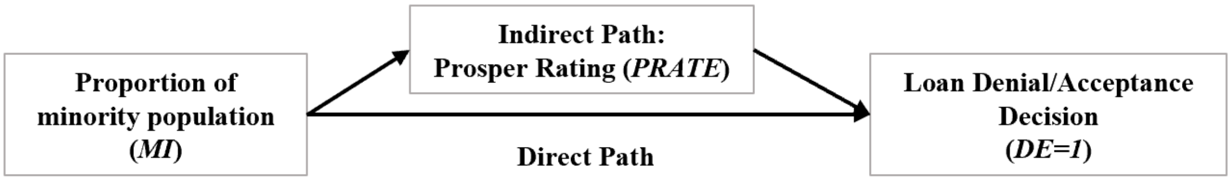
Panel B: Path analysis of loan interest rate on race through platform ratings



		Coefficients	T-stat
Direct Path	p (*MP, RATE*)	0.666***	(44.56)
Indirect Path	p (*MP, PRATE*)	-0.468***	(-46.34)
	p (*PRATE, RATE*)	-3.996***	(-2131)
Total magnitude of indirect effect		1.870***	(32.82)
Percentage of direct effect to the total effect		26%	
Percentage of indirect effect to the total effect		74%	
Control variables		Yes	
N		1,093,797	

This table reports result from a path analysis that examines the direct effect and the indirect effect through the Prosper ratings (*PRATE*) The second state dependent variable is the quoted interest rate (*RATE*). A recursive path model with observable variables is used. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The standard error is heteroskedasticity-robust.

Panel C: Path analysis of loan denial decisions on race through platform rating
(Note: 2nd stage regression uses a linear probability model)



		Coefficients	T-stat
Direct Path	p (*MP*, *DE*)	0.031***	(7.83)
Indirect Path	p (*MP*, *PRATE*)	-0.468***	(-46.34)
	p (*PRATE*, *DE*)	-0.011***	(-24.79)
Total magnitude of indirect effect		0.005***	(21.87)
Percentage of direct effect to the total effect		86%	
Percentage of indirect effect to the total effect		14%	
Control variables		Yes	
N		1,093,797	

This table reports result from a path analysis that examines the direct effect and the indirect effect through the internal developed credit rating by Prosper. A recursive path model with observable variables is used. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The standard error is heteroskedasticity-robust.

Panel D: Alternative specification for the test in Panel C
(Note: 2nd stage regression uses a linear probability model)

		Coefficients	T-stat
Direct Path	p (*MP*, *DE*)	0.024***	(7.81)
Indirect Path	p (*MP*, *RATE*)	2.985***	(80.23)
	p (*RATE*, *DE*)	0.002***	(27.93)
Total magnitude of indirect effect		0.006***	(26.40)
Percentage of direct effect to the total effect		80%	
Percentage of indirect effect to the total effect		20%	
Control variables		Yes	
N		1,093,797	

This table reports result from a path analysis that examines the direct effect and the indirect effect through the quoted interest rates by Prosper. A recursive path model with observable variables is used. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The standard error is heteroskedasticity-robust.

Table 7: The Effects of Alternative Information—County-level Regressions

	(1) *DECTY*	(2) *RATECTY*	(3) *DECTY*	(4) *RATECTY*	(5) *DECTY*	(6) *RATECTY*	(7) *DECTY*	(8) *RATECTY*
MP	0.087***	1.020***	0.060***	0.916***	0.042	2.389	-0.014	1.846
INFO × MP	**-0.055*****	**-0.703*****			**-0.053*****	**-0.585*****		
INFO_A × MP			**-0.024***	**-0.612****			**-0.015**	**-0.458****
INFO	0.017***	0.188			0.018***	0.160**		
INFO_A			0.008*	0.157			0.007*	0.122*
AMTCTY	0.006***	-0.128***	0.006***	-0.128***	0.007***	-0.140***	0.007***	-0.140***
TERMCTY	-0.001	0.099***	-0.001	0.099***	-0.000	0.102***	-0.000	0.102***
SCORECTY	0.000	-0.076***	0.000	-0.076***	-0.000	-0.077***	-0.000	-0.077***
INCSELFCTY	-0.013***	-0.147**	-0.013***	-0.149**	-0.009**	-0.109	-0.009**	-0.110
EMPLENCTY	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
PRIORCTY	-0.017*	-0.169	-0.017*	-0.165	-0.009	-0.038	-0.008	-0.033
PRIORACTY	-0.027	0.692**	-0.028	0.684**	-0.030	0.718**	-0.031*	0.708**
PRIORPCTY	0.002***	-0.028	0.002***	-0.028	0.002**	-0.032*	0.002**	-0.032*
PRIOROCTY	-0.001**	-0.002	-0.001**	-0.002	-0.001**	-0.012	-0.001**	-0.012
PRIORLCTY	0.016**	-0.106	0.016**	-0.106	-0.001	0.042	-0.001	0.040
INCOME	0.000***	-0.001	0.000***	-0.001	0.001	-0.015	0.001	-0.015
EDU	-0.171***	-1.900***	-0.169***	-1.880***	-0.231	-3.075	-0.230	-3.013
MALE	0.068	2.294**	0.068	2.320**	-0.462	5.781	-0.468	5.861
MARRIED	0.235***	-0.146	0.233***	-0.166	0.205	3.682*	0.201	3.628*
PEN	-0.014***	0.013	-0.014***	0.014	-0.017***	0.051***	-0.016***	0.051***
UNEMP	0.003***	-0.015	0.003***	-0.014	0.001	-0.128***	0.001	-0.124***
HPI	0.000***	-0.001	0.000***	-0.001	0.000***	-0.004**	0.000***	-0.005**
GDP	-0.000	0.106**	-0.000	0.106**	-0.014	-0.223	-0.014	-0.229
Constant	0.171*	72.202***	0.180*	72.206***	0.671**	74.712***	0.696**	74.841***
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	No	No	No	No
County Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes
SE Cluster	Year, State	Year, State	Year, State	Year, State	Year, County	Year, County	Year, County	Year, County
R-squared	0.47	0.62	0.47	0.62	0.60	0.70	0.60	0.70
N	12,600	12,600	12,600	12,600	12,600	12,600	12,600	12,600

This table reports the cross-sectional analyses by regressing county-level Prosper loan interest rate on the interaction terms between county-level race and corresponding state-level proxy of the availability of alternative information. Columns 1-4 include year and state fixed effects and columns 5-8 include year and county fixed effects. Columns 1, 3, 5, and 7 use the county-level loan failure rate as the dependent variable and columns 2, 4, 6, and 8 use county-level loan interest rate as the dependent variable. The original construction of the alternative information proxy, *Info*, is used in columns 1, 2, 5, and 6, while the alternative construction of the proxy is used in Columns 3, 4, 7, and 8. The detailed definitions of all other variables are provided in Appendix A. Continuous variables are winsorized at top and bottom 1%. Coefficients on the year, state and county indicator variables are not tabulated for brevity. The t-statistics reported in parentheses with robust standard errors clustered by year and state (columns 1-4) or year and county (columns 5-8). ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 8: The Effects of Alternative Information
Loan-level Regressions with Decomposition of the Alternative Information Proxy with Loan Interest Rate as Dependent Variable

	(1) *RATE*	(2) *RATE*	(3) *RATE*	(4) *RATE*	(5) *RATE*	(6) *RATE*	(7) *RATE*	(8) *RATE*
MP	0.9429***	0.6752***	0.7494***	0.8470***	7.2739***	7.1084***	6.4755**	5.7313**
INFO × MP	**-0.5653***				**-0.5220***			
ECOM × MP		**-0.2619**				**-0.2806*		
OFS × MP			**-0.3574**				**-0.3135**	
SNW × MP				**-0.4131*				**-0.2942**
INFO	0.1006				0.0870			
ECOM		-0.0895				-0.0780		
OFS			0.2155**				0.1860***	
SNW				0.0604				-0.0097
AMT	-0.0805***	-0.0805***	-0.0805***	-0.0805***	-0.0807***	-0.0807***	-0.0807***	-0.0807***
TERM	0.1296***	0.1296***	0.1296***	0.1296***	0.1294***	0.1294***	0.1294***	0.1294***
SCORE	-0.0760***	-0.0760***	-0.0760***	-0.0760***	-0.0761***	-0.0761***	-0.0761***	-0.0761***
INCSELF	-0.3253***	-0.3255***	-0.3255***	-0.3256***	-0.3261***	-0.3263***	-0.3262***	-0.3262***
EMPLEN	-0.0000	-0.0000	-0.0000	-0.0000	-0.0001	-0.0001	-0.0001	-0.0001
PRIOR	-0.4905***	-0.4907***	-0.4911***	-0.4910***	-0.4920***	-0.4922***	-0.4926***	-0.4924***
PRIORA	0.0386	0.0387	0.0383	0.0382	0.0455	0.0457	0.0452	0.0451
PRIORP	0.0679***	0.0679***	0.0680***	0.0680***	0.0682***	0.0682***	0.0682***	0.0682***
PRIORO	-0.0177***	-0.0177***	-0.0176***	-0.0177***	-0.0178***	-0.0177***	-0.0177***	-0.0177***
PRIORL	0.0298*	0.0298*	0.0296*	0.0297*	0.0313*	0.0313*	0.0311*	0.0312*
INCOME	-0.0008	-0.0007	-0.0010	-0.0008	-0.0381***	-0.0381***	-0.0388***	-0.0367***
EDU	-1.4175***	-1.3840***	-1.4084***	-1.3954***	-0.1694	-0.1898	-0.1402	0.0318
MALE	-0.5185	-0.6635	-0.4960	-0.5675	-11.3125	-12.5769	-12.2818	-12.3389
MARRIED	0.1552	0.0845	0.2021	0.1974	0.7490	0.8376	0.7038	0.6733
UNEMP	-0.0117	-0.0114	-0.0117	-0.0123	-0.1056***	-0.1067***	-0.1078***	-0.1055***
HPI	0.0017*	0.0017*	0.0018*	0.0017*	0.0001	0.0006	0.0006	0.0003
GDP	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Constant	69.6371***	69.8145***	69.5936***	69.6845***	74.3668***	75.1130***	74.7701***	74.8935***
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	No	No	No	No
County FE	No	No	No	No	Yes	Yes	Yes	Yes
SE Cluster	Year, State	Year, State	Year, State	Year, State	Year, County	Year, County	Year, County	Year, County
R^2	0.375	0.375	0.375	0.375	0.378	0.378	0.378	0.378
Observations	1,093,797	1,093,797	1,093,797	1,093,797	1,093,797	1,093,797	1,093,797	1,093,797

This table reports the cross-sectional analyses by regressing individual Prosper loan funding interest rate on the interaction terms between county-level race and corresponding state-level proxy of the availability of alternative information. The variables *INFO*, *ECOM*, *OFS*, and *SNW* are indicator variables as defined in Appendix A. All regressions include state and year fixed effects. The detailed definitions of all other variables are provided in Appendix A. Continuous variables are winsorized at top and bottom 1%. Coefficients on the year and state indicator variables are not tabulated for brevity. The t-statistics reported in parentheses with robust standard errors clustered by year and state. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 9: Additional Cross-sectional Tests on the Loan Level

	(1) RATE	(2) RATE	(3) RATE	(4) RATE	(5) RATE	(6) RATE
MP	0.5443***	0.5174***	0.7159***	0.6416***	0.5252***	0.5073***
PRIOR × MP	**-0.3581*** **					
PRIORO_High × MP		**-0.4056*** **				
SCORE_High × MP			**-0.5324** **			
AMT_High × MP				-0.3454***		
TERM_Long × MP					-0.1371	
INCOME_High × MP						0.0691
PRIOR	-0.2925***	0.0838***	-0.2577***	-0.1133***	-0.1148***	-0.1150***
PRIORO_High		-0.6388***				
SCORE_High			0.0060			
AMT_High				-0.7475***		
TERM_Long					3.0301***	
INCOME_High						-0.1261***
AMT	-0.0674***	-0.0675***	-0.0770***	-0.0254***	-0.0668***	-0.0668***
TERM	0.1245***	0.1242***	0.1296***	0.1254***	0.0000	0.1244***
SCORE	-0.0818***	-0.0817***	-0.0766***	-0.0818***	-0.0819***	-0.0819***
INCSELF	-0.3770***	-0.3721***	-0.3391***	-0.3713***	-0.3795***	-0.3795***
EMPLEN	0.0000	0.0000	-0.0001	-0.0000	-0.0000	-0.0000
PRIORA	-0.0551**	-0.1675***	0.1565***	0.0532**	0.0618***	0.0620***
PRIORP	0.1088***	0.1086***	0.0327***	0.0870***	0.0809***	0.0809***
PRIORO	-0.0096***	-0.0062***	-0.0090***	-0.0090***	-0.0089***	-0.0089***
PRIORL	0.0551***	0.0166*	0.1525***	0.0842***	0.0853***	0.0853***
INCOME	-0.0010	-0.0010	-0.0005	-0.0010	-0.0009	0.0004
EDU	-1.4655***	-1.4293***	-1.3299***	-1.4995***	-1.4590***	-1.0984***
MALE	-1.7764	-1.7454	-0.2579	-1.6621	-1.6873	-1.7564
MARRIED	0.0871	0.0685	0.1998	0.1036	0.1031	0.2710
UNEMP	-0.0249**	-0.0234**	-0.0132	-0.0242**	-0.0231**	-0.0223**
HPI	0.0008	0.0007	0.0017*	0.0008	0.0008	0.0009
GDP	0.0000**	0.0000*	0.0000	0.0000**	0.0000*	0.0000*
Constant	73.100***	73.0411***	68.1963***	72.7703***	77.5636***	72.9473***
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Year, State	Year, State	Year, State	Year, State	Year, State	Year, State
R^2	0.360	0.360	0.377	0.361	0.359	0.359
Observations	1,093,797	1,093,797	1,093,797	1,093,797	1,093,797	1,093,797

This table reports the cross-sectional analyses by regressing individual Prosper loan funding interest rate on the interaction terms between county-level race and several prior loans related characteristics, loan amount, loan term, credit score, and personal income. All variables are as defined in Appendix A. All regressions include state and year fixed effects. Loan interest rates are winsorized at top and bottom 1%. Coefficients on the year and state indicator variables are not tabulated for brevity. The t-statistics reported in parentheses with robust standard errors clustered by year and state. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively

Table 10: The Effect of Alternative Information—OLS Regression Estimates vs. Quantile Regression Estimates

Model	Dependent Variable	Explanatory Variable of Interest	Quantile Regression Estimates			OLS Estimates	
			0.25	0.5	0.75	Coef.	R^2
(1)	*RATECTY*	*MP*	0.938*** (11.86)	0.814*** (11.50)	0.781*** (8.15)	0.954*** (4.90)	0.60
(2)	***RATECTY***	***INFO × MP***	**-0.248* (-1.67)**	**-0.685*** (-4.54)**	**-0.966*** (-5.40)**	**-0.703*** (-2.85)**	**0.62**
(3)	*DECTY*	*MP*	0.050*** (10.44)	0.042*** (8.17)	0.055*** (6.21)	0.052*** (6.39)	0.46
(4)	***DECTY***	***INFO × MP***	**-0.018** (-2.55)**	**-0.044*** (-5.89)**	**-0.071*** (-5.18)**	**-0.055*** (-3.76)**	**0.47**
		Year Fixed Effects	Yes	Yes	Yes	Yes	
		State Fixed Effects	No	No	No	No	
		Control Variables	Yes	Yes	Yes	Yes	
		Observations	12,600	12,600	12,600	12,600	

