Asymmetric Information and Imperfect Competition: Evidence from the Korean Personal Loan Market

Meeroo Kim and Jangsu Yoon*

June 7, 2022

Abstract

This paper studies the effects of asymmetric information and oligopolistic competition between lenders on the personal loan market in South Korea. Focusing on personal loan contracts between banks and individual consumers, we establish an empirical model of the personal loan market considering potential adverse selection and moral hazard problems. Our model incorporates bank lending capacity with the bank's loan pricing and information asymmetries, specifying a dual screening device mechanism under which the banks endogenously determine both loan price and lending limit. Based on the Korea Credit Bureau (KCB) dataset in South Korea, we construct a structural model to describe loan demand, default, pricing, and lending limit, thereby quantitatively measuring the degrees of adverse selection and moral hazard. The estimated model finds empirical evidence of asymmetric information in the sense that unobserved components of the loan demand are positively correlated with the default. The propensity to default is also positively correlated to the loan price and the loan amount. The counterfactual analysis verifies that less asymmetric information and a more competitive lending market will provide lower loan prices and higher lending limits for loan takers but may increase the loan default rate.

Keywords: Asymmetric Information, Adverse Selection, Imperfect Competition
JEL Classification: D82, G21, L13, L22

*Meeroo Kim: Department of Macroeconomics and Financial Policies, Korea Development Institute, 263 Namsejongro, Sejong, 30149, Republic of Korea, (email: mrkim@kdi.re.kr); Jangsu Yoon: Department of Economics, University of Wisconsin-Milwaukee, 3210 North Maryland Avenue, Bolton Hall #882, Milwaukee, WI 53211, (email: yoon22@uwm.edu).

1 Introduction

1.1 Motivation

This paper provides an empirical study regarding how the asymmetric information and oligopolistic market structure jointly affect the personal loan market in South Korea. The principal-agent problem caused by information asymmetry and the corresponding welfare loss from market failure is well-known in many theoretical works of literature (e.g., Akerlof (1970), Rothchild and Stiglitz (1978), Azevedo and Gottlieb (2017), Mahoney and Weyl (2017), Lester et al. (2019)). So far, many articles have developed empirical analyses to quantify the impact of asymmetric information on consumer behavior in the selection markets. The empirical research about the selection market focused on verifying the existence of asymmetric information or measuring the size of inefficiency caused by asymmetric information (e.g., Chiappori and Salanié (2000), Karlan and Zinman (2009), Einav et al. (2012)). Our study contributes to the empirical literature by finding evidence of substantial adverse selection in the personal loan market in South Korea.

Our research is distinguishable from the previous literature by considering lenders (i.e., banks) with market power. Although many previous works pay attention to the loan markets as a representative selection market, there are few empirical applications considering both asymmetric information and imperfect competition of lenders (e.g., Crawford et al. (2018), Ioannidou et al. (2021)). Less empirical research is due to a lack of a qualified dataset and difficulty designing an econometric model that identifies the parameter of asymmetric information. Regarding information asymmetries, researchers face concerns identifying the effect of personal riskiness on the contract terms since the borrowers' hidden riskiness is even unobservable to lenders. Under the imperfect competition, a critical empirical issue is that each lender's optimal loan contract terms to the same customer may differ. Since we observe only the realized loan contract terms, the researcher must consider how to identify the counterfactual loan contract terms suggested by "unselected" lenders.

We overcome the challenges using a novel large-scale Korea Credit Bureau (KCB) dataset containing the consumer's detailed personal characteristics and the list of personal loan history. The dataset is a random sample containing 10% of the entire population in South Korea. Based on the sample size extracted from the whole population, the structural estimation investigated in the current paper is free from any representativeness issue. Furthermore, with a large sample size and diverse loan demand-supply relevant variables, we can check the robustness of our estimation results, allowing for flexible model specifications. The dataset enables us to identify the borrower's unobserved heterogeneity using the sample with multiple loan contracts. We can compute the counterfactual loan contract terms utilizing many observable predictors based on the borrower's income, credit score, and consumption patterns. Based on the dataset, our paper provides more abundant and robust empirical implications on the personal loan market with selection.

Our paper contributes to the literature in several ways. First, to the best of our knowledge, our study is the first empirical analysis on the consumer credit market considering (1) asymmetric information between borrowers and lenders: adverse selection and moral hazard, (2) interactive two-dimensional screening devices, and (3) oligopolistic competition between lenders. Based on the personal loan market data in South Korea, we construct a structural model to describe loan demand, default decision of consumers, loan pricing, and the bank's lending limit that regulates the maximum amount of a personal loan. The structural parameters in the model quantitatively measure the empirical evidence of adverse selection and moral hazard and enable a counterfactual analysis to predict the welfare outcome from policy changes in the financial market. The estimation results reveal the existence of asymmetric information because the unobserved component of loan demand is positively correlated with the default decision (adverse selection), and the default probability is higher as the remaining loan amount is more massive (moral hazard).

Second, our findings particularly highlight the role of the loan amount as a screening device. A key motivation of our paper is to investigate the information asymmetry under the interaction of two screening devices: loan price and loan amount. The loan price has been a primary source of screening device in the theoretical and empirical literature of asymmetric information (Stiglitz and Weiss (1981), Bester (1985), Chiappori and Salanié (2000), Einav et al. (2012), Crawford et al. (2018)). The primary mechanism of Stiglitz and Weiss (1981) predicts that a high-risk borrower who expects a higher return will likely accept a high-interest rate loan, while a low-risk borrower would not. We show that the structural model with a loan interest rate but without a loan amount channel may underestimate the magnitude of the loan interest rate effect. For example, our estimation predicts a 0.3% increase in loan demand and a 0.09%P (14.39%) increase in loan default rate with respect to loan price and lending limit increases by 0.1 standard deviations. However, the conventional model specification without endogenous loan amount predicts a 0.01% decrease in loan demand and a 0.05%P (8.53%) increase in loan default rate to the same changes. A bank with a high lending capacity can increase the loan price as a higher loan amount can compensate for the consumer disutility from the higher interest rate. Our estimation implies the necessity of modeling the endogenous loan amount for future works.

Third, we contribute to analyzing the impact of oligopolistic market structure on loan performances. In South Korea, the five largest commercial banks account for more than 50% of the market share in the personal loan market. Since there are few commercial banks for consumer financing, our model does not rely on the competitive market assumption. Based on the main estimation results, we provide counterfactual analyses about how the market structure interacts with asymmetric information. For example, our finding presents that more competition among banks may offer better loan contract terms for consumers. When the loan prices-amounts correlation among banks decreases by half, the model predicts a 9% decrease in loan price, a 0.5% increase in lending limit, a 0.3% increase in loan demand, and a 0.04%P default rate decrease. We believe our result may provide an implication for assessing the impact of new online banks operating without physical branches.

We construct a structural model with multiple stages to incorporate the interactive role of two screening devices into the conventional loan demand and supply models. Each personal loan contract consists of three phases. In the first stage, a consumer applies for taking out a personal loan from banks in her residence. Banks simultaneously decide a pair of loan prices and loan amounts by considering their lending capacity and consumer characteristics. We model a general consumer-bank-specific loan demand and supply function and use the estimates to predict counterfactual loan prices and quantities. In the second stage, a consumer decides from which bank to borrow. The estimation follows a discrete choice model used for demand estimation in a similar spirit as Berry (1994) and Berry et al. (1995). In the third stage, a consumer chooses to default or not, conditional on loan demand and the corresponding loan price and loan amount. The procedure is repeated for those who want to take out another personal loan. Then, we derive a conditional likelihood function from the above three stages and estimate model parameters.

Our methodology extends the structural model of Crawford et al. (2018) in two ways. First, our structural model incorporates the bank's lending capacity and loan pricing, thereby endogenizing the loan amount and price. The previous literature assumes that the loan amount is exogenous conditional on loan approval. In our model, banks compete by both loan prices and lending limits to maximize the expected profit from customers. The assumption of two-dimensional screening devices allows a more flexible price schedule for a personal loan contract between a bank and a consumer. We answer why the convex loan pricing pattern predicted under a single screening device is not apparent in South Korea. The compromise between loan price and lending limit in a loan contract implies that a convex pricing schedule is not necessarily optimal in a personal loan contract despite the exclusive loan contract property in South Korea.

Second, we design a more flexible cost structure for lenders and potential loan prices and amounts correlated among banks. The marginal cost of lending increases in the loan amount under the lending capacity. Since a few large commercial banks occupy a substantial market share in the personal loan market in South Korea, the model considers that the banks' decisions on loan contract terms can be intertwined. Our model categorizes 19 banks into four groups: five large, six middle, six regional, and two online banks. Our estimation finds significant correlations in loan prices and amounts within and between groups. The correlation is more substantial with adjacent categories, and the five largest banks' decisions have only a mere correlation with other smaller banks' decisions. The strategic interactions of banks enable us to conduct a more detailed counterfactual analysis regarding the change in market structure. Our model reveals how the lender's market power and the information asymmetry mutually interact with each other. We present the counterfactual outcome of loan price, loan amount, loan demand probability, and default rate concerning the changes in the market environment.

1.2 Previous Literature

Most theoretical and empirical literature in markets with asymmetric information focuses on a single screening device and assumes a competitive market for loan suppliers. Einav et al. (2012) pay attention to loan demand estimation in the automobile loan market where the loan interest rate is the only screening device, and consumers do not face borrowing constraints. Starc (2014) specifies a Medigap market with the imperfect competition of insurers, but the insurance premium is the only screening device. One of the closest papers to our approach is Crawford et al. (2018) because they consider the market power in oligopolistic loan suppliers and jointly analyze both the demand and supply side in Italian credit markets. But the loan price is still the only screening device in the previous literature, and the loan amount is exogenously given. Ioannidou et al. (2021) extend the setup established by Crawford et al. (2018) to measure the effectiveness of collateral in the credit market by considering both secured and unsecured loans.

Our paper contributes to the growing empirical literature dealing with asymmetric information in credit and insurance markets. The first trend of previous papers empirically tests the existence of asymmetric information in insurance markets. The empirical results of the tests are quite mixed. Puelz and Snow (1994), Finkelstein and Poterba (2004), Cohen (2005), and He (2009) find an evidence of asymmetric information in some markets, while Cawley and Philipson (1999), Chiappori and Salanié (2000), and Cardon and Hendel (2001) do not find any evidence of asymmetric information.[1]

In terms of asymmetric information in consumer credit markets, previous articles adopt various methods. Karlan and Zinman (2009), Agarwal et al. (2016) use a large-scale randomized experiment, while Adams et al. (2009) and Dobbie and Skiba (2013) exploit regulatory and institutional features to identify the moral hazard and adverse selection separately. Additionally, Davidoff and Welke (2004) find an advantageous selection in the U.S. reverse mortgage market. Agarwal et al. (2016) find that less credit-worthy applicants are more likely to choose a credit contract with a lower collateral requirement and a higher interest rate in the home equity loan market. Edelberg (2004) finds robust evidence of adverse selection when high-risk borrowers pledge less collateral and pay higher interest rates, even after controlling for income levels, loan size, risk aversion, and evidence of moral hazard.

Our structural model provides empirical evidence of the economic theory that explains market outcomes in the selection markets. For instance, Rothchild and Stiglitz (1978) developed a simple equilibrium framework in the selection market, considering both single and dual screening devices. Azevedo and Gottlieb (2017) provided the conditions on equilibrium in the market with adverse selection under the perfect competition between loan suppliers, and Mahoney and Weyl (2017) used a model to describe the selection market with imperfect competition and derived comparative statics to suggest policy implications. Lester et al. (2019) provide a theoretical background of how adverse selection, screening, and imperfect competition generate an equilibrium contract, especially when both price and quantity of the product are screening devices. Their paper implies that more competitive loan suppliers or less information asymmetry between borrowers-lenders may not necessarily improve social welfare. Considering both asymmetric information and imperfect competition, Lester et al. (2019) explain how the market power of banks and asymmetric information interact with each other. Our paper also considers the selection market with imperfect competition and dual screening devices. The structure motivates us to study how the market outcome, comparative

[1] Chiappori and Salanié (2000) also develop a general "positive correlation test" to check for the presence of asymmetric information in a contractual relationship within a competitive market.

statics, and policy implications vary on a more general theoretical framework.

Regarding the bank's lending capacity, Duca and Rosenthal (1994) and Acolin et al. (2016) discuss the role of borrowing constraints on homeownership in the United States, focusing on the housing loan market. The housing loan market uses collateral and minimum down payment requirements as screening devices. However, a personal loan market in the current paper is a market without requiring any specific collateral, so we cannot analyze it in the same context. Considering a lending market without collateral, we pay attention to the role of loan price and amount in the personal loan market of South Korea. For example, a bank may not necessarily offer a high loan price to a high-risk borrower if lending less loan to the consumer is more likely to control the default risk. Similarly, a bank can suggest a high-interest rate to a low-risk borrower by relaxing the lending constraints. A bank can manage the consumer's default risk by a nonlinear price and quantity schedule.[2]

The primary role of the loan amount constraint in our model is not pure credit rationing in previous literature but a borrowing constraint and a screening device. In most previous research, the borrowing constraint is presented by a form of credit rationing, which is an optimal behavior of lenders under information asymmetry even if lenders have enough financial resources (Stiglitz and Weiss (1981), Besanko and Thakor (1987), Williamson (1987), Berger and Udell (1992)). In contrast to many previous papers regarding pure credit rationing, the setup in our study presumes insufficient financial resources. There are several reasons for this setup. In South Korea, the Loan-Deposit Ratio (LDR) of banks is regulated by the government, so the financial source of banks for lending is limited in reality. Banks are also required to preserve the Debt Service Ratio (DSR) and the ratio of unsecured loans to the total amount of loans within a fixed level for risk management. The restriction on the maximum loan amount in a contract implies a borrowing constraint, which has been one of the non-price risk management terms in lending markets (Duca and Rosenthal (1994), Acolin et al. (2016)).

In the following sections, we introduce the source of data and summary statistics in Section 2. In Section 3, a theoretical background of the structural model will be introduced, and Section 4 provides an econometric specification of the structural model. Section 5 shows estimation results about structural parameters and counterfactual analysis, and discuss the following empirical/policy implications. Section 7 is a concluding remark.

2 Data

We use a unique dataset of personal loans in South Korea to investigate the effects of asymmetric information between consumers and banks under imperfect competition among banks. One of the major credit rating companies in South Korea, the Korea Credit Bureau (KCB), provides the

[2]The multiple screening devices are also related to multi-dimensional private information in previous literature. See Finkelstein and McGarry (2006) and Fang and Wu (2018). But this literature focuses on high-risk-averse type consumers (advantageous selection) who are not distinguishable from risky consumers (adverse selection). The context is different from our motivation because our paper focuses on borrowing constraints that lead a non-risky consumer to choose the same action as a risky consumer.

dataset.[3] The KCB dataset is a random sample containing 10% of the whole population in South Korea. The KCB rates credit scores (CB Score) for the entire population in South Korea by collecting credit information. The definition of credit information follows the Credit Information Use and Protection Act (Amended by Act No. 16188) and Enforcement Decree of the Credit Information Use and Protection Act (Amended by Presidential Decree No. 30893). The provided credit information includes the following:

1. Individual identifiers (Article 2-3 of the Enforcement Decree)

2. The type, period, and amount of commercial transactions, including loans, guarantees, provision of collateral, current account transactions, credit cards, installment financing, facility leases, and financial transactions (Article 2-6 of the Enforcement Decree)

3. Information on delinquency, dishonor, and default (Article 2-8 of the Enforcement Decree)

4. Occupation, total amount of assets, liabilities, and income (Article 2-10 of the Enforcement Decree and Addendum 1)

and the KCB maintains up-to-date information as consumers use financial services, including personal loans, car loans, and mortgages.

The provided variables include all the "hard information" that a bank considers for a personal loan contract in the real world. A bank requests a customer's information to credit rating companies whenever a customer tries to take a personal loan. The bank uses the credit information to determine whether to approve the loan application. Conditional on the approval, the bank offers the loan price and the maximum lending limit. After signing the loan contract, the bank reports the contract details back to the credit rating company. During the repayment period, the bank records if the personal loan payments are on time so that the credit rating company adjusts the customer's credit score.

Our dataset is not free from previous literature's "soft information" issue. Soft information refers to all the information observable to the banks but unobservable to the researcher (e.g., underwriting process or interview). In the personal loan market of South Korea, most banks use their credit evaluation system in addition to the KCB dataset to assess loan applicants, mainly if they are existing customers. Although many banks accept personal loan applications on their websites after 2016 and make decisions mechanically based on hard information, our model does not exclude potential soft information. We control the unobserved heterogeneity by a fixed-effects panel model since our dataset includes individuals who submitted multiple loan applications during the observed periods. About 24% of our observations (277,616 observations across the whole sample periods) are from consumers with more than one personal loan records.

This paper pays attention to unsecured personal loans out of many other loan types. Banks do not require loan-takers to pledge any collateral for taking a personal loan. The main reason is

[3] There are two major credit rating companies in Korea: Nice Investors Service Corporation and Korea Credit Bureau. Our dataset is a proprietary dataset from the KCB ([Dataset] (2019).)

that the adverse selection or moral hazard aspects are fundamentally different depending on loan types. Notably, a collateralized loan and a personal loan have some distinct attributes. First, a collateralized loan takes collateral as a screening device to mitigate information asymmetry in lending markets. The bank's lending limit also relies on the value of the collateral. However, the personal loan market uses loan price and amount as screening devices. Second, a bank usually requires collateral when the loan amount is sufficiently large, e.g., mortgages or car loans. Consumers tend to take a personal loan when they need a relatively small amount of finance promptly. Our raw dataset provides the average loan price and amount for loan-takers. The average loan price is 0.0448 (4.48% per year), and the average loan amount is 37,320,000 Korean won (KRW), around 37,320 US dollars.

The Korean consumer credit market is *effectively* exclusive: the structure is institutionally non-exclusive but practically working exclusively. There is no benefit for a loan-taker to linearize the loan price even though a loan-taker can simultaneously make loan contracts with several banks.[4] The exclusivity in a loan contract is a crucial issue in contract theory because the property restricts the possible set of contracts (Chiappori and Salanié (2000), Chiappori and Salanié (2013)). For example, a convex price scheme is suitable for the exclusive market structure: the loan interest rate rises with the loan amount. The credit market in South Korea has exclusive property because the Korea Federation of Banks (KFB) registers all consumer characteristics and contract details to a unified electronic system. Then all the KFB members share the collected information, including the previous loan contracts and payment history. If a consumer with unpaid debts applies for an additional loan, the KFB provides information about all the existing unpaid loans to prevent excessive loans beyond the consumer's affordable leverage. Thus, the bank decisions on a loan application are highly correlated. If a bank rejects one's loan application, it is unlikely for the same individual to get approval from other banks. Even if some banks approve the application, the consumer is usually offered a meager loan amount with a high interest rate.

The loan interest rate roughly consists of four components: the Korea Interbank Offered Rates (KORIBOR), education tax, credit risk cost, and profit. The exact formula of determining the loan interest rate is confidential information for each bank, so the details are not publicly available.

$$\text{Loan Interest Rate} = \text{KORIBOR} + \text{Tax} + \text{Credit Risk Cost} + \text{Profit}$$

The KORIBOR is the average leading interest rate determined by a group of major Korean banks. The education tax is constant (0.5%) throughout the data period. The credit risk cost is the average default price for the bank: the cost is higher for loans with a higher expected default rate and is lower for loans with a high recovery rate. The credit risk cost is the source of the positive correlation between a loan interest rate and the lending limit. A bank charges a higher interest rate to customers with lower credit ratings. At the same time, a bank offers a lower interest rate

[4] In contract theory, a bank may use a convex price scheme under the exclusive market by proposing a higher loan interest rate for a larger loan amount. If the credit market is non-exclusive in this situation, a borrower can linearize the price by borrowing a small loan from a large number of banks.

for small loans, thereby implying a positive relationship between a loan interest rate and the loan amount, given the customer's credit rating.

2.1 Loan Information

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.
Panel A. Loan level			
Amount Granted (Thousand KRW)	908,941	37,320	35,777.5
Interest Rate (%)	908,941	4.45	2.11
Default Rate (%)	908,941	0.98	0.31
Panel B. Bank level			
Number of Branches (Province level)	17	78.67	69.36
Total Deposit (2018, Billion KRW)	19	68,018.8	76,620
Deposit Interest Rate (2018, %)	19	1.45	0.23
Personal Loans (2018, Billion KRW)	19	33,254.5	40,995.2
Business Loans (2018, Billion KRW)	17	44,160.2	41,585.5
Debt Ratio	19	1,158.72	295.33
Return on Equity	19	4.02	2.38
Panel C. Consumer level			
Income (Thousand KRW)	908,941	51,558	39,680.8
Debt to Income Ratio (DTI)	908,941	0.13	0.28
Type of Job	908,941	0.26	0.44
Age	908,941	45.61	10.50
Credit Score	908,941	851.50	110.7
Credit Card Spendings (Thousand KRW)	908,941	25,620	38,423.4
Debit Card Spendings (Thousand KRW)	908,941	3,454	22,013.7

Note: Table 1 shows summary statistics for information contained in our dataset. Panel A describes loan level information. Panel B. describes bank level information, while panel C describes customer level information.

