EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF WORKING PAPER SERIES

EIEF Working Paper 25/02

February 2025

Does Distance *Still* Matter? Evidence from 30 Years of Mortgage Lending

By

Menaka V. Hampole
(Yale School of Management)

Adam Jørring
(University of Massachusetts Amherst)

Joao Monteiro
(Einaudi Institute for Economics and Finance)

Does Distance *Still* Matter? Evidence from 30 Years of Mortgage Lending*

Menaka V. Hampole Adam Jørring Joao Monteiro

February 4, 2025

Abstract

Mortgage lending by banks is still predominantly a local business. Local branches have a significant advantage in attracting and originating mortgages. More than 25% of borrowers are less than two miles from their branch, and more than 50% are within 10 miles. Distance is associated with lower credit quality, and the additional risk is priced: higher distance is associated with a lower probability of approval, a higher interest rate spread, and a higher ex-post likelihood of delinquency. We document business cycle patterns as the distance rises in booms and falls in busts but find that the sensitivity of distance for loan approval and pricing has persisted over the past 30 years despite the dramatic rise in securitization and online banking. Using a quantitative spatial model of the mortgage market, we explore the economic mechanisms driving our results. Through the lens of the model, we find that decreasing search costs lowers the match quality of borrower search, which counteracts the effects of increasing screening technology. Our findings highlight that local bank branches are still highly relevant despite the "fintech" revolution.

Keywords: Mortgages, Distance, Financial Intermediation

JEL Classification Codes: G21, G28

*We thank Effi Benmelech, Luigi Guiso, and Mounu Prem for helpful comments and discussions. Hampole is at Yale School of Management. Email: menaka.hampole@yale.edu. Jørring is at University of Massachusetts Amherst. Email: ajorring@umass.edu. Monteiro is at Einaudi Institute for Economics and Finance. Email: joao.monteiro@eief.it.

1 Introduction

What role do local bank branches play in mortgage lending? Traditionally, the theoretical banking literature has emphasized that branches give banks a local advantage by facilitating screening, monitoring, and the collection of soft information (Diamond, 1984, 1991; Stein, 2002). However, a growing empirical literature studying small business loans, beginning with the seminal work by Petersen and Rajan (2002), finds that, with increasing use of information technology, banks rely less on soft information and increase lending to borrowers farther away from their branches.[1] We might expect to observe a similar trend in the geographic scope of mortgage lending—away from local markets toward national markets—especially as mortgage lending has been impacted not only by the IT revolution but also by the rise of securitization, which further diminishes banks' incentive to screen, monitor, and collect soft information (Keys et al., 2010).

The geographic scope of the U.S. mortgage market is the topic of an ongoing debate in the literature with implications for both antitrust regulation and macroeconomic stability.[2] If, on the one hand, mortgage lending is local, then antitrust regulators should be concerned if mergers increase local concentration (as opposed to the current regulatory regime where only changes in the national level of concentration can trigger antitrust enforcement). Furthermore, if lending is local, then changing interest rates and deposit shocks could impact local credit supply through the deposit channel as outlined in Drechsler et al. (2017) and Granja et al. (2022). On the other hand, if the mortgage market is entirely national, these concerns are moot, financial stability is unaffected, and the current antitrust regulation is sufficient.

In this paper, we document three facts about the geographic scope of mortgage markets. First, we find that distance does indeed matter in mortgage lending, even in the modern economy today. In line with the predictions from the traditional banking literature mentioned above, we find that local lenders have a significant advantage in terms of attracting, approving, and originating mortgages from local borrowers. Consequently, the majority of mortgages are originated to borrowers who are less than 16 kilometers (10 miles) from their branch. Our second fact is that the distance between borrower and

[1] The astute reader will notice that our title is a tribute to the seminal work by Petersen and Rajan (2002).

[2] Prior to 2008, The Federal Reserve considered mortgage markets to be geographically local. However, with the approval of Bank of America's acquisition of Countrywide, the Fed declared the mortgage market to be "national in scope" (Federal Reserve System, 2008), and, consequently, local mortgage concentration does not impact evaluations of bank mergers. This view is based on the finding that there is no correlation between local concentration and interest rates (Amel et al., 2018). However, recent work by Buchak and Jørring (2024) challenges this view and argues that local lenders do have market power based on the finding that higher local concentration leads to higher upfront fees.

lender varies pro-cyclically over the business cycle, rising in booms and falling during the busts. This finding is consistent with the idea that credit quality deteriorates as soft information is harder to collect when the distance increases. Our third fact is that, across many dimensions, the role of distance is the same today as it was 30 years ago. For example, we find that the sensitivity between distance and whether a loan application is approved has not changed structurally over the 30-year period. This third fact might be the most surprising of the three, given the dramatic time-series changes in IT use and securitization described above.

We begin our paper by assembling a dataset covering the near universe of U.S. mortgage applications from 1994 to 2023. As the focus of our paper is the role of bank branches, in our baseline analysis, we study applications made to banks. (When studying the economic mechanism and in additional robustness tests, we include applications made to non-bank mortgage lenders.) Our main variable of interest is the geographical distance between the borrower and all the possible banks to which she could submit a mortgage application. Specifically, we calculate the distance between the borrower and every single branch in a given year. This is a massive computational exercise, and our paper is, to the best of our knowledge, the first paper to calculate these distances for the U.S. mortgage market.[3]

Using this dataset, we document a novel stylized fact that motivates our analysis. We show that most bank mortgage lending is to borrowers who live close to one of the bank's branches. More than a quarter of borrowers live within two miles of their closest branch, and more than half live within 10 miles. Surprisingly, this pattern holds true even today in the modern economy. In fact, in 2023 (the last year in our sample), the median distance between a bank and its borrowers was lower than in the late-1990s and 2000s. In other words, the median geographical distance between a bank and the mortgage borrower is lower today than it was *before* the rise of both securitization and online banking.

Having established our motivating fact, we turn to a formal analysis of the role of local bank branches. Our paper proceeds in three steps: First, we explore the role of branches in a cross-sectional analysis. Then, we ask whether the role of branches has changed over time in a time series analysis. Third, we use heterogeneity across borrower and lender characteristics to explore the economic mechanisms driving our results.

In the first part of the analysis, we ask whether local branches give banks an advan-

[3]For example, in 2023, there are 4.9 million mortgage applications and 70 thousand branches yielding 347 billion borrower-branch pairs. It takes about nine days to calculate all the distances when running the code on a Windows Datacenter server with 1TB ram and two AMD EPYC 9334 processors when running the code in parallel across 128 cores. The closest related paper that we know of is van der Plaat (2020) who calculates the distance between a borrower's county and her bank's county.

tage in their mortgage lending business. Specifically, we go through each step in the "mortgage production function" and ask whether having a local branch gives the bank an advantage at this step. The first step is attracting the borrower, i.e., "drumming up demand." Here, we find a significant local advantage. For a given mortgage application, the average bank is over 200 times more likely to receive this application if the bank has a branch within 5 km of the applicant (relative to the bank's average market share vis-a-vis its competitors). Consequently, the average bank has a 5.75 percentage-point higher market share among loans from borrowers who live within a 5 km radius of one of its branches (relative to the baseline). This result shows that local branches play a key role in creating local demand for their products.

Why is borrower demand decreasing in distance? While the previous literature has typically emphasized search costs (e.g., Agarwal et al. 2024), we find an additional channel that affects demand: loan approval varies with distance. Conditional on receiving an application, if the borrower is within 5 km of the bank's nearest branch, the bank is 3.2 percentage points more likely to approve the mortgage application. Interestingly, this result is not simply driven by the location of the borrowers. One might expect, for example, that it is harder to approve applications from rural borrowers if they have less available paperwork or documentation of income, and simultaneously, the closest bank branch is further away for rural applications. However, the result holds even when comparing two borrowers with the exact same hard information. That is, when we compare two borrowers who live in the same location and have the same observed credit risk, the applicant who applies to a local bank has a higher probability of having their application approved. Rather than being driven by differences in borrower locations, the result suggests that local loan officers possess valuable insights about the local market or have an easier time collecting the necessary information to approve the application. One such example could be local differences in variation in assessment values, which could make regulatory constraints more likely to bind, as suggested in recent work by Jiang and Zhang (2022).

Once the branch has accepted the application, the third and final step before the bank can originate the mortgage is for the applicant to accept the bank's offer. Also, here, having a local branch proves advantageous for banks. Conditional on having their application approved, borrowers who live within 5 km from the branch are 3.5 percentage points more likely to accept the offer.

Next, we ask whether distance correlates with the value of the originated mortgage. On average, we find that banks charge lower prices (lower interest rate spreads and lower fees) to local borrowers, consistent with the higher demand from local borrowers described above. Local borrowers are also less likely to become delinquent on their mortgages. Interestingly, these results hold after controlling for all available hard information on credit risk. Specifically, we control for each LLPA grid point, which is the grid of FICO and LTV that the GSEs use to price risk. Combined, these two results indicate that banks internalize the added risk from lending at a distance and price the added risk.

Having established that branches provide a local advantage, in the second step of our analysis, we ask whether this local advantage has changed over time. For each of the steps in the mortgage production function, we estimate the yearly sensitivity of distance. We begin by asking whether the local advantage in attracting mortgage applications has disappeared over time. The time-series evidence reveals three distinct periods in the data: First, from 1994 to 2000, the local advantage falls, consistent with the increase in cross-border banking facilitated by the The Riegle–Neal Interstate Banking and Branching Efficiency Act of 1994 (IBBEA).[4] Then, from 2000 until 2007, the local advantage rose steadily until it dropped in 2008 with the onset of the great financial crisis. Finally, it has remained stable from 2008 until 2023.

Next, we examine the correlation between distance and mortgage approval. This relationship is remarkably stable over time. Across the sample, a one-percent increase in distance is associated with a 50 bps decrease in the approval probability. In other words, the sensitivity of distance to mortgage approval has not changed over time. The correlation between distance and borrower acceptance, on the other hand, has declined over time (in absolute terms). That is, the benefit of having a local branch in terms of getting the borrower to accept a mortgage offer has declined over time.

The sensitivity of mortgage pricing with respect to distance is positive in all but two years and is positive and statistically significant in most years. Interestingly, we do not find any time trend across the 30 years of data. That is, the degree to which distance is priced has not changed over the sample. Since 2018, the HMDA data includes fees that lenders charge upfront, for example origination charges and discount points. Using the definition of lender fees from Buchak and Jørring (2024), we find that in each year from 2018 to 2023, lenders charge higher upfront fees as a percent of the home value when the borrower is further away from the lender.

Finally, we explore the economic mechanisms driving our results. Overall, we find that distance is priced, even conditional on all other available hard information, highlighting that distance contains information about the borrowers' credit quality. To shed light on this result, we explore heterogeneity across borrower types. Specifically, we compare the correlation between distance and the steps of the mortgage production process

[4] IBBEA, which was signed into law in 1994, allowed interstate banking (effective in 1995) and interstate branching (effective in 1997) (Johnson and Rice, 2008; Rice and Strahan, 2010).

for different groups of borrowers. We find that low-income or younger borrowers are far more likely to accept a mortgage contract with a nearby lender than a mortgage contract with a faraway lender. The terms of the mortgage contract explain the higher sensitivity of low-income and younger borrowers. Low-income or younger borrowers face larger increases in interest rate spreads and origination fees than high-income or older borrowers for the same increase in distance. The cross-sectional differences in the role of distance also translate into differences in the time series. The correlation between mortgage approval and distance varies similarly with the business cycle for different groups of borrowers. In contrast, only low-income borrowers faced higher interest spreads or larger origination fees after the Great Recession due to an increase in distance.

What economic mechanism can explain the evidence from the cross-section of mortgages and the time series? We construct a spatial model of the mortgage market to answer this question. The model has a finite number of identical locations, each with a continuum of borrowers. There are also lenders in space willing to provide borrowers with funds to purchase a house. However, borrowers may default with some exogenous probability. The probability of default is their type and is private information. Therefore, our model is a model of adverse selection.

The model is static and has three subperiods. In the morning, borrowers learn about their type and must decide which lender they want to obtain a mortgage from. Borrowers are unsophisticated. Borrowers believe all lenders will offer them the same interest rate. They face search costs that increase with distance. Consequently, borrowers want to choose the closest bank, subject to exogenous taste shocks. At the end of the morning, the model produces market shares for each lender in each location.

In the afternoon, lenders receive mortgage applications. For each application, the lender knows the borrower's location and receives a signal about the borrower's type. The signal's precision decreases with distance, so it can be thought of as hard information, which, as distance increases, represents a smaller share of total information. When distance is zero, the lender has full information about the borrower's type. The lender uses the signal and the distance between the borrower and the lender to compute the optimal interest rate by solving a standard screening problem. The interest rate decreases with the signal and increases with distance. Moreover, there are gains from trade, as borrowers value the house more than the funds required to acquire it. Hence, the lender can extract surplus from the borrower to compensate for the risk.

In the evening, borrowers receive a mortgage offer. They understand that their morning beliefs were wrong and accept the offer if it yields a utility larger than the reservation utility. However, some borrowers in some locations may reject the mortgage offer, and not all mortgage applications produce mortgages. For example, a borrower with a high type (low probability of default) who applied to a lender far away may receive an interest rate that is too high. Therefore, the mortgage will not be originated.

The model generates three objects. First, it produces market shares, which are decreasing functions of distance. Second, it produces interest rates, which increase with distance. Third, it produces mortgage origination probabilities, which decline with distance. Therefore, the model can explain all the facts we document for the cross-section of US mortgages.

We then turn to the time series. Between 1994 and 2023, we assume the productivity of the screening technology increases. Therefore, the precision of the signal the lender receives increases for a given distance. This assumption is consistent with the hardening of information documented in Liberti and Petersen (2019). The same period is also characterized by decreased search costs (e.g., as borrowers search for lenders using the Internet).

As the precision of the signal increases, lenders offer lower interest rates to borrowers. Consequently, the probability a mortgage is originated increases even more large distances. Therefore, the average distance between a borrower and a lender increases. Moreover, distance becomes less valuable to the lender, and hence, the elasticity of the interest rate to distance decreases. This prediction aligns with most of the literature, like Petersen and Rajan (2002), but is inconsistent with our empirical findings. However, the model can resolve this puzzle. If search costs decrease, lenders receive more applications from borrowers that are far away. Consequently, the average interest rate they charge increases, which then leads to a decline in the probability a mortgage is originated. Therefore, even though the average distance in all mortgage applications increases, the average distance in all mortgages decreases. Why does this take place? The intuition is simple – as search costs decrease, the probability the lender (she) meets a borrower she believes will become delinquent is higher. Moreover, distance is still priced in mortgage contracts, even with lower search costs. Hence, the model can reconcile higher screening productivity with a non-increasing average distance as long as search costs decrease.

Combined, our findings highlight that the impact of the so-called "fintech revolution" on the role of bank branches may not be as dramatic as previously argued. Instead, we argue that the bank branch is well and alive and still highly relevant for mortgage lending, even in the modern economy. To paraphrase Mark Twain, we find that the rumors of the death of the bank branch are highly exaggerated.

Our paper's main contribution is to provide empirical evidence on the role of local bank branches in mortgage lending. We build on a longstanding theoretical literature in banking that argues that bank branches provide value by facilitating screening, monitoring, and the collection of soft information (Diamond, 1984, 1991; Stein, 2002). We contribute to this literature by testing the theoretical predictions in the context of the largest consumer finance market in the world: the U.S. residential mortgage market.[5]

Beginning with the seminal work by Petersen and Rajan (2002) (henceforth PR) who studies data from the National Survey of Small Business Finance (SSBF) from 1973 to 1993, a longstanding literature has documented that banks are more likely to lend to small businesses that the banks are geographically close to. The same relationship has been documented in the context of small business loans in Belgium (Degryse and Ongena, 2005), SME lending from a large lender in the U.S. (Agarwal and Hauswald, 2010), and corporate lending in Norway (Herpfer et al., 2023). A key insight from PR is that banks will increase lending to businesses farther away as increased use of IT will lower the reliance on soft information, and our paper contributes to this literature by documenting that this pattern has not materialized in the mortgage market, despite the increased use of IT in the mortgage market (Berger, 2003; Jiang et al., 2023).[6]

In terms of pricing and credit risk, early studies, including PR, Degryse and Ongena (2005) and Agarwal and Hauswald (2010) find that loans made a higher distance carried lower interest rates, consistent with the idea that, in equilibrium, lending at a distance was only done to businesses with lower credit risk. In contrast, we find that mortgages at higher distances have *higher* credit risk and *higher* interest rates and fees. On this dimension, our results are closer to the more recent evidence from Granja et al. (2022), who studies small business lending in the boom and bust period around the great financial crisis and find that lending at a distance is associated with higher credit risk, although they find that banks did not charge higher interest rates for distant loans.

Our paper is most closely related to a recent debate on the geographic scope of U.S. mortgage markets and the role of local branches in mortgage lending. Historically, the Federal Reserve considered U.S. mortgage markets to be local. However, in 2008, with the approval of Bank of America's acquisition of Countrywide, the Fed declared the mortgage market to be "national in scope" (Federal Reserve System, 2008). This view is based

[5]Besides screening and monitoring, banks also acquire information to soften competition and expand market shares, as shown in Hauswald and Marquez (2006). In the context of mortgage lending, Loutskina and Strahan (2011) find that concentrated lenders invest more in information collection (than diversified lenders), which allows them to better price risk. For a recent survey on the role of information in lending, see Liberti and Petersen (2018).

[6]Relatedly, in an analysis of recent SSBF data from 1993 to 2003, Brevoort and Wolken (2008) finds that distance increased in the first half of the decade but decreased in the second half and that the median distance still remains very low. Similarly, Adams et al. (2023) shows that the increase in the average distance for small business lending is driven by a few specialized lenders who lend nationwide. Besides these small lenders, they find that small business lending still remains dependent on local banks.

on the finding that there is no correlation between local concentration and interest rates (Amel et al., 2018). The opposing view—that mortgage markets are still local—is based on the results that the closing of local branches curtails local lending (Nguyen, 2019) and that local lenders have market power and charge higher all-in costs (Buchak and Jørring, 2024). Furthermore, the lack of geographical variation in interest rates might be due to political pressure, as argued in Hurst et al. (2016).[7]

The geographic scope of the U.S. mortgage market has implications for both antitrust regulation and macroeconomic stability. Under the current regulatory regime, bank mergers are evaluated on whether they increase local concentration in deposit markets but not whether they increase local concentration in mortgage markets (Federal Reserve System, 2021). However, if mortgage lenders do have market power and influence local lending, then changing interest rates and deposit shocks could impact local credit supply (as shown in the context of small business lending in Granja et al. 2022), house prices (Favara and Giannetti, 2017), and the transmission of monetary policy (Scharfstein and Sunderam, 2016; Drechsler et al., 2017).

Finally, at a broader level, our results on the role of geographical distance between lenders and borrowers connect to the papers studying this relationship in bank branch networks (Ho and Ishii, 2011; Koont, 2023), syndicated cross-border lending (Giannetti and Laeven, 2012; De Haas and Van Horen, 2013; Kleimeier et al., 2013; Cerutti et al., 2015), finance and development (Sussman and Zeira, 1995; Jayaratne and Strahan, 1996), and trade (Helpman et al., 2008; Head and Mayer, 2014).

The rest of the paper is organized as follows. Section 2 describes our data sources and the algorithm we use to compute distances and provides the motivating stylized facts. Section 3 presents our results for the cross-section of mortgage applications, and Section 4 presents our results about the time-varying role of distance. We present a model that rationalizes the cross-sectional and time-series evidence in Section 5. Section 6 concludes.

2 Data

We combine several standard data sets used in the household finance and banking literature. In this Section, we describe these data, how we combine them, and present some summary statistics.

[7] Relatedly, Granja and Paixão (2019) find that banks price deposits uniformly across branches.

2.1 Data Sources

HMDA. Our primary data source is the mortgage-level application and acceptance data collected under the Home Mortgage Disclosure Act (HMDA), which covers the near universe of US mortgage applications. The HMDA data, extensively used in the literature, includes lender identification, the application's outcome, loan type, purpose, size, year of origination, and location at the census tract level. It also contains limited demographic information on applicants, notably race and income.[8] Since 2018, HMDA has recorded several further variables that we will use in our analysis: loan interest rate, non-interest rate charges (including origination charges, discount points, and lender credits), loan-to-value (LTV) ratios, and debt-to-income (DTI) ratios.

Fannie Mae and Freddie Mac single-family loan origination and performance data. Fannie Mae and Freddie Mac provide information on the GSEs' portfolios of 30-year single-family conforming fixed-rate mortgages. As with the HMDA sample, we restrict the sample to 30-year mortgages originated for purchases of owner-occupied homes. These loans are fully amortizing and have full documentation. The loan-level origination dataset provides interest rates, FICO scores, LTVs, and DTIs, as well as loan size, type, purpose, and location. It also identifies the originator that sold the loan to the GSE in cases where the originator had a sufficiently high origination market share in the reporting period.

In part of our analysis, we merge the HMDA dataset with the GSE dataset. For this merged sample, we observe the credit score at origination and track loan performance over time (to gather information on delinquencies). We match the HMDA and GSE datasets using the following conservative procedure, as in Buchak and Jørring (2024). First, we restrict the GSE loans to those matching the criteria listed above for the HMDA loans (i.e., loans for the purchase of owner-occupied homes, etc.). We then match loans on location (state, metropolitan statistical area, and ZIP Code), the exact loan amount and interest rate, and the purchaser type (i.e., Fannie Mae or Freddie Mac). To ensure the highest-quality match, we exclude all loans with duplicate observations and match without replacement.

[8] To study comparable mortgages, we follow the literature and restrict our baseline sample to 30-year conventional, first-lien mortgages originated for purchases of owner-occupied single-family homes. We exclude restricted government-insured loans, such as FHA, VA, FSA, and RHS loans. We also exclude mortgages with "exotic" features, such as reverse mortgages, an open-end line of credit (e.g., HELOCs), interest-only mortgages, and mortgages with prepayment penalties, intro-rate periods, balloon payments, or other non-amortizing features. In focusing on conforming loans, which are eligible to receive credit guarantees from the GSEs (Fannie Mae and Freddie Mac), our analysis largely avoids the issue of unobserved heterogeneity in credit quality driving pricing differences.

