

MEASURING THE WELFARE COST OF ASYMMETRIC INFORMATION IN CONSUMER CREDIT MARKETS*

Anthony A. DeFusco[†] Huan Tang[‡] Constantine Yannelis[§]

Abstract

Information asymmetries are 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 asymmetric information 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 asymmetric information generates large equilibrium price distortions, we find only small overall welfare losses, particularly for high-credit-score borrowers.

*This version: September 12, 2021. We thank Scott Baker, Effi Benmelech, Greg Buchak, Doug Diamond, Liran Einav, Arpit Gupta, Zhiguo He, Sasha Indarte, Dirk Jenter, Gregor Matvos, Filippo Mezzanotti, Holger Mueller, Daniel Paravisini, Raghu Rajan, Amir Sufi, Kairong Xiao, Anthony Zhang, Yiwei Zhang and numerous other 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, IDC Herzliya Arison School of Business, 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, the Red Rock Finance Conference, Stanford Business School, the Stanford Institute for Theoretical Economics (SITE) Financial Regulation session, University of Michigan (Ross), University of Pennsylvania (Wharton), and Washington University in St. Louis. Yannelis gratefully acknowledges financial support from the Booth School of Business at the University of Chicago. The experiment we study in this paper has been registered with the AEA RCT Registry (RCT ID: AEARCTR-0008183), available at <https://www.socialscisearch.org/trials/8183>.

[†]Northwestern University, Kellogg School of Management and NBER

[‡]London School of Economics

[§]University of Chicago, Booth School of Business and NBER

I INTRODUCTION

Asymmetric information 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 asymmetric information 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 asymmetric information 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 or moral hazard relative to 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 asymmetric information 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 asymmetric information builds on key insights from the empirical literature on selection in insurance markets. As emphasized by Einav, Finkelstein and Cullen (2010), the central force that generates inefficiency in asymmetrically informed markets is that firms’ marginal costs are increasing in price. In a credit market context this can arise either because riskier borrowers are willing to accept higher interest rates—adverse selection—or because higher interest rates induce individual borrowers to default—moral hazard. Regardless of its source, this tight link between price and marginal cost leads to inefficiently high equilibrium pricing, as firms, who cannot price on the basis of 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

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. For a broader overview of the fintech sector and its relation to more traditional financial intermediation see Philippon (2016).

²Statistics on the size of the fintech lending market in China are taken from <https://www.wdzt.com/zhuanli/2018report>. Since 2017, the Chinese market has contracted considerably due to regulatory pressure but remains the largest fintech lending market globally.

is uniquely determined by the slopes of the consumer's demand curve and the firms' average and marginal cost curves. Empirical estimates of these curves thus constitute sufficient statistics for the welfare losses generated by asymmetric information. 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 information asymmetries: 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 asymmetric information 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. Both of these curves can be easily constructed from our empirical estimates of how demand and costs change in response to exogenous variation in interest rates.

The socially efficient interest rate, however, is not determined by demand and average cost. Rather, it is set to equate the borrowers' willingness to pay to the lenders' marginal cost. While marginal costs can be recovered directly from the estimated demand and average cost curves, calculating borrowers' willingness to pay is less straightforward. Unlike in insurance markets, where the premiums that consumers pay are sunk prior to the realization of any uncertainty, in credit markets borrowers can default and end up paying less than the quoted interest rate. Because

of this, the demand curve, which captures the maximum *quoted* rate borrowers are willing to accept, does not accurately measure ex-ante willingness to pay. To measure willingness to pay, the demand curve must be adjusted downward to reflect the fact that borrowers who expect to default anticipate not having to pay the full quoted rate.

In markets where adverse selection is the only form of information asymmetry, we show that this adjustment factor is uniquely determined by the lenders' marginal cost curve. The intuition for this result is clear. The incremental cost of providing credit to one additional borrower is equal to that borrower's expected charge-off rate. The expected charge-off rate, in turn, is exactly equal to the portion of the quoted interest payments that the borrower does not expect to pay. Thus, scaling the demand curve downward by the lenders' marginal cost curve provides an estimate of the borrowers' willingness to pay curve. With the demand, marginal cost, and willingness to pay curves in hand, it is then straightforward to calculate both the efficient interest rate and the deadweight loss due to adverse selection.

In markets that feature both adverse selection and moral hazard, we show that this same approach will produce an upper bound for the pooled welfare loss due to both forces. A key distinction between adverse selection and moral hazard is that, under moral hazard, the incremental cost to the lender of providing credit to one additional borrower does not simply reflect that borrower's expected charge-offs. To attract a marginal borrower, the lender must lower its price. Under moral hazard, this change in price will also reduce the default rate of all inframarginal borrowers willing to borrow at the previously higher price. Therefore, the lender's marginal cost at any given price is lower than the expected charge-off rate of the set of borrowers who are indifferent at that price. This means that scaling demand by the lender's marginal costs will over-estimate borrowers' willingness to pay and therefore any loss of consumer surplus resulting from inefficiently high equilibrium pricing. While an upper bound offers less precision than an exact point estimate, such a bound is nonetheless still informative about the potential gains to be achieved by policies that seek to alleviate information asymmetries.

Applying this approach to the data from our experiment in the Chinese fintech lending market, we estimate that the competitive equilibrium interest rate in this market is approximately 30 percent on an annual basis. This falls roughly halfway between the average interest rates offered by the platform in the two experimental treatment arms. Consistent with the presence of information asymmetries, 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, asymmetric information leads to a large equilibrium price distortion.

Despite this apparently large price distortion, our bottom-line estimates of the welfare losses induced by asymmetric information are quite small. The reason for this is that demand is rel-

atively 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 due to asymmetric information 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. If moral hazard is present, this estimate is an upper bound for the true welfare loss, which could be even smaller.

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 effect of the interest rate on the lender's costs is larger among observably-riskier borrowers and therefore that the welfare losses from asymmetric information 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. This makes it difficult to argue strongly in favor of policies like interest rates subsidies, loan guarantees, or expanded creditor recourse in this market on the basis of asymmetric information alone.

Our 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 and moral hazard. Recent examples of this include [Stroebe \(2016\)](#) 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, [Karlan and Zinman \(2009\)](#) on microloans, and [Indarte \(2021\)](#) on consumer bankruptcy. Our paper contributes to this literature by illustrating how similar variation can be used not just to document the presence of asymmetric information, 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 [Chiapori 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 asymmetric information in consumer credit markets. As pointed out by [Einav, Finkelstein and Mahoney \(2021\)](#), one important feature that distinguishes credit markets from other selection markets is that the “price” borrowers promise to pay is not sunk conditional on their product choice. Lenders do not always receive the quoted interest rate, and the quoted interest rate itself may directly affect ex-post default probabilities due to moral hazard. Our paper is the first to explicitly trace out the implications of this distinction for empirical applications of the [Einav, Finkelstein and Cullen \(2010\)](#) framework to credit markets.³

While our results indicate that the welfare losses arising from information asymmetries 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 asymmetric information 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.

³The only other attempt to apply a framework for welfare analysis similar to [Einav, Finkelstein and Cullen \(2010\)](#) to a credit market that we are aware of is in [Lieberman et al. \(2019\)](#), who study how the deletion of default histories in credit records affected welfare in the Chilean consumer credit market. However, in their application, the fact that lenders do not always receive their quoted price is ignored. Our paper highlights how failing to take into account this feature of credit markets will lead to both an incorrect equilibrium pricing condition—lenders should set the *expected* interest rate equal to average costs rather than the *quoted* interest rate—and an incorrect estimate of consumer surplus—the demand curve will over-estimate willingness to pay when borrowers expect to default. In addition, our paper demonstrates how the possibility for moral hazard, which [Lieberman et al. \(2019\)](#) assume away, affects the interpretation of the resulting welfare estimates. Beyond these conceptual contributions, our empirical application is also better suited for estimating borrower demand and lender cost curves given that we can directly observe individual interest rates and charge-offs, which [Lieberman et al. \(2019\)](#) cannot observe and need to impute based on changes in predicted and realized defaults alone.

II CONCEPTUAL FRAMEWORK

This section presents a simple equilibrium model of consumer credit markets under asymmetric information. We begin by considering a baseline model in which adverse selection is the only form of information asymmetry. This baseline setup draws heavily on the framework originally developed by [Einav, Finkelstein and Cullen \(2010\)](#) to study insurance markets. In this setting, we show that exogenous variation in interest rates, along with information on borrower take-up and lender charge-offs are sufficient to estimate the welfare losses arising from asymmetric information. We then modify the baseline framework to incorporate moral hazard. Under both adverse selection and moral hazard we show that the same information can be used to identify an upper bound on the total welfare losses due to asymmetric information.