The KCB dataset contains information about detailed personal loan contracts. The sample period is from 2013 to 2019, and the regions include seven metropolitan cities in South Korea: Seoul, Incheon, Busan, Daegu, Daejeon, Gwangju, and Ulsan. The population size of cities is as follows: Seoul 10.048 million, Incheon 2.918 million, Busan 3.515 million, Daegu 2.489 million, Daejeon 1.522 million, Gwangju 1.473 million, and Ulsan 1.171 million, according to the 2015 census data. We implicitly assume that markets are segregated since each metropolitan city represents the capital city of the province.

There are 19 banks in total, but not all banks operated during the whole sample period. For example, two online banks in the dataset started their businesses in 2017. The final sample consists of 49 region-year and 774 bank-region-year combinations. The personal loan in our dataset refers

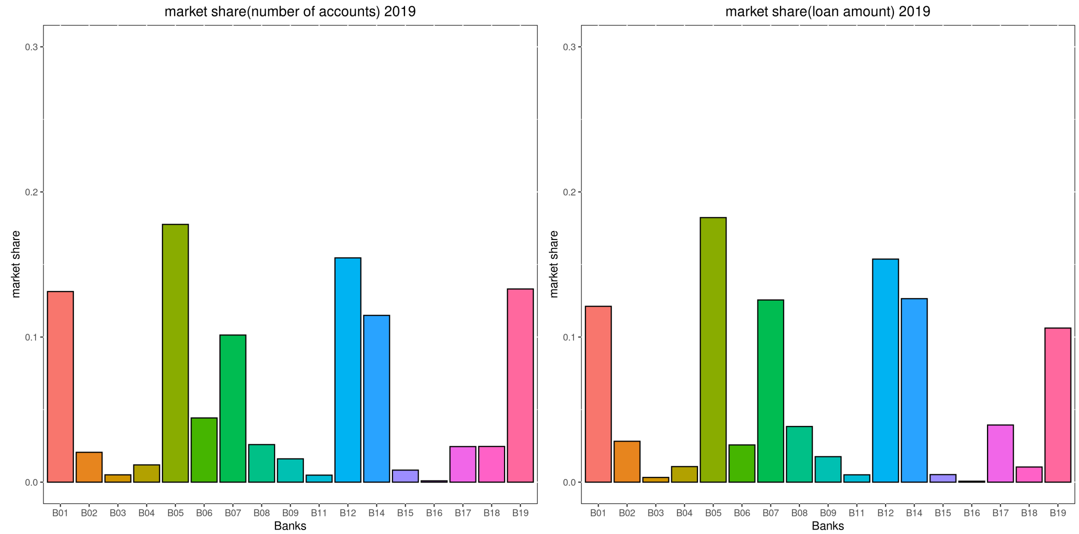

Figure 1: Market Share of Banks in 2019

Note: The horizontal axis indicates 19 banks, and the vertical axis shows the market share. The large-size banks include B01, B05, B07, B12, and B14. The middle-size banks include B02, B06, B10, B11, and B17. The regional banks are B03, B04, B08, B09, B15, and B16, and the online banks are B18 and B19. The first panel computes the market share based on the number of loan contracts. The second panel uses the total loan size to calculate the market share.

to a type of loan that is not required to pledge any collateral. We are particularly interested in one-year maturity personal loans observed from the seven largest cities in South Korea. The sample size in Table 1 is around 900,000 observed for seven years.[5]

Here we define a market as a combination of each city and year. For example, we regard Seoul - the capital city of South Korea - in 2013 and Seoul in 2014 as different markets. The setup follows a traditional approach to defining the market as Berry et al. (1995) suggested.

For each loan record, the dataset contains loan interest rate, loan amount, maturity, and the bank identifier. Based on loan characteristics combined with customer characteristics, we predict the loan interest rate and the loan amount suggested by "unselected" banks. The counterfactual loan prices and amounts provide information on the indirect utility function parameters in the loan demand function. Section 4 introduces details in estimation process.

Table 1 provides summary statistics of the KCB dataset. The average loan amount is around 35 million KRW, about 68% of their annual income. The average loan interest rate in the dataset is 4.45%, which is a little bit higher than the average loan price of collateralized loans (3~4%) in the same period. Banks impose a higher loan interest rate on personal loans since an unsecured loan does not require any collateral to pledge, thereby the recovery rate is lower once the borrower defaults.

[5] The number of observations increases if we include loans with more than one year of maturity. According to the dataset, there is no significant loan interest rate change concerning maturity, and our estimation results do not dramatically change by adding more observations.

Figure 1 displays the market share of the personal loan market in 2019. The five largest banks possess a substantial market share, and a newly rising online bank (B19) also takes a significant portion of the market. The dataset comprises six regional banks, six medium-sized banks, five large banks, and two online banks. The bank size category definition is based on the following criteria. First, a large bank refers to a bank with the largest total assets. The five largest banks account for 50.9% of the personal loan market share. According to the Korean Statistical Information Service (KOSIS), the five largest commercial banks managed 135,528.6 billion KRW out of 265,972.7 billion KRW of total personal loans in the 4th quarter of 2020.[6] The six regional banks have branches only in the specific province in South Korea, and two online banks do not have physical branches. The remaining six banks operating in all the regions in South Korea belong to the middle size banks.

Figures 4 and 5 in Appendix A.1 show the trend of market shares between 2013 to 2018. Before the rise of online banks in 2017, there was a more substantial difference between large-size banks and other banks. The figures present the oligopolistic market structure in the personal loan market in South Korea.

2.2 Bank Characteristics

The bank characteristics inform the lending capacity of banks. The Financial Supervisory Service of South Korea regulated the LDR of banks not to exceed 100% during the data periods (2013~2019). The regulation implements that the bank's total lending amount cannot exceed the whole deposit. The policy is the primary source that banks compete with through loan interest rates and loan amounts.

During the data periods, the average deposit interest rate is 1.45%, which is way lower than the average loan price of 4.42% in Table 1 Panel A. The government regulation to control LDR creates a significant gap between the deposit interest rate and loan interest rate. The banks compete for deposits by the deposit interest rate (e.g., Egan et al. (2017)). The average personal loan price is relatively high to keep the LDR at a low level.

Table 1 Panel B presents summary statistics about bank-level variables, including the number of branches, total deposit, average deposit interest rate, personal loan and business loan amounts, debt ratio, and return on equity. We observe 17 offline commercial banks and two online banks that provide personal and business loan services. An online bank denotes a bank operating without a physical branch, while consumers freely make transactions using smartphone applications or any ATM. In addition to the information in the table, the bank's number of employees, deposits, and loan amounts are available at the district level. This bank-specific information also matches with aggregate consumer information, including the population, average income, unemployment rate, and other financial and occupational variables at the district level.

Figure 2 shows the average loan prices and amounts across banks in our dataset. We find a considerable variation in loan prices and quantities. For example, the large-size banks do not have many outliers in loan contract terms. On the other hand, some regional banks (B15 and B16) keep

[6] The statistic is available at https://kosis.kr/.

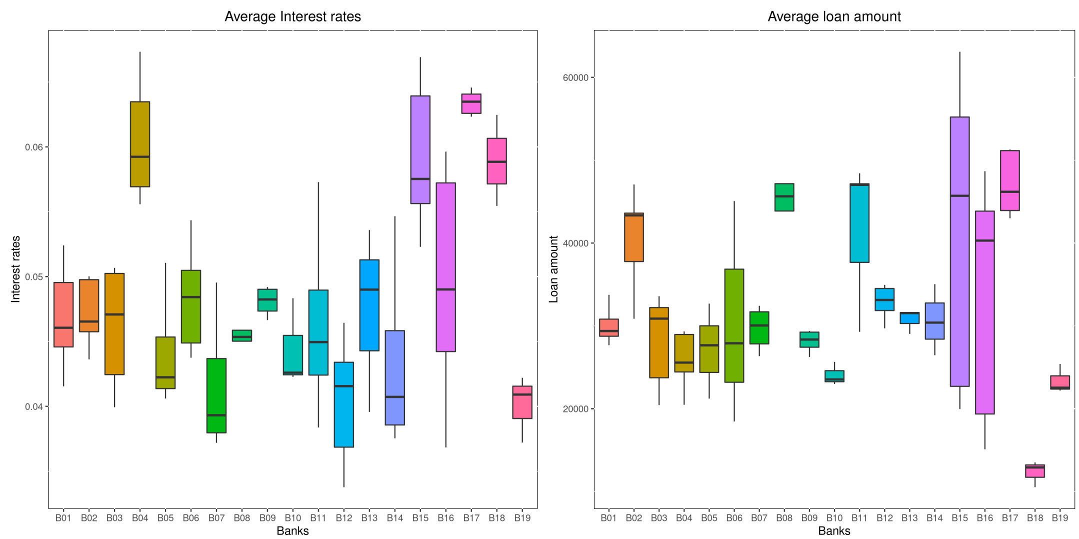

Figure 2: Average Loan Price and Amount

Note: The horizontal axis indicates 19 banks, and the vertical axis shows the average loan price and amount. The large-size banks include B01, B05, B07, B12, and B14. Middle-sized banks include B02, B06, B10, B11, B13, and B17. The regional banks are B03, B04, B08, B09, B15, and B16, and the online banks are B18 and B19. The loan amount in the second panel follows the unit of thousands of KRW.

both price and quantity at a higher level, and an online bank (B19) uses a strategy to lower the loan price and amount.

2.3 Consumer Characteristics

The KCB dataset also contains detailed information on potential loan takers in the Korean credit market, including customers who made contracts with banks and those who did not. The characteristics include age, income, occupation, residence, credit score, and variables relevant to consumer behavior. The key variables capturing an individual's consumption level are debit and credit card spending. The information indirectly reveals the consumption behavior of loan-takers because most financial transactions in large cities are through debit or credit card payments. According to the Bank of Korea statistics in 2018, debit and credit cards spending accounted for 80.2% of household consumption.

We also observe each individual's residence and credit activities, including delinquencies. The customer information matches the bank-level data and regional aggregate economic variables, including the number of bank branches, average income, and unemployment rate.

The credit score is a numerical value ranging from 0 to 999. The credit rating company KCB rates credit risks based on the observable characteristics of each consumer. In general, the credit score and the default rate have a negative correlation. The literature describes the default as specific days of loan repayment delinquency. In this dataset, a customer is regarded as committing a default decision if the length of repayment delinquency continues for more than 30 business days.

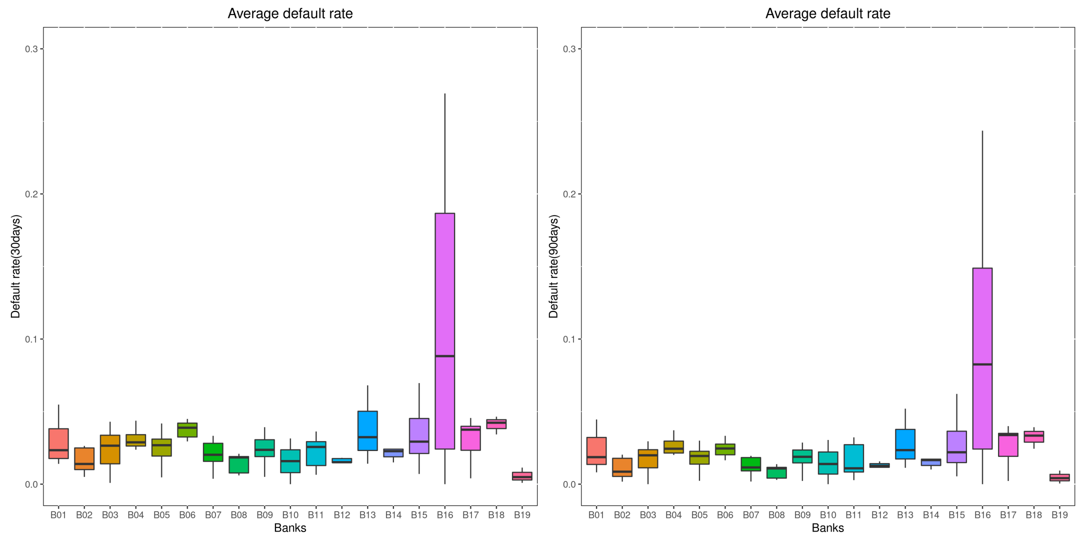

Figure 3: Average Default Rate Across Banks

Note: The horizontal axis indicates 19 banks, and the vertical axis shows the average default rate. The large-size banks include B01, B05, B07, B12, and B14. Middle-sized banks include B02, B06, B10, B11, B13, and B17. The regional banks are B03, B04, B08, B09, B15, and B16, and the online banks are B18 and B19. In the first panel, the definition of default is the loan delinquency for 30 days, and the second panel uses the loan delinquency for 90 days.

The average default rate is about 0.9%.

Table 1 Panel C shows consumer characteristics relevant to our analysis. Our dataset's average income level is about 51 million KRW, and their debt-to-income ratio (DTI) is about 0.13. Blue-collar and white-collar types classify the type of jobs. The consumer's average credit score is around 850, while the maximum rating is 999. The sample average score is relatively high because we focus on the primary banking sector.

Figure 3 shows the histogram of bank-specific default rate distributions. Despite some variations, the default rates are stable at around 1%. We present the average default rates using two different definitions. The first panel uses 30-days loan delinquency, and the second panel uses 90-days instead.

Figures 6 to 10 in Appendix A.2 display basic relationships affecting the default decision. According to the figures, the default rate is higher for those with higher loan interest rates, lower loan amounts, lower credit scores, and more credit/debit card spending. The trends indirectly verify the intuition that banks are screening high-risk customers by raising the loan price and lowering the lending limit. There is no specific correlation between age and default rate, but we find that age groups 30s and 60s have relatively higher default rates. The trend fits the standard life-cycle model's prediction.

3 The Model

This section presents the structural model to identify asymmetric information in the personal loan market. The model setup is similar to Crawford et al. (2018) but we additionally specify the bank's lending limit.

3.1 Basic Setup

We construct a model to measure the effects of asymmetric information between lenders-borrowers and imperfect competition among lenders. For market $m \in \{1, \ldots, M\}$ and year $t \in \{1, \ldots, T\}$, there are active banks (lenders) $j = 1, \ldots, J_{mt}$ and many customers (borrowers) $i = 1, \ldots, I_{mt}$ for each market. Each customer contacts all the J_{mt} banks for quotes. Receiving the application, each bank requests the customer's credit information (Z^C_{ijmt}) and confirms the bank's lending capacity (Z^B_{jmt}). Then, using the information set $Z_{ijmt} \equiv \left(Z^{C'}_{ijmt}, Z^{B'}_{jmt}\right)'$, banks simultaneously suggest a set of loan interest rates and maximum loan amount pairs to the customer to maximize the expected profit. Considering offers from J_{mt} different banks, the borrower chooses a bank j and the corresponding contract (P_{ijmt}, Q_{ijmt}) that maximizes her utility. If all banks reject the loan application or none of the received offers are satisfactory to the customer, there is no contract. Each bank j's market share depends on the number of borrowers who decide to take a loan from the bank j.

The model developed in our paper has similarities with the previous literature of Starc (2014) and Crawford et al. (2018) but postulates a fundamental distinction: the loan amount Q_{ijmt} in our model is endogenous. In the personal loan market in South Korea, no bank offers a loan without setting a limit. Every bank has a lending limit due to regulations to control the LDR and the ratio of unsecured loans to the entire loan. Each bank allocates a limited budget to maximize its expected profit. Banks also compete in both price and amount to attract more consumers. Thus, the loan demand utility is a function of both P_{ijmt} and Q_{ijmt}. We specify a system of equations to describe the equilibrium loan price and amount, considering both consumer preference and the market structure of loan suppliers.

Our model assumes that banks use the loan price and amount as screening devices, while previous literature only focused on the loan price. The loan amount is a primary source of borrowing constraints, and the upper bound depends on the lending limit of the banks and consumer characteristics observable to banks. The loan amount works as a screening device because the upper limit of the loan implies the maximum loss from the consumer's default, particularly in the unsecured personal loan market without collateral. The interactive role of dual screening devices provides a better model fit to explain a loan contract in South Korea under adverse selection.[7]

[7] The assumption of multiple screening devices implies that banks may not implement convex price schedules even under exclusive contracts. A risky consumer may put up a tighter borrowing constraint to avoid a high interest rate. See Rothchild and Stiglitz (1978) and Chiappori and Salanié (2000) for a detailed theoretical foundation. Furthermore, Rothchild and Stiglitz (1978) posited the potential non-existence of equilibria under price and quantity contracts in a competitive insurance market.

Our model considers oligopolistic loan suppliers facing price and quantity competition instead of assuming a competitive loan market. Determining the loan price and amount under oligopolistic competition depends on each bank's market power. The banks face uncertainties for every stage of a loan contract. At the pre-contract stage, a bank manages a potential information asymmetry issue related to loan applicants' unobserved riskiness. A bank also competes with other banks to attract customers with better contract conditions. Some loan takers may not pay the money back at the post-contract stage and declare a default. Banks optimally choose their lending limits and the corresponding price schedule based on their information set, considering the potential risk. We empirically design a correlation structure that affects all banks' quotes and estimate the effect of loan supply market structure on consumer loan contracts.

3.2 Consumer Demand and Default on Bank Loans

The section provides a model to describe the consumer utility function for loan demand and default. In the model described below, we assume that each customer i living in market m in year t applies for a loan, and exclusively makes a contract with a bank among the banks operating in market m. We start by the loan demand side. Suppose a borrower i has a set of loan contract offers (P_{ijmt}, Q_{ijmt}) from all banks $j = 1, \ldots, J_{mt}$. P_{ijmt} and Q_{ijmt} are functions of the bank's information set Z_{ijmt}, following the specification presented below.

$$
\begin{aligned}
P_{ijmt} &= g_j^P \left(Z_{ijmt}, \varepsilon_i^P, u_{ijmt}^P \right) \\
Q_{ijmt} &= g_j^Q \left(Z_{ijmt}, \varepsilon_i^Q, u_{ijmt}^Q \right),
\end{aligned} \tag{1}
$$

where g_j^P and g_j^Q are bank-specific optimal policy functions of hard information Z_{ijmt} and all the other information $\left(\varepsilon_i^P, \varepsilon_i^Q, u_{ijmt}^P, u_{ijmt}^Q\right)$. u_{ijmt}^P and u_{ijmt}^Q include soft information and the bank-region-year-level shocks that influence the resulting loan price and amount. Thus, u_{ijmt}^P and u_{ijmt}^Q are potentially correlated with Z_{ijmt}. We also allow for the correlation of u_{ijmt}^P and u_{ijmt}^Q across banks $j = 1, \ldots, J_{mt}$ to identify the competition effect among lenders. ε_i^P and ε_i^Q summarize the remaining individual-specific adjustment factors affecting the loan price and amount after controlling Z_{ijmt}. The functions g_j^P and g_j^Q reflect the bank's mechanism of how each loan applicant-bank pair decides the personal loan terms using the hard and soft information.

The joint distribution of ε_i^P and ε_i^Q reveals the correlation of loan price and amount not captured by observables Z_{ijmt}. Compared with Crawford et al. (2018), our specification endogenizes both P_{ijmt} and Q_{ijmt}. The borrower's potential riskiness may imply a higher loan price and a lower lending limit (high ε_i^P and low ε_i^Q). In comparison, positive factors like long transaction history can result in a lower loan price and a higher lending limit (low ε_i^P and high ε_i^Q). We will discuss the equilibrium loan price and amount considering the loan supply side in Section 3.3.

Conditional on the set of offers (P_{ijmt}, Q_{ijmt}) for $j = 1, \ldots, J_{mt}$, the borrower's valuation of the bank j's offer in market m in year t is a function of Z_{ijmt} and unobservable factors. We approximate the loan demand utility U_{ijmt}^D by a linear function of observable covariates.

$$U_{ijmt}^D = X_{jmt}^{D'}\beta^D + \xi_{jmt}^D + \alpha_1^D P_{ijmt} + \alpha_2^D Q_{ijmt} + Y_{ijmt}^{D'}\gamma^D + \varepsilon_i^D + \nu_{ijmt}, \tag{2}$$

where X_{jmt}^D is a vector of bank-specific covariates that affect the loan demand, ξ_{jmt}^D is unobserved preference on bank j in market m in year t, and Y_{ijmt}^D includes other individual-level covariates. ν_{ijmt} summarizes unobserved shocks to the loan demand. ε_i^D is the individual propensity on loan demand and is unobservable to the banks. ε_i^D correlates with $\left(\varepsilon_i^P, \varepsilon_i^Q\right)$ in equation (1) since a borrower with high demand on a personal loan may accept a higher loan price for borrowing a larger loan amount. Thus, we expect $cov\left(\varepsilon_i^D, \varepsilon_i^P\right) > 0$ and $cov\left(\varepsilon_i^D, \varepsilon_i^Q\right) < 0$. For an outside option $j = 0$, $U_{i0mt}^D = \nu_{i0mt}$ normalizes the utility of not choosing any bank. Under the given loan contract schedule P_{ijmt} and Q_{ijmt} from equation (1), the borrower chooses a bank $j \in \{0, 1, \ldots, J_{mt}\}$ that maximizes U_{ijmt}^D.