Summary of Deposits. We use the Summary of Deposits (SOD) dataset, compiled annually by the Federal Deposit Insurance Corporation (FDIC). The SOD provides comprehensive information on the deposit holdings of every bank branch in the United States that is insured by the FDIC. This data set does not have information on shadow banks not insured by the FDIC. This data set includes detailed branch-level data such as deposit amounts, geographic location (county, state, and ZIP code), institution type, and branch characteristics.

Supplemental data. We also obtain information on bank branches from the Housing and Urban Development Department (HUD). HUD provides information on all closed branches for all financial institutions that provide mortgages, even those not insured by the FDIC. These data also include information on the opening and closing dates of the branches. We use the Business Registry to obtain more data on bank branches. We also rely on the "Avery file" crosswalk from Neil Bhutta's website.[9]

2.2 Algorithm to Compute Distance

Our goal is to compute the distance between the mortgage applicant and the potential lender. To do this, we need the physical location of both the borrower and the lender. The HMDA data does not have information on the borrower's residence, so we use the location of the property the borrower wants to purchase. As we restrict our sample to owner-occupied single-family homes, we will not likely capture borrowers who buy multiple homes. Using the HMDA data merged with Census data, we use the centroid of the census tract as the location of the borrower.[10]

We use the SOD files to obtain the addresses of all bank branches for banks that are FDIC insured. We then use the USPTO's website to get each bank branch's coordinates (latitude and longitude). We then merge the HMDA application data with the SOD location information using the crosswalk developed by Neil Bhutta.

For each mortgage application, we may have several different bank branches for the same bank. To compute the distance between the property and each bank branch, we use

[9] The data is available here. We use the HMDA lender file, or "Avery" file, that contains matching information for all lenders who have ever filed a HMDA report.

[10] U.S. Census tracts are small, relatively permanent subdivisions of a county designed to provide a stable set of geographic units for statistical analysis. Census tracts are generally designed to include 4,000 residents, though they may range from 1,200 to 8,000 people. The variation accounts for differences in population density across urban, suburban, and rural areas. We also account for the fact that Census tracts vary from one Census to another. More information on Census tracts can be found here. We use the Gazetteer to compute the coordinates (latitude and longitude) of the centroid of the census tracts. The files can be found here.

the equirectangular approximation, which is similar to Pythagoras' theorem with an adjustment for the curvature of the Earth.[11] We then define distance as the smallest distance between the property and all bank branches of the bank in the mortgage application.

2.3 Stylized Facts

Our sample covers all mortgages for which we can compute the distance between the borrower and the lender. We observe almost 200 million mortgages between 2000 and 2023, ranging from 5.1 million in 2006 to 12 million in 2003. We also observe the lender's identity, and we have between 1,958 lenders in 2021 and 4,800 lenders in 2008.[12]

Figure 1 plots the median distance between borrowers and lenders for all mortgage applications between 1994 and 2023. Distance seems procyclical, as it increases in periods of expansion and falls during recessions. For example, between 1994 and 2001, median distance increased sharply, only to fall equally sharply during the 2001-2002 recession. The increasing path of distance until 2001 is consistent with the findings in PR, who report that between 1973 and 1993, the distance between borrowers and lenders increased. Therefore, distance seems to move with credit risk, which tends to rise during expansions.[13] With the exception of the spikes in the expansion periods, distance fluctuates around a constant value, under 20 km. Following the Great Recession, distance does not fluctuate and remains almost constant. As the cost of transmitting soft information declined after 2010, it is somewhat surprising that the period was not characterized by a decrease in distance.

The distribution of distance between borrowers and lenders is also stable over time.[14] In Figure 2, we decompose the number of mortgages per year in bins according to the distance between the borrower and the lender. The share of mortgages with a distance under 2.5 km (or 1.6 miles) is very high, ranging from around 28 percent in 1994 to 17 percent in 2023. Between 1994 and 2001, as median and average distance increased, the share

[11] Suppose we have two locations (x_1, y_1) and (x_2, y_2) where x_1 denotes latitude of point 1 and y_2 denotes longitude of point 2. The distance between point 1 and point 2 is given by $d = R \times \sqrt{\phi^2 + \lambda^2}$, where R is the radius of the Earth in km, $\phi = (y_2 - y_1) \times \cos((x_1 + x_2)/2)$ and $\lambda = (x_2 - x_1)$. For small distances, $\cos((x_1 + x_2)/2) = 1$ and this is Pythagoras' theorem.

[12] We present the number of mortgages and lenders for all years in Figures A.2 and A.3, respectively.

[13] In Figure A.1, we plot median distance along with the Federal Funds Rate. Both time series behave similarly, and the time series correlation between the two is 0.4.

[14] We plot the distribution of distance in Figure A.4. We show that the distribution is usually bimodal and is stable over time. We also compute the spatial distribution of distance, which we present in Figure A.5. We find that the dispersion of average distance across US counties is very stable over time. Therefore, our results are not driven by significant changes in spatially heterogeneous access to mortgage services. Moreover, we also find that the dispersion of average distance across lenders is stable over time, as shown in Figure A.6.

of mortgages under 2.5 km declined. However, following 2001, the share increased once more - in 2001, the share of mortgages with a distance under 2.5 km was 15 percent, two percentage points lower than the share in 2023. Therefore, as the banking sector adopts technologies that allow potential customers to interact with banks remotely, mortgages remain a proximity-based business.

Table I presents some summary statistics for all mortgages in our sample in 2023. The median distance between borrowers and lenders is 16 km, while the average distance is 385 km, which highlights that many mortgages have very high values for distance. The standard deviation of distance is around two times the mean. 72 percent of loans are accepted in 2023, and the average loan amount is around 320 thousand dollars. The average applicant has an annual income of 230 thousand dollars and wishes to buy a property worth 648 thousand dollars. The average interest rate is 6.8 percent, and the average interest rate spread is 0.53 percentage points.

3 Cross-Sectional Evidence

The banking literature has long recognized the importance of soft information (Petersen and Rajan, 1994; Stein, 2002; Brickley et al., 2003) and has argued that the geographical distance between the borrower and the lender reflects the lender's ability to obtain soft information about the borrower (Petersen and Rajan, 2002; Berger, 2003; Liberti and Mian, 2008). Therefore, distance is priced in mortgage markets as a proxy for soft information.

In this Section, we use cross-sectional variation to study how distance is correlated with the several steps in the production of a mortgage. The expected value of a mortgage, conditional on distance, to a lender can be written as

$$\begin{aligned}\mathbb{E}[\text{Value}|\text{Distance}] = \mathbb{P}\left(\text{Lender receives application}|\text{Distance}\right) \\ \times \mathbb{P}\left(\text{Lender approves application}|\text{Distance, Receipt}\right) \\ \times \mathbb{P}\left(\text{Borrower accepts}|\text{Distance, Receipt, Approval}\right) \\ \times \mathbb{E}[\text{Value}|\text{Distance, Receipt, Approval, Acceptance}]. \qquad (1)\end{aligned}$$

The value of a mortgage can be written as a product of four terms: (1) the probability a lender receives an application, conditional on distance, (2) the probability a lender approves the application, conditional on distance and receipt of an application, (3) the probability the borrower accepts the terms, conditional on distance, receipt of an application, and approval, and (4) the expected value, conditional on distance, receipt and approval of an application, and acceptance on the part of the borrower. In this Section, we estimate the first three terms.

The fourth term in equation (1) represents the expected value to the lender. The value depends on three terms: (1) the origination fees, (2) the interest rate spread, and (3) the probability the borrower becomes delinquent. Each of these three terms may also depend on distance, as we show in this Section.

As a motivating example, consider a borrower (he) who can apply to a lender (she) nearby or a far-away lender. The borrower has a given quality, which is private information. The lenders observe a signal about the borrower's quality, and the precision of the signal decreases with geographic distance. Suppose the borrower is of high quality, i.e., has a low probability of default. In this case, the borrower may wish to reveal his type and will apply for a mortgage with the lender who is closest to him. Now consider a low-quality borrower who wishes to hide his type. The lender closest to him has a good signal about his type, so the borrower will apply for a mortgage with a far-away lender to hide his type. However, the lender also understands the borrower's incentives, and upon observing a borrower who comes from afar, she will adjust her expectations about his quality and either reject his application or charge a higher interest rate. Therefore, the lender will price distance in the mortgage contract as the quality of the borrower determines the distance.

3.1 Market Share of Banks

We begin by studying whether distance plays a role in the production of mortgages. To examine this link, we compute each lender's share in mortgages. For every lender j, we draw circles of radii r around every branch. We then compute the share of mortgages generated inside the circle directed at lender j using all mortgages, all approved mortgages, and all rejected mortgages. We conduct this exercise for increasing values of r.[15] This produces a data set at the year-lender-distance level, which contains information on the lender's share of mortgages in a particular geographic area. We define the distance levels as $d = 1, \ldots, 11$, where $d = 1$ represents a radius of $r = 5$ km and $d = 11$ represents the full sample.

Using the data set we produced, we estimate the following equation

$$s_{jdt} = \lambda_{j,t} + \sum_{m \neq 11} \gamma_m \cdot \mathbf{1}\{d = m\} + \varepsilon_{jdt}, \tag{2}$$

[15] We draw circles of radii of 5, 10, 25, 50, 100, 150, 200, 500, 1,000, and 2,000 km. Note that this exercise means we expand the circles' radii, so all mortgages we use when $r = 5$ are also used when we set $r = 10$. Moreover, when we set $r = \infty$, we include all mortgage applications made that year.

where the outcome variable is the share of mortgages in year t in a distance d for lender j. We include lender-year fixed effects. We also include a series of distance fixed effects, using $d = 11$ (the largest possible radius) as the reference group. Therefore, we can interpret γ_m as the difference in the share of mortgages for a lender between a distance m and the full sample. For example, for $m = 1$, the coefficient γ_1 captures the "local" advantage of lenders. We cluster the standard errors at the lender lender.

Discussion of Assumptions. The main concern with estimating equation (2) is that lenders may have different strategies to attract and screen applicants. Some lenders may rely more on soft information than others, leading to a higher share for lower distances. Therefore, including the lender-time fixed effects is important to absorb differences across lenders.

Results. Figure 3 presents the results of estimating equation (2). The coefficients display a decreasing trend, which shows that the lender's share decreases with distance. For example, the coefficient associated with the lowest distance ($\leq$ 5 km) is 0.058, which implies that the lender's market share is 5.75 percentage points larger in the 5 km radius around its branches when compared with its market share for the whole US. The average unconditional market share for all lenders in 2000 is 0.04 percent, so our model predicts that the market share in the 5 km radius around a lender's branches is two orders of magnitude larger when compared to the market share in the whole country which reflects the local advantage of lenders. Moreover, the local advantage of lenders is not driven by specific lenders as that would be captured by the lender fixed effects.

We can also estimate equation (2) using market shares in approved or rejected mortgages. We present these results in Figure B.1. For each distance bin, the coefficients tend to be larger for approved mortgages than rejected ones. This is consistent with the hypothesis that low-quality borrowers prefer to search for a mortgage further away. As low-quality borrowers are more likely to observe a rejection in their mortgage application, this reduces the local lender's market share in rejected applications.

3.2 Production of Mortgages

We start by studying how distance shapes the production of mortgages. We focus on how the presence of soft information introduces a local advantage for lenders and study how it shapes the correlation between distance and mortgage approval. We define mortgage approval as an indicator variable that takes the value of 1 if the lender approves the mortgage and zero if otherwise. For all approved mortgages, we define mortgage acceptance as an indicator variable that takes the value of 1 if the borrower accepts the terms of the mortgage contract, and zero if otherwise.[16] Using our full sample of mortgage applications we estimate the regression

$$Y_{ijt} = \alpha_{c(i),t} + \mu_{s(i),j,t} + \beta X_{ijt} + + \sum_{m \neq 11} \gamma_m \cdot \mathbf{1}\{i \in m\} + \varepsilon_{ijt}, \tag{3}$$

where the outcome variable is either an indicator variable associated with the approval of mortgage application i made to lender j in year t or an indicator variable associated with the acceptance of the mortgage application. We include a county-year fixed effect $\alpha_{c(i),t}$ where $c(i)$ is the county in which the property is located, as well as a lender-state-year fixed effect $\mu_{s(i),j,t}$, where $s(i)$ denotes the state. We also include a vector of mortgage-level controls X_i, which includes the logarithm of the loan amount for $t \leq 2017$ and, for $t \geq 2018$, also includes the logarithm of the value of the property, the logarithm of the applicant's income, and a series of debt-to-income fixed effects. The coefficients of interest are the γ_m, which multiply an indicator variable that takes the value of 1 if the distance associated with the mortgage application is inside a bin m, and zero if otherwise. For example, when $m = 1$, the coefficient γ_1 captures the difference in the outcome variable between mortgages with a distance under 5 km and the full sample of mortgages. We cluster the errors at the county level.[17]

Discussion of Assumptions. Our specification of the model in equation (3) includes both county-year and lender-state-year fixed effects. Local shocks may be an important driver of the approval or acceptance of mortgage applications. For example, as the industrial composition of rural areas differs from that of cities, shocks to agriculture may decrease the quality of borrowers in rural areas. As rural areas are more likely to be under-served by lenders, shocks to agriculture may create a negative correlation between mortgage approval and distance. Hence, it is essential to fully absorb these local shocks with the fixed effects. We therefore compare two individuals from the same county to estimate γ_m. Borrowers may also select lenders based on their expectations about the probability of acceptance. Therefore, we include lender-state fixed effects to account for

[16] We define an approved mortgage as one in which the action type is either 1 or 2, conditional on the action type being between 1 and 3. We define an accepted mortgage as one in which the action type is 1, conditional on the action type being between 1 and 2.

[17] In all of our empirical analysis, we prefer to cluster the errors at the county level, because we are concerned about local shocks. However, for the analysis where the outcome variable is the share of banks, our data is at the bank-distance bin level and so does not have information at the county level. Therefore, for those regressions, we cluster at the lender level.

this sorting, as lender behavior may also exhibit spatial variation.

Results. We present the results of estimating equation (3) in Figure 4. In Panel (a), we find that the probability that the lender approves a mortgage application declines with distance. The coefficient associated with the lowest distance bin is close to 0.04, implying that mortgage applications under 5 km have a likelihood of approval that is 3.4 percentage points - or 8 percent - larger than the likelihood of approval for mortgages exceeding 2,000 km. The effect persists until 500 km, after which the possibility of approval is invariant to distance. The results in Panel (a) are consistent with the literature, which predicts that low-quality borrowers attempt to disguise their type by applying for a mortgage with a lender that is far away and thus does not have any soft information on the borrower. Lenders understand borrowers' incentives and are, therefore, more likely to reject a mortgage application from a borrower far away.

Panel (b) shows that a borrower's probability of accepting a mortgage proposal also decreases with distance. The coefficient associated with a distance under 5 km is 0.031, implying that a borrower's likelihood of accepting a mortgage contract is 3.1 percentage points - or 4.5 percent - for short distances relative to distances exceeding 2,000 km. The effects are less persistent than those we report for mortgage approval, as the probability of acceptance is flat for distances exceeding 25 km. The findings are also consistent with theory - as lenders understand borrowers who live far away are more likely to be of low quality, they offer worse terms, and borrowers are more likely to reject the proposal. Therefore, the results in Panel (b) suggest that the terms of the mortgage contract should be worse the larger the distance.

Heterogeneity. We also whether the role of distance in the production of mortgages is heterogeneous across borrowers. We estimate equation (3) for different types of borrowers. In Figure B.3, we present the results based on the income of the borrower. Both high- and low-income borrowers exhibit paths for the probability of having a mortgage approved by the lender. However, in terms of mortgage acceptance, high-income borrowers are as likely to accept a mortgage contract with a distance under 2.5 km as they are of accepting a mortgage contract offered by a lender that is 2,000 km away. In contrast, low-income borrowers are far more likely to accept a mortgage contract offered by a nearby lender. In Figure B.4, we present the results based on age. Younger borrowers and older borrowers see their mortgage applications approved at similar rates for the same distance bins, but younger borrowers are more likely to accept mortgage contracts from nearby lenders.

3.3 Pricing and Outcomes of Mortgages

Our results in Figure 4 suggest that lenders price the fact that low-quality borrowers may attempt to hide their type by making a mortgage application with a lender who is far away. To show this link, we study the correlation between distance and the terms of the mortgage. We focus on two elements of the mortgage contract: (1) the interest rate spread, calculated as the interest rate relative to the prime mortgage rate reported in Freddie Mac's weekly Primary Mortgage Market Survey (PMMS), and (2) the origination fees as a share of the total loan value. Our interpretation of the correlations in Figure 4 also relies on the assumption that low-quality borrowers are associated with larger distances. We can test this assumption by studying the correlation between the probability of delinquency and distance.

We do not observe the terms of the mortgage contract or possible delinquent behavior on the part of the borrower for all mortgages in our sample. We, therefore, rely on the sub-sample of mortgages in the HMDA-GSE data for the period between 2018 and 2023.[18] Using all mortgages in the sub-sample, we estimate the following equation

$$Y_{ijt} = \alpha_{c(i),q(i),t} + \mu_{j,t} + \beta X_{ijt} + \gamma \log \text{Distance}_{ijt} + \varepsilon_{ijt}, \tag{4}$$

where the outcome variable is of the four following variables: (1) the interest rate spread of mortgage i granted by lender j in year t, (2) origination fees as a share of the loan value, (3) an indicator variable that takes the value of 1 if the borrower becomes delinquent at least once in the seven years after origination and zero if otherwise, or (4) an indicator variable that takes the value of 1 if the borrower becomes delinquent at least once in the three years after origination and zero if otherwise. We include lender fixed effects and LLPA-county fixed effects ($q(i)$ denotes the LLPA group of the borrower in mortgage i). LLPA groups, or Loan Level Price Adjustment groups, are categories used in the mortgage industry to determine adjustments in pricing based on the credit score and the loan-to-value ratio. LLPA groups help lenders assess the risk associated with a loan and adjust interest rates or fees. We also include the logarithm of the loan amount as a control. The coefficient of interest is γ, which captures the semi-elasticity of the outcome variable to distance.

Discussion of Assumptions. Our running hypothesis is that mortgage lenders price the risk associated with distance in addition to the risk they observe from hard information about the borrower. To test the hypothesis, we must include all hard information available

[18] We include data starting in 2018 as we only observe origination fees after 2018.

to the lender in our regression. Consequently, we augment our fixed effect specification with county-LLPA group fixed effects. LLPA groups reflect all hard information available to the lender; therefore, this specification absorbs this variation. We are, thus, left with only soft information as we compare two individuals with the same measure of risk and in the same county. If lenders do not price soft information, we expect $\gamma = 0$ for the interest rate spread and origination fees. Moreover, if distance is not associated with lower-quality borrowers, we expect $\gamma = 0$ for the probability of delinquency.

Results. We present the results of estimating equation (4) in Table II. A higher distance between borrower and lender is associated with a higher interest rate spread and larger origination fees. A one standard deviation increase in the logarithm of distance is associated with a 4.5 percent increase in the interest spread - or a 0.4 basis point increase - and a 2.7 percent increase in the ratio of origination fees to loan value. Therefore, lenders understand that borrowers with a larger distance are riskier, even if they do not know why, and price the mortgages accordingly. However, they do not seem to fully price the added risk. A one standard deviation increase in the logarithm of distance is associated with a 3.2 percent increase in the probability of delinquency in seven years. Similarly, a one standard deviation increase in the logarithm of distance is associated with a 3 percent increase in the likelihood of delinquency in three years, so the increase in risk materializes in the short run.

The data on the cross-section of mortgages is consistent with a theory of hard and soft information. Soft information is hard to acquire and transmit, so only local lenders can produce it and cannot transmit it to other lenders. Consequently, low-quality borrowers may attempt to hide their type and apply for a mortgage with a lender that is far away. However, lenders understand borrowers' incentives and, therefore, view an increase in distance as an increase in risk. Consequently, they are less likely to approve a mortgage application with a high value for distance. On the intensive margin, they also price the added risk. However, the risk is not fully priced, as borrowers with a larger value for distance are more likely to become delinquent.

Heterogeneity. Distance is priced because lenders believe it is negatively correlated with the quality of borrowers. Consequently, the sensitivity of the terms of the mortgage contract to distance should be larger (in absolute value) for borrowers who are of lower quality. In Table III, we estimate equation (4) for different types of borrowers. We find that, for borrowers with an income above the median in their state, distance is not priced. For high-income borrowers, the correlation between interest rate spreads or origination fees and distance is not statistically different from zero. In contrast, low-income borrowers observe a positive and statistically significant correlation between distance and interest rate spreads or origination fees. A one standard deviation in the logarithm of distance for low-income borrowers is associated with a 6.7 percent increase in the interest rate spread and a 4 percent increase in origination fees as a share of loan value. We find similar results when we condition on the borrower's age - the correlation between distance and interest rate spreads is twice as large for young borrowers than it is for older borrowers. For example, for younger borrowers, a one standard deviation increase in the logarithm of distance is associated with a 6.1 percent increase in the interest rate spread and a 3.3 percent increase in origination fees.

The underlying assumption in our interpretation of the results in Table III is that low-income or younger borrowers have lower quality. We test this assumption by estimating equation (4) for different types of borrowers using delinquency as the outcome variable, and present the results in Table IV. We find that the correlation between distance and delinquency is larger for low-income borrowers or for younger borrowers. For low-income borrowers, a one standard deviation in the logarithm of distance is associated with a 4.3 percent increase in the probability of delinquency. For younger borrowers, a one standard deviation in the logarithm of distance is associated with a 3.4 percent increase in the probability of delinquency.