II.A Baseline Framework: Adverse Selection Only

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 potential 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. The probability that borrower i (with characteristics X_i) defaults on her loan is denoted by $\delta(X_i)$. Similarly, if borrower i defaults on her loan, the expected share of promised payments charged off by the lender is denoted by $\theta(X_i)$. As a baseline, we assume that there is no moral hazard so that both $\delta(X_i)$ and $\theta(X_i)$ are independent of the interest rate. In [Section II.C](#) below we modify this assumption and explore its implications for our results.

Demand for Loans. We denote the expected utility that a potential borrower 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

$$D(r) = \int \mathbb{1}(\rho(X) \geq r) dF(X). \quad (1)$$

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.⁴ The expected profits of each individual lender j are therefore given by

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

where c denotes all 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.⁵ The average expected cost curve facing lenders in the market is given by

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

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 posted interest rate.

The marginal cost curve, defined as the change in total costs when additional borrowers are selected into the market at a marginally lower price, can be similarly expressed as

$$MC(r) = \frac{\partial TC(r)}{\partial D(r)} = \frac{1}{D'(r)} \frac{\partial \int c(X) \mathbb{1}(\rho(X) \geq r) dF(X)}{\partial r} = \mathbb{E}[c(X) | \rho(X) = r]. \quad (4)$$

We say that the market is adversely selected when borrowers who choose to take up a loan at higher posted prices also have higher expected default costs. Under adverse selection, the

⁴We focus on a competitive equilibrium outcome in our analysis for two reasons. First, in the absence of asymmetric information, the competitive equilibrium outcome is socially efficient. This provides a useful benchmark for welfare analysis since any measured losses can be attributed entirely to information asymmetries 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. Consistent with this assumption, our estimate of the competitive equilibrium interest rate is very close to what the lender actually charges in the data. Nonetheless, the framework presented here is easily adapted to other market structures and would simply require the derivation of an alternative equilibrium pricing function. [Mahoney and Weyl \(2017\)](#) discuss the welfare implications of imperfect competition in selection markets in detail and show that the welfare losses from asymmetric information will generally be larger when measured relative to an imperfectly competitive benchmark.

⁵Here 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.

marginal and therefore average cost curve will be increasing in the interest rate (decreasing in quantity).

Rearranging terms in equation (2) yields the following expression for equilibrium interest rates:⁶

$$r = \frac{\int c(X)\mathbb{1}(\rho(X) \geq r)dF(X)}{\int (1 - \delta(X)\theta(X))\mathbb{1}(\rho(X) \geq r)dF(X)} = \frac{AC(r)}{(1+c)-AC(r)} \equiv \widetilde{AC}(r). \quad (5)$$

Throughout, we will refer to the function $\widetilde{AC}(\cdot)$ as the scaled average cost curve. In a competitive equilibrium, lenders set interest rates equal to this scaled version of average costs.⁷ 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 ex-ante willingness to pay for the loan is given by $m^L(X_i) - m^N(X_i) = (1 - \delta(X_i)\theta(X_i))\rho(X_i)L$. Total consumer surplus is thus given by

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

Similarly, producer surplus is equal to aggregate profits

$$PS = L \times \int (r - \delta(X)\theta(X)(1+r) - c)\mathbb{1}(\rho(X) \geq r)dF(X). \quad (7)$$

⁶Strictly 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 $\frac{MC(r)}{\int (1 - \delta(X)\theta(X))\mathbb{1}(\rho(X)=r)dF(X)} = \frac{MC(r)}{(1+c)-MC(r)} < r$ 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 \underline{r} such that $\frac{MC(r)}{(1+c)-MC(r)} > \underline{r}$ then $\frac{MC(r)}{(1+c)-MC(r)} > r$ for all $r > \underline{r}$, which guarantees uniqueness.

⁷One key feature that distinguishes credit markets from other selection markets like insurance 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 $\widetilde{AC}(\cdot)$ curve and is chosen to equate *expected* revenue with *expected* costs.

Total welfare is simply the sum of consumer and producer surplus

$$TS = CS + PS = \int \left[(m^L(X) - c(X)L) \mathbb{1}(\rho(X) \geq r) + m^N(X) \mathbb{1}(\rho(X) < r) \right] dF(X). \quad (8)$$

From equation (8) it is straightforward to see that it is socially efficient for potential borrower i to receive 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 receive 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 and willingness to pay. 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:

$$\begin{aligned} \frac{\partial TS(r)}{\partial D(r)} &= -\frac{1}{D'(r)} \frac{\partial \int (m^L(X) - m^N(X) - c(X)L) \mathbb{1}(\rho(X) \geq r) dF(X)}{\partial r} \\ &= \mathbb{E}[m^L(X) - m^N(X) | \rho(X) = r] - \mathbb{E}[c(X)L | \rho(X) = r]. \end{aligned} \quad (9)$$

The first term in this expression traces out how the average willingness to pay among marginal borrowers varies as a function of the interest rate being charged. Dividing this term by the initial loan principal converts the borrower’s willingness to pay into the same units as the interest rate and gives rise to what we will call the willingness to pay curve

$$WTP(r) = \frac{\mathbb{E}[m^L(X) - m^N(X) | \rho(X) = r]}{L} = \mathbb{E}[1 - \delta(X)\theta(X) | \rho(X) = r] \times r. \quad (10)$$

The second term measures the average dollar-valued cost of serving these marginal borrowers, $\mathbb{E}[c(X)L | \rho(X) = r] = MC(r) \times L$. Thus, under our notion of efficiency, allocating a loan to potential borrower i will increase total surplus if and only if

$$WTP(\rho(X_i)) \geq MC(\rho(X_i)). \quad (11)$$

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

Rearranging terms in equation (11) allows us to express this efficiency condition in terms of the quoted interest rate and marginal cost curves directly. That is, equation (11) is equivalent to

$$\rho(X_i) \geq \frac{MC(\rho(X_i))}{\mathbb{E}[1 - \delta(X)\theta(X) | \rho(X) = \rho(X_i)]} = \frac{MC(\rho(X_i))}{(1+c) - MC(\rho(X_i))} \equiv \widetilde{MC}(\rho(X_i)). \quad (12)$$

As with average costs, we will refer to the function $\widetilde{MC}(\cdot)$ 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 interest rate she is willing to accept is at least as great as the scaled marginal cost of providing her the loan.

II.B Estimating Welfare Losses Under Adverse Selection

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 x-axis measures the share of borrowers in the market (quantity). The y-axis measures the price, cost, and willingness to pay for the loan as a share of the initial loan amount. The scale of these axes is the same in both panels.

In Panel A., the demand curve plots the share of potential borrowers willing to take out a loan at each quoted price r . The defining characteristic of markets featuring adverse selection is that borrowers who are willing to accept higher quoted interest rates will also have higher expected default 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 \widetilde{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 \widetilde{AC} curve and the demand curve at point A gives the equilibrium price r^{EQ} and quantity Q^{EQ} . The intersection of the \widetilde{MC} curve and the demand curve at point B 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 ($Q^{EQ} < Q^{EF}$).

Panel B. illustrates the welfare losses that arise from this underprovision of credit. In this figure, the willingness to pay curve converts the quoted interest rate into the expected willingness to pay among the set of borrowers who are just indifferent between taking out a loan or not at that quoted rate. Because borrowers do not always end up paying the quoted rate, the willingness to pay curve is always beneath the demand curve. Marginal social surplus at any given quantity is measured by the vertical distance between the borrowers' willingness to pay curve and the lenders' marginal cost curve, which is also plotted in the figure. The shaded region DEF gives the deadweight loss of adverse selection. This area is equal to the loss in consumer surplus from borrowers who do not receive a loan in equilibrium but whose willingness to pay exceeds their marginal cost.

Estimating Welfare Losses Using Variation in Prices. As Figure I makes clear, knowledge of the demand curve, cost curves, and willingness to pay curve is sufficient for measuring the welfare loss of adverse selection. The key inputs required to estimate these curves are data on borrower take-up and loan charge-offs and an exogenous source of variation in offered interest rates.