The valuation in equation (2) is a linear function of P_{ijmt} and Q_{ijmt}. The demand utility function implies substitutability between loan price and the amount because a larger loan can compensate for a high loan price. The borrower chooses a bank based on both loan price and amount to maximize the utility of loan demand, conditional on observable bank-specific variables X_{jmt}^D and individual-bank-specific variables Y_{ijmt}^D. Note that the bank's information set Z_{ijmt} contains X_{jmt}^D and Y_{ijmt}^D. In our configuration, the parameters α_1^D and α_2^D represent the relative sensitivity of consumer demand to loan price and quantity. For example, suppose a case that α_1^D is highly negative, while α_2^D is close to zero. Then the model implies that borrowers respond to P_{ijmt}'s change more sensitively than the change of Q_{ijmt}. In this case, the borrower's bank choice should highly depend on the loan interest rate instead of the loan amount.[8]

Next, the borrower i can make a default decision after taking out a loan at bank j. We approximate the default utility function by a linear function of observable covariates. The valuation of default in market m and year t is specified by:

$$U_{ijmt}^F = X_{jmt}^{F'}\beta^F + \alpha_1^F P_{ijmt} + \alpha_2^F Q_{ijmt} + Y_{ijmt}^{F'}\gamma^F + \varepsilon_i^F, \tag{3}$$

where X_{jmt}^F are bank-specific factors that affect the default decision, Y_{ijmt}^F are other individual-bank level covariates related to the default, and ε_i^F is the unobserved individual propensity on default. The default probability potentially depends on the remaining loan amount, potential penalty of a default, and unexpected income shocks. More specifically, X_{jmt}^F summarizes the source of potential penalty to borrowers who declare a default, Y_{ijmt}^F includes credit score and variables regarding the type of loan contract, and P_{ijmt} and Q_{ijmt} present the remaining loan amount that the borrower will repay. The coefficients α_1^F and α_2^F capture the moral hazard effect because the borrower's effort to prevent a default may depend on the magnitude of (P_{ijmt}, Q_{ijmt}) under the moral hazard.

[8]The linearity of demand utility in the loan amount Q_{ijmt} may not be a correct specification because each borrower chooses an optimal loan amount that maximizes the expected lifetime utility. We observe that the lending limit is binding for most loan contracts, thereby assume that the demand utility is a non-decreasing function of the loan amount.

Many theoretical references, including Pauly (1978), Golosov and Tsyvinski (2007), and Bertola and Koeniger (2015), witness the theoretical background of moral hazard in the selection markets. The utility from no default is normalized by zero: the customer pays back the loan as long as $U^F_{ijmt} < 0$ and defaults if $U^F_{ijmt} > 0$.

We assume the additional information structure for the joint distribution of ε_i^P, ε_i^Q, ε_i^D and ε_i^F. The joint distribution of unobservable individual preference parameterizes information asymmetry between customers and banks. Consider the following joint normal distribution:

$$\begin{pmatrix} \varepsilon_i^P \\ \varepsilon_i^Q \\ \varepsilon_i^D \\ \varepsilon_i^F \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{PQ} & \rho_{PD}\sigma_D & \rho_{PF} \\ \rho_{PQ} & 1 & \rho_{QD}\sigma_D & \rho_{QF} \\ \rho_{PD}\sigma_D & \rho_{QD}\sigma_D & \sigma_D^2 & \rho_{DF}\sigma_D \\ \rho_{PF} & \rho_{QF} & \rho_{DF}\sigma_D & 1 \end{pmatrix} \right), \tag{4}$$

where the correlation coefficient $\rho_{DF} > 0$ presents the evidence of adverse selection.[9] That is, a positive correlation between unobservable preference on loan demand (hidden information) and the propensity to default is the degree of adverse selection. ρ_{PD} and ρ_{QD} show how the borrower's hidden preference on loan demand affects the loan price and amount. The unobservable demand shifter ε_i^D correlates with the loan price and amount shifters ε_i^P and ε_i^Q since a borrower is more likely to accept the high interest rate or small loan amount under the high demand. Similarly, ρ_{PF} and ρ_{QF} show the relation of default and loan price/amount. The larger loan amount typically causes more default. The variance of ε_i^F is normalized by one.

Our model also considers the existence of moral hazards in the personal loan market. We use the coefficients α_1^F and α_2^F in equation (3) to measure moral hazard. The moral hazard in the personal loan market reflects the intuition that a borrower's hidden action, for example, making fewer efforts to gain money, increases the probability of unexpected income shocks and leads to default. The default propensity increases as the moral hazard problem or the realized income shock are more serious. As the loan contract (P_{ijmt}, Q_{ijmt}) reflects the borrower's observable riskiness, α_1^F and α_2^F should be zero without moral hazard. The adverse selection and moral hazard parameters show the evidence of asymmetric information caused by hidden information and hidden action.

There are two econometric issues in the estimation of demand and default parameters. First, the dataset only includes (P_{ijmt}, Q_{ijmt}) that are actually chosen by a borrower, while the presented structural model requires information on the loan price P_{ijmt} and the loan amount Q_{ijmt} for all $j = 1, \ldots, J_{mt}$. In Section 4, we predict the unobserved offers from other banks. We specify a system of equations that determine the loan price and the amount. Then we use the fitted value of the model as a predictor of the unobserved loan contract terms. Second, (P_{ijmt}, Q_{ijmt}) are functions of soft information observed by banks but unobserved by econometricians. The omitted variables raise an issue of unobserved heterogeneity when we estimate the lending schedule g_j^P and g_j^Q and demand utility parameters. In Section 4.2, we use a panel model with fixed effects to resolve a

[9] Note that $\rho_{DF} < 0$ can be interpreted as evidence of advantageous selection. The selection happens if the lending cost for a lower credit borrower is high so that only well-prepared borrowers apply for taking out a personal loan.

potential endogeneity of P_{ijmt} and Q_{ijmt}. We exploit borrowers who are involved with more than one loan contract to separate unobserved heterogeneity from random errors.

3.3 Supply of Bank Loans

This section describes how the loan suppliers choose the equilibrium loan price P_{ijmt} and loan amount Q_{ijmt}. In contrast to previous literature that uses P_{ijmt} as the only screening device, we focus on both P_{ijmt} and Q_{ijmt}. Banks suggest an offer to each borrower considering the elasticity of demand and the elasticity of default to P_{ijmt} and Q_{ijmt}. For almost all cases in our dataset, the bank's offered lending limit is binding to the customer. Q_{ijmt} is the actual loan amount for borrowers who finally choose the bank j. The model described below discusses the price-quantity competition in the context of expected profit maximization.

Suppose a borrower i wants to take out a loan. Π_{ijmt} denotes the expected profit function of the bank j in market m in year t. Pr^D_{ijmt} is the probability that consumer i accepts the bank j's offer. If a bank j has a branch in the neighborhood m of consumer i, the objective function of each bank j follows the equation below.

$$\Pi_{ijmt} = Pr^D_{ijmt}\left(P_{ijmt}Q_{ijmt}\left(1 - F_{ijmt}\right) - C_{ijmt}\left(Q_{ijmt}\right)\right),$$
$$\text{where } C_{ijmt}\left(Q_{ijmt}\right) = c^a_{ijmt}Q_{ijmt} + \frac{1}{2}c^b_{ijmt}Q^2_{ijmt}.$$

P_{ijmt} and Q_{ijmt} denote a pair of the loan interest rate and the corresponding lending limit for consumer i, F_{ijmt} is bank j's expected value of borrower i's default probability, and C_{ijmt} is the cost function. The cost function is quadratic in Q_{ijmt}, thereby the marginal cost of lending increases in the loan amount Q_{ijmt}. The cost structure reflects the limited lending capacity. For example, a bank with less deposit tends to offer less loan amount due to the regulations. The expected profit function Π_{ijmt} is zero if the borrower i does not take out a loan from bank j ($Pr^D_{ijmt} = 0$) or if the bank j refuses to approve the loan application ($Q_{ijmt} = 0$).

Each bank simultaneously decides (P_{ijmt}, Q_{ijmt}) that maximizes the expected payoff function Π_{ijmt} at the timing of applicant i's application. Taking the derivative of Π_{ijmt} to choice variables provides the following two first order conditions:

$$P_{ijmt} = \frac{c^a_{ijmt} + \frac{1}{2}c^b_{ijmt}Q_{ijmt}}{1 - F_{ijmt} + F^{P'}_{ijmt}\mathcal{M}^P_{ijmt}} + \frac{\left(1 - F_{ijmt}\right)\mathcal{M}^P_{ijmt}}{1 - F_{ijmt} + F^{P'}_{ijmt}\mathcal{M}^P_{ijmt}} \tag{5}$$

$$Q_{ijmt} = \frac{1}{c^b_{ijmt}}\left(\pi_{ijmt} + \sqrt{\pi^2_{ijmt} + 2c^b_{ijmt}\mathcal{M}^Q_{ijmt}\left(P_{ijmt}\left(1 - F_{ijmt}\right) - c^a_{ijmt}\right)}\right), \tag{6}$$

where $\pi_{ijmt} = P_{ijmt}\left(1 - F_{ijmt} - F^{Q'}_{ijmt}\mathcal{M}^Q_{ijmt}\right) - c^a_{ijmt} - c^b_{ijmt}\mathcal{M}^Q_{ijmt}$ is the marginal profit from one more unit of Q_{ijmt}. $F^{P'}_{ijmt}$ and $F^{Q'}_{ijmt}$ are derivatives of the default probability F_{ijmt} satisfying

$$F^{P'}_{ijmt} = \frac{\partial F_{ijmt}}{\partial P_{ijmt}},\ F^{Q'}_{ijmt} = \frac{\partial F_{ijmt}}{\partial Q_{ijmt}},$$

and $\mathcal{M}^P_{ijmt}$ and $\mathcal{M}^Q_{ijmt}$ denote the inverse of the elasticity of choice probability with respect to P_{ijmt} and Q_{ijmt},

$$\mathcal{M}^P_{ijmt} = -\frac{Pr^D_{ijmt}}{\partial Pr^D_{ijmt}/\partial P_{ijmt}},\ \mathcal{M}^Q_{ijmt} = \frac{Pr^D_{ijmt}}{\partial Pr^D_{ijmt}/\partial Q_{ijmt}}.$$

Equations (5) and (6) provide a set of equations to display the optimal loan price and the corresponding lending limit chosen by the bank. The equations have several implications. First, the optimal price P_{ijmt} in equation (5) is the same as the Bertrand-Nash equilibrium price derived by Crawford et al. (2018), except that we assume a quadratic cost function instead of a linear cost function. The equilibrium price is the sum of effective marginal cost and effective markup. The marginal cost and the markup terms are weighted by a function of the default probability F_{ijmt} and the choice probability Pr^D_{ijmt}. For example, if a customer is less likely to default (F_{ijmt} is low), the effective marginal cost becomes lower. The exact value may vary due to the marginal default probability $F^{P'}_{ijmt} \geq 0$ weighted by the inverse elasticity of market share $\mathcal{M}^P_{ijmt} \geq 0$. The effective markup also depends on the elasticity parameter $\mathcal{M}^P_{ijmt}$, and the value vanishes as the market becomes more competitive ($\mathcal{M}^P_{ijmt} = 0$). Pr^D_{ijmt} converges to zero as the number of banks increases, thereby the optimal price will eventually converge to the effective marginal cost in the competitive market. The result coincides with the market equilibrium in the competitive market.

Second, the loan amount Q_{ijmt} is the other choice variable of the bank j to attract customers and reduce the potential loss when the default occurs. Following equation (6), the equilibrium lending limit Q_{ijmt} is also a function of F_{ijmt}, $\mathcal{M}^Q_{ijmt}$, and cost parameters $\left(c^a_{ijmt}, c^b_{ijmt}\right)$, provided that an interior solution exists. The existence of the solution relies on the condition $P_{ijmt}\left(1 - F_{ijmt}\right) \geq c^a_{ijmt}$, or lending $Q_{ijmt} > 0$ is more beneficial to the bank's profit. If $c^b_{ijmt} = 0$ and the marginal cost is fixed at c^a_{ijmt}, an interior solution may not exist. Note that the marginal profit of Q_{ijmt} is proportional to $\left(P_{ijmt}\left(1 - F_{ijmt} - \mathcal{M}^Q_{ijmt}F^{Q'}_{ijmt}\right) - c^a_{ijmt}\right)Q_{ijmt} + \mathcal{M}^Q_{ijmt}\left(P_{ijmt}\left(1 - F_{ijmt}\right) - c^a_{ijmt}\right)$ when $c^b_{ijmt} = 0$. Then, the optimal lending limit is unbounded as far as the expected marginal profit is greater than the marginal cost: $P_{ijmt}\left(1 - F_{ijmt} - \mathcal{M}^Q_{ijmt}F^{Q'}_{ijmt}\right) \geq c^a_{ijmt}$. The optimal lending limit is bounded when $\mathcal{M}^Q_{ijmt}F^{Q'}_{ijmt}$ is sufficiently large: under the fixed marginal cost, a bank has an incentive to regulate the lending amount when the increase in the lending limit significantly causes more defaults. If $c^b_{ijmt} > 0$, the increasing marginal cost and information asymmetry influence the lending limit together. Under the convex cost structure with $c^b_{ijmt} > 0$, the marginal profit will eventually move toward zero as the loan amount increases. For example, if the marginal cost of lending more loans is higher (c^b_{ijmt} is high), the lending limit is lower. The components of equation (6) respectively present the effect of asymmetric information and the effect of bank's lending capacity on the equilibrium loan amount.

To determine the interactive effect between asymmetric information and market power, we separately analyze the effect of asymmetric information or market power on the optimal loan price or loan amount. The asymmetric information effect is summarized by terms including F_{ijmt}, and the market power depends on the market share Pr^D_{ijmt}. We present the increase in market power by the increase in $\mathcal{M}^P_{ijmt}$ and $\mathcal{M}^Q_{ijmt}$, and the increase in information asymmetry by the change of default probabilities $F_{ijmt}^{P'}$ and $F_{ijmt}^{Q'}$. For the following analysis we assume that an interior solution of equation (6) exists, and the market power variables $\mathcal{M}_{ijmt}^{P}$ and $\mathcal{M}_{ijmt}^{Q}$ and asymmetric information variables $F_{ijmt}^{P'}$ and $F_{ijmt}^{Q'}$ are positively correlated each other.

1. The change of market power by fixing the degree of asymmetric information:
   (a) If $F_{ijmt}^{P'} = 0$, the optimal price increases as $\mathcal{M}_{ijmt}^{P}$ increases.
   (b) If $F_{ijmt}^{P'} > 0$, the optimal price is uncertain even if $\mathcal{M}_{ijmt}^{P}$ increases, because the effective markup increases while the effective marginal cost decreases. Specifically, the optimal price increases if $(1 - F_{ijmt})^2 > \left(c_{ijmt}^{a} + \frac{1}{2}c_{ijmt}^{b}Q_{ijmt}\right) F_{ijmt}^{P'}$ and decreases otherwise.
   (c) If $F_{ijmt}^{P'} < 0$, the optimal price increases as $\mathcal{M}_{ijmt}^{P}$ increases, because both effective marginal cost and effective markup increase.
   (d) If $F_{ijmt}^{Q'} = 0$, the optimal lending limit highly depends on the convexity of the marginal cost function. If $c_{ijmt}^{b} = 0$, the lending amount is unlimited for the approved customers. If $c_{ijmt}^{b} > 0$, however, the optimal lending limit decreases as $\mathcal{M}_{ijmt}^{Q}$ increases.
   (e) If $F_{ijmt}^{Q'} > 0$, the optimal lending limit decreases as $\mathcal{M}_{ijmt}^{Q}$ increases.
   (f) If $F_{ijmt}^{Q'} < 0$, the optimal lending limit is uncertain since the decreasing default risk compensates the increasing marginal cost of lending more loan amount. Specifically, the optimal lending limit increases if $F_{ijmt}^{Q'} < -c_{ijmt}^{b}/2P_{ijmt}$ and decreases otherwise.

2. The change of the degree of asymmetric information by fixing the market structure:
   (a) The optimal price decreases as $F_{ijmt}^{P'}$ increases under the given default rate F_{ijmt}.
   (b) The optimal price is uncertain if both the default rate F_{ijmt} and $F_{ijmt}^{P'}$ increases. The increasing default risk offsets the high price from the high default level.
   (c) The optimal Q_{ijmt} increases as $F_{ijmt}^{Q'}$ decreases under the given default rate F_{ijmt}.
   (d) The optimal Q_{ijmt} is uncertain if the default rate F_{ijmt} decreases but $F_{ijmt}^{Q'}$ increases. The rising default risk offsets the large loan amount from a low default level.

The welfare change corresponding with the optimal interest rate and the lending limit is uncertain in most cases. The social welfare measured by consumer surplus may not necessarily indicate an explicit welfare gain or loss under dual screening devices. For example, if $F_{ijmt}^{P'} > 0$ and $F_{ijmt}^{Q'} > 0$, the loan amount decreases, but the loan price may also be lower when $F_{ijmt}^{P'}$ is sufficiently high. Similarly, if $F_{ijmt}^{P'} < 0$ and $F_{ijmt}^{Q'} < 0$, the higher loan price and larger lending limit cannot confirm a welfare gain or loss. The direction of welfare change belongs to an empirical question. We introduce an econometric specification to examine the model in the next section.

4 Econometric Specifications

This section develops an econometric specification for structural estimation of the loan demand-supply parameters. We first predict the counterfactual loan prices and the amounts. Based on consumer characteristics, loan demand, and bank-specific variables, we establish a simulated likelihood function of consumer choice probabilities and estimate the consumer loan demand and default function parameters.

4.1 Price and Quantity Prediction

The estimation of the model described in Section 3 requires to recovering variables that are inherently unobserved from our dataset. As previously mentioned in the literature, including Crawford et al. (2018) and Ioannidou et al. (2021), the researcher cannot observe the offered loan price and amount except the bank that the customer made a contract with. One big challenge in estimation is to predict the loan price and amount provided by other active banks.

We begin with the realized loan contracts, focusing on the individuals with more than one loan contract. The repeated observations enable us to control individual-level heterogeneity on their decision of the loan amount and the interest rate. We specify a linear panel model specification of P_{ijmt} and Q_{ijmt} with fixed effects.

$$\begin{aligned} P_{ijmt} &= Z'_{ijmt}\beta^P + \varepsilon_i^P + u_{ijmt}^P \\ Q_{ijmt} &= Z'_{ijmt}\beta^Q + \varepsilon_i^Q + u_{ijmt}^Q, \end{aligned}$$

assuming that g_j^P and g_j^Q are linear functions. The regressors Z_{ijmt} include time-varying regressors (consumer and bank characteristics) and four-way fixed effects: individual, bank, region, and time dummies. For example, B_{ijmt} for $j = 1, \ldots, J_{mt}$ denotes the bank fixed effects. Then, $B_{ijmt} = (0, \ldots, 0)' \in \mathbb{R}^{J_{mt}-1}$ if $j = 1$ and $B_{ijmt} = (1, 0, \ldots, 0) \in \mathbb{R}^{J_{mt}-1}$ if $j = 2$. The individual-specific price-quantity adjustments ε_i^P and ε_i^Q are correlated with the individual-specific loan demand ε_i^D and default risk ε_i^F following the specification (4). ε_i^P and ε_i^Q follow a mean-zero normal distribution and do not depend on Z_{ijmt}, thereby separately identified from the individual fixed effects.

u_{ijmt}^P and u_{ijmt}^Q present the unexplained part of (P_{ijmt}, Q_{ijmt}) after controlling Z_{ijmt}. We consider the clustered u_{ijmt}^P and u_{ijmt}^Q across banks $j = 1, \ldots, J_{mt}$ since the banks' decisions are highly correlated. Banks share the same soft information about customers, and the oligopolistic market structure generates the correlation of (P_{ijmt}, Q_{ijmt}) across banks. For example, the correlation coefficient of u_{i1mt}^P and u_{i2mt}^P measures the closeness of bank 1 and bank 2 in their loan interest rate decisions for the same customer i. We do not assume the correlation of u_{ijmt}^P and $u_{ij'mt}^Q$ for any (j, j') since the correlation of ε_i^P and ε_i^Q captures the correlation of loan price and amount.