This table reports the comparison of quantile regression estimates and OLS regression estimates. OLS regression estimates indicate the marginal effects of explanatory variables of interest on the conditional mean of the dependent variable, while quantile regressions show the marginal effects of explanatory variables of interest on the conditional 0.25, 0.5, and 0.75 quantiles of the dependent variable. The list of control variables used in quantile regressions are the same as the list control variables used in OLS regressions (shown in Table 4, columns 3 and 4). Considering the convergence issue of regression estimates when including state fixed effects, only year fixed effects are included in quantile regressions. The detailed definitions of all variables are provided in Appendix A. Coefficients on the year indicator variables are not tabulated for brevity. The continuous variables are winsorized at top and bottom 1% to eliminate the confounding effects of outliers. The t-statistics reported in parentheses are based on robust standard errors clustered by year or state. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 11: The Effects of Alternative Information—Loan-level Regressions with Alternative Construction of the Information Proxy

	(1) *RATE*	(2) *RATE*	(3) *RATE*	(4) *RATE*
MP	0.8216**	0.6685***	1.1185***	0.7140***
INFO_A × MP	**-0.4576****			
ECOM_A × MP		**-0.2918****		
OFS_A × MP			**-0.7174*****	
SNW_A × MP				**-0.2978***
INFO	0.0742			
ECOM_A		0.0854*		
OFS_A			0.1131	
SNW_A				0.0493
AMTCTY	-0.0807***	-0.0807***	-0.0807***	-0.0807***
TERMCTY	0.1294***	0.1294***	0.1294***	0.1294***
SCORECTY	-0.0760***	-0.0760***	-0.0761***	-0.0760***
INCSELFCTY	-0.3249***	-0.3249***	-0.3250***	-0.3251***
EMPLENCTY	-0.0000	-0.0000	-0.0000	-0.0000
PRIORCTY	-0.491***	-0.4913***	-0.4917***	-0.4915***
PRIORACTY	0.0397	0.0398	0.0395	0.0394
PRIORPCTY	0.0681***	0.0681***	0.0682***	0.0682***
PRIOROCTY	-0.0178***	-0.0177***	-0.0177***	-0.0177***
PRIORLCTY	0.0299*	0.0294*	0.0297*	0.0294*
INCOME	-0.0007	-0.0006	-0.0008	-0.0006
EDU	-1.3326***	-1.2731***	-1.3448***	-1.2912***
MALE	-0.1574	-0.3214	-0.0493	-0.2814
MARRIED	0.2611	0.1340	0.3074	0.2600
UNEMP	-0.0122	-0.0101	-0.0128	-0.0122
HPI	0.0017**	0.0017*	0.0018**	0.0017*
GDP	0.0000	0.0000	0.0000	0.0000
Constant	69.3169***	69.4600***	69.2311***	69.4248***
Year Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
SE Cluster	Year, State	Year, State	Year, State	Year, State
R^2	0.375	0.375	0.375	0.375
Observations	1,093,797	1,093,797	1,093,797	1,093,797

This table reports the cross-sectional analyses by regressing individual Prosper loan funding interest rate on the interaction terms between county-level race and corresponding state-level proxy of the availability of alternative information. Unlike tables 7 and 8, the state-level proxies for alternative information are constructed differently (the detailed definitions are provided in Appendix A). Columns 1-4 include year and state fixed effects and columns 5-8 include year and county fixed effects. The detailed definitions of all other variables are provided in Appendix A. Continuous variables are winsorized at top and bottom 1%. Coefficients on the year, state and county indicator variables are not tabulated for brevity. The t-statistics reported in parentheses with robust standard errors clustered by year and state (columns 1-4) or year and county (columns 5-8). ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 12: Repeat of Main Empirical Tests Using Black/African American Population Proxy

	(1) *DECTY*		(2) *RATECTY*		(3) *DECTY*		(4) *RATECTY*		(5) *RATE*		(6) *RATE*	
AA	**0.048*****	**(4.29)**	**1.058*****	**(5.34)**	0.074***	(4.34)	1.608***	(5.06)	**0.650*****	**(4.86)**	0.876***	(5.95)
INFO × AA					**-0.034****	**(-2.30)**	**-0.737****	**(-2.37)**			**-0.315****	**(-2.19)**
INFO					0.011**	(2.41)	0.148	(1.16)			0.059	(0.98)
AMTCTY	0.006***	(8.68)	-0.128***	(-9.90)	0.006***	(8.69)	-0.128***	(-9.93)	-0.081***	(-25.01)	-0.081***	(-25.02)
TERMCTY	-0.001	(-1.50)	0.099***	(14.76)	-0.001	(-1.48)	0.099***	(14.78)	0.129***	(61.54)	0.129***	(61.55)
SCORECTY	-0.000	(-0.06)	-0.076***	(-30.46)	-0.000	(-0.05)	-0.076***	(-30.61)	-0.076***	(-93.66)	-0.076***	(-93.66)
INCSELFCTY	-0.013***	(-2.80)	-0.148**	(-2.09)	-0.013***	(-2.79)	-0.147**	(-2.08)	-0.325***	(-25.70)	-0.325***	(-25.70)
EMPLENCTY	-0.000	(-0.70)	-0.000	(-0.25)	-0.000	(-0.73)	-0.000	(-0.30)	-0.000	(-0.75)	-0.000	(-0.75)
PRIORCTY	-0.017*	(-1.75)	-0.166	(-1.00)	-0.017*	(-1.77)	-0.173	(-1.04)	-0.492***	(-9.61)	-0.492***	(-9.61)
PRIORACTY	-0.028	(-1.51)	0.679**	(2.11)	-0.028	(-1.47)	0.693**	(2.14)	0.042	(0.54)	0.042	(0.54)
PRIORPCTY	0.002***	(2.74)	-0.028	(-1.56)	0.002***	(2.70)	-0.028	(-1.56)	0.068***	(12.54)	0.068***	(12.53)
PRIOROCTY	-0.001**	(-2.56)	-0.002	(-0.23)	-0.001**	(-2.55)	-0.002	(-0.22)	-0.018***	(-7.50)	-0.018***	(-7.50)
PRIORLCTY	0.016**	(2.05)	-0.112	(-1.06)	0.016**	(2.08)	-0.109	(-1.04)	0.029*	(1.83)	0.030*	(1.83)
INCOME	0.000***	(3.17)	-0.002	(-1.00)	0.000***	(3.15)	-0.002	(-1.04)	-0.001	(-0.53)	-0.001	(-0.49)
EDU	-0.180***	(-5.69)	-1.979***	(-4.63)	-0.180***	(-5.70)	-1.982***	(-4.64)	-1.442***	(-5.59)	-1.475***	(-5.64)
MALE	0.074	(1.13)	2.549**	(2.48)	0.074	(1.13)	2.554**	(2.49)	0.371	(0.29)	0.429	(0.34)
MARRIED	0.226***	(8.14)	0.483	(1.14)	0.225***	(8.13)	0.470	(1.12)	0.236	(0.72)	0.241	(0.75)
PEN	-0.013***	(-21.26)	0.015*	(1.70)	-0.013***	(-21.28)	0.015*	(1.73)	0.008	(0.99)	0.008	(0.95)
UNEMP	0.003***	(3.87)	-0.018	(-1.50)	0.003***	(3.81)	-0.019	(-1.56)	-0.005	(-0.57)	-0.006	(-0.62)
HPI	0.000***	(5.70)	-0.001	(-0.52)	0.000***	(5.75)	-0.000	(-0.45)	0.003***	(2.98)	0.003***	(3.00)
GDP	0.001	(0.27)	0.112**	(2.25)	0.001	(0.29)	0.113**	(2.29)	0.000	(0.21)	0.000	(0.26)
Constant	0.178*	(1.94)	72.206***	(33.91)	0.172*	(1.88)	71.809***	(34.63)	68.898***	(46.54)	68.831***	(46.43)
Control Variables	Yes		Yes		Yes		Yes		Yes		Yes	
Year Fixed Effects	Yes		Yes		Yes		Yes		Yes		Yes	
State Fixed Effects	Yes		Yes		Yes		Yes		Yes		Yes	
SE Cluster	Year, State		Year, State		Year, State		Year, State		Year, State		Year, State	
R-squared	0.47		0.62		0.47		0.62		0.38		0.38	
N	12,600		12,600		12,600		12,600		1,093,797		1,093,797	

This table reports the main empirical analyses by replacing the main proxy for race with one that measures only the Black/African American population percentage of a county. Columns 1-4 repeat the tests at county level and columns 5 and 6 repeat the main tests at loan level. The detailed definitions of all variables are provided in Appendix A. Continuous variables are winsorized at top and bottom 1%0.01. Coefficients on the year and state indicator variables are not tabulated for brevity. The t-statistics reported in parentheses with robust standard errors clustered by year and state. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Chapter 3: Digital Inclusion and Financial Inclusion: Evidence from Peer-to-peer Lending

Abstract

I examine the influence of digital inclusion on financial inclusion. Using evidence from a sizable P2P lender in the U.S., I document that digital inclusion is positively associated with P2P lending penetration, with such relation more pronounced in county-years with more vulnerable/excluded populations. The results are robust to the use of the instrumental variable (2SLS) approach, alternative measurements, weighted least squares regression, additional controls, and single-year analysis. In addition, I document that higher risk borrowing is less likely to be denied in county-years with higher digital inclusion. This study emphasizes the crucial role of digital inclusion in financial inclusion.

Keywords: FinTech; Peer-to-peer lending; Digital inclusion; Financial inclusion; Credit scoring; Non-traditional information

JEL codes: G23 G28 J15 L86

1. Introduction

Financial inclusion refers to the provision of various types of financial services to underserved populations. Such financial services include, for example, the provision of deposit or transaction accounts, access to credit, and mobile payment (e.g., Demirgüç-Kunt et al. 2017; Fernandez-Olit et al. 2020). Although financial inclusion has improved evidently during the past decade worldwide, salient gaps in access to financial services still exist. For example, according to the 2021 World Bank Global Financial Inclusion database, despite salient gaps, the global average percentage of adults with an account increased from 51% in 2011 to 76% in 2021 (Demirgüç-Kunt et al. 2022). Digital inclusion, a public policy designed to provide high speed internet infrastructure for historically digitally excluded populations, has gained much attention from policymakers as a mechanism to facilitate financial inclusion. For example, Villasenor et al. (2015) emphasize the importance of digital infrastructures, such as reliable and accessible internet services, for digital financial services to reach the excluded populations. Similarly, Demirgüç-Kunt et al. (2022) consider digital inclusion an effective way to boost the account ownership of hard-to-reach populations. Nevertheless, despite the increasing policy attention, few existing studies examine how digital inclusion is related to financial inclusion.

I explore the relationship between digital inclusion and one key dimension of financial inclusion, namely, access to credit. Specifically, I examine how digital inclusion is related to the credit access facilitated by peer-to-peer (P2P) lenders. Unlike traditional banks, P2P lenders match potential borrowers and lenders through a platform and pre-screen borrowers with big data and machine learning algorithms (e.g., Philippon 2016; Fuster et al. 2019, 2022; Wu and Zhang 2021). Researchers have been increasingly interested in how P2P lending can facilitate financial inclusion. For example, Philippon (2019) argues that P2P lenders can reduce lending discrimination due to their effective utilization of non-traditional data. Studies also document that P2P lending tends to penetrate areas that have lost bank branches and areas with highly concentrated banking markets (e.g., Jagtiani and Lemieux 2017; Hodula 2022).

I posit that digital inclusion is positively related to P2P lending penetration (or P2P penetration). On the supply side, people may be excluded by banks due to their poor (or absence of) standard credit bureau scores. These underserved populations may benefit from digital inclusion thanks to P2P lenders' effective utilization of non-traditional information, which is often collected via online activities. For example, Berg et al. (2020) show that online digital footprints can match the information content of credit bureau scores. Similarly, Yu and Zhang (2021) show that with the right algorithm, even small samples can perform relatively well in internet loan credit risk evaluation. In this sense, digital inclusion can increase P2P lenders' ability to assess the creditworthiness of the underserved population, hence increasing the credit supply and P2P penetration. On the demand side, people doing more online activities are more likely to be aware of the P2P lending options, especially those whom banks exclude. Hence, digital inclusion can facilitate borrowers' active search for credit providers online, leading to increased P2P lending penetration. Consistent with this reasoning, I also predict that the relationship between digital inclusion and P2P lending should be more pronounced in areas where there are more underserved populations.

I examine the relationship between digital inclusion and P2P lending penetration using a large sample of loan applications data from Prosper Marketplace (Prosper), which is considered as one of the largest P2P lenders in the U.S.[53] I measure the level of digital inclusion at the U.S. county-year level based on Federal Communications Commission (FCC) Form 477 data.

[53] A 2020 market research report shows that LendingClub and Prosper are the two largest P2P lenders in the U.S. (https://mangosoft.tech/blog/top-5-peer-to-peer-lending-companies-2020-full-market-research)

Specifically, I use the number of residential fixed broadband connections with a downstream speed of at least 10 Mbps per 1,000 housing units as the main proxy for digital inclusion. For robustness checks, I also use an alternative measure of digital inclusion, which is the first principal component of a factor analysis on three variables, including a measure of high-speed internet connections, a measure of basic internet connections, and a measure of households' high-speed internet subscriptions. Because digital inclusion is measured at the county level, I collapse the Prosper loan level data into 28,027 county-year observations spanning from 2009 to 2020, so that P2P penetration is also measured at the county-year level. My primary measure of P2P penetration is calculated as the county-year-level successfully funded Prosper loans in dollar amounts per capita. In robustness tests, I also use four alternative measures of P2P penetration. Consistent with my prediction, I find consistent evidence that digital inclusion is positively associated with the extent of P2P lending penetration. This effect is also economically significant. Using the primary measure of digital inclusion and P2P penetration, I find that a one standard deviation increase in digital inclusion is associated with a 13.35% to 24.52% increase in P2P penetration.

I conduct several additional analyses to strengthen the validity of my main findings. First, I employ four alternative measures of P2P penetration: the count of successfully funded Prosper loans per capita, the rank of successfully funded Prosper loans amounts per capita, the amount of successfully funded Prosper loans divided by local commercial banks' consumer loans, and the amount of successfully funded Prosper loans divided by local commercial banks' total loans. Using all four alternative measures, I find that digital inclusion remains significantly positively associated with P2P penetration. Second, I employ instrumental variable estimation to control for potential endogeneity. I use the county-year level of basic internet connections as an external instrument for measures of P2P lending penetration. I find consistent results using the instrumental variable approach. Third, I conduct a battery of sensitivity tests to ascertain the robustness of the main findings. Specifically, my results remain robust to using weighted-least squares regressions, the use of several additional county-year levels of local commercial bank controls, and the use of the alternative measure of digital inclusion. Finally, to further alleviate the concern that my results are confounded by the upward trending of both P2P penetration and digital inclusion over the years, I conduct the regression analyses on a year-by-year basis. I find that digital inclusion remains positively and significantly related to P2P penetration with the trending concerns eliminated.

To support my main reasoning that digital inclusion positively influences P2P penetration by increasing the information availability and credibility for excluded populations, I perform cross-sectional tests to examine whether the influence of digital inclusion on P2P penetration is more pronounced in areas with more vulnerable and/or excluded populations. Specifically, I document a stronger influence of digital inclusion on P2P penetration in county-years with a lower level of commercial banks' total loans per capita and consumer loans per capita. I also find that the relation between digital inclusion and P2P penetration is stronger in counties with greater minority proportions. These results suggest that digital inclusion plays a key role in financial inclusion, particularly in regions with more vulnerable and/or underserved populations.

Finally, I explore further the consequences of the influence of digital inclusion on P2P lending penetration. I find that higher-risk listings are less likely to be denied when digital inclusion is higher and that P2P lenders tend to rely less on credit bureau information when determining the P2P loan interest rates. The consequences tests support my reasoning that when digital inclusion is higher, P2P lenders and investors can utilize more alternative information, which makes higher-risk borrowing easier.