Combining data on take-up and charge-offs with exogenous variation in the offered rate allows for the direct estimation of both the demand curve $D(r)$ and the average cost curve $AC(r)$. The marginal cost curve does not require separate estimation, as it can be constructed directly from these two curves. To see this, note that the marginal cost curve can always be written as

$$MC(r) = \frac{\partial TC(r)}{\partial D(r)} = \frac{\partial (AC(r) \times D(r))}{\partial D(r)} = \frac{1}{D'(r)} \frac{\partial (AC(r) \times D(r))}{\partial r}. \quad (13)$$

Similarly, the willingness to pay curve can be constructed from the marginal cost curve and also does not require separate estimation. To see this, note from the definitions of the marginal cost and willingness to pay curves in equations (4) and (10) that

$$WTP(r) = \mathbb{E}[1 - \delta(X)\theta(X) | \rho(X) = r] \times r = (1 + c - MC(r)) \times r. \quad (14)$$

Finally, with the average and marginal cost curves in hand, it is straightforward to construct their scaled counterparts $\widetilde{AC}(r)$ and $\widetilde{MC}(r)$ using the definitions in equations (5) and (12). These curves together provide sufficient information to calculate the size of the deadweight loss triangle in Figure I. In Sections III and IV we will discuss in detail both the data and the source of variation in interest rates that we use to estimate all of these curves in the market for online consumer loans in China.

II.C Extended Framework: Adverse Selection and Moral Hazard

In the baseline 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.⁸

In this section, we extend the baseline framework to incorporate moral hazard. Under both adverse selection and moral hazard, we show that the estimation approach outlined in [Section II.B](#) no longer identifies the exact welfare losses arising from asymmetric information. Instead, this approach will identify an upper bound on the pooled welfare losses generated by both adverse selection and moral hazard. While an upper bound offers less precision than an exact point estimate, such a bound is nonetheless still informative about the potential welfare gains to be achieved by policies that seek to alleviate information asymmetries. This is particularly true in applications such as ours where the estimated upper bound is small.

Incorporating Moral Hazard. To incorporate moral hazard into the baseline framework, we allow borrower-level costs to explicitly depend on the interest rate: $c(X_i, r) = c + \delta(X_i, r)\theta(X_i, r)$. We say that there is moral hazard when default costs are increasing in the interest rate at the individual level (i.e., when $\partial c(X_i, r)/\partial r > 0$). Under moral hazard, the expression for the average cost curve takes the same basic form as in equation (3) and can be rewritten as

$$AC(r) = \frac{1}{D(r)} \int c(X, r) \mathbb{1}(\rho(X) \geq r) dF(X) = \mathbb{E}[c(X, r) | \rho(X) \geq r]. \quad (15)$$

The expression defining the marginal cost curve also looks similar:

$$MC(r) = \frac{\partial TC(r)}{\partial D(r)} = \frac{1}{D'(r)} \frac{\partial \int c(X, r) \mathbb{1}(\rho(X) \geq r) dF(X)}{\partial r}. \quad (16)$$

Unlike in equation (4), however, this expression no longer simplifies to be equal to the average cost among marginal borrowers. This is because, under both adverse selection and moral hazard, increasing quantity by lowering the price will affect total costs by altering both the pool of bor-

⁸A 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.

rowers who select into the market and the individual-level costs of all borrowers already in the market at the chosen price.

Decomposing the last term in equation (16) into these separate components allows us to rewrite the marginal cost curve as

$$MC(r) = \underbrace{-\frac{1}{D'(r)} \int c(X, r) \mathbb{1}(\rho(X) = r) dF(X)}_{\text{Adverse Selection}} + \underbrace{\frac{1}{D'(r)} \int \frac{\partial c(X, r)}{\partial r} \mathbb{1}(\rho(X) \geq r) dF(X)}_{\text{Moral Hazard}}. \quad (17)$$

The first term, which is equal to $\mathbb{E}[c(X, r) | \rho(X) = r]$, is directly analogous to the marginal cost curve from the baseline case featuring only adverse selection. This term is positive and reflects the fact that lowering the interest rate will increase total costs by drawing in a set of new borrowers each of whom have some positive expected cost. The second term is new and comes about as a result of moral hazard. This term is negative and reflects the fact that lowering the interest rate will decrease total costs by lowering the expected costs of all inframarginal borrowers who were already willing to borrow at the previously higher rate. Thus, the lenders' marginal cost at any given price will be lower than the expected cost of serving the set of borrowers who are indifferent at that price. This is the key distinction between adverse selection and moral hazard.

In the presence of moral hazard, the key insights from the baseline framework hold as long as the marginal cost curve is upward sloping in price (downward sloping in quantity). To see why this is generally the case, observe that the first term in equation (17) is, by definition, increasing in r under either adverse selection or moral hazard. The second term is also likely to be increasing in r .⁹ At high interest rates, the reduction in total costs achieved by lowering the rate is relatively small as there are few inframarginal borrowers in the market at those prices. In contrast, when the interest rate is low, lowering it further will decrease costs among a relatively larger set of inframarginal borrowers. This means that the second term is less negative at high interest rates than it is at lower rates and therefore increasing in r .

Under moral hazard, the expressions for the equilibrium interest rate, consumer surplus, producer surplus, and total surplus are the same as those in section [Section II.A](#), but with the cost function $c(X_i)$ replaced by $c(X_i, r)$. Similarly, the condition determining the constrained efficient allocation in equation (11) is also the same, but with marginal costs defined as in equation (17).¹⁰ However, because of this change to the marginal cost curve, the condition on posted prices

⁹Technically speaking, the sign of the derivative of the second term in equation (17) with respect to the interest rate is indeterminate and depends on the shapes of the demand and borrower-level cost curves. Under linear demand, which is what we will assume in our empirical analysis, a sufficient (but not necessary) condition for this term to be increasing in interest rates is that $c(X_i, r)$ is concave or linear in r .

¹⁰To see this, note that under both adverse selection and moral hazard the derivative of total costs with respect

that implements the efficient outcome needs to be altered. In particular, the inequality in equation (12) can no longer be expressed merely as a function of the marginal cost curve. Instead, the analog to this equation under both moral hazard and adverse selection is given by

$$\rho(X_i) \geq \frac{MC(\rho(X_i))}{(1+c) - \mathbb{E}[c(X, \rho(X)) | \rho(X) = \rho(X_i)]}. \quad (18)$$

In the presence of moral hazard, the last term in the denominator of this expression, $\mathbb{E}[c(X, \rho(X)) | \rho(X) = \rho(X_i)]$, is greater than $\widetilde{MC}(\rho(X_i))$, which means that the right hand side of the inequality is greater than $\widetilde{MC}(\rho(X_i))$.

Estimating Welfare Losses Under both Adverse Selection and Moral Hazard. The same basic approach to estimation outlined in Section II.B can still be applied in the presence of moral hazard. In particular, the demand and average cost curves can still be estimated directly by combining data on take-up and charge-offs with a source of exogenous variation in offered interest rates. The identity in equation (13) also continues to apply, which means that the marginal cost curve can still be constructed from the estimated demand and average cost curves. The key distinction is that, under moral hazard, the marginal cost curve is no longer equal to $\mathbb{E}[c(X, r) | \rho(X) = r]$. This introduces two closely related sources of bias in the resulting welfare estimates.

First, solving for the efficient outcome by equating the demand curve with the scaled marginal cost curve will produce an estimated efficient price (quantity) that is too low (high). Mathematically, this is because the right hand side of equation (18) is greater than $\widetilde{MC}(\rho(X_i))$. Intuitively, this is because the price that implements the efficient outcome is the one that equates the marginal borrowers' willingness to pay, which depends only on their expected default costs, with the lenders' incremental cost of serving those borrowers, which also incorporates the effect of the price on the default costs of inframarginal borrowers. Replacing the right hand side of equation (18) with $\widetilde{MC}(\rho(X_i))$ is equivalent to assuming that when the lender lowers the price to a level that attracts a given set of marginal borrowers, those marginal borrowers also directly value the resulting reduction in inframarginal borrowers' costs. This will over-estimate the marginal borrowers' willingness to pay at any given price and therefore lead to a lower (higher) implied efficient price (quantity). In Figure I, this is equivalent to saying that the \widetilde{MC} curve should be rotated counterclockwise around point C and therefore that point B should lie further up the demand curve and closer to point A.

to quantity can be written as $\frac{\partial TS(r)}{\partial D(r)} = -\frac{1}{D'(r)} \int [(m^L(X) - m^N(X) - c(X, r)L)\mathbb{1}(\rho(X) = r) + \frac{\partial c(X, r)L}{\partial r} \mathbb{1}(\rho(X) \geq r)] dF(X) = \mathbb{E}[m^L(X) - m^N(X) | \rho(X) = r] - MC(r) \times L$, where $MC(r)$ is defined as in equation (17). Thus it is still true that allocating a loan to potential borrower i will increase total surplus if and only if $WTP(\rho(X_i)) \geq MC(\rho(X_i))$.

