MEASURING THE WELFARE EFFECTS OF ADVERSE SELECTION IN CONSUMER CREDIT MARKETS∗

Anthony A. DeFusco† Huan Tang‡ Constantine Yannelis§

Abstract

Adverse selection is known in theory to lead to inefficiently low credit provision, yet empirical estimates of the resulting welfare losses are scarce. This paper leverages a randomized experiment conducted by a large fintech lender to estimate welfare losses arising from selection in the market for online consumer credit. Building on methods from the insurance literature, we show how exogenous variation in interest rates can be used to estimate borrower demand and lender cost curves and recover implied welfare losses. While adverse selection leads to large equilibrium price distortions, we find only small overall welfare losses, particularly for high-credit-score borrowers.

∗This version: April 2, 2021. This draft is preliminary and results are subject to change. We thank Scott Baker, Effi Benmelech, Dirk Jenter, Gregor Matvos, Filippo Mezzanotti, Holger Mueller, Daniel Paravisini, Amir Sufi, Kairong Xiao and Anthony Zhang and numerous colleagues for helpful conversations, comments, and suggestions. We are grateful for comments from seminar and conference participants at the University of Chicago (Booth), Columbia Business School, Copenhagen Business School, Harvard University (HBS), the Hong Kong University of Science and Technology Business School, the London School of Economics, the Midwest Finance Association, the Moscow Higher School of Economics, Norges Bank, Northwestern University (Kellogg), the Philadelphia Federal Reserve Bank, Stanford Business School, University of Michigan (Ross), and Washington University in St. Louis. Yannelis gratefully acknowledges financial support from the Booth School of Business at the University of Chicago.

†Northwestern University, Kellogg School of Management: anthony.defusco@kellogg.northwestern.edu
‡London School of Economics: huan.ht.tang@gmail.com
§University of Chicago, Booth School of Business and NBER: constantine.yannelis@chicagobooth.edu
I Introduction

Adverse selection is known in theory to have important effects on credit markets, leading to inefficiently-low credit provision and even market unraveling (Akerlof, 1970; Jaffee and Russell, 1976; Stiglitz and Weiss, 1981). While recent empirical work has documented the presence of adverse selection in a number of consumer credit markets, there is less work linking theory and empirics to measure the welfare consequences of this market imperfection.

In this paper, we show how methods from the empirical literature on selection in insurance markets can be adapted to the case of credit markets and use those methods to estimate the welfare losses arising from adverse selection in the market for online “fintech” consumer credit in China. Three features of this market motivate our focus on it. First, fintech lending, which we define to include any lending activity facilitated by stand-alone online platforms, is relatively new.¹ Lenders operating in this market may therefore be at a larger informational disadvantage and suffer from a greater degree of adverse selection than those in other, more established markets. Second, the Chinese fintech market in particular is by far the most important globally. At its peak in 2017, total outstanding loan volume in this market exceeded $210 billion, constituting more than half of all fintech lending worldwide.² Finally, our collaboration with a specific lending platform operating in this market allows us to cleanly identify the presence of adverse selection and transparently estimate its welfare costs in a manner that is often difficult to do in other settings.

The approach we use to estimate the welfare losses from adverse selection builds on key insights from the empirical literature on selection in insurance markets. As emphasized by Einav, Finkelstein and Cullen (2010), the defining characteristic of adversely selected markets is that firms’ marginal costs are increasing in price. In a credit market context, this is because the borrowers who are willing to pay the highest interest rates also have the highest expected default costs. This tight link between willingness to pay and marginal costs leads to inefficiently high equilibrium pricing, as firms, who cannot observe individual consumers’ marginal costs, will instead set a single price that equates the expected return from selling their product to the average cost of the pool of consumers who choose to buy it. The core insight from Einav, Finkelstein and Cullen (2010) is that the welfare loss generated by this inefficient pricing is uniquely determined by the slopes of the consumer demand curve and the firms’ average and marginal cost curves.

¹Our definition of fintech credit follows that of Cornelli et al. (2020) and includes what is often called “market place” or “peer-to-peer” lending but does not include direct lending by large technology firms for whom such activity represents only a small part of the overall business.

²Statistics on the size of the fintech lending market in China are taken from https://www.wdzj.com/zhuanti/2018report. Since 2017, the Chinese market has contracted considerably due to regulatory pressure but remains the largest fintech lending market globally.
Empirical estimates of these curves thus constitute sufficient statistics for the welfare losses generated by adverse selection. Our paper adapts this insight to the context of a credit market and estimates these curves in the market for Chinese fintech consumer loans.

While the approach we use is general and can be applied to any credit market, the key empirical advantage of our setting is that we are able to leverage exogenous variation in interest rates arising from a randomized experiment. The experiment we analyze was conducted by a large and popular online lending platform during the first quarter of 2018. The platform randomly selected approximately 11,000 loan applicants who had been approved for credit during this period to be part of the experiment. These loan applicants were then randomly split into two equally sized groups and offered different interest rates. Borrowers in the control group were offered standard financing terms, while those in the treatment group received a roughly 40 percent reduction in borrowing costs. We use this exogenous variation, along with information on borrower take-up, default, and recovery rates, to estimate how borrower demand and lender costs vary with interest rates.

We find that both demand and costs are sensitive to interest rates. Estimates from our baseline specification indicate that a 10 percentage point increase in the offered interest rate decreases applicant take-up by approximately 4.3 percentage points and increases borrower charge-offs by about 1 percentage point. Together, these results indicate the presence of adverse selection: borrowers who endogenously select into taking up a loan when offered an exogenously higher interest rate have, on average, higher expected charge-off rates.

The welfare consequences of this adverse selection depend on the degree to which it distorts equilibrium pricing. As in the insurance literature, we measure this distortion using a competitive equilibrium benchmark. That is, we assume that, in equilibrium, lenders set the interest rate to equate the expected return on lending to the average cost of the pool of borrowers who endogenously select into the market at the chosen rate. This equilibrium rate is determined by the intersection of the borrowers’ demand curve with the lenders’ average cost curve. The socially efficient price, however, is determined not by the intersection of demand with average cost, but instead by the intersection of demand with the lenders’ marginal cost curve. Both of these quantities can be easily calculated using our empirical estimates of how demand and costs change in response to exogenous variation in interest rates.

Our estimates imply that the competitive equilibrium interest rate in this market is approximately 30 percent, which falls roughly halfway between the average interest rates offered by the platform in the two experimental treatment arms. Consistent with adverse selection, our estimate of the socially efficient price is substantially smaller than this. We estimate that a social planner seeking to maximize total surplus in this market would set an interest rate equal to only about 9 percent. Thus, adverse selection leads to a large equilibrium price distortion.
Despite this apparently large price distortion, our bottom-line estimates of the welfare losses induced by adverse selection are actually quite small. The reason for this is that demand is relatively inelastic. Given our estimated demand curve, borrowers facing the higher equilibrium interest rate are only about 10 percentage points less likely to take-up a loan than if they were to face the much lower efficient interest rate. Thus, the large price distortion leads to only a relatively small quantity distortion and therefore small overall welfare losses. We estimate that the per-applicant deadweight loss from adverse selection is equal to approximately 0.8 percent of the typical loan amount. Given the average loan size in our sample, this equates to a welfare loss of only $50 ($7.20) per loan applicant.

While the overall welfare losses we document are small, there is interesting heterogeneity across the distribution of ex-ante credit risk. To show this, we repeat our main analysis for subsamples split according to the applicants’ assigned credit rating. We find that demand is equally sensitive to interest rates across high- and low-credit-score borrowers, but that the average charge-off rate among those who choose to take up the loan is substantially more responsive to interest rates for borrowers with lower credit scores. These findings imply that the correlation between demand and unobservable lending costs is higher among observably-riskier borrowers and therefore that the welfare losses from adverse selection should also be larger for this group. Consistent with this, our estimates of the welfare losses for low-credit-score borrowers are roughly four times as large as those for high-credit-score borrowers and twice as large as the results from the pooled sample. Nonetheless, at only $100 ($14.40) per loan applicant, these losses are still small in an absolute sense, making it difficult to argue strongly in favor of policies like interest rates subsidies or loan guarantees in this market on the basis of adverse selection alone.

This paper joins a growing empirical literature on information asymmetries in consumer credit markets. Many papers in this literature have similarly exploited exogenous variation in contract terms arising from either pure randomization or natural experiments to document the presence of adverse selection. Recent examples of this include Stroebel (2016) and Gupta and Hansman (2019) on mortgages, Adams, Einav and Levin (2009) and Einav, Jenkins and Levin (2012) on auto loans, Ausubel (1999) and Agarwal, Chomsisengphet and Liu (2010) on credit cards, Hertzberg, Liberman and Paravisini (2018) on maturity choice, Dobbie and Skiba (2013) on payday loans, and Karlan and Zinman (2009) on microloans. Our paper contributes to this literature by illustrating how similar variation can be used not just to document the presence of adverse selection, but also to estimate its effect on both equilibrium pricing and market efficiency. Our finding that average charge-offs are more responsive to interest rates among observably higher-risk borrowers also echoes one of the key results of Agarwal et al. (2018), who use a similar framework to study the pass-through of bank funding costs to borrowers and document that banks’ marginal profits decrease more quickly in response to credit limit increases among
lower-credit-score borrowers.

As discussed above, our paper is also closely related to the recent empirical literature on adverse selection in insurance markets, which draws inspiration from earlier work such as Chiappori and Salanie (2000) and is reviewed extensively by Einav and Finkelstein (2011). In particular, our conceptual framework closely parallels and builds on the work of Einav, Finkelstein and Cullen (2010), who show how exogenous variation in premiums can be used to measure the welfare losses arising from adverse selection in health insurance markets. A key contribution of our paper is to show how similar methods can be used to estimate the welfare cost of selection in consumer credit markets. While our results indicate that the welfare losses of adverse selection are small in the market for Chinese fintech loans, it is possible that these losses are much greater in other consumer loan markets. In applying these methods to this specific market, we hope to illustrate their general applicability and foster future work measuring these efficiency losses in other credit market contexts.