This study makes several contributions. First, this study is among the first to explore the relationship between digital inclusion and P2P penetration. I note that Corrado and Corrado (2015) and Zhong and Jiang (2021), using P2P lending data from Europe and China, respectively, document a positive relation between P2P lending participation and internet development. This study is different from Corrado and Corrado (2015) and Zhong and Jiang (2021) in the following aspects: 1) I focus on the U.S. credit market that contains diverse ethnic groups, which are well documented to be prone to racial lending discrimination (e.g., Bartlett et al. 2022; Butler et al. 2022), allowing me to use cross-sectional tests to support the role of digital inclusion in facilitating P2P lending penetration, particularly in areas with more vulnerable populations; 2) I perform cross-sectional tests to show that the influence of digital inclusion on P2P inclusion is more salient in areas that are underserved by traditional banks, and 3) I provide several additional consequences tests. Second, this study extends the literature that examines the determinants of P2P lending growth. For example, existing studies have attributed the rapid growth of P2P lending to regulatory arbitrage (e.g., Buchak et al. 2018; Tang 2019; De Roure et al. 2022), technological advantages (e.g., Fuster et al. 2019; Hau et al. 2019; Frost et al. 2019; Berg et al. 2020), and market structure (Balyuk et al. 2020). This study adds to the literature by documenting digital inclusion as a plausible factor for the growth of P2P lending. This study also provides new insights concerning the effectiveness of microfinance institutions by documenting the role of digital inclusion in facilitating more effective credit information that can be utilized by both banks and P2P lenders. Third, this study emphasizes the role of P2P lending in decreasing the information friction in the consumer credit market.

This study corroborates policymakers' digital inclusion policy initiatives[54] by providing evidence of a significant economic implication, namely, financial inclusion. By studying digital inclusion and financial inclusion in the P2P lending context, this study also serves to generalize the statistical inferences to the traditional banking sector, implying that the promotion of digital inclusion and better utilization of big data in the traditional banking sector can help facilitate financial inclusion.

The structure of the paper is as follows. Section 2 discusses the related literature and provides the hypothesis development; section 3 describes empirical design; section 4 describes the sample and main empirical results; section 5 presents the results of cross-sectional analyses; section 6 reports the results of consequences tests, and section 7 offers concluding remarks.

2. Related Literature and Hypothesis Development

According to the existing literature, financial inclusion can be referred to as the provision of access to the financial system, particularly for the poor and disadvantaged populations (e.g., Leyshon and Thrift 1995; Kempson and Whyley 1999). Such access ranges from basic financial services, such as having a deposit/transaction account that can be used to receive payments and save money, to advanced financial services like access to credit and the use of formal insurance products (Demirgüç-Kunt et al. 2017). In an early study focusing on Great Britain, Kempson and Whyley (1999) report that households with a lower income, a lower level of education, single parent, and classified as African-Caribbean or Black, Pakistani, or Bangladeshi are associated with a higher likelihood of financial exclusion. Despite the increasing policy and academic attention paid to the issue of financial inclusion over the past decades, recent evidence still shows concerns

[54] For example, in November 2021, U.S. President Joe Biden signed the Infrastructure Investment and Jobs Act into law and provided $65 billion for broadband: https://broadbandusa.ntia.doc.gov/resources/federal/federal-funding

regarding the lack of financial inclusion around the world. According to the Global Findex 2021 survey, despite the recent progress, only 53% of the global adult population have borrowed any money in the past 12 months, including the use of a credit card (Demirgüç-Kunt et al. 2022). Also, the survey indicates that women, the lower-income populations, the young, and those outside the workforce continue to have lower account ownership rates (Demirgüç-Kunt et al. 2022).

In addition, financial inclusion is important—existing studies document a positive real impact of financial inclusion on various economic outcomes, such as poverty, inequality, and financial stability. Bruhn and Love (2014) explore how access to finance influences poverty by utilizing exogenous bank branch openings in previously underserved areas in Mexico. They find that access to finance reduces poverty through the channel of labor market activity. They also document that the effect of financial inclusion on reducing poverty is more pronounced among low-income populations and those areas with lower prior bank penetration. In a cross-country study, Beck et al. (2007) document that financial inclusion, measured by private credit, reduces income inequality. Moreover, several studies suggest that financial inclusion, such as increased access to bank deposit and borrowing, can increase the financial system resilience in times of financial stress (e.g., Han and Melecky 2013; Hannig and Jansen 2010; Khan 2012; Ozili 2018).

Increasingly, policymakers are focusing on digital inclusion as a means of facilitating financial inclusion. Villasenor et al. (2015) emphasize the importance of digital infrastructure, such as reliable and accessible internet services, for digital financial services to reach hard-to-reach groups. Demirgüç-Kunt et al. (2022) also consider digital inclusion a valuable tool for increasing account ownership of the excluded populations. Lapukeni (2015) suggests that the increasing mobile phone subscriptions in Africa indicate opportunities for information and communication technology to facilitate financial inclusion. However, the related empirical evidence on the relationship between digital inclusion and financial inclusion is scarce. Among the exceptions, Corrado and Corrado (2015) examine the potential determinants of financial inclusion and document that people who are more "internet-connected" are also more likely to be financially included. They reason that the new information technologies can allow easier information flows, which reduces costs and simplifies credit and deposit-taking. Another exception is Zhong and Jiang (2021), who explore the P2P lending markets in China and find that P2P lending participation is positively related to the degree of internet development and negatively related to the access level of traditional finance. This study is different from Corrado and Corrado (2015) and Zhong and Jiang (2021), who focus on Europe and China, respectively, in the following aspects. First, I focus on the U.S. market, which contains diverse ethnic groups, allowing me to explore whether the influence of digital inclusion is stronger for minority groups. Second, I provide several cross-sectional tests to reinforce the channel that digital inclusion can promote individuals' use of the internet and dissemination of non-traditional information, which can be effectively utilized by P2P lenders when evaluating the creditworthiness of the previously excluded populations. Third, I explore the potential consequences of digital inclusion on the determination of loan denial and loan interests.

Digital inclusion can be positively associated with financial inclusion. On the supply side, traditional banks may reject applicants if they have subpar or no credit scores from the major credit bureaus. P2P lenders can effectively utilize non-traditional information, which is frequently gathered via online activities, for these marginalized people. For instance, Berg et al. (2020) demonstrate that the informational content of credit bureau scores may be matched with online digital footprints. Yu and Zhang (2021) demonstrate that with the correct algorithm, even small samples can evaluate credit risk for internet loans reasonably well. In this way, digital inclusion can improve the capacity of P2P lenders to judge the creditworthiness of the underserved population, thus increasing P2P penetration. On the demand side, people with more online footprints are more exposed to advertisements and information about internet finance providers, increasing the likelihood of seeking credit from P2P lenders. Based on the above discussion and prior findings, I propose the following hypothesis:

H1: Digital inclusion is positively related to P2P lending penetration.

3. Research Design

3.1 Institutional Background and Measurements

I use five measures of P2P lending penetration. All the measures are based on loan applications (or loan listings) data from Prosper Marketplace. Prosper adopts a business model in which all the listed loan interest rates are pre-determined by Prosper according to its credit pricing system that incorporates both credit bureau scores and alternative information. Investors then bid for loan amounts. Under this business model, Prosper utilizes alternative information such as "group affiliations, the general economic environment, and competitive conditions." [55] In this sense, I expect that digital inclusion could promote Prosper's use of such alternative information, thus increasing the likelihood of supplying loan credits to previously underserved populations.

The primary measure of P2P penetration is the county-year-level total amount of successfully funded Prosper loans per capita (*P2PPEN*). The secondary measure of P2P penetration, which captures the penetration in the dimension of loan numbers, is calculated as the county-year-level number of successfully funded Prosper loans per capita (*P2PPEN_CT*). Also, considering that the main measure has a nontrivial number of large outliers, I use a third measure

[55] A more illustrative version of the Prosper credit rating system is shown in Table 1, Panel A and Panel B.

that is constructed as the rank of *P2PPEN* across all county-years (*P2PPEN_RK*). This measure minimizes the bias of the prevalence of extreme values of *P2PPEN*. In addition, to minimize the possible bias that county-level P2P penetration could also be driven by some unobservable factors that influence the overall creditworthiness at the county level, I use two more proxies that measure P2P penetration relative to the local commercial bank loans. Specifically, *P2PPEN_CS* is calculated as the county-year-level successfully funded Prosper loans amount divided by the county-year-level commercial bank consumer loans amount. *P2PPEN_TL* is calculated as the county-year-level successfully funded Prosper loans amount divided by the county-year-level commercial bank total loans amount.

My primary measure for digital inclusion is based on Federal Communications Commission (FCC)'s Form 477 data, which contains internet access service data filed by all facilities-based broadband providers at the U.S. county level.[56] I use the residential fixed broadband connections with a downstream speed of at least 10 Mbps per 1,000 housing units as my primary proxy for digital inclusion (*DI*), which is coded from 0 to 5, with 5 indicating the highest level of digital inclusion. In specific, let X be the number of high-speed internet connections per 1,000 housing units. I denote $DI = 0$ if $X = 0$; $DI = 1$ if $0 < X <= 200$; $DI = 2$ if $200 < X <= 400$; $DI = 3$ if $400 < X <= 600$; $DI = 4$ if $600 < X <= 800$; and $DI = 5$ if $800 < X$. This coding methodology is recommended and used by the FCC. For robustness purposes, I also use a secondary measure of digital inclusion (*DI_ALT*), which is the first principal component extracted from a principal component analysis (PCA) of the following three variables: the main measure of digital inclusion (*DI*), a measure of the county-year-level basic internet connections (*T1_INT*, the data is also from FCC Form 477, which measures the number of fixed internet connections greater

[56] Source: https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477

than 300 kbps per 1,000 households), and the number of county-year households with high-speed internet subscriptions (*HI_SUB*, the data is from the U.S. Census Bureau).

3.2 Empirical Models

To test H1, I estimate the following cross-sectional baseline regression, with robust standard errors of the estimates clustered by year and county:

$$PEPPEN = \alpha_0 + \alpha_1 DI + \alpha_2 V + \alpha_3 D + \alpha_4 E + \alpha_5 B + YEAR\_FE + STATE\_FE + \varepsilon \quad (1)$$

where *PEPPEN* is the dependent variable. In additional tests, I also use *P2PPEN_CT*, *P2PPEN_RK*, *P2PPEN_TL*, and *PEPPEN_CS* as dependent variables. The main explanatory variable is *DI*, which is the measure of digital inclusion. I also use *DI_ALT* for robustness checks. *V* is a vector of loan-level characteristics, all collapsed into the county-year level, *D* is a vector of county-year-level demographic variables, *E* is a vector of county-year-level macroeconomic variables, and *B* is a vector of county-year-level commercial bank variables. I use year fixed effects (*YEAR_FE*) to control for the confounding effects of different years, mitigating the concern that digital inclusion and P2P penetration can vary concurrently with time. I use state fixed effects (*STATE_FE*) to control for unobservable cross-state time-invariant differences that can potentially confound the main results. I provide detailed definitions of all the variables in the Appendix. H1 hypothesizes that digital inclusion is positively associated with P2P lending penetration; hence, I expect α_1 to be positive.

I select loan characteristics controls based on Prosper's credit pricing system, considering that the local P2P penetration level can be closely related to the quality of P2P loan applications. According to Prosper's credit pricing system, after the initial screening for minimum credit scores and identity and anti-fraud checks, Prosper uses a credit pricing matrix (10 by 10) to determine the initial credit rating for loan applications. One dimension of the matrix is credit bureau scores—I use the FICO score (*SCORE*) to control for this dimension. The other dimension of the matrix is a "custom risk model" based on possible data related to the borrowers.[57] I use the borrower's self-disclosed income range (*INRANGE*), self-disclosed employment tenure (*LENEMP*), and prior interactions with the platforms (*PRIOR*) to control for this dimension. Because Prosper adjusts the baseline interest rates based on the loan term, I also control for the loan term (*TERM*). Finally, loan interest rate (*RATE*) is used to control for the overall credit quality. In regression analysis, all loan characteristics controls are collapsed at the county-year level.

Next, I include a comprehensive set of county-year-level demographic controls and county-year-level macroeconomic controls to further control for the potential confounding factors that could both determine the local level of digital inclusion and P2P lending penetration. In terms of demographic controls, I control for county-year-level personal income per capita (*INCOME*), county-year-level proportion of the adult population receiving post-secondary education (*EDU*), county-year-level proportion of the population that is not White (*RACE*), county-year-level proportion of the population that is male (*GENDER*), and county-year-level proportion of the adult population that is currently married (*MARRY*). In terms of macroeconomic controls, I control for the county-year-level unemployment rate (*UNEMP*), county-year-level housing pricing index (*HPI*), and the natural logarithm of county-year-level GDP per capita (*LNGDP*).

In addition, to further alleviate the concern that certain aspects of local population credit quality are not fully captured with the comprehensive set of control variables, I include variables that control for the local commercial banking environment. When constructing these controls, I require only community banks to minimize the bias of large banks operating across counties and states. In the baseline regressions, I control for the county-year-level total commercial bank loans

[57] Prosper claims that the "custom risk model" uses historical Prosper data and is built on the Prosper borrower population, i.e., the model inputs all historical Prosper loan records and makes predictive analysis.

per capita (*TLPC*) and the county-year-level total commercial bank assets per capita (*TAPC*). More bank controls are added in additional analysis, including the county-year commercial bank loan loss provisions scaled by lagged total loans (*LLP*), county-year-level commercial bank loan loss allowances scaled by lagged total loans (*LLA*), county-year-level commercial bank net loan charge-offs scaled by lagged total loans (*NCO*), county-year-level commercial bank non-performing loans scaled by lagged total loans (*NPL*), and county-year-level commercial bank Tier 1 capital ratio (*T1R*).

4. Results

4.1 Sample

I retrieve detailed P2P listings data from 2009 to 2020[58] from Prosper Marketplace. The dataset contains both funded and unfunded loan listings. I conduct the following sample selection process: First, I restrict the sample to the contiguous United States. Second, I match each listing's city and state to U.S. counties and delete unmatched observations. Third, I collapse the data by county and year, match the collapsed data to the county-year-level digital inclusion measures, and delete unmatched observations. Fourth, I match the data with county-level demographic data obtained from the U.S. Census 5-Year ACS. Fifth, I match the data with macroeconomic variables including county-level unemployment rates, the housing price index, and the natural logarithm of GDP per capita. Sixth, I match the data with county-year-level commercial bank-related variables obtained from Call Reports.[59] Considering the prevalence of outliers, I winsorize P2P lending

[58] Considering that starting from 2020, Prosper added a new product—home equity line of credit—I also use a cutoff sample period up to the end of 2019. I find consistent results on all tests.

[59] Call Report is short for Consolidated Reports of Condition and Income. All national banks, state member banks, insured state nonmember banks, and savings associations are required to submit Call Report data to bank regulators.

penetration and digital inclusion measures at the top and bottom 1% levels.[60] The final sample contains 22,698 county-year observations for 3,000 counties from 2009 to 2020.

4.2 Descriptive Statistics

Table 1 reports the descriptive statistics and correlations of the regression variables for the full sample. Table 1 Panel A reports the descriptive statistics for regression variables at the county level. In terms of the P2P penetration variables, the mean (median) county-level amount of successfully funded P2P loans per capita is 4.61 (3.37); the mean (median) county-level number of successfully funded P2P loans per 10,000 people is 3.81 (2.96); the mean (median) county-level amount of successfully funded P2P loans per thousand dollars of local community bank consumer loans is 110.51 (7.61); and the mean (median) county-level amount of successfully funded P2P loans per thousand dollars of local community bank total loans is 1.27 (0.31). With respect to the digital inclusion measurements, the mean (median) county-level rating of the residential high-speed internet is 2.34 (2.00), indicating that on average, about 4 out of 10 housing units have residential fixed connections of at least 10 Mbps (download)/1 Mbps (upload).

In addition, the mean (median) county-level P2P lending interest rate is 17.53% (15.98%). The mean (median) county-level loan term is 42.33 (42.45) months. The mean county-level percentage of P2P loans that have prior borrowing records on the platform is 21.90% (16.67%), suggesting that on average, 21.9% of the loans are from repeat borrowers. The mean (median) P2P successfully funded loans' FICO score is 717.49 (716.50). Tang (2019), who uses data from LendingClub loans from 2009 to 2012, reports a mean FICO score of 652 for all loan listings, and a mean interest rate of 13.3% for funded loans, suggesting that applicants' characteristics of these two platforms are comparable despite different sample periods.