Second, constructing the willingness to pay curve using equation (14) will produce upwardly biased estimates of consumer surplus. The reason for this is the same. Under moral hazard, scaling the interest rate by $(1+c-MC(r))$ will over-estimate the marginal borrowers' willingness to pay since $MC(r) < \mathbb{E}[c(X, r) \mid \rho(X) = r]$. In Figure I, this is equivalent to saying that the willingness to pay curve should be rotated counter-clockwise around point G in panel B and therefore that the vertical difference between the willingness to pay and marginal cost curves should be smaller at all quantities less than the efficient quantity.

Taken together, these two sources of bias imply that the area of the deadweight loss triangle as calculated following the estimation approach outlined in Section II.B will be too large relative to its true value. Thus, while this approach to estimation produces exact welfare estimates under adverse selection alone, in the presence of both adverse selection and moral hazard, it will instead provide an upper bound on the true welfare losses arising from asymmetric information. The key empirical challenge is that it is not possible to separately estimate both terms in equation (17) using only ex-ante variation in offered interest rates. This is the classic difficulty pointed out by Karlan and Zinman (2009) that arises any time one attempts to empirically separate adverse selection and moral hazard. Intuitively, this challenge arises because the second term in equation (17), which measures the effect of the interest rate on the default costs of inframarginal borrowers, can only be estimated using variation in interest rates that occurs *after* borrowers have selected into the market. Unfortunately, such variation is not typically available in most empirical settings. Nonetheless, there are many applications, both experimental and quasi-experimental, that feature exogenous variation in ex-ante offered rates. In these settings, our approach will always identify an upper bound on the pooled welfare losses arising from both adverse selection and moral hazard. In the next section, we discuss the data and source of variation in interest rates we use to estimate this upper bound in the market for online consumer loans in China.

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 United States and accounted for 38 percent of total lending in that segment as of 2018.¹¹

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 that is akin to a standard risk premium. It 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

¹¹Source: <https://newsroom.transunion.com/fintechs-continue-to-drive-personal-loans-to-record-levels>.

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 risk premium 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.¹² 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 risk premium 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.¹³ 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 (1–4, 1 being the best) 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

¹²Note that to empirically estimate the borrower demand curve, one must first define the relevant market. Broadly speaking, we view the relevant market to be the set of borrowers who are potentially interested in taking out a fintech loan in China at interest rates within the range of those offered in our experiment. We proxy for borrower interest using the fact that all borrowers in the experiment chose to apply for credit with our lender prior to knowing the price that they would be offered. Broadening the market definition to include borrowers who did not apply would likely lead to a lower estimated demand elasticity as doing so would include many borrowers who would never take a loan at any plausible interest rate. In our empirical analysis, we find small overall welfare losses driven by relatively inelastic borrower demand. Using a broader market definition would thus likely lead to qualitatively similar conclusions with even smaller estimated welfare losses.

¹³The fact that we study an experimental price cut rather than a price increase alleviates concerns regarding borrowers moving to another lender since the treatment group receives a price that is likely not on offer at other lenders.

groups.¹⁴ 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, the 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).¹⁵ 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

The key outcome variables required to estimate the welfare loss arising from asymmetric information 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 show that the marginal cost curve $MC(r)$ and willingness to pay curve $WTP(r)$ can both be constructed directly from the demand and average cost curves and therefore do not require separate estimation. Similarly, the scaled average cost curve $\widetilde{AC}(r)$ and scaled marginal cost curve $\widetilde{MC}(r)$ can be derived directly from their unscaled counterparts. As discussed in Section II, knowledge of the demand curve, cost curves, and willingness to pay curve is sufficient for calculating an upper bound on the total welfare losses due to asymmetric information. This section describes our approach to estimating these curves 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,

¹⁴In 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.

¹⁵In Appendix Figure A.1 we show that not only the mean, but the full distribution of these variables is similar across the two groups.

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.¹⁶ 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.

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

$$d_i = \alpha_d + \beta_d r_i + \epsilon_i \quad (19)$$

$$c_i = \alpha_c + \beta_c r_i + v_i. \quad (20)$$

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 subset 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 (19) and (20). To see this, note from the definition of $\widetilde{AC}(r)$ in equation (5) that

$$\widetilde{AC}(r) = \frac{\alpha_c + \beta_c r}{(1+c) - (\alpha_c + \beta_c r)}, \quad (21)$$

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

$$MC(r) = \frac{\partial TC(r)}{\partial D(r)} = \frac{\partial (AC(r) \times D(r))}{\partial D(r)} = \frac{\alpha_d \beta_c}{\beta_d} + \alpha_c + 2\beta_c r, \quad (22)$$

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

¹⁶Specifically, 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.

marginal cost curve can then be written as

$$\widetilde{MC}(r) = \frac{\alpha_m + \beta_m r}{(1+c) - (\alpha_m + \beta_m r)}. \quad (23)$$

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 $\widetilde{AC}(r) = D^{-1}(r)$ allows us to solve for this point analytically, yielding

$$r^{EQ} = \frac{\phi_c - \sqrt{\psi_c}}{2\beta_c} \quad \text{and} \quad Q^{EQ} = \alpha_d + \beta_d r^{EQ}, \quad (24)$$

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 $\widetilde{MC}(r) = D^{-1}(r)$, yielding

$$r^{EF} = \frac{\phi_m - \sqrt{\psi_m}}{2\beta_m} \quad \text{and} \quad Q^{EF} = \alpha_d + \beta_d r^{EF}, \quad (25)$$

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 asymmetric information is given by the area underneath the borrowers' willingness to pay curve and above the lenders' marginal cost curve between the equilibrium and efficient quantities. This area is represented by the shaded region between points D, E and F in [Figure I](#). Calculating this area requires an estimate of the borrowers' willingness to pay curve. As discussed in [Section II](#), this curve can be estimated by converting each maximum acceptable quoted interest rate along the borrowers' demand curve into an ex-ante willingness to pay using the expression in [equation \(14\)](#)

$$WTP(r) = (1 + c - MC(r)) \times r = (1 + c - \alpha_m)r - \beta_m r^2. \quad (26)$$

When the willingness to pay curve is roughly linear, this area can be well-approximated using a standard [Harberger \(1964\)](#) triangle formula:

$$DWL \approx \frac{1}{2}(Q^{EF} - Q^{EQ})(WTP(r^{EQ}) - MC(r^{EQ})). \quad (27)$$

The third term in this expression highlights the core distinction between insurance and credit markets. In insurance markets, borrowers always pay the quoted premium. This means that the

height of the deadweight loss triangle can be measured as in [Einav, Finkelstein and Cullen \(2010\)](#) using the distance between the equilibrium price and the lenders' marginal cost at that price. In credit markets, however, borrowers can default and therefore do not always end up actually paying the quoted interest rate. This means that the equilibrium quoted interest rate does not accurately measure borrowers' willingness to pay. The third term in equation (27) incorporates this fact by converting the equilibrium price into a measure of ex-ante willingness to pay before calculating the height of the triangle.

While the approximation in equation (27) is useful and intuitive, we could also calculate the area corresponding to the deadweight loss exactly by integrating between the willingness to pay and marginal cost curves over the range $[Q^{EQ}, Q^{EF}]$ using the expressions for those curves given by equations (22) and (26). In [Section V](#) below, we provide estimates of the welfare loss calculated using both approaches 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 asymmetric information 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 $(r^{EQ} - \bar{r})^2 + (Q^{EQ} - \bar{Q})^2$, where \bar{r} and \bar{Q} are the average interest rate and take-up rate across all loan applicants in the sample.¹⁷ To calculate standard errors for the demand and cost curves

¹⁷Since our data come from an experiment in which the lender is actively changing prices, it is not clear which

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.

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 asymmetric information. 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 asymmetric information.

V.A Main Estimates

[Table II](#) presents our main results.¹⁸ In Panel A., we report estimates of the demand curve from various versions of the basic regression specification given by equation (19). 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

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.