The remainder of this paper is organized as follows. Section II presents the conceptual framework we use to measure the welfare losses arising from adverse selection in consumer credit markets. Section III describes our empirical setting and data. Section IV discusses how we map the theory to the data to estimate welfare losses. Section V presents our main results, and Section VI concludes.

II CONCEPTUAL FRAMEWORK

We begin by presenting a simple model of adverse selection in consumer credit markets. Our setup draws heavily on the framework originally developed by Einav, Finkelstein and Cullen (2010) to study adverse selection in insurance markets. Though stylized, this framework incorporates the most relevant features of adversely selected credit markets and shows how information on interest rates, lender costs, and borrower take-up can be used to estimate the welfare losses arising from adverse selection.

Setup. We model a credit market in which a fixed population of potential borrowers decide whether to accept take-it-or-leave-it loan offers from lenders. For the sake of simplicity, we consider one-period loans with a fixed loan amount equal to $L$. Lenders choose the interest rate $r$ at which to offer these loans and borrowers decide whether to accept a loan at the lender’s chosen rate. The population of potential borrowers is heterogeneous and characterized by the cumulative distribution function $F(X)$, where $X$ is a vector of borrower characteristics.
Demand for Loans. We denote the utility that a potential borrower \(i\) (with characteristics \(X_i\)) derives from accepting a loan by \(u^L(X_i, r)\) and assume that \(u^L(X_i, r)\) is strictly decreasing in \(r\). Similarly the utility of not accepting a loan is denoted by \(u^N(X_i)\).

Given these assumptions, potential borrower \(i\) will choose to take out a loan if and only if \(u^L(X_i, r) > u^N(X_i)\). Let \(\rho(X_i) \equiv \max\{r : u^L(X_i, r) > u^N(X_i)\}\) denote the maximum interest rate at which potential borrower \(i\) is willing to take out a loan. Aggregate loan demand is thus given by

\[
D(r) = \int 1(\rho(X) \geq r) dF(X).
\] (1)

Market Structure, Supply, and Equilibrium. The supply of loans is determined by the outcome of a Bertrand competition between \(N \geq 2\) identical risk-neutral lenders, each of whom choose interest rates independently to maximize profits taking as given the choices of other lenders. In equilibrium all lenders choose the same interest rate, earn zero profits individually, and split aggregate profits evenly.\(^3\) The expected profits of each individual lender \(j\) are therefore given by

\[
\Pi_j = \frac{L}{N} \times \int (r - \delta(X) \theta(X)(1 + r) - c)1(\rho(X) \geq r) dF(X) = 0,\] (2)

where \(\delta(X_i)\) and \(\theta(X_i)\) denote borrower \(i\)’s default probability and expected share of promised payments charged off in default, and \(c\) denotes all other costs of lending that do not vary across borrowers such as the lender’s cost of funds and customer acquisition costs (expressed as a fraction of the loan amount).

Let \(c(X_i) = c + \delta(X_i) \theta(X_i)\) denote the expected (per-dollar) cost of lending to borrower \(i\), which consists of both the fixed costs and expected charge-offs.\(^4\) The average expected cost curve facing lenders in the market is given by

\[
AC(r) = \frac{1}{D(r)} \int c(X)1(\rho(X) \geq r) dF(X) = \mathbb{E}[c(X) | \rho(X) \geq r].\] (3)

As in any selection market, the key feature of this expected cost curve is that it is governed entirely by the characteristics of the individuals who endogenously choose to take up a loan offer at the

\(^3\)We focus on a competitive equilibrium outcome in our analysis for two reasons. First, in the absence of adverse selection, the competitive equilibrium outcome is socially efficient. This provides a useful benchmark for welfare analysis since any measured losses can be attributed entirely to selection rather than other sources of inefficient pricing. Second, perfect competition is also likely to be an accurate benchmark in our empirical application. The market we study features hundreds of lenders who are competing with each other and selling relatively undifferentiated products. Nonetheless, the framework here is easily adapted to other market structures and would simply require the derivation of an alternative equilibrium pricing function.

\(^4\)Here and throughout the paper we use the term “fixed costs” to refer to any costs that do not vary across borrowers but do scale with the aggregate number of loans originated.
posted interest rate. We can similarly denote the marginal expected cost curve by

\[ MC(r) = \mathbb{E} [c(X) | \rho(X) = r]. \tag{4} \]

We say that the market is adversely selected when the marginal and therefore average cost curve is increasing in the interest rate (decreasing in quantity). That is, when borrowers who choose to take up a loan at higher posted prices also have higher expected default costs.

Rearranging terms in equation (2) yields the following expression for equilibrium interest rates:

\[ r = \frac{\int c(X)\mathbb{1}(\rho(X) \geq r)dF(X)}{\int (1 - \delta(X)\theta(X))\mathbb{1}(\rho(X) \geq r)dF(X)} = \frac{AC(r)}{(1 + c) - AC(r)} \equiv \overline{AC}(r). \tag{5} \]

Throughout, we will refer to the function \( \overline{AC}(\cdot) \) as the scaled average cost curve.\footnote{Strictly speaking, to guarantee the existence and uniqueness of equilibrium we must impose two additional simplifying assumptions on the demand and marginal cost curves. First, we assume that there exists some interest rate \( \bar{r} \) such that \( D(r) > 0 \) and \( \frac{MC(r)}{(1 - \delta(X)\theta(X))\mathbb{1}(\rho(X) = r)dF(X)} = \frac{MC(r)}{(1 + c) - MC(r)} < r \) for all \( r > \bar{r} \). That is, we assume it is profitable to lend to those borrowers with the highest willingness to pay. Second, we assume that if there exists \( r \) such that \( \frac{MC(r)}{(1 + c) - MC(r)} > r \) then \( \frac{MC(r)}{(1 + c) - MC(r)} > r \) for all \( r > r \), which guarantees uniqueness.}

Because the sign of the slope of the scaled average cost curve is inherited from the average cost curve, it is equivalent to say that the market is adversely selected when scaled average costs are increasing in interest rates (decreasing in quantity).

**Welfare and Efficiency.** We use a money-metric notion of utility to measure consumer surplus. The money-metric value of a given allocation is the minimum monetary compensation that would be required for a consumer to attain the same level of utility as consuming that allocation directly.

Let \( m^L(X_i) \) be the money-metric value of being allocated a loan for borrower \( i \) and \( m^N(X_i) \) the money-metric value of being allocated no loan. We assume that borrowers are risk neutral, so that the willingness to pay for the loan is given by \( (1 - \delta(X_i)\theta(X_i))\rho(X_i)L = m^L(X_i) - m^N(X_i) \). Total consumer surplus is thus given by

\[ CS = \int \left[ (m^L(X) - (1 - \delta(X)\theta(X))rL)\mathbb{1}(\rho(X) \geq r) + m^N(X)\mathbb{1}(\rho(X) < r) \right]dF(X). \tag{6} \]

\footnote{One key distinction between insurance markets and credit markets is that, due to borrower default, lenders do not always actually receive their quoted price. Because of this, the break-even quoted interest rate will be a mark-up over average cost. This mark-up is reflected in the denominator of the \( \overline{AC}(\cdot) \) curve and is chosen to equate expected revenue with expected costs.}
Similarly, producer surplus is equal to aggregate profits

\[ PS = L \times \int (r - \delta(X)\theta(X)(1 + r) - c) \mathbb{1}(\rho(X) \geq r) dF(X). \]  

(7)

Total welfare is simply the sum of consumer and producer surplus

\[ TS = CS + PS = \int \left[ (m^L(X) - c(X)L) \mathbb{1}(\rho(X) \geq r) + m^N(X) \mathbb{1}(\rho(X) < r) \right] dF(X). \]  

(8)

From equation (8) it is straightforward to see that it is socially efficient for potential borrower \( i \) to take out a loan if and only if \( m^L(X_i) - m^N(X_i) \geq c(X_i)L \). That is, it is socially efficient for potential borrower \( i \) to take out a loan if and only if her willingness to pay for the loan is at least as great as the social cost of providing it to her.

In contexts where the quoted interest rate is the only instrument available to affect the equilibrium allocation of loans, it may not be possible to achieve this first best outcome since there may be multiple potential borrowers who are willing to accept the same quoted interest rate but who have different expected costs. Therefore, we will work with a constrained efficient allocation as our benchmark. Specifically, we will call an allocation “efficient” if it maximizes total surplus subject to the constraint that interest rates are the only tool available for screening.

More formally, this constrained efficient outcome can be achieved by maximizing total surplus with respect to quantity, noting that quantity can only be altered through the choice of the interest rate. Taking the partial derivative of (8) with respect to quantity yields the following expression:

\[ \frac{\partial TS(r)}{\partial D(r)} = \frac{\partial TS(r) / \partial r}{D'(r)} = - \frac{1}{D'(r)} \int (m^L(X) - m^N(X) - c(X)L) \mathbb{1}(\rho(X) = r) dF(X) 
\]

\[ = \mathbb{E}[m^L(X) - m^N(X) | \rho(X) = r] - \mathbb{E}[c(X)L | \rho(X) = r]. \]  

(9)

Under this notion of efficiency, allocating a loan to potential borrower \( i \) increases total surplus if and only if

\[ \mathbb{E}[m^L(X) - m^N(X) | \rho(X) = \rho(X_i)] \geq \mathbb{E}[c(X)L | \rho(X) = \rho(X_i)]. \]  

(10)

In words, it is socially optimal for potential borrower \( i \) to receive a loan if and only if her expected willingness to pay is at least as great as the expected social cost of allocating a loan to all potential borrowers willing to accept the same quoted rate.