4 Time-Series Evidence

In recent years, soft information has become easier to collect and transmit. Liberti and Petersen (2019) refer to this process as the "hardening" of information. Consequently, soft information is easier to obtain, so distance should no longer play a role in mortgage markets. In our motivating example, the low-quality borrower's incentive to choose the far-away lender decreases as the lender observes a better signal about the borrower's quality. Consequently, over time, the hardening of information should increase the average distance between borrowers and lenders.[19] Moreover, distance's role in mortgage contracts should decrease as distance becomes a less relevant source of information. For example, the penalty for interest rate spreads for a higher distance should decline over time.

In the previous Section, we established a series of stylized facts about the role distance plays in the production, pricing, and outcomes of mortgages. In this Section, we provide

[19]This prediction is in line with the findings in PR, who document that the average distance between borrower and lender increased between 1973 and 1993.

evidence of the role of distance over time.

4.1 Market Share of Banks

Our goal is to understand how the local advantage of lenders varies over time. The local advantage of a lender is the difference between the lender's share in all mortgage applications in the vicinity of the branch and the same share on all mortgage applications in the US. We can estimate the local advantage of lenders by estimating $\gamma_{1,t}$ in equation (2). Therefore, to understand how the local advantage evolves, we estimate equation (2) for every year in our sample and present the results in Figure 5.

Between 1994 and 2000, the local advantage of lenders decreases by almost 50 percent. This period is characterized by an increase in the median distance between lenders and borrowers, as shown in Figure 1, which our model interprets as a decrease in the local advantage of banks. Between 2001 and 2009, we observe both an increase in the local advantage of banks and an increase in median distance.[20] Following 2009, the local advantage declines and remains stable until 2023.

The local advantage of lenders is characterized by stability over time, with sharp declines around recessions. The sharp decreases can be explained by waves of bankruptcy of mortgage providers. In recessions, smaller mortgage providers are more likely to exist. These smaller providers are also more likely to be local than national banks. Therefore, the average local advantage of banks will decrease. In contrast, the stability of the local advantage of lenders is at odds with some of the banking literature. One of the main drivers of local advantage is the ability of local banks to produce soft information. As soft information becomes less important (or easier to transmit), the local advantage of banks should decrease. Our results in Figure 5 do not support this hypothesis. Our results suggest that in 2023, like in 1994, the mortgage market remains local rather than national.

4.2 Production of Mortgages

In the previous Section, we showed that distance is negatively correlated with the probability a mortgage application is approved and that, conditional on approval on the part of the lender, distance is also negatively correlated with the probability the borrower accepts the terms. We now wish to study how this correlation evolves over time. Using our full

[20] The correlation between local advantage and distance is negative in 1994–2000 and positive in 2001–2009. One possible explanation is that the variation in distance was common across all lenders in 1994–2000 and heterogeneous across lenders in 2001–2009. If the increase in distance in 2001–2009 was driven by only a share of lenders, the variation in distance will be picked up by the lender fixed effects we include.

sample of mortgages, for each year t, we estimate the following regression

$$Y_{ijt} = \alpha_{c(i),t} + \mu_{s(i),j,t} + \beta_t X_i + \gamma_t \log \text{Distance}_{ijt} + \varepsilon_{ijt}, \tag{5}$$

where the outcome variable is either an indicator variable associated with the approval of mortgage application i made to lender j in year t or an indicator variable associated with the acceptance of the mortgage application. We include county and lender-state fixed effects, as well as a vector of mortgage-level controls. The coefficient of interest is γ_t which estimates the semi-elasticity of the outcome variable to distance.

Challenges to Identification. Our results in this Section are not meant to be interpreted as causal evidence of the effect of distance on mortgage approval or acceptance. Instead, we view them as evidence of the correlation between distance and mortgage approval and how it evolves over time. Nevertheless, to be able to interpret the coefficients, we need to take a stand on the direction of the bias. One potential challenge comes from geographic variation in the quality of the borrowers. Suppose low-quality borrowers live far away from bank branches. In that case, a negative correlation between distance and loan approval will be driven by a negative correlation between distance and quality. To the extent that the spatial sorting takes place at the county level, our inclusion of county-level fixed effects should address this issue. A second challenge comes from the fact that distance is an endogenous variable and is the product of a choice by the bank. If banks decide not to place branches in areas where they know borrowers have low quality, then low-quality borrowers will have to make mortgage applications to bank branches that are far away. If low-quality borrowers are less likely to see their mortgage approved, this introduces a reverse causality problem - higher distance drives lower mortgage approvals. Consequently, our coefficient will have a negative bias. The only way to fully address this reverse causality is to exploit exogenous variation in distance. However, even in the presence of the bias, our results are still informative because we observe how the correlation between distance and mortgage approval evolves over time. If banks' incentives to place branches are time-invariant, then the bias should not vary over time, and we can still compare the coefficients at different points in time.

Results. We present the results of estimating equation (5) on our full sample of mortgages in Figure 6. In Panel (a), we see that the correlation between distance and mortgage approval is negative, statistically significant, and stable over time. In 1994, a one standard deviation increase in the logarithm of distance is associated with a 0.68 percentage point decrease - or a 0.82 percent decline relative to the cross-sectional average in 1994 - in the probability that a mortgage application is approved. In 2000, a one standard deviation increase in the logarithm of distance is associated with a 2.64 percent decline in the probability of mortgage approval; in 2023, the effect was 1.93 percent. Therefore, there is always an economically significant correlation between distance and mortgage approval.

In Panel (b), we present the results for mortgage acceptance. The correlation between mortgage acceptance and distance is, like the correlation between approval and distance, negative and statistically significant for most years. However, the coefficients are converging towards zero following 2015 (even though the coefficient is still statistically significant from zero in 2023). Moreover, the value of the coefficients in Panel (b) is roughly half the value of the coefficients we report in Panel (a). In 1994, a one standard deviation increase in the logarithm of distance is associated with a 0.33 percent decrease in the probability a mortgage contract is accepted. In 2023, a one standard deviation increase in the logarithm of distance is only associated with a 0.18 percent decline in mortgage acceptance. Therefore, unlike what we find for mortgage approval, the role distance plays in shaping mortgage acceptance is decreasing in importance over time.

Robustness. In Figure B.5, we present the results of estimating (5) for a variety of specifications in which the most restrictive one is the one we show in Figure 6. The correlation between distance and mortgage approval is negative and statistically significant across all models and years. However, for models that do not include lender fixed effects, the correlation between mortgage approval and distance converges to zero over time. Once we include lender fixed effects, the correlation is relatively constant over time. This is evidence of selection into specific lenders - low-quality borrowers are likely choosing specific lenders, and once we account for this selection, the correlation is constant over time.

Our cross-sectional results in Section 3 are consistent with the banking literature, which predicts a negative correlation between distance and mortgage approval. However, the literature also predicts that, as soft information becomes easier to collect and to transmit, the correlation should become smaller. Our results in Figure 6 show that this prediction did not come to pass. Instead, the distance is as correlated with mortgage approval with 2023 as it was in 1994, if not more.[21] In contrast, the correlation between mortgage accep-

[21] One possible explanation for this lack of convergence is the negative bias induced by the endogenous choice of location by the lenders. However, if that was the case, and if that bias was constant over time, the coefficients would not converge to zero but would decrease in absolute value, which we also do not find. Another possible issue can arise if there is substantial heterogeneity in borrower quality within counties. The within-county heterogeneity implies that county fixed effects are not enough to absorb variations in lender quality, which are not captured by distance. To address this, we estimate equation 2 including as controls the logarithm of the applicant's income, the logarithm of the value of the property, and a series of

tance and distance has become weaker. Therefore, two potential borrowers from the same county who apply to the same lender and differ only in their distance to the branch of the lender do not accept the mortgage contract with different probabilities in 2023. There are two potential explanations for this fact. One possibility is that lenders reject the borrower with the largest value of distance, which is consistent with Panel (a) of Figure 6. Another possibility is that lenders no longer price distance in mortgage contracts. We investigate the second possibility in the next subsection.

Heterogeneity. We estimate equation (5) for different types of borrowers. In Figure B.7, we present results for low- and high-income borrowers. The correlation between distance and mortgage approval is larger (in absolute value) for low-income borrowers, as we also show in Figure B.3. However, the correlation for both groups moves in similar ways over time. Similarly, the correlation between distance and mortgage acceptance evolves in a similar for low- and high-income borrowers. We also estimate equation (5) for young and older borrowers, as we show in Figure B.8. For both mortgage approval and mortgage acceptance, the correlation with distance varies in similar ways for both young and old borrowers.

4.3 Pricing of Mortgages

We now turn to the pricing of mortgages. We established in Section 3 that lenders price distance, even conditional on hard information about the borrower. Our goal is to understand how this pricing varies over time. We therefore estimate, for each year t, on the HMDA-GSE sub-sample of mortgages, the following equation:

$$Y_{ijt} = \alpha_{c(i),q(i),t} + \mu_j + \beta_t X_i + \gamma_t \log \text{Distance}_{ijt} + \varepsilon_{ijt}, \tag{6}$$

where the outcome variable is either the spread for mortgage i given by lender j in year t or the ratio of origination fees to the total value of the loan. We include lender fixed effects and LLPA-county fixed effects. We also include the logarithm of the loan amount as a control. The coefficient of interest is the γ_t, which captures the semi-elasticity of the spread to distance. Therefore, we think of equation (6) as the time-series version of equation (4).

fixed effects for levels of the debt-to-income ratio. We can only estimate this for 2018–2023, and we present the results in Figure B.6. We find that the coefficient associated with distance does not change with the inclusion of the additional controls, which suggests that there is no substantial heterogeneity in borrower quality within counties or that this variation is not correlated with distance.

Results. We present the results of estimating equation (6) in Figure 7. In Panel (a), we find that larger distances are, for most years, associated with larger interest rate spreads. Moreover, following 2008, the correlation is usually positive and statistically significant. In 2000, a one standard deviation increase in the logarithm of distance is associated with a one basis point increase in the interest rate spread - or a 1.64 percent increase relative to the unconditional mean. In 2023, a one standard deviation increase in the logarithm of distance is associated with a 2.44 percent increase in the interest rate spread. The results for origination fees in Panel (b) are very similar, as the correlation between distance and origination fees is positive and statistically significant. In 2018, a one standard deviation increase in the logarithm of distance is associated with a 2.19 percent increase in the ratio of origination fees to total loan value relative to the unconditional mean; in 2023, the effect had grown to 3.34 percent.

Our results in Figure 7 show that distance is always priced. Lenders understand that higher distances may be associated with low-quality borrowers and adjust spreads and origination fees to reflect the added level of risk. However, the price of distance is not decreasing over time, which is the pattern a decline of soft information should imply. Instead, our findings are consistent with a model in which soft information is still important and is, therefore, priced in mortgage contracts.

Heterogeneity. In Figure B.10, we present the results of the estimation of (6) for low- and high-income borrowers. We find that most of the time-series variation in the correlation between distance and interest rate spreads is driven by low-income borrowers. In most periods, the correlation is positive and statistically significant, suggesting that, for low-income borrowers, distance is priced in mortgage contracts. For high-income borrowers, there is no correlation between distance and origination fees. In contrast, for low-income borrowers, increases in distance are associated with higher origination fees. We also present results by borrower age in Figure B.11. There is no difference between young and old borrowers in terms of the sensitivity of origination fees to distance. In contrast, only young borrowers display a positive association between interest rate spreads and distance.

4.4 Delinquency

We now turn to delinquency rates. In Table II, we showed that, for the 2018–2023 period, higher distance was associated with a larger incidence of delinquent behavior on the part of borrowers. In line with our previous empirical exercise on the pricing of distance, our goal is to understand if the correlation between distance and delinquency is stable over time. We, therefore, estimate equation (6) using the indicator variable for delinquency as the outcome variable and present the results in Figure 8.

In Panel (a), we focus on delinquency in the seven years following the mortgage contract's origination. The correlation between distance and delinquency is positive and, for some periods, statistically significant.[22] It increased substantially before the Great Recession, which is explained by the increase in delinquency during the recession. In 2000, a one standard deviation increase in the logarithm of distance is associated with a 5.6 percent increase relative to the unconditional mean in the probability of delinquency. In 2023, a one standard deviation increase in the logarithm of distance is associated with a 2.3 percent increase in the likelihood of delinquency. Therefore, if we assume that ex-post delinquency is a good proxy for ex-ante borrower quality (conditional on controls), higher distances correlate with lower borrower quality. Our results hold if we instead focus on the probability of delinquent behavior in the three years following the origination of the mortgage contract, as we show in Panel (b).

Heterogeneity. In Figure B.13, we present the results of the estimation of (6) using delinquency as the outcome variable for different groups of borrowers. We find that, over time, for a given value of distance, low-income borrowers are more likely to become delinquent than high-income borrowers. Similarly, the association between distance and delinquency is stronger for younger borrowers than it is for older borrowers.

Overall, our delinquency results align with the results on the likelihood of a mortgage application being approved and with the results on the pricing of distance. Despite the hardening of information, mortgage lenders still price distance and associate it with low-quality borrowers.

5 Model

In this Section, we present a simple spatial model of the mortgage market. The model features multiple locations and lenders, as well as endogenous participation into the mortgage market and dispersion of interest rates within and across locations.

[22] These results are robust to the specification we use, as we show in Figure B.12. Most of the coefficients are positive, and most are also statistically significant.

5.1 Environment

The model is static. Space is a set $\mathcal{O} \subset \mathbb{R}^2$ and there are N locations. Each location has a continuum of potential borrowers that want to obtain a mortgage to buy a house. Locations are identical. There is a finite number $B < N$ of lenders (she), each in a particular location. Let b identify both a lender and her location. For each lender b and location i, let d_{bi} denote the distance between the lender and the location.[23] Let the matrix $\mathbb{D}$ be a $N \times B$ matrix where element (i, b) is d_{ib}.

Timing. There are three subperiods. In the first subperiod, the borrowers learn about their quality and choose to which lender they apply for a mortgage. In the second subperiod, the lenders receive applications, observe signals, and set interest rates. At the end of the period, the borrowers decide whether or not to accept the mortgage contract.

Borrowers. Each borrower is characterized by their quality $\theta \sim \mathcal{U}[\theta_0, \theta_1]$ where $\theta_0 > 0$ and $\theta_1 < 1$. In this model, a borrower's quality is the probability he does not become delinquent. We assume delinquency is exogenous and so this is a model of adverse selection. The borrower makes two choices: (1) which lender to apply for a mortgage, and (2) whether or not to accept the application. In the first subperiod, each borrower j in location i chooses a lender b. We assume the borrower's are unsophisticated and believe all lenders will offer the same interest rate $R = b/\theta$. The ex-ante utility borrower j in location i derives from applying to lender b is given by

$$U_{jib}(\theta) = \bar{u} - \theta \frac{b}{\theta} - \frac{1}{\mu} d_{ib} + \varepsilon_{ijb},$$

where $1 < b < \bar{u} < 2$ is the utility the borrower has from owning a house, θ is the probability the borrower does not become delinquent, $\mu > 0$ is a parameter governing the search costs, and ε_{ijb} is a taste shock drawn from a common extreme value distribution. There is an outside option $U_{ji0} = 0$. If the borrower does receive a mortgage offer, his utility is then $u(R, \theta) = \bar{u} - \theta R$, as search costs are sunk. Therefore, this is a model in which borrowers are not time-consistent and may make mistakes in predicting interest rates.

Lender. All lenders are identical, apart from their location, and so we describe only the problem of one lender b. She obtains applications from borrowers and decides on

[23] Suppose location i has coordinates (x_i, y_i) and bank b has coordinates (x_b, y_b). Distance is computed using Pythagoras's theorem and so $d_{ib} = \sqrt{(x_i - x_b)^2 + (y_i - y_b)^2}$.

the interest rate she charges. Once the lender receives a mortgage application, she also receives a signal s about the quality of the borrower. The signal is drawn from a uniform distribution - $s|\theta \sim \mathcal{U}[e^{-(1/\beta)d}\theta + (1 - e^{-(1/\beta)d})\theta_0, e^{-(1/\beta)d}\theta + (1 - e^{-(1/\beta)d})\theta_1]$, where $\beta > 0$ governs the precision of the signal and d represents distance. The lender observes the signal and the distance relative to the borrower. She then sets an interest rate, which the borrower may then accept or reject. We assume there is an infinitesimally small cost of making an offer, so that the lender does not make offers she knows will be rejected. If the offer is accepted, the payoff of the lender is $-1 + \theta R$.

Note that, as all lenders are identical apart from location, the solution to their problem depends only on the signal s and distance d. Therefore, the entire structure of the model is condensed in the matrix $\mathbb{D}$.

Discussion. In our model, we consider only the case of adverse selection, as borrowers have different types leading to different delinquency rates. Consequently, the problem of the lender becomes a simple screening problem which then generates dispersion in interest rates. Crucially, borrowers cannot choose to default and default is effectively exogenous. If they could choose, then the model would also feature a moral hazard dimension which would not alter our results but would add greater complexity.

The crucial assumption we make that allows us to solve the model is that borrowers are unsophisticated and not time consistent. As borrowers are unsophisticated, they believe all lenders will make the same offer and that this is offer takes the form b/θ. This assumption then implies that the ex-ante utility of the borrower for a given bank is not a function of his quality. If the ex-ante utility was a function of quality, the problem would become far harder to solve as borrowers would behave strategically and lenders, understanding the incentives of borrowers, would react accordingly. We also assume that the ex-post utility is different than the ex-ante utility. The assumption of time inconsistency implies that borrowers don't anticipate the possibility that they might have to reject the mortgage offer. In the model, this implies we can generate dispersion in interest rates as well as endogenous participation in the mortgage market.

The model also features interest rate dispersion within and across locations for the same signal. There is dispersion within locations as there is heterogeneity across potential borrowers, which matches empirical evidence. There is also dispersion across locations for individuals that are otherwise identical. The second type of dispersion is consistent with the pricing of distance we document in Table II.

We also don't feature entry.[24] Lenders cannot choose to enter more locations. If they

[24] One possibility would be use a version of the model of bank spatial competition presented in Oberfield

could, this would lead to a reduction in distance, which is inconsistent with the empirical evidence showing that distance is constant following the Great Recession. Nevertheless, as long as lenders don't enter all locations, we would still match all of the cross-sectional evidence and most of the time-series evidence for some values of μ and β.

5.2 Solution

We solve the model backwards, starting from the last subperiod.

Problem of borrower with mortgage offer. If a borrower receives an offer R, he will accept the offer if and only if $R \leq \frac{\bar{u}}{\theta}$, which creates an upper bound for the interest rate the lender can charge.

Problem of the lender. The lender understands that the borrower will reject interest rates above the threshold and so, conditional on a given signal s and a distance d, sets interest rates to maximize her expected payoff

$$\text{Payoff}(R, s, d) = \int_{\theta_0}^{\theta_1} [-1 + \theta R] \times \mathbf{1}\left\{\theta \leq \frac{\bar{u}}{R}\right\} \times f_{\theta|s;d}(\theta|s; d)d\theta.$$

The expected payoff depends on three terms: (1) the payoff conditional on acceptance, (2) the probability that the offer is accepted, and (3) the conditional density of the borrowers' type, conditional on the signal received by the lender and distance. The following Proposition summarizes the solution to the lenders's problem.

Proposition 1 *The solution of the lender's problem is summarized by the optimal interest rate $R^\star(s, d)$ and the conditional density of θ $f_{\theta|s;d}$ given the signal s and distance d. These functions are given by*

$$R^\star(s, d) = \max\left\{\frac{\bar{u}}{\theta_{max}(s, d)}, \frac{\sqrt{2\bar{u}(1 - 0.5\bar{u})}}{\theta_{min}(s, d)}\right\}, \tag{7}$$

$$f_{\theta|s;d}(\theta|s; d) = \begin{cases} \frac{1}{\theta_{max}(s,d) - \theta_{min}(s,d)}, & \textit{if } \theta \in [\theta_{min}(s, d), \theta_{max}(s, d)] \\ 0, & \textit{if otherwise,} \end{cases} \tag{8}$$

et al. (2024).

where

$$\theta_{min}(s,d) = \max\left\{\theta_0, \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-(1/\beta)d}}\right\},$$
$$\theta_{max}(s,d) = \min\left\{\theta_1, \frac{s-(1-e^{-(1/\beta)d})\theta_0}{e^{-(1/\beta)d}}\right\}.$$

Therefore, $\theta|s$ follows a uniform distribution.

Proof. The proof is in Appendix C. ■

We plot the optimal interest rate for different signals and levels of distance in Figure 9. The interest rate is decreasing with the signal. If the lender has a higher belief about the borrower's type, she will charge a lower interest so that the probability the borrower rejects the mortgage proposal is lower. The lender wants to do this because she can extract more surplus from the borrower as $\bar{u} > 1$, i.e. because there are gains from trade. Conditional on a given signal, the optimal interest rate increases with distance as the precision of the signal becomes worse.[25] Finally, in location 0, where $d = 0$, the signal is perfect and so the lender has degenerate beliefs about θ. Hence, she is able to extract all of the surplus from the borrower and charges $R^\star(s,0) = \bar{u}/s$. Note that the interest rate charged when distance is zero matches the ex-ante beliefs of borrowers.

Problem of the borrower before choosing a lender. Given the structure of the problem, the optimal choice of the borrower in location i is given by

$$P_{ib} = \frac{\exp\{\bar{u} - b - (1/\mu)d_{ib}\}}{1 + \sum_{b'}\exp\{\bar{u} - b - (1/\mu)d_{ib'}\}}$$
$$P_{i0} = \frac{1}{\sum_{b'}\exp\{\bar{u} - b - (1/\mu)d_{ib'}\}}$$

which do not depend on quality. Therefore, the market share of bank b in location i is given by $P_{ib}/\sum_{b'}P_{ib'}$. Moreover, participation in the search for mortgage is given by P_{i0}.

Not all mortgage proposals are accepted by borrowers. For a given distance d, conditional on receiving an application, the probability a mortgage proposal is then accepted by the borrower is

$$\mathcal{P}(d) \equiv \int_{\theta_0}^{\theta_1} F_{\theta|s;d}\left(\frac{\bar{u}}{R^\star(s,d)}|s;d\right) f_s(s;d)ds. \tag{9}$$

[25] In Proposition 2, we present the expression for the elasticity of the optimal interest rate with respect to distance.