Define vectors $\mathbb{P}_{imt} \equiv (P_{i1mt}, \ldots, P_{iJ_{mt}mt})'$ and $\mathbb{Q}_{imt} \equiv (Q_{i1mt}, \ldots, Q_{iJ_{mt}mt})'$ for the loan price and loan amount. The system of equations $(\mathbb{P}'_{imt}, \mathbb{Q}'_{imt})'$ follow a multivariate linear panel data model. Under the model specification,

$$\begin{pmatrix} \mathbb{P}_{imt} \\ \mathbb{Q}_{imt} \end{pmatrix} = \begin{pmatrix} \mathbb{Z}_{imt} & 0 \\ 0 & \mathbb{Z}_{imt} \end{pmatrix} \begin{pmatrix} \beta^P \\ \beta^Q \end{pmatrix} + \begin{pmatrix} \varepsilon_i^P \\ \varepsilon_i^Q \end{pmatrix} + \begin{pmatrix} \mathbb{U}_{imt}^P \\ \mathbb{U}_{imt}^Q \end{pmatrix}, \tag{7}$$

where $\mathbb{Z}_{imt} = [Z_{i1mt}, \ldots, Z_{iJ_{mt}mt}]'$, $\mathbb{U}^P_{imt} = \left(u^P_{i1mt}, \ldots, u^P_{iJ_{mt}mt}\right)'$ and $\mathbb{U}^Q_{imt} = \left(u^Q_{i1mt}, \ldots, u^Q_{iJ_{mt}mt}\right)'$. $\mathbb{U}^P_{imt}$ and $\mathbb{U}^Q_{imt}$ follow a mean-zero joint normal distribution with an unknown variance-covariance matrix. Assuming that unobservable individual preference and the bank-specific contract adjustments are independent,

$$\begin{pmatrix} \varepsilon^P_i \\ \varepsilon^Q_i \end{pmatrix} + \begin{pmatrix} \mathbb{U}^P_{imt} \\ \mathbb{U}^Q_{imt} \end{pmatrix} \sim N\left(0, \Omega_1\right) + N\left(0, \Omega_2\right) \sim N\left(0, \Omega_1 + \Omega_2\right),$$

where Ω_1 is a variance-covariance matrix of ε^P_i and ε^Q_i. The covariance of ε^P_i and ε^Q_i means the correlation of P_{ijmt} and Q_{ijmt} adjustments within the same individual. The realized values of ε^P_i and ε^Q_i do not vary across time, and the values are constant for the same individual's other loan contracts. Ω_2 is a variance-covariance matrix of $\mathbb{U}^P_{imt}$ and $\mathbb{U}^Q_{imt}$. The structure reflects the competition effect among banks. For example, consider a case with two banks 1 and 2, then,

$$\Omega_2 = \begin{pmatrix} \sigma^2_{P1} & \sigma_{P1,P2} & 0 & 0 \\ \sigma_{P1,P2} & \sigma^2_{P2} & 0 & 0 \\ 0 & 0 & \sigma^2_{Q1} & \sigma_{Q1,Q2} \\ 0 & 0 & \sigma_{Q1,Q2} & \sigma^2_{Q2} \end{pmatrix}, \Omega_1 + \Omega_2 = \begin{pmatrix} 1+\sigma^2_{P1} & \sigma_{P1,P2} & \rho_{PQ} & 0 \\ \sigma_{P1,P2} & 1+\sigma^2_{P2} & 0 & \rho_{PQ} \\ \rho_{PQ} & 0 & 1+\sigma^2_{Q1} & \sigma_{Q1,Q2} \\ 0 & \rho_{PQ} & \sigma_{Q1,Q2} & 1+\sigma^2_{Q2} \end{pmatrix},$$

where Ω_2 is a diagonal matrix if all banks independently and separately determine the loan price and amount. $\sigma_{P1,P2}$ and $\sigma_{Q1,Q2}$ present clustering effects.

The estimation of β^P and β^Q follows a high-dimensional linear panel fixed effects estimator after standardizing observable covariates. The maximum likelihood estimator based on Fernández-Val and Weidner (2018) provides estimates for β^P, β^Q and unknown components of $\Omega_1 + \Omega_2$. We assumed $\sigma_{Pi,Pj} = 0$ if no customer took a personal loan from both banks i and j. The predicted loan price and amount, including the counterfactual ones, are $\hat{\mathbb{P}}_{imt} = \mathbb{Z}_{imt}\hat{\beta}^P$ and $\hat{\mathbb{Q}}_{imt} = \mathbb{Z}_{imt}\hat{\beta}^Q$ for customer i.

The parameter vectors $\hat{\beta}^P$ and $\hat{\beta}^Q$ do not include the individual-level fixed effects for customers with only one personal loan contract. We simulate the fixed effect for the customer i whose single loan history is with bank j. The fixed effect estimates are $\widehat{FE}^P_i = P_{ijmt} - Z'_{ijmt}\hat{\beta}^P - \left(\tilde{\varepsilon}^P_i + \tilde{u}^P_{ijmt}\right)$ and $\widehat{FE}^Q_i = Q_{ijmt} - Z'_{ijmt}\hat{\beta}^Q - \left(\tilde{\varepsilon}^Q_i + \tilde{u}^Q_{ijmt}\right)$, where $\tilde{\varepsilon}^P_i + \tilde{u}^P_{ijmt}$ and $\tilde{\varepsilon}^Q_i + \tilde{u}^Q_{ijmt}$ are simulated by a random draw from the estimated normal distribution $N\left(0, \hat{\Omega}_1 + \hat{\Omega}_2\right)$. Then for the customer i with single loan history, $\hat{\mathbb{P}}_{imt} = \mathbb{Z}_{imt}\hat{\beta}^P + \widehat{FE}^P_i$ and $\hat{\mathbb{Q}}_{imt} = \mathbb{Z}_{imt}\hat{\beta}^Q + \widehat{FE}^Q_i$.

4.2 Demand and Default

The next step of the estimation follows a standard backward induction approach. The analysis comprises two phases. In the first stage, based on the results in price-quantity predictions, we derive the propensity of demand as a function of structural parameters in equation (2). The default probability is also a function of structural parameters based on equation (3). We will use the simulated maximum likelihood method to recover structural parameters in demand and default equations. In the second stage, we compute equations (5) and (6) to retrieve the additional structural parameters on the supply side.

We construct a likelihood for the consumer demand probability assuming the conditional logit specification. Denote Pr^D_{ijmt} by the probability of customer i to choose bank j in market m in time t, conditional on the values of P_{ijmt} and Q_{ijmt}. Then, equation (2) is

$$\begin{aligned} U^D_{ijmt} =& X^{D'}_{jmt}\beta^D + \xi^D_{jmt} + \alpha^D_1 P_{ijmt} + \alpha^D_2 Q_{ijmt} + Y^{D'}_{ijmt}\gamma^D + \varepsilon^D_i + \nu_{ijmt} \\ =& X^{D'}_{jmt}\beta^D + \xi^D_{jmt} + \alpha^D_1 \left(Z'_{ijmt}\hat{\beta}^P + \tilde{\varepsilon}^P_i + \tilde{u}^P_{ijmt} \right) \\ &+ \alpha^D_2 \left(Z'_{ijmt}\hat{\beta}^Q + \tilde{\varepsilon}^Q_i + \tilde{u}^Q_{ijmt} \right) + Y^{D'}_{ijmt}\gamma^D + \varepsilon^D_i + \nu_{ijmt}, \end{aligned}$$

thereby the utility function follows a linear function of covariates. Suppose $Z_{ijmt} = \left(Z^P_{jmt}, Z^P_{ijmt}\right) = \left(Z^Q_{jmt}, Z^Q_{ijmt}\right)$ and the corresponding parameter vectors $\hat{\beta}^P = \left(\hat{\beta}^P_1, \hat{\beta}^P_2\right)$ and $\hat{\beta}^Q = \left(\hat{\beta}^Q_1, \hat{\beta}^Q_2\right)$ to separate the bank-specific covariates and individual-level covariates. The first set of variables includes the bank deposit rate, total deposit, and debt ratio. The second group of covariates includes the number of bank branches around the customer's residence, the customer's consumption pattern, age, income, and other consumer-level characteristics. Arranging terms again,

$$\begin{aligned} U^D_{ijmt} =& \underbrace{\left(X^{D'}_{jmt}\beta^D + \xi^D_{jmt} + Z^{P'}_{jmt}\eta^P_1 + Z^{Q'}_{jmt}\eta^Q_1 \right)}_{\equiv \zeta^D_{jmt}} + Z^{P'}_{ijmt}\eta^P_2 + Z^{Q'}_{ijmt}\eta^Q_2 + Y^{D'}_{ijmt}\gamma^D \\ &+ \left(\tilde{\varepsilon}^P_i + \tilde{u}^P_{ijmt} \right)\alpha^D_1 + \left(\tilde{\varepsilon}^Q_i + \tilde{u}^Q_{ijmt} \right)\alpha^D_2 + \varepsilon^D_i + \nu_{ijmt}, \end{aligned}$$

where $\eta^P_1 = \alpha^D_1\hat{\beta}^P_1$, $\eta^Q_1 = \alpha^D_2\hat{\beta}^Q_1$, $\eta^P_2 = \alpha^D_1\hat{\beta}^P_2$, and $\eta^Q_2 = \alpha^D_2\hat{\beta}^Q_2$. The individual level price-quantity adjustments $\tilde{\varepsilon}^P_i$ and $\tilde{\varepsilon}^Q_i$ are drawn from $N\left(0, \hat{\Omega}_1\right)$. The correlated random errors $\tilde{u}^P_{ijmt}$ and $\tilde{u}^Q_{ijmt}$ for $j = 1, \ldots, J_{mt}$ are drawn from $N\left(0, \hat{\Omega}_2\right)$. Let ν_{ijmt} follow the Type 1 extreme distribution in the spirit of the conditional logit approximation. Then,

$$\begin{aligned} Pr^D_{ijmt} &= P\left(U^D_{ijmt} \geq U^D_{ij'mt} | X^D_{mt}, Y_{imt} \right) \text{ for all } j' \neq j \\ &= \int \left[\frac{\exp\left(\zeta^D_{jmt} + Y'_{ijmt}\eta^D + \varepsilon^D_i \right)}{1 + \sum_{\iota=1}^{J_{mt}} \exp\left(\zeta^D_{\iota mt} + Y'_{i\iota mt}\eta^D + \varepsilon^D_i \right)} \right] f\left(\varepsilon^D_i | \varepsilon^P_i, \varepsilon^Q_i \right) d\varepsilon^D_i, \end{aligned} \tag{8}$$

where $X^D_{mt} = \left(X^{D'}_{1mt}, \ldots, X^{D'}_{J_{mt}mt} \right)'$ include the aggregate bank specific observables, and $Y_{imt} = \left(Y'_{i1mt}, \ldots, Y'_{iJ_{mt}mt} \right)'$ with $Y_{ijmt} = \left(Z^{P'}_{ijmt}, Z^{Q'}_{ijmt}, Y^{D'}_{ijmt}, \left(\tilde{\varepsilon}^P_i + \tilde{u}^P_{ijmt} \right), \left(\tilde{\varepsilon}^Q_i + \tilde{u}^Q_{ijmt} \right) \right)'$ is the vector of individual-bank specific regressors. X^D_{mt} and Y_{imt} also include explanatory variables to predict the loan price and amount. The corresponding parameter vector of Y_{ijmt} is $\eta^D = \left(\eta^{P'}_2, \eta^{Q'}_2, \gamma^{D'}, \alpha^D_1, \alpha^D_2 \right)'$ in equation (8). The bank j's information set for customer i is $\left(X^D_{mt}, Y_{imt} \right)$ for $j = 1, \ldots, J_{mt}$. The estimation of ζ^D_{jmt} follows the contraction method suggested by Berry et al. (1995). $f\left(\varepsilon^D_i | \varepsilon^P_i, \varepsilon^Q_i \right)$

is the conditional normal density function following the joint normal specification in (4). Define

$$\Sigma_{PQ} \equiv \begin{pmatrix} 1 & \rho_{PQ} \\ \rho_{PQ} & 1 \end{pmatrix}, \; \Sigma_{D,PQ} \equiv \begin{pmatrix} \rho_{PD}\sigma_D \\ \rho_{QD}\sigma_D \end{pmatrix}$$

$$\Sigma_{PQD} \equiv \begin{pmatrix} 1 & \rho_{PQ} & \rho_{PD}\sigma_D \\ \rho_{PQ} & 1 & \rho_{QD}\sigma_D \\ \rho_{PD}\sigma_D & \rho_{QD}\sigma_D & \sigma_D^2 \end{pmatrix}, \; \Sigma_{F,PQD} \equiv \begin{pmatrix} \rho_{PF} \\ \rho_{QF} \\ \rho_{DF}\sigma_D \end{pmatrix},$$

then the conditional distribution $\varepsilon_i^D|\varepsilon_i^P, \varepsilon_i^Q$ follows $N\left(\mu_{PQ}, \sigma_{PQ}^2\right)$, where $\mu_{PQ} = \Sigma'_{D,PQ}\Sigma_{PQ}^{-1}\begin{pmatrix} \varepsilon_i^P \\ \varepsilon_i^Q \end{pmatrix}$ and $\sigma_{PQ}^2 = \sigma_D^2 - \Sigma'_{D,PQ}\Sigma_{PQ}^{-1}\Sigma_{D,PQ}$.

We compute the default probability in similar way. The conditional distribution of ε_i^F on $\left(\varepsilon_i^P, \varepsilon_i^Q, \varepsilon_i^D\right)$ follows a joint normal distribution $\varepsilon_i^F|\varepsilon_i^P, \varepsilon_i^Q, \varepsilon_i^D \sim N\left(\mu_{PQD}, \sigma_{PQD}^2\right)$, where $\mu_{PQD} = \Sigma'_{F,PQD}\Sigma_{PQD}^{-1}\begin{pmatrix} \varepsilon_i^P \\ \varepsilon_i^Q \\ \varepsilon_i^D \end{pmatrix}$ and $\sigma_{PQD}^2 = 1 - \Sigma'_{F,PQD}\Sigma_{PQD}^{-1}\Sigma_{F,PQD}$. Then the default probability conditional on demand is

$$\begin{aligned} Pr_{ijmt,F=1|D=1}^F &= P\left(U_{ijmt}^F \geq 0 | D = 1, X_{jmt}^F, P_{ijmt}, Q_{ijmt}, Y_{ijmt}^F, \varepsilon_i^P, \varepsilon_i^Q\right) \\ &= \int \Phi_{\varepsilon_i^F|\varepsilon_i^P,\varepsilon_i^Q,\varepsilon_i^D}\left(\frac{X_{jmt}^{F\prime}\beta^F + \alpha_P^F P_{ijmt} + \alpha_Q^F Q_{ijmt} + Y_{ijmt}^{F\prime}\gamma^F + \mu_{PQD}}{\sqrt{\sigma_{PQD}^2}}\right) \\ &\quad \times f\left(\varepsilon_i^D|\varepsilon_i^P, \varepsilon_i^Q, D = 1\right) d\varepsilon_i^D, \end{aligned}$$

where Φ is the CDF of a standard normal distribution. Based on the demand and default probabilities, we derive the following log-likelihood function:

$$\log \mathcal{L} = \sum_{i=1}^{n} d_{ijmt}\left(\log\left(Pr_{ijmt}^D\right) + f_{ijmt}\log\left(Pr_{ijmt}^F\right) + (1 - f_{ijmt})\log\left(1 - Pr_{ijmt}^F\right)\right), \tag{9}$$

where d_{ijmt} is a dummy variable if consumer i chooses bank j in market m in time t, and f_{ijmt} is the dummy variable indicating the default. Solving the log-likelihood function in equation (9), we estimate the structural parameters in demand and default functions. Specifically, the simulated maximum likelihood estimation derives the estimates on on ζ_{jmt}^D, η^D, β^F, α_P^F, α_Q^F, and variance-covariance matrix parameters σ_D and ρ_{DF}. The fitted value ζ_{jmt}^D that contains an additional parameter β^D, and we can estimate β^D in the second stage.

Denote $\hat{\zeta}_{jmt}^D$ by the fitted value of ζ_{jmt}^D derived from the simulated maximum likelihood estimation of equation (9). Recall

$$\begin{aligned} \zeta_{jmt}^D &= X_{jmt}^{D\prime}\beta^D + \xi_{jmt}^D + Z_{jmt}^{P\prime}\eta_1^P + Z_{jmt}^{Q\prime}\eta_1^Q \\ &= \zeta_0 + X_{jmt}^{D\prime}\beta^D + Z_{jmt}^{P\prime}\eta_1^P + Z_{jmt}^{Q\prime}\eta_1^Q + \tilde{\zeta}_{jmt}, \end{aligned}$$

where the bank specific fixed effect can be decomposed by $\xi^D_{jmt} = \zeta_0 + \tilde{\zeta}_{jmt}$. ζ_0 is the mean of the fixed effect and $\tilde{\zeta}_{jmt}$ is the mean-zero structural error term. We treat the equation as a linear regression model with the dependent variable $\hat{\zeta}^D_{jmt}$. The standard least-squares estimator provides estimates for the bank specific parameter β^D.

4.3 Supply Side

The estimation on the supply side is to fit the data into the model in Section 3.3. Recall that the equilibrium loan price and quantity equations we obtained from the first-order conditions have parameters regarding the cost structure of banks. Using both equilibrium equations, we compute the individual-level cost function parameters c^a_{ijmt} and c^b_{ijmt}. Solving two equations (5) and (6) with two unknowns c^a_{ijmt} and c^b_{ijmt}, we find the following equations:

$$\begin{aligned} c^a_{ijmt} &= \left(2 + \frac{Q_{ijmt}}{\mathcal{M}^Q_{ijmt}}\right)\left(1 - F_{ijmt} - F^{P'}_{ijmt}P_{ijmt}\right)\mathcal{M}^P_{ijmt} + P_{ijmt}\left(1 - F_{ijmt} + F^{Q'}_{ijmt}Q_{ijmt}\right) \\ c^b_{ijmt} &= 2\left(\left(\frac{1}{\mathcal{M}^Q_{ijmt}} + \frac{1}{Q_{ijmt}}\right)\left(1 - F_{ijmt} - F^{P'}_{ijmt}P_{ijmt}\right)\mathcal{M}^P_{ijmt} - P_{ijmt}F^{Q'}_{ijmt}\right), \end{aligned} \tag{10}$$

thereby the bank's cost structure is identified by the demand and default parameters.