Rearranging terms in equation (10) allows us to express this efficiency condition in terms of
the quoted interest rate and marginal cost curves directly. That is, equation (10) is equivalent to

$$\rho(X_i) \geq \frac{\mathbb{E}[c(X) | \rho(X) = \rho(X_i)]}{\mathbb{E}[1 - \delta(X)\theta(X) | \rho(X) = \rho(X_i)]} = \frac{MC(\rho(X_i))}{(1 + c) - MC(\rho(X_i))} \equiv \overline{MC}(\rho(X_i)).$$

(11)

As with average costs, we will refer to the function $\overline{MC}(\cdot)$ as the scaled marginal cost curve. This condition states that it is socially optimal for potential borrower $i$ to receive a loan if and only if the maximum quoted rate she is willing to accept is at least as great as the scaled marginal cost of providing her the loan.

**Graphical Representation.** Figure I provides a graphical illustration of how the framework above can be used to empirically quantify the welfare cost of adverse selection. The y-axis measures the price (or cost) of the loan. The x-axis measures the share of borrowers in the market (quantity). The demand curve plots the share of potential borrowers willing to take out a loan at each potential price $r$.

The defining characteristic of markets featuring adverse selection is that borrowers with the highest willingness to pay for credit will also have the highest expected costs. This is represented in the figure by the downward-sloping scaled marginal cost curve. At high prices (low quantities), only those with high expected costs choose to take up the loan. As the price is lowered, the marginal borrower drawn into the market has lower expected costs. Because inframarginal borrowers always have higher expected costs than marginal borrowers, average cost will always be greater than marginal cost. This is represented in the figure by the relatively less steeply-sloped scaled average cost curve.

The welfare loss of adverse selection arises from the fact that lenders are unable to set interest rates based on each individual borrower’s (unobservable) marginal cost and must instead set a single price for everyone. In equilibrium, this price is equal to $\overline{AC}$. Because average cost is always greater than marginal cost, fewer borrowers than would be efficient end up receiving a loan. In the figure, the intersection of the $\overline{AC}$ curve and the demand curve at B gives the equilibrium price $r^{EQ}$ and quantity $Q^{EQ}$. The intersection of the $\overline{MC}$ curve and the demand curve at C gives the efficient price $r^{EF}$ and quantity $Q^{EF}$. The equilibrium price is higher than the efficient price, resulting in an underprovision of credit ($Q^{EQ} < Q^{EF}$). The shaded region BCD gives the deadweight loss of adverse selection.

As the figure makes clear, knowledge of the demand and cost curves is sufficient for measuring the welfare loss of adverse selection. In the next section, we will discuss the data and source of variation in interest rates we use to estimate these curves in the market for online consumer loans in China.
**Moral Hazard**  In the basic framework outlined above, we assume that interest rates do not have a causal effect on default costs at the individual borrower level and that any relationship between interest rates and lender costs is therefore driven by borrower selection. However, an alternative and non-mutually exclusive reason why lenders’ costs would be increasing in interest rates is because higher interest payments induce individuals to default—an effect commonly referred to as moral hazard.\(^7\) In Appendix C we extend the basic framework to incorporate moral hazard and show that in the presence of such effects our empirical estimates will identify an upper bound on the pooled welfare losses generated by both adverse selection and moral hazard.

To accommodate moral hazard we allow borrower-level costs to explicitly depend on the interest rate, \(c(X_i, r) = c + \delta(X_i, r)\theta(X_i, r)\), and then re-derive the conditions for equilibrium and efficiency. While most of the basic intuition from the pure adverse selection case remains, the derivation of the socially efficient outcome in equation (9) and the condition on pricing that implements this outcome in equation (10) both change in important ways. In particular, under moral hazard increasing the number of borrowers in the market by lowering the interest rate will affect total surplus not just by lowering the expected default cost of the marginal borrower but also by lowering the expected costs of all inframarginal borrowers willing to borrow at the previously higher rate. It is the latter effect that leads our estimates to overstate the magnitude of the pooled welfare losses when both adverse selection and moral hazard are present.

### III Empirical Setting and Data

**III.A Institutional Background**

Fintech lending platforms allow individuals and small businesses to borrow without the presence of traditional financial intermediaries by directly connecting borrowers with potential investors. These platforms first emerged in the United Kingdom in 2005. Following more than a decade of rapid growth, fintech lenders are now a significant supplier of unsecured consumer credit, with global origination volumes reaching $297 billion in 2018. China, the United States, and the United Kingdom are the three largest markets in the industry and collectively account for roughly two-thirds of total lending volume (Cornelli et al., 2020). According to TransUnion, fintech lenders are now the most popular source for unsecured consumer installment loans in the

---

\(^7\)A large number of models can generate a causal relationship between interest rates and default costs. These range from models of strategic behavior on the part of borrowers to models of liquidity default in which higher payments are more difficult to service. For our purposes the key distinction is that these mechanisms will all lead lender costs to be increasing in interest rates even holding the composition of the borrower pool constant. For simplicity and to distinguish it from adverse selection, we refer to this effect as moral hazard.
United States and accounted for 38 percent of total lending in that segment as of 2018.⁸

Unlike in developed countries, where online lenders compete with traditional banks for credit-worthy borrowers, fintech platforms in China mostly serve those who are underserved by banks. Due to an under-developed credit scoring system, Chinese banks are reluctant to lend to individuals and small firms. Unmet credit demand from these groups has fostered the growth of the Chinese online lending industry.

Chinese fintech platforms function similarly to their US counterparts. To apply for a loan, potential borrowers must first create an account and provide personal information to the platform. The information borrowers provide is then analyzed and sometimes combined with other third-party data to create an internal credit score. Because there are no official credit scores in China, Chinese platforms rely heavily on these internal scores both to determine whether borrowers qualify for credit and to set the terms of their loan. If borrowers accept the offered loan terms, the loan requests are then posted online where potential investors can choose to fund them. Investors can fund loans either at the individual level or by investing in portfolios of loans through wealth management products provided by the platform.

### III.B Our Lender and Experiment

Our data come from a randomized experiment conducted by a major Chinese lending platform. This platform provides small consumer installment loans with maturities ranging from 3 to 24 months. In the experiment we study, all loans have a 12-month maturity, which is the platform’s most popular product. Borrowers who are approved for a loan through the platform are offered a maximum loan size and quoted a total cost of borrowing. The loan size borrowers are offered depends on their assigned credit score, but all borrowers are quoted the same borrowing cost. While borrowers can in principle choose to borrow less than the offered maximum, the vast majority (91 percent) of those who take-up the loan borrow the full amount. For this reason, we abstract from the choice of loan size in our main analysis.

The total cost of borrowing consists of three components. The first is a base interest rate that is constant across all borrowers who apply at a given point in time but varies with the date of application. The base interest rate is remitted to the investor(s) who fund the loan in 12 monthly installments and guaranteed by the platform. The second component of the borrowing cost is an addition to the base interest rate called the “quality assurance fee.” The quality assurance fee varies across applicants at a point in time as a function of the credit rating assigned by the platform and is paid directly to the platform. The third component of the borrowing cost is an origination fee that is paid in equal monthly installments directly to the lending platform during the first

three months of the loan. This fee also varies by date of application and borrower credit rating. While both the quality assurance and origination fees depend on borrowers’ credit ratings, they are set in a manner that holds the total cost of borrowing constant across ratings. Borrowers who receive a loan offer do not directly see each individual component of the total borrowing cost. Instead, they are presented with a full 12-month payment schedule that incorporates these financing charges along with the amortizing portion of the principal.

The experiment we analyze was designed independently of this study by the platform to examine applicants’ sensitivity to borrowing costs and was implemented from January to March, 2018. The platform randomly selected 11,180 potential borrowers into the experiment who had successfully applied and qualified for credit during this window. Selected applicants were then randomly divided into two equal-sized groups: one control group and one treatment group. Applicants in the control group were offered standard financing terms, while the treatment group received a 50 percent reduction in the portion of the borrowing cost coming from the quality assurance and origination fees. Ignoring discounting and summing the financing charges across all 12 months of the loan, the average total cost of borrowing was equal to 36 percent of the original principal balance in the control group and 21.5 percent in the treatment group. Throughout the rest of the paper we will refer to this total cost of borrowing as the interest rate and to the two groups as the “High-Price” and “Low-Price” groups respectively.

### III.C Descriptive Statistics and Balance Checks

The data we use are at the loan offer level and contain three types of information. The first includes the basic terms of the offer: loan size and interest rate. The second includes information about the applicant, such as the platform’s internal credit rating (A–D) and basic demographics like age, gender, marital status, education, and city of residence tier (1–6). The third includes the full repayment history for those who take up the offer and recovery amounts on any defaulted loans.

Table I compares the mean characteristics of applicants across the High-Price and Low-Price groups. The results indicate that the randomization was successful. With the exception of the interest rate, which is mechanically different across treatment and control, there is no statistically or economically significant difference across the two groups in the covariates we examine: loan size, age, the fraction of male borrowers, the fraction of single borrowers, education level, fraction of borrowers in the six city tiers, and the fraction of borrowers in the four credit rating categories. The average applicant is 30 years old and receives a loan offer of ¥6300 ($900).  

---

9. In this table and throughout our analysis we include only applicants with non-missing values for all covariates. This reduces the sample size from 11,180 to 10,991.

10. In Appendix Figure A.1 we show that not only the mean, but the full distribution of these variables is similar.
The sample skews male (77 percent) and is roughly evenly split between married and single applicants. Approximately 45 percent of applicants have received their highest degree from either a vocational school or a traditional college and 41 percent of them live in first- or second-tier cities. While comprehensive data on the characteristics of borrowers participating in the online Chinese lending market are hard to come by, Appendix Table A.1 shows that the borrowers in our sample are roughly similar to those borrowing from other large lending platforms along the characteristics we can observe.