¹⁸The 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 ([Stock and Yogo, 2005](#); [Lee et al., 2020](#)).

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.¹⁹ 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. Across all specifications, the point estimates are nearly identical to those obtained in column (1). The consistency of these results across specifications echos the finding of near complete covariate balance from [Section III](#).

Panel B. reports analogous results for the average cost curve obtained by estimating equation (20). 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 relationship could arise either because riskier borrowers are willing to accept higher interest rates—adverse selection—or because higher interest rates increase the default rate of all individual borrowers—moral hazard. Regardless of its source, this positive coefficient indicates the presence of asymmetric information. 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 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.²⁰

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](#)

¹⁹All control variables are demeaned prior to estimation to ensure that the intercept term can be interpreted similarly across specifications.

²⁰The 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.

using the coefficient estimates from Panels A. and B. The results indicate that asymmetric information 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 .²¹

The bottom two rows of the table report estimates of the implied welfare loss arising from competitive pricing in the presence of asymmetric information. We construct these estimates in two ways. First, we use the approximation formula from equation (27). Second, we calculate the exact welfare loss by integrating between the willingness to pay and 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 asymmetric information 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.²² Given the average loan size reported in Table I, 0.8 percent equates to a welfare loss of only ¥50, or approximately \$7.20 per applicant. The 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 asymmetric information generates only modest quantity distortions and therefore small overall welfare losses. If moral hazard is present, then these numbers should be interpreted as upper bounds for the true welfare loss of asymmetric information, which could be even smaller.

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 in both panels measures the price, cost, or willingness to pay for the loan and the x-axis measures the share of borrowers who take-up the loan. Panel A. illustrates the effect of asymmetric information on the equilibrium allocation of credit. The solid blue line plots our estimated demand curve. The dashed and solid orange lines plot the scaled average and marginal cost curves, respectively.²³ The equilibrium and efficient outcomes, which are determined by the intersection of the demand curve with the scaled average

²¹We calculate the interest rate elasticity at the midpoint between the equilibrium and efficient allocations, yielding $\frac{(0.703-0.608)/[(0.703+0.608)/2]}{(0.085-0.304)/[(0.085+0.304)/2]} = -0.13$

²²Table 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.

²³In 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.

and marginal cost curves, are also indicated in the figure. Panel B. illustrates the welfare losses that arise from the underprovision of credit due to asymmetric information. In this panel, the solid blue line plots the borrowers' willingness to pay curve and the solid orange line plots the lenders marginal cost curve. The shaded region measures the welfare loss due to asymmetric information. Because the willingness to pay curve is nearly linear between the equilibrium and efficient quantities, 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 (27).

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 equilibrium and efficient outcomes. 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 efficient outcome and therefore also the area under the willingness to pay curve and above the marginal cost curve between the efficient and equilibrium quantities. 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 the best rating—rating category 1. 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 2–4. 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 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 observed charge-off rate among 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 either the correlation between demand and unobservable lending costs is higher among observably higher-risk borrowers or that raising the interest rate increases default rates at the individual-borrower level more for borrowers who are observably riskier. As a result of this, the welfare losses of asymmetric information should also be larger for the higher-risk 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 1.4–1.8 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 asymmetric information remains 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 asymmetric information in the new and growing fintech consumer credit market. Leveraging a randomized experiment conducted by a popular Chinese lending platform, we document that asymmetric information 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 at most 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, loan guarantees, or expanded creditor recourse to address information asymmetries in this

market.

While we find small welfare losses from asymmetric information in the specific 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 more broadly. The same methodology, perhaps with slight modifications, could also be applied to measure or bound 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 information asymmetries generate large or small welfare losses. Beyond measuring the welfare consequences of asymmetric information 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

- Adams, William, Liran Einav, and Jonathan Levin. 2009. "Liquidity Constraints and Imperfect Information in Subprime Lending." *American Economic Review*, 99(1): 49–84.
- Agarwal, Sumit, Souphala Chomsisengphet, and Chunlin Liu. 2010. "The Importance of Adverse Selection in the Credit Card Market: Evidence from Randomized Trials of Credit Card Solicitations." *Journal of Money, Credit and Banking*, 42(4): 743–754.
- Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel. 2018. "Do Banks Pass through Credit Expansions to Consumers Who want to Borrow?" *Quarterly Journal of Economics*, 133(1): 129–190.
- Akerlof, George A. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84(3): 488–500.
- Ausubel, Lawrence M. 1999. "Adverse Selection in the Credit Card Market." Working Paper.
- Chiappori, Pierre-André, and Bernard Salanie. 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy*, 108(1): 56–78.
- Cornelli, Giulio, Jon Frost, Leonardo Gambacorta, Raghavendra Rau, Robert Wardrop, and Tania Ziegler. 2020. "Fintech and Big Tech Credit: A New Database." BIS Working Papers No. 887.
- Dobbie, Will, and Paige Marta Skiba. 2013. "Information Asymmetries in Consumer Credit Markets: Evidence from Payday Lending." *American Economic Journal: Applied Economics*, 5(4): 256–282.
- Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *Quarterly Journal of Economics*, 125(3): 877–921.
- Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2021. "The IO of Selection Markets." NBER Working Paper No. 29039.
- Einav, Liran, and Amy Finkelstein. 2011. "Selection in Insurance Markets: Theory and Empirics in Pictures." *Journal of Economic Perspectives*, 25(1): 115–138.
- Einav, Liran, Mark Jenkins, and Jonathan Levin. 2012. "Contract Pricing in Consumer Credit Markets." *Econometrica*, 80(4): 1387–1432.
- Gupta, Arpit, and Christopher Hansman. 2019. "Selection, Leverage, and Default in the Mortgage Market." Working Paper.
- Harberger, Arnold C. 1964. "The Measurement of Waste." *American Economic Review*, 54(3): 58–76.
- Hertzberg, Andrew, Andres Liberman, and Daniel Paravisini. 2018. "Screening on Loan Terms: Evidence from Maturity Choice in Consumer Credit." *Review of Financial Studies*, 31(9): 3532–3567.

- Indarte, Sasha.** 2021. “Moral Hazard versus Liquidity in Household Bankruptcy.” Working Paper.
- Jaffee, Dwight M., and Thomas Russell.** 1976. “Imperfect Information, Uncertainty, and Credit Rationing.” *Quarterly Journal of Economics*, 90(4): 651–666.
- Karlan, Dean, and Jonathan Zinman.** 2009. “Observing the Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment.” *Econometrica*, 77(6): 1993–2008.
- Lee, David S., Marcelo J. Moreira, Justin McCrary, and Jack Porter.** 2020. “Valid t -ratio Inference for IV.” arXiv:2010.05058v1 [econ.EM].
- Liberman, Andres, Christopher A. Neilson, Luis Opazo, and Seth Zimmerman.** 2019. “The Equilibrium Effects of Information Deletion: Evidence from Consumer Credit Markets.” Working Paper.
- Mahoney, Neale, and Glen Weyl.** 2017. “Imperfect Competition in Selection Markets.” *The Review of Economics and Statistics*, 99(4): 637–651.
- Philippon, Thomas.** 2016. “The Fintech Opportunity.” NBER Working Paper No. 22476.
- Stiglitz, Joseph E., and Andrew Weiss.** 1981. “Credit Rationing in Markets with Imperfect Information.” *American Economic Review*, 71(3): 393–410.
- Stock, James H., and Motohiro Yogo.** 2005. “Testing for Weak Instruments in Linear IV Regressions.” In *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, ed. James H. Stock and Donald Andrews, Chapter 5. Cambridge University Press.
- Stroebel, Johannes.** 2016. “Asymmetric Information about Collateral Values.” *Journal of Finance*, 71(3): 1071–1112.