---

[60] While not tabulated, I document similar empirical results with a winsorization at the top and bottom 2% levels, considering the prevalence of outliers of the P2P penetration measures.

Moreover, Table 1 Panel B reports Spearman (upper diagonal) and Pearson (lower diagonal) correlations between the variables in my analyses at the county level. Consistent with my predictions, each of the five measures of P2P loan penetration is positively associated with the two proxies for digital inclusion. Because these are pairwise univariate correlations, I defer inferences to the multivariate tests reported in the following section.

[Insert Table 1 Here]

4.3 Empirical Results

4.3.1 Main Analysis: Digital Inclusion and P2P Lending Penetration

This section reports the results of the test of H1, which examines the association between digital inclusion and P2P lending penetration. Table 2 presents the regression results. In Column (1), I regress P2P penetration (*P2PPEN*) on digital inclusion without loan characteristics, demographic, macroeconomic, and local banking environment control variables. In Column (2), I report the results including only loan characteristics and demographic control variables. In Column (3), I report the results including loan characteristics, demographic, and macroeconomic control variables, and in Column (4), I report the results including loan characteristics, demographic, macroeconomic, and local banking environment control variables. In all four columns, I include year and state fixed effects with standard errors clustered by year and county. In all columns, I report a positive and significant coefficient (all with p-value < 0.01) on *DI*, indicating that P2P lending penetration increases with county-level digital inclusion. The relation between digital inclusion and P2P lending penetration is also economically significant. From Column (1) to Column (4), a one standard deviation increase in *DI* is associated with a 24.52%, 13.35%, 15.68%, and 15.18% increase, respectively, in P2P penetration proxied by *P2PPEN*.[61] Overall, the results reported in Table 2 indicate that digital inclusion plays an economically significant role in facilitating P2P lending penetration.

[Insert Table 2 Here]

4.3.2 Digital Inclusion and P2P Penetration—Alternative Measures of P2P Penetration

In this section, I report the regression results of the baseline model with all the loan characteristics, demographic, and macroeconomic control variables, using four alternative measures of P2P lending penetration. As aforementioned, the four measures include the number of P2P loans per capita (*P2PPEN_CT*), an ordinal measure of *P2PPEN* (*P2PPEN_RK*), the ratio of P2P loans to bank total loans (*P2PPEN_TL*), and the ratio of P2P loans to bank consumer loans (*P2PPEN_CS*). In Table 3 Column (1), I report the results using *P2PPEN_CT* as the measure of P2P penetration. Column (2) shows the results using *P2PPEN_RK* as the measure for P2P penetration. Column (3) shows the results using *P2PPEN_TL* as the measure for P2P penetration. Column (4) shows the results using *P2PPEN_CS* as the measure for P2P penetration. In all four columns, I include state and year fixed effects with standard errors clustered by county and year. In all columns, I report a positive and significant coefficient (all with p-value < 0.01) on *DI*, indicating that P2P lending penetration, proxied by these alternative measures, increases with county-level digital inclusion. With these measures, the relationship between digital inclusion and P2P lending penetration is still economically significant. From Column (1) to Column (4), a one standard deviation increase in *DI* is associated with a 12.46%, 6.24%, 56.03%, and 73.79% increase, respectively. The results shown in Table 3 suggest that digital inclusion plays a

[61] Using Column (1) as an example, the impact of a one standard deviation increase in *DI* on *P2PPEN* is computed as 0.9953 (the coefficient of *DI*) × 1.1357 (the sample standard deviation of *DI*) ÷ 4.6092 (the sample mean of *P2PPEN*) × 100% = 24.52%.

significant role in the four other different dimensions of P2P penetration. In particular, the large economic magnitude of the coefficients in Column (3) and Column (4) indicates that digital inclusion may contribute to the competitive advantages of P2P lenders relative to local community banks.

[Insert Table 3 Here]

4.3.3 Instrumental Variable (2SLS) Estimation

The main results of this study may be biased due to omitted correlated variables because it is difficult to control for all plausible cross-county characteristics that are potentially related to both the level of digital inclusion and the P2P lending penetration. To mitigate potential endogeneity concerns, I employ instrumental variable (2SLS) estimation.

I use the county-year-level basic internet connections (*T1_INT*) as an external instrument for measures of P2P lending penetration.[62] A good instrument should be highly correlated with digital inclusion, but not have a direct effect on P2P penetration (Roberts and Whited 2012). On the exclusion criterion, I argue that the variations in basic internet connections across counties are unlikely a direct determinant for P2P penetration for the following important reasons. First, digital inclusion is a dynamic concept, which depends on continuous advances in technology. With the development of technology, online service providers continuously update their service that relies on increasing internet speeds (e.g., Gant et al. 2010). For example, in 2015, the FCC set "26-Mbps down/3-Mbps up" as the new speed benchmark for broadband service, which is also considered as the basic requirement for individuals' quality use of online services such as e-commerce and socialization (e.g., Rhinesmith 2016; Reisdorf and Rhinesmith 2020; Sanders and Scanlon 2021). Second, throughout the past decade, studies suggest that it is the variations in access to high-speed

[62] Instrumental analyses using two measures of P2P penetration, *P2PPEN* and *P2PPEN_CT*, are displayed in Table 4. While not tabulated, the results are similar using other alternative measure of P2P penetration.

internet, rather than basic internet connections, that contribute to social and digital inclusion. For example, Wallace et al. (2017) indicate that it is high-speed internet (as opposed to basic internet connections) that plays a key role in the local development of digital and social inclusion. Rodriguez et al. (2022) suggest that bandwidth limitation could exclude patients with slower internet speeds from accessing healthcare services, highlighting the role of high-speed internet. In an earlier study focusing on access to information by public libraries, Jeager et al. (2012) argue that with the continuous development of the internet and computer technologies, it is the high-speed internet that serves to close the digital divide and promote digital literacy. Lastly, my data suggests that the variations in basic internet connections across U.S. counties are relatively small (i.e., the standard deviation of *T1_INT* is 0.81, compared to a much larger standard deviation of high-speed internet coverage, which is 1.14). Therefore, it is unlikely that basic internet connections (which a large proportion of individuals in the U.S. have access to) contribute to the differences in people's access to sophisticated online activities and services. In other words, *T1_INT* is unlikely to be directly related to P2P penetration and therefore meets the exclusion criterion. Moreover, *T1_INT* is positively associated with the county-level coverage of high-speed internet (*DI*) in that by the measurement methodology, *T1_INT* is a necessary condition of *DI*. Therefore, *T1_INT* is likely to be a valid instrument.

The results of the first-stage regressions are reported in Table 4, columns (1) and (3). Consistent with my expectation, *T1_INT* is significantly and positively associated with *P2PPEN* and *P2PPEN_CT.*[63] I then use the predicted values of digital inclusion from the first-stage regressions as the instrument in the second stage and test the prediction in H1. I present the results

[63] As suggested by Roberts and Whited (2012), I formally test the strength of the instrumental variable by computing the partial F-statistic for the instrument used in the first-stage regressions. The partial F-statistic is 3705.05 in both analyses, considerably higher than the suggested minimum benchmark of 8.96 for a model with one instrument, as reported by Stock and Yogo (2005). Overall, it is likely that the analyses do not suffer from a weak instrument problem.

in Table 4, columns (2) and (4). The results show that the predicted value of digital inclusion (*Pred_DI*) is significantly positively associated with P2P penetration, which is consistent with the results of the test of H1 reported in Table 2 and Table 3. Overall, the results from the instrumental variable estimation mitigate concerns that the main results are driven by potentially omitted correlated variable problems.

[Insert Table 4 Here]

4.3.4 Additional Tests

In this section, I report the results of several additional robustness and sensitivity tests. First, in Table 5, columns (1) through (5), I report the regression results using the alternative measure of digital inclusion (*DI_ALT*). Throughout the five regression analyses, *DI_ALT* remains positively significantly related to five different measures of P2P lending penetration (all with p-values < 0.01). Second, as the dataset for this study is an unbalanced panel with missing values, I employ a weighted least squares (WLS) approach so that each county or year in the sample receives equal weight in the regression estimation and no single county or single year drives the result (Dittmar et al. 2003). The WLS approach results are shown in Table 6 columns (1) and (2). In Column (1), the weighting scheme is the inverse of year frequency and in Column (2), the weighting scheme is the inverse of county frequency. The findings are robust for these weighting schemes. Third, in Table 6 columns (3) and (4), I include six additional county-level community commercial bank variables to further control for local banking environments. In particular, I include bank loan loss provisions to control for local credit quality because loan loss provisions can contain information about the current and expected local market conditions (Khan and Ozel 2016). Other local bank variables, including *CSLPC, LLA, NCO, NPL,* and *T1R*, are also selected to control for local creditworthiness. With the six additional commercial bank-related variables, the main relationship between digital inclusion and P2P lending penetration remains significantly positive. Last, another concern of the documented main results is that the positive relationship between digital inclusion and financial inclusion may be affected by the upward trend of both P2P penetration and digital inclusion over the sample years. To alleviate this concern, I conduct regression analyses on a year-by-year basis. The results of the year-by-year analysis are reported in Table 7. Table 7 columns (1) through (5) report the regression results for each of the years from 2015 to 2019 using the main measure of digital inclusion. Table 7 columns (6) through (10) report the year-by-year regression results using a dummy variable measure of digital inclusion. Across all tests, I find that digital inclusion remains positively and significantly related to P2P penetration with the trending concerns eliminated.

[Insert Table 5, 6, and 7 Here]

5. Cross-sectional Analyses

In the main analysis, I find robust evidence that digital inclusion is positively associated with the extent of P2P penetration. Next, I examine whether the influence of digital inclusion on P2P penetration is systematically stronger in regions where there are more excluded and/or vulnerable populations. To do so, I modify equation (1) to include the conditional variable (*CONV*) and its interaction with *DI*, and estimate the following cross-sectional regression, with robust standard errors clustered by year and county:

$$PEN = \alpha_0 + \alpha_1 DI \times CONV + \alpha_2 V + \alpha_3 D + \alpha_4 E + \alpha_5 B + YEAR\_FE + STATE\_FE + \varepsilon \qquad (2)$$

where *CONV* stands for one of the four conditional variables, all in dummy form. The first conditional variable is *BP1*, which is a dummy variable that equals 1 if the county-level commercial bank total loans per capita is in the highest quantile, and 0 if it is in the lowest quantile. The second conditional variable is *BP2*, which is a dummy variable that equals 1 if the county-level commercial bank consumers loans per capita is in the highest quantile, and 0 if it is in the lowest quantile. *BP1* and *BP2* measure the county-year level of bank penetration. The third conditional variable is *AA*, which is a dummy variable that equals 1 if the county-year-level proportion of African American population is in the highest quantile, and 0 if it is in the lowest quantile. The fourth conditional variable is *HL*, a dummy variable that equals 1 if the county-level proportion of the Hispanic/Latino population is in the highest quantile, and 0 if it is in the lowest quantile. All other variables in equation (2) are the same as those in equation (1).

5.1. Cross-sectional Analyses with Bank Penetration as Conditional Variables

Prior studies document that P2P lending tends to penetrate areas that have poor bank branch coverages and highly concentrated banking markets (Jagtiani and Lemieux 2017; Hodula 2022). If digital inclusion influences P2P penetration through its effect on P2P lenders' ability to utilize the ubiquitous online information that potentially works as well as standard credit bureau scores (e.g., Berg et al. 2020; Yu and Zhang 2021), then I expect that the incremental impact of digital inclusion on P2P penetration to be larger for areas where individuals are more likely to be excluded by traditional banks due to their lack of quality credit bureau scores. Hence, using *BP1* and *BP2* as the conditional variable, I expect α_1 to be negative. I report the results in Table 8 using both conditional variables. Consistent with my expectations, I find that the coefficients of the interaction terms are negative and significant in all six columns, indicating that the positive association between digital inclusion and P2P penetration is significantly more pronounced when the extent of local bank penetration is lower. These findings are consistent with digital inclusion being more influential in facilitating P2P penetration in areas where people are more likely to be excluded by traditional financial institutions.

[Insert Table 8 Here]

5.2. Cross-sectional Analyses with Minority Proportion as Conditional Variables

Next, I explore whether the relationship between digital inclusion and P2P penetration is significantly stronger in county-years with higher proportions of the minority population. A large existing literature explores the relationship between race and financial exclusion. For example, Black et al. (1978) document that minority borrowers are rejected more often than non-minority borrowers. Schafer and Ladd (1981) find that Black mortgage applicants had higher frequency of loan denial than Whites. Incorporating a thorough list of control variables, Munnell et al. (1996) find that the probability of loan denial for a minority applicant is 8.3 percentage points higher than that for a non-minority applicant. Despite policymakers' efforts in promoting fair lending through various regulations, recent studies continue to find evidence of lending discrimination in traditional financial institutions (e.g., Ghent et al. 2014; Reid et al. 2017; Bayer et al. 2018; Ambrose et al. 2021; Fairlie et al. 2021). Consistent with the statistical discrimination theory, when the observable credit signals for minorities are noisier than for non-minorities, lenders put less weight on the observable signals for minorities and may find it less costly to use group characteristics, such as race, to proxy for creditworthiness (Carr and Megbolugbe 1993; Ladd 1998). In this sense, the use of non-traditional data (such as phone bills, shopping histories, subscriptions, or browsing histories) and machine learning in consumer credit can reduce racial discrimination against minorities (Philippon 2019). Therefore, I expect that digital inclusion, through its influence on local online activities which can aid P2P lenders' use of non-traditional information, should play a larger role in facilitating P2P penetration in counties with a larger proportion of the minority population. Hence, using *AA* and *HL* as the conditional variable, I expect α_1 to be positive.

I report the results in Table 9 using both conditional variables. In columns (1) and (4), I document that the positive relation between digital inclusion and P2P penetration is significantly more pronounced in county-years with a higher proportion of the minority population. In addition, in columns (2), (3), (5), and (6), I find that relation between digital inclusion and P2P penetration is only highly statistically significant in county-years with the highest quantile of minority populations proportion. Overall, the results reported in Table 9 corroborate my prediction that digital inclusion plays a crucial role in financial inclusion, particularly in regions with more vulnerable populations.

[Insert Table 9 Here]

6. Consequences Tests

Finally, I explore further the consequences of the influence of digital inclusion on P2P penetration. In specific, I predict that, in areas with higher digital inclusion: 1) higher risk listings are less likely to be denied, and 2) P2P lenders incorporate more alternative information beyond credit bureau scores. The test results corroborate my predictions. In Table 10 Column (1), as a baseline analysis, I report a positive relationship between loan interest rate and loan denial rate. In Column (2), I interact the loan interest rates with a dummy variable *HI*, which equals 1 if the county-year has an above median *DI*, and 0 otherwise. I find a significantly negative moderating effect of *HI*. In columns (3) and (4), I find that the relation between loan interest rate and denial rate is only significant in county-years with poorer digital inclusion. Finally, I find that credit bureau score significantly predicts loan interest rates (Column (5)), however, such prediction power is attenuated in areas with higher digital inclusion (Column (6)). Overall, the consequences tests support the reasoning that when digital inclusion is higher, non-traditional information may play a larger role in credit assessments, making higher-risk borrowing easier.

[Insert Table 10 Here]

7. Conclusion

Financial inclusion, in general, refers to the provision of various types of financial services, such as deposit or transaction accounts, access to credit, and mobile payment, to the excluded population. In this paper, I study the influence of digital inclusion on financial inclusion, focusing on one important dimension of financial inclusion, namely the access to credit facilitated by P2P lenders. P2P lenders, unlike traditional banks, match potential lenders and borrowers via a platform and pre-screen borrowers with machine learning and big data techniques, utilizing ubiquitous non-traditional online data (Philippon 2016; Fuster et al. 2019, 2022). In this sense, I expect that digital inclusion, which can increase individuals' use of the internet and promote online activities such as online shopping and social media usage, can positively influence the extent of P2P lending penetration.