IV Estimation

They key outcome variables required to estimate the welfare loss arising from adverse selection in this market are borrower take-up and loan charge-offs. Combining data on these outcomes with exogenous variation in interest rates allows us to estimate both the demand curve $D(r)$ and the average cost curve $AC(r)$. Below, we will show that the scaled average cost curve $\overline{AC}(r)$ and scaled marginal cost curve $\overline{MC}(r)$ can both be constructed directly from the demand and average cost curves and therefore do not require separate estimation. As discussed in Section II, knowledge of $D(r)$, $\overline{AC}(r)$, and $\overline{MC}(r)$ is sufficient for calculating both the equilibrium and efficient outcomes and therefore also any deadweight loss due to selection. This section describes our approach to estimating these quantities empirically.

IV.A Estimating the Demand and Cost Curves

Since our data come from a randomized experiment, estimating demand and average cost is relatively straightforward. We measure demand using an indicator $d_i$ for whether loan applicant $i$ takes up the loan, which maps directly into the notion of demand discussed in Section II. The cost $c_i$ of providing a loan to borrower $i$ consists of two components. The first component, charge-offs, varies across borrowers and is taken directly from the data. Specifically, charge-offs are measured as the realized share of total promised payments never received or recovered by the lender. The second component, fixed costs, does not vary across borrowers and is not directly measured in the data. For our baseline estimates, we calibrate this measure to external information on the lender’s cost of funds and administrative costs.\footnote{Specifically, we assume a 4.35 percent cost of funds, which is equal to the average one-year benchmark rate set by the People’s Bank of China during our sample period. On top of this, we also add administrative costs equal to 10 percent of the original loan balance, which is on par with estimates of customer acquisition and operational expenses provided to us by the lender.} In Section IV.C below, we provide an alternative approach that allows for estimation of these fixed costs in contexts when external targets for calibration are not available.

across the two groups.
With these measures of borrower-level demand and cost in hand, we can then estimate the demand and average cost curves using the following two equations

\[ d_i = \alpha_d + \beta_d r_i + \epsilon_i \]  

(12)

\[ c_i = \alpha_c + \beta_c r_i + \nu_i. \]  

(13)

These equations are estimated via two-stage least squares where the interest rate \( r_i \) is instrumented using an indicator variable for whether the borrower was assigned to the High-Price or Low-Price group. Random assignment ensures that this instrument is valid and allows us to estimate these two curves using variation in interest rates that is orthogonal to unobserved drivers of both demand and costs. As in the theory, we estimate the demand curve in the sample of all loan applicants and the average cost curve in the sample of loan applicants who chose to take-up the loan at the offered interest rate.

The scaled average and marginal cost curves \( \widetilde{AC}(r) \) and \( \widetilde{MC}(r) \) can be computed directly using the coefficient estimates from equations (12) and (13). To see this, note from the definition of \( \widetilde{AC}(r) \) in equation (5) that

\[ \widetilde{AC}(r) = \frac{\alpha_c + \beta_cr}{(1+c) - (\alpha_c + \beta_cr)}, \]  

(14)

where \( c \) denotes the component of costs that does not vary across borrowers. Similarly, using the definition of marginal costs we can calculate

\[ MC(r) = \frac{\partial TC(r)}{\partial D(r)} = \frac{\partial (AC(r) \times D(r))}{\partial D(r)} = \frac{\alpha_d \beta_c}{\beta_d} + \alpha_c + 2\beta_cr, \]  

(15)

where \( TC(r) \) denotes the lender’s total cost. Letting \( \alpha_m = \frac{\alpha_d \beta_c}{\beta_d} + \alpha_c \) and \( \beta_m = 2\beta_c \), the scaled marginal cost curve can then be written as

\[ \widetilde{MC}(r) = \frac{\alpha_m + \beta_mr}{(1+c) - (\alpha_m + \beta_mr)}. \]  

(16)

**IV.B Calculating Welfare Losses**

With the demand and scaled cost curves available we can compute both the equilibrium and efficient prices and quantities. The equilibrium price and quantity are determined by the intersection of the demand and scaled average cost curves. Setting \( \widetilde{AC}(r) = D(r) \) allows us to solve for this
point analytically, yielding

\[ r^{EQ} = \frac{\phi_c - \sqrt{\psi_c}}{2\beta_c} \quad \text{and} \quad Q^{EQ} = \alpha_d + \beta_d \times r^{EQ}, \]

(17)

where \( \phi_c = 1 + c - \alpha_c - \beta_c \) and \( \psi_c = \phi_c^2 - 4\alpha_c \beta_c \). The efficient price and quantity are similarly determined by equating \( MC(r) = D(r) \), yielding

\[ r^{EF} = \frac{\phi_m - \sqrt{\psi_m}}{2\beta_m} \quad \text{and} \quad Q^{EF} = \alpha_d + \beta_d \times r^{EF}, \]

(18)

where \( \phi_m = 1 + c - \alpha_m - \beta_m \) and \( \psi_m = \phi_m^2 - 4\alpha_m \beta_m \).

The welfare loss arising from adverse selection is given by the area of the shaded region between points B, C and D in Figure I. When the scaled marginal cost curve is roughly linear this area can be well approximated using a standard Harberger (1964) triangle formula:

\[ DWL \approx \frac{1}{2}(Q^{EF} - Q^{EQ})(r^{EQ} - \overline{MC}(r^{EQ})). \]

(19)

Alternatively, because equations (12) and (16) provide exact formulas for the demand and \( \overline{MC}(\cdot) \) curves, we can also compute this area directly by integrating between these curves over the range \([Q^{EQ}, Q^{EF}]\). In Section V below, we provide estimates of the welfare loss calculated in both ways along with the parameters of the demand and cost curves from which these estimates are constructed.

IV.C Estimation with Unknown Fixed Costs

Estimating the welfare costs of adverse selection requires us to take a stand on the value of \( c \), the component of costs that does not vary across borrowers. The value of this parameter governs the intercept of the average cost curve and therefore both the equilibrium and efficient outcomes. For our baseline estimates, we calibrate this value to external information on the lender’s fixed costs. However, in many other contexts reliable information on lenders’ fixed costs may be difficult to obtain. In this section, we describe an alternative approach that, with one additional assumption, allows us to estimate this parameter directly. This approach serves as a useful validation of any particular calibration of \( c \) and is also valuable in situations where external information on lender costs is unavailable.

The key assumption we make to allow for estimation of \( c \) is that the equilibrium we observe in the data corresponds to the competitive equilibrium benchmark against which we evaluate any potential welfare losses. Since any particular value of \( c \) produces only one equilibrium outcome
\((r^{EQ}, Q^{EQ})\), it is possible to estimate \(c\) by choosing the value that most closely replicates the observed price and quantity outcome in the data.

We implement this idea by performing a grid search over a large set of possible values for the fixed cost parameter ranging from 0 to 0.25 with a step size of 0.0001. Our chosen value of \(c\) is the one that minimizes the squared euclidean distance between the model-implied equilibrium outcome and the mean interest rate and take-up rate in the data. That is, we choose \(c\) to minimize 
\[
(r^{EQ} - \overline{r})^2 + (Q^{EQ} - \overline{Q})^2,
\]
where \(\overline{r}\) and \(\overline{Q}\) are the average interest rate and take-up rate across all loan applicants in the sample.\(^{12}\) To calculate standard errors for the demand and cost curves we bootstrap from the observed sample of loan applicants, drawing 100 random samples with replacement and re-estimating both the fixed cost parameter and the demand and cost curves at each iteration. All other implied quantities are then calculated as described above using the chosen value of \(c\) as an input into the estimation.

\section*{V Results}

As a starting point for our empirical analysis, Figure II plots the raw data for the two key outcome variables of interest. Panel A. plots the take-up rate. Each bar in the figure represents the share of applicants in the Low-Price or High-Price group that ultimately decided to take up the loan offer. For applicants in the Low-Price group, the take-up rate was approximately 65 percent. This is substantially higher than the 58 percent take-up rate among applicants in the High-Price group and indicates that borrower demand is indeed sensitive to interest rates. Panel B. plots the charge-off rate. On average, borrowers in the High-Price group who choose to take up a loan fail to pay back roughly 13 percent of the promised payments. This is significantly higher than the corresponding charge-off rate in the Low-Price group, which was approximately 11 percent. Taken together, these results indicate the presence of adverse selection. Borrowers who choose to take up the loan offer when the interest rate is higher have, on average, higher expected costs. This implies that the average cost curve is downward-sloping in quantity, as shown in Figure I. In the next section, we estimate these demand and cost curves directly and use them to construct our measure of the welfare losses arising from adverse selection.

\(^{12}\)Since our data come from an experiment in which the lender is actively changing prices, it is not clear which of the two prices (treatment or control) would map most directly to the competitive equilibrium market price. For this reason, we choose to match the mean outcome across treatment arms. Results are similar if we instead match only the mean interest rate and take-up rate in the control group, where the lender did not alter its baseline pricing.
Table II presents our main results.\textsuperscript{13} In Panel A., we report estimates of the demand curve from various versions of the basic regression specification given by equation (12). Column 1, which includes no controls, confirms the evidence of a downward sloping demand curve from Figure II. The coefficient estimate in the top row indicates that a 10 percentage point increase in offered interest rates reduces applicant take-up by approximately 4.3 percentage points. This is roughly the same number that would be obtained by simply dividing the 6.3 percentage point difference in take-up rates across the two treatment groups by the 14.5 percentage point mean difference in interest rates reported in Table I.

In columns (2)–(4) we add a series of control variables to the specification.\textsuperscript{14} Column (2) adds controls for borrower demographics, which include a linear term in age and indicator variables for gender, marital status, and highest degree completed. Column (3) further controls for geographic location by adding indicators for the city tier in which the borrower lives. Finally, in column (4) we add controls for both loan size and the borrower’s credit rating. These controls are potentially important given that our conceptual framework assumes a constant loan size and that interest rates can vary across borrowers with different credit ratings. Across all specifications, the point estimates are nearly identical to those obtained in column (1). The consistency of these results across specifications echoes the finding of near complete covariate balance from Section III.