To understand equation (9), suppose a lender receives a signal s. She then sets the interest rate according to the optimal policy in equation (7). The probability this proposal is accepted by the borrower is the probability the borrower's type does not exceed $\bar{u}/R^{\star}(s,d)$, which we can compute using the conditional distribution in Proposition 1. The lender receives a continuum of signals, and so to compute the total probability, we need to add these probabilities weighted by the density of the signal.[26]

We can interpret $\mathcal{P}(d)$ as the probability a mortgage is originated with distance d, conditional on a mortgage application. Therefore, the expression in (9) maps well to our empirical evidence. We can also define the average distance between all borrowers with a mortgage and a lender b as

$$\mathcal{D}(b) \equiv \frac{1}{N}\sum_i P_{ib} \times \mathcal{P}(d_{ib}) \times d_{ib}. \tag{10}$$

5.3 Cross-sectional implications

We now turn to the model's implications for the cross-sectional mortgage data. We focus on three key empirical facts: (1) the negative correlation between distance and the market share of lenders, (2) the negative correlation between distance and mortgage origination, and (3) the positive correlation between interest rates and distance. We present the model's results in Figure 10.

As distance increases, the lender's probability of receiving a mortgage application decreases. This finding matches the results in Figure 3 – lenders' market share declines with distance. Distance increases also lead to a lower probability of mortgage origination. In the model, borrowers further away from the lender are less likely to accept the mortgage contract because they are more likely to receive an interest rate they deem too high. Therefore, the model matches the patterns we document in Figure 4 – the probability that a mortgage is originated declines with distance.

Finally, the model generates a positive relation between interest rates and distance, except for small distances. For large enough distances, increases in distance lead to higher interest rates, conditional on the same signal s, which is consistent with our empirical findings in Table II. The higher interest rate is driven by the fact that the precision of the signal the lender receives decreases exponentially with distance. However, the lender still bears risk because of the participation constraint of the borrower and so the lender cannot extract all of the gains from trade. When distance is small, distance increases may lead to a reduction in the interest rate because the lender prefers to keep the probability

[26] We present the expression for the density of the signal in equation (C.1).

of origination high, requiring a lower interest rate.

5.4 Time-series implications

We now turn to the time series. The period between 1994 and 2023 is characterized by improvements in the technology lenders use to screen potential borrowers. In the model, the technology gains can be modeled as an increase in β - conditional on a given distance d, increases in β lead to an increase in the precision of the model. Therefore, the lender can charge a lower interest rate, increasing the probability that a loan is originated. Consequently, mortgage applications from borrowers far away become more likely to be accepted, leading to an increase in distance. We present the solution of the model for different values of β in Figure 11.

The probability the lender receives an application is identical for all values of β, as we assume this variable only depends on distance and α. In panel (b), we plot the probability a mortgage is originated as a function of distance. Screening technology improvements lead to mortgage origination increases, as the lender can charge a lower interest rate. In fact, and as we see in panel (c), the elasticity of the interest rate to distance declines with improvements in screening technology. In our model, the elasticity is driven entirely by lenders pricing distance as a source of risk. Finally, the increases in the productivity of the screening technology lead to increases in distance.

The results in Figure 11 are inconsistent with our empirical findings. In particular, the model predicts a decrease in distance, which is inconsistent with our findings in Figure 1, where we show that distance remained relatively constant from 2008 onwards. The model also predicts a reduction in the pricing of distance, which is not inconsistent with our empirical findings.

The last 30 years have been characterized by more than improvements in the productivity of screening technology. Lenders can now advertise their products to potential borrowers at a lower cost. Moreover, borrowers may also search for lenders at a lower cost. Therefore, competition in the mortgage market has become fiercer. In our model, this increase in competition can be captured by the parameter μ, where increases in μ imply a decrease in search costs. The implications for lenders are not trivial. On the one hand, borrowers that are far away are more likely to match with the lender. On the other hand, borrowers that are close may match with other lenders. Therefore, the local advantage of the lender may decrease, which is consistent with our findings in Figure 5. We present the solution of the model for different values of μ in Figure 12.

As μ increases, the probability a lender receives a mortgage application from a borrower that is close by decreases, while this probability increases for borrowers that are far away. In other words, as competition is fiercer, the lender's market share decreases overall but decreases by more for larger distances. Conditional on receiving an application, neither the probability of origination nor the elasticity of interest rates to distance change. However, as competition becomes fiercer due to lower search costs, the local advantage of lenders becomes more relevant, leading to a decrease in distance.

The model shows improvements in the productivity of the screening technology cannot rationalize the data on US mortgages. However, if productivity improvements occur as borrowers' search costs decrease, the model can explain why both distance and its price in mortgage contracts have remained stable.

6 Conclusion

What role do local bank branches play in mortgage lending? In this paper, we document three facts about the geographic scope of mortgage markets. In line with the predictions from the traditional banking literature mentioned above, we find that local lenders have a significant advantage in terms of attracting, approving, and originating mortgages from local borrowers. Our second fact is that the distance between borrower and lender varies pro-cyclically over the business cycle, rising in booms and falling during the busts. This finding is consistent with the idea that credit quality deteriorates as soft information is harder to collect when the distance increases. Our third fact is that, across many dimensions, the role of distance is the same today as it was 30 years ago.

We assemble a dataset covering the near universe of U.S. mortgage applications from 1994 to 2023. Our main variable of interest is the geographical distance between the borrower and all the possible banks to which she could submit a mortgage application. For a given mortgage application, we find the average bank is over 200 times more likely to receive this application if the bank has a branch within 5 km of the applicant (relative to the bank's average market share vis-a-vis its competitors). We also show that distance is negatively correlated with the likelihood that a mortgage application is approved by the lender. Conditional on approval, distance is priced in the mortgage contract - an increase in distance is associated with a higher interest spread and larger origination fees. We also find that distance is positively correlated with delinquent behavior.

The role that distance plays in mortgage contracts is also stable over time, although it moves with the business cycle. The sensitivity of distance to mortgage approval has not changed over time. Similarly, the sensitivity of interest rate spreads or origination fees to distance is table over time.

To explain both the cross-section and the time series of US mortgages, we construct a spatial model of the mortgage market where borrowers search for lenders and lenders screen potential borrowers. The model explains the cross-sectional stylized facts. The model is also able to explain the puzzle in the time series - why don't increases in the productivity of screening lead to increases in distance? Through the lens of the model, we find that decreasing search costs lowers the match quality of borrower search, which counteracts the effects of increasing screening technology. Quantitatively, these two effects negate each other.

Combined, our findings highlight that the impact of the so-called "fintech revolution" on the role of bank branches may not be as dramatic as previously argued. Instead, we argue that the bank branch is well and alive and still highly relevant for mortgage lending, even in the modern economy.

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*, 2023, *156* (C).

Agarwal, Sumit and Robert Hauswald, "Distance and Private Information in Lending," *The Review of Financial Studies*, 04 2010, *23* (7), 2757–2788.

_ , John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao, "Searching for Approval," *Econometrica*, 2024, *92* (4), 1195–1231.

Amel, Dean F, Elliot Anenberg, Rebecca Jorgensen et al., "On the Geographic Scope of Retail Mortgage Markets," *Board of Governors of the Federal Reserve System, FEDS Notes, June*, 2018, *15.*

Berger, Allen, "The Economic Effects of Technological Progress: Evidence from the Banking Industry," *Journal of Money, Credit and Banking*, 2003, *35* (2), 141–76.

Brevoort, Kenneth P. and John D. Wolken, "Does distance matter in banking?," Finance and Economics Discussion Series 2008-34, Board of Governors of the Federal Reserve System (U.S.) 2008.

Brickley, James A, James S Linck, and Clifford W Smith Jr, "Boundaries Of the Firm: Evidence From The Banking Industry," *Journal of Financial Economics*, 2003, *70* (3), 351–383.

Buchak, Greg and Adam Jørring, "Competition with Multi-Dimensional Pricing: Evidence from U.S. Mortgages," *Working Paper*, 2024.

Cerutti, Eugenio, Galina Hale, and Camelia Minoiu, "Financial crises and the composition of cross-border lending," *Journal of International Money and Finance*, 2015, *52* (C), 60–81.

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

Diamond, Douglas W., "Financial Intermediation and Delegated Monitoring," *The Review of Economic Studies*, 07 1984, *51* (3), 393–414.

_ , "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt," *Journal of Political Economy*, 1991, *99* (4), 689–721.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl, "The Deposits Channel of Monetary Policy," *The Quarterly Journal of Economics*, 2017, *132* (4), 1819–1876.

Favara, Giovanni and Mariassunta Giannetti, "Forced Asset Sales and the Concentration of Outstanding Debt: Evidence from the Mortgage Market," *The Journal of Finance*, 2017, *72* (3), 1081–1118.

Federal Reserve System, "Order Approving the Acquisition of a Savings Association and Other Nonbanking Activities," *Federal Reserve Board Notice*, 2008, *June 05.*

_ , "How do the Federal Reserve and the U.S. Department of Justice, Antitrust Division, analyze the competitive effects of mergers and acquisitions under the Bank Holding Company Act, the Bank Merger Act and the Home Owners Loan Act?," *https://www.federalreserve.gov/bankinforeg/competitive-effects-mergers-acquisitions-faqs.htm*, 2021.

Giannetti, Mariassunta and Luc Laeven, "The flight home effect: Evidence from the syndicated loan market during financial crises," *Journal of Financial Economics*, 2012, *104* (1), 23–43.

Granja, João and Nuno Paixão, "Market Concentration and Uniform Pricing: Evidence from Bank Mergers," *Working Paper*, 2019.

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

Haas, Ralph De and Neeltje Van Horen, "Running for the Exit? International Bank Lending During a Financial Crisis," *The Review of Financial Studies*, 01 2013, *26* (1), 244–285.

Hauswald, Robert and Robert Marquez, "Competition and Strategic Information Acquisition in Credit Markets," *The Review of Financial Studies*, 02 2006, *19* (3), 967–1000.

Head, Keith and Thierry Mayer, "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 131–195.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating Trade Flows: Trading Partners and Trading Volumes," *The quarterly journal of economics*, 2008, *123* (2), 441–487.

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

Ho, Katherine and Joy Ishii, "Location and competition in retail banking," *International Journal of Industrial Organization*, 2011, *29* (5), 537–546.

Hurst, Erik, Benjamin J. Keys, Amit Seru, and Joseph Vavra, "Regional Redistribution through the US Mortgage Market," *American Economic Review*, October 2016, *106* (10), 2982–3028.

Jayaratne, Jith and Philip E. Strahan, "The Finance-Growth Nexus: Evidence from Bank Branch Deregulation*," *The Quarterly Journal of Economics*, 08 1996, *111* (3), 639–670.

Jiang, Erica Xuewei and Anthony Lee Zhang, "Collateral Value Uncertainty and Mortgage Credit Provision," *Working Paper*, 2022.

Jiang, Sheila, , Adam Jørring, and Douglas Xu, "Bank Technology Adoption and Loan Production in the U.S Mortgage Market," *Working Paper*, 2023.

Johnson, Christian and Tara Rice, "Assessing a Decade of Interstate Bank Branching," *Washington and Lee Law Review*, 2008, *65* (1).

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *The Quarterly Journal of Economics*, 02 2010, *125* (1), 307–362.

Kleimeier, Stefanie, Harald Sander, and Sylvia Heuchemer, "Financial crises and cross-border banking: New evidence," *Journal of International Money and Finance*, 2013, *32* (C), 884–915.

Koont, Naz, "The Digital Banking Revolution: Effects on Competition and Stability," *Working Paper*, 2023.

Liberti, Jose M and Atif R Mian, "Estimating The Effect of Hierarchies on Information Use," *The Review of Financial Studies*, 2008, *22* (10), 4057–4090.

Liberti, José María and Mitchell A Petersen, "Information: Hard and soft," *Review of Corporate Finance Studies*, 2019, *8* (1), 1–41.

Liberti, José María and Mitchell A Petersen, "Information: Hard and Soft," *The Review of Corporate Finance Studies*, 11 2018, *8* (1), 1–41.

Loutskina, Elena and Philip E. Strahan, "Informed and Uninformed Investment in Housing: The Downside of Diversification," *The Review of Financial Studies*, 02 2011, *24* (5), 1447–1480.

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.

Oberfield, Ezra, Esteban Rossi-Hansberg, Nicholas Trachter, and Derek Wenning, "Banks in Space," March 2024, (32256).

Petersen, Mitchell A and Raghuram G Rajan, "The Benefits of Lending relationships: Evidence from Small business Data," *The Journal of Finance*, 1994, *49* (1), 3–37.

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

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

Scharfstein, David and Adi Sunderam, "Market power in mortgage lending and the transmission of monetary policy," *Unpublished working paper.*, 2016, *Harvard University.*

Stein, Jeremy C., "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms," *The Journal of Finance*, 2002, *57* (5), 1891–1921.

Sussman, Oren and Joseph Zeira, "Banking and Development," *Centre for Economic Policy Research Discussion Paper Series*, 1995, (1127).

van der Plaat, Mark, "Loan sales and the tyranny of distance in U.S. residential mortgage lending," MPRA Paper 107519, University Library of Munich, Germany Sep 2020.

Tables and Figures

FIGURE 1: Median Distance of Originated Mortgages

This Figure presents the median distance of U.S. mortgages originated by banks between 1994 and 2023. The Figure presents the median distance between the borrower's census tract and the bank's closest branch (in km). The grey bars represent recessions, as defined by the NBER. See Section 2 for sample description.

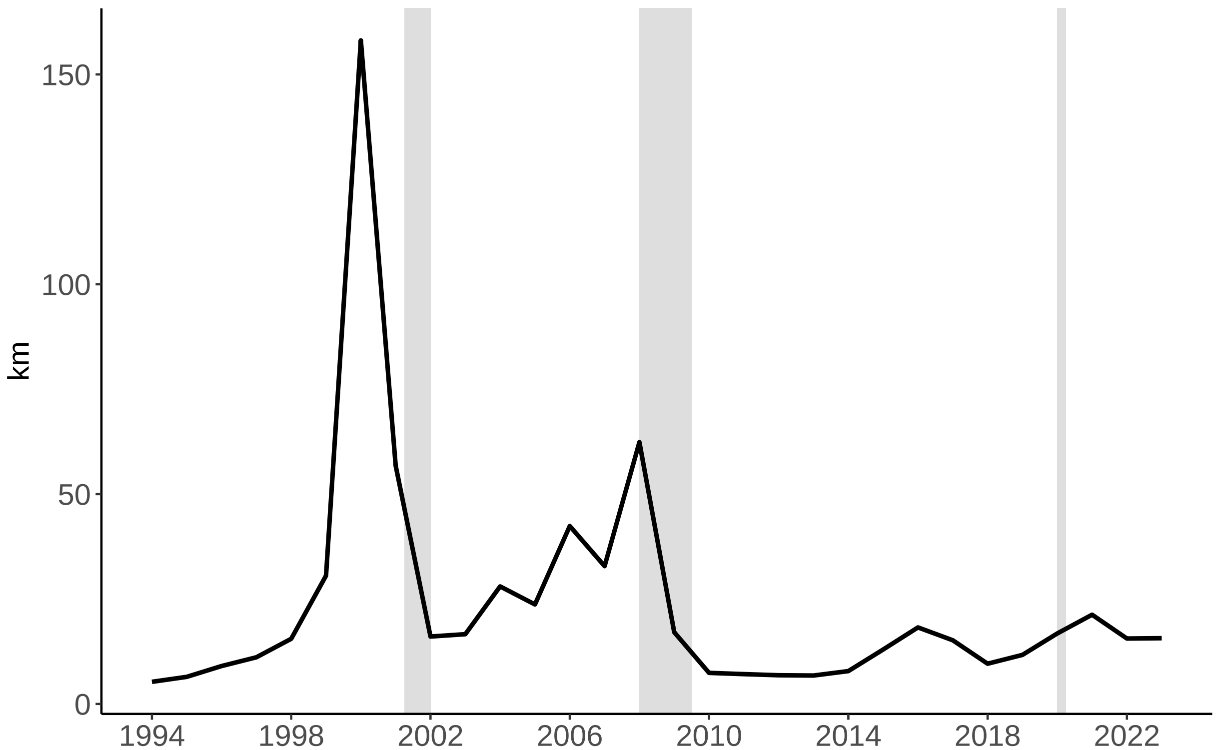


FIGURE 2: Decomposition of Mortgages by Distance

This Figure decomposes the number of mortgages (as a share of the total number of mortgages) according to the distance between the borrower and the lender. We consider eight bins: (1) distances under 2.5 km, (2) distances between 2.5 km and 5 km, (3) distances between 5 km and 10 km, (4) distances between 10 and 25 km, (5) distances between 25 km and 50 km, (6) distances between 50 km and 100 km, (7) distances between 100 km and 150 km, and (8) distances greater than 150 km.

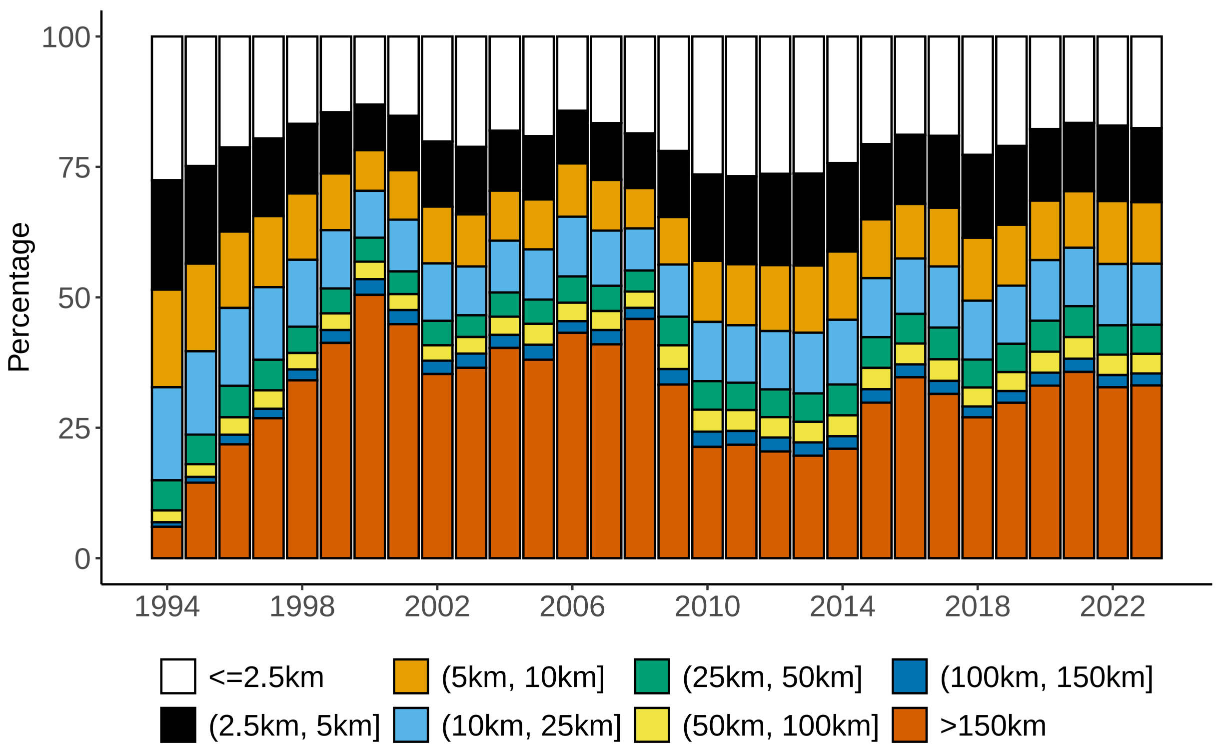


FIGURE 3: Share of Mortgages and Distance

This Figure presents the results of estimating equation (2) for all years between 1994 and 2023. The outcome variable is the share of all mortgages in a given radius which are directed at a given lender. We include lender-time fixed effects. We consider ten radii with which we draw circles around bank branches: 5, 10, 25, 50, 100, 150, 200, 500, 1,000, and 2,000 km. We therefore end up with 11 distance bins, as we use the full sample as the eleventh bin. We include a series of fixed effects for the bins, using the eleventh bin ($> 2,000$ km) as the reference group. We plot the fixed effects associated with the distance bins along with a 95% confidence interval. Errors are clustered by lender.

