# IMES DISCUSSION PAPER SERIES

**Distance and Loan Pricing Revisited: A Structural Estimation Approach**

Mitsuru Katagiri and Naoki Wakamori

**Discussion Paper No. 2026-E-11**

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

2-1-1 NIHONBASHI-HONGOKUCHO
CHUO-KU, TOKYO 103-8660
JAPAN

**IMES Discussion Paper Series 2026-E-11**
**August 2026**

# Distance and Loan Pricing Revisited: A Structural Estimation Approach

**Mitsuru Katagiri* and Naoki Wakamori****

**Abstract**

This paper examines how geographic distance between bank branches and firms affects firms' borrowing costs. Building on descriptive evidence that highlights the role of nearby competing banks in determining loan rates, we develop a model where firms choose their lenders and banks compete in loan rates, and we structurally estimate the model using matched firm- and bank-branch-level microdata from Japan. The estimates suggest that distance affects loan rates by increasing both banks' and borrowers' costs. Counterfactual simulations indicate that the impact of branch closures on small businesses' borrowing costs depends on local banking market conditions.

**Keywords:** Banking; Loan pricing; Spatial price discrimination
**JEL classification:** G21, L13, L14, D43

* Waseda University and TDB-CAREE. (E-mail: mitsuru.katagiri@waseda.jp)
** Keio University and TDB-CAREE. (E-mail: nwakamo@keio.jp)

This paper was prepared in part while Wakamori was a visiting scholar at the Institute for Monetary and Economic Studies, Bank of Japan. We thank Nicola Cetorelli, Yusuke Imani, Anil Kashyap, Yoshiaki Ogura, Manju Puri, Rebecca De Simone, Iichiro Uesugi, and participants at the Econometric Society World Congress 2025, SWET 2025, and the seminars at Japan Society of Monetary Economics, TDB-CAREE, RIETI, Bank of Japan, and Waseda University. Katagiri and Wakamori gratefully acknowledge financial support from the JSPS KAKENHI Grant Numbers 21H00704 and 22H00837. The views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan.

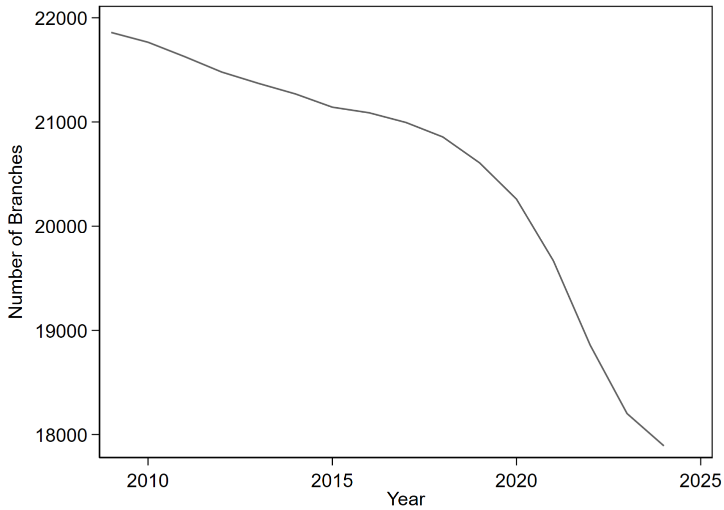
# 1 Introduction

The geographic distance between borrowers and lenders has been shown to be an important determinant of loan pricing. In recent years, borrower-lender distance has been widening in many countries due to bank branch closures.[1] This trend is driven, for example, by bank consolidation (Keil and Ongena, 2024) and by a decline in the franchise value of retail deposits associated with digitization (Narayanan et al., 2025). It raises policy concerns about declining competition in loan markets, as reduced competition among banks tends to increase borrowing costs for small businesses (Rice and Strahan, 2010). Despite ongoing digitization in the banking sector, recent empirical studies nonetheless show that geographic proximity still matters for corporate lending to small businesses (Nguyen, 2019; Granja et al., 2022; Adams et al., 2023; Herpfer et al., 2023). In light of these developments, a key question is what mechanisms determine loan rates given the geographic configuration of a borrower, its lender, and competing banks.

The existing empirical literature has extensively studied the role of distance in loan pricing, primarily using reduced-form approaches. This paper revisits this issue using a structural estimation approach with three main advantages. First, we address the endogeneity arising from the simultaneity of firms' lender choices and banks' pricing decisions, both of which may depend on distance. To this end, we explicitly model these decisions within a unified framework and estimate the model using a likelihood function constructed from the implied choices of lenders and loan rates. Second, our framework allows us to separately identify the effects of distance on lenders' costs and borrowers' costs, which can be interpreted as monitoring and transportation costs, respectively. Given our assumptions on bank competition, distance to the lender and to nearby competing banks affect loan rates differently, enabling us to quantify the contribution of each channel. Third, our model enables counterfactual analysis of policy-relevant scenarios. We examine how loan rates adjust endogenously to branch closures while accounting for firms' lender choices, which is essential because firms must select new lenders when their current branches close.

[1] See Keil and Ongena (2024) for the U.S. and Legran et al. (2024) for Europe.

Figure 1: Number of Bank Branches in Japan




*Note:* The figure shows the total number of bank branches in Japan from 2009 to 2024.

We begin with an empirical analysis of the role of distance to the lending bank and to competing banks, following Degryse and Ongena (2005). We combine firm-level data with bank-branch-level information from Japan, where more than 20 percent of branches have closed over the past 15 years, as shown in Figure 1, a pattern similar to that observed in other advanced economies. Our results are consistent with the existing literature (e.g., Degryse and Ongena, 2005; Agarwal and Hauswald, 2010; Herpfer et al., 2023): distance to the lending bank branch has a modest effect on loan rates, whereas distance to rival bank branches has a larger effect. These patterns inform how we specify the mode of competition in our structural model.

Guided by these empirical patterns, we develop a structural model of firms' lender choices and banks' loan pricing. Given their geographic configuration, firms choose a bank branch based on loan rates and distance (capturing transportation costs). Banks compete in setting loan rates à la Bertrand, taking into account rivals' behavior and their own costs, which depend on borrower characteristics, distance to the borrower (capturing monitoring costs), and idiosyncratic components. Under this assumption on the mode of competition, equilibrium loan rates are characterized by a second-price auction, as in Allen, Clark and Houde (2019), which is consistent with the empirical patterns we document. In other words, equilibrium rates are largely determined by the best alternative offer available to the firm, implying that loan rates depend primarily on the proximity of rival banks rather than that of the chosen lender.

We estimate the model using maximum simulated likelihood and the main findings are summarized as follows. First, distance affects loan pricing through both firms' costs and banks' costs. The estimates imply that a one-kilometer increase in the distance to a rival bank branch raises borrowing costs by approximately 8.7 basis points, which is economically significant in Japan's low-interest environment. In contrast, the effect of distance to the lending branch is more moderate, underscoring the central role of bank competition in determining loan rates. Second, our results indicate that reduced-form estimates may be biased due to endogeneity. In particular, the effect of distance to rival bank branches is substantially attenuated in reduced-form regressions: the estimated impact of a one-kilometer increase in rival distance is at most 1.8 basis points, which is much smaller than the 8.7 basis points implied by our structural model.

We use the estimated model to conduct counterfactual simulations that assess how hypothetical branch closures affect firms' borrowing costs in Japan. Since consolidation-driven branch closures in Japan are broadly consistent with global trends, our policy exercises have implications for other countries facing similar challenges. As a policy experiment, we close all branches of the second-largest bank (in terms of the number of borrowers) in each prefecture. While our results indicate that such consolidation can lead to sizable increases in loan rates, the magnitude of these effects varies substantially across firms and regions, reflecting differences in the availability of nearby alternative banks. Although bank consolidation may yield benefits such as reduced operating costs, these findings imply that it can also impose welfare costs by increasing firms' borrowing costs. This suggests that policymakers should consider regional bank competition when anticipating the adverse effects of bank consolidation.

**Related Literature**

This paper contributes to three strands of the literature. First, our study builds on the empirical literature examining the role of distance in loan pricing. Degryse and Ongena

(2005) use data from a single bank in Belgium and show that firms' borrowing costs depend on distance to both the lending bank and competing banks. Agarwal and Hauswald (2010) obtain similar results for the U.S. banking sector, using declined loan application data to address sample selection bias arising from endogenous lender-borrower matching. Herpfer, Mjøs and Schmidt (2023) address the endogeneity of distance by exploiting infrastructure projects as exogenous changes in accessibility and show that geographic proximity matters for loan pricing in Norway.[2] Despite the expansion of geographic scope of bank lending through digitization (Petersen and Rajan, 2002) and the share of FinTech lending is increasing (Gopal and Schnabl, 2022), recent empirical studies, including Nguyen (2019), Granja et al. (2022), Adams et al. (2023), and Herpfer et al. (2023), show that geographic proximity still matters for corporate lending to small businesses. Building on prior reduced-form studies that establish the importance of geographic proximity in corporate lending, we develop and structurally estimate a model where firms' bank choice and loan pricing are jointly determined, allowing us to unravel the mechanisms underlying distance effects by separately quantify lenders' and borrowers' costs associated with distance.

Second, this paper contributes to the recent literature applying structural estimation to banking, including Egan, Hortaçsu and Matvos (2017), Pavanini, Ioannidou and Peng (2022), and Wang, Whited, Wu and Xiao (2022). Among these studies, our paper is closely related to approaches that exploit firm-level decision data, rather than market-level aggregated data. In particular, our paper is similar to Crawford, Pavanini and Schivardi (2018) in exploiting matched firm-bank data, although we abstract from asymmetric information and focus instead on how geographic distance affects lender choice and loan pricing. Our estimation framework, in turn, closely follows Allen, Clark and Houde (2019), which studies the Canadian mortgage market by modeling Bertrand competition among banks. In addition, our motivation is similar to that of Yasuda (2005), who shows that relationship length affects both underwriter choice and underwriting fees, and separately identifies these two channels. In our setting, distance plays an analogous role, influenc-

[2] In the Japanese context, Ono, Saito, Sakai and Uesugi (2023) show that increase in lender distance due to bank mergers induce firms to switch lenders.

ing both firms' lender choices and loan rates, and our framework allows us to disentangle these effects.

Third, this paper is related to the literature on the effects of bank consolidation (e.g., Sapienza, 2002; Maudos and de Guevara, 2007; Degryse et al., 2011; Erel, 2011; Nguyen, 2019; Huber, 2021). In contrast to these studies, we do not identify the causal effects of bank consolidation. Instead, we conduct counterfactual simulations to quantify its impact on loan pricing while accounting for firms' endogenous bank choice and the competitive structure each firm faces. Our counterfactual analysis is also related to the merger simulation literature in industrial organization based on the framework developed by (Berry et al., 1995). Unlike standard approaches relying on aggregate market share data, we exploit firm-level data to account for rich heterogeneity in firms' choice sets. This feature is particularly important in local banking markets, where firms face geographically differentiated lender options and respond differently to changes in branch networks.

The remainder of the paper is organized as follows. Section 2 describes the data used in the empirical analysis and presents reduced-form evidence to motivate the structural model. Section 3 introduces the structural model of banking, and Section 4 presents the empirical results, including counterfactual simulations. Section 6 concludes.

## 2 Data and Motivating Facts

This study primarily utilizes two datasets: the Teikoku Databank database and Kinyu Meikan (Financial Directory). First, We describe these data sources and their structure and then present descriptive statistics. Appendix A provides a more detailed description of the data, including the construction of the variables used for estimation. Next, we perform descriptive and reduced-form analyses to motivate our structural models. In our analysis, we particularly focus on how firms' choice of lending branches, as well as their loan pricing, are affected by the geographical distance between firms and bank branches.

## 2.1 Data Sources

Our firm-level data is provided by Teikoku Databank LTD., a leading business credit bureau in Japan. We utilize various datasets from the Teikoku Databank: the corporate financial database, the corporate profile database, and the corporate credit research database. The corporate financial database contains financial information on firms, including interest payments, profitability, and asset and loan sizes. The corporate profile database provides firm-level information, including its address. The corporate credit research database provides the bank-branch code of the bank branch with which the firm primarily transacts.

Financial Directory (Kinyu Meikan) is an annual dataset that comprehensively lists branches of financial institutions in Japan. Although Japanese private financial institutions can be categorized in various ways, they are generally classified into five main groups: large banks, first-tier regional banks, second-tier regional banks, credit associations, and credit cooperatives. We refer to these bank categories as "bank type." The bank type is a potentially important variable in loan pricing, as the Japanese loan market is loosely segmented by bank type for firms with different firm size. The dataset contains detailed information on each branch, including its address and bank-branch code. For our analysis, we collected data from 2011 to 2020 and structured it as a panel dataset.