Panel B. reports analogous results for the average cost curve obtained by estimating equation (13). The slope coefficient reported in the top row of the first column indicates that a 10 percentage point increase in offered interest rates increases the average charge-off rate among those who accept the loan offer by roughly 1 percentage point. This positive coefficient indicates the presence of adverse selection—borrowers who choose to take up a loan at higher interest rates have, on average, higher expected costs. As with the demand curve, estimates of the cost curve are nearly identical across specifications including various controls.

Column (5) reports results from a version of the specification that is identical to that in column (4), but for which we estimate the fixed cost parameter rather than calibrating it. In this specification, the fixed cost parameter is chosen to minimize the difference between the model-implied equilibrium price and quantity and the observed averages of these variables as described

\textsuperscript{13}The first stage results corresponding to each instrumental variables regression in Table II are reported in Appendix Table A.2 for the demand curve and Appendix Table A.3 for the cost curve. Unsurprisingly given the randomization, these results indicate the presence of a strong and stable first stage. The average interest rate in the High-Price group is estimated to be 14.5 percentage points higher than that of the Low-Price group. This difference is statistically significant at the one percent level in all specifications with F-statistics far above conventional thresholds for rejecting the presence of weak instruments (\cite{Stock and Yogo, 2005; Lee et al., 2020}).

\textsuperscript{14}All control variables are demeaned prior to estimation to ensure that the intercept term can be interpreted similarly across specifications.
in Section IV. Reassuringly, the results from this specification are very similar to those from column (4), suggesting that our calibration of the fixed cost parameter in columns (1)–(4) is reasonable.\footnote{The estimated value of the fixed cost parameter in column (5) is equal to 0.1309, which is indeed close to our calibration of 0.1435 in columns (1)–(4). In Figure A.2, we plot the value of the objective function at all candidate fixed costs we consider. This figure shows that the objective function is convex over the range of values we consider and therefore that our estimated fixed cost represents a unique solution to the minimization problem over this range.}

In Panel C, we combine the demand and cost curve estimates from each specification to produce several implied quantities of interest. The top four rows report both the equilibrium and efficient price and quantity. These outcomes are determined by the intersection of the demand curve with the $\widetilde{AC}$ and $\widetilde{MC}$ curves, respectively. We calculate them as discussed in Section IV using the coefficient estimates from Panels A. and B. The results indicate that adverse selection leads to a large equilibrium price distortion. The equilibrium price under perfect competition is approximately 22 percentage points higher than the efficient level. This large price difference, however, leads to a relatively small difference in quantities. Borrower take-up at the equilibrium price is only about 10 percentage points lower than it would be at the efficient price. This corresponds to an interest rate elasticity of demand equal to approximately $-0.13.\footnote{We calculate the interest rate elasticity at the midpoint between the equilibrium and efficient allocations, yielding $\left(\frac{0.703-0.608}{0.703+0.608}/2\right)\left(\frac{0.085-0.304}{0.085+0.304}/2\right) = -0.13$}

The bottom two rows of the table report estimates of the implied welfare loss arising from competitive pricing in the presence of adverse selection. We construct these estimates in two ways. First, we use the approximation formula from equation \eqref{eq:approximation}. Second, we calculate the exact welfare loss by integrating between the demand and scaled marginal cost curves over the interval containing the competitive and efficient quantities. In both cases we report welfare losses per ¥100 originated. The estimates are nearly identical whether we construct them using the approximation formula or the exact solution. For example, the numbers reported in column (1) indicate that the welfare cost of adverse selection is equal to roughly 0.8 percent of the loan amount. These estimates are statistically significant with ninety-five percent confidence intervals ranging from roughly 1 to 2.8 percent.\footnote{Table A.4 reports bootstrapped confidence intervals for the implied welfare losses as well as all other quantities reported in Panel C. of Table II. To construct these confidence intervals we draw 1,000 independent samples with replacement from our data and reestimate the demand curve, cost curve, and implied quantities in each sample. The upper and lower bounds of the ninety-five percent confidence interval are given by the 97.5th and 2.5th percentiles of the distribution of estimates across these random samples.} Given the average loan size reported in Table I, 0.8 percent equates to a welfare loss of ¥50, or approximately $7.20 per applicant. This small overall welfare loss is driven by the fact that borrower demand is inelastic to interest rates. That is, the inefficiently high pricing due to adverse selection generates only modest quantity distortions and therefore small overall welfare losses.

Figure III presents a graphical representation of the results from column (1) that is a direct
empirical analog to Figure I. As in that figure, the y-axis measures the price (or cost) of the loan and the x-axis measures the share of borrowers who take-up the loan. The solid blue line plots our estimated demand curve. The dashed and solid orange lines plot the scaled average and marginal cost curves, respectively. The equilibrium and efficient outcomes, which are determined by the intersection of the demand curve with the scaled average and marginal cost curves, are also indicated in the figure. The shaded region measures the welfare loss due to adverse selection. Because the scaled marginal cost curve is nearly linear between the equilibrium and efficient prices, this area is roughly triangular. This explains why the estimated welfare loss is similar whether we measure it exactly or approximate it using the formula from equation (19).

Figure III also provides a useful visual gauge of how far out of sample we must extrapolate the estimated demand and cost curves to calculate the area corresponding to the welfare loss. This is important given the linearity assumptions we impose when estimating the demand and average cost curves. Reassuringly, estimating the equilibrium outcome requires essentially no out-of-sample extrapolation. In our experiment, the mean interest rates in the High- and Low-Price groups were 36 and 21.5 percent, respectively. Our estimate of the equilibrium interest rate of 30.4 percent falls nearly in the middle of this range. Calculating the efficient outcome, however, does require some extrapolation. The efficient price of 8.5 percent is roughly 13 percentage points lower than the mean interest rate observed in the Low-Price group. This means that we must rely on functional form assumptions to estimate the area under the demand curve and above the scaled marginal cost curve between these two prices. Nonetheless, we view the cost of needing to impose these assumptions to be relatively small given the otherwise minimal assumptions we make and the benefits that arise from the ability to draw meaningful and transparent welfare conclusions.

V.B Heterogeneity by Credit Scores

While the overall welfare losses we document are small, there is interesting heterogeneity across the distribution of ex-ante credit risk. To show this, Table III repeats our main analysis for subsamples split according to the applicants’ assigned credit rating. Columns (1) and (2) restrict attention to applicants assigned an “A” rating. These applicants are judged by the platform to pose the lowest ex-ante credit risk based on observable characteristics and constitute roughly half of the sample. Columns (3) and (4) focus on the remaining half of applicants who received lower ratings of B–D. As in Table II, we report estimates for the demand curve, average cost curve, and implied welfare losses separately in Panels A., B., and C., respectively.

The demand curve results from Panel A. indicate that high- and low-credit-score applicants are equally sensitive to interest rates. As in the main analysis, we find that a 10 percentage point

---

18In Figure A.3, we also plot the estimated average cost curve from which the two scaled cost curves are derived along with the actual data points being used to estimate these curves.
increase in the offered interest rate reduces applicant take-up by approximately 0.4–0.45 percentage points, regardless of the applicant’s credit rating. This range is similar whether we exclude control variables, as in columns (1) and (3), or include the full set of non-credit score controls, as in columns (2) and (4). Thus, increasing the interest rate has a similar effect on the overall level of demand in the two credit score groups.

The cost curve results from Panel B., however, indicate that increasing the interest rate shifts the composition of borrowers who choose to take up the loan very differently across the two groups. For low-credit-score borrowers, we estimate that a 10 percentage point increase in offered interest rates raises the average charge-off rate among those who choose to take up the loan by 1.2–1.3 percentage points. These estimates are roughly 20–30 percent larger than the effect in the pooled sample and are statistically significant at conventional levels. In contrast, offering higher interest rates has relatively little effect on costs for the high-credit-score group. We estimate statistically insignificant coefficients that are only about half as large as the effect for low-credit-score borrowers.

Taken together, these findings imply that the correlation between demand and unobservable lending costs is higher among observably higher-risk borrowers. As a result of this, the welfare losses of adverse selection are also larger for this group. Panel C. reports the implied welfare losses separately for each credit score group along with the associated equilibrium and efficient outcomes. For low-credit-score borrowers, we estimate welfare losses equal to about 1.4–1.6 percent of the loan amount. This is roughly twice the magnitude of the welfare losses estimated in the pooled sample and more than four times as large as the losses among high-credit-score borrowers. Nonetheless, despite these large relative differences across credit scores, the overall magnitude of the welfare loss due to adverse selection remain small. For example, our highest estimate for low-credit-score borrowers still only implies a per-applicant deadweight loss of about ¥100 ($14.40).

VI Concluding Remarks

This paper provides estimates of the welfare losses arising from adverse selection in the new and growing fintech consumer credit market. Leveraging a randomized experiment conducted by a popular Chinese lending platform, we document that adverse selection is present in this market and leads to a large equilibrium price distortion. However, due to inelastic borrower demand, this price distortion generates a relatively small distortion in equilibrium quantities. As a result, the overall welfare losses we estimate are fairly small, amounting to only 0.8 percent of the typical loan amount on a per-applicant basis. This loss is larger among observably higher-risk borrowers, though still small in absolute terms. The small overall welfare losses we document make it hard
to argue for typical policy remedies, such as interest rate subsidies or loan guarantees, to address adverse selection in this market.