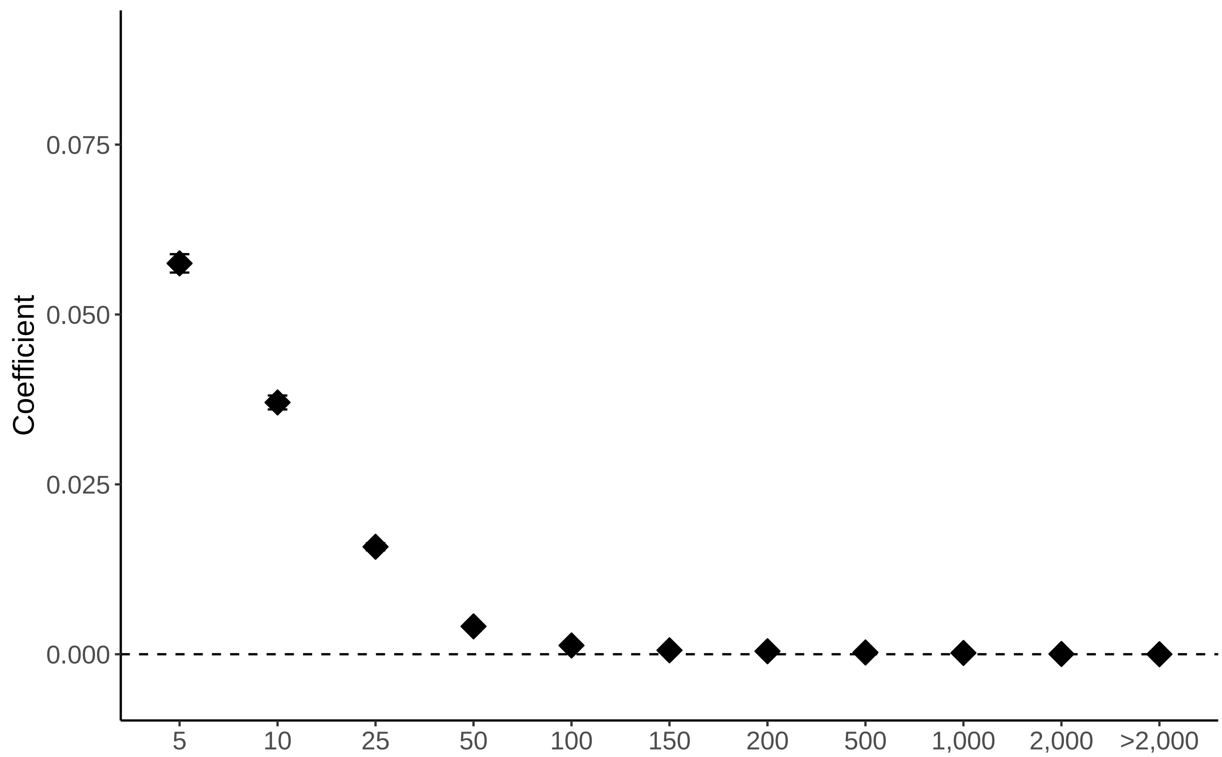
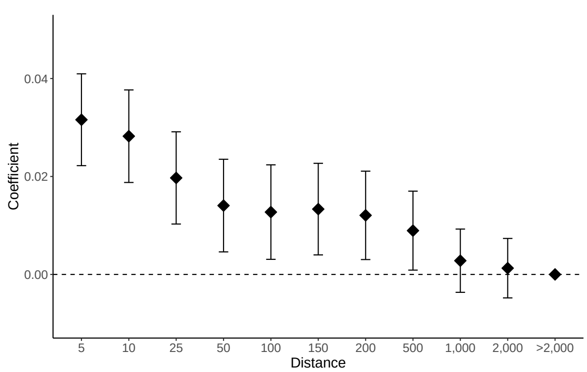


FIGURE 4: Mortgage Approval and Acceptance

This Figure presents the results of estimating equation (3) for our full sample of mortgages. The outcome variable is either mortgage approval - an indicator variable that takes the value of 1 if the mortgage application is approved by the lender or zero if otherwise - or mortgage acceptance - an indicator variable that takes the value of 1 if the mortgage contract is then accepted by the borrower, or zero if otherwise (and is not defined for mortgages that where denied by the lender). We include county-year and lender-state-year fixed effects as well as the logarithm of the loan amount as a control. We present coefficients associated with the distance bins - 5, 10, 25, 50, 100, 150, 200, 500, 1,000, and 2,000 km - using the eleventh bin ($> 2,000$ km) as the reference group. We plot the fixed effects associated with the distance bins along with a 95% confidence interval. Errors are clustered by lender.

(a) Approval



(b) Acceptance

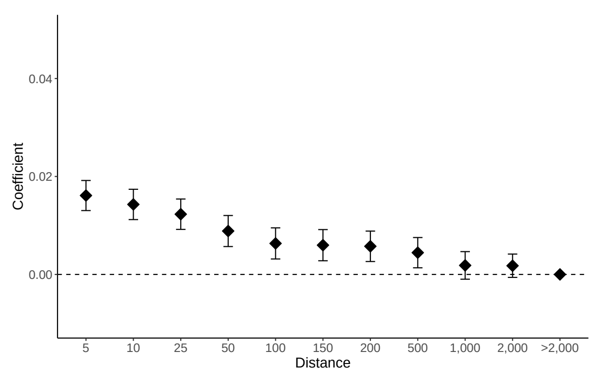


FIGURE 5: Evolution of Local Advantage

This Figure presents the results of estimating equation (2) for each year between 1994 and 2023. The outcome variable is the share of all mortgages in a given radius which are directed at a given lender. We include lender fixed effects. We consider ten radii with which we draw circles around bank branches: 5, 10, 25, 50, 100, 150, 200, 500, 1,000, and 2,000 km. We therefore end up with 11 distance bins, as we use the full sample as the eleventh bin. We include a series of fixed effects for the bins, using the eleventh bin ($> 2,000$ km) as the reference group. We plot the fixed effect associated with the the first distance bin (≤ 5 km) for all years along with a 95 percent confidence interval. Standard errors are clustered by lender.

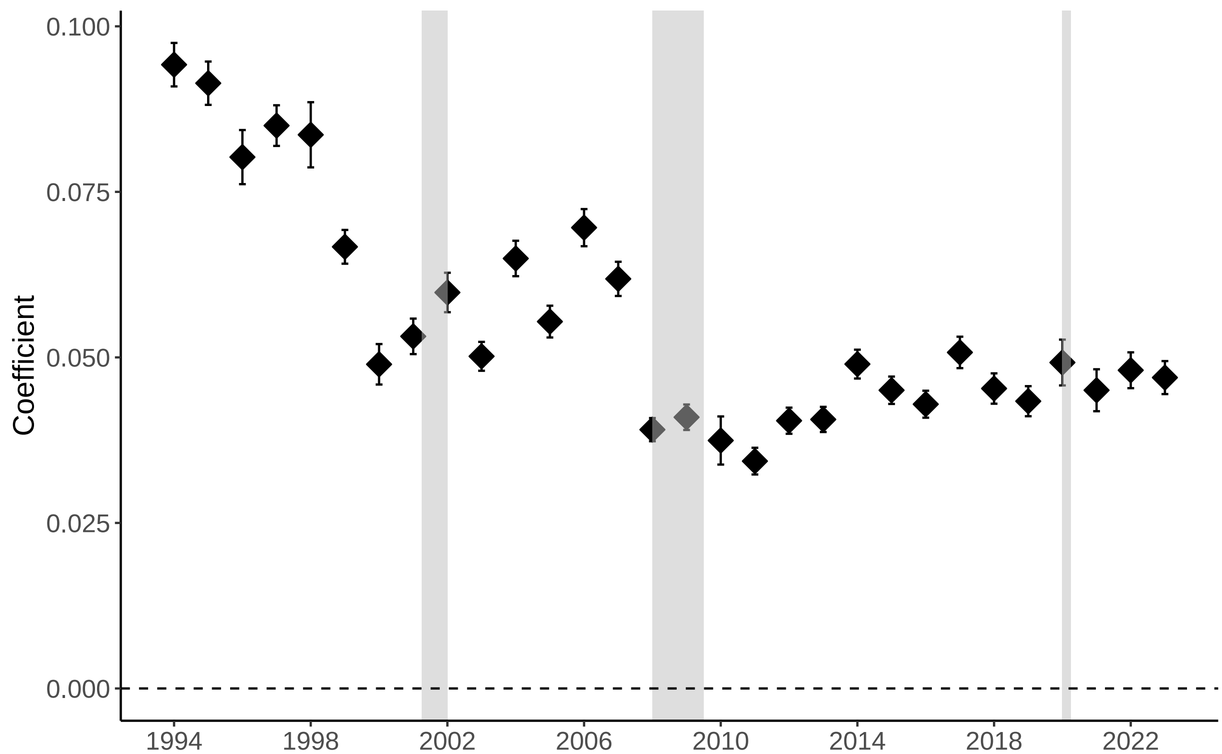
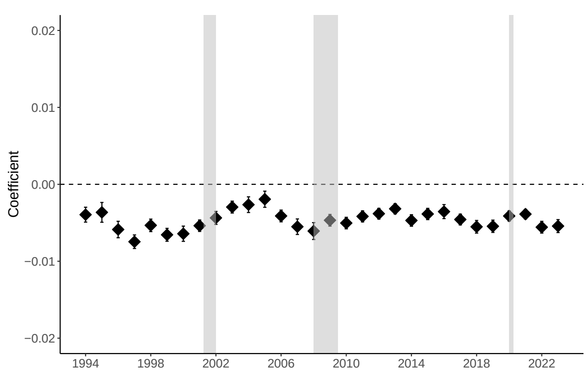


FIGURE 6: Evolution of Correlation Between Approval and Acceptance, and Distance

This Figure presents the results of estimating equation (5) year by year on our full sample of mortgages between 1994 and 2023. The outcome variable is either mortgage approval - an indicator variable that takes the value of 1 if the mortgage application is approved by the lender or zero if otherwise - or mortgage acceptance - an indicator variable that takes the value of 1 if the mortgage contract is then accepted by the borrower, or zero if otherwise (and is not defined for mortgages that where denied by the lender). We include county and lender-state fixed effects as well as the logarithm of the loan amount as a control. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Standard errors are clustered by county.

(a) Approval



(b) Acceptance

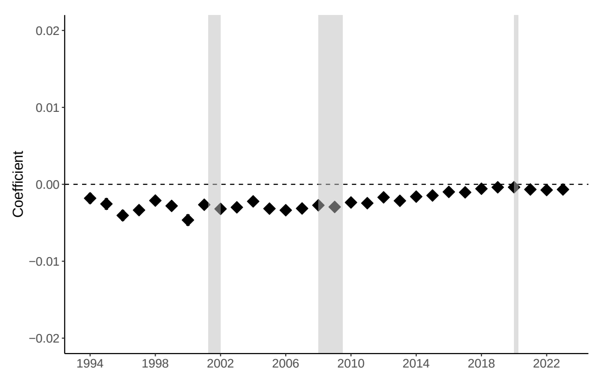
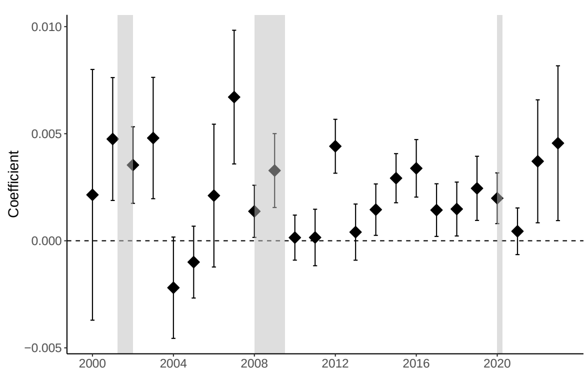


FIGURE 7: Evolution of Correlation Between Mortgage Characteristics and Distance

This Figure presents the results of estimating equation (6) year by year on the merged HMDA-GSE subsample of mortgages between 2000 and 2023 for Panel (a), and between 2018 and 2023 for Panel (b). The outcome variable is: (1) the interest rate spread, calculated as the interest rate relative to the prime mortgage rate reported in Freddie Mac's weekly Primary Mortgage Market Survey (PMMS), or (2) origination fees as a share of loan value. We include county-LLPA group and lender fixed effects as well as the logarithm of the loan amount as a control. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Errors are clustered by county. The two plots have different vertical axes.

(a) Spread



(b) Origination Fees

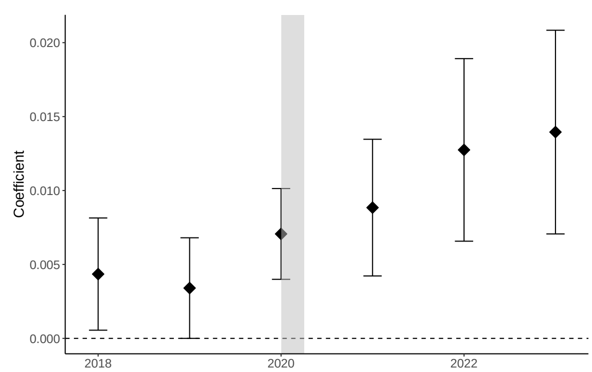


FIGURE 8: Evolution of Correlation Between Delinquency and Distance

This Figure presents the results of estimating equation (6) year by year on the merged HMDA-GSE subsample of mortgages between 2000 and 2023. The outcome variable is: (1) an indicator variable which takes the value of one if the borrower is delinquent at least once in the seven years following the issuance of the mortgage contract, and zero if otherwise, and (2) an indicator variable which takes the value of one if the borrower is delinquent at least once in the three years following the issuance of the mortgage contract, and zero if otherwise. We include county-LLPA group and lender fixed effects as well as the logarithm of the loan amount as a control. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Errors are clustered by county.

(a) Delinquency in 7 years (b) Delinquency in 3 years

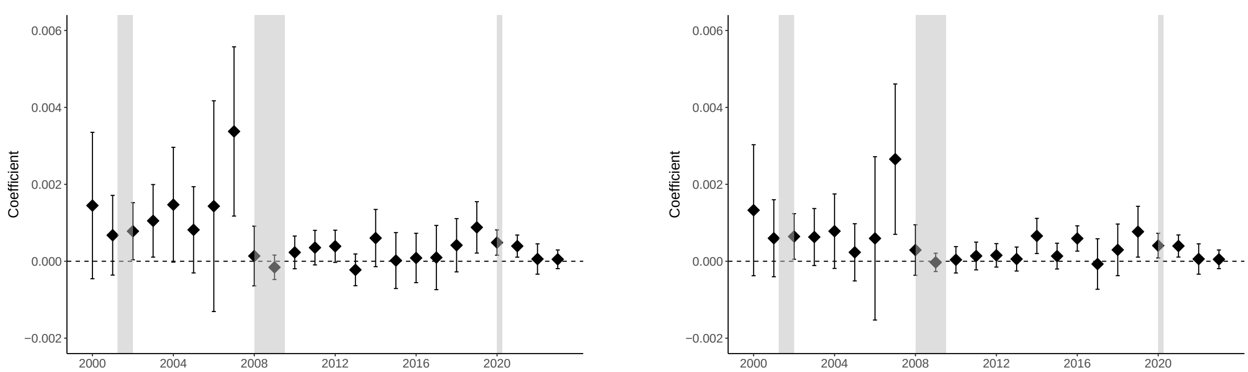


FIGURE 9: Optimal Interest Rate as Function of Distance

This Figure presents the optimal interest rate $R^{\star}(s, d)$, as defined in equation (7), for different values of the signal and for three different locations.

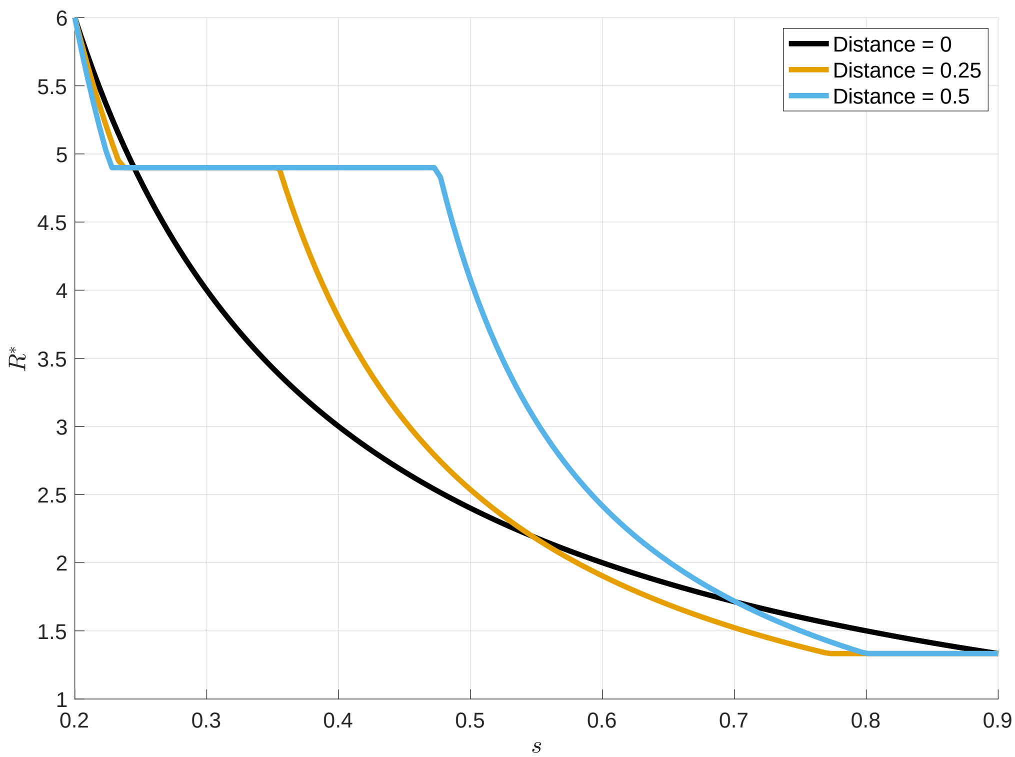
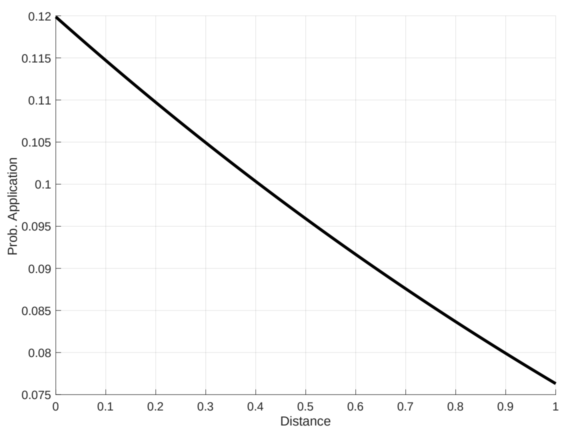


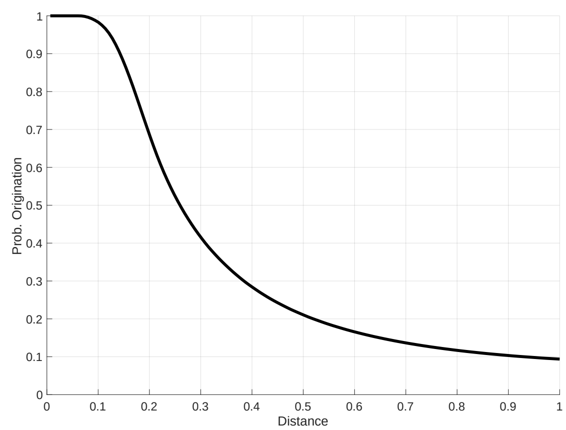
FIGURE 10: Cross-Sectional Implications of Model

This Figure presents the cross-sectional implications of the model. We plot the probability the lender receives an application $e^{-\alpha d}$, the probability of origination conditional on receiving an application $\mathcal{P}(d)$, and the optimal interest rate $R^{\star}(s,d)$ as functions of distance.

(a) Probability of receiving application



(b) Probability of origination



(c) Interest Rate

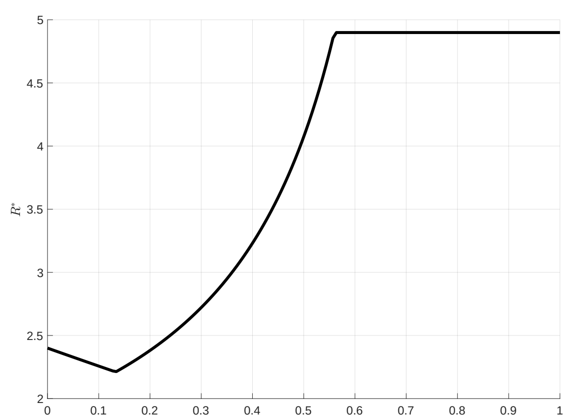
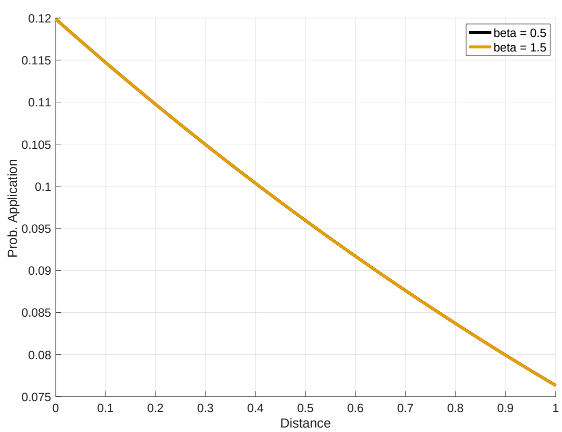


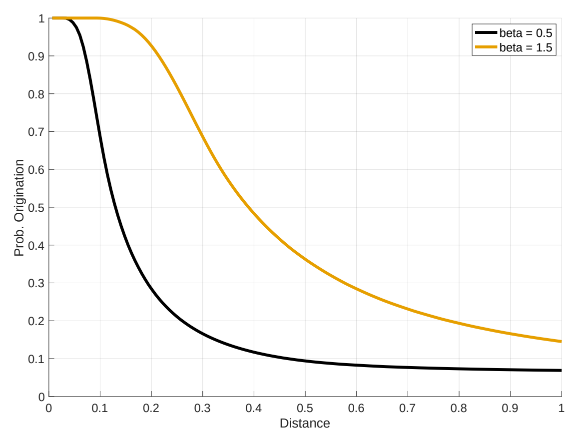
FIGURE 11: Time Series Implications of Improvements in Screening Technology

This Figure presents equilibia in our model for different values of β. We present four results: (1) the probability the lender receives an application as a function of distance, (2) the probability a mortgage is approved as a function of distance, (3) the elasticity of the optimal interest rate with respect to distance for different values of β, and (4) the average distance $\mathcal{D}$ for different values of β.