These two databases are merged using the bank-branch code, as in Ono et al. (2023). We then calculate the distance between each firm and its lending branch, as well as the distance between the firm and the 30 nearest bank branches as a potential "rival branch."[3] Because our focus is on bank accessibility for small businesses in regional areas, we examine only firms in those areas that borrow from regional banks. Specifically, we exclude firms from our sample if: (i) they are located in Tokyo or Osaka, or (ii) they borrow from one of the four large banking groups.[4] Appendix A provides more details about our

[3]As described in Appendix A, when a firm borrows from multiple bank branches, we treat the main bank as the lender and include the other lending branches among rival branches if they are located near the firm.

[4]The four large banking groups are Mizuho, SMBC, MUFG, and Resona. We also exclude firms that borrow from credit cooperatives, as they tend to be too small—primarily individual businesses—to be comparable with other firms.

Table 1: Descriptive Statistics

| | Obs. | Mean | S.D. | p5 | p50 | p95 |
|---|---|---|---|---|---|---|
| Lending Bank | | | | | | |
| Distance (km) | 871,527 | 3.18 | 3.75 | 0.27 | 2.00 | 9.95 |
| Distance to rivals | | | | | | |
| The closest (km) | 871,527 | 1.41 | 1.54 | 0.14 | 0.90 | 4.52 |
| Second closest (km) | 871,475 | 2.17 | 2.81 | 0.31 | 1.36 | 6.65 |
| Third closest (km) | 871,352 | 3.01 | 5.61 | 0.47 | 1.78 | 8.78 |
| Forth closest (km) | 870,860 | 3.63 | 6.34 | 0.61 | 2.15 | 10.56 |
| Fifth closest (km) | 870,242 | 4.29 | 10.22 | 0.73 | 2.47 | 12.26 |
| Other variables | | | | | | |
| Interest Rate (%) | 871,527 | 1.86 | 1.26 | 0.38 | 1.65 | 4.04 |
| Total Assets (bill. yen) | 871,527 | 1.12 | 10.02 | 0.02 | 0.23 | 3.75 |
| ROA | 871,527 | 0.03 | 0.09 | -0.09 | 0.02 | 0.17 |
| Capital Ratio | 871,527 | 0.31 | 0.22 | 0.03 | 0.27 | 0.73 |
| Liquidity Ratio | 871,527 | 0.60 | 0.23 | 0.21 | 0.62 | 0.94 |
| Long-Term Loan Ratio | 871,527 | 0.77 | 0.28 | 0.14 | 0.88 | 1.00 |
| Loan Size (bill. yen) | 871,527 | 0.36 | 1.81 | 0.004 | 0.07 | 1.37 |

database.

## 2.2 Descriptive Statistics

Table 1 reports the descriptive statistics of our merged dataset. The sample consists of 871,527 observations from 2011 to 2020, with approximately 90,000 firms observed each year. The table shows that the average distance to the lending branch is about 3.2 km, with a standard deviation of 3.8 km and a median of 2.0 km.[5] Compared with previous studies for other countries, the average distance to the lender in Japan is longer than the 2.3 km reported for Belgium in Degryse and Ongena (2005), but shorter than the 10.1 km reported for Norway in Herpfer et al. (2023).

The table also reports the distances to the closest, second-, third-, fourth-, and fifth-

[5]Some previous studies, including Herpfer et al. (2023), use travel time between firms and bank branches, rather than geographic distance, as a measure of accessibility. We cannot compute this measure because we are not allowed to take our confidential data outside the TDB data center. However, geographic distance is a good proxy for accessibility in Japan, given that transportation infrastructure is well developed even in regional areas.

closest branches, excluding the lending branch. The average distance to the closest bank branch (excluding the lending branch) is 1.4 km, with a median of 0.9 km, while the average distance to the second-closest branch is 2.2 km, with a median of 1.4 km. The average distance to the fifth-closest bank branch is 3.72 km, implying that, on average, more than five financial institutions are located within a 4.0 km radius.

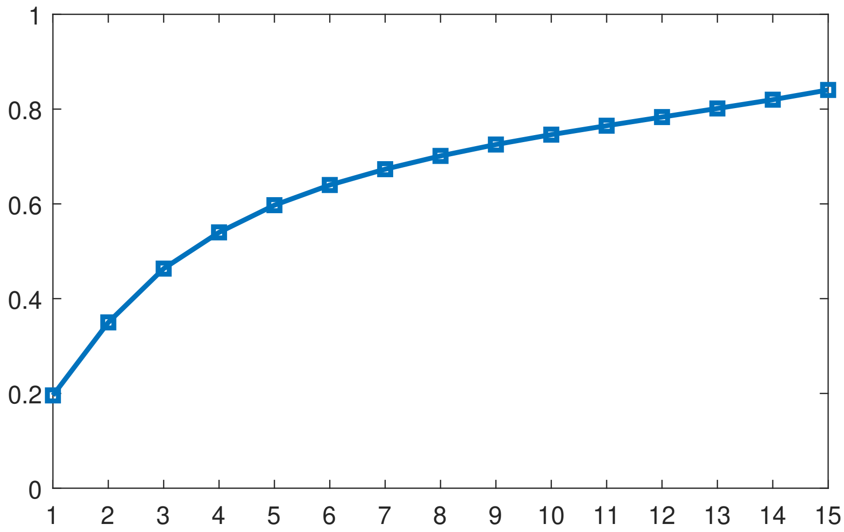
Regarding firms' borrowing costs, the table shows that the average interest rate is only 1.86%, with a median of 1.65%, reflecting the low interest rate environment in Japan. Although not shown in the table, average loan rates in Japan have declined over the past few decades, suggesting that controlling for time trends is important for the estimation. The table also reports summary statistics for additional financial variables, such as asset size, ROA, and capital ratios, which are included as determinants of interest rates in our estimation. The average asset size, 1.12 billion yen, is somewhat larger than that of Japanese firms overall, likely because our sample does not include very small individual firms. The liquidity ratio, long-term loan ratio, and loan size also exhibit substantial heterogeneity across firms, reflecting differences in firms' financial conditions and borrowing relationships.

## 2.3 Does Distance Matter When Choosing a Lending Branch?

We begin our descriptive analysis by asking the following question: Do firms actually borrow from nearby bank branches? In other words, does the distance between borrowers and lenders matter for firms' lender choices? If greater borrower-lender distance entails costs for borrowers and/or lenders, as suggested by the theoretical and empirical literature, we would expect firms to be more likely to borrow from nearby banks.

As a first step toward answering this question, Figure 2 reports the number of firms borrowing from the $x$th closest bank branch and plots the corresponding cumulative shares. The figure shows that approximately 20% of firms choose the nearest branch as their lender, similar to the share reported for Belgium by Degryse and Ongena (2005). Moreover, approximately 84% of firms select their lender from among the fifteen closest bank branches, suggesting that proximity to bank branches is an important determinant of firms' lender choices.

Figure 2: Rank of Lending Branch by Distance (Cumulative)



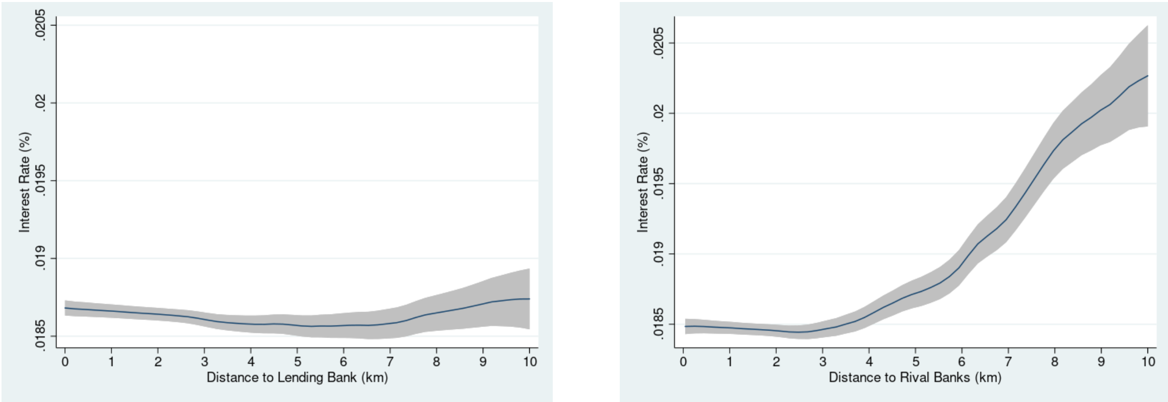

*Note:* The figure shows the cumulative share of firms selecting the x-th closest bank branch as their main lender. It implies that around 20% of firms choose the nearest branch as their lender.

Table 2: Firm's Lender Choice by Loan Size

| | (i) | (ii) | (iii) |
|---|---|---|---|
| Dep. Var. | $Select_{i,3}$ | $Select_{i,5}$ | $Select_{i,10}$ |
| LoanSize | -0.173$^{***}$ | -0.184$^{***}$ | -0.193$^{***}$ |
| | (0.001) | (0.001) | (0.011) |
| # of Rivals | -0.128$^{***}$ | -0.134$^{***}$ | -0.150$^{***}$ |
| | (0.001) | (0.001) | (0.001) |
| $N$ | 880,345 | 880,345 | 880,345 |

*Note:* Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. $Select_{i,j}$ for $j = 3, 5$, and $10$ is a dummy variable equal to one if firm $i$ chooses a lending branch from among the $j$ nearest bank branches. The table reports logit estimation results for lender choice as a function of loan size and the number of nearby rival branches.

Figure 3: Distance and Borrowing Costs




*Note:* The figure shows the relationship between borrowing costs and distance to the lending and rival bank branches. The horizontal axis represents the distance to the lending branch (the left panel) and the distance to the average distance to the five nearest branches (the right panel).

While Figure 2 suggests that distance is an important factor in firms' lender choices, it also indicates substantial heterogeneity across firms in the extent to which proximity is valued. What explains this heterogeneity? If the costs associated with greater distance reflect transportation or monitoring costs, as suggested by the existing literature, firms with larger loan volumes may be less sensitive to distance due to economies of scale. To examine this hypothesis, we define $Select_{i,j}$ for $j = 3, 5$, and $10$ as dummy variables that take the value one if firm $i$ selects its lending branch from among the $j$ nearest branches. We then conduct a logit analysis in which these indicators are regressed on loan volume and the number of nearby branches. Table 2 shows that, after controlling for the availability of nearby branches, firms with larger loan volumes are less likely to choose their lender from among nearby branches for all cases of $j = 3, 5$, and $10$. These results imply that distance is more important for firms with smaller loan volumes when selecting their lender, consistent with the hypothesis that borrowers' distance-related costs exhibit economies of scale. Building on this empirical result, we allow the cost of distance to vary with loan volume in the structural model in the next section.

## 2.4 Does Distance Matter When Setting Loan Rates?

Next, to understand the role of distance for banks in setting loan rates, we examine the relationship of loan rates with distance to the lending branch, as well as rival branches. Figure 3 illustrates these relationships using local polynomial regressions. The horizontal axis in the left and right panel represents the distance to the lending branch and the distance to rival branches, respectively.[6] The figure shows that, while there is no clear relationship between loan rates and the distance to the lending branch, loan rates increase significantly with the distance to nearby rival branches. More specifically, when the distance to nearby branches increases from 3 km to 8 km, borrowing costs rise by approximately 15 bps on average. This finding suggests that the distance to nearby bank branches—potential competitors to the current lending bank—is an important determinant of firms' borrowing costs.