While we find small welfare losses of adverse selection in the market we study, an additional contribution of the paper is to demonstrate how a versatile methodological approach to welfare analysis adapted from the empirical literature on selection in insurance markets can be applied in the context of consumer credit markets. The same methodology, perhaps with slight modifications, could also be applied to measure efficiency losses in the mortgage, auto, credit card, and student loan markets, among others. The key input required to apply these methods in a credit market context is data on interest rates, loan take-up, and charge-offs, all of which are readily available in many settings. We hope that future work will apply this methodology in diverse credit market settings to gain insights into the conditions under which adverse selection generates large or small welfare losses. Beyond measuring the welfare consequences of selection in such settings, there is also significant scope for further work linking theory and data to study optimal policy interventions in cases where these losses are substantial.
REFERENCES


NOTE.— This figure provides a graphical illustration of the welfare cost of adverse selection. The x-axis measures the share of potential borrowers in the market. The y-axis measures the price (or cost) of the loan. The market depicted in the figure features adverse selection because the scaled marginal cost curve is downward-sloping. Borrowers with the highest willingness to pay for credit are also those with the highest expected costs. The competitive equilibrium is determined by the intersection of the demand and $\overline{AC}$ curves (point B). The efficient allocation is determined by the intersection of the demand and $\overline{MC}$ curves (point C). The shaded region (BCD) depicts the welfare loss arising from the underprovision of credit due to adverse selection.
FIGURE II
Mean Take-up and Charge-off Rates by Treatment Arm

Panel A. Take-up Rate

Panel B. Charge-off Rate

NOTE.—This figure presents the average take-up rates (Panel A) and charge-off rates (Panel B) in the two treatment groups. Each bar plots the mean outcome across borrowers in a given treatment group along with its 95% confidence interval. Take-up rates are calculated in the full sample of loan applicants and charge-off rates are calculated among only those applicants who choose to take up the loan.
FIGURE III
Empirical Estimates of Welfare Cost of Adverse Selection

Note.—This figure presents the empirical analog to figure I. The x-axis measures the share of potential borrowers in the market. The y-axis measures the price (or cost) of the loan. The curves plotted in this figure are derived from the point estimates in column 1 of Table II as described in Section IV. The competitive equilibrium is determined by the intersection of the demand and $\bar{AC}$ curves. The efficient allocation is determined by the intersection of the demand and $MC$ curves. The equilibrium price and quantities $(P^{EQ}, Q^{EQ})$ and efficient price and quantities $(P^{EF}, Q^{EF})$ are denoted in the figure. The shaded region (BCD) depicts the welfare loss arising from the underprovision of credit due to adverse selection.
### Table I: Summary Statistics and Covariate Balance Test

<table>
<thead>
<tr>
<th></th>
<th>Low Price</th>
<th>High Price</th>
<th>Difference</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Loan terms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest Rate (%)</td>
<td>21.5</td>
<td>36.0</td>
<td>-14.5</td>
<td>-291.45***</td>
</tr>
<tr>
<td>Loan size (¥100’s)</td>
<td>62.7</td>
<td>62.5</td>
<td>0.2</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>Basic Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>30.4</td>
<td>30.5</td>
<td>-0.1</td>
<td>-0.61</td>
</tr>
<tr>
<td>Male (%)</td>
<td>77.3</td>
<td>77.2</td>
<td>0.1</td>
<td>0.08</td>
</tr>
<tr>
<td>Single (%)</td>
<td>50.4</td>
<td>51.3</td>
<td>-0.8</td>
<td>-0.86</td>
</tr>
<tr>
<td><strong>Highest Degree Completed (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior-High School</td>
<td>12.8</td>
<td>13.0</td>
<td>-0.2</td>
<td>-0.31</td>
</tr>
<tr>
<td>Senior-High School</td>
<td>42.4</td>
<td>41.6</td>
<td>0.8</td>
<td>0.85</td>
</tr>
<tr>
<td>Vocational School</td>
<td>31.2</td>
<td>32.2</td>
<td>-0.9</td>
<td>-1.06</td>
</tr>
<tr>
<td>Bachelor’s or Higher</td>
<td>13.5</td>
<td>13.2</td>
<td>0.3</td>
<td>0.52</td>
</tr>
<tr>
<td><strong>City Tier (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tier 1</td>
<td>12.4</td>
<td>11.9</td>
<td>0.5</td>
<td>0.76</td>
</tr>
<tr>
<td>Tier 2</td>
<td>29.1</td>
<td>29.4</td>
<td>-0.3</td>
<td>-0.38</td>
</tr>
<tr>
<td>Tier 3</td>
<td>21.2</td>
<td>21.4</td>
<td>-0.2</td>
<td>-0.28</td>
</tr>
<tr>
<td>Tier 4</td>
<td>23.4</td>
<td>23.1</td>
<td>0.4</td>
<td>0.44</td>
</tr>
<tr>
<td>Tier 5</td>
<td>2.8</td>
<td>2.8</td>
<td>-0.1</td>
<td>-0.18</td>
</tr>
<tr>
<td>Tier 6</td>
<td>11.1</td>
<td>11.4</td>
<td>-0.2</td>
<td>-0.37</td>
</tr>
<tr>
<td><strong>Credit Rating (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A Rating</td>
<td>50.8</td>
<td>50.0</td>
<td>0.8</td>
<td>0.87</td>
</tr>
<tr>
<td>B Rating</td>
<td>15.3</td>
<td>16.0</td>
<td>-0.8</td>
<td>-1.12</td>
</tr>
<tr>
<td>C Rating</td>
<td>26.4</td>
<td>25.8</td>
<td>0.6</td>
<td>0.66</td>
</tr>
<tr>
<td>D Rating</td>
<td>7.5</td>
<td>8.1</td>
<td>-0.6</td>
<td>-1.19</td>
</tr>
<tr>
<td><strong>Number of Observations</strong></td>
<td>5,479</td>
<td>5,512</td>
<td>10,991</td>
<td>10,991</td>
</tr>
</tbody>
</table>

**Note.**—This table reports the average characteristics for borrowers in the two treatment groups and the differences in the averages. Except for age and loan size, all covariates are indicator variables. We multiply the indicator variables by 100 so that each mean represents the percentage of borrowers in a given category. Columns (1) and (2) report, respectively, average characteristics for borrowers who face the high price and low prices. Column (3) presents the difference between columns (1) and (2) and column (4) reports the t-statistics from a two-sided t-test for equality of means across the two groups. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.
### TABLE II
DEMAND, AVERAGE COST, AND WELFARE ESTIMATES

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. Demand</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest Rate</td>
<td>-0.433***</td>
<td>-0.429***</td>
<td>-0.430***</td>
<td>-0.425***</td>
<td>-0.425***</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.063)</td>
<td>(0.063)</td>
<td>(0.063)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.740***</td>
<td>0.739***</td>
<td>0.739***</td>
<td>0.737***</td>
<td>0.737***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>10,991</td>
<td>10,991</td>
<td>10,991</td>
<td>10,991</td>
<td>10,991</td>
</tr>
<tr>
<td><strong>Panel B. Average Cost</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest Rate</td>
<td>0.096**</td>
<td>0.094**</td>
<td>0.093**</td>
<td>0.090**</td>
<td>0.090*</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.044)</td>
<td>(0.044)</td>
<td>(0.043)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.238***</td>
<td>0.238***</td>
<td>0.238***</td>
<td>0.239***</td>
<td>0.227***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>6,761</td>
<td>6,761</td>
<td>6,761</td>
<td>6,761</td>
<td>6,761</td>
</tr>
<tr>
<td>Demographics</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loan Size and Rating</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated Fixed Cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Panel C. Implied Quantities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equilibrium Price</td>
<td>0.304</td>
<td>0.304</td>
<td>0.304</td>
<td>0.304</td>
<td>0.288</td>
</tr>
<tr>
<td>Equilibrium Quantity</td>
<td>0.608</td>
<td>0.608</td>
<td>0.608</td>
<td>0.608</td>
<td>0.615</td>
</tr>
<tr>
<td>Efficient Price</td>
<td>0.085</td>
<td>0.088</td>
<td>0.091</td>
<td>0.095</td>
<td>0.081</td>
</tr>
<tr>
<td>Efficient Quantity</td>
<td>0.703</td>
<td>0.701</td>
<td>0.700</td>
<td>0.697</td>
<td>0.703</td>
</tr>
<tr>
<td>Welfare Loss (per ¥100): Approximate</td>
<td>0.825</td>
<td>0.798</td>
<td>0.780</td>
<td>0.745</td>
<td>0.736</td>
</tr>
<tr>
<td>Welfare Loss (per ¥100): Exact</td>
<td>0.828</td>
<td>0.801</td>
<td>0.782</td>
<td>0.748</td>
<td>0.738</td>
</tr>
</tbody>
</table>

**Note.** — This table reports estimates of the demand curve, average cost curve, and implied welfare loss of adverse selection. Panel A. reports the results from estimating the demand equation using the full sample of loan applicants. Panel B. reports the results from estimating the cost equation in the sample of applicants who take up the loan. In all specifications, we instrument for the interest rate using an indicator variable for whether the applicant was assigned to the High-Price or Low-Price group. Across columns we gradually add controls for demographics, borrower geography, credit rating, and loan size. Demographics include a linear term in age as well as indicator variables for gender, marital status and highest degree completed. Geographic controls include a series of indicator variables for city tier. We control for loan size using a linear term and credit rating with a series of indicator variables. All control variables are demeaned prior to estimation so that the intercept term can be interpreted similarly across specifications. Panel C. reports implied quantities of interest for welfare analysis. These quantities are calculated from the coefficients in Panels A. and B. as described in Section IV. In columns (1)–(4) the fixed cost parameter is calibrated to external estimates. In column (5) the fixed cost parameter is estimated to minimize the squared euclidean distance between the model-implied equilibrium outcome and the mean observed interest rate and take-up rate across the two treatment arms. Standard errors in column (5) are calculated by bootstrapping from the observed sample of loan applicants, drawing 100 random samples with replacement and re-estimating all parameters at each iteration. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.
TABLE III
HETEROGENEITY BY BORROWER CREDIT RATING
IN DEMAND, AVERAGE COST, AND WELFARE ESTIMATES