We estimate the cost parameters using the plug-in estimators. The components of c^a_{ijmt} and c^b_{ijmt} include $\mathcal{M}^P_{ijmt} = -\frac{Pr^D_{ijmt}}{\partial Pr^D_{ijmt}/\partial P_{ijmt}}$, $\mathcal{M}^Q_{ijmt} = \frac{Pr^D_{ijmt}}{\partial Pr^D_{ijmt}/\partial Q_{ijmt}}$, F_{ijmt}, $F^{P'}_{ijmt}$, and $F^{Q'}_{ijmt}$. We use simulation-based estimators to approximate all components. Let the number of random draws $S = 1000$ be a sufficiently large number.

$$\begin{aligned} \hat{Pr}^D_{ijmt} &= \frac{1}{S}\sum_{s=1}^{S}\left[\frac{\exp\left(\hat{\zeta}^D_{jmt} + Y'_{ijmt}\hat{\eta}^D + \varepsilon^D_{i,s}\right)}{1 + \sum_{\iota=1}^{J_{mt}}\exp\left(\hat{\zeta}^D_{\iota mt} + Y'_{i\iota mt}\hat{\eta}^D + \varepsilon^D_{i,s}\right)}\right] \\ \frac{\partial \hat{Pr}^D_{ijmt}}{\partial P_{ijmt}} &= \frac{1}{S}\sum_{s=1}^{S}\frac{\partial}{\partial P_{ijmt}}\left[\frac{\exp\left(\hat{\zeta}^D_{jmt} + Y'_{ijmt}\hat{\eta}^D + \varepsilon^D_{i,s}\right)}{1 + \sum_{\iota=1}^{J_{mt}}\exp\left(\hat{\zeta}^D_{\iota mt} + Y'_{i\iota mt}\hat{\eta}^D + \varepsilon^D_{i,s}\right)}\right] \\ \frac{\partial \hat{Pr}^D_{ijmt}}{\partial Q_{ijmt}} &= \frac{1}{S}\sum_{s=1}^{S}\frac{\partial}{\partial Q_{ijmt}}\left[\frac{\exp\left(\hat{\zeta}^D_{jmt} + Y'_{ijmt}\hat{\eta}^D + \varepsilon^D_{i,s}\right)}{1 + \sum_{\iota=1}^{J_{mt}}\exp\left(\hat{\zeta}^D_{\iota mt} + Y'_{i\iota mt}\hat{\eta}^D + \varepsilon^D_{i,s}\right)}\right], \end{aligned}$$

where $\varepsilon^D_{i,s}$ is a random draw from the estimated conditional normal distribution $\varepsilon^D_i|\varepsilon^P_i, \varepsilon^Q_i$. The default probability and the derivatives are estimated in a similar way.

$$\hat{F}_{ijmt} = \frac{1}{S}\sum_{s=1}^{S}\left[\Phi_{\varepsilon_i^F|\varepsilon_i^P,\varepsilon_i^Q,\varepsilon_i^D}\left(\frac{X_{jmt}^{F'}\hat{\beta}^F + \hat{\alpha}_P^F P_{ijmt} + \hat{\alpha}_Q^F Q_{ijmt} + Y_{ijmt}^{F'}\hat{\gamma}^F + \hat{\mu}_{PQD,s}}{\sqrt{\hat{\sigma}_{PQD}^2}}\right)\right]$$

$$\hat{F}_{ijmt}^{P'} = \frac{1}{S}\sum_{s=1}^{S}\frac{\hat{\alpha}_P^F}{\sqrt{\hat{\sigma}_{PQD}^2}}\left[\phi_{\varepsilon_i^F|\varepsilon_i^P,\varepsilon_i^Q,\varepsilon_i^D}\left(\frac{X_{jmt}^{F'}\hat{\beta}^F + \hat{\alpha}_P^F P_{ijmt} + \hat{\alpha}_Q^F Q_{ijmt} + Y_{ijmt}^{F'}\hat{\gamma}^F + \hat{\mu}_{PQD,s}}{\sqrt{\hat{\sigma}_{PQD}^2}}\right)\right]$$

$$\hat{F}_{ijmt}^{Q'} = \frac{1}{S}\sum_{s=1}^{S}\frac{\hat{\alpha}_Q^F}{\sqrt{\hat{\sigma}_{PQD}^2}}\left[\phi_{\varepsilon_i^F|\varepsilon_i^P,\varepsilon_i^Q,\varepsilon_i^D}\left(\frac{X_{jmt}^{F'}\hat{\beta}^F + \hat{\alpha}_P^F P_{ijmt} + \hat{\alpha}_Q^F Q_{ijmt} + Y_{ijmt}^{F'}\hat{\gamma}^F + \hat{\mu}_{PQD,s}}{\sqrt{\hat{\sigma}_{PQD}^2}}\right)\right],$$

where $\hat{\mu}_{PQD,s}$ is a function of $\varepsilon_{i,s}^D$. The simulated values $\varepsilon_{i,s}^D$ are random draws from the probability distribution with the density function $f\left(\varepsilon_i^D|\varepsilon_i^P,\varepsilon_i^Q,D=1\right)$. We observe the conditioning values $D=1$, $\varepsilon_i^P=\tilde{\varepsilon}_i^P$, and $\varepsilon_i^Q=\tilde{\varepsilon}_i^Q$ from the price-quantity prediction part. We predict the cost parameters for other banks using the predicted values $\left(\hat{P}_{ijmt},\hat{Q}_{ijmt}\right)$.

$$\hat{c}_{ijmt}^a = \left(2+\frac{\hat{Q}_{ijmt}}{\hat{\mathcal{M}}_{ijmt}^Q}\right)\left(1-\hat{F}_{ijmt}-\hat{F}_{ijmt}^{P'}\hat{P}_{ijmt}\right)\hat{\mathcal{M}}_{ijmt}^P + \hat{P}_{ijmt}\left(1-\hat{F}_{ijmt}+\hat{F}_{ijmt}^{Q'}\hat{Q}_{ijmt}\right)$$

$$\hat{c}_{ijmt}^b = 2\left(\left(\frac{1}{\hat{\mathcal{M}}_{ijmt}^Q}+\frac{1}{\hat{Q}_{ijmt}}\right)\left(1-\hat{F}_{ijmt}-\hat{F}_{ijmt}^{P'}\hat{P}_{ijmt}\right)\hat{\mathcal{M}}_{ijmt}^P - \hat{P}_{ijmt}\hat{F}_{ijmt}^{Q'}\right),$$

where $\hat{\mathcal{M}}_{ijmt}^P$, $\hat{\mathcal{M}}_{ijmt}^Q$, $\hat{F}_{ijmt}$, $\hat{F}_{ijmt}^{P'}$, and $\hat{F}_{ijmt}^{Q'}$ are predicted values under $\left(\hat{P}_{ijmt},\hat{Q}_{ijmt}\right)$. The predicted total cost and the marginal cost of bank j to offer a loan contract (P_{ijmt},Q_{ijmt}) for a customer i are

$$\hat{C}_{ijmt}\left(Q_{ijmt}\right) = \hat{c}_{ijmt}^a Q_{ijmt} + \frac{1}{2}\hat{c}_{ijmt}^b Q_{ijmt}^2$$

$$\widehat{MC}_{ijmt}\left(Q_{ijmt}\right) = \hat{c}_{ijmt}^a + \hat{c}_{ijmt}^b Q_{ijmt},$$

using $\hat{c}_{ijmt}^a$ and $\hat{c}_{ijmt}^b$. The estimated marginal cost in the loan supply market approximates the amount of markup caused by oligopolistic competition between lenders. In addition to the strategic interaction parameters that indicate market power among lenders, markup estimates provide how market power affects the extra benefit of lenders in a personal loan market.

5 Estimation Results

This section presents the estimation procedure and results. The data covers personal loans initiated between 2013 and 2019 in South Korea. We first standardize variables in Table 1 using the sample means and standard deviations. Then we predict the counterfactual interest rates and loan amounts using the estimated parameters from the maximum likelihood estimation. We estimate $\left(\varepsilon_i^P,\varepsilon_i^Q\right)$ for customers with multiple personal loan contracts or draw $\left(\varepsilon_i^P,\varepsilon_i^Q\right)$ for single-contract customers.

Next, $\left(\mathbb{U}^P_{imt}, \mathbb{U}^Q_{imt}\right)$ also follows a joint normal distribution. The correlation of $u^P_{ij_1mt}$ and $u^P_{ij_2mt}$ (or $u^Q_{ij_1mt}$ and $u^Q_{ij_2mt}$) is identified by customers who took out multiple loans from both banks j_1 and j_2. We assumed that the correlation of $u^P_{ij_1mt}$ and $u^P_{ij_2mt}$ is zero if no individual took out personal loans from both banks j_1 and j_2. In total, 15 pairs of banks out of 171 possible pairs have zero correlations.

The demand and default utility function parameters include coefficients for consumer-specific variables, bank-specific variables, and individual-bank interactive variables. The demand utility depends on the current consumption and the possibility of consumption smoothing. Current job type (blue collar or white collar), income, debt, and the consumption level measured by credit/debit card spending approximate the utility from consumption. The credit score and age variables inform the individual's consumption-smoothing flexibility. The number of banks around the individual's residence also affects the demand utility, indicating how approachable each bank is to the customer.

The default utility shares the same variables with the demand utility but has additional variables since bank characteristics affect the loan taker's default decision. We use the bank's overall debt ratio, return on equity, average loan maturity, year and region fixed effects, and bank size dummies (large, medium, regional, and online banks) to capture the heterogeneity across banks. The simulated log-likelihood function in equation (9) jointly estimates demand and default utility function parameters.

We first estimate the structural model following our main specification. Then we compare our structural estimates with the estimates from the alternative specification that the loan price is the only screening device (Crawford et al. (2018)).

5.1 Main Estimation

Table 2 presents the summary of estimates. Panel A displays the customer's loan demand function parameters, and Panel B shows default parameters. We included the three-way fixed effects by bank-region-year dummies for each stage to control unobserved heterogeneity.

Panels A and B in Table 2 show the estimates of demand and default function parameters. In both panels, all explanatory variables are highly significant except the customer's income in demand function. As expected, the loan demand for bank j has positive correlations with the customer's age, previous consumption level, the current debt (DTI), credit score, and the number of bank j branches around the customer's residence. The result matches the previous literature analyzing the customer's loan demand structure (e.g., Perraudin and Sorensen (1992)). The customer with a lower personal saving rate is more likely to take out a loan. The customer's yearly income negatively correlates with the loan demand, implying that a high-income customer can flexibly respond to the liquidity demand shocks without taking out a loan. However, the magnitude of the income effect is not statistically significant.

Table 2 Panel A shows that the customer's credit score positively correlates with the loan demand. We believe the higher loan approval rate for customers with higher credit scores attributes to the positive coefficient of the credit score variable. For individuals who did not finally take out

Table 2: Structural Estimates of the Main Specification

Variable	Estimates	Std. Err
Panel A. Individual-specific Variables in Demand		
Age	0.012^{***}	0.004
Income	-0.006	0.009
Type of Job	0.024^{***}	0.003
Debt to Income (DTI)	0.082^{***}	0.008
Credit Score	0.067^{***}	0.013
Amount of Credit Card Used in Last Year ($)	0.190^{***}	0.048
Amount of Debit Card Used in Last Year ($)	0.083^{***}	0.021
Individual-bank-specific Variables in Demand		
Num. of Banks Neighborhood	0.069^{***}	0.002
Price and Quantity Effects		
Loan Price	-0.005^{***}	0.001
Loan Amount	0.008^{***}	0.001
Panel B. Individual-specific Variables in Default		
Age	0.053^{***}	0.003
Income	-0.037^{***}	0.005
Type of Job	-0.015^{***}	0.003
Credit Score	-0.332^{***}	0.003
Amount of Credit Card Used in Last Year ($)	0.080^{***}	0.010
Amount of Debit Card Used in Last Year ($)	0.255^{***}	0.012
Bank-specific Variables in Default		
Debt Ratio	0.045^{***}	0.004
Return on Equity	0.028^{***}	0.003
Maturity	0.037^{***}	0.003
Price and Quantity Effects		
Loan Price	0.322^{***}	0.004
Loan Amount	0.128^{***}	0.003
Year Dummies included	Yes	Yes
Region Dummies included	Yes	Yes
Bank Size Dummies included	Yes	Yes

Notes: Panel A. in Table 2 reports the parameter estimates of the customer's demand function. And following Panel B. in Table 2 reports the parameter estimates of the customer's default function. Both structural parameters are estimated by the combination of BLP contraction method and simulated maximum likelihood estimation using the structural model of the current paper.

a loan in our dataset, we cannot identify if the individuals did not have to apply for loans or banks refused the loan applications.

The demand estimation also verifies the intuitive relation of personal loan terms and loan demand. The higher loan price or lower loan amount decreases the loan demand. The estimates demonstrate that a bank holding a large share of deposits may attract potential customers by increasing the maximum loan amount to compensate for relatively higher loan pricing. The loan amount is a more critical factor in the loan demand than the loan price in terms of magnitudes.

In Table 2 Panel B, the estimates show the relation of default and consumer characteristics conditional on taking out a loan. The propensity to consumption correlates with a higher chance of default. The group of customers with high income, white-collar occupations, and high credit scores are less likely to default, while the same group's loan demand is relatively lower than other groups. The default estimation also found significantly positive correlations between loan price/amount and the default propensity. As the repayment gets more difficult, the customer is less likely to avoid the default under a negative income shock. According to Crawford et al. (2018), the loan price and amount parameters in the default equation indicate moral hazards. The estimates also show that the loan price's correlation with a default is more substantial than the loan amount side correlation. Compared to reducing the interest rate, increasing the lending limit is safer for banks as the strategy focusing on the loan amount leads to a larger market share and a lower default rate.

Next, Table 3 shows the estimates of the variance-covariance matrix revealing asymmetric information structure. The positive and significant correlation coefficient, ρ_{DF}, in the joint normal distribution of the unobserved propensity to demand loan and unobserved propensity to default shows a substantial adverse selection. The parameter implies that those with a higher tendency to demand a personal loan are more likely to default on the loan, and the implication is consistent with the classical adverse selection story. We also considered various fixed effects to control possible soft information that may affect pricing the loans.

Table 3: Structural Estimates of Information Asymmetry Parameters

Variable	Estimates	Std. Err
Var-Cov Matrix of unobserved heterogeneity		
σ_D	2.621***	0.0006
ρ_{PD}	0.010***	0.0012
ρ_{QD}	-0.001	0.0013
ρ_{PF}	-0.001	0.0031
ρ_{QF}	0.005*	0.0033
ρ_{DF}	0.418***	0.0086

Notes:

$$\begin{pmatrix} \varepsilon_i^P \\ \varepsilon_i^Q \\ \varepsilon_i^D \\ \varepsilon_i^F \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{PQ} & \rho_{PD}\sigma_D & \rho_{PF} \\ \rho_{PQ} & 1 & \rho_{QD}\sigma_D & \rho_{QF} \\ \rho_{PD}\sigma_D & \rho_{QD}\sigma_D & \sigma_D^2 & \rho_{DF}\sigma_D \\ \rho_{PF} & \rho_{QF} & \rho_{DF}\sigma_D & 1 \end{pmatrix}\right)$$

Panel A. in Table 3 reports the result of the variance-covariance matrix of unobserved heterogeneity through the first step estimation of the structural model.

The correlation coefficients ρ_{PD} and ρ_{QD} capture how the hidden loan demand shifter correlates with loan price and amount. According to ρ_{PD} and ρ_{QD} signs, customers with higher loan demand are more likely to accept a higher loan price and a lower loan amount. The estimate for ρ_{PD} is significantly positive, implying heterogeneous loan demand sensitivity to the change in loan price across customers. Even though a high loan price discourages the loan utility, a customer with high ε_i^D is not much influenced compared with customers with low ε_i^D. However, the estimate for ρ_{QD} is negative but not statistically significant. The result implies that the relationship between loan amount and loan demand is relatively more homogeneous across customers. That is, a customer's marginal utility from the loan amount is similar to other customers' marginal utilities regardless of the size of the hidden loan demand factors.

Similarly, ρ_{PF} and ρ_{QF} indicate the correlation between the hidden default risk and loan price/amount. The estimated ρ_{PF} is negative, and ρ_{QF} is positive, but both estimates are not statistically significant. The loan price/quantity effects on default probability are homogenous across customers. Still, our estimation concludes that a customer who took out a more sizable loan has a higher default risk.

5.2 Alternative Specification without Loan Amount

The main estimation in Table 2 verifies that both loan price and amount are critical factors explaining the loan demand and default. This subsection provides an alternative estimation table to present why endogenizing the loan amount in our structural model is essential. Table 8 presents the structural estimates from the specification without considering the lending capacity. We treat the loan amount as an exogenous regressor; thereby, the alternative specification's demand and default utility functions do not have the loan amount as a direct component. The alternative structural model naturally excludes the information asymmetry parameters ρ_{PQ}, ρ_{QD}, and ρ_{QF}.

In Table 8 Panels A and B, we find that the estimates are similar to Table 2 estimates. We focus on how the loan price effects in Panels A and B change across the model specifications. The loan price effect on the demand utility is -0.005 in the main model, while the same effect becomes -0.003 in the alternative specification. The magnitude is only 60% of the main model's price effect. The comparison shows that the estimation without considering the loan amount may underestimate the loan price effect on demand. The main model's implication is that a larger loan amount can compensate a higher loan price. Without the lending capacity constraint, there is no interactive relation between the loan price and amount. Therefore, the loan price effect conditional on the loan amount in the main model specification should be more significant than the estimate from the alternative model. The counterfactual analysis in Section 6 shows that the underestimated loan price effect can predict a misleading welfare effect regarding the bank's marginal cost change.

Table 8 Panel B also presents that the estimated loan price effect on default is less significant than the estimate from the main specification. The current loan price and amount separately influence the customer's default decision. The loan price is the monthly payment for the loan, while the loan amount is the total principal balance for the repayment. Without considering the loan amount, the model may underestimate the effect of a high loan price, usually combined with a low-level borrowing limit.

Table 9 presents a larger degree of adverse selection compared to Table 3. The result verifies that the bank's lending limit has a role as the second screening device, together with the loan price. The estimated correlation between the unobservable loan demand shifter and the hidden default risk ρ_{DF} is around 12% lower under Table 3. Despite the Korean personal loan market suffering from a substantial degree of asymmetric information between borrowers and lenders, we find the lenders keep endeavoring to control the potential default risk using all the available screening devices.

6 Counterfactual Analysis

This section provides counterfactual analyses based on the structural estimation in the previous section. The model estimates enable researchers to examine the impact of new economic policies. A new policy may affect the degree of asymmetric information or change the competition structure of lenders. Since Lester et al. (2019) implies that the policy impact on social welfare is not monotone, we empirically verify the quantitative effects of some counterfactual scenarios.

We consider four counterfactual cases to predict how social welfare varies under the degrees of information asymmetry and market structure change. We follow the first and second counterfactual setups from Crawford et al. (2018) and add two more counterfactuals regarding the shift in market structure. For the first case, we estimate the impact when the adverse selection parameter ρ_{DF} increases by double. Due to a higher correlation between a consumer's unobserved demand propensity and default risk, the bank's screening process becomes more strict. The second case assumes the situation when the bank's marginal cost increases. A bank's marginal cost change corresponds to financial distress or stricter government regulation of the financial market. For the third case, we merge the two largest banks and predict the effect of the market structure change. The case captures the possibility that the market structure has become less competitive. In addition to the merger effect, we also provide a counterfactual outcome by combining the merger effect and the adverse selection effect. Lastly, we decrease the correlation coefficients of loan prices and amounts across banks. Compared to the third scenario, the last experiment considers that the loan prices and quantities assessed by banks become less correlated as the strategic interaction effect is weaker.

For all counterfactual scenarios, we report the average change in loan price, loan amount, demand and default probabilities.

6.1 Adverse Selection

The first counterfactual scenario is to double the adverse selection parameter ρ_{DF}. The increase in ρ_{DF} implies that the demand utility component unobserved by lenders highly correlates with the default utility. We fix the marginal costs of banks and market structure parameters, changing the degree of adverse selection only. The initial impact of the change in ρ_{DF} is the change in $\left(\varepsilon_i^P, \varepsilon_i^Q, \varepsilon_i^D, \varepsilon_i^F\right)$. Following the new joint distribution of $\left(\varepsilon_i^P, \varepsilon_i^Q, \varepsilon_i^D, \varepsilon_i^F\right)$, we simulate new values of $\left(\varepsilon_i^D, \varepsilon_i^F\right)$ and estimate the updated equilibrium demand probability and default rate. Then, the new equilibrium loan price and amount follow

$$\bar{P}_{ijmt} = \frac{\hat{c}^a_{ijmt} + \frac{1}{2}\hat{c}^b_{ijmt}\bar{Q}_{ijmt}}{1 - \bar{F}_{ijmt} + \bar{F}^{P'}_{ijmt}\bar{\mathcal{M}}^P_{ijmt}} + \frac{\left(1 - \bar{F}_{ijmt}\right)\bar{\mathcal{M}}^P_{ijmt}}{1 - \bar{F}_{ijmt} + \bar{F}^{P'}_{ijmt}\bar{\mathcal{M}}^P_{ijmt}}$$

$$\bar{Q}_{ijmt} = \frac{1}{\hat{c}^b_{ijmt}}\left(\hat{\pi}_{ijmt} + \sqrt{\hat{\pi}^2_{ijmt} + 2\hat{c}^b_{ijmt}\bar{\mathcal{M}}^Q_{ijmt}\left(\bar{P}_{ijmt}\left(1 - \bar{F}_{ijmt}\right) - \hat{c}^a_{ijmt}\right)}\right), \tag{11}$$

where $\bar{F}_{ijmt}$ is the new borrower-specific default rate, $\bar{F}^{P'}_{ijmt}$ is the marginal default probability under the new default parameter ρ_{DF}, and $\hat{\pi}_{ijmt} = \bar{P}_{ijmt}\left(1 - \bar{F}_{ijmt} - \bar{F}^{Q'}_{ijmt}\bar{\mathcal{M}}^Q_{ijmt}\right) - \hat{c}^a_{ijmt} - \hat{c}^b_{ijmt}\bar{\mathcal{M}}^Q_{ijmt}$ is the estimated marginal profit of $\bar{Q}_{ijmt}$. $\bar{\mathcal{M}}^P_{ijmt} = -\bar{Pr}^D_{ijmt}/\partial\bar{Pr}^D_{ijmt}/\partial P_{ijmt}$ and $\bar{\mathcal{M}}^Q_{ijmt} = \bar{Pr}^D_{ijmt}/\partial\bar{Pr}^D_{ijmt}/\partial Q_{ijmt}$ are also computed at the new demand level. The marginal cost components $\left(\hat{c}^a_{ijmt}, \hat{c}^b_{ijmt}\right)$ come from equation (10) and are fixed at the current level:

$$\hat{c}^a_{ijmt} = \left(2 + \frac{Q_{ijmt}}{\hat{\mathcal{M}}^Q_{ijmt}}\right)\left(1 - \hat{F}_{ijmt} - \hat{F}^{P'}_{ijmt}P_{ijmt}\right)\hat{\mathcal{M}}^P_{ijmt} + P_{ijmt}\left(1 - \hat{F}_{ijmt} + \hat{F}^{Q'}_{ijmt}Q_{ijmt}\right)$$

$$\hat{c}^b_{ijmt} = 2\left(\left(\frac{1}{\hat{\mathcal{M}}^Q_{ijmt}} + \frac{1}{Q_{ijmt}}\right)\left(1 - \hat{F}_{ijmt} - \hat{F}^{P'}_{ijmt}P_{ijmt}\right)\hat{\mathcal{M}}^P_{ijmt} - P_{ijmt}\hat{F}^{Q'}_{ijmt}\right).$$

Since the equilibrium $\left(\bar{P}_{ijmt}, \bar{Q}_{ijmt}\right)$ are nonlinear mappings, we find the fixed point of equation (11) to obtain $\left(\bar{P}_{ijmt}, \bar{Q}_{ijmt}\right)$. For all observations who decide to take a loan, we can find a fixed point $\left(\bar{P}_{ijmt}, \bar{Q}_{ijmt}\right)$ located at nonnegative values. We compare the new equilibrium loan price and amount with the current loan contract terms. The counterfactual demand and default probabilities also follow the same process, comparing $\bar{Pr}^D_{ijmt}$ and Pr^D_{ijmt} for demand, $\bar{F}_{ijmt}$ and F_{ijmt} for default.