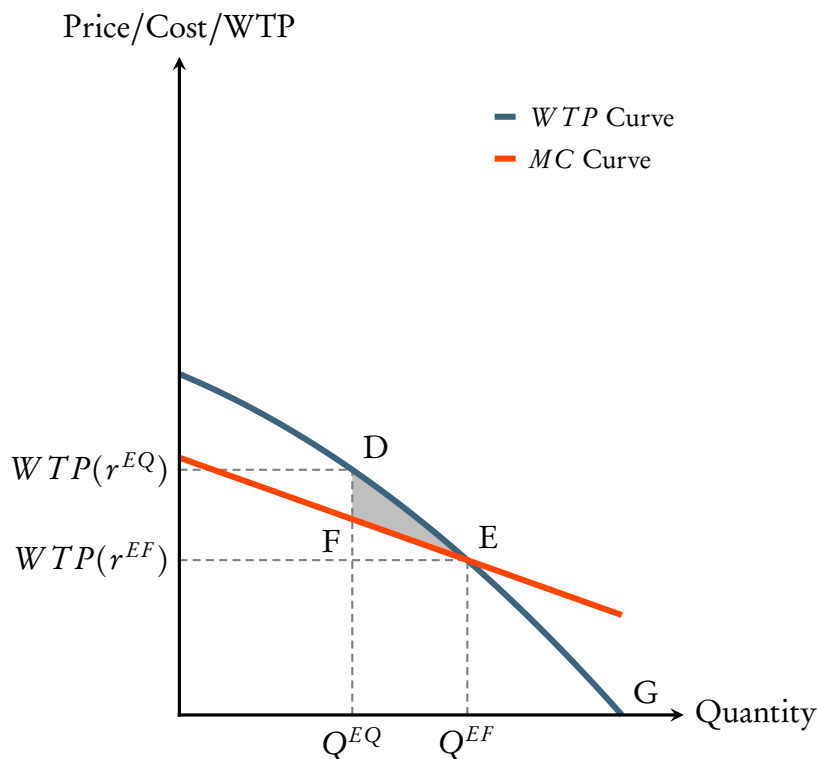
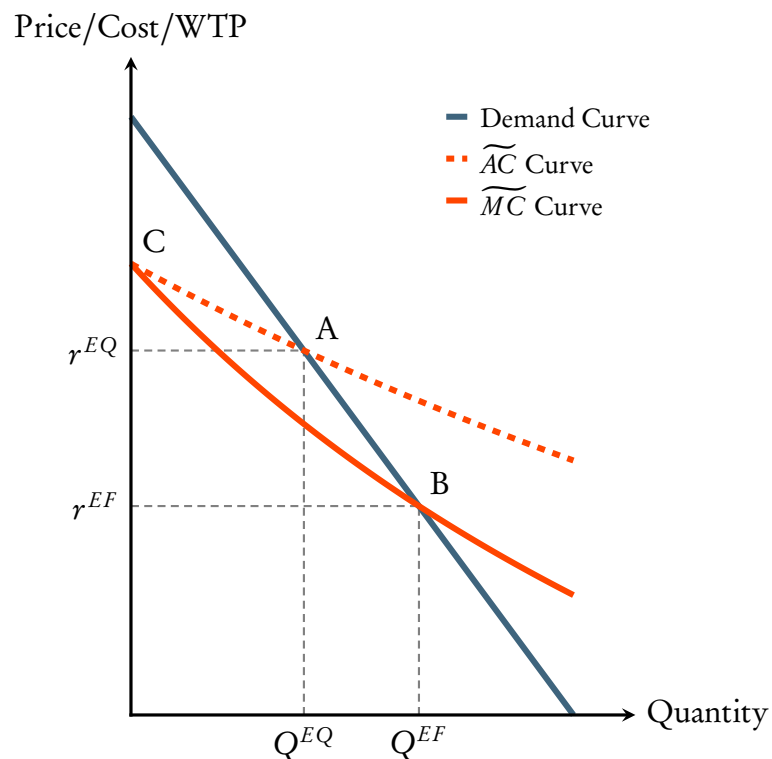


FIGURE I
Welfare Cost of Adverse Selection

NOTE.— This figure provides a graphical illustration of the welfare cost of adverse selection. In both panels, the x-axis measures the share of potential borrowers in the market and the y-axis measures the price, cost, or willingness to pay for the loan as a share of the initial loan amount. The scale of these axes is the same in both panels. Panel A. illustrates the effect of adverse selection on the equilibrium market outcome under perfect competition. The market depicted in the figure features adverse selection because the scaled marginal cost curve is downward-sloping. Borrowers who select into the market at the highest posted interest rates are also those with the highest expected costs. The competitive equilibrium is determined by the intersection of the demand and \widetilde{AC} curves (point A). The efficient allocation is determined by the intersection of the demand and \widetilde{MC} curves (point B). Under perfect competition, adverse selection leads to inefficiently high equilibrium pricing and an underprovision of credit. Panel B. illustrates the welfare losses that arise from this inefficiently high pricing. The welfare loss can be calculated as the difference between consumers' willingness to pay and lenders' marginal cost summed across all consumers who are priced out of the market due to adverse selection. The shaded region (DEF) depicts these losses.

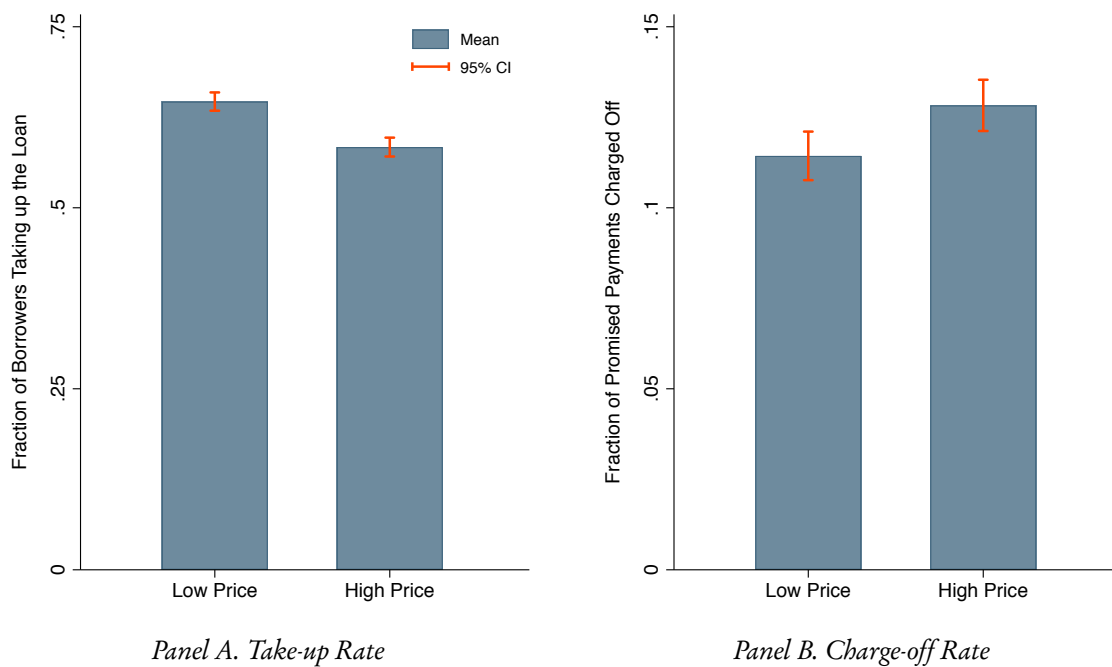
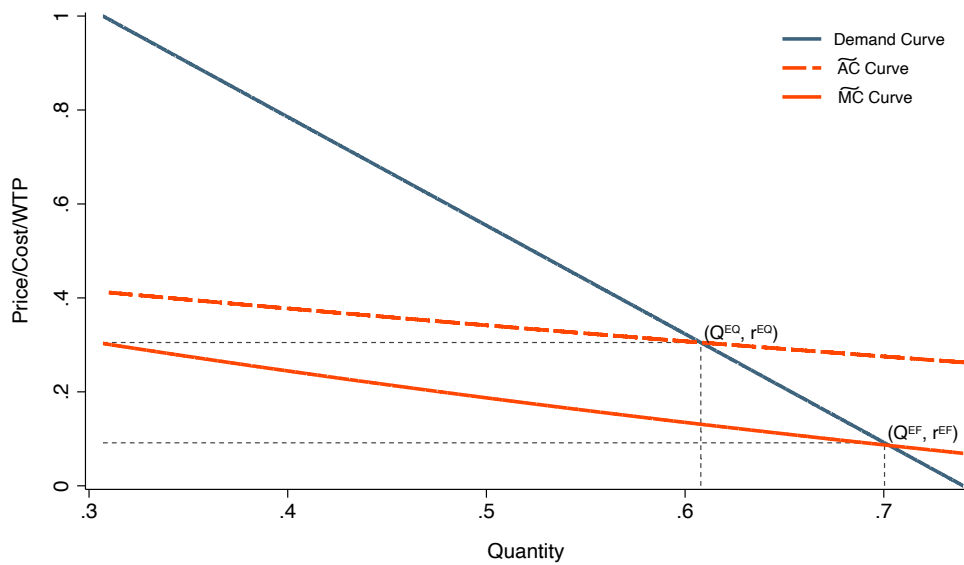
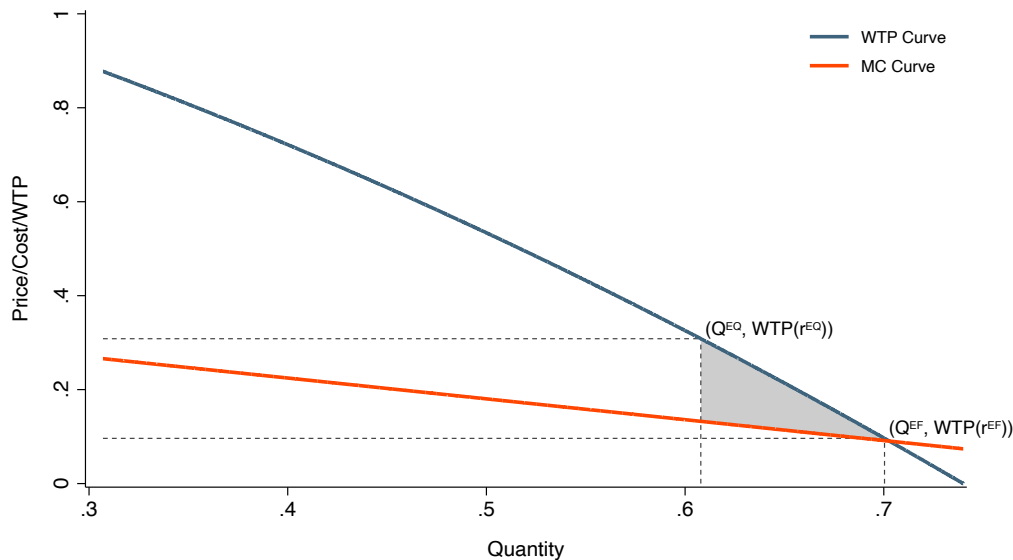