Using a sample of P2P listings data from 2009 to 2020 from Prosper Marketplace and digital inclusion proxies based on the Federal Communications Commission (FCC) Form 477 data, I find that digital inclusion is positively associated with P2P lending penetration. The main results are robust to the use of four alternative measures of P2P lending penetration, instrumental variable estimation to mitigate endogeneity concerns, an alternative measure of digital inclusion, weighted least-square regressions, more control variables of local banking environments, and year-by-year regression analysis. Moreover, I find that the relationship between digital inclusion and P2P lending penetration is more pronounced in areas with less bank penetration and greater minority population proportion. In consequences tests, I document that higher-risk borrowing is less likely to be denied in county-years with higher digital inclusion. Overall, this study provides evidence that digital inclusion plays a crucial role in P2P lending inclusion.

This study contributes to the stream of literature on the relationship between digital inclusion and financial inclusion. The existing empirical evidence in this literature is scarce; an exception is Zhong and Jiang (2021) and Corrado and Corrado (2015), who find a positive relation between P2P penetration and internet development in China and Europe, respectively. This study differs from these studies in that it focuses on the U.S. market, which is known to be prone to racial discrimination, allowing cross-sectional tests using minority proportions as conditional variables. This study is also different in that through the cross-sectional tests, it reinforces the channel that digital inclusion can promote individuals' use of the internet and dissemination of non-traditional information, which can be effectively utilized by P2P lenders when evaluating the creditworthiness of the previously excluded populations. This paper also suggests that digital inclusion can be another plausible determinant of the growth of P2P lending. Finally, I emphasize the role of P2P lenders in mitigating information frictions in the credit market, facilitated by digital inclusion.

References

Ambrose, B. W., J. N. Conklin, and L. A. Lopez. 2021. Does borrower and broker race affect the cost of mortgage credit? P. Strahan, ed. *The Review of Financial Studies* 34 (2): 790–826.

Balyuk, T., A. N. Berger, and J. Hackney. 2020. What is fueling FinTech lending? The role of banking market structure. *SSRN Electronic Journal*.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace. 2022. Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics* 143 (1): 30–56.

Bayer, P., F. Ferreira, and S. L. Ross. 2018. What drives racial and ethnic differences in high-cost mortgages? The role of high-risk lenders. *The Review of Financial Studies* 31 (1): 175–205.

Beck, T., A. Demirgüç-Kunt, and R. Levine. 2007. Finance, inequality and the poor. *Journal of Economic Growth* 12 (1): 27–49.

Berg, T., V. Burg, A. Gombović, and M. Puri. 2020. On the rise of FinTechs: Credit scoring using digital footprints. A. Karolyi, ed. *The Review of Financial Studies* 33 (7): 2845–2897.

Black, H., R. L. Schweitzer, and L. Mandell. 1978. Discrimination in mortgage lending. *The American Economic Review* 68 (2): 186–191.

Bruhn, M., and I. Love. 2014. The real impact of improved access to finance: Evidence from Mexico: Impact of access to finance on poverty. *The Journal of Finance* 69 (3): 1347–1376.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru. 2018. FinTech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics* 130 (3): 453–483.

Butler, A. W., E. J. Mayer, and J. P. Weston. 2022. Racial disparities in the auto loan market. *The Review of Financial Studies* 36 (1): 1-41.

Carr, J. H., and I. F. Megbolugbe. 1993. The Federal Reserve bank of Boston study on mortgage lending revisited. *Journal of Housing Research* 4 (2): 277–313.

Corrado, G., and L. Corrado. 2015. The geography of financial inclusion across Europe during the global crisis. *Journal of Economic Geography* 15 (5): 1055–1083.

De Roure, C., L. Pelizzon, and A. Thakor. 2022. P2P Lenders versus banks: Cream skimming or bottom fishing? A. Ellul, ed. *The Review of Corporate Finance Studies* 11 (2): 213–262.

Demirgüç-Kunt, A., L. Klapper, and D. Singer. 2017. Financial inclusion and inclusive growth: A review of recent empirical evidence. *World Bank Policy Research Working Paper* 8040.

Demirgüç-Kunt, A., L. Klapper, D. Singer, and S. Ansar. 2022. *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. World Bank Publications.

Dittmar, A., J. Mahrt-Smith, and H. Servaes. 2003, International corporate governance and corporate cash holdings. *Journal of Financial and Quantitative Analysis* 38 (1): 111-133.

Fairlie, R., A. Robb, and D. T. Robinson. 2021. Black and white: Access to capital among minority-owned start-ups. *Management Science* 68 (4): 2377–3174.

Fernández-Olit, B., J. M. Martín Martín, and E. Porras González. 2019. Systematized literature review on financial inclusion and exclusion in developed countries. *International Journal of Bank Marketing* 38 (3): 600–626.

Frost, J., L. Gambacorta, Y. Huang, H. S. Shin, and P. Zbinden. 2019. BigTech and the changing structure of financial intermediation. *Economic Policy* 34 (100): 761–799.

Fuster, A., M. Plosser, P. Schnabl, and J. Vickery. 2019. The role of technology in mortgage lending. *The Review of Financial Studies* 32 (5): 1854–1899.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2022. Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance* 77 (1): 5–47.

Gant, J., N. Turner-Lee, Y. Li, and J. Miller. 2010. *National Minority Broadband Adoption: Comparative Trends in Adoption, Acceptance and Use.* Washington, DC: Joint Center for Political and Economic Studies.

Ghent, A. C., R. Hernández-Murillo, and M. T. Owyang. 2014. Differences in subprime loan pricing across races and neighborhoods. *Regional Science and Urban Economics* 48: 199–215.

Han, R., and M. Melecky. 2013. *Financial inclusion for financial stability*. Policy Research Working Papers. The World Bank.

Hannig, A., and S. Jansen. 2010. Financial inclusion and financial stability: Current policy issues. *SSRN Electronic Journal*.

Hau, H., Y. Huang, H. Shan, and Z. Sheng. 2019. How FinTech enters China's credit market. *AEA Papers and Proceedings* 109: 60–64.

Hodula, M. 2022. Does Fintech credit substitute for traditional credit? Evidence from 78 countries. *Finance Research Letters* 46: 102469.

Jaeger, P. T., J. C. Bertot, K. M. Thompson, S. M. Katz, and E. J. DeCoster. 2012. The intersection of public policy and public access: Digital divides, digital literacy, digital inclusion, and public libraries. *Public Library Quarterly*, 31 (1), 1-20.

Jagtiani, J., and C. Lemieux. 2017. Fintech lending: Financial inclusion, risk pricing, and alternative information. *SSRN Electronic Journal*.

Kempson, E., and C. Whyley. 1999. *Kept out or opted out? Understanding and combating financial exclusion*. Bristol, UK: The Policy Press.

Khan, H. R. 2012. Financial inclusion and financial stability: Are they two sides of the same coin? *Address by Shri HR Khan, Deputy Governor of the Reserve Bank of India, at BANCON, 1-12.*

Khan, U., and N. B. Ozel. 2016. Real activity forecasts using loan portfolio information. *Journal of Accounting Research* 54 (3): 895–937.

Ladd, H. F. 1998. Evidence on discrimination in mortgage lending. *Journal of Economic Perspectives* 12 (2): 41–62.

Lapukeni, A. F. 2015. Financial inclusion and the impact of ICT: An overview. *American Journal of Economics* 5 (5): 495-500.

Leyshon, A., and N. Thrift. 1995. Geographies of financial exclusion: Financial abandonment in Britain and the United States. *Transactions of the Institute of British Geographers* 20 (3): 312.

Munnell, A. H., G. M. B. Tootell, L. E. Browne, and J. McEneaney. 1996. Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review* 86 (1): 25–53.

Ozili, P. K. 2018. Impact of digital finance on financial inclusion and stability. *Borsa Istanbul Review* 18 (4): 329–340.

Philippon, T. 2016. *The FinTech opportunity*. Cambridge, MA: National Bureau of Economic Research.

Philippon, T. 2019. *On Fintech and financial inclusion*. Cambridge, MA: National Bureau of Economic Research.

Reid, C. K., D. Bocian, W. Li, and R. G. Quercia. 2017. Revisiting the subprime crisis: The dual mortgage market and mortgage defaults by race and ethnicity. *Journal of Urban Affairs* 39 (4): 469–487.

Reisdorf, B., and C. Rhinesmith. 2020. Digital inclusion as a core component of social inclusion. Social inclusion, 8 (2), 132-137.

Rhinesmith, C. 2016. *Digital Inclusion and Meaningful Broadband Adoption Initiatives*. Benton Institute for Broadband & Society.

Roberts, M. R., and T. M. Whited. 2012. Endogeneity in empirical corporate finance. *SSRN Electronic Journal*.

Rodriguez, J. A., C. Shachar, and D. W. Bates. 2022. Digital inclusion as health care—supporting health care equity with digital-infrastructure initiatives. *New England Journal of Medicine*, 386 (12), 1101-1103.

Sanders, C. K., and E. Scanlon. 2021. The digital divide is a human rights issue: Advancing social inclusion through social work advocacy. *Journal of Human Rights and Social Work*, 6, 130-143.

Schafer, R., and H. F. Ladd. 1981. *Discrimination in mortgage loans*. Cambridge, MA: MIT Press.

Stock, J. and M. Yogo. 2005. Testing for weak instruments in linear IV regression. In *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, 80-108. Cambridge: Cambridge University Press.

Tang, H. 2019. Peer-to-peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies* 32 (5): 1900–1938.

Villasenor, J. D., D. M. West, and R. J. Lewis. 2015. *The 2015 Brookings Financial and Digital Inclusion Project Report*. Washington, D.C: The Brookings Institution.

Wallace, C., K. Vincent, C. Luguzan, L. Townsend, and D. Beel. 2017. Information technology and social cohesion: A tale of two villages. *Journal of Rural Studies*, 54: 426-434.

Wu, Y., and T. Zhang. 2021. Can credit ratings predict defaults in peer-to-peer online lending? Evidence from a Chinese platform. *Finance Research Letters* 40: 101724.

Yu, L., and X. Zhang. 2021. Can small sample dataset be used for efficient internet loan credit risk assessment? Evidence from online peer to peer lending. *Finance Research Letters* 38: 101521.

Zhong, W., and T. Jiang. 2021. Can internet finance alleviate the exclusiveness of traditional finance? Evidence from Chinese P2P lending markets. *Finance Research Letters* 40: 101731.

Appendix: Variable Definitions

P2PPEN	=	The main measure of P2P penetration, which is calculated as the county-year-level successfully funded Prosper loans amount per capita.
P2PPEN_CT	=	The second measure of P2P penetration, which is calculated as the county-year-level number of successfully funded Prosper loans per capita, scaled by 10,000.
P2PPEN_RK	=	The third measure of P2P penetration, which is calculated as the rank of a county-year's *P2PPEN* across all county-years. This measure minimizes the bias of the prevalence of extreme values of *P2PPEN*.
P2PPEN_CS	=	The fourth measure of P2P penetration, which is calculated as the county-year-level successfully funded Prosper loans amount divided by the county-year-level commercial bank consumer loans amount.
P2PPEN_TL	=	The fifth measure of P2P penetration, which is calculated as the county-year-level successfully funded Prosper loans amount divided by the county-year-level commercial bank total loans amount.
DI	=	The main proxy for digital inclusion, which measures the county-year-level residential fixed connections of at least 10 Mbps (download) / 1 Mbps (upload) (i.e., high-speed internet) per 1,000 households. The tables are based on Federal Communications Commission (FCC) Form 477 data. Data Link: https://www.fcc.gov/form-477-census-tract-data-internet-access-services
T1_INT	=	A measure of the level of basic internet connections (i.e., a measure of residential fixed internet connections of more than 200 kbps per 1,000 households). Data is from FCC Form 477. Because the basic connections are prevalent across the U.S., I do not use this directly as the measure for digital inclusion. Instead, this measure is used in a principal component analysis (PCA).
HI_SUB	=	The number of county-year households with high-speed internet subscriptions scaled by county-year population. The data from 2017 to 2020 is queried from the U.S. Census 5-Year American Community Survey (ACS) Application Programming Interface (API). The data from 2013 to 2016 is queried from the U.S. Census 1-Year ACS API. However, the 1-Year ACS only covers census blocks with populations of 65,000 or more, which is why I only use this measure in an auxiliary PCA.
DI_ALT	=	The alternative measure of digital inclusion, which is the first principal component of a factor analysis on three variables, namely *DI*, *T1_INT*, and *HI_SUB*. These three variables used in factor analysis are described in detail in this appendix.
DI_HIGH = 1	=	A dummy variable measure of digital inclusion, which equals 1 if *DI* is in the highest quintile across all county-years, and equals 0 if *DI* is in the lowest quintile across all county-years.
DI_HIGH_YR = 1	=	A dummy variable measure of digital inclusion within a single year, which equals 1 if *DI* is in the highest quintile across all counties in the same year, and equals 0 if *DI* is in the lowest quintile across all counties in the same year. This measure is **ONLY** used in the year-by-year regressions in Table 8.
AA = 1	=	A dummy variable measure of county-level Black/African Americans (AA) population, which equals 1 if the county-level proportion of Black/AA population is in the highest quantile, and 0 if it is in the lowest quantile.
HL = 1	=	A dummy variable measure of county-level Hispanic/Latino population, which equals 1 if the county-level proportion of Hispanic/Latino population is in the highest quantile, and 0 if it is in the lowest quantile.

BP1 = 1	=	A dummy variable measure of county-level commercial bank loans penetration, which equals 1 if the county-level commercial bank total loans per capita is in the highest quantile, and 0 if it is in the lowest quantile.
BP2 = 1		A dummy variable measure of county-level commercial bank loans penetration, which equals 1 if the county-level commercial bank consumer loans per capita is in the highest quantile, and 0 if it is in the lowest quantile.
RATE	=	The county-year mean of Prosper loan borrowers' interest rates. This is used as a proxy for county-year level of P2P borrowers' credit risk.
SCORE	=	The county-year mean of Prosper loan borrowers' credit bureau score.
TERM	=	The county-year mean of Prosper loan terms.
INRANGE	=	The county-year mean of Prosper loan borrowers' self-disclosed income range.
LENEMP	=	The county-year mean of Prosper loan borrowers' self-disclosed months of employment.
PRIOR	=	The county-year proportion of Prosper loan borrowers who had previously borrowed from the platform.
INCOME	=	The county-year-level personal income per capita in thousands. *Source:* U.S. Bureau of Economic Analysis (BEA).
EDU	=	The county-year-level proportion of adult population receiving post-secondary education. *Source*: U.S. Census 5-Year ACS
RACE	=	The county-year-level proportion of population that is not White. *Source*: U.S. Census 5-Year ACS
GENDER	=	The county-year-level proportion of population that is male. *Source*: U.S. Census 5-Year ACS
MARRY	=	The county-year-level proportion of adult population that is currently married. *Source*: U.S. Census 5-Year ACS
TLPC	=	The county-year-level total commercial bank loans per capita (Call Report).
TAPC	=	The county-year-level total commercial bank assets per capita (Call Report).
CSLPC	=	The county-year-level total commercial bank consumer loans per capita (Call Report).
LLP	=	The county-year-level commercial bank loan loss provisions scaled by one-year lagged total loans (Call Report).
LLA	=	The county-year-level commercial bank loan loss allowances scaled by one-year lagged total loans (Call Report).
NCO	=	The county-year-level commercial bank net loan charge-offs scaled by one-year lagged total loans. Net charge-offs equal to charge-offs minus recoveries (Call Report).
NPL	=	The county-year-level commercial bank non-performing loans scaled by one-year lagged total loans (Call Report).
T1R	=	The county-year-level commercial bank Tier 1 capital ratio (Call Report).
UNEMP	=	The county-year-level unemployment rate. *Source:* U.S. Bureau of Labor Statistics.
HPI	=	The county-year-level housing price index. *Source:* Federal Housing Finance Agency.
LNGDP	=	The natural logarithm of county-year-level GDP per capita. *Source:* U.S. Bureau of Economic Analysis.