Next, to more formally examine the relationship between loan rates and distance, we conduct a regression analysis. Specifically, following Degryse and Ongena (2005), we estimate the following econometric model:

$$\text{Interest Rate}_{it} = \beta_0 + \beta_1 D_{it}^{Lender} + \beta_2 D_{it}^{Rival,k} + \text{ Controls}_{it} + \text{Year}_t + \epsilon_{it}$$

where Interest $\text{Rate}_{it}$ denotes the loan rate for firm $i$ in year $t$, and $D_{it}^{Lender}$ is the distance between firm $i$ and its lending branch in year $t$. As proxies for the distance to rival branches, $D_{it}^{Rival,k}$, we consider four alternative measures indexed by $k = 1, 2, 3, 4$. First, $D_{it}^{Rival,1}$ is the distance to the nearest bank branch for firm $i$ (or the second-nearest branch if the lending branch is the nearest). Second, $D_{it}^{Rival,2}$ is the distance to the nearest branch within the same bank type. For example, when the lender for firm $i$ is a tier-one regional bank, $D_{it}^{Rival,2}$ denotes the distance to the nearest branch of tier-one regional banks (or the second-nearest branch if the lending branch is the nearest). This second measure captures the possibility that firms may prefer particular types of banks over others. Third, $D_{it}^{Rival,3}$ is the average distance to the three closest bank branches for firm $i$,

[6] As a proxy for the distance to rival branches, we use the average distance to the five nearest branches. Using alternative proxies, such as the average distance to the three nearest branches, yields qualitatively similar results.

Table 3: Distance and Borrowing Cost

| | (i) | (ii) | (iii) | (iv) |
|---|---|---|---|---|
| $D^{Lender}$ | .0012*** | .0016*** | .0003 | .0002 |
| | (.0004) | (.0004) | (.0004) | (.0003) |
| $D^{Rival,1}$ | .0086*** | | | |
| | (.0009) | | | |
| $D^{Rival,2}$ | | .0078*** | | |
| | | (.0004) | | |
| $D^{Rival,3}$ | | | .0137*** | |
| | | | (.0006) | |
| $D^{Rival,4}$ | | | | .0131*** |
| | | | | (.0005) |
| Controls | √ | √ | √ | √ |
| Year FE | √ | √ | √ | √ |
| # of Obs. | 871,527 | 700,010 | 871,220 | 869,934 |

*Note:* Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. The table the estimation results, where each column uses a different measure of distance to rival bank branches: (1) $D^{Rival,1}$ is the distance to the nearest bank branch for firm $i$ (or the second-nearest branch if the lending branch is the nearest), (2) $D^{Rival,2}$ is the distance to the nearest branch within the same bank type, (3) $D^{Rival,3}$ is the average distance to the three closest bank branches for firm $i$, (4) $D^{Rival,4}$ is the average distance to the five closest bank branches for firm $i$. The estimation also includes dummy variables for the bank type of the lending branch, return on assets (ROA), capital ratios, liquidity ratios, the logarithm of total assets, and long-term loan ratios for firm $i$ in year $t$. We also include time fixed effects, denoted by $\text{Year}_t$, to capture macroeconomic factors affecting interest rates.

and fourth, $D_{it}^{Rival,4}$ is the average distance to the five closest bank branches for firm $i$. The third and fourth measures are supposed to address the possibility that the relevant rival branch is not necessarily the nearest one, but rather one of several branches in the vicinity. $\text{Controls}_{it}$ includes dummy variables for the bank type of the lending bank, return on assets (ROA), capital ratios, liquidity ratios, the logarithm of total assets, and long-term loan ratios for firm $i$ in year $t$. We also include time fixed effects, denoted by $\text{Year}_t$, to capture macroeconomic factors affecting interest rates, including the declining trend in interest rates in Japan.

Table 3 reports the estimation results. Each column employs a different measure of distance to rival bank branches. When the distance to the nearest bank branch, $D^{Rival,1}$, is used (the first column), the coefficients on both the distance to the lending branch and the distance to the rival branch are positive and statistically significant; however, the effect of rival distance is larger in magnitude and more statistically significant. Using the distance to the nearest branch within the same bank type, $D^{Rival,2}$, in the second column yields qualitatively similar results. When average distances to nearby branches are used instead (the third and fourth columns), the estimated effects of rival distance become larger and more statistically significant, while the effects of distance to the lending branch become smaller and statistically insignificant. As discussed above, average distances to nearby branches may more appropriately capture rival distance, as they account for the possibility that the nearest bank branch is not a relevant competitor. Consistent with this interpretation, the results in the third and fourth columns indicate that distance to the lending branch is not a significant determinant of borrowing costs once distance to rival branches is controlled for. Overall, these results suggest that distance to rival branches plays a more important role in determining equilibrium loan rates, highlighting the importance of competition among bank branches in loan pricing.

We also examine heterogeneity across firms in the relationship between loan rates and distance. Motivated by the evidence that firms exhibit different preferences for proximity, we test whether firms that are more sensitive to distance when choosing their lender are also more sensitive to distance in borrowing costs. To do so, we re-estimate regression (1) using a subsample of firms that borrow from one of the five nearest bank branches and compare the results with the baseline estimates obtained from the full sample estimation in Table 3. The estimation results in Table 4 show that, while most coefficients are qualitatively similar to those in Table 3, a key difference emerges in columns (3) and (4): the coefficient for the distance to the lending branch is *negative* and statistically significant. That is, conditional on borrowing from one of the five nearest branches, firms face lower borrowing costs when their lending branch is located farther away. Although this result contrasts with the estimates based on the full sample, it is more consistent with findings in the previous literature, such as Degryse and Ongena (2005), Agarwal and Hauswald (2010), and Herpfer et al. (2023). In the next section, we use a structural model of bank competition to discuss a possible mechanism underlying these reduced-form results in

Table 4: Distance and Borrowing Cost for Firms Borrowing from Five Nearest Banks

| | (i) | (ii) | (iii) | (iv) |
|---|---|---|---|---|
| $D^{Lender}$ | .0039*** | .0058*** | -.0041** | -.0049*** |
| | (.0013) | (.0011) | (.0011) | (.0011) |
| $D^{Rival,1}$ | .0090*** | | | |
| | (.0014) | | | |
| $D^{Rival,2}$ | | .0080*** | | |
| | | (.00051) | | |
| $D^{Rival,3}$ | | | .0183*** | |
| | | | (.00088) | |
| $D^{Rival,4}$ | | | | .0171*** |
| | | | | (.00067) |
| Controls | √ | √ | √ | √ |
| Year FE | √ | √ | √ | √ |
| # of Obs. | 520,534 | 405,960 | 519,613 | 518,127 |

*Note:* Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively. The table the estimation results, where each column uses a different measure of distance to rival bank branches: (1) $D^{Rival,1}$ is the distance to the nearest bank branch for firm $i$ (or the second-nearest branch if the lending branch is the nearest), (2) $D^{Rival,2}$ is the distance to the nearest branch within the same bank type, (3) $D^{Rival,3}$ is the average distance to the three closest bank branches for firm $i$, (4) $D^{Rival,4}$ is the average distance to the five closest bank branches for firm $i$. The estimation also includes dummy variables for the bank type of the lending branch, return on assets (ROA), capital ratios, liquidity ratios, the logarithm of total assets, and long-term loan ratios for firm $i$ in year $t$. We also include time fixed effects, denoted by $\text{Year}_t$, to capture macroeconomic factors affecting interest rates. The sample for this analysis is restricted to firms borrowing from one of their five nearest bank branches.

Tables 3 and 4.

# 3 Structural Model of Bank Competition

This section presents our structural model to investigate the relationship between distance and firms' borrowing costs. In the model, firms choose among nearby bank branches based on offered interest rates, geographic distance, and other branch characteristics such as bank type. Banks compete in interest rates under Bertrand competition, and the resulting equilibrium lending rates are analogous to the outcomes of an English auction. In what follows, first, we describe the bank's loan supply. We then describe how firms choose their lenders and how their lending rates are determined in equilibrium. Finally, we discuss parameter identification and construct a likelihood function to estimate those structural parameters. To simplify notation, we suppress the time subscript $t$ throughout this section whenever there is no confusion.

## 3.1 Bank's Lending Cost and Distance

When bank $j$ extends loans to firm $i$ in its vicinity, it incurs the lending cost $\underline{r}_{ij}$. In the model, the lending cost is assumed to be decomposed as:

$$\underline{r}_{ij} \equiv c_i + \omega_{ij}, \tag{1}$$

where $c_i$ is a deterministic firm-specific lending cost shared by all banks and $\omega_{ij}$ is a stochastic and idiosyncratic lending cost for bank $j$ to extend loans to firm $i$. We assume that both are observed by all banks and the borrower, but not by econometricians. The firm-specific lending cost is modeled as a function of firm $i$'s characteristics and is defined as $c_i \equiv x_i\beta$, where $x_i$ is a vector of firm $i$'s characteristics, including profitability, capital ratios, and asset size.

The stochastic and idiosyncratic lending cost $\omega_{ij}$ is assumed to follow a Type I extreme value distribution,

$$\omega_{ij} \sim \text{TIEV}(\xi_{ij}, \sigma_\omega). \tag{2}$$

where $\sigma_w$ is a scale parameter that determines the variance of the distribution. We assume that distance from borrowers affects banks' lending costs by influencing the location parameter, $\xi_{ij}$, which determines the mean of the cost distribution. Specifically, we assume that $\xi_{ij}$ is a function of the distance between bank $j$ and firm $i$, $d_{ij}$, bank type $T_j$, and the interaction between bank type and firm size $Size_i$:

$$\xi_{ij} = \xi(d_{ij}, T_j, T_j \times Size_i).$$

The empirical and theoretical literature argues that lending costs may increase with distance because banks must incur higher monitoring costs for more distant borrowers (e.g., Petersen and Rajan, 2002). This implies that $\partial \xi_{ij}/\partial d_{ij} > 0$, so that greater distance from borrowers increases the idiosyncratic lending cost $\omega_{ij}$ in Equation (1) by shifting the mean of the distribution, $\xi_{ij}$. We further allow $\xi_{ij}$ to depend on bank type $T_j$ and its interaction with firm size, reflecting segmentation in lending markets across bank types and firm sizes.

## 3.2 Firm Preference and Equilibrium Interest Rate

Firm $i$ maximizes the following indirect utility by choosing a lender to finance its business investment,

$$\max_{j \in \mathcal{J}_i} \; v_{ij} - r_{ij}$$

where $\mathcal{J}_i$ is the set of nearby bank branches, $v_{ij}$ is firm $i$'s perceived benefit on borrowing from branch $j$, and $r_{ij}$ is the interest rate offered by branch $j$ to firm $i$. The firm's perceived benefit $v_{ij}$ is assumed to be a function of distance,

$$v_{ij} = v(d_{ij}),$$

where $d_{ij}$ is the distance between firm $i$ and bank branch $j$.[7] The previous literature argues that borrowers may incur higher costs when borrowing from more distant lenders due, for example, to higher transportation costs (e.g., Degryse and Ongena, 2005). If greater distance entails higher borrowing costs, we expect $\partial v(d_{ij})/\partial d_{ij} \leq 0$. Note that the effect of distance on $v_{ij}$ captures costs and benefits *for borrowers*, over and above the effect of distance on lending costs *for banks*, which is reflected in $r_{ij}$.

In addition, the estimation results in Table 2 indicate that firms with larger loan volumes are less sensitive to distance when choosing their lender. To capture such heterogeneity by loan size, we allow the marginal effect of distance on $v_{ij}$ to vary with loan size. In particular, if distance is less important for firms with larger loans, we expect

$$\frac{\partial^2 v(d_{ij})}{\partial d_{ij}\,\partial LoanSize_{ij}} \geq 0,$$

where $LoanSize_{ij}$ denotes the loan size for firm $i$ and branch $j$. The left-hand side is expected to be positive, as $v_{ij}$ and $r_{ij}$ represent the benefit and cost *per dollar of the loan*. That is, given that transportation costs are largely invariant to loan size, their importance per dollar borrowed should decline as loan size increases due to the economies of scale.

Firm $i$ chooses its lender from nearby branches. We assume that the lending branch has bargaining power in loan negotiations and sets the loan rate to maximize its profit under Bertrand competition. Accordingly, the equilibrium interest rate, $r_i^*$, is determined through an English auction, in which the most efficient branch offers the interest rate that makes firm $i$ indifferent between borrowing from the most efficient branch and the second-efficient branch under the zero-profit condition.

To characterize the equilibrium interest rate more formally, bank branches in $\mathcal{J}_i$ are sorted in descending order according to the benefit they provide to firm $i$ under $r_{ij} = \underline{r}_{ij}$ (i.e., under the zero-profit condition) and are numbered sequentially.

$$v_{i(1)} - \underline{r}_{i(1)} > v_{i(2)} - \underline{r}_{i(2)} > \cdots > v_{i(n)} - \underline{r}_{i(n)}$$

[7] Given that the borrower's perceived utility determines loan demand in the model, the above specification may appear overly simple, as it does not include firm-specific factors such as the borrower's growth expectations. However, as discussed below, this specification, which depends only on distance, is an innocuous simplification because any firm-specific factors in $v_{ij}$ do not affect the equilibrium interest rate in the model and therefore have no impact on the empirical results.