FIGURE II
Mean Take-up and Charge-off Rates by Treatment Arm

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 among 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.



Panel A. Equilibrium and Efficient Allocations



Panel B. Welfare Losses

FIGURE III
Empirical Estimates of Welfare Cost of Asymmetric Information

NOTE.—This figure presents an empirical analog to Figure I. In both panels, the x-axis measures the share of potential borrowers in the market and the y-axis measures the price, cost, or willingness to pay for the loan as a share of the initial loan amount. The curves plotted in this figure are derived from the point estimates in column 1 of Table II as described in Section IV. Panel A. illustrates the effect of asymmetric information on the equilibrium market outcome under perfect competition. The competitive equilibrium is determined by the intersection of the demand and \widetilde{AC} curves. The efficient allocation is determined by the intersection of the demand and \widetilde{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. Panel B. illustrates the welfare losses that arise from asymmetric information. The welfare loss can be calculated as the difference between consumers' willingness to pay and lenders' marginal cost summed across all consumers who are priced out of the market due to asymmetric information. The shaded region depicts these losses.

TABLE I
SUMMARY STATISTICS AND COVARIATE BALANCE TEST

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

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

	(1)	(2)	(3)	(4)	(5)
<i>Panel A. Demand</i>					
Interest Rate	-0.433*** (0.064)	-0.429*** (0.063)	-0.430*** (0.063)	-0.425*** (0.063)	-0.425*** (0.066)
Constant	0.740*** (0.019)	0.739*** (0.019)	0.739*** (0.019)	0.737*** (0.019)	0.737*** (0.020)
Number of Observations	10,991	10,991	10,991	10,991	10,991
<i>Panel B. Average Cost</i>					
Interest Rate	0.096** (0.044)	0.094** (0.044)	0.093** (0.044)	0.090** (0.043)	0.090* (0.049)
Constant	0.238*** (0.013)	0.238*** (0.013)	0.238*** (0.013)	0.239*** (0.013)	0.227*** (0.014)
Number of Observations	6,761	6,761	6,761	6,761	6,761
Demographics		X	X	X	X
Geography			X	X	X
Loan Size and Rating				X	X
Estimated Fixed Cost					X
<i>Panel C. Implied Quantities</i>					
Equilibrium Price	0.304	0.304	0.304	0.304	0.288
Equilibrium Quantity	0.608	0.608	0.608	0.608	0.615
Efficient Price	0.085	0.088	0.091	0.095	0.081
Efficient Quantity	0.703	0.701	0.700	0.697	0.703
Welfare Loss (per ¥100): Approximate	0.835	0.806	0.787	0.750	0.743
Welfare Loss (per ¥100): Exact	0.849	0.820	0.800	0.762	0.754

NOTE.—This table reports estimates of the demand curve, average cost curve, and implied welfare loss of asymmetric information. 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

	Category 1		Category 2-4	
	(1)	(2)	(3)	(4)
<i>Panel A. Demand</i>				
Interest Rate	-0.448*** (0.088)	-0.448*** (0.088)	-0.417*** (0.091)	-0.402*** (0.091)
Constant	0.746*** (0.026)	0.778*** (0.026)	0.733*** (0.028)	0.722*** (0.029)
Number of Observations	5,543	5,543	5,448	5,448
<i>Panel B. Average Cost</i>				
Interest Rate	0.056 (0.056)	0.060 (0.056)	0.133** (0.068)	0.122* (0.067)
Constant	0.230*** (0.016)	0.227*** (0.016)	0.247*** (0.020)	0.237*** (0.021)
Number of Observations	3,432	3,432	3,329	3,329
Demographics		X		X
Geography		X		X
Loan Size		X		X
<i>Panel C. Implied Quantities</i>				
Equilibrium Price	0.272	0.270	0.344	0.317
Equilibrium Quantity	0.624	0.657	0.590	0.594
Efficient Price	0.156	0.140	0.016	0.019
Efficient Quantity	0.676	0.715	0.727	0.714
Welfare Loss (per ¥100): Approximate	0.256	0.325	1.725	1.425
Welfare Loss (per ¥100): Exact	0.258	0.327	1.790	1.468

NOTE.—This table reports estimates of the demand curve, average cost curve, and implied welfare loss of asymmetric information 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 the best credit rating (rating 1), whereas columns (3) and (4) include only those assigned ratings of 2, 3, or 4. 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 of 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.¹ 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](#).

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.

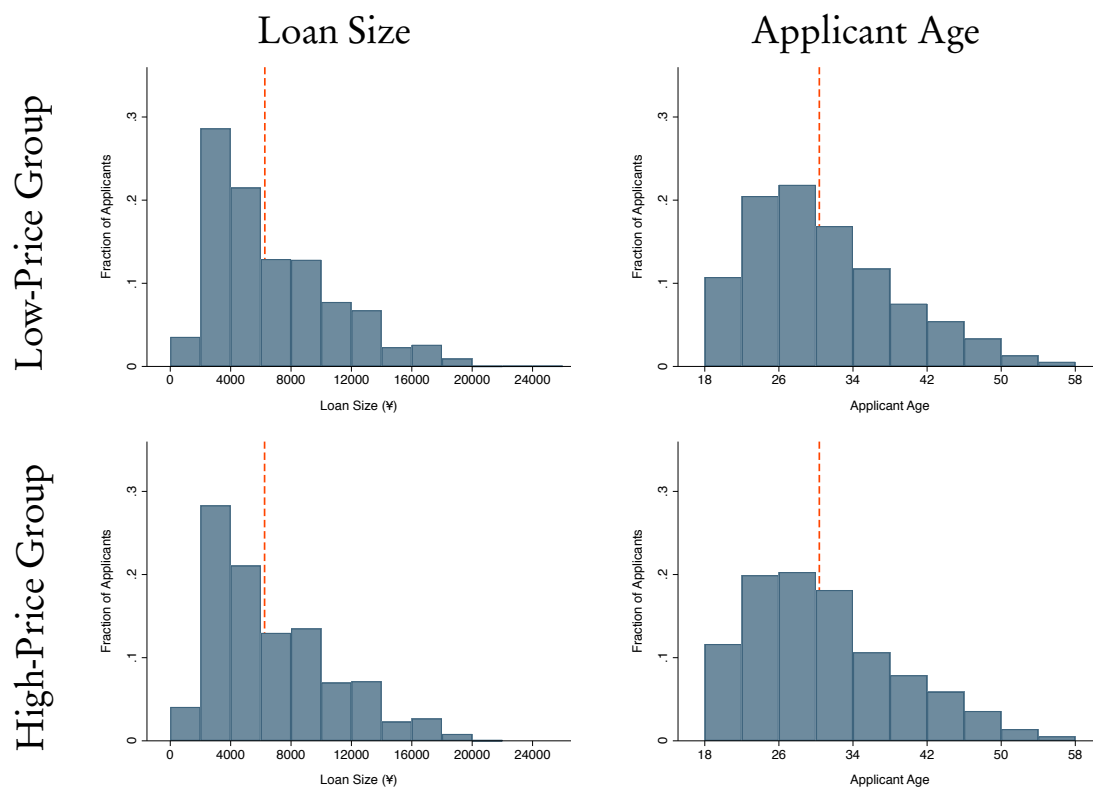


FIGURE A.1
Distributions of Loan Size and Applicant Age by Treatment Arm

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.