Table 1: Descriptive Statistics and Correlations

Panel A: Descriptive Statistics for Main Variables

Variable	N	Mean	p25	p50	p75	SD
P2PPEN	28,027	4.6092	0.7298	3.3655	7.0069	4.6380
P2PPEN_CT	28,027	3.8116	1.1495	2.9622	5.5972	3.2663
P2PPEN_CS[64]	20,670	110.5058	1.6055	7.6126	33.1046	430.7103
P2PPEN_TL	20,714	1.2690	0.0696	0.3080	1.0600	2.8581
DI	22,698	2.3399	1.0000	2.0000	3.1250	1.1357
DI_ALT	10,564	0.0000	-1.0809	0.0461	1.1668	1.5653
T1_INT	25,441	3.5809	3.0000	3.6667	4.1000	0.8099
HI_SUB	13,938	0.5329	0.4247	0.5393	0.6449	0.1496
RATE	28,030	0.1753	0.1404	0.1598	0.2000	0.0549
SCORE	28,029	717.4946	705.1539	716.5000	729.0000	25.1828
TERM	28,030	42.3300	36.0000	42.4533	45.0000	5.8611
INRANGE	28,030	4.0068	3.6667	4.0000	4.3750	0.6874
LENEMP	28,017	116.1023	82.5556	110.5000	139.0000	63.5752
PRIOR	28,030	0.2190	0.0000	0.1667	0.3333	0.2440
INCOME	27,496	40928	33500	38601	45377	12030
EDU	28,027	0.1281	0.0897	0.1179	0.1568	0.0527
RACE	28,027	0.1765	0.0555	0.1165	0.2500	0.1628
GENDER	28,027	0.4988	0.4879	0.4949	0.5035	0.0215
MARRY	28,024	0.2783	0.2306	0.2654	0.3135	0.0678
UNEMP	25,454	6.3383	4.2000	5.7000	8.0000	2.8580
HPI	23,774	142.9802	124.0400	137.9850	156.3700	28.7803
LNGDP	27,494	10.5511	10.2408	10.5349	10.8129	0.4768
TLPC	20,734	24.0319	3.1870	7.4139	15.5796	283.8860
TAPC	20,734	46.6463	5.5014	12.3523	24.9359	790.1261
CSLPC	20,734	3.1045	0.0907	0.3037	0.7826	51.2018
LLP	17,817	0.0043	0.0007	0.0019	0.0043	0.0089
LLA	17,817	0.0154	0.0111	0.0138	0.0177	0.0079
NCO	17,542	0.0068	0.0003	0.0012	0.0034	0.3272
NPL	17,542	0.0182	0.0000	0.0000	0.0071	1.4383
T1R	20,337	0.1718	0.1282	0.1500	0.1837	0.1901

[64] *P2PPEN_CS* and *P2PPEN_TL* are scaled by 1,000.

Table 1 (Continued)

Panel B: Pearson's (Lower Triangle) and Spearman's (Higher Triangle) Correlations for Selected Main Variables

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
(1)	*P2PPEN*		0.95	0.69	0.77	1.00	**0.51**	**0.29**	-0.47	0.03	0.37	0.26	0.20	0.13	0.33	0.14	0.04	0.04	0.04	-0.56	0.23	0.18
(2)	*P2PPEN_CT*	0.95		0.64	0.72	0.95	**0.45**	**0.25**	-0.38	-0.03	0.28	0.16	0.17	0.13	0.30	0.14	0.02	0.04	0.01	-0.54	0.21	0.18
(3)	*P2PPEN_CS*	0.27	0.26		0.90	0.69	**0.53**	**0.31**	-0.37	0.06	0.35	0.27	0.18	0.16	0.29	0.20	0.11	-0.01	0.15	-0.39	0.15	0.13
(4)	*P2PPEN_TL*	0.46	0.44	0.63		0.77	**0.49**	**0.21**	-0.40	0.04	0.37	0.25	0.21	0.12	0.20	0.08	0.10	0.02	0.10	-0.39	0.13	0.04
(5)	*P2PPEN_RK*	0.91	0.88	0.25	0.42		**0.51**	**0.29**	-0.46	0.03	0.37	0.26	0.20	0.12	0.33	0.14	0.04	0.04	0.04	-0.56	0.23	0.18
(6)	*DI*	**0.41**	**0.38**	**0.28**	**0.31**	**0.49**		0.93	-0.34	0.10	0.28	0.32	0.11	0.30	0.56	0.46	0.09	-0.07	0.26	-0.54	0.37	0.37
(7)	*DI_ALT*	**0.26**	**0.23**	**0.23**	**0.21**	**0.28**	0.92		-0.19	0.10	0.04	0.27	-0.07	0.15	0.62	0.66	-0.02	-0.07	0.17	-0.31	0.32	0.43
(8)	*RATE*	-0.39	-0.34	-0.11	-0.19	-0.53	-0.37	-0.20		-0.42	-0.17	-0.24	-0.17	-0.10	-0.24	-0.05	0.04	-0.09	-0.01	0.39	-0.26	-0.08
(9)	*SCORE*	0.00	-0.04	-0.01	-0.02	0.05	0.09	0.06	-0.45		0.09	0.16	0.08	0.14	0.14	0.06	0.00	0.02	0.02	-0.12	0.17	0.05
(10)	*TERM*	0.19	0.14	0.05	0.10	0.27	0.17	0.03	-0.20	0.11		0.16	0.14	0.06	0.14	0.04	-0.02	-0.02	0.03	-0.21	0.03	0.06
(11)	*INRANGE*	0.20	0.11	0.10	0.12	0.24	0.27	0.23	-0.26	0.15	0.13		0.10	0.18	0.32	0.23	0.13	-0.01	0.12	-0.23	0.26	0.20
(12)	*LENEMP*	0.06	0.05	0.00	0.02	0.11	0.03	-0.08	-0.13	0.09	0.07	0.09		0.08	-0.01	-0.09	0.01	-0.03	0.01	-0.09	-0.04	-0.05
(13)	*PRIOR*	-0.06	-0.05	0.00	-0.03	-0.03	0.12	0.08	-0.06	0.04	-0.03	0.09	0.04		0.24	0.18	0.08	-0.04	0.12	-0.24	0.22	0.12
(14)	*INCOME*	0.26	0.24	0.13	0.09	0.27	0.51	0.53	-0.21	0.10	0.06	0.27	-0.04	0.12		0.70	-0.05	-0.01	0.02	-0.56	0.44	0.66
(15)	*EDU*	0.18	0.16	0.15	0.13	0.14	0.49	0.64	-0.07	0.05	-0.01	0.21	-0.10	0.08	0.70		0.04	-0.12	0.15	-0.35	0.30	0.52
(16)	*RACE*	0.03	0.02	0.10	0.06	0.03	0.02	-0.13	0.02	-0.01	-0.03	0.07	0.02	0.02	-0.05	-0.02		-0.22	0.63	0.15	0.11	0.09
(17)	*GENDER*	0.01	0.01	-0.06	-0.03	0.02	-0.09	-0.12	-0.05	0.01	-0.01	0.01	0.01	-0.02	-0.06	-0.14	-0.09		-0.08	-0.13	0.13	-0.02
(18)	*MARRY*	0.02	0.01	0.12	0.07	0.03	0.23	0.11	-0.03	0.01	-0.02	0.08	-0.01	0.04	0.04	0.17	0.62	0.05		0.05	0.10	0.22
(19)	*UNEMP*	-0.45	-0.44	-0.08	-0.17	-0.56	-0.51	-0.33	0.43	-0.10	-0.18	-0.20	-0.06	-0.09	-0.42	-0.32	0.20	-0.07	0.07		-0.44	-0.34
(20)	*HPI*	0.22	0.20	0.14	0.09	0.24	0.39	0.33	-0.24	0.14	0.03	0.24	-0.04	0.17	0.46	0.31	0.06	0.12	0.12	-0.40		0.30
(21)	*LNGDP*	0.20	0.19	0.08	0.04	0.18	0.34	0.38	-0.10	0.04	0.02	0.17	-0.07	0.04	0.60	0.47	0.05	-0.01	0.18	-0.31	0.34	

This table provides the descriptive statistics (Panel A), Pearson's correlations (Panel B below the diagonal), and Spearman's correlations (Panel B above the diagonal) of the main variables used in this study. The detailed definitions of the variables are provided in the Appendix. All correlations with absolute values greater than 0.02 are statistically significant at the 0.01 level or better (two-tailed)

Table 2: P2P Penetration and Digital Inclusion—Baseline Regressions

	(1) *P2PPEN*	(2) *P2PPEN*	(3) *P2PPEN*	(4) *P2PPEN*
DI	**0.9953***** **(4.15)**	**0.5420***** **(3.43)**	**0.6364***** **(3.96)**	**0.6159***** **(3.61)**
RATE		-0.5095 (-0.72)	0.5466 (1.03)	0.4388 (0.61)
SCORE		-0.0027** (-3.03)	-0.0026** (-2.85)	-0.0040*** (-3.29)
TERM		0.0402*** (3.79)	0.0429** (3.07)	0.0411** (2.91)
INRANGE		0.5212** (2.43)	0.4139* (2.16)	0.3569* (2.06)
LENEMP		-0.0010 (-1.72)	-0.0012* (-2.15)	-0.0010 (-1.61)
PRIOR		-0.1377 (-1.37)	-0.0561 (-0.88)	-0.0127 (-0.17)
INCOME		0.0000*** (3.56)	0.0000* (2.14)	0.0000* (2.22)
EDU		5.7685* (2.04)	4.7307** (2.49)	4.2824** (2.28)
RACE		0.7380 (1.65)	0.8973 (1.72)	1.1072 (1.77)
GENDER		-8.0149** (-3.08)	-6.6329** (-2.68)	-7.2894** (-2.91)
MARRY		-3.4387*** (-3.39)	-2.6335** (-2.58)	-2.7499* (-2.22)
UNEMP			-0.0067 (-0.23)	-0.0096 (-0.30)
HPI			0.0109** (2.80)	0.0122** (3.03)
LNGDP			0.0822 (0.67)	-0.0757 (-0.56)
TLPC				0.0011*** (3.57)
TAPC				-0.0004*** (-3.52)
Constant	2.3817*** (4.25)	4.6886*** (4.18)	1.6143 (0.90)	4.4649** (2.37)
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
SE Cluster	County, Year	County, Year	County, Year	County, Year
R-squared	0.60	0.63	0.69	0.72
N	22698	22248	18921	14693

This table reports the regression of my main proxy for county-level P2P penetration on the proxy for digital inclusion and an array of control variables. All models include county and year fixed effects. Column 1 shows the results without control variables. Column 2 shows the results including P2P loan-related and demographic controls. Column 3 shows the results including P2P loan-related, demographic, and macroeconomic controls. Column 4 shows the results including P2P loan-related, demographic, macroeconomic, and county-level commercial bank-related controls. The dependent variable *P2PPEN* is winsorized at top and bottom 1% levels to eliminate prevalent outliers. The t-statistics reported in parentheses are based on robust standard errors clustered by year and county. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 3: Baseline Regressions using Four Alternative Measures of P2P Penetration

	(1) *P2PPEN_CT*	(2) *P2PPEN_RK*	(3) *P2PPEN_TL*	(4) *P2PPEN_CS*
DI	**0.4181***** **(3.43)**	**0.8521***** **(4.39)**	**0.6261***** **(3.90)**	**71.8006***** **(3.89)**
RATE	1.4526* (2.05)	-2.4140 (-1.44)	0.0675 (0.25)	19.4667 (0.52)
SCORE	-0.0037** (-3.13)	-0.0030 (-0.94)	-0.0035*** (-3.63)	-0.5264*** (-3.64)
TERM	0.0103* (2.15)	0.0886*** (4.29)	0.0164** (2.68)	1.2365 (1.79)
INRANGE	0.0737 (1.51)	0.7008** (2.53)	0.1396* (1.86)	14.0657 (1.77)
LENEMP	-0.0008* (-1.94)	-0.0011 (-1.08)	-0.0006 (-1.46)	-0.0704 (-1.49)
PRIOR	0.0598 (1.20)	0.2356 (1.17)	-0.0332 (-0.56)	-12.9203 (-1.70)
INCOME	0.0000 (1.76)	0.0000** (2.41)	-0.0000 (-1.83)	-0.0010 (-0.81)
EDU	2.4072* (2.10)	5.1534** (2.69)	4.7466* (2.00)	195.6148 (0.70)
RACE	0.7762* (2.08)	1.2214 (1.63)	0.5298 (0.94)	240.9028* (2.04)
GENDER	-5.9861** (-2.87)	-14.9940** (-3.00)	-3.1882 (-1.28)	-1.14e+03** (-3.04)
MARRY	-2.0524** (-2.71)	-2.3174 (-1.68)	-1.3181 (-1.21)	-71.4702 (-0.53)
UNEMP	-0.0218 (-0.97)	-0.0380 (-0.91)	-0.0557 (-1.54)	-8.5515 (-1.18)
HPI	0.0071** (2.30)	0.0124* (1.98)	0.0032 (0.78)	1.5900** (2.29)
LNGDP	0.0605 (0.57)	0.0667 (0.34)	-0.5972*** (-3.26)	-83.8487** (-3.19)
TLPC	0.0010*** (3.70)	0.0015** (2.30)	-0.0038 (-1.29)	-0.3712 (-1.28)
TAPC	-0.0003*** (-3.49)	-0.0005* (-2.09)	0.0013 (1.26)	0.1261 (1.27)
Constant	5.4332*** (3.43)	12.4389*** (3.94)	9.2721*** (3.74)	1510.1816*** (3.76)
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
SE Cluster	County, Year	County, Year	County, Year	County, Year
R-squared	0.71	0.85	0.24	0.18
N	14693	14693	14676	14640

This table reports the regression of four alternative measures of county-level P2P penetration on the proxy for digital inclusion and an array of control variables. All models include county and year fixed effects. All models include a full set of P2P loan-related, demographic, macroeconomic, and county-level commercial bank-related controls. Column 1 shows the results using *P2PPEN_CT* as the measure for P2P penetration. Column 2 shows the results using *P2PPEN_RK* as the measure for P2P penetration. Column 3 shows the results using *P2PPEN_TL* as the measure for P2P penetration. Column 4 shows the results using *P2PPEN_CS* as the measure for P2P penetration. The dependent variable *P2PPEN* is winsorized at top and bottom 1% levels to eliminate prevalent outliers. The t-statistics reported in parentheses are based on robust standard errors clustered by year and county. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 4: Relation between Digital Inclusion and P2P Penetration Tests —Instrumental Variable (2SLS) Estimation