Then, we can show that, in equilibrium under Bertrand competition, firm $i$ chooses its lender $b_i^*$ to realize the (potentially) highest benefit,

$$b_i^* = b_{i(1)} \equiv \underset{j \in \mathcal{J}_i}{\text{argmax}}\ v_{ij} - \underline{r}_{ij}$$

and the equilibrium interest rate $r_i^*$, which by definition is equal to the interest rate offered by the lending branch, $r_{i(1)}$, is determined as

$$r_i^* = r_{i(1)} = v_{i(1)} - v_{i(2)} + \underline{r}_{i(2)}, \tag{3}$$

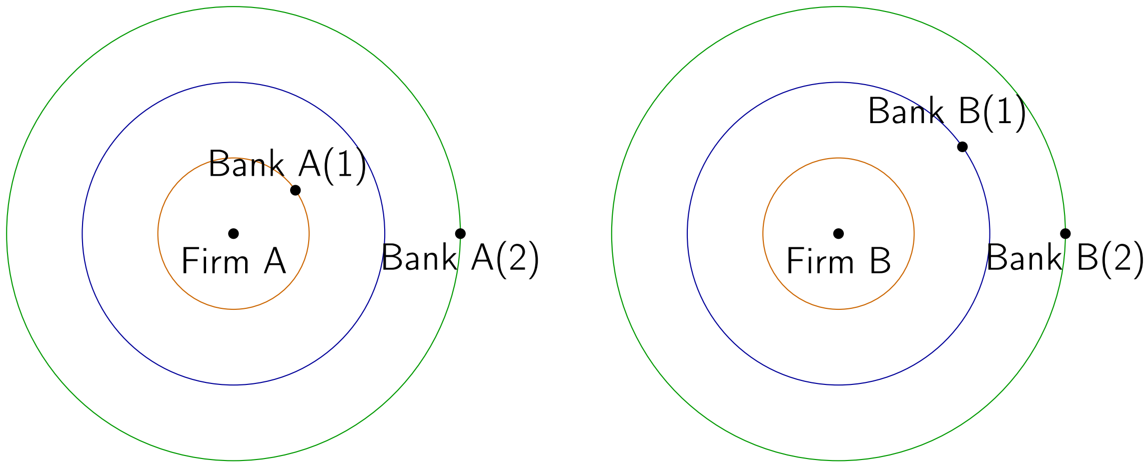
while all other banks offer interest rates equal to their zero-profit levels, $r_{i(k)} = \underline{r}_{i(k)}$ for $k = 2, \cdots, n$. The equilibrium condition for the interest rate in equation (3) follows from the fact that the equilibrium interest rate is determined through an English auction among banks. Specifically, the most efficient bank branch raises its offered interest rate, $r_{i(1)}$, up to the point at which firm $i$ is indifferent between borrowing from the most efficient branch and borrowing from the second-efficient branch under the zero-profit condition, that is, $v_{i(1)} - r_{i(1)} = v_{i(2)} - \underline{r}_{i(2)}$. The pair $(b_i^*, r_i^*)$ constitutes an equilibrium outcome, as neither the lending bank nor the other banks have an incentive to deviate from their offered interest rates, and firm $i$ likewise has no incentive to switch lenders.

A notable feature of the equilibrium condition for the interest rate in equation (3) is that the equilibrium interest rate does not depend on the lender's lending cost $\underline{r}_{i(1)}$, while it does depend on the *rival's* lending cost $\underline{r}_{i(2)}$. Consequently, the distance between firm $i$ and its lender affects the interest rate only through firm $i$'s perceived utility $v_{i(1)}$. This mechanism implies a negative effect of borrower-lender distance on the interest rate, which is consistent with the estimation results reported in previous empirical studies as well as those in Table 4.

## 3.3 Identification for Borrowers' and Lenders' Cost of Distance

In general, the distance between borrowers and lenders affects both lenders' costs (e.g., monitoring costs) and borrowers' costs (e.g., transportation costs). Separately identifying these two effects using only observed interest rates and observed borrower–lender

Figure 4: Identification for Lending and Borrowing Costs 2




distances is therefore challenging. Our model and data allow us to overcome this identification problem. In particular, our dataset contains information not only on the branch chosen by each firm, but also on nearby competing branches that were not selected. This feature is important because equilibrium pricing depends not only on the characteristics of the chosen lender, but also on the competitive pressure exerted by alternative lenders. The key intuition of our identification strategy is that, by comparing firms facing similar nearby competing branches, we can effectively control for lender-side cost variation and isolate the effect of distance on borrowers' perceived utility.

To illustrate our identification strategy, consider two otherwise similar firms, $A$ and $B$, operating in comparable environments. Suppose that, after controlling for observable factors, the distances from the two firms to their second-nearest (i.e., second-highest utility) branches are identical, while the distances to their respective lending branches differ. As Figure 4 demonstrates, Bank A(2) and Bank B(2) are the second-nearest branches for Firm A and Firm B, respectively, and are located at the same distance from the firms. In contrast, Bank A(1) is located closer to Firm A than Bank B(1) is to Firm B.

In equilibrium, the interest rates faced by Firms $A$ and $B$ are given by $r_A^* = v_{A(1)} - v_{A(2)} + \underline{r}_{A(2)}$ and $r_B^* = v_{B(1)} - v_{B(2)} + \underline{r}_{B(2)}$, respectively. Since the distances to the second-nearest branches are identical for firms $A$ and $B$, we have

$$r_A^* - r_B^* = v_{A(1)} - v_{B(1)}.$$

Therefore, conditional on Firms $A$ and $B$ being otherwise similar, any observed difference in equilibrium interest rates between the two firms reflects differences in firms' perceived utility associated with their respective lending branches, rather than differences in banks' lending costs.

Now suppose that firms' perceived utility is linear in distance as in our structural model. Under this assumption, the equation above implies that we can identify the effect of distance on firms' perceived utility by comparing the equilibrium interest rates faced by Firms $A$ and $B$. Once the effect of distance on firms' perceived utility is identified, we can then recover the distance-related component of banks' lending costs, $\underline{r}_{i(2)}$, from the equilibrium pricing condition in (3).

Note that, for expositional simplicity, Figure 4 illustrates the case in which the second-efficient branch coincides with the physically second-nearest branch. More generally, however, the relevant comparison branch in our model is defined by firms' utility rankings rather than geographic distance alone.

### 3.4 Likelihood Function for Estimating Structural Parameters

We construct a likelihood function to estimate the parameter values in the following two steps. First, the probability for firm $i$ to choose bank branch $m$ as its lender, $\mathcal{L}^1(b_i = m)$, is computed by:

$$\mathcal{L}^1(b_i = m) \quad = \quad \mathbf{Pr}\left[v_{im} - \underline{r}_{im} \geq v_{ij} - \underline{r}_{ij}, \quad \forall j \in \mathcal{J}_i \backslash \{m\}\right]$$

Next, under the condition that firm $i$ chooses branch $m$ as their lender, we compute the conditional probability of the equilibrium interest rate being equal to $r_i$, $\mathcal{L}^2(r_i|b_i = m)$. Given the equilibrium interest rate is determined by Equation (3), $\mathcal{L}^2(r_i|b_i = m)$ is computed by:

$$\begin{aligned} \mathcal{L}^2(r_i|b_i = m) &= \mathbf{Pr}\left[r_i = v_{im} - v_{i(2)} + \underline{r}_{i(2)}\right] \\ &= \mathbf{Pr}\left[\omega_{i(2)} = -v_{im} + v_{i(2)} + r_i - x_i\beta\right] \end{aligned}$$

where the variables $v_{i(2)}, \underline{r}_{i(2)}$, and $\omega_{i(2)}$ are variables for the second-efficient branch for firm $i$. Finally, by computing the product of $\mathcal{L}^1(b_i = m)$ and $\mathcal{L}^2(r_i|b_i = m)$, we can compute the likelihood that firm $i$'s lender is branch $m$ and the interest rate is $r_i$:

$$\mathcal{L}(b_i = m, r_i) = \mathcal{L}^1(b_i = m) \cdot \mathcal{L}^2(r_i|b_i = m)$$

and estimate the structural parameters using maximum likelihood estiamtion.

The main challenge in calculating $\mathcal{L}(b_i = m, r_i)$ is that evaluating $\mathcal{L}^2(r_i|b_i = m)$ requires variables associated with the second-efficient branch, namely $v_{i(2)}$, $\underline{r}_{i(2)}$, and $\omega_{i(2)}$. Although we can observe which bank is the lender for firm $i$, we cannot identify which branch is the second-efficient branch for that firm. To address this challenge, our strategy is to account for all cases in which one of the nearby banks other than the lender serves as the second-efficient branch for firm $i$. Specifically, to compute $\mathcal{L}^2(r_i|b_i = m)$, we first calculate the probability that each branch $k \in \mathcal{J}_i \backslash \{m\}$ is the second-efficient branch. Then, conditional on branch $k$ being the second-efficient branch, we compute the probability that the equilibrium interest rate equals firm $i$'s observed interest rate, $r_i$. Finally, we sum these probabilities over all nearby branches $k \in \mathcal{J}_i \backslash \{m\}$, weighting each term by the likelihood that branch $k$ is the second-efficient branch.

Let branches $m$ and $k$ denote the lender and the second-efficient branch for firm $i$, respectively. Then, conditional on branch $m$ being the lender for firm $i$, the conditional probability that branch $k$ is the second-efficient branch is given by:

$$\begin{aligned}
\mathbf{Pr}\left[v_{ik} - \underline{r}_{ik} \geq v_{il} - \underline{r}_{il}, \quad \forall l \in \mathcal{J}_i \backslash \{m, k\}\right] &= \mathbf{Pr}\left[v_{im} - r_i \geq v_{il} - \underline{r}_{il}, \quad \forall l \in \mathcal{J}_i \backslash \{m, k\}\right] \\
&= \mathbf{Pr}\left[\omega_{il} \geq -v_{im} + v_{il} + r_i - x_i\beta, \quad \forall l \in \mathcal{J}_i \backslash \{m, k\}\right] \\
&= \prod_{l \in \mathcal{J}_i \backslash \{m,k\}} \{1 - G_{il}(\psi_{iml})\}
\end{aligned}$$

where $\psi_{imj} \equiv -v_{im} + v_{ij} + r_i - x_i\beta$, and $G_{ij}$ is a cdf of TIEV$(\xi_{ij}, \sigma_w)$ for firm $i$ and branch $j$. The equality in the first line utilizes the equilibrium condition for the interest rate in equation (3). Notably, no variable specific to the second-efficient branch $k$ appears on the right-hand side of the first line. This is because, under an English auction, all relevant information about the second-efficient branch is fully summarized by the observed equilibrium interest rate $r_i$. The second and third lines follow from the assumption that the lending cost $\underline{r}_{il}$ is additively separable as described in equation (1), and the idiosyncratic component $\omega_{il}$ follows TIEV$(\xi_{il}, \sigma_w)$.

Given the probability that each branch $k \in \mathcal{J}_i \backslash \{m\}$ is the second-efficient branch for firm $i$, the conditional likelihood that firm $i$'s interest rate equals $r_i$, denoted by $\mathcal{L}^2(r_i|b_i = m)$, is given by:

$$\mathcal{L}^2(r_i|b_i = m) = \sum_{k \in \mathcal{J}_i \backslash \{m\}} g_{ik}(\psi_{imk}) \left[ \prod_{l \in \mathcal{J}_i \backslash \{m,k\}} \{1 - G_{il}(\psi_{iml})\} \right]$$

Here, $g_{ij}$ denotes the probability density function (pdf) of $\text{TIEV}(\xi_{ij}, \sigma_w)$ for firm $i$ and bank branch $j$. According to the equilibrium condition for the interest rate in equation (3), the term $g_{ik}(\psi_{imk})$, where $\psi_{imk} \equiv -v_{im} + v_{ik} + r_i - x_i\beta$, represents the conditional density of the equilibrium interest rate being equal to $r_i$, conditional on branches $m$ and $k$ being the lender and the second-efficient branch, respectively. Thus, $\mathcal{L}^2(r_i|b_i = m)$ can be computed as a weighted average of $g_{ik}(\psi_{imk})$, where the weight for each branch $k$ corresponds to the likelihood that it is the second-efficient branch for firm $i$.