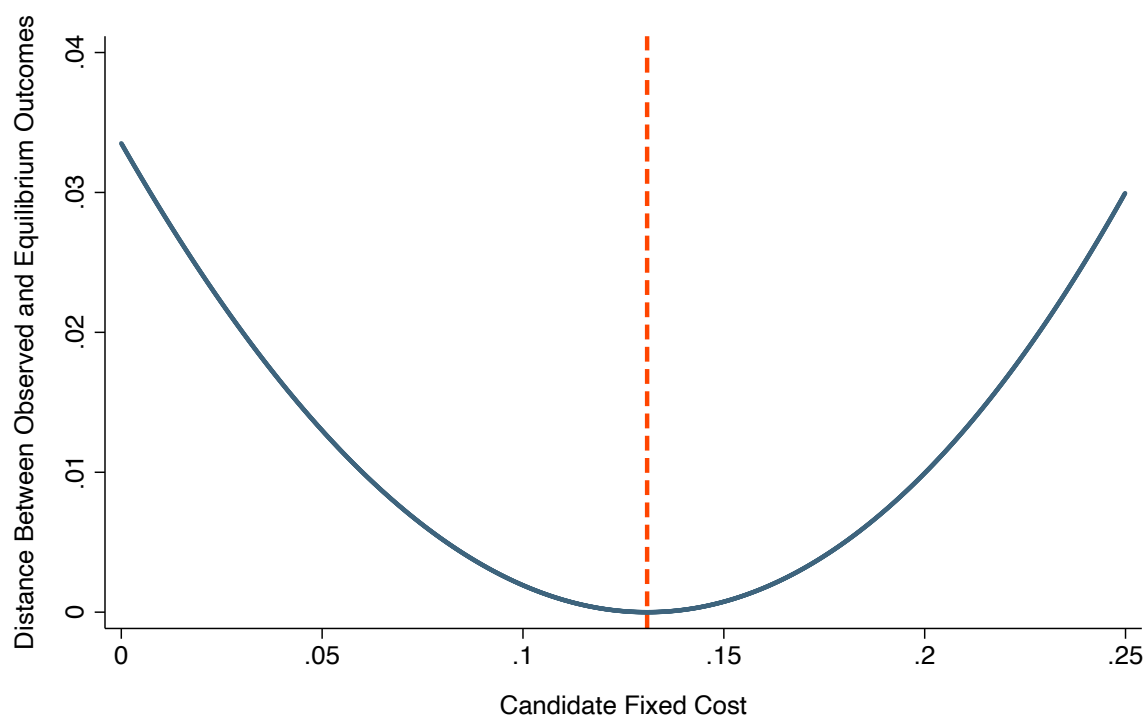


FIGURE A.2
Objective Function Values for Fixed Cost Estimation

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).

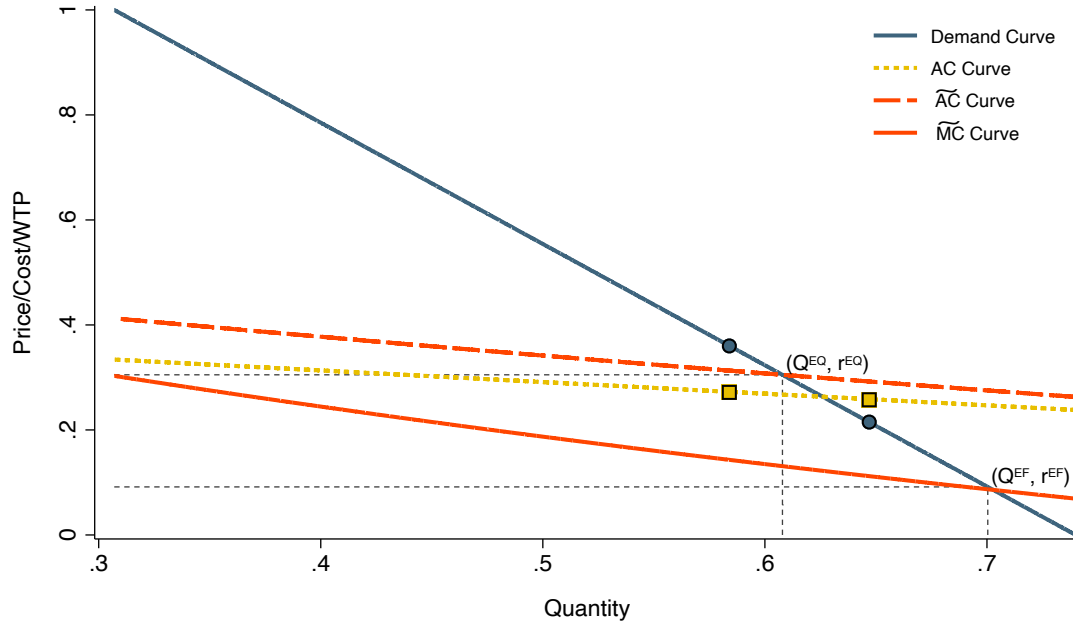


FIGURE A.3
Equilibrium and Efficient Allocations Under Asymmetric Information

NOTE.—This figure presents an extended version of Panel A. of [Figure III](#). The x-axis measures the share of potential borrowers in the market. The y-axis measures price, cost, or willingness to pay for the loan as a share of the initial loan amount. 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 \widetilde{AC} and \widetilde{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 \widetilde{AC} curves. The efficient allocation is determined by the intersection of the demand and \widetilde{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.

TABLE A.1
LOAN AND BORROWER REPRESENTATIVENESS

	Our Sample	Industry Peers
<i>Loan Characteristics</i>		
Loan Size (¥)	6,255	18,000
Maturity (months)	12	12.65
Investment Return (%)	10.83	9.81
<i>Borrower Characteristics</i>		
Age Under 40 (%)	85	83
Male (%)	77	65
Single (%)	51	53

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/>. For example, statistics on loan characteristics are from <https://www.wdzj.com/zhuanti/2018report/> and borrower characteristics are from <https://www.wdzj.com/news/yc/2326325.html>.

TABLE A.2
FIRST STAGE ESTIMATES FOR DEMAND CURVE

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

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

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

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

	(1)	(2)	(3)	(4)	(5)
Equilibrium Price	0.304 [0.294, 0.317]	0.304 [0.294, 0.317]	0.304 [0.294, 0.316]	0.304 [0.294, 0.317]	0.288 [0.278, 0.299]
Equilibrium Quantity	0.608 [0.597, 0.618]	0.608 [0.598, 0.618]	0.608 [0.598, 0.618]	0.608 [0.597, 0.618]	0.615 [0.605, 0.625]
Efficient Price	0.085 [-0.131, 0.286]	0.088 [-0.133, 0.294]	0.091 [-0.128, 0.294]	0.095 [-0.129, 0.299]	0.081 [-0.141, 0.285]
Efficient Quantity	0.703 [0.614, 0.781]	0.701 [0.612, 0.782]	0.700 [0.612, 0.780]	0.697 [0.608, 0.777]	0.703 [0.615, 0.783]
Welfare Loss (per ¥100): Approximate	0.835 [0.025, 2.684]	0.806 [0.012, 2.647]	0.787 [0.009, 2.625]	0.750 [0.004, 2.555]	0.743 [0.005, 2.547]
Welfare Loss (per ¥100): Exact	0.849 [0.006, 3.306]	0.820 [0.001, 3.291]	0.800 [0.001, 3.202]	0.762 [0.000, 3.127]	0.754 [0.000, 3.066]
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 II](#). 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.