	(1) First Stage	(2) Second Stage	(3) First Stage	(4) Second Stage
Dependent Var	*DI*	*P2PPEN*	*DI*	*P2PPEN_CT*
Pred_DI		**0.6161***** **(8.27)**		**0.4912***** **(9.54)**
T1_INT	**0.5573***** **(60.87)**		**0.5573***** **(51.4)**	
RATE	-0.1544 (-1.12)	0.3815 (0.61)	-0.1973 (-1.36)	1.4625*** (4.06)
SCORE	-0.0011*** (-4.63)	-0.0039*** (-3.63)	-0.0010*** (-3.87)	-0.0036*** (-5.41)
TERM	0.0001 (0.14)	0.0407*** (8.92)	-9.3E-05 (-0.09)	0.0101*** (3.18)
INRANGE	0.0348*** (4.52)	0.3463*** (9.87)	0.0492*** (6.03)	0.0695*** (3.08)
LENEMP	-0.0003*** (-3.48)	-0.0009** (-2.54)	-0.0004*** (-4.56)	-0.0008*** (-3.13)
PRIOR	-0.0382* (-1.69)	-0.0158 (-0.15)	-0.0221 (-0.93)	0.0617 (0.95)
INCOME	0.0000 (1.05)	0.0000*** (7.11)	1.03E-05*** (12.19)	0.0000*** (3.65)
EDU			3.8484*** (-20.5)	1.8507*** (2.66)
RACE	0.3556*** (6.54)	3.8167*** (4.33)	0.3084*** (5.47)	0.7650*** (3.88)
GENDER	-2.9233*** (-9.13)	0.9584*** (3.89)	-4.1519*** (-11.81)	-5.7308*** (-4.99)
MARRY	0.8652*** (8.12)	-7.2392*** (-4.91)	1.5355*** (13.08)	-2.1324*** (-5.54)
UNEMP	0.0076** (2.46)	-2.5344*** (-5.17)	-0.0072** (-2.29)	-0.0210** (-2.28)
HPI	0.0022*** (7.95)	-0.0088 (-0.63)	0.0020*** (5.99)	0.0070*** (6.25)
LNGDP	0.0966*** (6.38)	0.0116*** (9.10)	0.0928*** (5.61)	0.049222 (0.85)
TLPC	-0.0001 (-0.64)	-0.0620 (-0.89)	-3.3E-05 (-0.39)	0.00102*** (2.72)
TAPC	0.0000 (0.48)	0.0010** (2.15)	6.67E-06 (0.22)	-0.00034*** (-2.58)
Constant	2.681 (0.12)	-0.0003** (-1.98)	-0.7436** (-2.24)	3.3759*** (3.25)
State and Year FE	Yes	Yes	Yes	Yes
SE Cluster	County, Year	County, Year	County, Year	County, Year
Cragg-Donald F		3705.05		3705.05
R-squared	0.20	0.86	0.20	0.87
N	14,693	14,693	14,693	14,693

This table reports the regression results of the relation between digital inclusion and P2P penetration based on an instrumental variable (2SLS) approach. The instrument is *T1_INT*. Column 1 shows the results of the first stage regression. In Column 2, I report the second-stage results using the predicted value of *DI* from the first stage. Columns 3 and 4 show the 2SLS results using *P2PPEN_CT* as the second stage dependent variable. The t-statistics reported in parentheses are based on robust standard errors clustered by county and state. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 5: Alternative Measure of Digital Inclusion

	(1)	(2)	(3)	(4)	(5)
Dependent Var:	*P2PPEN*	*P2PPEN_CT*	*P2PPEN_RK*	*P2PPEN_TL*	*P2PPEN_CS*
DI_ALT	**0.5936***** **(5.65)**	**0.4403***** **(5.90)**	**0.8997***** **(6.47)**	**0.7008***** **(7.03)**	**86.9603***** **(6.01)**
RATE	1.3865 (0.45)	4.4411 (1.68)	1.9932 (0.38)	-1.2007 (-1.40)	-207.5215 (-0.86)
SCORE	-0.0058 (-1.41)	-0.0076** (-2.92)	-0.0116 (-1.69)	-0.0100** (-3.14)	-1.3554** (-2.59)
TERM	0.0619*** (7.12)	0.0086 (0.78)	0.1057** (3.85)	0.0384*** (7.02)	3.9181** (3.35)
INRANGE	1.6284*** (5.69)	0.4425** (3.87)	2.5419*** (8.64)	0.6400** (3.58)	76.5669* (2.52)
LENEMP	-0.0028* (-2.51)	-0.0020** (-2.59)	-0.0028* (-2.46)	-0.0026* (-2.24)	-0.2682* (-2.07)
PRIOR	0.2603 (0.83)	0.2697 (0.91)	-0.1485 (-0.35)	0.1533 (0.63)	2.7798 (0.07)
INCOME	0.0000 (1.75)	0.0000 (1.18)	0.0000* (2.29)	-0.0000* (-2.33)	-0.0028 (-1.30)
EDU	2.9454 (1.30)	1.1633 (0.72)	-0.1443 (-0.05)	8.3349* (2.15)	317.5817 (0.52)
RACE	2.8659** (2.59)	1.6757* (2.46)	2.3574* (2.17)	1.7062 (1.69)	420.0026 (1.87)
GENDER	-3.6657 (-0.91)	-4.9295 (-1.80)	-13.6145* (-2.50)	-4.6672 (-1.04)	-1.68e+03* (-2.20)
MARRY	-4.4585** (-2.63)	-2.8498* (-2.48)	-2.5841 (-1.36)	-3.7366 (-1.83)	-152.6710 (-0.58)
UNEMP	-0.0675 (-0.90)	-0.0551 (-1.12)	-0.0496 (-0.52)	0.0959 (0.82)	16.5225 (0.64)
HPI	0.0035 (0.57)	0.0017 (0.38)	0.0004 (0.04)	-0.0037 (-0.59)	0.8230 (0.65)
LNGDP	-0.0386 (-0.17)	0.1549 (0.89)	0.2792 (1.03)	-0.7714** (-2.88)	-125.7353* (-2.45)
TLPC	0.0031 (1.52)	0.0025 (1.79)	0.0038 (1.57)	-0.0107 (-1.58)	-1.1579 (-1.68)
TAPC	-0.0009 (-1.72)	-0.0007* (-2.17)	-0.0010 (-1.70)	0.0031 (1.58)	0.3453 (1.70)
Constant	3.2178 (0.74)	8.9409** (2.88)	17.2055* (2.57)	14.2713** (3.15)	2512.3378** (2.65)
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
R-squared	0.62	0.59	0.63	0.31	0.27
N	6651	6651	6651	6649	6626

This table reports the results of robustness tests using the alternative measure of digital inclusion. Columns 1 to 5 show results using an alternative measure of digital inclusion (*DI_ALT*), with *P2PPEN*, *P2PPEN_CT*, *P2PPEN_RK*, *P2PPEN_TL*, and *P2PPEN_CS*, respectively, as dependent variables. All variables are defined in the Appendix. The t-statistics reported in parentheses are based on robust standard errors clustered by year and county. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 6: Additional Robustness Tests

Model Specification	(1) Weighted by 1/Freq. of Year	(2) Weighted by 1/Freq. of County	(3) Additional Bank Controls	(4) Additional Bank Controls
Dependent Var:	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN_CT*
DI	**0.5374****	**0.6462*****	**0.6241*****	**0.4313*****
	(3.14)	**(3.54)**	**(3.77)**	**(3.60)**
RATE	0.2855	0.3305	0.3285	1.9301*
	(0.46)	(0.46)	(0.30)	(1.94)
SCORE	-0.0034**	-0.0030**	-0.0059**	-0.0050**
	(-2.69)	(-2.86)	(-3.08)	(-2.85)
TERM	0.0352**	0.0436**	0.0466**	0.0121**
	(2.72)	(2.84)	(3.10)	(2.36)
INRANGE	0.2528*	0.3797*	0.5550*	0.1149
	(1.94)	(2.16)	(2.20)	(1.56)
LENEMP	-0.0007	-0.0009	-0.0017*	-0.0013**
	(-1.39)	(-1.49)	(-2.10)	(-2.52)
PRIOR	-0.0218	0.0178	0.1190	0.1185
	(-0.31)	(0.23)	(1.26)	(1.44)
INCOME	0.0000**	0.0000**	0.0000	0.0000
	(2.35)	(2.32)	(1.84)	(1.37)
EDU	3.2609*	3.2018*	5.7173**	3.4909**
	(1.88)	(1.85)	(3.06)	(3.02)
RACE	0.7827	0.7072	1.3637*	0.9138*
	(1.43)	(1.24)	(1.90)	(2.14)
GENDER	-6.9503**	-10.0073**	-6.3692**	-5.6175**
	(-2.78)	(-2.96)	(-2.31)	(-2.83)
MARRY	-2.1178*	-1.9602	-3.3139**	-2.3784**
	(-1.87)	(-1.81)	(-2.52)	(-2.88)
UNEMP	0.0054	-0.0305	-0.0197	-0.0284
	(0.24)	(-0.78)	(-0.44)	(-0.92)
HPI	0.0132***	0.0112**	0.0104*	0.0054
	(4.41)	(2.61)	(1.94)	(1.37)
LNGDP	-0.0975	-0.0515	0.0047	0.1346
	(-0.86)	(-0.37)	(0.03)	(1.13)
TLPC	0.0009**	0.0011***	0.0013	0.0013
	(3.04)	(3.90)	(1.23)	(1.50)
TAPC	-0.0003**	-0.0004***	-0.0004	-0.0004*
	(-2.85)	(-3.81)	(-1.75)	(-2.19)
CSLPC			-0.0001	-0.0002
			(-0.04)	(-0.08)
LLP			-17.1431**	-10.2497**
			(-2.84)	(-2.68)
LLA			-10.0958**	-9.5346**
			(-2.32)	(-2.71)
NCO			-0.0626	-0.0357
			(-0.24)	(-0.17)
NPL			0.0074	-0.0018
			(0.16)	(-0.05)
T1R			-0.0738	-0.1357
			(-0.28)	(-0.66)
Constant	4.0279**	4.7102*	4.7145*	6.0555**
	(2.51)	(2.26)	(1.94)	(2.95)
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R-squared	0.75	0.70	0.70	0.69
N	14693	14693	12436	12436

This table reports the results of robustness and sensitivity tests. Columns 1 and 2 show the results of weighted least-squares regressions. Columns 3 and 4 show the results with six additional commercial bank-related controls, including *CSLNS*, *LLP*, *LLP*, *NCO*, *NPL*, and *T1R*. The t-statistics reported in parentheses are based on robust standard errors clustered by year and county. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 7: Year-by-year Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	P2PPEN	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN*
Year	*2019*	*2018*	*2017*	*2016*	*2015*	*2019*	*2018*	*2017*	*2016*	*2015*
DI	**0.7616*****	**0.8862*****	**0.8858*****	**0.5171*****	**1.1919*****					
	(6.23)	**(6.73)**	**(5.75)**	**(4.13)**	**(6.29)**					
DI_HIGH_YR						**1.8838*****	**2.7556*****	**2.2997*****	**0.9446****	**3.0394*****
						(4.54)	**(5.44)**	**(4.40)**	**(2.01)**	**(4.79)**
RATE	-3.1315	-3.0626	-3.6901	-7.0246*	-32.7792***	-2.6898	-5.7260	-1.8117	-5.9915	-20.2142
	(-0.91)	(-0.69)	(-0.73)	(-1.80)	(-2.93)	(-0.49)	(-0.65)	(-0.20)	(-0.67)	(-1.09)
SCORE	-0.0054	0.0086	-0.0029	-0.0040	-0.0456***	-0.0096	0.0047	0.0096	0.0031	-0.0520*
	(-1.00)	(1.23)	(-0.32)	(-0.61)	(-3.00)	(-1.13)	(0.34)	(0.66)	(0.22)	(-1.93)
TERM	0.0819***	0.0664***	0.0791***	0.1270***	0.1771***	0.1197***	0.0997*	0.0720	0.1261***	0.1597**
	(3.74)	(2.69)	(2.61)	(5.28)	(4.29)	(3.07)	(1.93)	(1.37)	(2.85)	(2.08)
INRANGE	1.4786***	2.0735***	2.0023***	1.5462***	1.7819***	1.6976***	2.1358***	2.1863***	1.4360***	1.7884**
	(7.80)	(8.54)	(6.19)	(6.14)	(3.88)	(6.00)	(4.51)	(4.06)	(3.07)	(2.11)
LENEMP	-0.0016	-0.0034	-0.0019	-0.0061***	-0.0055	-0.0026	-0.0048	-0.0061*	-0.0086***	-0.0063
	(-0.93)	(-1.41)	(-1.58)	(-3.80)	(-1.35)	(-0.85)	(-0.90)	(-1.90)	(-2.95)	(-0.90)
PRIOR	0.5403	-0.4693	0.9887	-0.4102	0.9454	-0.1643	-1.1374	1.2138	-0.0759	4.2916
	(1.27)	(-0.81)	(1.22)	(-0.65)	(0.48)	(-0.22)	(-0.97)	(0.87)	(-0.07)	(1.02)
INCOME	0.0000	0.0000	0.0000	0.0000	0.0000*	-0.0000	-0.0000	-0.0000	0.0000	0.0000
	(0.73)	(1.01)	(1.08)	(0.62)	(1.88)	(-0.83)	(-0.08)	(-0.16)	(0.06)	(1.55)
EDU	4.1411	6.8704**	4.7402	8.8677***	10.3287**	5.5808	6.1162	2.1788	13.5955***	5.7759
	(1.59)	(2.34)	(1.45)	(2.67)	(2.44)	(1.34)	(1.29)	(0.43)	(2.62)	(0.99)
RACE	1.5574**	2.1880**	3.1565***	2.2010***	5.0697***	1.4766	2.5639*	4.0190**	2.4800*	7.9536***
	(2.11)	(2.54)	(3.27)	(2.62)	(3.68)	(1.11)	(1.67)	(2.41)	(1.82)	(3.30)
GENDER	1.0062	-2.6608	-9.5152**	-11.8431***	-4.3251	-5.5062	-9.2467	-13.4854*	-7.9088	-14.8235
	(0.25)	(-0.60)	(-2.04)	(-2.73)	(-0.60)	(-0.70)	(-1.03)	(-1.93)	(-0.74)	(-0.89)
MARRY	-3.8963***	-4.4910**	-4.4438**	-4.6796**	-7.8870***	-4.3797	-5.4609*	-6.1916*	-4.4903	-14.0623***
	(-2.60)	(-2.46)	(-2.16)	(-2.41)	(-3.01)	(-1.52)	(-1.66)	(-1.65)	(-1.36)	(-3.17)
UNEMP	-0.1602**	-0.1748**	-0.2474**	-0.1553**	-0.1723**	-0.0866	-0.1987	-0.2458	0.0456	-0.0714
	(-2.20)	(-2.22)	(-2.55)	(-2.49)	(-2.10)	(-0.64)	(-1.19)	(-1.36)	(0.37)	(-0.45)
HPI	0.0062	-0.0030	-0.0106**	-0.0070	-0.0131	0.0100	0.0021	0.0007	0.0051	0.0141
	(1.36)	(-0.63)	(-1.98)	(-1.21)	(-1.44)	(1.30)	(0.25)	(0.08)	(0.55)	(0.95)
LNGDP	0.2286	-0.0175	0.0673	-0.1295	0.2033	0.0711	-0.4373	0.3465	-0.0916	0.3253
	(0.97)	(-0.06)	(0.23)	(-0.49)	(0.52)	(0.15)	(-0.90)	(0.72)	(-0.18)	(0.48)
TLPC	0.0017	0.0042	0.0048	0.0025	0.0039	0.0022*	0.0065	0.0029	0.0002	0.0021
	(1.29)	(1.00)	(0.97)	(0.98)	(0.56)	(1.86)	(1.09)	(0.34)	(0.03)	(0.15)
TAPC	-0.0009	-0.0007	-0.0014	-0.0008	-0.0012	-0.0008	-0.0013	0.0009	0.0022	0.0024
	(-0.85)	(-0.34)	(-1.09)	(-1.12)	(-0.51)	(-0.90)	(-0.52)	(0.29)	(0.75)	(0.47)
CSLPC	-0.0011	-0.0098	-0.0056	-0.0022	-0.0015	-0.0028	-0.0131	-0.0226*	-0.0254*	-0.0371**
	(-0.17)	(-0.82)	(-0.58)	(-0.45)	(-0.12)	(-0.47)	(-1.20)	(-1.76)	(-1.71)	(-2.09)
LLP	-25.9101	-20.5902	54.1211	-20.3573	-45.2881	-31.6526	-28.8715	-22.2483	6.5830	13.4179