Finally, by computing the product of $\mathcal{L}^1(b_i = m)$ and $\mathcal{L}^2(r_i|b_i = m)$, we can compute the likelihood for estimation as follows,

$$\begin{aligned}
& \mathcal{L}(b_i = m, r_i) \\
= \; & \mathcal{L}^1(b_i = m) \times \mathcal{L}^2(r_i|b_i = m) \\
= \; & \mathbf{Pr}\left[v_{im} - \underline{r}_{im} \geq v_{ij} - \underline{r}_{ij}, \quad \forall j \in \mathcal{J}_i \backslash \{m\}\right] \times \sum_{k \in \mathcal{J}_i \backslash \{m\}} g_{ik}(\psi_{imk}) \left[ \prod_{l \in \mathcal{J}_i \backslash \{m,k\}} \{1 - G_{il}(\psi_{iml})\} \right] \\
= \; & G_{im}(r_i - x_i\beta) \sum_{k \in \mathcal{J}_i \backslash \{m\}} g_{ik}(\psi_{imk}) \left[ \prod_{l \in \mathcal{J}_i \backslash \{m,k\}} \{1 - G_{il}(\psi_{iml})\} \right] \qquad (4)
\end{aligned}$$

where $\psi_{imj} \equiv -v_{im} + v_{ij} + r_i - x_i\beta$. The equality in the last line follows from the fact that the equilibrium interest rate is determined by an English auction, as described in equation (3). Note that the likelihood function depends only on observable variables for firm $i$, the lender $m$, and the nearby branches in $\mathcal{J}_i$. In the next section, we use this likelihood function to estimate the structural parameters by maximum likelihood estimation.

# 4 Structural Estimation

In this section, we conduct a quantitative analysis using the structural model described in the previous section. First, we parameterize the relationship between distance and lending costs, as well as perceived utility. Second, prior to estimating the model parameters using data, we conduct a Monte Carlo simulation to examine the model's properties. Third, we estimate the model parameters using our dataset and interpret the estimation results. Finally, we conduct a counterfactual simulation of bank branch closures to assess their impact on the equilibrium interest rate.

## 4.1 Parameterization

To investigate the structural relationship between distance and interest rates, we assume the following specifications of firm $i$'s utility, $v_{ij}$, when it borrows from bank branch $j$, as well as for lending costs, $c_i$ and $\xi_{ij}$:

$$\begin{aligned}
v_{ij} &\equiv \gamma_D\, d_{ij} + \gamma_{DL}\, d_{ij} \cdot \log(\text{LoanSize}_i), \\
c_i &\equiv \beta_S \text{FirmSize}_i + \beta_R \text{ROA}_i + \beta_Q \text{Liquidity}_i + \beta_C \text{CapitalRatio}_i + \beta_L \text{LongTermRatio}_i \\
\xi_{ij} &\equiv \theta_D\, d_{ij} + \theta_{D2}\, d_{ij}^2/1000 + \sum_{b\in\{1,2\}} \left[\theta_{T,b} \text{BankType}_{b,j} + \theta_{TS,b} \text{BankType}_{b,j} \cdot \text{FirmSize}_i\right]
\end{aligned}$$

In this specification, the distance between borrowers and lenders, $d_{ij}$, affects interest rates through both borrowers' perceived utility and banks' lending costs, captured by $\gamma_D$, $\gamma_{DL}$, $\theta_{D1}$, and $\theta_{D2}$. The effect of distance on lending costs is allowed to be non-linear, as captured by the quadratic term. In addition, the effect of distance on borrowers' perceived utility varies with loan size, as captured by the interaction term coefficient $\gamma_{DL}$, consistent with the evidence in Table 2, which shows that preferences regarding distance differ across firms with different loan sizes. With respect to bank type, we classify all banks into three categories: first-tier regional banks, second-tier regional banks, and credit associations. We identify level effects for each bank type and their interactions with firm size by setting first-tier regional banks as the base group and labeling second-tier regional banks and credit associations as $b = 1$ and $b = 2$, respectively.

## 4.2 Monte Carlo Simulation

Prior to estimating the model parameters using data, we conduct a Monte Carlo simulation to examine the model's properties. First, by performing maximum likelihood estimation of the structural parameters using simulated data, we assess whether the parameter values are properly identified by the likelihood function developed in the previous section. Second, by conducting reduced-form estimations using the simulated data, we aim to interpret the reduced-form estimation results in Section 2 through the lens of the structural assumptions of our model. The Monte Carlo simulation proceeds as follows. We first generate 10,000 artificial firms, specifying their geographic characteristics and financial environments. Given the equilibrium conditions of the model, these artificial firms optimally choose their lender, and the interest rate is determined through an English auction à la Bertrand competition among bank branches. See Appendix B for more details.

First, the maximum likelihood estimation using the simulated data indicates that the parameter values are properly identified by the likelihood function developed in the previous section. Although the detailed results are reported in Appendix B, the maximum likelihood estimation based on our likelihood function—which depends only on observable variables for firm $i$, the lender $m$, and the nearby branches in $\mathcal{J}_i$—successfully recovers the calibrated parameter values used in the Monte Carlo simulation. In particular, consistent with the argument in Section 3.3, the maximum likelihood estimation is able to separately and precisely identify the parameters governing the effects of distance on borrowers' costs, $\gamma_D$, and those on lenders' costs, $\theta_D$.

Second, we conduct reduced-form estimations using the simulated data to interpret them through the lens of our structural model. Specifically, using the data generated through the Monte Carlo simulation, we estimate the following reduced-form specification:

$$\text{Interest Rate}_i = \beta_0 + \beta_1 D_i^{Lend} + \beta_2 D_i^{Rival} + \epsilon_i \tag{5}$$

Table 5: Estimated Coefficients Using Simulated Data

| | [i] $(\gamma_D, \theta_D) = (0.0, 0.1)$ | [ii] $(\gamma_D, \theta_D) = (-0.1, 0.0)$ | [iii] $(\gamma_D, \theta_D) = (-0.1, 0.1)$ |
|---|---|---|---|
| $\beta_1$ | 0.028 | -0.072 | -0.042 |
| | [0.022, 0.033] | [-0.077, -0.066] | [-0.048, -0.034] |
| $\beta_2$ | 0.065 | 0.064 | 0.103 |
| | [0.045, 0.084] | [0.045, 0.084] | [0.084, 0.125] |

Note: The table reports estimation results using artificial data generated by Monte Carlo simulations. Columns 1, 2, and 3 correspond to three cases with different parameter values for the effect of distance on lending costs, $\theta_D$, and on perceived utility, $\gamma_D$. The reported values are the medians of the estimated coefficients from 100 simulations, with the 5th and 95th percentiles shown in brackets.

where $D_{\text{Lend},i}$ denotes the distance to the lending branch, and $D_{\text{Rival},i}$ denotes the distance to the nearest bank branch for firm $i$ (or the distance to the second-nearest bank branch when the lender is the nearest branch). Because the definitions of the explanatory variables are identical to those used in the estimation in Section 2, the coefficients $\beta_1$ and $\beta_2$ are comparable to the corresponding reduced-form estimates in that section.

Prior to discussing the estimation results, it is useful to theoretically examine how the structural parameters, $\theta_D$ and $\gamma_D$, map into the coefficients, $\beta_1$ and $\beta_2$, in our model. The equilibrium condition in equation (3) implies that the equilibrium interest rate is determined through an English auction and can be written as, $r_i^* = v_{i(1)} - v_{i(2)} + \underline{r}_{i(2)}$. Accordingly, the *true* marginal effects of distance to the lender and to the rival branch on interest rates, $\beta_1$ and $\beta_2$, are analytically given by:

$$\beta_1 \equiv \frac{\partial r_i^*}{\partial D^{Lend}} = \gamma_D \quad \text{and} \quad \beta_2 \equiv \frac{\partial r_i^*}{\partial D^{Rival}} = \theta_D - \gamma_D \tag{6}$$

This implies that $\gamma_D$ has negative and positive effects on the marginal effects of $D^{Lend}$ and $D^{Rival}$, respectively, while $\theta_D$ has positive effects on the marginal effects of $D^{Rival}$.

Table 5 shows the estimation results using the artificial data generated by the Monte Carlo simulation. Columns 1, 2, and 3 show the three cases with different parameter values for the effects of distance on the lending cost, $\theta_D$, and that on the borrower's perceived utility, $\gamma_D$. By differentiating these parameter values, we can assess how those differences affect reduced-form estimation results. In all cases, the first and second rows show the estimated values for the marginal effects of distance to the lending branch and the rival branch, $\beta_1$ and $\beta_2$ in equation (5), respectively. The reported values are the medians of the estimated coefficients from 100 simulations, with the 5th and 95th percentiles shown in brackets.

Some comments on the estimation results are in order. First, the estimation results for different calibrated parameter values are broadly consistent with the theoretical implications in equation (6). Specifically, when comparing cases (i) and (iii), an increase in the absolute value of $\gamma_D$ (i.e., a stronger effect of distance on the borrower's perceived utility) makes the estimated values of $\beta_1$ and $\beta_2$ more negative and more positive, respectively. Likewise, when comparing cases (ii) and (iii), an increase in the absolute value of $\theta_D$ (i.e., a stronger effect of distance on the lender's cost) leads to a more positive estimated value of $\beta_2$. Second, the estimation results in Table 5 provide useful insights for interpreting the reduced-form estimations in Section 2, as well as those in previous empirical studies. The reduced-form results reported in Tables 3 and 4 indicate that distance to the rival branch has a positive and relatively large effect on loan rates, whereas distance to the lending branch has a slightly negative effect. These patterns are more consistent with case (iii) than with cases (i) or (ii), suggesting that geographic distance affects both lenders' costs and borrowers' utility.

Third, Table 5 implies that reduced-form estimations may yield biased estimates of the marginal effects of distance. Specifically, the estimates of $\beta_1$ and $\beta_2$ are positively and negatively biased, respectively. For example, equation (6) implies that the estimated values of $\beta_1$ and $\beta_2$ in case (i) should be centered around $0.0$ and $0.1$. However, the estimated values in case (i) are $0.028$ and $0.065$, indicating that the reduced-form estimates of $\beta_1$ and $\beta_2$ are positively and negatively biased. The mechanisms underlying these biases can be understood as follows. First, the estimate of $\beta_1$ is positively biased because distance to the lender is positively correlated with unobservable lending costs of the second-efficient branch, $\omega_{i(2)}$. That is, when a nearby branch's lending cost is high for idiosyncratic reasons (i.e., a high $\omega_{i(2)}$), firms optimally choose a more distant branch as their lender (i.e., a larger $d_{i(1)}$), generating a positive correlation between $\omega_{i(2)}$ and $d_{i(1)}$ and leading $\beta_1$ to be positively biased. Second, the estimate of $\beta_2$ is negatively biased due probably to attenuation bias arising from an errors-in-variables problem. As discussed above, we cannot identify the second-efficient branch from the data. Consequently, the distance to the second-nearest bank serves as a reasonable proxy for $d_{i(2)}$ but contains measurement error, which biases the estimate of $\beta_2$ toward zero.

In sum, the Monte Carlo simulation indicates that distance affects loan rates by influencing both lenders' costs and borrowers' costs. The estimation using simulated data shows that reduced-form estimation is subject to bias due to firms' endogenous lender choice, whereas the maximum likelihood estimation can properly identify the structural parameter values. Hence, a structural estimation approach offers advantages in measuring the marginal effects of distance on loan rates more precisely and provides a useful tool for conducting counterfactual analyses.

## 4.3 Estimation Results

In this subsection, we estimate the parameter values using maximum likelihood, based on the likelihood function presented in the previous section. As we describe in Section 2 and Appendix A, our dataset consists of the following items for firms and bank branches, which provide sufficient information to compute likelihood in equation (4) for each firm. First, the firm-level microdata provide firm $i$'s borrowing cost $r_i$, the address of its headquarters, and a vector of characteristics $x_i$, such as profitability, firm size, and debt levels. Second, the financial institution directory provides the bank branch addresses and the bank types of the lender, $T_m$, as well as other banks $T_j$ where $j \in \mathcal{J}_i \backslash \{m\}$. Third, by combining those two datasets, we can calculate the distance between firm $i$ and its lender, $d_{im}$, as well as other bank branches in its neighborhood, $d_{ij}$, where $j \in \mathcal{J}_i \backslash \{m\}$.

Table 6 reports parameter estimates obtained by maximizing the likelihood function in equation (4), based on the model specified in Section 4.1. Columns (i) and (ii) present estimates without and with the interaction term between distance and loan volume in firms' perceived utility, respectively. Column (iii) reports estimates based on the subsample of firms that borrow from a single bank branch, as a robustness check.