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Demand</th>
<th>Panel B: Average Cost</th>
<th>Panel C: Implied Quantities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>-0.448***</td>
<td>-0.448***</td>
<td>-0.417***</td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
<td>(0.088)</td>
<td>(0.091)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.746***</td>
<td>0.778***</td>
<td>0.733***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>5,543</td>
<td>5,543</td>
<td>5,448</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>0.056</td>
<td>0.060</td>
<td>0.133**</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.056)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.230***</td>
<td>0.227***</td>
<td>0.247***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>3,432</td>
<td>3,432</td>
<td>3,329</td>
</tr>
<tr>
<td>Demographics</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Loan Size</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Equilibrium Price</td>
<td>0.272</td>
<td>0.270</td>
<td>0.344</td>
</tr>
<tr>
<td>Equilibrium Quantity</td>
<td>0.624</td>
<td>0.657</td>
<td>0.590</td>
</tr>
<tr>
<td>Efficient Price</td>
<td>0.156</td>
<td>0.140</td>
<td>0.016</td>
</tr>
<tr>
<td>Efficient Quantity</td>
<td>0.676</td>
<td>0.715</td>
<td>0.727</td>
</tr>
<tr>
<td>Welfare Loss (per ¥100): Approximate</td>
<td>0.263</td>
<td>0.329</td>
<td>1.661</td>
</tr>
<tr>
<td>Welfare Loss (per ¥100): Exact</td>
<td>0.263</td>
<td>0.330</td>
<td>1.676</td>
</tr>
</tbody>
</table>

NOTE.—This table reports estimates of the demand curve, average cost curve, and implied welfare loss of adverse selection separately by borrower credit rating. Panel A. reports the results from estimating the demand equation using the full sample of loan applicants. Panel B. reports the results from estimating the cost equation in the sample of applicants who take up the loan. Columns (1) and (2) include only loan applicants assigned an A credit rating, whereas columns (3) and (4) include only those assigned ratings of B, C, or D. In all specifications, we instrument for the interest rate using an indicator variable for whether the applicant was assigned to the High-Price or Low-Price group. Columns (2) and (4) include controls for demographics, borrower geography and loan size. Demographics include a linear term in age as well as indicator variables for gender, marital status and highest degree completed. Geographic controls include a series of indicator variables for city tier. We control for loan size using a linear term. All control variables are demeaned prior to estimation so that the intercept term can be interpreted similarly across specifications. Panel C. reports implied quantities of interest for welfare analysis. These quantities are calculated from the coefficients in Panels A. and B. as described in Section IV. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.
MEASURING THE WELFARE EFFECTS OF ADVERSE SELECTION IN CONSUMER CREDIT MARKETS

Online Appendices

Anthony A. DeFusco  Huan Tang  Constantine Yannelis
A ADDITIONAL DESCRIPTIVE STATISTICS

A.1 Continuous Covariate Distributions

Figure A.1 extends the covariate balance tests from Table I by showing the full distribution for the two continuous covariates in our analysis: loan size and borrower age. Each panel of the figure plots the distribution of the indicated variable separately for the High-Price and Low-Price groups. The vertically dashed orange line marks the mean value of the variable as reported in Table I. As is clear from the figure, the two treatment arms are similar along these dimensions both in terms of means and distributions. This provides further assurance that the randomization in the experiment we study was successful.

A.2 Sample Representativeness

In Table A.1 we compare the loans and borrowers in our sample to those from the ten largest online lending platforms in China, based on the transaction volume in the first quarter of 2018. The comparison data for other lending platforms were hand collected from WDZJ, an online platform used to compare lenders and loan terms. The summary statistics on loan terms and borrower characteristics for our sample are reported in the first column and those for other lending platforms in the second.

There is no evident first-order difference between loans offered on the platform we study and those offered by industry peers—except for loan size. The average loan size in our sample (¥6,255) is much lower than the industry average of ¥18,000. Borrowers in our sample are, however, similar to those borrowing on other platforms in terms of age distribution and marital status. As in our sample, the set borrowers using other platforms also tilts male, although slightly less so (65 versus 77 percent).

B SUPPLEMENTARY RESULTS

B.1 First Stage Regression Results

Table A.2 and Table A.3 report the first stage regression results for the demand and cost curve regressions, respectively. In both tables, the dependent variable is the interest rate and the excluded instrument is an indicator variable for whether the borrower was assigned to the High-Price group. The samples used to estimate these regressions are the same as in Table II and control
variables are introduced in the same order across columns. The bottom panel of each table reports the F-statistic from a test of the significance of the excluded instrument. The results from both tables indicate that being assigned to the High-Price group leads to a statistically significant increase of 14.5 percentage points in the interest rate. Given random assignment, this difference is identical to the simple mean difference in interest rates across the two groups as reported in the first row of Table I.

B.2 Objective Function Convexity for Fixed Cost Estimation

In column 5 of Table II we estimate the fixed cost parameter \( c \) by minimizing the squared euclidean distance between the model-implied equilibrium outcome, \((r^{EQ}, Q^{EQ})\), and the mean interest rate and take-up rate in the data. Figure A.2 plots the value of this objective function at each candidate fixed cost we consider on an evenly spaced grid ranging from 0 to 0.25 with a step size of 0.0001. The vertically dashed orange line marks the estimated value for the fixed cost parameter, which is equal to 0.1309. As the figure makes clear, the objective function is convex over the range of values we consider. This implies that our estimated fixed cost represents a unique solution to the minimization problem over the range of values we consider.

B.3 Confidence Intervals for Implied Quantities

Table A.4 reports bootstrapped ninety-five percent confidence intervals for our baseline welfare estimates as well as all other implied quantities of interest reported in Table II. To construct these confidence intervals, we draw 1,000 independent samples from our data and then reestimate the demand curve, cost curve, and all implied quantities separately in each sample. The upper and lower bound of each confidence interval is given by the 97.5th and 2.5th percentiles of the distribution of estimates across bootstrap replicates. For example, the numbers reported in the second row of the first column indicate that the implied equilibrium price fell below 0.294 or above 0.317 in only 5 percent of the 1,000 random samples in which we estimated it. For reference, the table also repeats our baseline estimates from the full sample as reported in Panel C. of Table II.

C Moral Hazard

Our baseline conceptual framework assumes that interest rates do not have a causal effect on on default costs at the individual borrower level and that any relationship between interest rates and lender costs is therefore driven by borrower selection. However, an alternative and non-mutually exclusive reason why lenders’ costs would be increasing in interest rates is because higher interest

---

1The first stage regressions for both the demand and cost curve in column 5 of Table II are identical to those in column 4 and are omitted from these tables.
payments induce individuals to default—an effect commonly referred to as moral hazard. In the presence of both moral hazard and adverse selection, our empirical analysis remains unchanged. However, the welfare loss measure we estimate will identify an upper bound on the pooled welfare losses generated by both adverse selection and moral hazard.

To show this, we incorporate moral hazard into the basic framework by allowing borrower-level costs to explicitly depend on the interest rate:

\[ c(X_i, r) = c + \delta(X_i, r)\theta(X_i, r) \]

The expression for the average cost curve can then be rewritten as

\[ AC(r) = \frac{1}{D(r)} \int c(X, r)\mathbb{1}(\rho(X) \geq r)dF(X). \]  

(1)

The marginal cost curve, defined as the change in total costs when additional borrowers are selected into the market at a marginally lower price, is now composed of two terms. The first term captures the costs associated with lending to these marginal borrowers. The second term reflects the impact of the price change on the costs of inframarginal borrowers due to moral hazard. This term naturally appears when \( c(X_i, r) \) is a function of \( r \). Formally, to derive the expression for marginal costs, we start from its definition,

\[ MC(r) = \frac{\partial TC(r)}{\partial D(r)} = \frac{1}{D'(r)} \frac{\partial TC(r)}{\partial r} = \frac{1}{D'(r)} \frac{\partial}{\partial r} \int c(X, r)\mathbb{1}(\rho(X) \geq r)dF(X). \]

(2)

Note that price affects total costs by altering both the pool of borrowers who select into the market, \( \mathbb{1}(\rho(X) \geq r) \), and the individual-level costs of all borrowers in the market at a particular price, \( c(X_i, r) \). Decomposing the last term in Equation (2) into these separate components allows us to re-write the marginal cost curve as

\[ MC(r) = -\frac{1}{D'(r)} \int c(X, r)\mathbb{1}(\rho(X) = r)dF(X) + \frac{1}{D'(r)} \int \frac{\partial c(X, r)}{\partial r}\mathbb{1}(\rho(X) \geq r)dF(X). \]

(3)

The first term, which is equal to \( \mathbb{E}[c(X, r)|\rho(X) = \rho(X_i)] \), is directly analogous to \( MC(r) \) in the baseline case, and the second represents the moral hazard component.