Table 4 presents the impact of higher adverse selection. We report the changes in loan price, loan size (lending capacity), demand probability, and default probability. We show both the percentage point (%P) change and the percentage (%) change for loan price and demand probability. For the loan size variation, the unit measure is a million KRW, which approximately corresponds to a thousand US dollars. The last row of the table shows the percentage point variation of the default probability. Since the baseline default probability is relatively low (0.9~1.0%), the percentage changes are sometimes too significant for values.

We not only report the counterfactual outcome based on our primary model (Table 2) but also the comparable predictions from the alternative model (Table 8). The comparison between the two outcomes highlights the importance of considering endogenous loan amounts to analyze the personal loan market.

According to Table 4, the higher adverse selection causes the personal loan market to be depressed. Due to the increasing default risk, banks decrease the loan size and increase the loan price. On average, banks increase the loan price by 4.14% and reduce the lending amount by 2%. Meanwhile, the disadvantageous loan contract for consumers lowers the demand probability by 5.9%. The average default rate increases since the loan price effects are more critical than the loan size

Table 4: Summary Statistics for Counterfactual Changes in Outcomes: Higher Adverse Selection

Variable	Main Model		Exogenous Loan Amount	
	Mean	SD	Mean	SD
Price variation (%P)	0.120	0.629	0.127	0.574
Price variation (%)	4.141	4.651	4.401	4.681
Loan size variation (million KRW)	−1.066	8.537	4.635	25.424
Loan size variation (%)	−1.989	10.468	7.672	7.367
Demand variation (%P)	−0.681	4.668	−0.358	4.182
Demand variation (%)	−5.901	8.022	−2.931	6.988
Default variation (%P)	0.158	0.204	0.117	0.178

Notes: The variables report the average changes of loan price, loan amount, demand and default probabilities. The "Main Model" columns show predictions based on the main model in Table 2, and the "Exogenous Loan Amount" columns are predictions using the alternative model in Table 8.

effects. Fixing the current market structure, more severe adverse selection may cause a recession in the personal loan market.

Assuming the exogenous loan size, the predicted market outcome is not as bad as the primary model's outcome. The notable difference in prediction is from the estimated loan amount. The main specification expects a 2% decrease in the lending amount, while the alternative specification with only the loan price channel predicts a 7% increase in the loan amount. We believe that the loan-price-only model underestimates the effect of loan size on the default probability. Since the loan price is the only screening device to control the default risk, the banks do not fully consider the impact of more loans on default. Then, banks allow more sizable loans to compensate for higher loan prices and attract more customers. The two counterfactual outcomes explain why the dual screening device model provides more reliable predictions.

6.2 Marginal Cost

The following exercise increases the marginal cost of lenders. We consider that the government regulation lowers the maximum LDR, putting pressure on additional room for liquidity. The customer's annual income also restricts the maximum loan amount that a customer can take out. For example, in the personal loan market of South Korea, the maximum loan amount was 100% of the customer's yearly income in 2021. Before then, the borrowing limit was 200~300% of the annual income, depending on the occupation. These regulations naturally prevent customers from taking too many loans regardless of their credit score or expected default rate.

Since the total cost of lending is potentially a quadratic function of Q_{ijmt}, we increased c^b_{ijmt} so that the consumer-specific marginal cost increases by 10%. The change in c^b_{ijmt} affects both equilibrium loan price and amount in equation (11). The changed price and quantity affect the components of demand and default utilities.

Table 5: Summary Statistics for Counterfactual Changes in Outcomes: Higher Marginal Costs

Variable	Main Model		Exogenous Loan Amount	
	Mean	SD	Mean	SD
Price variation (%P)	1.862	1.183	1.436	0.716
Price variation (%)	40.264	28.064	31.497	22.073
Loan size variation (million KRW)	−1.911	18.293	−5.246	29.721
Loan size variation (%)	−5.013	11.199	−17.626	35.796
Demand variation (%P)	−2.974	4.917	−1.520	4.176
Demand variation (%)	−12.092	18.041	−6.349	6.984
Default variation (%P)	0.568	0.263	0.132	0.156

Notes: The variables report the average changes of loan price, loan amount, demand and default probabilities. The "Main Model" columns show predictions based on the main model in Table 2, and the "Exogenous Loan Amount" columns are predictions using the alternative model in Table 8.

Table 5 finds that the marginal costs substantially influence the equilibrium loan price and the amount. The exponent CB directly affects the marginal cost of lending, and the total cost increases quickly as the loan amount gets larger. As a result, the loan price rises by 40%, and the loan amount decreases by 5%. The considerable change in the loan contract terms discourages customers from taking out a loan. The default rate surges up by 0.568% as the loan price effect dominates the loan size effect. Again, we confirm that the tightened financial market deteriorates the social welfare of the personal loan market.

If we assume the exogenous loan size, the predicted market outcome is slightly different. We observe relatively more minor price changes and more significant loan amount decreases. The overall default rate increase and demand reduction are milder than the primary model's prediction. The difference in predictions comes from the role of the loan size, as the loan size does not affect much for demand and default decisions in the alternative model.

6.3 Merging Banks

The following counterfactual analysis explores the effect of a merger. The largest two banks with the highest market shares in our dataset are B05 and B12. We assume that two banks merge and suggest uniform loan price and amount for the same individual. Each of the other banks maximizes the expected profit function

$$\Pi_{ijmt} = Pr^{D}_{ijmt} \left(P_{ijmt} Q_{ijmt} \left(1 - F_{ijmt}\right) - C_{ijmt}\left(Q_{ijmt}\right)\right),$$

but we assume that B05 and B12 choose $\bar{P}_{i5mt} = \bar{P}_{i12mt}$ and $\bar{Q}_{i5mt} = \bar{Q}_{i12mt}$ that maximize $\Pi_{i5mt} + \Pi_{i12mt}$. Using the baseline equilibrium price and quantity components $\bar{\mathcal{M}}^P_{ijmt}$, $\bar{\mathcal{M}}^Q_{ijmt}$, $\bar{F}^{P'}_{ijmt}$, $\bar{F}^{Q'}_{ijmt}$, and $\bar{F}_{ijmt}$, we compute the new equilibrium price and quantity for banks B05 and B12:

$$
\begin{aligned}
\bar{P}_{i5mt} = \bar{P}_{i12mt} =& \frac{1}{2}\left(\frac{\hat{c}^a_{i5mt} + \frac{1}{2}\hat{c}^b_{i5mt}\bar{Q}_{i5mt} + \left(1 - \bar{F}_{i5mt}\right)\bar{\mathcal{M}}^P_{i5mt}}{1 - \bar{F}_{i5mt} + \bar{F}^{P'}_{i5mt}\bar{\mathcal{M}}^P_{i5mt}}\right) \\
&+ \frac{1}{2}\left(\frac{\hat{c}^a_{i12mt} + \frac{1}{2}\hat{c}^b_{i12mt}\bar{Q}_{i5mt} + \left(1 - \bar{F}_{i12mt}\right)\bar{\mathcal{M}}^P_{i12mt}}{1 - \bar{F}_{i12mt} + \bar{F}^{P'}_{i12mt}\bar{\mathcal{M}}^P_{i12mt}}\right) \\
\bar{Q}_{i5mt} = \bar{Q}_{i12mt} =& \frac{1}{2}\left(\frac{1}{\hat{c}^b_{i5mt}}\left(\bar{\pi}_{i5mt} + \sqrt{\bar{\pi}^2_{i5mt} + 2\hat{c}^b_{i5mt}\bar{\mathcal{M}}^Q_{i5mt}\left(\bar{P}_{i5mt}\left(1 - \bar{F}_{i5mt}\right) - \hat{c}^a_{i5mt}\right)}\right)\right) \\
&+ \frac{1}{2}\left(\frac{1}{\hat{c}^b_{i12mt}}\left(\bar{\pi}_{i12mt} + \sqrt{\bar{\pi}^2_{i12mt} + 2\hat{c}^b_{i12mt}\bar{\mathcal{M}}^Q_{i12mt}\left(\bar{P}_{i5mt}\left(1 - \bar{F}_{i12mt}\right) - \hat{c}^a_{i12mt}\right)}\right)\right),
\end{aligned}
$$

where $\bar{\pi}_{i5mt}$ and $\bar{\pi}_{i12mt}$ depend on the common equilibrium price $\bar{P}_{i5mt}$. The fixed point of the mapping finds the equilibrium price and quantity. After computing the loan contract terms, the rescaled demand utilities generate new equilibrium demand probabilities and market shares. The default rates also change, corresponding to equilibrium loan prices and amounts.

In addition to the sole impact of a merger, we also combine the merger effect and a higher adverse selection. A more serious adverse selection and less competitive market structure may cause some loss in social welfare, Lester et al. (2019) predict that a higher adverse selection may relieve the welfare loss from market power. We increase ρ_{DF} by double and add the merger of banks B05 and B12.

Table 6 summarizes the effect of merging the two largest banks. The first and third columns present the merging impact only, and the second column combines the merge and increase in adverse selection parameter. The situation corresponds to more information asymmetries and market power. As Crawford et al. (2018) already verified, the combination effect does not worsen the market outcome. Due to high default risk, banks cannot easily increase the loan price and amount. Since the loan price's marginal effect on default is relatively larger, the average loan price is lower. Instead, there is a reduction in the average loan size since the reduced loan price can generate extra capacity to lower the lending limit. As a result, the demand probability increases, and the default rate decreases. The overall market outcome is not much worse than before.

The counterfactual scenario has a relatively trivial impact on market outcomes if we consider the merging effect only. We find that the consequences of the merge cause milder effects when we assume the exogenous loan amount. For all factors, including the loan price, loan amount, demand, and default probabilities, the variations under the endogenous loan amount are more substantial.

As we observe more detailed variations across banks, banks B05 and B12 account for most market outcomes variations. For example, other large-sized banks, B01, B07, and B14, decrease the average loan price by 2~5% to compete with B05 and B12. Other smaller banks do not directly

Table 6: Summary Statistics for Counterfactual Changes in Outcomes: Merging Two Largest Banks

Variable	Main Model		Exogenous Loan Amount
	Merge	Merge+Adverse Selection	Merge
Price variation (%P)	0.092	-0.034	0.073
Price variation (%)	2.975	-1.089	2.146
Loan size variation (million KRW)	2.101	-0.550	1.504
Loan size variation (%)	7.373	-0.633	5.277
Demand variation (%P)	0.069	0.274	0.038
Demand variation (%)	0.933	3.705	0.340
Default variation (%P)	0.122	-0.045	0.054

Notes: The variables report the average changes of loan price, loan amount, demand and default probabilities. The "Main Model" columns show predictions based on the main model in Table 2, and the "Exogenous Loan Amount" columns are predictions using the alternative model in Table 8. "Merge" column measures the impact of merger only, while "Merge+Adverse Selection" column presents the combined effect of two largest banks' merger and doubled adverse selection (ρ_{DF}).

compete with B05 and B12, so they instead increase the loan price by 0.6~1.4% or decrease it by 1%. The magnitude of change is not comparable to the large-sized bank's adjustment. However, the model predicts that banks B05 and B12 will increase the loan price by 13% after the merge. We conclude that banks B05 and B12 cause the overall increase in price (2.975%) with high market shares.

6.4 Price/Quantity Shock Correlation

The last counterfactual analysis is to reduce the correlation of loan prices and amounts across banks. The merger analysis in Subsection 6.3 measures the increase in market power by the behavior of the two largest banks. In our model, the loan prices and amounts correlation across banks can also present the market's competitiveness. If the market is perfectly competitive, the marginal costs across banks are equivalent. Thus, the idiosyncratic price shocks do not affect the equilibrium price.

Under the market power, we observe that the idiosyncratic price/quantity shocks are correlated. For example, the estimated price shock correlation between middle-sized and regional banks is 0.26, and the quantity shock correlation is 0.25. The large-sized banks show only little price/quantity correlations with all the middle-sized, regional, and online banks. That is because large banks only compete in the same group. Meanwhile, online banks have only 0.04 and 0.06 price correlations with large and middle-sized banks but 0.10 with regional banks. The shocks are more highly correlated between similar-sized banks.

We reduce all the correlation coefficients of shocks by half and derive new predicted counterfactual loan prices and amounts for unselected banks. The situation implies that the market structure is more competitive. The new loan price and quantity adjust the estimated demand probability and default rate.

Table 7: Summary Statistics for Counterfactual Changes in Outcomes: Lower Correlation of Loan Price/Amount Shocks Across Banks

Variable	Main Model		Exogenous Loan Amount	
	Mean	SD	Mean	SD
Price variation (%P)	-2.512	0.904	-1.213	0.532
Price variation (%)	-9.765	4.649	-12.678	4.171
Loan size variation (million KRW)	0.610	1.745	6.162	45.316
Loan size variation (%)	0.477	0.179	18.260	6.571
Demand variation (%P)	0.069	0.470	0.801	4.182
Demand variation (%)	0.823	0.081	9.553	6.988
Default variation (%P)	-0.039	0.024	0.023	0.017

Notes: The variables report the average changes of loan price, loan amount, demand and default probabilities. The "Main Model" columns show predictions based on the main model in Table 2, and the "Exogenous Loan Amount" columns are predictions using the alternative model in Table 8.

Table 7 presents the predicted market outcome when the correlation of loan contract term shocks is half of the existing correlation. The primary model predicts the intuitive consequences. As the market becomes more competitive, the loan price decreases, and the lending limit increases. As the loan market is more approachable, more consumers demand personal loans, and the default rate slightly decreases.

There is no dramatic difference when comparing the results to the alternative model prediction. However, the alternative model predicts a much more significant increase in the loan size. Although both models derive similar counterfactual outcomes, the endogenous loan amount channel generally implies milder changes.

7 Conclusion

This paper investigates the personal loan market in South Korea equipped with dual screening devices and oligopolistic competition between lenders. Compared to previous literature focusing on a single screening device, our structural model captures the bank's lending capacity and develops a mechanism to explain how the banks endogenously determine the loan price and lending limit. Under the interaction of information asymmetry and market power, the lenders optimally adjust the loan price and amount. Therefore, the banks do not necessarily have to increase the loan price for the risky consumers.

Based on a large-scale dataset of KCB in South Korea, we construct a structural model to describe personal loan demand and default on the demand side and loan supply on the supply side. The estimated model shows the substantial adverse selection in the personal loan market. We also verify that the bank's lending amount restriction to the consumers works as another screening device in reality. Without considering the lending capacity channel, the model estimates different predictions on the personal loan market outcomes.

There are some potential extensions for future work. First, we can analyze the effects of asymmetric information in a dynamic contract context. In particular, we may extend the structural model in which the loan suppliers decide loan price and quantity schedule based on a contingent dynamic contract. Second, we can develop the model by considering the general lending market and figure out how the market outcome varies under the existence of collateral and endogenous lending limits. Lastly, we find a hierarchical structural model where the lending market is divided by primary and secondary financial institutions. The extended model may explain how heterogeneous the market outcome is and its interaction with asymmetric information and market power. We leave the topics for future research.

Acknowledgements: The authors appreciate Bernard Salaniè, Bruce Hansen, and Lorenzo Magnolfi for their helpful suggestions and opinions. We thank Isabella Vesely for excellent research assistance. We also benefited from seminar participants at UW-Milwaukee, Sejong University, Hanyang University, Kyunghee University, Yonsei University, NASMES 2019, AMES 2019, KERIC 2020, KASIO, and discussions with colleagues. Jangsu Yoon acknowledges the financial support from Yoshio Niho Faculty Excellence Fund.

References

Acolin, A., J. Bricker, P. Calem, and S. Wachter (2016). Borrowing constraints and homeownership. *American Economic Review 106*(5), 625–29.

Adams, W., L. Einav, and J. Levin (2009). Liquidity constraints and imperfect information in subprime lending. *American Economic Review 99*(1), 49–84.

Agarwal, S., S. Chomsisengphet, and C. Liu (2016). An empirical analysis of information asymmetry in home equity lending. *Journal of Financial Services Research 49*(1), 101–119.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Azevedo, E. M. and D. Gottlieb (2017). Perfect competition in markets with adverse selection. *Econometrica 85*(1), 67–105.

Berger, A. N. and G. F. Udell (1992). Some evidence on the empirical significance of credit rationing. *Journal of Political Economy 100*(5), 1047–1077.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Bertola, G. and W. Koeniger (2015). Hidden insurance in a moral-hazard economy. *The RAND Journal of Economics 46*(4), 777–790.

Besanko, D. and A. V. Thakor (1987). Collateral and rationing: sorting equilibria in monopolistic and competitive credit markets. *International Economic Review*, 671–689.

Bester, H. (1985). Screening vs. rationing in credit markets with imperfect information. *The American Economic Review 75*(4), 850–855.

Cardon, J. H. and I. Hendel (2001). Asymmetric information in health insurance: evidence from the national medical expenditure survey. *RAND Journal of Economics*, 408–427.

Cawley, J. and T. Philipson (1999). An empirical examination of information barriers to trade in insurance. *American Economic Review 89*(4), 827–846.

Chiappori, P.-A. and B. Salanié (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy 108*(1), 56–78.

Chiappori, P.-A. and B. Salanié (2013). Asymmetric information in insurance markets: Predictions and tests. *Handbook of Insurance*, 397–422.

Cohen, A. (2005). Asymmetric information and learning: Evidence from the automobile insurance market. *Review of Economics and Statistics 87*(2), 197–207.

Crawford, G. S., N. Pavanini, and F. Schivardi (2018). Asymmetric information and imperfect competition in lending markets. *American Economic Review 108*(7), 1659–1701.

[Dataset] (2019). Korea credit bureau personal loan data 2013–2019. Available at Korea Credit Bureau (http://www.koreacb.com/en/), 15, Gukjegeumyung-ro 6-gil, Yeongdeungpo-gu, Seoul, Korea, Accessed May 20, 2019.

Davidoff, T. and G. Welke (2004). Selection and moral hazard in the reverse mortgage market. *working paper*.

Dobbie, W. and P. M. Skiba (2013). Information asymmetries in consumer credit markets: Evidence from payday lending. *American Economic Journal: Applied Economics 5*(4), 256–82.

Duca, J. V. and S. S. Rosenthal (1994). Borrowing constraints and access to owner-occupied housing. *Regional Science and Urban Economics 24*(3), 301–322.

Edelberg, W. (2004). Testing for adverse selection and moral hazard in consumer loan markets. *working paper*.

Egan, M., A. Hortaçsu, and G. Matvos (2017). Deposit competition and financial fragility: Evidence from the us banking sector. *American Economic Review 107*(1), 169–216.

Einav, L., M. Jenkins, and J. Levin (2012). Contract pricing in consumer credit markets. *Econometrica 80*(4), 1387–1432.

Fang, H. and Z. Wu (2018). Multidimensional private information, market structure, and insurance markets. *The RAND Journal of Economics 49*(3), 751–787.

Fernández-Val, I. and M. Weidner (2018). Fixed effects estimation of large-t panel data models. *Annual Review of Economics 10*, 109–138.

Finkelstein, A. and K. McGarry (2006). Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review 96*(4), 938–958.

Finkelstein, A. and J. Poterba (2004). Adverse selection in insurance markets: Policyholder evidence from the uk annuity market. *Journal of Political Economy 112*(1), 183–208.

Golosov, M. and A. Tsyvinski (2007). Optimal taxation with endogenous insurance markets. *The Quarterly Journal of Economics 122*(2), 487–534.

Hausman, J. A. (1983). Specification and estimation of simultaneous equation models. *Handbook of Econometrics 1*, 391–448.

He, D. (2009). The life insurance market: Asymmetric information revisited. *Journal of Public Economics 93*(9-10), 1090–1097.

Ioannidou, V., N. Pavanini, and Y. Peng (2021). Collateral and asymmetric information in lending markets. *Journal of Financial Economics*.

Karlan, D. and J. Zinman (2009). Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica 77*(6), 1993–2008.

Koopmans, T. C. (1949). Identification problems in economic model construction. *Econometrica, Journal of the Econometric Society*, 125–144.

Lester, B., A. Shourideh, V. Venkateswaran, and A. Zetlin-Jones (2019). Screening and adverse selection in frictional markets. *Journal of Political Economy 127*(1), 338–377.

Mahoney, N. and E. G. Weyl (2017). Imperfect Competition in Selection Markets. *The Review of Economics and Statistics 99*(4), 637–651.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica 76*(5), 945–978.

Pauly, M. V. (1978). Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection. In P. Diamond and M. Rothchild (Eds.), *Uncertainty in Economics*, pp. 307–331. Academic Press.

Perraudin, W. R. M. and B. E. Sorensen (1992). The credit-constrained consumer: An empirical study of demand and supply in the loan market. *Journal of Business & Economic Statistics 10*(2), 179–192.