First, the estimation results indicate that the distance between borrowers and lenders affects interest rates through both borrowers' perceived utility and banks' lending costs.

Table 6: Estimation Results of Structural Parameters

| | Baseline (i) | With interaction (ii) | Single lender (iii) |
|---|---|---|---|
| Borrower's utility | | | |
| $\gamma_{Dist}$ | -0.022 (0.014) | -0.241 (0.005) | -0.279 (0.013) |
| $\gamma_{Dist \times LoanSize}$ | | 0.019 (0.006) | 0.023 (0.016) |
| Lender's cost | | | |
| $\theta_{Dist}$ | 0.065 (0.005) | 0.059 (0.006) | 0.085 (0.011) |
| $\theta_{Dist^2}$ | -0.198 (0.004) | -0.113 (0.013) | -0.198 (0.020) |
| $\theta_{Type,1}$ | -1.625 (0.017 | -1.551 (0.018) | -1.526 (0.051) |
| $\theta_{Type \times Size,1}$ | 0.186 (0.012) | 0.179 (0.012) | 0.205 (0.034) |
| $\theta_{Type,2}$ | -2.567 (0.010) | -2.511 (0.009) | -2.223 (0.031) |
| $\theta_{Type \times Size,2}$ | 0.289 (0.007) | 0.284 (0.007) | 0.274 (0.022) |
| $\beta_{FirmSize}$ | -0.183 (0.007) | -0.165 (0.007) | -0.164 (0.023) |
| $\beta_{ROA}$ | 1.398 (0.009) | 1.410 (0.008) | 1.000 (0.023) |
| $\beta_{Liquidity}$ | 0.264 (0.020) | 0.261 (0.020) | 0.395 (0.032) |
| $\beta_{CapitalRatio}$ | -0.473 (0.011) | -0.538 (0.010) | -0.089 (0.139) |
| $\beta_{LongTermRatio}$ | 0.351 (0.011) | 0.352 (0.011) | 0.455 (0.021) |
| Num. of Obs. | 871,527 | 871,527 | 174,776 |

*Note*: The table reports parameter estimates obtained by maximizing the likelihood function in equation (4), based on the model specified in Section 4.1. Columns (i) and (ii) present estimates without and with the interaction term between distance and loan volume in firms' perceived utility, respectively. Column (iii) reports estimates based on the subsample of firms that borrow from a single bank branch, as a robustness check. The numbers in parentheses show the standard errors.

Column (i) reports negative and positive estimates for $\gamma_{Dist}$ and $\theta_{Dist}$, respectively, indicating that greater distance lowers borrowers' perceived utility while increasing banks' lending costs. These estimates are consistent with the existing literature, which argues that greater distance increases monitoring costs for lenders and transportation costs for borrowers. Second, the results in Column (ii) suggest that the effects of distance on perceived utility are heterogeneous across firms. Given the positive and significant interaction term between distance and loan size, $\gamma_{Dist\times LoanSize}$, the marginal effect of distance on perceived utility is $-0.0491(= -0.241 + 0.019 \times 10.1)$ for smaller firms with loan volumes at the 25th percentile, whereas it is $-0.0054(= -0.241 + 0.019 \times 12.4$ for larger firms with loan volumes at the 75th percentile. This pattern suggests that the negative effects of distance on borrowers are more pronounced for smaller firms, consistent with the reduced-form evidence reported in Table 2. Moreover, the estimate of $\gamma_{Distance}$ in Column (i), which imposes homogeneous sensitivity to distance, is negative but imprecisely estimated, whereas the estimate in Column (ii) is more precise. Accordingly, in the counterfactual analysis, we adopt the specification that allows for heterogeneous effects of distance by loan size.

To assess the economic magnitude, we translate these estimates into marginal effects per kilometer. The estimate in Column (i) implies that a 1 km increase in the distance to the rival branch (i.e., the second-most efficient branch) raises borrowing costs by approximately 8.7 bps $(= \theta - \gamma)$.[8] In contrast, the effect of distance to the lending branch is more moderate, estimated at $-2.2$ bps $(= \gamma)$ per additional kilometer. These magnitudes are economically meaningful given the low interest rate environment in Japan. Turning to heterogeneity across firms, based on the estimates in Column (ii), the effects of distance to both the rival branch and the lending branch vary across firms with different loan volumes. Specifically, for small firms with loan volumes at the 25th percentile, a 1 km increase in the distance to the lending branch and the rival branch raises borrowing costs by about $4.91$ bps and $10.81(= 5.90 + 4.91)$ bps, respectively. For large firms with loan volumes at the 75th percentile, the corresponding effects are $0.54$ bps

[8]Note that although the coefficient on the quadratic term, $\theta_{Distance,2}$, is negative and statistically significant, its magnitude is small unless the rival bank is located at a sufficiently large distance. This suggests that the effect of distance on borrowing costs is approximately linear.

and $6.44(= 5.90 + 0.54)$ bps, respectively. These results suggest that borrowing costs for smaller firms are more sensitive to geographic distance than those for larger firms.

While these effects of distance are broadly consistent with the reduced-form estimates reported in Section 2, their magnitudes differ substantially. In particular, the reduced-form estimates imply effects of at most 1.8 bps for distance to rival branches and $-0.5$ bps for distance to lending branches, which are much smaller in absolute value than the corresponding structural estimates. The Monte Carlo simulation in the previous section suggests that such attenuation arises from a simultaneity problem, as reduced-form regressions do not account for firms' endogenous lender choice. Therefore, to assess the effects of bank consolidation more accurately, counterfactual simulations based on the estimated structural model are particularly informative.

One concern with our structural estimation is that, while the model assumes a single lending relationship, many firms in the data borrow from multiple bank branches. In our sample, 20.0% of firms borrow from a single branch, whereas a substantial fraction borrow from multiple branches, with approximately 33% borrowing from more than three branches. In the baseline specification, we interpret a firm's main bank as the lending branch in the model and implicitly assume that it primarily determines the firm's overall borrowing cost, while other banks act as potential competitors. Under this assumption, we use the firm-level average borrowing rate as the dependent variable, since the data do not contain loan-level interest rates or loan shares. If borrowing from secondary lenders is quantitatively important and their rates differ substantially from that of the main bank, the observed average rate may be measured with error relative to the model-implied rate, potentially attenuating the estimated effects. In addition, firms that borrow from multiple banks may face a different competitive environment, which could alter the role of geographic distance.

To address these concerns, we re-estimate the model using only firms that borrow from a single bank branch, for which the mapping between the model and the data is most direct. The results, reported in Column (iii), are quantitatively similar to the baseline estimates and confirm that geographic distance affects both lenders' costs and borrowers' perceived utility. This finding indicates that our main results are not driven by multi-bank borrowing and that the mapping from the model-implied rate to the observed firm-level borrowing cost is not materially distorted by the presence of secondary lending relationships. While the single-branch subsample represents a subset of firms, the close correspondence of the estimates suggests that the baseline results capture a general mechanism linking geographic proximity and borrowing costs.

# 5 Counterfactual Simulation

This section conducts a counterfactual simulation based on the estimated structural parameters reported in Table 6. Motivated by policymakers' concerns about rising borrowing costs following bank branch closures, our counterfactual analysis examines how hypothetical large-scale closures affect firms' borrowing costs by influencing the geographic distance between lenders and borrowers. Specifically, the counterfactual simulation assumes that, in each prefecture in Japan, the bank ranked second by market share in the number of borrowers (hereafter, the"second-largest bank") closes all of its branches. Following these closures, we quantify the extent to which equilibrium interest rates increase as a result of reduced competition. Although this scenario is inspired by the fact that many bank consolidations involve first-tier banks acquiring second-tier banks, we do not view it as a realistic or likely outcome in Japan. Rather, it serves as a useful benchmark for assessing the quantitative impact of bank consolidation on borrowing costs and the underlying mechanisms.

A sketch of our counterfactual analysis is as follows. First, for all firms $i$ and their potential lenders $j \in \mathcal{J}_i$ in our sample, we draw random shocks $\omega_{ij}$ following the extreme value distribution in equation (2) until we obtain $M$ draws that are consistent with the observed lender choices and interest rates in 2019.[9] Note that, since numerous combinations of shocks can rationalize the observed lender choices and interest rates, different branches are selected as the second- and third-efficient branches across draws. Sec-

[9]Because the interest rate is a continuous variable, the probability of observing the exact realized interest rate is, by definition, zero. We therefore assume the presence of a measurement error that follows a Normal distribution for the observed interest rate. We treat a simulated shock draw as consistent with the observation if the implied measurement error is less than 10% of the observed interest rate.

ond, we select the second-largest bank in each prefecture by the number of borrowers in 2019 and assume that it closes all of its branches in the prefecture where its headquarters are located. Third, we calculate the increase in borrowing costs for firm $i$ following the branch closures. In our model, branch closures affect the equilibrium interest rate for firm $i$ in the following two cases. The first case is the closure of the lending branch. In this case, the firm selects the second-efficient branch as a new lender, and the third-efficient branch (or the fourth-efficient branch if the third-efficient branch is closed) becomes a new rival branch. The second case is the closure of the second-efficient branch. In this case, the third-efficient branch becomes a new rival branch. The equilibrium interest rate changes in the second case, even though the lender remains unchanged, as the interest rate is determined by an English auction. The increase in borrowing costs for firm $i$ is then defined as the average increase over all $M$ draws.[10]

## 5.1 Firm-level Analysis

Table 7 presents summary statistics for the estimated increase in borrowing costs in the counterfactual analysis. The table shows that the average increase in the equilibrium interest rate is 42 bps, which is economically significant given Japan's low-interest environment. While bank consolidation has many benefits, including reduced operating costs, this result suggests that it may also be associated with sizable increases in firms' borrowing costs due to declined competition among banks. The table further shows that the 25th percentile is only 11 bps, whereas the 75th percentile exceeds 63 bps, highlighting substantial heterogeneity in the magnitude of the increase in borrowing costs across firms.

To examine the sources of heterogeneity in the effects across firms, we conduct the following regression analysis:

$$\Delta r_i = \beta_0 + \beta_L Cl_{Lend,i} + \sum_{j=1}^{30} \beta_{R,j} Cl_{Rival,i,j} + \nu_i \tag{7}$$

[10]When computing the average over $M$ draws, we weight each draw by the density of the measurement error described in the previous footnote. In our analysis, we set $M = 20$ for tractability; increasing $M$ further does not materially affect the quantitative results.

Table 7: Changes in Borrowing Costs in Counterfactual Analysis

| | Mean | Std. Dev. | Median | [0.25, 0.75] |
|---|---|---|---|---|
| Borrowing cost increase (%) | 0.42 | 0.44 | 0.30 | [0.11, 0.63] |

*Note:* The table reports summary statistics for changes in borrowing costs in the counterfactual analysis.

where $\Delta r_i$ is the change in the interest rate for firm $i$ in the counterfactual simulation, $Cl_{Lend,i}$ is a dummy variable that takes the value one if the lending branch for firm $i$ is closed, and $Cl_{Rival,i,j}$ is a dummy variable that takes the value one if the $j$th nearest branch for firm $i$ is closed. In our counterfactual analysis, the interest rate changes only if the lending branch or nearby rival branches close. Therefore, in the regression above, $\beta_L$ and $\beta_{R,j}$ capture the average effects of closures of the lending branch and nearby rival branches, respectively.

Table 8 and Figure 5 show the estimation results for equation (7). First, Table 8 indicates that firms whose lending branches are closed would face a significant increase in borrowing costs. Specifically, the estimated value of $\beta_L$ in column (1) implies that the interest rate increases by 87 bps on average when the lending branch is closed. Closures of lending branches, by definition, increase the distance to rival banks, thereby leading to higher equilibrium interest rates in our model. Furthermore, firms select their lenders based not only on distance but also on unobservable factors $\omega_{ij}$ in equation (2). Therefore, closures of lending branches mean that firm $i$ loses lenders with favorable shocks, leading to further increases in borrowing costs.