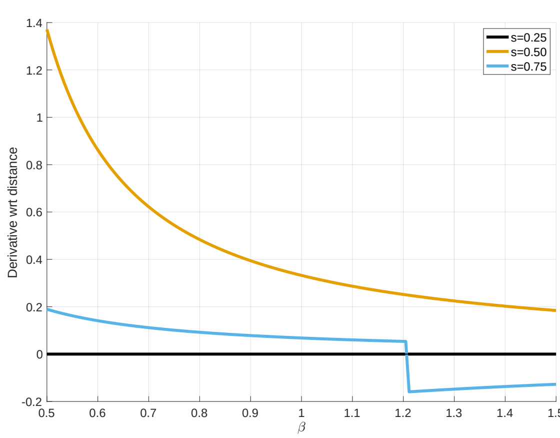
(a) Probability of receiving application



(b) Probability of origination



(c) Elasticity of Interest Rate to Distance



(d) Distance

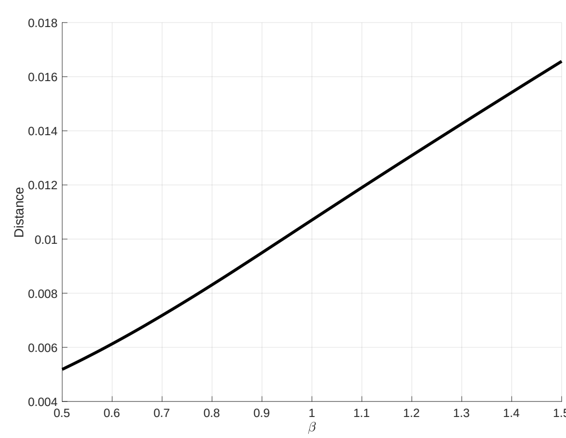
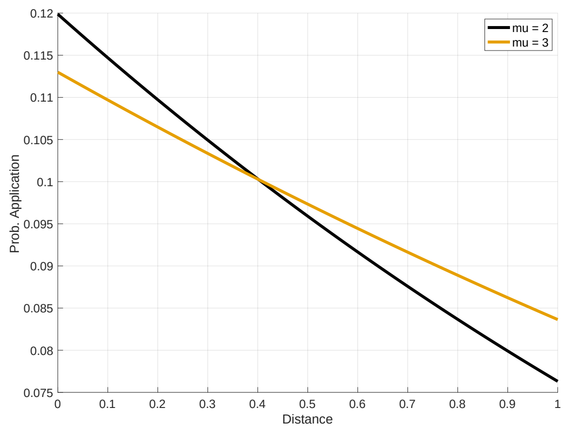


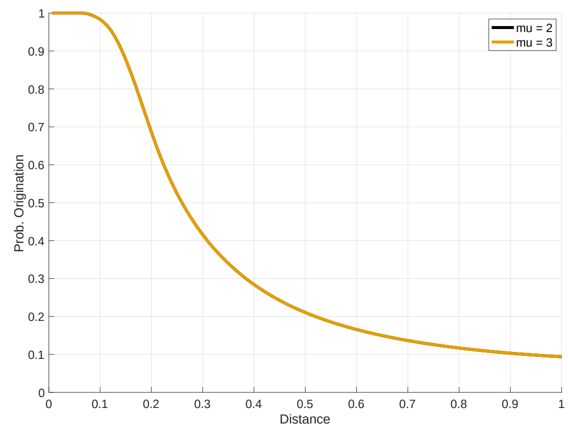
FIGURE 12: Time Series Implications of Lower Search Costs

This Figure presents equilibria in our model for different values of μ. We present four results: (1) the probability the lender receives an application as a function of distance, (2) the probability a mortgage is approved as a function of distance, (3) the elasticity of the optimal interest rate with respect to distance for different values of μ, and (4) the average distance $\mathcal{D}(b)$ for different values of μ.

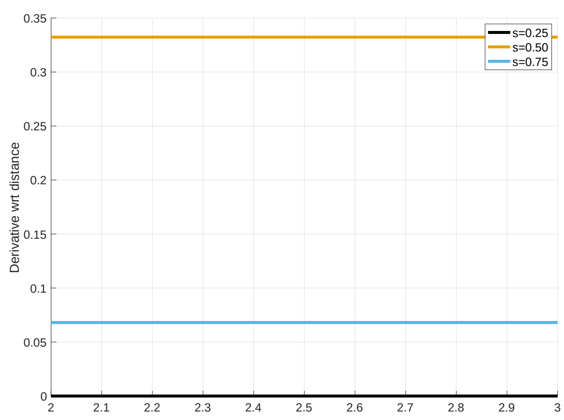
(a) Probability of receiving application



(b) Probability of origination



(c) Elasticity of Interest Rate to Distance



(d) Distance

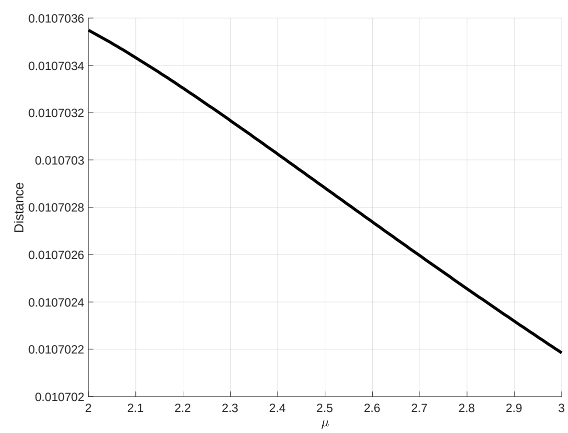


TABLE I: Summary Statistics

This Table presents summary statistics for 4,913,454 mortgages in our sample in 2023. We present statistics for the following variables: (1) distance in km between the borrower and the lender, (2) acceptance, which takes the value of 1 if the mortgage application is approved by the bank and accepted by the borrower, and 0 if otherwise, (3) the loan amount, in thousands of dollars, (4) the interest rate spread in percentage points, measured as the difference between the interest rate on the mortgage and the US 30-year treasury rate, (5) the applicant's annual income, measured in thousands of dollars, (6) the interest rate of the mortgage in percentage points, (7) loan costs in dollars, (8) discount points in dollars, (9) the value of the property in thousands of dollars, and (10) total points and fees in dollars.

	Mean	Median	25th Percentile	75th Percentile	Std. Dev.
Distance (km)	385.23	15.67	3.57	403.94	773.12
Accepted loans	0.72	1.00	0.00	1.00	0.45
Loan amount	321.41	205.00	95.00	365.00	1,657.89
Spread	0.53	0.39	-0.12	1.06	2.26
Applicant income	233.15	104.00	66.00	170.00	75,101.44
Interest rate	6.83	6.75	6.00	7.55	1.79
Loan costs	7,048.39	5,677.49	3,357.94	9,357.29	7,052.15
Discount points	4,293.07	2,835.19	1,187.03	5,668.98	20,054.64
Property value	648.25	385.00	255.00	625.00	4,604.76
Points and fees	1,412.90	987.75	325.75	1,836.00	1,786.22

TABLE II: Correlation Between Mortgage Characteristics and Distance

This Table presents the results of estimating equation (4) on the HMDA-GSE sub-sample of mortgages between 2018 and 2023. The outcome variable is: (1) the interest rate spread, calculated as the interest rate relative to the prime mortgage rate reported in Freddie Mac's weekly Primary Mortgage Market Survey (PMMS), (2) origination fees as a share of loan value, (3) an indicator variable which takes the value of one if the borrower is delinquent at least once in the seven years following the issuance of the mortgage contract, and zero if otherwise, and (4) an indicator variable which takes the value of one if the borrower is delinquent at least once in the three years following the issuance of the mortgage contract, and zero if otherwise. We include county-LLPA- group-year and lender-year fixed effects as well as the logarithm of the loan amount as a control. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We present the coefficients associated with the logarithm of the distance between the borrower and the lender. Standard errors are clustered by county. $***, **, *$ denote significance at a 1, 5, and 10 percent level, respectively.

	Spread	Origination Fees	Delinquency	Delinquency (3 years)
Log Distance	0.0018***	0.0078***	0.0004***	0.0004***
	(0.0016)	(0.0002)	(0.0001)	(0.0001)
Average	0.11 p.p.	0.66%	3.66%	3.44%
Lender-Year FE	✓	✓	✓	✓
County-LLPA-Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
R^2	0.46	0.31	0.13	0.13
N (million)	2.52	2.52	2.52	2.52

TABLE III: Correlation Between Prices and Distance - Role of Borrower Characteristics

This Table presents the results of estimating equation (4) on the HMDA-GSE sub-sample of mortgages between 2018 and 2023. The outcome variable is: (1) the interest rate spread, calculated as the interest rate relative to the prime mortgage rate reported in Freddie Mac's weekly Primary Mortgage Market Survey (PMMS), or (2) origination fees as a share of loan value. We include county-LLPA- group-year and lender-year fixed effects as well as the logarithm of the loan amount as a control. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We present the coefficients associated with the logarithm of the distance between the borrower and the lender. We split the sample according to the income of the applicant or the age of the applicant. For income, we divide the sample according to the median income for each state-year pair. For age, we divide the sample into applicants with age lower or equal to 34 and applicants older than 34. Standard errors are clustered by county. $***$, $**$, $*$ denote significance at a 1, 5, and 10 percent level, respectively.

Panel A: Interest Rate Spread

	Income of Applicant		Age of Applicant	
	Low	High	Younger than 34	Older than 34
Log Distance	0.0035^{***}	-0.0001	0.0020^{***}	0.0010^{*}
	(0.0006)	(0.0005)	(0.0006)	(0.0005)
Average	0.14 p.p.	0.08 p.p.	0.09 p.p.	0.12 p.p.
Lender-Year FE	✓	✓	✓	✓
County-LLPA-Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
R^2	0.48	0.47	0.47	0.47
N (million)	1.23	1.25	0.72	1.80

Panel B: Origination Fees as Share of Loan Value

	Income of Applicant		Age of Applicant	
	Low	High	Younger than 34	Older than 34
Log Distance	0.0126^{***}	0.0005	0.0070^{***}	0.0076^{***}
	(0.0020)	(0.0012)	(0.0015)	(0.0017)
Average	0.82%	0.51%	0.55%	0.70%
Lender-Year FE	✓	✓	✓	✓
County-LLPA-Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
R^2	0.32	0.29	0.32	0.32
N (million)	1.23	1.25	0.72	1.80

TABLE IV: Correlation Between Delinquency and Distance - Role of Borrower Characteristics

This Table presents the results of estimating equation (4) on the HMDA-GSE sub-sample of mortgages between 2018 and 2023. The outcome variable is an indicator variable which takes the value of one if the borrower is delinquent at least once in the seven years following the issuance of the mortgage contract, and zero if otherwise. We include county-LLPA- group-year and lender-year fixed effects as well as the logarithm of the loan amount as a control. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We present the coefficients associated with the logarithm of the distance between the borrower and the lender. We split the sample according to the income of the applicant or the age of the applicant. For income, we divide the sample according to the median income for each state-year pair. For age, we divide the sample into applicants with age lower or equal to 34 and applicants older than 34. Standard errors are clustered by county. $***, **, *$ denote significance at a 1, 5, and 10 percent level, respectively.

	Income of Applicant		Age of Applicant	
	Low	High	Younger than 34	Older than 34
Log Distance	0.0007***	0.0003***	0.0005***	0.0004***
	(0.0002)	(0.0001)	(0.0002)	(0.0001)
Average	4.37%	2.96%	3.57%	3.69%
Lender-Year FE	✓	✓	✓	✓
County-LLPA-Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
R^2	0.14	0.16	0.18	0.14
N (million)	1.23	1.25	0.72	1.80

Online Appendix

A Additional Details about Data, Tables and Figures

FIGURE A.1: Evolution of Median Distance and Federal Funds Rate

This Figure presents the median distance (in km) of mortgages for which we can compute the distance between the borrower and the lender between 1994 and 2023, as well as the average annual Federal Funds Rate. The grey bars represent recessions, as defined by the NBER. The left vertical axis is associated with distance, while the right vertical axis is associated with the Federal Funds Rate.

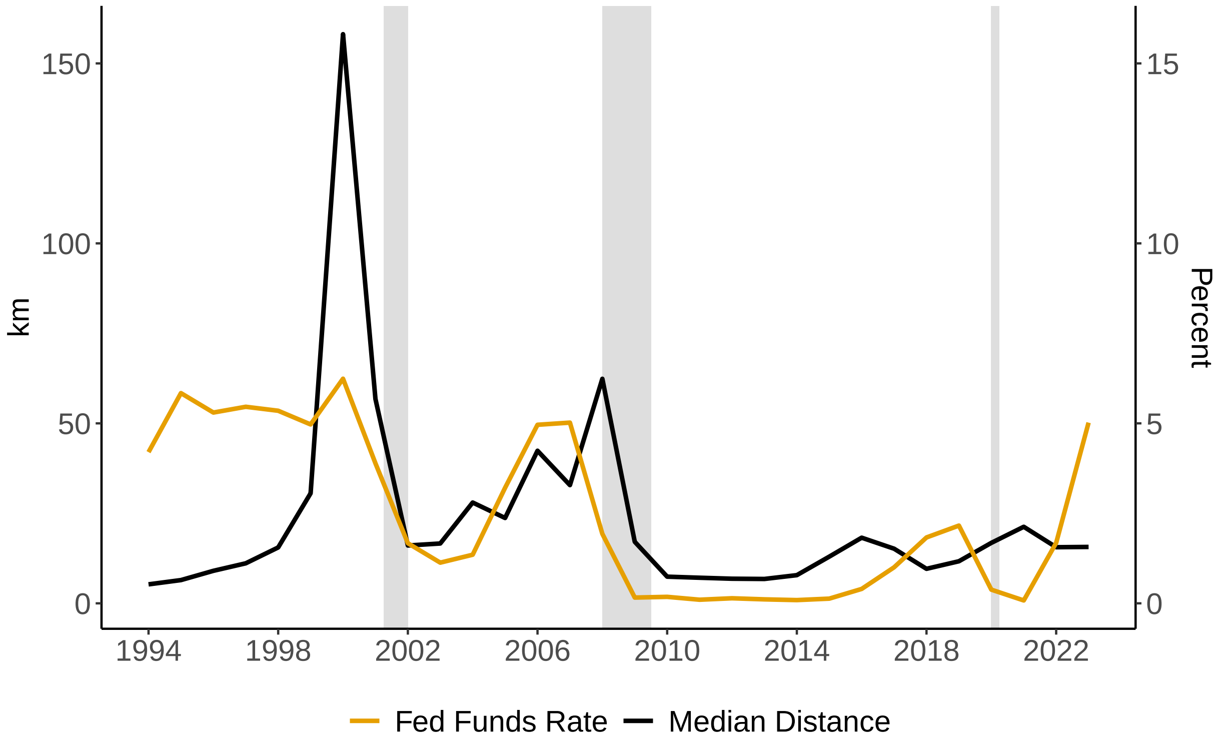


FIGURE A.2: Number of Mortgages

This Figure presents the number of mortgages for which we can compute the distance between the borrower and the lender between 2000 and 2023.

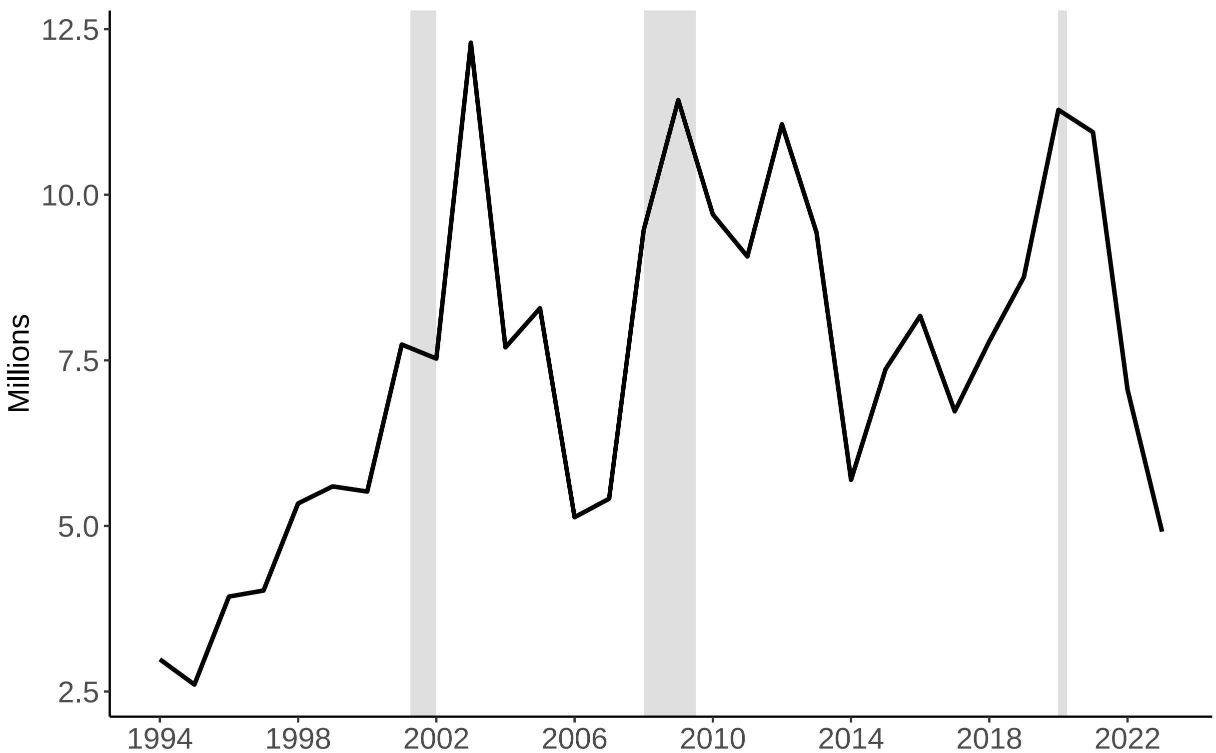


FIGURE A.3: Number of Lenders

This Figure presents the number of lenders for mortgages for which we can compute the distance between the borrower and the lender between 2000 and 2023.

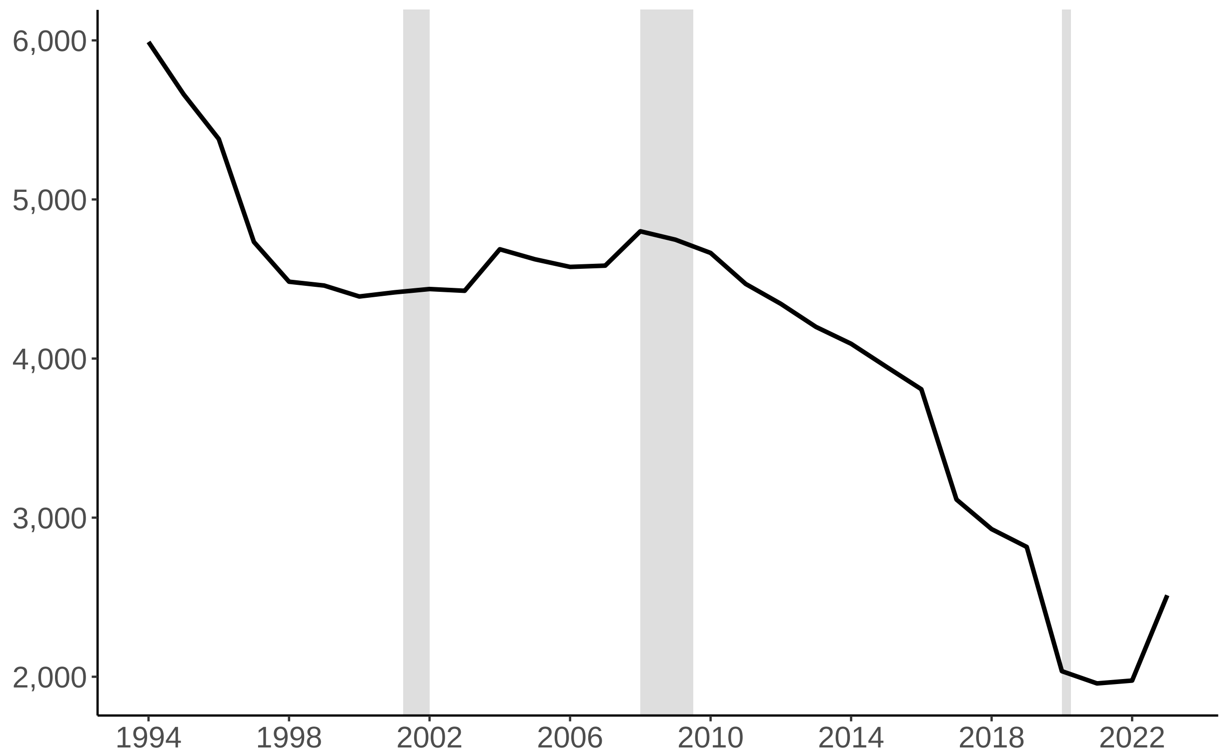
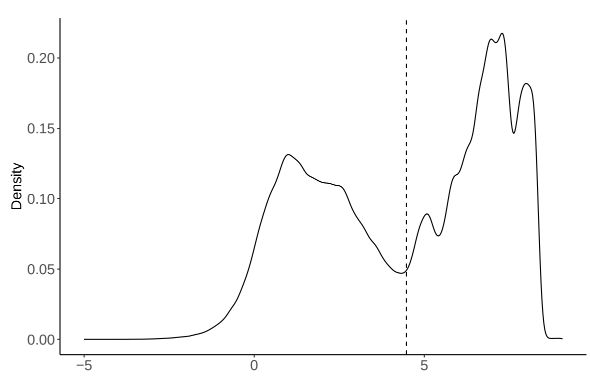


FIGURE A.4: Distribution of Distance

This Figure presents the distribution of the logarithm of the distance between the borrower and the lender for all mortgages in our sample.

(a) 2000



(b) 2023

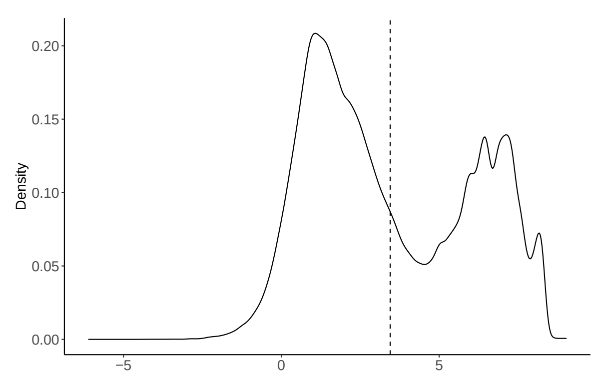


FIGURE A.5: Spatial Distribution of Distance

This Figure presents the spatial distribution of distance between borrower and lender in our sample of mortgages. For each year and county, we compute the average distance between borrower and lender. For each year, we then plot the distribution across counties of the logarithm of the average distance.