The expressions for the equilibrium interest rate, consumer surplus, producer surplus, and total surplus are the same as those in section Section II, but with the cost function \( c(X_i) \) replaced by \( c(X_i, r) \). The condition determining the constrained efficient allocation, however, needs to be modified in the presence of moral hazard. In particular, the derivative of total surplus with respect to quantity now includes an additional term that captures the impact of the price change
on the costs of inframarginal borrowers

\[
\frac{\partial TS(r)}{\partial D(r)} = -\frac{1}{D'(r)} \int \left[ (m^L(X) - m^N(X) - c(X, r)L) \mathbb{I}(\rho(X) = r) \\
+ \frac{\partial c(X, r)L}{\partial r} \mathbb{I}(\rho(X) \geq r) \right] dF(X).
\] (4)

Noting that \(m^L(X) - m^N(X) = (1 - \delta(X, r)\theta(X, r))\rho(X)L\) and rearranging terms in equation (4), we obtain

\[
\frac{\partial TS(r)}{\partial D(r)} = -L \left( MC(r) + \frac{1}{D'(r)} \int \rho(X)(1 - \delta(X, r)\theta(X, r)) \mathbb{I}(\rho(X) = r) dF(X) \right).
\] (5)

Thus, allocating a loan to potential borrower \(i\) increases total surplus if and only if

\[
\rho(X_i) \geq \frac{MC(\rho(X_i))}{1 + c - \mathbb{E}[c(X, r)|\rho(X) = \rho(X_i)]}.
\] (6)

Since \(\mathbb{E}[c(X, r)|\rho(X) = \rho(X_i)] \geq MC(\rho(X_i))\), the right hand side of equation (6) is larger than \(MC(\rho(X_i))\). Therefore, in the presence of moral hazard, the welfare loss estimated in Section IV.A provides an upper bound for the pooled losses due to adverse selection and moral hazard.
FIGURE A.1
Distributions of Loan Size and Applicant Age by Treatment Arm

NOTE.—This figure plots histograms showing the distribution of loan size (first column) and applicant age (second column) in the Low-Price (top row) and High-Price (bottom row) experimental treatment arms. The vertically dashed orange line marks the mean value of the variable in each sample.
Distance Between Observed and Equilibrium Outcomes

Candidate Fixed Cost

FIGURE A.2
Objective Function Values for Fixed Cost Estimation

Note.—This figure plots the value of the objective function used to estimate the fixed cost parameter $c$ at each candidate fixed cost on an evenly spaced grid ranging from 0 to 0.25 with a step size of 0.0001. For each candidate fixed cost, we estimate the demand and average cost curves and compute the implied equilibrium outcome $(r^{EQ}, Q^{EQ})$. The objective function is the squared euclidean distance between this implied equilibrium outcome and the observed outcome in the data, which we take to be the average observed interest rate and take-up rate across the two treatment arms. The vertically dashed orange line marks the estimated value for the fixed cost parameter (i.e., the point at which the distance between the equilibrium and observed outcomes is minimized).
NOTE.—This figure presents an extended version of Figure III. The x-axis measures the share of potential borrowers in the market. The y-axis measures the price (or cost) of the loan. The demand and average cost curves plotted correspond to those estimated in column 1 of Table II. The shaded circles and squares plot the actual data used to estimate these two curves. The two circles indicate the average observed interest rate and take-up rate in the High- and Low-Priced groups, whereas the two squares indicate the average observed cost (fixed cost plus charge-off rate) in each group. The scaled cost curves \( \overline{AC} \) and \( \overline{MC} \) are derived from the demand and average cost curves as described in Section IV. The competitive equilibrium is determined by the intersection of the demand and \( \overline{AC} \) curves. The efficient allocation is determined by the intersection of the demand and \( \overline{MC} \) curves. The equilibrium price and quantities \((P^E_Q, Q^E_Q)\) and efficient price and quantities \((P^E_F, Q^E_F)\) are denoted in the figure. The shaded region \((BCD)\) depicts the welfare loss arising from the underprovision of credit due to adverse selection.
## TABLE A.1
### Loan and Borrower Representativeness

<table>
<thead>
<tr>
<th></th>
<th>Our Sample</th>
<th>Industry Peers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Loan Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loan Size (¥)</td>
<td>6,255</td>
<td></td>
</tr>
<tr>
<td>Maturity (months)</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Investment Return (%)</td>
<td>10.83</td>
<td></td>
</tr>
<tr>
<td><strong>Borrower Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Under 40 (%)</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>Male (%)</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Single (%)</td>
<td>51</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE.**—Summary statistics for industry peers are calculated based on the ten largest online lending platforms in China by transaction volume as of the first quarter of 2018. Data are hand collected from a portal website for online lending platforms: [https://www.wdzj.com/](https://www.wdzj.com/). For example, statistics on loan characteristics are from [https://www.wdzj.com/zhuanti/2018report/](https://www.wdzj.com/zhuanti/2018report/) and borrower characteristics are from [https://www.wdzj.com/news/yc/2326325.html](https://www.wdzj.com/news/yc/2326325.html).
TABLE A.2
FIRST STAGE ESTIMATES FOR DEMAND CURVE

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interest Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-Price Group</td>
<td>0.145***</td>
<td>0.145***</td>
<td>0.145***</td>
<td>0.145***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.215***</td>
<td>0.215***</td>
<td>0.215***</td>
<td>0.215***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Demographics</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Loan Size and Rating</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-statistic</td>
<td>84,945</td>
<td>84,945</td>
<td>85,015</td>
<td>87,371</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>10,991</td>
<td>10,991</td>
<td>10,991</td>
<td>10,991</td>
</tr>
</tbody>
</table>

**NOTE.**—This table reports first stage results for the demand curve regressions. The first stage coefficient estimate on the excluded instrument is reported in the top row. The F-statistic from a test of the significance of the excluded instrument is reported in the bottom of the panel for each specification. Across columns we gradually add controls for demographics, borrower geography, credit rating, and loan size. Demographics include a linear term in age as well as indicator variables for gender, marital status and highest degree completed. Geographic controls include a series of indicator variables for city tier. We control for loan size using a linear term and credit rating with a series of indicator variables. All control variables are demeaned prior to estimation so that the intercept term can be interpreted similarly across specifications. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.
TABLE A.3
FIRST STAGE ESTIMATES FOR AVERAGE COST CURVE

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interest Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-Price Group</td>
<td>0.145***</td>
<td>0.145***</td>
<td>0.145***</td>
<td>0.145***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.211***</td>
<td>0.211***</td>
<td>0.211***</td>
<td>0.211***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Demographics</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loan Size and Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-statistic</td>
<td>59,143</td>
<td>59,083</td>
<td>59,092</td>
<td>60,439</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>6,761</td>
<td>6,761</td>
<td>6,761</td>
<td>6,761</td>
</tr>
</tbody>
</table>

NOTE.—This table reports first stage results for the average cost curve regressions. The first stage coefficient estimate on the excluded instrument is reported in the top row. The F-statistic from a test of the significance of the excluded instrument is reported in the bottom of the panel for each specification. Across columns we gradually add controls for demographics, borrower geography, credit rating, and loan size. Demographics include a linear term in age as well as indicator variables for gender, marital status and highest degree completed. Geographic controls include a series of indicator variables for city tier. We control for loan size using a linear term and credit rating with a series of indicator variables. All control variables are demeaned prior to estimation so that the intercept term can be interpreted similarly across specifications. Significance levels 10%, 5%, and 1% are denoted by *, **, and ***, respectively.
### TABLE A.4
Bootstrapped Confidence Intervals for Implied Quantities of Interest

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equilibrium Price</td>
<td>0.304</td>
<td>0.304</td>
<td>0.304</td>
<td>0.304</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>[0.294, 0.317]</td>
<td>[0.294, 0.317]</td>
<td>[0.294, 0.316]</td>
<td>[0.294, 0.317]</td>
<td>[0.278, 0.299]</td>
</tr>
<tr>
<td>Equilibrium Quantity</td>
<td>0.608</td>
<td>0.608</td>
<td>0.608</td>
<td>0.608</td>
<td>0.615</td>
</tr>
<tr>
<td></td>
<td>[0.597, 0.618]</td>
<td>[0.598, 0.618]</td>
<td>[0.598, 0.618]</td>
<td>[0.597, 0.618]</td>
<td>[0.605, 0.625]</td>
</tr>
<tr>
<td>Efficient Price</td>
<td>0.085</td>
<td>0.088</td>
<td>0.091</td>
<td>0.095</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>[-0.131, 0.286]</td>
<td>[-0.133, 0.294]</td>
<td>[-0.128, 0.294]</td>
<td>[-0.129, 0.299]</td>
<td>[-0.141, 0.285]</td>
</tr>
<tr>
<td>Efficient Quantity</td>
<td>0.703</td>
<td>0.701</td>
<td>0.700</td>
<td>0.697</td>
<td>0.703</td>
</tr>
<tr>
<td></td>
<td>[0.614, 0.781]</td>
<td>[0.612, 0.782]</td>
<td>[0.612, 0.780]</td>
<td>[0.608, 0.777]</td>
<td>[0.615, 0.783]</td>
</tr>
<tr>
<td>Welfare Loss (per ¥100): Approximate</td>
<td>0.825</td>
<td>0.798</td>
<td>0.780</td>
<td>0.745</td>
<td>0.736</td>
</tr>
<tr>
<td></td>
<td>[0.015, 2.720]</td>
<td>[0.013, 2.689]</td>
<td>[0.012, 2.663]</td>
<td>[0.009, 2.571]</td>
<td>[0.009, 2.510]</td>
</tr>
<tr>
<td>Welfare Loss (per ¥100): Exact</td>
<td>0.828</td>
<td>0.801</td>
<td>0.782</td>
<td>0.748</td>
<td>0.738</td>
</tr>
<tr>
<td></td>
<td>[0.007, 2.763]</td>
<td>[0.002, 2.733]</td>
<td>[0.001, 2.708]</td>
<td>[0.000, 2.609]</td>
<td>[0.000, 2.546]</td>
</tr>
</tbody>
</table>

Demographics X X X X
Geography X X X
Loan Size and Rating X X
Estimated Fixed Cost X

**Note.**—This table reports bootstrapped ninety-five percent confidence intervals (in brackets) for each point estimate reported in Panel C. of Table A.3. Confidence intervals are constructed by drawing 1,000 independent samples with replacement and reestimating the demand curve, cost curve, and all implied quantities separately in each sample. The ninety-five percent confidence interval is given by the 97.5th and 2.5th percentiles of the distribution of estimates across bootstrap samples. Point estimates for each implied quantity are also reported for reference and constructed as described in Section IV. Controls are introduced across columns in the same order as in Table II. Demographic controls include a linear term in age as well as indicator variables for gender, marital status and highest degree completed. Geographic controls include a series of indicator variables for city tier. We control for loan size using a linear term and credit rating with a series of indicator variables. All control variables are demeaned prior to estimation. In columns (1)–(4) the fixed cost parameter is calibrated to external estimates. In column (5) the fixed cost parameter is estimated to minimize the squared euclidean distance between the model-implied equilibrium outcome and the mean observed interest rate and take-up rate across the two treatment arms.