Second, the regression analysis of equation (7) also implies that closures of nearby branches increase borrowing costs. Figure 5 plots the estimated values of $\beta_{R,j}$ for $j = 1, \cdots, 17$ in equation (7), which capture the marginal effects of closures of the $j$th nearest branches. The figure shows that the closure of the nearest bank branch for firm $i$ raises its borrowing cost by 17 bps on average, and that the marginal effects decline gradually as the closed branch becomes more distant (i.e., as $j$ increases). In the counterfactual analysis, the interest rate increases only if the closed branch is the lending branch or the second-efficient branch. In that sense, it may appear somewhat surprising that the closure of relatively distant branches still has economically significant effects on borrow-

Table 8: Effects of Closure of Lending Branch

| Dependent Var. | Model (i) $\Delta r_i$ | Model (ii) $\Delta r_i$ |
|---|---|---|
| $Cl_{Lend}$ | 0.789*** | 1.028*** |
| | (0.004) | (0.014) |
| $Cl_{Lend} \times$ # of Rivals | | -0.0153*** |
| | | (-0.0008) |
| $N$ | 68,345 | 68,329 |

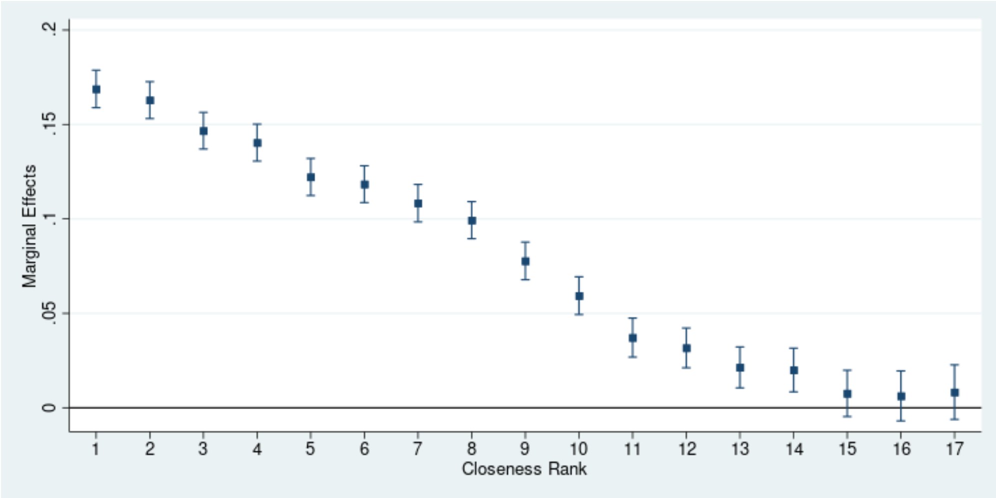
*Note:* The table shows the estimation results for $\beta_L$, the marginal effects of lending branches' closure, in (7). Columns (1) and (2) are the estimates in the specification without and with interaction terms. The numbers in parentheses show the standard errors. Significance levels are denoted by $< 0.10$ (*), $< 0.05$ (**), and $< 0.01$ (***).

ing costs (for example, the closure of the 10th-nearest branch raises borrowing costs by around 6 bps). However, given that some firms choose relatively distant branches as their lending branches as shown in Figure 2, it is not surprising that some distant branches constitute the second-efficient branches for firm $i$ as well.

For firms whose lending branch, or second-efficient branch, is closed, the magnitude of the increase in borrowing costs should depend on the availability of alternative nearby branches. We examine this hypothesis by allowing $\beta_L$ to vary with measures of availability captured by the number of bank branches in vicinity. Column (2) of Table 8 reports the estimation results for the interaction terms between $Cl_{Lend,i}$ and the number of bank branches within 15 km. The results indicate that the availability of nearby alternative branches significantly affects the marginal effects of lending-branch closures. In particular, for each additional nearby branch, the impact of a lending-branch closure is mitigated by 1.5 bps, underscoring the importance of nearby branches as alternative lenders or rivals.[11] Figure 6 presents the histogram of $\Delta r_i$ by the number of nearby bank branches. Consistent with the estimation results above, the figure indicates that

[11]While we do not report the estimation results for the interaction terms between $\beta_{R,j}$ and the number of nearby banks, these interaction terms also have statistically significant effects on changes in borrowing costs.

Figure 5: Effects of Nearby Branches' Closures




Note: The figure plots the estimated values of $\beta_{R,j}$ for $j = 1, \cdots, 17$ in (7), which capture the marginal effects of closures of the $j$th nearest branches.

the availability of nearby branches significantly affects the impact of branch closures on borrowing costs. Specifically, the magnitude of the impact is relatively large when the number of nearby branches is 10 or fewer (left panel), becomes substantially smaller when the number is between 11 and 20 (middle panel), and is quite limited when the number exceeds 20 (right panel).

## 5.2 Prefecture-level Analysis

Our counterfactual analysis assumes a common scenario across all prefectures in which the second-largest bank closes all of its branches. Nonetheless, the impact on borrowing costs varies across regions, reflecting the heterogeneous effects across firms discussed above.

Regarding the heterogeneous effects across regions, the left panel of Figure 7 shows the average increase in borrowing costs by region in our counterfactual simulation and indicates substantial heterogeneity across regions. As an example of this heterogeneity, the right panel of the figure presents the histogram of $\Delta r_i$ for two representative prefectures that experience large and small impacts in the counterfactual analysis. The panel indicates that firms in Kanagawa Prefecture, which is adjacent to Tokyo and characterized by a highly competitive banking market, face much smaller increases in borrowing

Figure 6: Borrowing Costs and Number of Nearby Banks


Note: The figure presents the histogram of $\Delta r_i$ by the number of nearby bank branches.

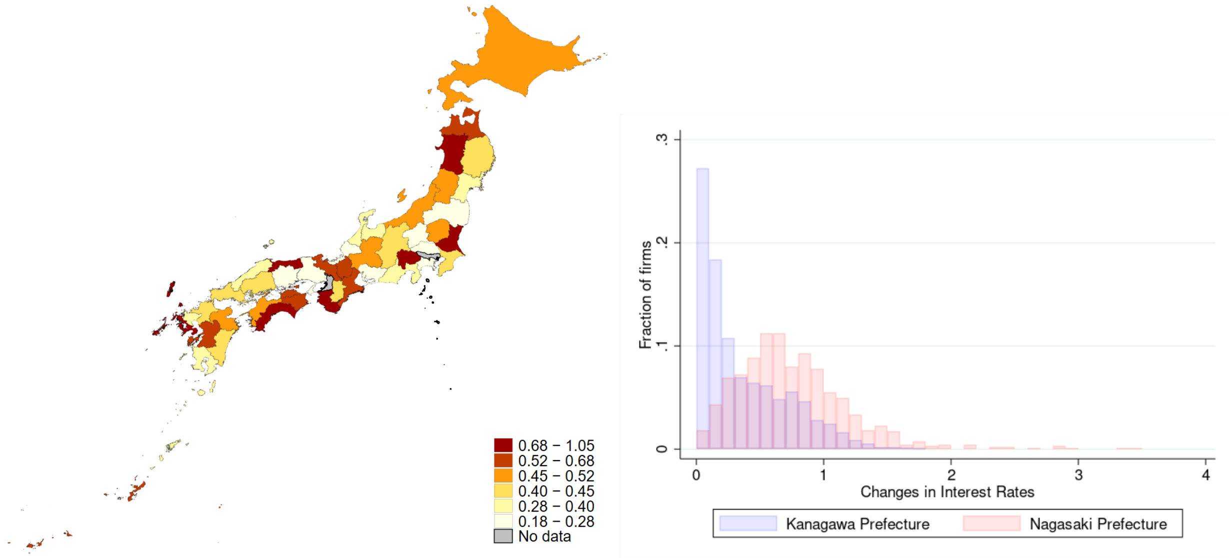
costs than firms in Nagasaki Prefecture, where alternative banking options are relatively limited.

How can policymakers infer the heterogeneous effects of bank consolidation across regions? Although the number of firms whose lending branches or second-efficient branches are closed can predict these effects, such information is not readily available to policymakers. Instead, we propose using the market shares of the first- and second-largest banks in each prefecture as practical predictors of the regional impact of bank consolidation. To examine this hypothesis, we run a prefecture-level regression of average changes in borrowing costs on the market shares of the first- and second-largest banks. The estimation results are as follows:

$$\Delta r_p = \beta_0 + \underset{[0.15]}{0.55} \times Share_{1,p} + \underset{[0.15]}{1.03} \times Share_{2,p} \qquad adj.R^2 = .65$$

where $\Delta r_p$ is the average change in borrowing costs in prefecture $p$, and $Share_{1,p}$ and $Share_{2,p}$ denote the market shares of the first- and second-largest banks, respectively, measured by the number of borrowers in prefecture $p$. Numbers in brackets report standard errors. The results indicate that higher market shares of either the first- or second-largest bank are associated with larger increases in borrowing costs following bank consolidation. This finding suggests that policymakers can use these readily observable concentration measures to assess and predict the regional impact of branch closures.

Figure 7: Changes in Borrowing Costs by Region




Note: The left panel shows the average increase in borrowing costs by region in our counterfactual simulation. The right panel focuses on two representative prefectures that experience large and small impacts in the counterfactual analysis, Kanagawa and Nagasaki, and presents the histogram of $\Delta r_i$.

Why do the first- and second-largest banks' shares predict the effects of branch closures across regions? First, a higher second-largest bank's share, $Share_{2,p}$, leads to a larger impact simply because we assume that the second-largest bank closes all of its branches in our counterfactual analysis. That is, when the second-largest bank's share is high, its branches are more likely to be the lending branch or the second-efficient branch for firms, thereby generating larger increases in borrowing costs. Second, a higher first-largest bank's share, $Share_{1,p}$, also predicts a larger impact because it proxies the availability of nearby alternative branches. When the first-largest bank's share is high, branch closures by the second-largest bank make the regional banking market effectively monopolistic, thereby leading to larger increases in borrowing costs.

In sum, our counterfactual analysis indicates that if bank consolidation entails large-scale bank branch closures, it may substantially increase firms' borrowing costs. The analysis also shows that the impact on borrowing costs is highly heterogeneous across firms and regions, reflecting differences in the availability of nearby alternative banks. Accordingly, the share of the largest bank, in addition to the share of the bank conducting the consolidation, predicts the regional impact of branch closures. This finding suggests that policymakers can use these measures to predict, in advance, the effects of bank consolidation on regional borrowing costs.

# 6 Conclusion

In this paper, we examine how firms' borrowing costs are affected by geographic distance from bank branches. We develop a model where firms endogenously choose their lenders by considering loan rates determined à la Bertrand competition among banks, and estimate the structural parameters using a dataset that combines firm- and bank-level microdata from Japan. The estimation results indicate that distance affects loan pricing by influencing both banks' and borrowers' costs, and that a simple reduced-form approach may yield biased estimates due to firms' endogenous lender choice. Counterfactual simulations indicate that the impact of bank consolidation depends on regional competitive conditions.

# References


**Adams, Robert M., Kenneth P. Brevoort, and John C. Driscoll**, "Is lending distance really changing? Distance dynamics and loan composition in small business lending," *Journal of Banking & Finance*, None 2023, *156* (C), None.

**Agarwal, Sumit and Robert Hauswald**, "Distance and Private Information in Lending," *The Review of Financial Studies*, July 2010, *23* (7), 2757–2788.

**Allen, Jason, Robert Clark, and Jean-François Houde**, "Search Frictions and Market Power in Negotiated-Price Markets," *Journal of Political Economy*, 2019, *127* (4), 1550–1598.

**Berry, Steven, James Levinsohn, and Ariel Pakes**, "Automobile Prices in Market Equilibrium," *Econometrica*, July 1995, *63* (4), 841–890.

**Crawford, Gregory S., Nicola Pavanini, and Fabiano Schivardi**, "Asymmetric Information and Imperfect Competition in Lending Markets," *American Economic Review*, July 2018, *108* (7), 1659–1701.

**Degryse, Hans and Steven Ongena**, "Distance, Lending Relationships, and Competition," *Journal of Finance*, 2005, *60* (1), 231–266.

— **, Nancy Masschelein, and Janet Mitchell**, "Staying, Dropping, or Switching: The Impacts of Bank Mergers on Small Firms," *The Review of Financial Studies*, 2011, *24* (4), 1102–1140.

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

**Erel, Isil**, "The Effect of Bank Mergers on Loan Prices: Evidence from the United States," *The Review of Financial Studies*, None 2011, *24* (4), 1068–1101.

**Gopal, Manasa and Philipp Schnabl**, "The Rise of Finance Companies and FinTech Lenders in Small Business Lending," *The Review of Financial Studies*, None 2022, *35* (11), 4859–4901.