(a) 2000 (b) 2023

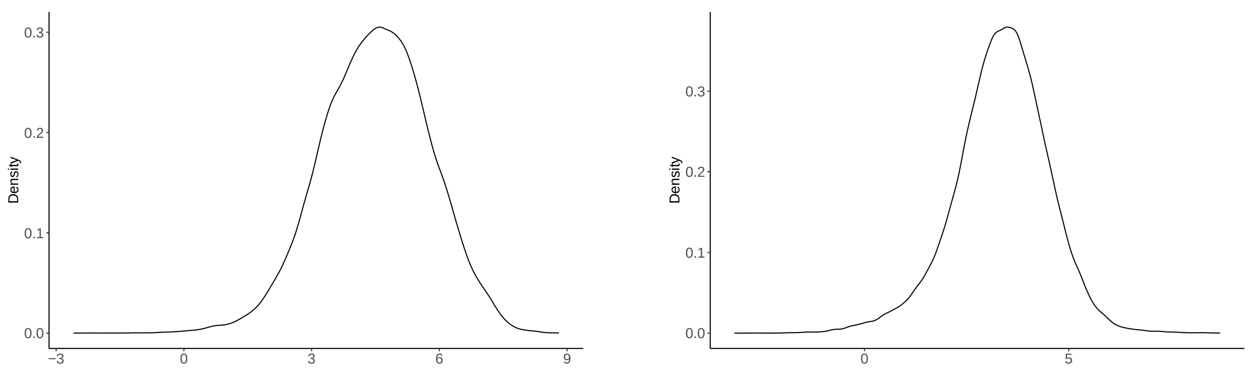
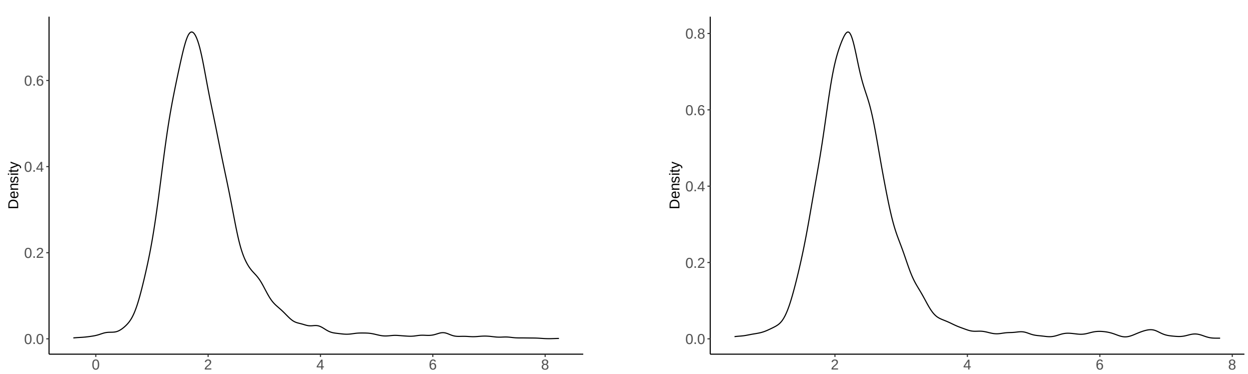


FIGURE A.6: Distribution of Distance Across Lenders

This Figure presents the distribution across lenders of distance between borrower and lender in our sample of mortgages. For each year and lender, we compute the average distance between borrower and lender. For each year, we then plot the distribution across lenders of the logarithm of the average distance.

(a) 2000 (b) 2023



B Additional Results

B.1 Market Share of Banks

FIGURE B.1: Share of Mortgages and Distance

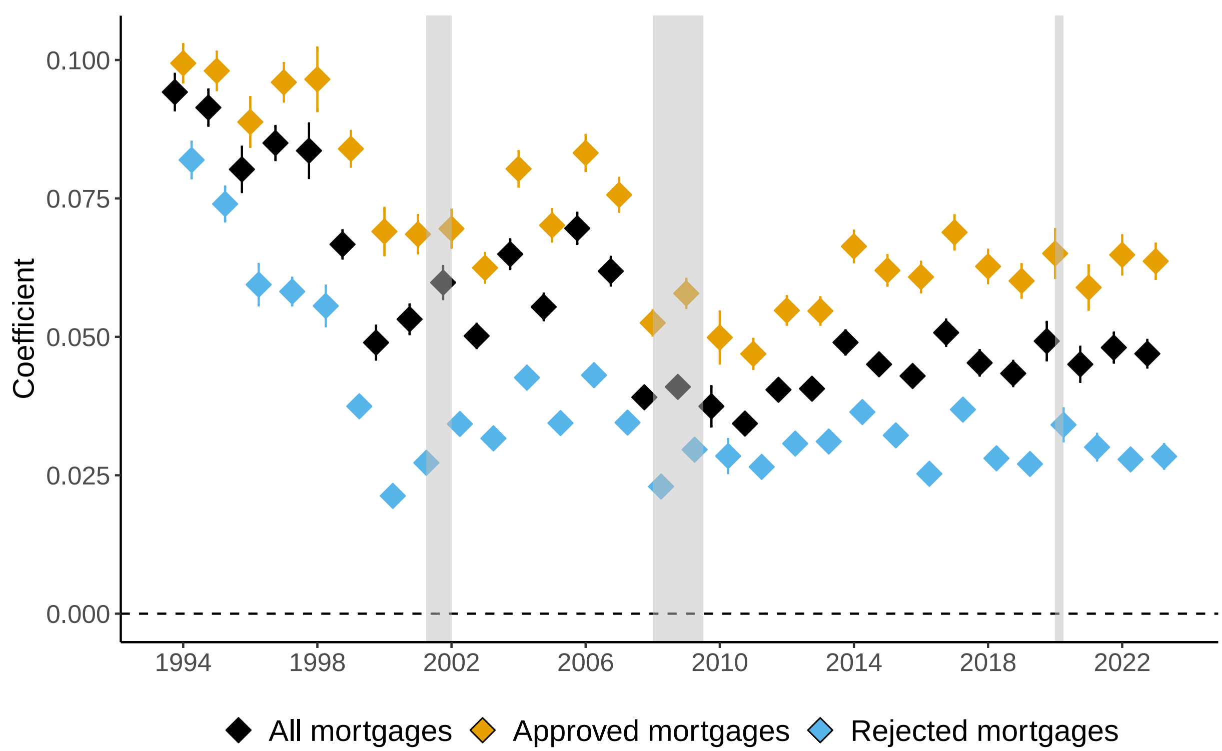
This Figure presents the results of estimating equation (2) for each year between 2000 and 2023. We consider three outcome variables: (1) the share of all mortgages in a given radius which are directed at a given lender, (2) the share of all approved mortgages in a given radius which are directed at a given lender, and (3) the share of all rejected mortgages in a given radius which are directed at a given lender. We include lender fixed effects. We consider ten radii with which we draw circles around bank branches: 5, 10, 25, 50, 100, 150, 200, 500, 1,000, and 2,000 km. We therefore end up with 11 distance bins, as we use the full sample as the eleventh bin. We include a series of fixed effects for the bins, using the eleventh bin ($> 2,000$ km) as the reference group. We plot the fixed effect associated with the the first distance bin ($\leq$ 5 km) for all years along with a 95 percent confidence interval. Errors are clustered by lender.

(a) 2000

(b) 2023

FIGURE B.2: Evolution of Local Market Share

This Figure presents the results of estimating equation (2) for each year between 2000 and 2023. We consider three outcome variables: (1) the share of all mortgages in a given radius which are directed at a given lender, (2) the share of all approved mortgages in a given radius which are directed at a given lender, and (3) the share of all rejected mortgages in a given radius which are directed at a given lender. We include lender fixed effects. We consider ten radii with which we draw circles around bank branches: 5, 10, 25, 50, 100, 150, 200, 500, 1,000, and 2,000 km. We therefore end up with 11 distance bins, as we use the full sample as the eleventh bin. We include a series of fixed effects for the bins, using the eleventh bin ($> 2,000$ km) as the reference group. We plot the fixed effect associated with the the first distance bin (≤ 5 km) for all years along with a 95 confidence interval. Errors are clustered by lender.

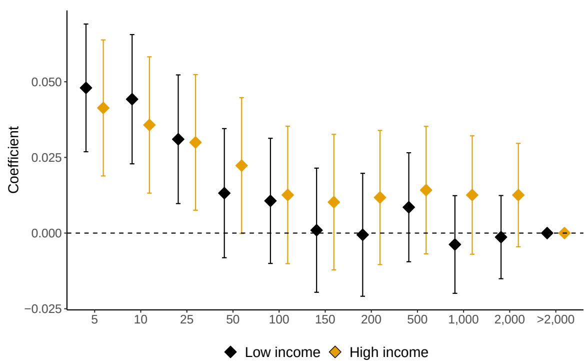


B.2 Mortgage Approval

FIGURE B.3: Mortgage Approval and Acceptance - Role of Borrower Income

This Figure presents the results of estimating equation (3) for 2000. The outcome variable is either mortgage approval - an indicator variable that takes the value of 1 if the mortgage application is approved by the lender or zero if otherwise - or mortgage acceptance - an indicator variable that takes the value of 1 if the mortgage contract is then accepted by the borrower, or zero if otherwise (and is not defined for mortgages that where denied by the lender). We include county and lender-state fixed effects as well as the logarithm of the loan amount as a control. We present coefficients associated with the distance bins - 5, 10, 25, 50, 100, 150, 200, 500, 1,000, and 2,000 km - using the eleventh bin ($> 2,000$ km) as the reference group. We split the sample according to the income of the applicant. For income, we divide the sample according to the median income for each state pair. We plot the fixed effects associated with the distance bins along with a 95% confidence interval. Errors are clustered by lender.

(a) Approval



(b) Acceptance

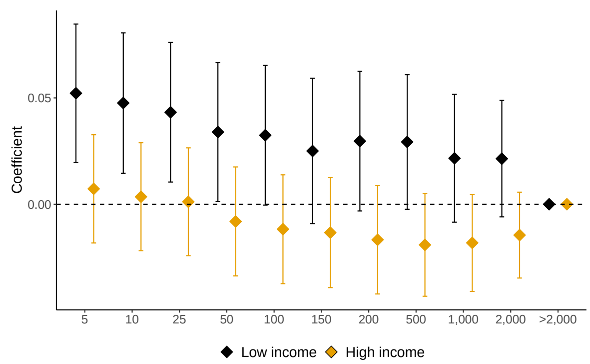


FIGURE B.4: Mortgage Approval and Acceptance - Role of Borrower Age

This Figure presents the results of estimating equation (3) for 2000. The outcome variable is either mortgage approval - an indicator variable that takes the value of 1 if the mortgage application is approved by the lender or zero if otherwise - or mortgage acceptance - an indicator variable that takes the value of 1 if the mortgage contract is then accepted by the borrower, or zero if otherwise (and is not defined for mortgages that where denied by the lender). We include county and lender-state fixed effects as well as the logarithm of the loan amount as a control. We present coefficients associated with the distance bins - 5, 10, 25, 50, 100, 150, 200, 500, 1,000, and 2,000 km - using the eleventh bin ($> 2,000$ km) as the reference group. We split the sample according to the age of the applicant. For age, we divide the sample into applicants with age lower or equal to 34 and applicants older than 34. We plot the fixed effects associated with the distance bins along with a 95% confidence interval. Errors are clustered by lender.

(a) Approval

(b) Acceptance

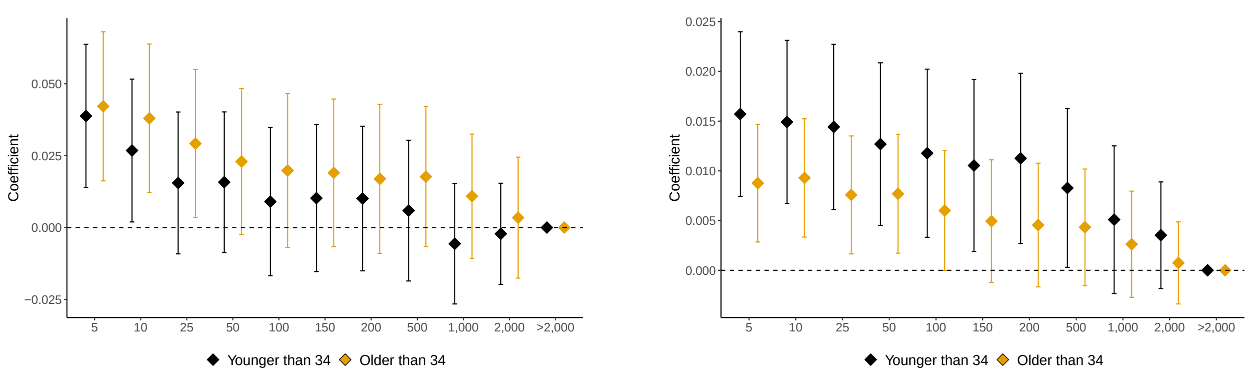


FIGURE B.5: Evolution of Correlation Between Mortgage Approval and Distance

This Figure presents the results of estimating equation (5) year by year on our full sample of mortgages between 2000 and 2023. The outcome variable is an indicator variable that takes the value of 1 if the mortgage is approved by the bank and accepted by the borrower and 0 if otherwise. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. We present results for five different specifications: (1) including no controls and no fixed effects, (2) including the logarithm of the loan amount as a control, (3) further including county fixed effects, (4) further including lender fixed effects, and (5) further including lender-state fixed effects. Errors are clustered by county.

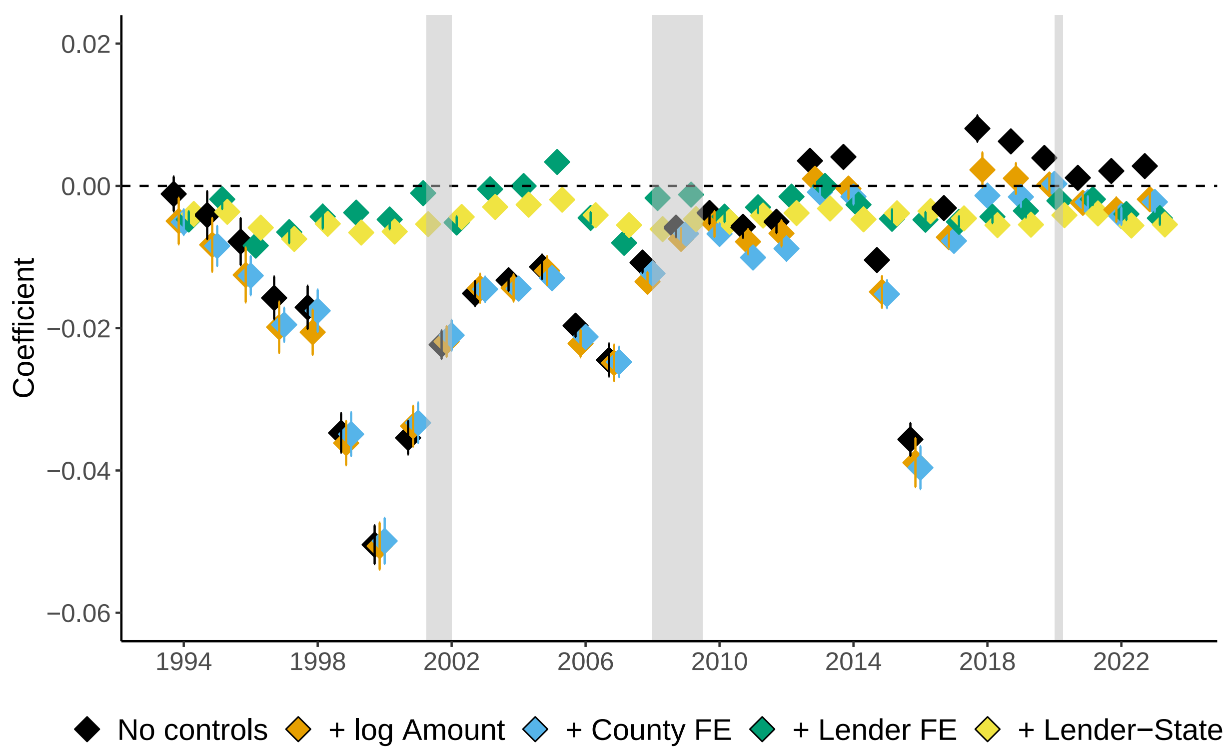


FIGURE B.6: Evolution of Correlation Between Mortgage Approval and Distance

This Figure presents the results of estimating equation (5) year by year on our full sample of mortgages between 2018 and 2023. The outcome variable is an indicator variable that takes the value of 1 if the mortgage is approved by the bank and accepted by the borrower and 0 if otherwise. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. We present results for five different specifications: (1) including county and lender-state fixed effects and the logarithm of the loan amount as a control (the baseline regression in Figure 6), (2) further including the logarithm of the applicant's income and the logarithm of the property value, and (3) further including debt-to-income fixed effects. Errors are clustered by county.

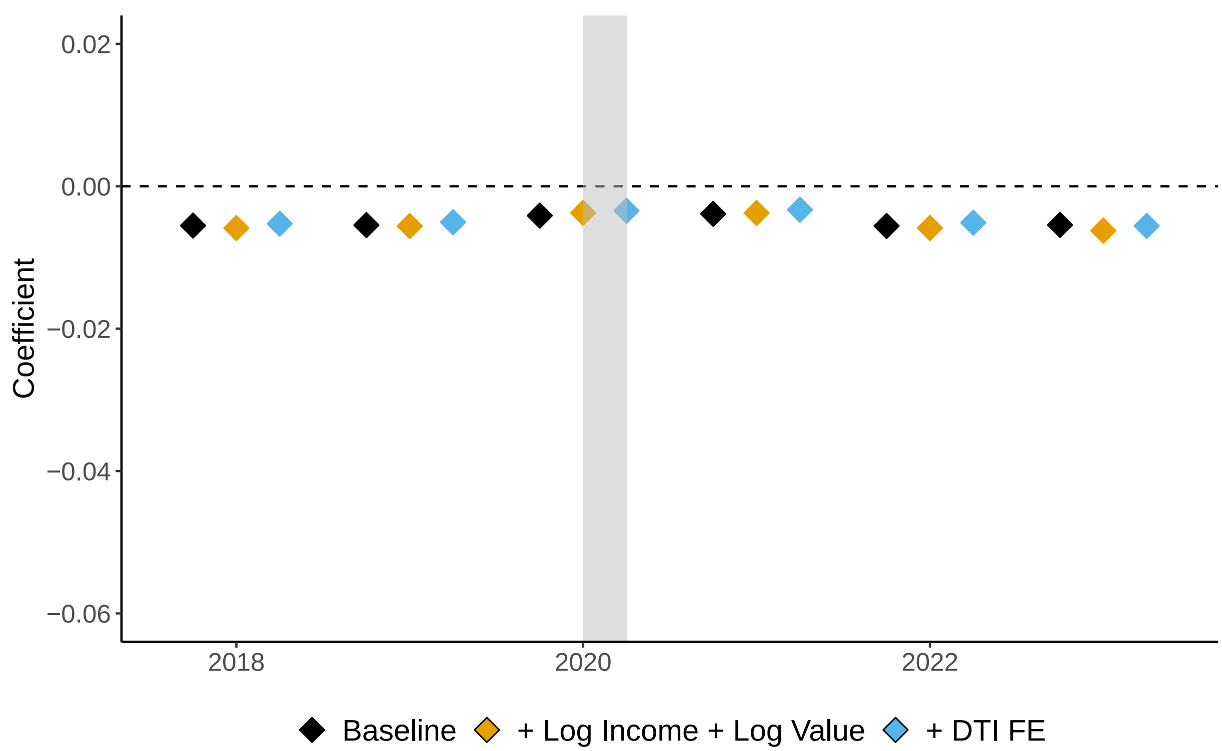
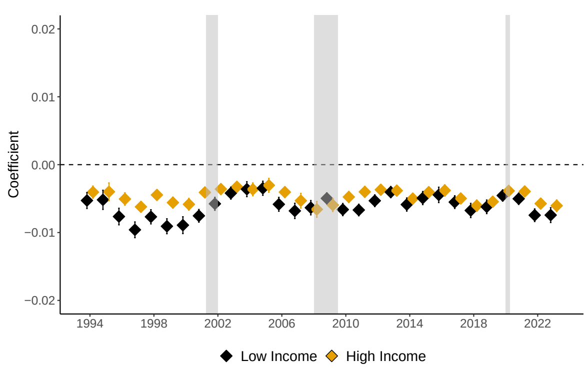


FIGURE B.7: Evolution of Correlation Between Mortgage Approval and Acceptance, and Distance - Role of Borrower Income

This Figure presents the results of estimating equation (5) year by year on our full sample of mortgages between 1994 and 2023. The outcome variable is either mortgage approval - an indicator variable that takes the value of 1 if the mortgage application is approved by the lender or zero if otherwise - or mortgage acceptance - an indicator variable that takes the value of 1 if the mortgage contract is then accepted by the borrower, or zero if otherwise (and is not defined for mortgages that where denied by the lender). We include county and lender-state fixed effects as well as the logarithm of the loan amount as a control. We split the sample according to the income of the applicant. For income, we divide the sample according to the median income for each state-year pair. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Standard errors are clustered by county.