Puelz, R. and A. Snow (1994). Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market. *Journal of Political Economy 102*(2), 236–257.

Rothchild, M. and J. Stiglitz (1978). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. In P. Diamond and M. Rothchild (Eds.), *Uncertainty in Economics*, pp. 257–280. Academic Press.

Starc, A. (2014). Insurer pricing and consumer welfare: evidence from medigap. *The RAND Journal of Economics 45*(1), 198–220.

Stiglitz, J. E. and A. Weiss (1981). Credit rationing in markets with imperfect information. *The American Economic Review 71*(3), 393–410.

Williamson, S. D. (1987). Costly monitoring, loan contracts, and equilibrium credit rationing. *The Quarterly Journal of Economics 102*(1), 135–145.

A Summary Statistics of the KCB Dataset

A.1 Market Structure

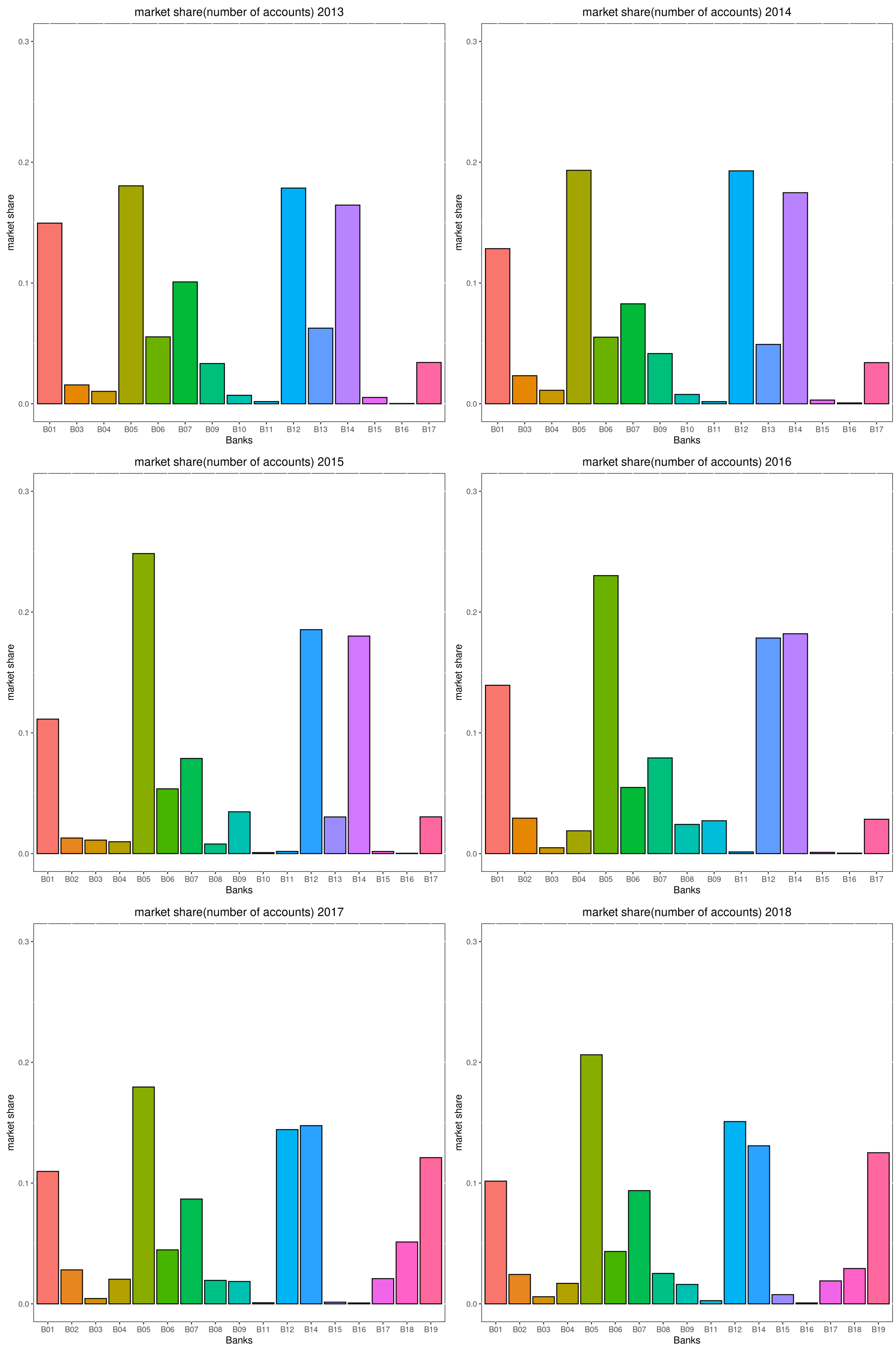

Figure 4: Market Share in the Personal Loan Market of South Korea (2013~2018)

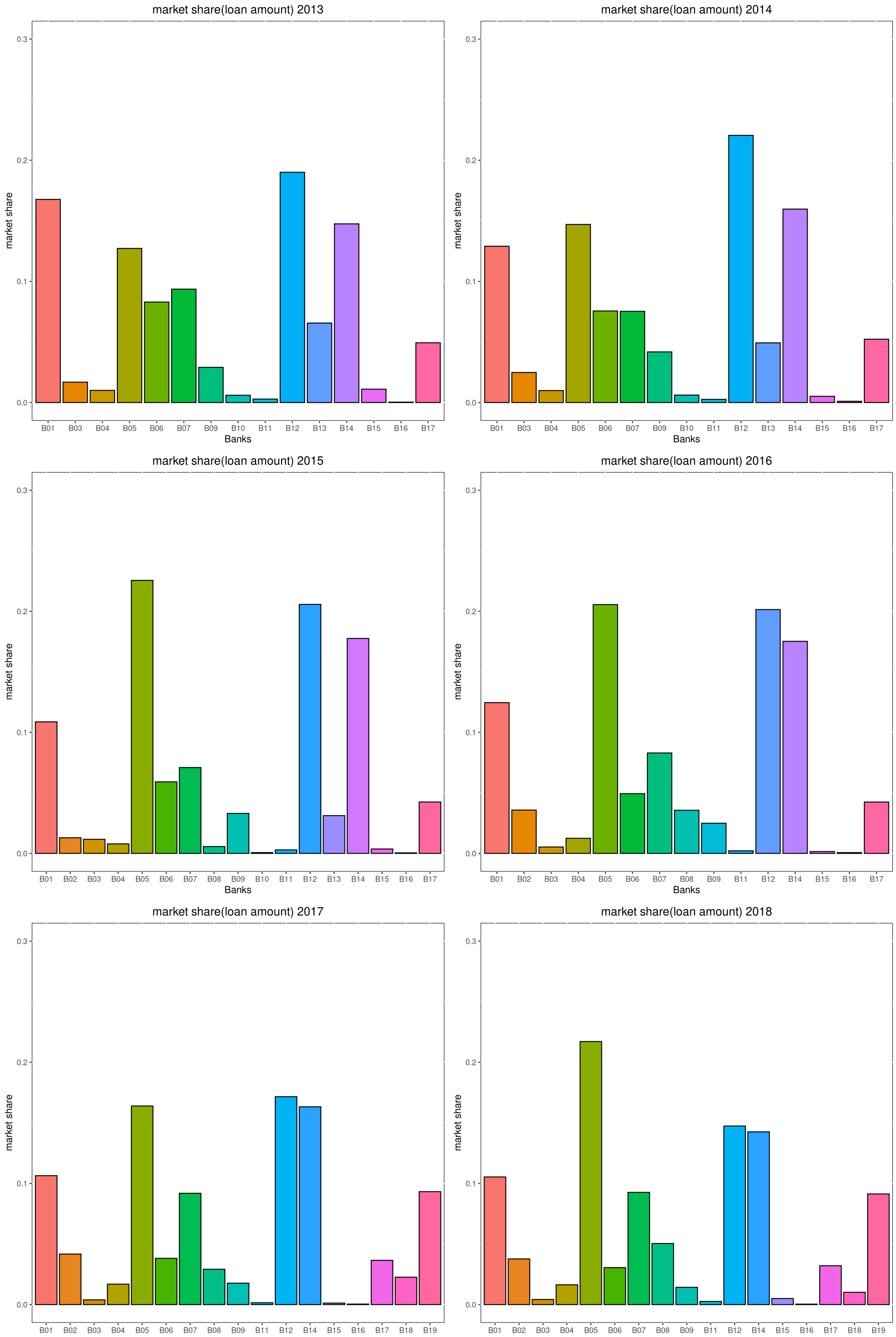

Figure 5: Market Share in the Personal Loan Market of South Korea (2013~2018)

A.2 Relationship between Consumer Characteristics and Default

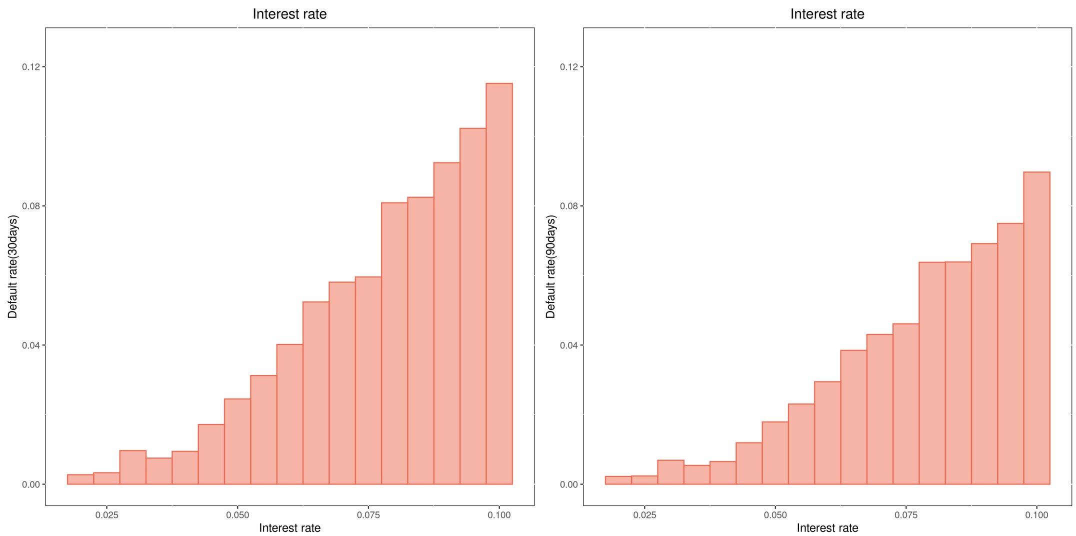

Figure 6: The Relationship between Loan Price and Default

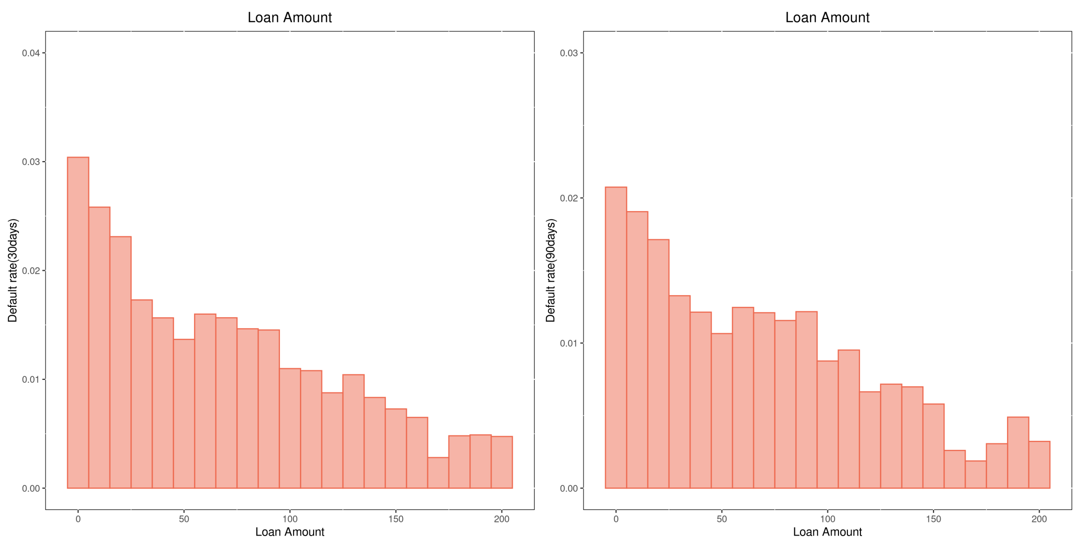

Figure 7: The Relationship between Loan Amount and Default

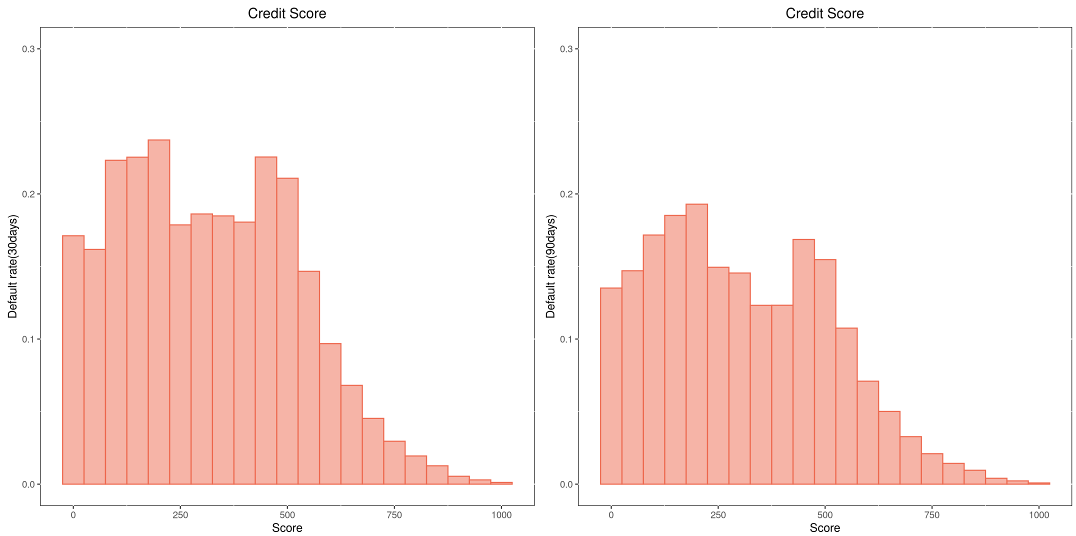

Figure 8: The Relationship between Credit Score and Default

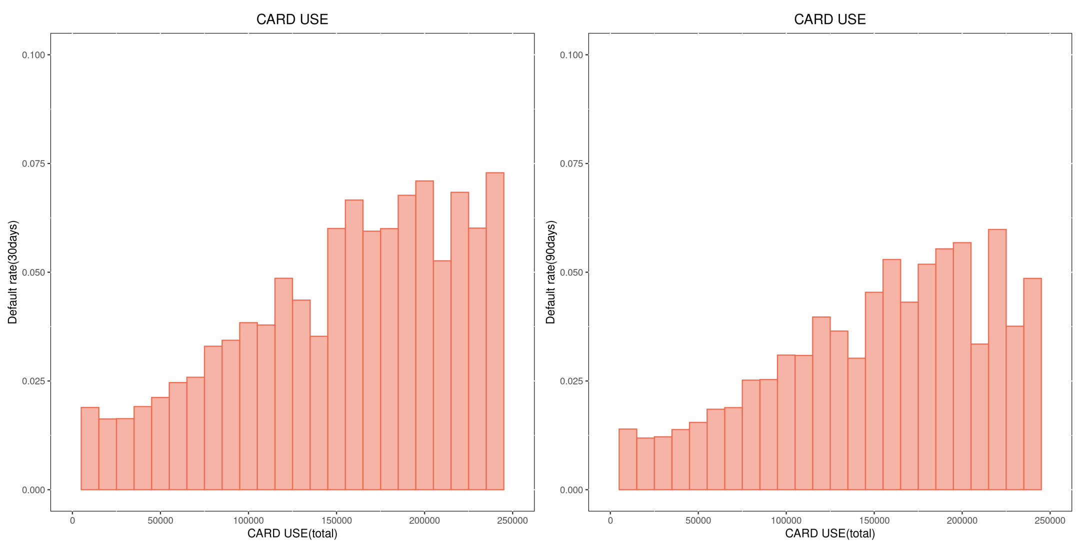

Figure 9: The Relationship between Card Use and Default

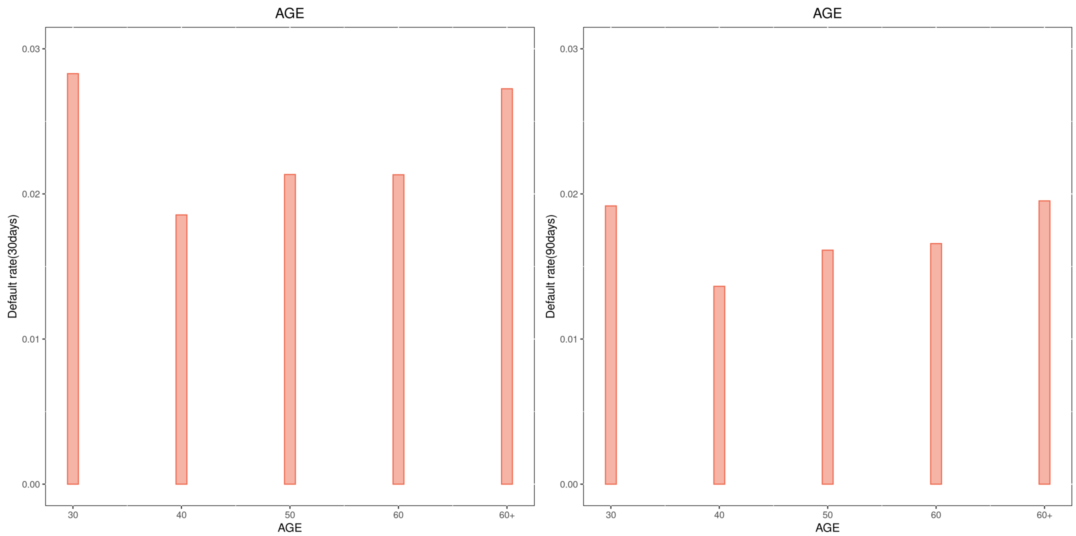

Figure 10: The Relationship between Consumer Age and Default

B Robustness Check

B.1 Alternative Specification

The following two tables, Tables 8 and 9 present the estimation results using the loan price channel only. The estimation process follows a similar step as Crawford et al. (2018). The tables support the explanations in Section 5.2.

Table 8: Structural Estimates of the Alternative Model

Variable	Estimates	Std. Err
Panel A. Individual-specific Variables in Demand		
Age	0.010***	0.002
Income	−0.008	0.007
Type of Job	0.013***	0.002
Debt to Income (DTI)	0.087***	0.003
Credit Score	0.053***	0.002
Amount of Credit Card Used in Last Year ($)	0.157***	0.028
Amount of Debit Card Used in Last Year ($)	0.206***	0.004
Individual-bank-specific Variables in Demand		
Num. of Banks Neighborhood	0.054***	0.001
Price Effect		
Loan Price	−0.003***	0.001
Panel B. Individual-specific Variables in Default		
Age	0.062***	0.004
Income	−0.024***	0.005
Type of Job	0.017***	0.003
Credit Score	−0.289***	0.003
Amount of Credit Card Used in Last Year ($)	0.176***	0.008
Amount of Debit Card Used in Last Year ($)	0.412***	0.010
Bank-specific Variables in Default		
Debt Ratio	0.057***	0.004
Return on Equity	0.029***	0.004
Maturity	0.049***	0.003
Price Effect		
Loan Price	0.258***	0.003
Year Dummies included	Yes	Yes
Region Dummies included	Yes	Yes
Bank Size Dummies included	Yes	Yes

Notes: Panel A. in Table 8 reports the parameter estimates of the customer's demand function. And following Panel B. in Table 8 reports the parameter estimates of the customer's default function. Both structural parameters are estimated by the combination of BLP contraction method and simulated maximum likelihood estimation.

Table 9: Structural Estimates of Information Asymmetry Parameters (Alternative Specification)

Variable	Estimates	Std. Err
Var-Cov Matrix of unobserved heterogeneity		
σ_D	2.070***	0.0006
ρ_{PD}	0.011***	0.0013
ρ_{PF}	0.017***	0.0032
ρ_{DF}	0.471***	0.0094

Notes:

$$\begin{pmatrix} \varepsilon_i^P \\ \varepsilon_i^D \\ \varepsilon_i^F \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{PD}\sigma_D & \rho_{PF} \\ \rho_{PD}\sigma_D & \sigma_D^2 & \rho_{DF}\sigma_D \\ \rho_{PF} & \rho_{DF}\sigma_D & 1 \end{pmatrix} \right)$$

Panel A. in Table 9 reports the result of the variance-covariance matrix of unobserved heterogeneity through the first step estimation of the structural model.