**Granja, João, Christian Leuz, and Raghuram G. Rajan**, "Going the Extra Mile: Distant Lending and Credit Cycles," *Journal of Finance*, April 2022, *77* (2), 1259–1324.

**Herpfer, Christoph, Aksel Mjøs, and Cornelius Schmidt**, "The Causal Impact of Distance on Bank Lending," *Management Science*, February 2023, *69* (2), 723–740.

**Hosono, Kaoru, Daisuke Miyakawa, Taisuke Uchino, Makoto Hazama, Arito Ono, Hirofumi Uchida, and Iichiro Uesugi**, "Natural Disasters, Damage to Banks, and Firm Investment," *International Economic Review*, 2016, *57* (4), 1335–1370.

**Huber, Kilian**, "Are Bigger Banks Better? Firm-Level Evidence from Germany," *Journal of Political Economy*, None 2021, *129* (7), 2023–2066.

**Keil, Jan and Steven Ongena**, "The demise of branch banking – Technology, consolidation, bank fragility," *Journal of Banking & Finance*, 2024, *158*, 107038.

**Legran, Daniel, Olli Tuomikoski, and Johanna Spitz**, "The changing landscape of bank offices in the euro area," 2024. Blog posts by European Central Bank on 16 February 2024.

**Maudos, Joaquin and Juan Fernandez de Guevara**, "The cost of market power in banking: Social welfare loss vs. cost inefficiency," *Journal of Banking & Finance*, July 2007, *31* (7), 2103–2125.

**Narayanan, Rajesh P., Dimuthu Ratnadiwakara, and Philip Strahan**, "The Decline of Bank Branching," May 2025.

**Nguyen, Hoai-Luu Q.**, "Are Credit Markets Still Local? Evidence from Bank Branch Closings," *American Economic Journal: Applied Economics*, January 2019, *11* (1), 1–32.

**Ono, Arito, Yukiko Saito, Koji Sakai, and Iichiro Uesugi**, "Does Geographical Proximity Matter in Small Business Lending? Evidence from Changes in Main Bank Relationships," *Asia-Pacific Journal of Financial Studies*, 2023, (52), 819–855.

**Pavanini, Nicola, Vasso Ioannidou, and Yushi Peng**, "Collateral and Asymmetric Information in Lending Markets," *Journal of Financial Economics*, 2022, *144* (1), 93–121.

**Petersen, Mitchell A. and Raghuram G. Rajan**, "Does Distance Still Matter? The Information Revolution in Small Business Lending," *Journal of Finance*, December 2002, *57* (6), 2533–2570.

**Rice, Tara and Philip E. Strahan**, "Does Credit Competition Affect Small-Firm Finance?," *Journal of Finance*, June 2010, *65* (3), 861–889.

**Sapienza, Paola**, "The Effects of Banking Mergers on Loan Contracts," *Journal of Finance*, February 2002, *57* (1), 329–367.

**Wang, Yifei, Toni M. Whited, Yufeng Wu, and Kairong Xiao**, "Bank Market Power and Monetary Policy Transmission: Evidence from a Structural Estimation," *Journal of Finance*, August 2022, *77* (4), 2093–2141.

**Yasuda, Ayako**, "Do Bank Relationships Affect the Firm's Underwriter Choice in the Corporate-Bond Underwriting Market?," *Journal of Finance*, June 2005, *60* (3), 1259–1292.

# Appendix A: Data Sources

As stated in the main text, this study primarily utilizes two datasets: the Teikoku Databank database and Kinyu Meikan (Financial Directory). This appendix provides a more detailed description of these datasets, including the construction of the variables used for the estimation.

## A.1: Firm-level Data

Our firm-level data are provided by Teikoku Databank Ltd., a leading business credit bureau in Japan. We use several datasets from Teikoku Databank, including the corporate financial database, the corporate profile database, and the corporate credit research database.

First, the corporate profile database provides firm-level information, including firms' addresses. Because addresses are recorded in terms of latitude and longitude, we use this information to calculate the geographic distance between firms and their lenders, as well as the distance to nearby bank branches. Second, the corporate credit research database reports the branch code of the bank branch with which a firm primarily transacts. When a firm transacts with multiple bank branches, the branches are listed in descending order of loan volume. Following the methodology of Hosono et al. (2016), we designate the top-listed bank branch as the firm's lender in our analysis. Finally, the corporate financial database contains firms' financial information, including interest payments, profitability, and asset and loan sizes. Using these data, we construct the firm-level variables used for estimation as follows. We first define firm $i$'s loan size as the sum of long- and short-term loans:

$$\text{LoanSize}_{it} = \text{Long-Term Loan}_{it} + \text{Short-Term Loan}_{it}.$$

We then compute the interest rate and the long-term loan ratio by dividing interest payments and long-term loans, respectively, by outstanding loans:

$$\begin{aligned} \text{InterestRate}_{it} &= \frac{\text{Interest Payment}_{it}}{\text{LoanSize}_{i,t-1}}, \\ \text{LongTermRatio}_{t} &= \frac{\text{Long-Term Loan}_{t}}{\text{LoanSize}_{t}}. \end{aligned}$$

Next, we define firm size as the logarithm of total assets: $\text{FirmSize}_{it} = \log(\text{Total Assets}_{it})$. Finally, we define return on assets (ROA), liquidity ratio, and capital ratio as the ratios of operating profits, liquid assets, and net assets to total assets, respectively.

## A.2: Bank-level Data

Financial Directory (Kinyu Meikan) is an annual dataset that comprehensively lists branches of financial institutions in Japan. First, the dataset contains address information for all bank branches in Japan. Since we need to convert these addresses into latitude and longitude to calculate geographic distances between firms and bank branches, we use the "CSV Geocoding Service" provided by the University of Tokyo.[12] Second, the dataset contains information on bank categories. Hereafter, we refer to these bank categories as "bank type." As explained in the main text, we use this information to control for the possibility that firms prefer a particular bank type over others. Third, the dataset includes bank and branch codes for all bank branches in Japan, which we use to match address and bank-type information for bank branches with the firm-level data.

In our analysis, the presence of rival bank branches in the vicinity plays an important role in loan pricing. To capture the effects of rival banks more accurately, we make the following adjustments. First, we exclude branch codes from our sample if they are categorized as "branches-in-branch." In Japan, for logistical reasons, some banks retain branch codes when closing branches by relocating their addresses to an existing branch. If such de facto closed branches were counted as nearby rival branches, this would result in double-counting. Hence, we exclude all branches that belong to the same bank and share the same address. Second, we exclude branches that do not engage in corporate lending. Specifically, we remove branches from the set of potential rival branches if they do not serve as a lender for any firm in any period in our sample.

## A.3: Combining the Two Databases

These two databases are merged using the bank-branch code, as in Ono et al. (2023). Using the latitude and longitude for firms and bank branches, we calculate distances between firms and their lenders, as well as between firms and 30 nearby bank branches that could serve as potential rivals.

## A.4: Sample Selection

Because our focus is on bank accessibility for small businesses in regional areas, our analysis examines only firms located in regional areas that borrow from regional banks. First, we exclude firms located in Tokyo or Osaka. Since bank access is substantially better in these very large cities, we consider it difficult to analyze firms in these areas alongside those in other regions. Second, we exclude firms that borrow from the four largest banking groups—namely, Mizuho, MUFG, SMBC, and Resona—and focus instead on

[12] At `https://geocode.csis.u-tokyo.ac.jp/geocode-cgi/geocode.cgi?action=start`, this service is available for free.

firms that borrow from tier-one regional banks, tier-two regional banks, or credit associations.[13] We also exclude branches of the four large banking groups from the list of nearby banks, as they are unlikely to serve as realistic alternatives for small businesses in regional areas.

In addition to the above sample selection based on region and bank type, we exclude firms as outliers if at least one of the following conditions is satisfied: (i) the interest rate is below zero or above the 99th percentile; (ii) the ROA is below the 1 percentile or above the 99 percentile; (iii) the capital ratio is below zero or above the 99 percentile; (iv) the distance to the lender is above the 99 percentile; and (v) the distance to the nearest bank branch is above the 99 percentile. In addition, in the regression analyses using the average distance to the three or five nearest banks, we exclude firms if these variables exceed the 99 percentile.

Finally, we set the sample period from 2009 to 2020, as the Japanese government introduced policy support for commercial banks at the onset of the COVID-19 pandemic. Because firms issue their financial statements in different months, we construct an annual panel dataset by labeling, for example, data from January 2009 to December 2009 as 2009 data.

# Appendix B: Monte Carlo Simulation

This appendix provides details of the Monte Carlo simulation in Section 5.2 and shows that our likelihood estimation using simulated data successfully identifies the parameter values. In the Monte Carlo simulation, we assume for simplicity that firms are characterized only by distance to bank branches and firm size. Specifically, we parameterize $v_{ij}$, $c_i$, and $\xi_{ij}$ as:

$$v_{ij} \equiv \gamma_D d_{ij}, \;\; c_i \equiv \beta_S \log(\text{Asset}_i), \;\; \text{and} \;\; \xi_{ij} \equiv \theta_D d_{ij}$$

Thus, $\gamma_D$, $\beta_S$, and $\theta_D$ are the only parameters in the simplified model. We set the parameter values to $\gamma_D = -0.1$, $\beta_S = -0.1$, and $\theta_D = 0.1$ and examine whether the likelihood estimation recovers these true values.

The Monte Carlo simulation proceeds as follows. First, we generate 10,000 artificial firms, specifying their geographic characteristics and firm size. For the distance to bank branches, we assume that each firm has fifteen nearby branches, and the distance to each branch is drawn from a log-normal distribution. The mean and standard deviation of the log-normal distribution are set to 0.99 and 0.48 to match our sample moments. For firm size, we draw each firm's size from a log-normal distribution with mean and stan-

[13]We also exclude firms that borrow from credit cooperatives or government-affiliated financial institutions, as these firms tend to be too small—primarily individual businesses—to be comparable with other firms.

Table 9: Estimated Coefficients Using Simulated Data

| $\gamma_D$ | $\beta_S$ | $\theta_D$ |
|---|---|---|
| -0.100 | -0.100 | 0.100 |
| [-0.107, -0.092] | [-0.103, -0.096] | [0.092, 0.107] |

Note: The table reports the results for the maximum likelihood estimation using artificial data generated by Monte Carlo simulations. Columns 1, 2, and 3 show the estimation results for $\gamma_D$, $\beta_S$, and $\theta_D$, respectively. The table reports the medians of the estimated coefficients from 100 simulations, with the 5th and 95th percentiles shown in brackets.

dard deviation set to -1.26 and 1.61, respectively, again to match our sample. As the unobservable component of lending costs, we generate random variables following a Type I extreme value distribution, TIEV$(\xi_{ij}, \sigma_\omega)$, where $\sigma_\omega = 1.0$, and construct the lending cost $\underline{r}_{ij}$ based on equation (1) in the main text.

Second, given the equilibrium conditions of the model, the artificial firms optimally choose their lenders, and the interest rate is determined through an English auction à la Bertrand competition among bank branches. By sorting the values of $v_{ij} - \underline{r}_{ij}$ for firm $i$ and bank branch $j$, we identify the lender as well as the second-efficient rival branch for firm $i$. The equilibrium interest rate for firm $i$ is then determined as $r_i^* = v_{i(1)} - v_{i(2)} + \underline{r}_{i(2)}$ where variables with (1) and (2) denote those for the lender and the rival branch, respectively.

Third, we estimate the parameter values by maximizing the likelihood function. Specifically, we construct the likelihood function based on equation (4) in the main text using the artificially generated data on firm $i$'s lender choice, lending rate, and distance to nearby banks. Note that we assume that econometricians cannot identify which branch is firm $i$'s second-efficient bank, as in the estimation using real data.

Table 9 presents the estimation results for $\gamma_D$, $\beta_S$, and $\theta_D$. The table reports the medians of the parameter estimates from 100 simulations, with the 5th and 95th percentiles shown in brackets. The results indicate that the maximum likelihood estimation using the simulated data correctly identifies the parameter values. Specifically, although our likelihood function depends only on observable variables for firm $i$, the lender $m$, and the nearby branches in $\mathcal{J}_i$, it successfully recovers the calibrated parameter values used in the Monte Carlo simulation, namely $\gamma_D = -0.1$, $\beta_S = -0.1$, and $\theta_D = 0.1$. In particular, consistent with the argument in Section 3.3, the maximum likelihood estimation is able to separately and precisely identify the parameters governing the effects of distance on borrowers' costs, $\gamma_D$, and on lenders' costs, $\theta_D$.