	(-1.34)	(-1.07)	(1.62)	(-0.76)	(-1.36)	(-0.91)	(-0.78)	(-0.50)	(0.14)	(0.22)
LLA	-13.9580	5.3075	3.0329	-11.5438	-20.0865	-28.5952*	-7.4634	-3.1803	-7.1016	-22.1762
	(-1.48)	(0.52)	(0.26)	(-1.03)	(-1.24)	(-1.83)	(-0.51)	(-0.20)	(-0.37)	(-0.78)
NCO	4.4284***	-33.0319	-79.2284*	17.8871	0.7575	3.9389	-31.5966	-0.1813	44.8315	-44.1303
	(3.41)	(-1.49)	(-1.95)	(0.74)	(0.03)	(0.54)	(-0.58)	(-0.00)	(0.63)	(-1.05)
NPL	-4.7033***	5.8731	-3.6941	0.0000	0.0000	-4.0878	5.6139	-9.5765	0.0000	0.0000
	(-3.19)	(1.49)	(-0.55)	(.)	(.)	(-0.49)	(0.58)	(-0.87)	(.)	(.)
T1R	0.8133	-0.1089	-0.1719	0.0544	0.6630	1.9236	0.5317	-0.2871	-0.1569	0.1153
	(0.73)	(-0.24)	(-0.51)	(0.06)	(0.44)	(0.67)	(1.31)	(-1.18)	(-0.12)	(0.05)
Constant	-5.9214	-11.1703*	2.6106	6.8872	32.3943**	1.0597	0.3463	-5.7846	-2.1752	39.1605*
	(-1.10)	(-1.80)	(0.34)	(1.19)	(2.55)	(0.12)	(0.03)	(-0.48)	(-0.17)	(1.70)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	County	County	County	County	County	County	County	County	County	County
R-squared	0.42	0.44	0.42	0.42	0.41	0.55	0.56	0.49	0.52	0.50
N	1553	1563	1559	1607	1629	569	540	640	574	599

This table reports the regression of P2P penetration on digital inclusion for each year separately from 2015 to 2019, using both the continuous measure of digital inclusion (*DI*, columns 1 to 5) and a dummy variable measure of digital inclusion (*DI_HIGH_YR*, columns 6 to 10). All models include state and year fixed effects. All variables are defined in the Appendix. The t-statistics reported in parentheses are based on robust standard errors clustered by county. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 8: Cross-sectional Tests I—The Moderating Role of Local Bank Penetration

Proxy for BP	(1) *P2PPEN* *BP1*	(2) *P2PPEN_CT* *BP1*	(3) *P2PPEN_RK* *BP1*	(4) P2PPEN *BP2*	(5) P2PPEN_CT *BP2*	(6) P2PPEN_RK *BP2*
DI	0.6977***	0.4932***	0.9948***	0.6695***	0.4758***	0.8969***
	(4.39)	(4.09)	(4.85)	(3.84)	(3.70)	(4.27)
BP = 1	0.5636**	0.5289***	1.0856**	0.2829	0.3163	0.4220
	(2.58)	(3.26)	(3.16)	(1.15)	(1.70)	(1.17)
BP × DI	**-0.3215*** **	**-0.2618*** **	**-0.6048*** **	**-0.2755** **	**-0.2289** **	**-0.4541*** **
	(-3.42)	**(-3.75)**	**(-4.30)**	**(-2.81)**	**(-3.19)**	**(-3.34)**
RATE	0.4477	1.3686	-2.4831	-0.0264	1.1465	-3.7170
	(0.58)	(1.81)	(-1.66)	(-0.02)	(1.13)	(-1.77)
SCORE	-0.0048***	-0.0044***	-0.0060	-0.0068**	-0.0056**	-0.0084
	(-3.48)	(-3.78)	(-1.63)	(-2.98)	(-2.76)	(-1.71)
TERM	0.0374**	0.0066	0.0812***	0.0319**	0.0044	0.0709***
	(2.63)	(1.08)	(3.48)	(2.57)	(0.76)	(3.70)
INRANGE	0.3894*	0.0788	0.7744**	0.3397*	0.0423	0.6861**
	(2.06)	(1.43)	(2.71)	(2.05)	(0.90)	(2.78)
LENEMP	-0.0008	-0.0006	-0.0003	-0.0005	-0.0004	-0.0002
	(-1.16)	(-1.19)	(-0.46)	(-0.93)	(-0.99)	(-0.20)
PRIOR	-0.0258	0.0408	0.2729	-0.0403	0.0096	0.1945
	(-0.21)	(0.43)	(1.03)	(-0.27)	(0.10)	(0.80)
INCOME	0.0000	0.0000	0.0000	0.0000*	0.0000	0.0000**
	(1.40)	(1.22)	(1.73)	(2.11)	(1.80)	(2.89)
EDU	5.8288*	3.1939*	6.4750**	4.5651*	2.4261	3.5149
	(2.24)	(2.08)	(2.63)	(1.85)	(1.60)	(1.51)
RACE	1.0373	0.7677	1.2344	1.6607	1.0515*	1.3873
	(1.44)	(1.72)	(1.27)	(1.75)	(1.85)	(1.28)
GENDER	-7.1536*	-6.1683*	-16.2371**	-10.4957***	-7.8046**	-21.8518***
	(-1.97)	(-2.00)	(-2.69)	(-3.44)	(-3.15)	(-4.02)
MARRY	-2.7061*	-1.8560*	-2.2617	-2.4296	-1.3975	-0.1982
	(-1.86)	(-2.04)	(-1.24)	(-1.57)	(-1.42)	(-0.11)
UNEMP	-0.0112	-0.0249	-0.0225	-0.0258	-0.0347	-0.0532
	(-0.30)	(-0.91)	(-0.51)	(-0.60)	(-1.18)	(-0.96)
HPI	0.0133**	0.0071*	0.0133*	0.0125**	0.0069**	0.0107
	(2.95)	(2.18)	(1.97)	(2.77)	(2.27)	(1.56)
LNGDP	0.1812	0.2250	0.4145	0.1170	0.1771	0.3684
	(0.90)	(1.48)	(1.48)	(0.64)	(1.28)	(1.34)
TLPC	0.0013***	0.0011***	0.0020**	0.0013***	0.0011**	0.0019**
	(3.88)	(3.51)	(2.54)	(3.46)	(3.14)	(2.37)
TAPC	-0.0004***	-0.0004***	-0.0007**	-0.0004***	-0.0004**	-0.0006*
	(-3.68)	(-3.28)	(-2.36)	(-3.28)	(-2.94)	(-2.18)
Constant	2.2175	4.3734*	11.4117**	-0.0264	1.1465	-3.7170
	(0.88)	(2.01)	(2.75)	(-0.02)	(1.13)	(-1.77)
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	County Year	County Year	County Year	County Year	County Year	County Year
R-squared	0.73	0.72	0.86	0.73	0.72	0.86
N	7295	7295	7295	7315	7315	7315

This table reports the cross-sectional analyses by regressing P2P penetration on the interaction terms between digital inclusion and dummy variables of county-year-level proxy of traditional bank penetration. In columns 1 to 3, bank penetration is proxied by *BP1*. In columns 4 to 6, bank penetration is proxied by *BP2*. All variables as defined in the Appendix. All regressions include state and year fixed effects. Coefficients on the year and state indicator variables are not tabulated for brevity. The t-statistics reported in parentheses are based on robust standard errors clustered by year and county. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 9: Cross-sectional Tests II—The Moderating Role of Local Minority Population Proportion

	(1)	(2)	(3)	(4)	(5)	(6)
	P2PPEN	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN*	*P2PPEN*
Sub-Sample:	Full	*AA* = 1	*AA* = 0	Full	*HL* = 1	*HL* = 0
DI	0.2665*	**0.7274*****	**0.3568****	0.0012	**0.8666*****	**0.1962**
	(1.90)	**(3.33)**	**(2.66)**	(0.01)	**(3.58)**	**(1.52)**
AA × DI	**0.5814*****					
	(4.48)					
HL × DI				**1.1614*****		
				(7.67)		
AA = 1	-0.8898**					
	(-2.78)					
HL = 1				-2.2427***		
				(-6.87)		
RATE	-0.4087	-0.3147	-0.6689	1.3150	4.0202	0.1243
	(-0.49)	(-0.36)	(-0.49)	(1.12)	(1.61)	(0.17)
SCORE	-0.0031**	-0.0052*	-0.0020	-0.0030**	-0.0071**	0.0002
	(-2.30)	(-2.03)	(-1.39)	(-2.68)	(-2.59)	(0.21)
TERM	0.0482**	0.0437**	0.0516**	0.0369**	0.0547**	0.0308*
	(2.92)	(2.36)	(3.05)	(2.93)	(2.71)	(2.16)
INRANGE	0.3262*	0.3532*	0.2883	0.2582*	0.3119	0.2212*
	(2.01)	(1.84)	(1.81)	(1.84)	(1.58)	(2.07)
LENEMP	-0.0010	-0.0003	-0.0014	-0.0005	-0.0009	-0.0003
	(-1.22)	(-0.41)	(-1.34)	(-1.07)	(-0.86)	(-0.58)
PRIOR	-0.0691	-0.1786	-0.0535	-0.0198	-0.0350	-0.0304
	(-0.66)	(-1.44)	(-0.29)	(-0.26)	(-0.18)	(-0.23)
INCOME	0.0000	0.0000*	0.0000	0.0000	-0.0000	0.0000*
	(1.39)	(2.07)	(0.92)	(0.44)	(-0.29)	(2.18)
EDU	4.8926**	7.2344**	1.8145	5.8876*	7.5289**	2.9860
	(2.59)	(2.56)	(0.85)	(2.03)	(2.37)	(0.94)
GENDER	-10.8965**	-9.7851*	-10.0237	-8.3351**	-8.4070**	-7.0063
	(-2.62)	(-2.26)	(-1.07)	(-2.74)	(-2.46)	(-1.83)
MARRY	-0.9163	-0.3316	-1.0465	-1.7024	-4.0520**	0.6379
	(-0.87)	(-0.28)	(-0.53)	(-1.72)	(-2.28)	(0.58)
UNEMP	0.0292	0.0465	0.0314	-0.0179	0.0143	-0.0808
	(0.98)	(1.09)	(0.83)	(-0.57)	(0.40)	(-1.40)
HPI	0.0083*	0.0137*	-0.0027	0.0112**	0.0165***	-0.0074
	(1.84)	(2.16)	(-0.42)	(2.56)	(3.70)	(-1.47)
LNGDP	0.0512	-0.1835	0.2351	-0.0154	0.1471	-0.4574*
	(0.30)	(-0.97)	(1.00)	(-0.08)	(0.48)	(-2.16)
TLPC	0.0028	0.0004	-0.0020	-0.0049	-0.0018	-0.0083
	(1.21)	(0.11)	(-0.27)	(-1.54)	(-0.44)	(-0.80)
TAPC	-0.0009	-0.0002	0.0057	0.0025	0.0012	0.0050
	(-0.86)	(-0.10)	(0.83)	(1.46)	(0.55)	(0.72)
Constant	4.9912**	5.6673	4.1625	5.6527*	4.1444	8.1485**
	(2.26)	(1.64)	(0.81)	(2.12)	(0.98)	(2.59)
State and Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.72	0.75	0.68	0.74	0.78	0.68
N	6436	3553	2883	7033	3718	3310

This table reports the cross-sectional analyses by regressing P2P penetration on the interaction terms between digital inclusion and dummy variables of county-year-level proxy of minority population proportion in columns 1 and 4, and reports subsample regressions results in columns 2, 3, 5, and 6. All variables as defined in the Appendix. All regressions include state and year fixed effects. Coefficients on the year and state indicator variables are not tabulated for brevity. The t-statistics reported in parentheses are based on robust standard errors clustered by year and county. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 10: Consequences Tests

	(1)	(2)	(3)	(4)	(5)	(6)
	DENIAL	*DENIAL*	*DENIAL*	*DENIAL*	*RATE*	*RATE*
Sub Sample	Full	Full	*DI_HIGH*=1	*DI_HIGH*=0	Full	Full
RATE	**0.3051*****	**0.3495*****	0.0780	0.3355***		
	(3.27)	**(3.48)**	(0.55)	(3.01)		
SCORE	0.0001	0.0001	-0.0000	0.0000	**-0.0009*****	**-0.0009*****
	(0.75)	(0.51)	(-0.02)	(0.13)	**(-44.56)**	**(-36.43)**
DI_HIGH		0.0390**				-0.0727***
		(2.38)				(-2.62)
DI_HIGH × RATE		**-0.2125****				
		(-2.38)				
DI_HIGH × SCORE						**0.0001*****
						(2.64)
TERM	0.0006	0.0009	0.0009	0.0010	0.0003***	0.0003***
	(1.08)	(1.49)	(1.23)	(1.22)	(4.80)	(3.42)
INRANGE	0.0154***	0.0163***	-0.0016	0.0201***	-0.0060***	-0.0061***
	(3.01)	(3.04)	(-0.23)	(3.17)	(-8.05)	(-7.81)
LENEMP	0.0001***	0.0002***	0.0001	0.0002***	0.0000	0.0000
	(2.78)	(3.09)	(1.29)	(2.82)	(0.64)	(0.58)
PRIOR	-0.2138***	-0.2139***	-0.1427***	-0.2316***	-0.0250***	-0.0258***
	(-17.59)	(-17.31)	(-8.89)	(-14.81)	(-11.44)	(-11.49)
INCOME	0.0000	0.0000	0.0000	0.0000	-0.0000	-0.0000
	(1.18)	(0.76)	(0.96)	(0.42)	(-1.09)	(-0.80)
EDU	-0.2592***	-0.2341***	-0.1138**	-0.3577***	-0.0121	-0.0153*
	(-4.60)	(-3.87)	(-2.51)	(-2.73)	(-1.52)	(-1.71)
RACE	0.0124	0.0168	0.0089	0.0149	0.0021	0.0031
	(0.66)	(0.84)	(0.55)	(0.36)	(0.73)	(0.97)
GENDER	-0.0273	-0.0721	-0.1297	0.0002	0.0088	0.0097
	(-0.21)	(-0.52)	(-1.18)	(0.00)	(0.46)	(0.45)
MARRY	-0.0763*	-0.0718	-0.0465	-0.1023	-0.0052	-0.0059
	(-1.81)	(-1.57)	(-1.32)	(-1.09)	(-0.79)	(-0.81)
UNEMP	0.0038**	0.0038**	0.0035***	0.0069***	0.0001	0.0000
	(2.55)	(2.46)	(2.67)	(2.90)	(0.33)	(0.13)
HPI	0.0000	0.0000	0.0002***	-0.0000	-0.0000	-0.0000
	(0.36)	(0.45)	(2.95)	(-0.17)	(-0.50)	(-0.53)
LNGDP	0.0057	0.0053	0.0010	0.0102	0.0003	0.0002
	(0.91)	(0.77)	(0.19)	(0.76)	(0.30)	(0.15)
TLPC	-0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000
	(-1.32)	(-1.49)	(-1.62)	(0.19)	(0.20)	(0.30)
TAPC	0.0000	0.0000*	0.0000**	-0.0000	-0.0000	-0.0000
	(1.61)	(1.83)	(2.13)	(-0.00)	(-0.13)	(-0.23)
Constant	0.0584	0.0745	0.2224	0.0826	0.8387***	0.8670***
	(0.39)	(0.48)	(1.17)	(0.36)	(43.23)	(37.65)
State and Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.29	0.29	0.41	0.18	0.69	0.68
N	16402.00	14693.00	7883.00	6809.00	16402.00	14693.00

This table reports the regression results of consequences tests. Column 1 shows the baseline results of regressing county-level loan denial rate on county-level loan interest rate. Column 2 shows the results with the interaction of *DI_HIGH* and loan interest rate. Columns 3 and 4 show the regression results of loan denial rate on loan interest rate in the subsample in which *DI_HIGH* = 1 and the subsample in which *DI_HIGH* = 0, respectively. Column 5 shows the baseline results of regressing county-year average loan interest rate on county-year average borrowers' credit score. Column 6 shows the results with the interaction of *DI_HIGH* and county-level average credit score. All models include state and year fixed effects. The t-statistics reported in parentheses are based on robust standard errors clustered by county. ***, **, and * denote significance at the 1%, 5%, and 10% levels (two-tailed), respectively.