B.2 Reduced Form Evidence

This section presents some reduced form empirical results to describe a relation of default and observable characteristics. In order to check whether asymmetric information exists or not, we run simple binary choice models: linear probability model, probit, and logit model with and without instrumental variables. The potential endogeneity of the loan interest rate and the loan amount is controlled by bank-specific instrumental variables. We used the number of bank employees, the amount of total bank asset, and the ratio of individual and firm loans as instrumental variables. We use a binary specification:

$$F_{ijmt} = 1\left\{X'_{ijmt}\gamma^F + \varepsilon_m^F + \varepsilon_j^F + \varepsilon_t^F + \epsilon_{ijmt} \geq 0\right\},$$

where ε_m^F is a region-specific effect, ε_j^F is a bank-specific effect, and ε_t^F is a time-specific effect. $F_{ijmt} = 1$ if a consumer i makes a default. The model is designed to figure out the correlation of a borrower's default decision with observable characteristics.

Tables 10~12 show an empirical evidence of how the default is correlated to the interest rate and the loan amount. There are significantly positive correlations with both regressors, and the interest rate is more significant variable compared to the loan amount. The reduced-form IV approach finds a negative correlation of the loan amount and the default probability.

	Dependent variable: Default		
	Linear Probability	Control Function	2SLS
Interest Rate	0.597^{***} (0.013)	0.733^{***} (0.216)	0.733^{***} (0.216)
log(Loan Amount)	0.007^{***} (0.0003)	-0.048^{***} (0.009)	-0.048^{***} (0.009)
Age	0.0001 (0.00003)	-0.0001^{**} (0.0001)	-0.0001^{**} (0.0001)
log(Income)	0.001 (0.001)	0.049^{***} (0.008)	0.049^{***} (0.008)
Credit Score	-0.0002^{***} (0.0000)	-0.0001^{***} (0.00003)	-0.0001^{***} (0.00003)
Type of Job	0.001^{*} (0.001)	-0.002^{**} (0.001)	-0.002^{**} (0.001)
Debt to Income (DTI)	0.005^{***} (0.001)	0.013^{***} (0.002)	0.013^{***} (0.002)
log(Credit Cards Usage in Last Year)	-0.002^{***} (0.0001)	-0.002^{***} (0.0002)	-0.002^{***} (0.0002)
log(Debit Cards Usage in Last Year)	0.001^{***} (0.0001)	0.001^{***} (0.0001)	0.001^{***} (0.0001)
Maturity	0.0001^{***} (0.00001)	0.0002^{***} (0.00004)	0.0002^{***} (0.00004)
Control Function for Interest Rates		0.136 (0.216)	
Control Function for Loan Amount		0.055^{***} (0.009)	
Constant	0.079^{***} (0.007)	-0.013 (0.051)	-0.013 (0.051)
Regional, Year, Bank Fixed Effects	Yes	Yes	Yes
Observations	343,864	343,864	343,864

Table 10: Reduced Form Evidence of Asymmetric Information: Control both the loan interest rate and the loan amount

	Dependent variable: Default		
	Linear Probability	Control Function	2SLS
Interest Rate	0.546^{***}	0.958^{***}	0.958^{***}
	(0.012)	(0.211)	(0.211)
Age	0.0001^{***}	0.00002	0.00002
	(0.00003)	(0.00004)	(0.00004)
log(Income)	0.006^{***}	0.011^{***}	0.011^{***}
	(0.001)	(0.003)	(0.003)
Credit Score	-0.0002^{***}	-0.0002^{***}	-0.0002^{***}
	(0.0000)	(0.00002)	(0.00002)
Type of Job	0.001	-0.0001	-0.0001
	(0.001)	(0.001)	(0.001)
Debt to Income (DTI)	0.006^{***}	0.006^{***}	0.006^{***}
	(0.001)	(0.001)	(0.001)
log(Credit Cards Usage in Last Year)	-0.002^{***}	-0.002^{***}	-0.002^{***}
	(0.0001)	(0.0002)	(0.0002)
log(Debit Cards Usage in Last Year)	0.001^{***}	0.001^{***}	0.001^{***}
	(0.0001)	(0.0001)	(0.0001)
Maturity	0.0001^{***}	0.00005^{*}	0.00005^{*}
	(0.00001)	(0.00003)	(0.00003)
Control Function for Interest Rates		-0.413^{*}	
		(0.212)	
Constant	0.083^{***}	-0.015	-0.015
	(0.007)	(0.051)	(0.051)
Regional, Year, Bank Fixed Effects	Yes	Yes	Yes
Observations	343,864	343,864	343,864

Table 11: Reduced Form Evidence of Asymmetric Information: Control the loan interest rate only

	Dependent variable: Default		
	Linear Prob.	Control Function	2SLS
log(Loan Amount)	0.004^{***}	-0.054^{***}	-0.054^{***}
	(0.0003)	(0.009)	(0.009)
Age	0.0002^{***}	-0.00004	-0.00004
	(0.00003)	(0.00005)	(0.00005)
log(Income)	-0.003^{***}	0.046^{***}	0.046^{***}
	(0.001)	(0.008)	(0.008)
Credit Score	-0.0003^{***}	-0.0001^{***}	-0.0001^{***}
	(0.0000)	(0.00002)	(0.00002)
Type of Job	0.002^{***}	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Debt to Income (DTI)	0.006^{***}	0.014^{***}	0.014^{***}
	(0.001)	(0.001)	(0.001)
log(Credit Cards Usage in Last Year)	-0.002^{***}	-0.002^{***}	-0.002^{***}
	(0.0001)	(0.0001)	(0.0001)
log(Debit Cards Usage in Last Year)	0.001^{***}	0.001^{***}	0.001^{***}
	(0.0001)	(0.0001)	(0.0001)
Maturity	0.0002^{***}	0.0003^{***}	0.0003^{***}
	(0.00001)	(0.00003)	(0.00003)
Control Function for Loan Amount		0.058^{***}	
		(0.009)	
Constant	0.217^{***}	0.154^{***}	0.154^{***}
	(0.006)	(0.012)	(0.012)
Regional, Year, Bank Fixed Effects	Yes	Yes	Yes
Observations	343,864	343,864	343,864

Table 12: Reduced Form Evidence of Asymmetric Information: Control the loan amount only

C Supplementary Analysis for Loan Price and Amount Predictions

C.1 Theoretical Model Background

Consider a simple utility maximization problem of a consumer i by

$$\max_{\{c_{ijm(t+s)},F_{ijm(t+s-1)},j\}_{s=0}^{\infty}} E_t \sum_{s=0}^{\infty} \beta_i^s U\left(c_{ijm(t+s)}, F_{ijm(t+s-1)}\right)$$

such that

$$\begin{aligned}
&c_{ijmt} + \left(1 - F_{ij'm(t-1)}\right)\left(1 + P_{ij'm(t-1)}\right) Q_{ij'm(t-1)} \leq w_{it} + \left(1 - F_{ij'm(t-1)}\right) Q_{ijmt} \\
&Q_{ijmt} \leq \bar{Q}_{ijmt} \\
&c_{ijmt} \geq 0,\ Q_{ijmt} \geq 0,
\end{aligned}$$

where c_{ijmt} is consumer i's consumption level in time t when the lender is bank j, $F_{ijmt} \in \{0,1\}$ is a default indicator, w_{it} is consumer i's income in time t, $\bar{Q}_{ijmt}$ is the lending limit set by the bank j, and β_i is an individual-specific time discount factor. The consumer i makes a payment on the loan borrowed in period t ($F_{ijmt} = 0$) or fails to repay the loan ($F_{ijmt} = 1$). We assume that $F_{ijmt} = 1$ implies a permanent exit from the loan market: $F_{ijmt'} = 1$ for $t' \geq t$ and $Q_{ijmt'} = 0$ for $t' \geq t + 1$.

Under the given state variables $\left(P_{ij'm(t-1)}, Q_{ij'm(t-1)}, w_{it}\right)$, a borrower i in time t decides whether to repay the previous loan or not. If a borrower i chooses a default, $c_{ijmt'} = w_{it'}$ for $t' \geq t$. If the previous loan is paid off, then the consumer i can choose a lender j for the current period t, the loan amount Q_{ijmt}, and the corresponding loan price P_{ijmt}. The loan price schedule P_{ijmt} depends on the loan size Q_{ijmt} and is offered by each bank j. For simplicity, we assume that there is no saving decision ($Q_{ijmt} \geq 0$) and no-Ponzi scheme condition is imposed. If the optimal Q_{ijmt} has an interior solution, the choice variables $\left(Q_{ijmt}, Q_{ijm(t+1)}, \ldots\right)$ follow the Euler equation

$$\begin{aligned}
&U'\left(w_{it} + Q_{ijmt} - \left(1 + P_{ij'm(t-1)}\right) Q_{ij'm(t-1)}\right) \\
=&\beta_i E_t U'\left(w_{it+1} + Q_{ij''m(t+1)} - \left(1 + P_{ijmt}\right) Q_{ijmt}\right)\left(1 + P_{ijmt} + P'_{ijmt} Q_{ijmt}\right),
\end{aligned} \tag{12}$$

where P'_{ijmt} is the derivative of the price schedule P_{ijmt} with respect to Q_{ijmt}. The optimal policy function Q_{ijmt} is a function of the past debt $\left(1 + P_{ij'm(t-1)}\right) Q_{ij'm(t-1)}$, the current income w_{it}, the expected future income w_{it+1}, and individual-level heterogeneity β_i. We summarize the relevant variables for optimal Q_{ijmt} by Z^C_{ijmt}. If the lending limit $\bar{Q}_{ijmt}$ is binding, Q_{ijmt} also relies on the bank-specific variables Z^B_{jmt}. The optimal price $P_{ijmt} = P_{ijmt}\left(Q_{ijmt}\right)$ is also a function of Z_{ijmt}.

Consider the optimal policy functions Q_{ijmt} and P_{ijmt} satisfying the equation (12). The policy functions provide a mechanism of choosing a lender and making a default decision. Define a value function of choosing bank j in time t by

$$U_{ijmt}^{D} = \max_{\left\{c_{ijmt},F_{ijmt},\left\{c_{ij'm(t+s)},F_{ij'm(t+s)},j'\right\}_{s=1}^{\infty}\right\}} \left[U\left(c_{ijmt},0\right) + E_t \sum_{s=1}^{\infty} \beta_i^s U\left(c_{ij'm(t+s)},F_{ijm(t+s-1)}\right)\right], \tag{13}$$

where c_{ijmt} is a function of the policy functions P_{ijmt} and Q_{ijmt}. The borrower i in time t takes out a loan from a bank $j^* = \arg\max_{j\in\{1,\dots,J_{mt}\}} U_{ijmt}^{D}$. The default decision in period $t+1$ is made if the utility from the one-shot deviation is greater than $\max_{j'\in\{1,\dots,J_{mt+1}\}} U_{ij'm(t+1)}^{D}$. The deviation from the equilibrium path in time t ($F_{ijmt} = 1$) allows more consumption in time t but consumption smoothing is deprived in the later periods.

$$\begin{cases} F_{ijmt} = 1 & \text{if } E_{t+1}\sum_{s=1}^{\infty}\beta_i^{s-1}U\left(w_{it+s},1\right) \geq \max_{j'\in\{1,\dots,J_{mt+1}\}} U_{ij'm(t+1)}^{D} \\ F_{ijmt} = 0 & \text{otherwise.} \end{cases} \tag{14}$$

The consumption after a default is the same as the income w_{it} since there is not a saving option. The default decision in time t follows a threshold rule that the utility of default U_{ijmt}^{F} is greater than zero: $U_{ijmt}^{F} = E_{t+1}\sum_{s=1}^{\infty}\beta_i^{s-1}U\left(w_{it+s},1\right) - \max_{j'\in\{1,\dots,J_{mt+1}\}} U_{ij'm(t+1)}^{D} > 0$. The empirical model to capture the process of consumer's loan contract, loan demand, and default.

Remark C.1. $\bar{Q}_{ijmt}$ regulates the maximum loan amount for borrower i. The lending limit theoretically works to satisfy the transversality condition for the optimal dynamic programming problem. But we specify the borrowing constraint not only for a theoretical reason but also for an empirical context. In many cases, according to our dataset, $\bar{Q}_{ijmt}$ is the actual loan amount when a borrower i chooses a bank j. The borrowers are prone to take a loan out up to the bank's lending limit for most of the loan contracts.

C.2 Simultaneous Equations Model

This section suggests a simultaneous equations model to describe how the loan price and borrowing limit offers are made by banks. In the pre-contract stage, a consumer i who wants to use a personal loan service asks for her borrowing limit and the loan price schedule from J_{mt} banks. Let Q_{ijmt}^{D} and Q_{ijmt}^{S} denote, respectively, the loan demand and supply for consumer i, bank j, market m, and time t. Note that Q_{ijmt}^{D} and Q_{ijmt}^{S} are not the actual demand and supply, but a bank j's expected demand and supply based on its information set. The loan supply Q_{ijmt}^{S} is a function of the bank's lending capacity Z_{jmt}^{B} and the loan demand Q_{ijmt}^{D} is a function of consumer characteristics Z_{ijmt}^{C}. Z_{ijmt}^{C} includes all demand-default relevant covariates about a consumer i within bank j's information set. The loan price P_{ijmt}^{D} and P_{ijmt}^{S} are also included in loan demand and supply functions in a similar manner of Q_{ijmt}^{D} and Q_{ijmt}^{S}. In equilibrium, $Q_{ijmt}^{D} = Q_{ijmt}^{S} = Q_{ijmt}$ and $P_{ijmt}^{D} = P_{ijmt}^{S} = P_{ijmt}$ holds and a bank j obtains equilibrium loan price and quantity from a consumer i to choose bank j.

Considering loan price and quantity as a result of strategic interactions between banks, equilibrium P_{ijmt} and Q_{ijmt} rely on $\mathbb{P}_{i,-j,mt} \equiv \left(P_{i1mt}, \ldots, P_{i(j-1)mt}, P_{i(j+1)mt}, \ldots, P_{iJ_{mt}mt}\right)'$ and $\mathbb{Q}_{i,-j,mt} \equiv \left(Q_{i1mt}, \ldots, Q_{i(j-1)mt}, Q_{i(j+1)mt}, \ldots, Q_{iJ_{mt}mt}\right)'$. Then the simultaneous equations with demand-supply shocks $\mathbb{U}_{ijmt} = \left(\epsilon^D_{ijmt}, \epsilon^S_{ijmt}\right)'$ are represented by

$$
\begin{aligned}
Q^D_{ijmt} &= \beta^Q_0 + \beta^Q_1 P^D_{ijmt} + Z^{C'}_{ijmt}\beta^Q_2 + \mathbb{P}'_{i,-j,mt}\beta^Q_3 + \mathbb{Q}'_{i,-j,mt}\beta^Q_4 + \epsilon^D_{ijmt} && (15)\\
Q^S_{ijmt} &= \gamma^Q_0 + \gamma^Q_1 P^S_{ijmt} + Z^{B'}_{jmt}\gamma^Q_2 + \mathbb{P}'_{i,-j,mt}\gamma^Q_3 + \mathbb{Q}'_{i,-j,mt}\gamma^Q_4 + \epsilon^S_{ijmt} && (16)\\
Q_{ijmt} &= Q^D_{ijmt} = Q^S_{ijmt}\\
P_{ijmt} &= P^D_{ijmt} = P^S_{ijmt}
\end{aligned}
$$

for $j = 1, \ldots, J_{mt}$. We assume that $\left(\epsilon^D_{i1mt}, \epsilon^S_{i1mt}, \ldots, \epsilon^D_{iJ_{mt}mt}, \epsilon^S_{iJ_{mt}mt}\right)'$ follows a $2J_{mt}$-dimensional joint normal distribution. Koopmans (1949), Hausman (1983), and Matzkin (2008) discussed identification and estimation of simultaneous equations model, and identification of the linear simultaneous equations introduced above is attained by Koopmans (1949).[10] Thus equilibrium loan price and quantity for a consumer i are derived as a solution of the provided system of simultaneous equations.

Remark C.2. The prediction of $P_{ij'mt}$ and $Q_{ij'mt}$ for $j' \neq j$ provided by simultaneous equations (15) and (16) may not be consistent with equations (5) and (6) that derive optimal P_{ijmt} and $\bar{Q}_j$ by maximizing the expected profit function. Example C.1 provides a simple linear example that exemplifies the equivalence of our prediction method and theoretical analysis in Section 3.3.

One important issue in solving the simultaneous equations is missing observations. The reduced form of the specified model presents

$$\left(\gamma^Q_1 - \beta^Q_1\right) P_{ijmt} = \left(\beta^Q_0 - \gamma^Q_0\right) + Z^{C'}_{ijmt}\beta^Q_2 - Z^{B'}_{jmt}\gamma^Q_2 + \tilde{\epsilon}^P_{ijmt},$$

where $\tilde{\epsilon}^P_{ijmt} = \mathbb{P}'_{i,-j,mt}\left(\beta^Q_3 - \gamma^Q_3\right) + \mathbb{Q}'_{i,-j,mt}\left(\beta^Q_4 - \gamma^Q_4\right) + \left(\epsilon^D_{ijmt} - \epsilon^S_{ijmt}\right)$, and Q_{ijmt} can be similarly represented as a function of $\{Z_{ijmt}, \mathbb{P}_{i,-j,mt}, \mathbb{Q}_{i,-j,mt}, \mathbb{U}_{ijmt}\}$. Researchers can only observe the actual loan price and quantity that result in a loan contract, while counterfactual offers made by other banks are not observable. Without the missing values $\mathbb{P}_{i,-j,mt}$ and $\mathbb{Q}_{i,-j,mt}$, the simultaneous equations are not identified even if bank specific price or quantity shifters exist.

Note that $E\left[\tilde{\epsilon}^P_{ijmt}|Z_{ijmt}\right] = \mu^\epsilon + Z^{C'}_{ijmt}\beta^\epsilon_2 + Z^{B'}_{jmt}\gamma^\epsilon_2$ for some μ^ϵ, β^ϵ_2 and γ^ϵ_2 so that each endogenous error term can be approximated by a linear function of Z_{ijmt}. That is, $\{P_{ijmt}, Q_{ijmt}\}$ for $j = 1, \ldots, J_{mt}$ can be represented by a system of linear equations with an unknown variance-covariance matrix for error terms. To simplify the model setup, we use a linear specification. Assume that the

[10] Detailed literature review on the simultaneous equations model is available in Hausman (1983), and a general nonadditive simultaneous equations model is discussed in Matzkin (2008).

loan price and quantity are presented by the sum of predicted value and measurement error. Then,

$$
\begin{aligned}
P_{ijmt} &= \tilde{P}_{ijmt} + u^P_{ijmt} \\
&= \left(Z^{B'}_{jmt} \delta^P_1 + Z^{C'}_{ijmt} \delta^P_2 + \varepsilon^P_i \right) + u^P_{ijmt} \\
Q_{ijmt} &= \bar{Q}_j \left(Z_{ijmt} \right) = \tilde{Q}_{ijmt} + u^Q_{ijmt} \\
&= \left(Z^{B'}_{jmt} \delta^Q_1 + Z^{C'}_{ijmt} \delta^Q_2 + \varepsilon^Q_i \right) + u^Q_{ijmt},
\end{aligned}
$$

so that $\tilde{P}_{ijmt}$ and $\tilde{Q}_{ijmt}$ are the predicted values of P_{ijmt} and Q_{ijmt}, considering bank specific fixed effect and the individual specific fixed effect. Z^B_{jmt} and Z^C_{ijmt} are respectively bank specific regressors and individual-bank specific regressors to predict the loan interest rate, and ε^P_i and ε^Q_i are the individual specific fixed effect that summarizes information that are observed by the bank but not observed by econometricians. In particular, Z^C_{ijmt} is a vector of borrower specific observables including i's credit score, age, education, and Z^B_{jmt} is a vector of bank specific determinants that affect the maximum loan amount. Similarly, the borrowing limit set by the banks are also predicted by the information in the loan application and other bank specific variables.

Example C.1. In this example, we exemplify that a pair of (P^*, Q^*) that satisfies the following demand and supply schedule generates a profit maximizing outcome. Consider a simple loan demand and supply function:

$$
\begin{aligned}
P^D &= aQ^D \\
P^S &= -c + bQ^S,
\end{aligned}
$$

where P^D and Q^D denote a demand schedule with $a > 0$ and P^S and Q^S present a supply schedule with $b, c > 0$ and $b > a$. A consumer is willing to pay higher interest with respect to higher loan amount, and a bank also provides a loan price and quantity following a linear schedule. Let (P^*, Q^*) satisfy both equations $P^* = aQ^*$ and $P^* = -c + bQ^*$.

The loan amount cannot exceed Q^* because $P^S > P^D$ at the $Q > Q^*$ so that a consumer cannot accept the offer from a bank. Conditional on $Q \leq Q^*$, a bank maximizes the profit function $\Pi = (P - mc) Q$ by exploiting its market power and setting the price by $P = aQ > P^S$. Then Π is a quadratic function of Q and is maximized at the boundary points: $Q = 0$ or $Q = Q^*$. The optimal Q that maximizes the bank's profit function is determined by the solution of simultaneous equations (loan demand and supply) as far as a loan contract between a borrower and a lender is made.