(a) Approval



(b) Acceptance

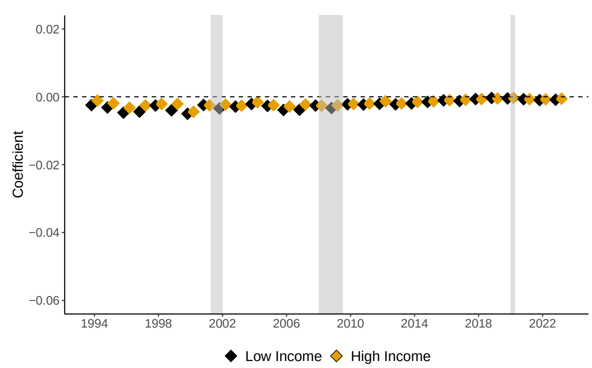
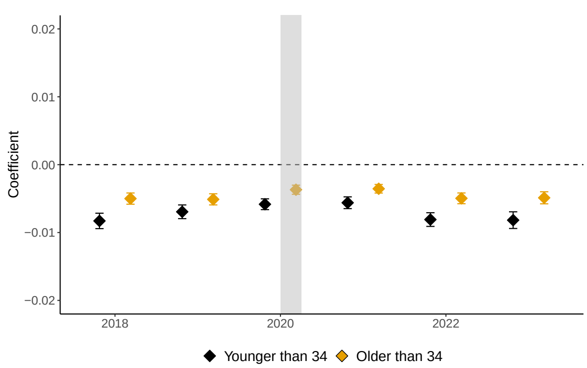


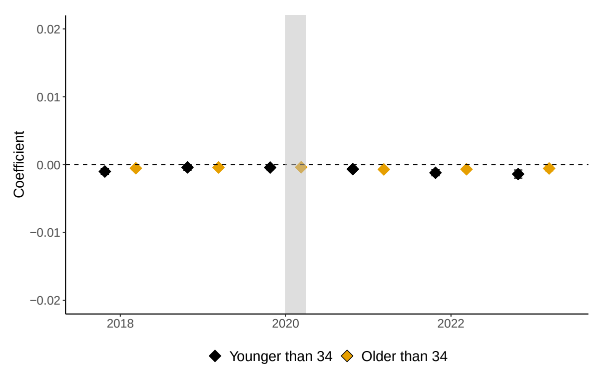
FIGURE B.8: Evolution of Correlation Between Mortgage Approval and Acceptance, and Distance - Role of Borrower Age

This Figure presents the results of estimating equation (5) year by year on our full sample of mortgages between 1994 and 2023. The outcome variable is either mortgage approval - an indicator variable that takes the value of 1 if the mortgage application is approved by the lender or zero if otherwise - or mortgage acceptance - an indicator variable that takes the value of 1 if the mortgage contract is then accepted by the borrower, or zero if otherwise (and is not defined for mortgages that where denied by the lender). We include county and lender-state fixed effects as well as the logarithm of the loan amount as a control. We split the sample according to the age of the applicant. For age, we divide the sample into applicants with age lower or equal to 34 and applicants older than 34. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Standard errors are clustered by county.

(a) Approval



(b) Acceptance



B.3 Interest Rate Spreads

FIGURE B.9: Evolution of Correlation Between Spread and Distance

This Figure presents the results of estimating equation (6) year by year on the merged HMDA-GSE subsample of mortgages between 2000 and 2023. The outcome variable is the interest rate spread, calculated as the interest rate relative to the prime mortgage rate reported in Freddie Mac's weekly Primary Mortgage Market Survey (PMMS). We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. We present results for five different specifications: (1) including no controls and no fixed effects, (2) including the logarithm of the loan amount as a control, (3) further including county fixed effects, (4) further including county-LLPA group fixed effects, and (5) further including lender fixed effects. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Errors are clustered by county.

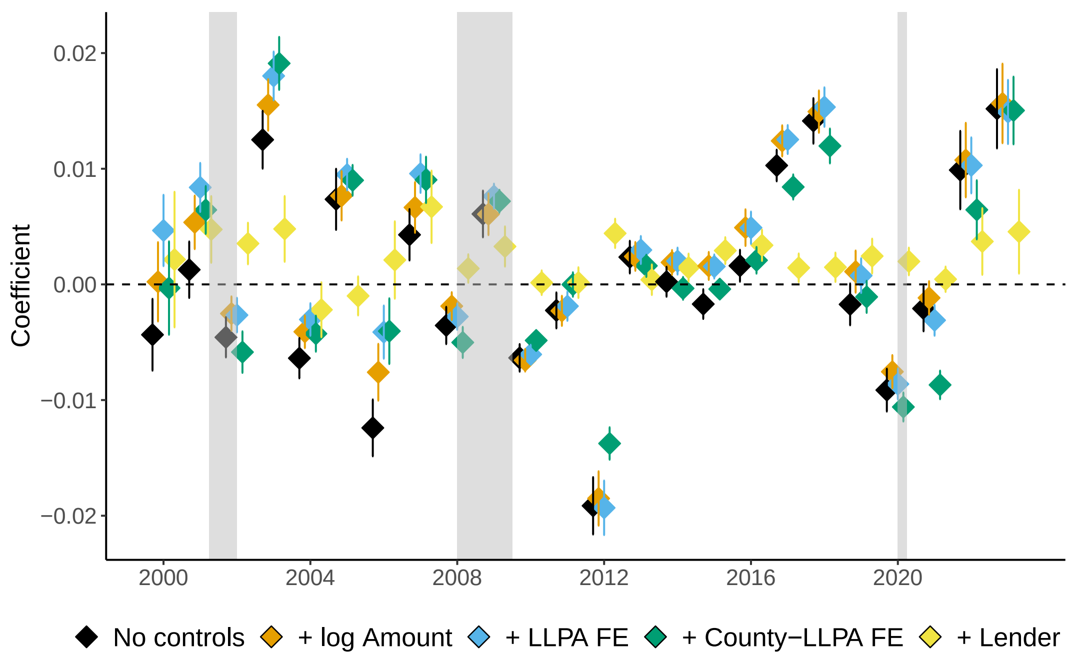
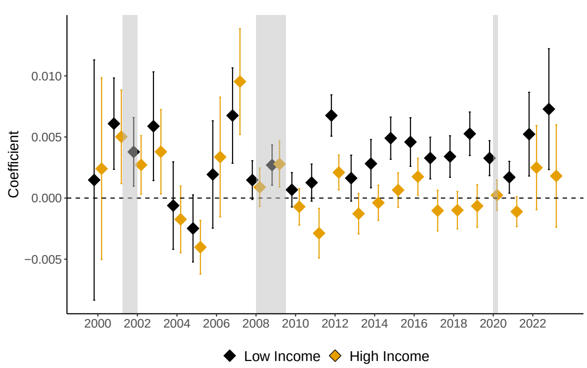


FIGURE B.10: Evolution of Correlation Between Mortgage Characteristics and Distance - Role of Borrower Income

This Figure presents the results of estimating equation (6) year by year on the merged HMDA-GSE subsample of mortgages between 2000 and 2023 for Panel (a), and between 2018 and 2023 for Panel (b). The outcome variable is: (1) the interest rate spread, calculated as the interest rate relative to the prime mortgage rate reported in Freddie Mac's weekly Primary Mortgage Market Survey (PMMS), or (2) origination fees as a share of loan value. We include county-LLPA group and lender fixed effects as well as the logarithm of the loan amount as a control. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We split the sample according to the income of the applicant. For income, we divide the sample according to the median income for each state-year pair. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Errors are clustered by county. The two plots have different vertical axes.

(a) Spread



(b) Origination Fees

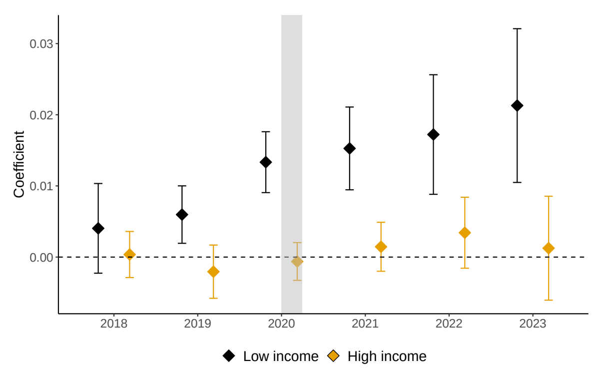
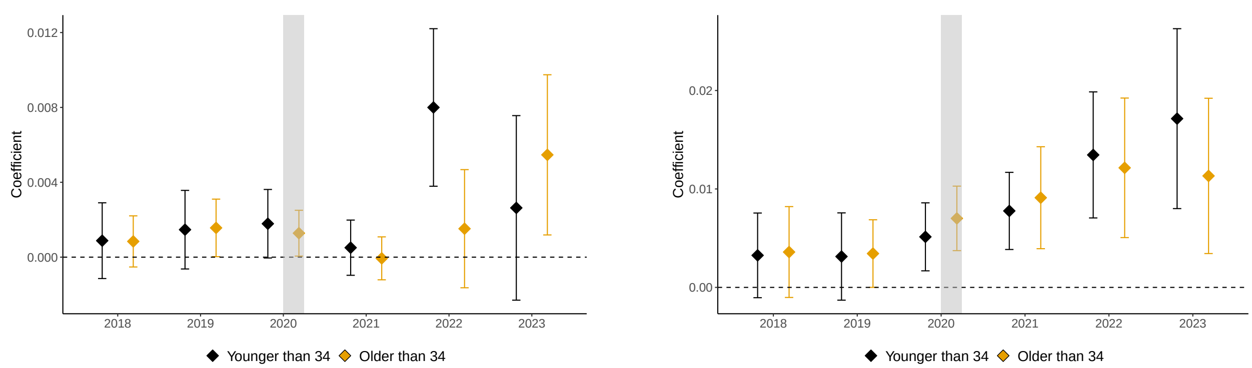


FIGURE B.11: Evolution of Correlation Between Mortgage Characteristics and Distance - Role of Borrower Age

This Figure presents the results of estimating equation (6) year by year on the merged HMDA-GSE subsample of mortgages between 2000 and 2023 for Panel (a), and between 2018 and 2023 for Panel (b). The outcome variable is: (1) the interest rate spread, calculated as the interest rate relative to the prime mortgage rate reported in Freddie Mac's weekly Primary Mortgage Market Survey (PMMS), or (2) origination fees as a share of loan value. We include county-LLPA group and lender fixed effects as well as the logarithm of the loan amount as a control. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We split the sample according to the age of the applicant. For age, we divide the sample into applicants with age lower or equal to 34 and applicants older than 34. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Errors are clustered by county. The two plots have different vertical axes.

(a) Spread

(b) Origination Fees



B.4 Delinquency

FIGURE B.12: Evolution of Correlation Between Delinquency and Distance

This Figure presents the results of estimating equation (6) year by year on the merged HMDA-GSE subsample of mortgages between 2000 and 2023. The outcome variable is an indicator variable which takes the value of one if the borrower is delinquent at least once in the seven years following the issuance of the mortgage contract, and zero if otherwise. We present results for five different specifications: (1) including no controls and no fixed effects, (2) including the logarithm of the loan amount as a control, (3) further including county fixed effects, (4) further including county-LLPA group fixed effects, and (5) further including lender fixed effects. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Errors are clustered by county.

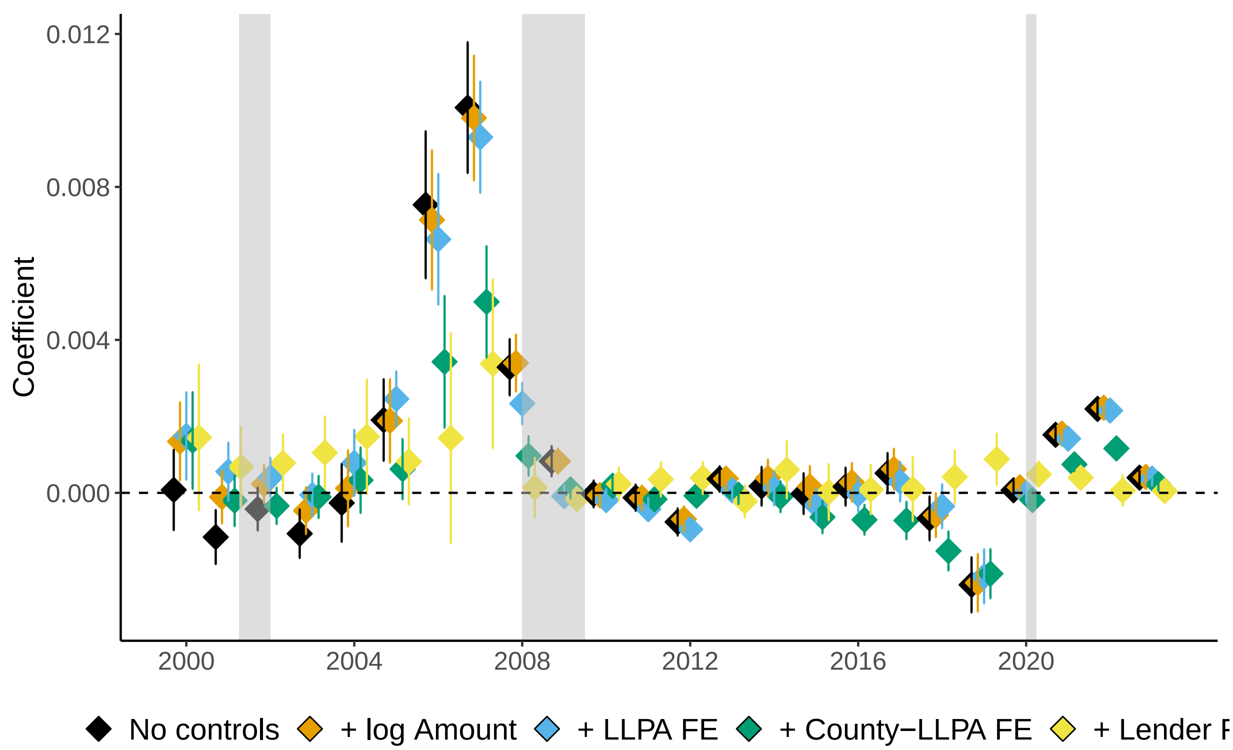
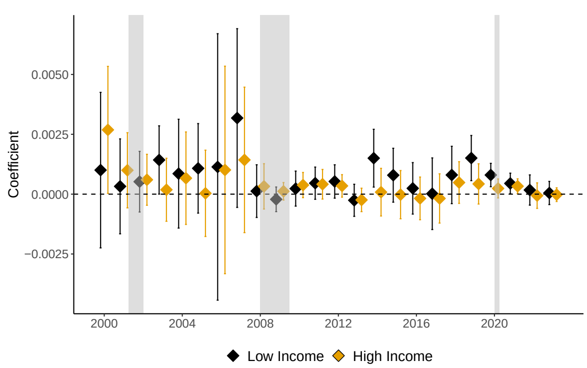


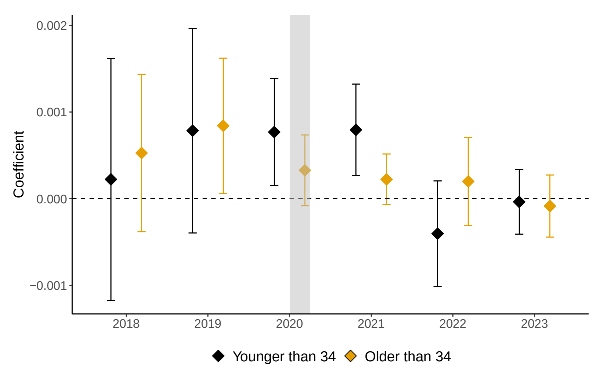
FIGURE B.13: Evolution of Correlation Between Delinquency and Distance - Role of Borrower Characteristics

This Figure presents the results of estimating equation (6) year by year on the merged HMDA-GSE subsample of mortgages between 2000 and 2023. The outcome variable is an indicator variable which takes the value of one if the borrower is delinquent at least once in the seven years following the issuance of the mortgage contract, and zero if otherwise. We include county-LLPA group and lender fixed effects as well as the logarithm of the loan amount as a control. The LLPA groups are constructed according to the Fannie Mae and Freddie Mac rules using the borrower's FICO score and the loan-to-value ratio. We split the sample according to the income of the applicant or the age of the applicant. For income, we divide the sample according to the median income for each state-year pair. For age, we divide the sample into applicants with age lower or equal to 34 and applicants older than 34. We present the coefficients associated with the logarithm of the distance between the borrower and the lender, as well as a 95 percent confidence interval. Errors are clustered by county. The two plots have different vertical axes.

(a) Income



(b) Age



C Model

C.1 Proofs

C.1.1 Proof of Proposition 1

The expected payoff of the lender is given by

$$\text{Payoff}(R, s, d) = \int_{\theta_0}^{\theta_1} [-1 + \theta R] \times \mathbf{1}\left\{\theta \leq \frac{\bar{u}}{R}\right\} \times f_{\theta|s;d}(\theta|s;d)d\theta$$

and so we need to compute the conditional distribution of θ and the integration bounds.

Conditional distribution. Our goal is to compute the conditional distribution of θ given s, which is given by

$$f_{\theta|s}(\theta|s;d) = \frac{f_{\theta,s}(\theta, s;d)}{f_s(s;d)}.$$

The first step is to compute the conditional distribution of θ, given s. The conditional distribution of s given θ is

$$f_{s|\theta}(s|\theta;s) = \begin{cases} \frac{1}{e^{-(1/\beta)d}(\theta_1-\theta_0)}, & s \in \left[e^{-(1/\beta)d}\theta + (1-e^{-(1/\beta)d})\theta_0, e^{-(1/\beta)d}\theta + (1-e^{-(1/\beta)d})\theta_1\right] \\ 0, & \text{otherwise.} \end{cases}$$

The joint distribution of the two random variables is given by

$$f_{\theta,s}(\theta, s;d) = f_\theta(\theta)f_{s|\theta}(s|\theta;d) = \frac{1}{e^{-(1/\beta)d}(\theta_1-\theta_0)^2},$$

which represents the first ingredient of the formula. The second ingredient is the distribution of s. This is relatively straightforward, as we only need to integrate $f_{\theta,s}(\theta, s;d)$ over all θ values that are consistent with s. Let $\mathcal{A}(s;d)$ be the set of θ values such that $s \in [e^{-(1/\beta)d}\theta + (1-e^{-(1/\beta)d})\theta_0, e^{-(1/\beta)d}\theta + (1-e^{-(1/\beta)d})\theta_1]$. Let's rewrite the support condition:

$$
\begin{aligned}
e^{-(1/\beta)d}\theta + (1-e^{-(1/\beta)d})\theta_0 \leq s \leq e^{-(1/\beta)d}\theta + (1-e^{-(1/\beta)d})\theta_1 \\
\Leftrightarrow (1-e^{-(1/\beta)d})\theta_0 - s \leq -e^{-(1/\beta)d}\theta \leq (1-e^{-(1/\beta)d})\theta_1 - s \\
\Leftrightarrow -(1-e^{-(1/\beta)d})\theta_0 + s \geq e^{-(1/\beta)d}\theta \geq -(1-e^{-(1/\beta)d})\theta_1 + s \\
\Leftrightarrow \frac{s-(1-e^{-(1/\beta)d})\theta_0}{e^{-(1/\beta)d}} \geq \theta \geq \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-d}} \\
\Leftrightarrow \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-(1/\beta)d}} \leq \theta \leq \frac{s-(1-e^{-(1/\beta)d})\theta_0}{e^{-(1/\beta)d}}.
\end{aligned}
$$

However, we need to make sure that this is consistent with the actual bounds for θ, which are known. Therefore, the bounds are given by

$$
\mathcal{A}(s;d) \equiv \left[\max\left(\theta_0, \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-(1/\beta)d}}\right), \min\left(\theta_1, \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-(1/\beta)d}}\right)\right].
$$

Given the expression for $\mathcal{A}(s;d)$, we can then compute the distribution of s as

$$
\begin{aligned}
f_s(s;d) &= \int_{\mathcal{A}(s;d)} f_{\theta,s}(\theta,s;d)d\theta \\
&= \int_{\mathcal{A}(s;d)} \frac{1}{e^{-(1/\beta)d}(\theta_1-\theta_0)^2}d\theta \\
&= \frac{1}{e^{-(1/\beta)d}(\theta_1-\theta_0)^2}\left\{\min\left(\theta_1, \frac{s-(1-e^{-(1/\beta)d})\theta_0}{e^{-(1/\beta)d}}\right) - \max\left(\theta_0, \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-(1/\beta)d}}\right)\right\}.
\end{aligned}
$$

Now, we can finally compute the conditional distribution, Using Bayes' rule, it follows that the conditional distribution of $\theta|s$ is given by

$$
\begin{aligned}
f_{\theta|s}(\theta|s;d) &= \frac{f_{\theta,s}(\theta,s;d)}{f_s(s;d)} \\
&= \frac{1}{\min\left(\theta_1, \frac{s-(1-e^{-(1/\beta)d})\theta_0}{e^{-(1/\beta)d}}\right) - \max\left(\theta_0, \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-(1/\beta)d}}\right)},
\end{aligned}
$$

which implies that the conditional distribution of θ given s is uniformly distributed over $\mathcal{A}(s;d)$. Define

$$\theta_{\min}(s,d) = \max\left\{\theta_0, \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-(1/\beta)d}}\right\},$$
$$\theta_{\max}(s,d) = \min\left\{\theta_1, \frac{s-(1-e^{-(1/\beta)d})\theta_0}{e^{-(1/\beta)d}}\right\}.$$

Payoff of the lender. Define $\theta_R \equiv \bar{u}/R$. The borrower will only accept if $\theta \leq \theta_R$. We can now compute the expected payoff. Note that there are two regions: (1) the acceptance region from $\theta_{\min}(s,d)$ to θ_R, and (2) the rejection region from θ_R to $\theta_{\max}(s,d)$. In the end, the lenderswill choose something inside the interval $[\theta_{\min}(s,d), \theta_{\max}(s,d)]$. To see this, suppose she sets $\theta_R < \theta_{\min}(s,d)$. No one will accept the loan so the payoff is zero. Suppose she sets $\theta_R > \theta_{\max}(s,d)$. Then, everyone will accept the loan. However, if the increases interest rates by a small share such that θ_R is still above the limit, everyone will still accept the loan but profits will increase. Define $\theta_{\text{upper}}(s,d) \equiv \min\{\theta_R, \theta_{\max}(s,d)\}$ The payoff of the lender is

$$\begin{aligned}
\text{Payoff}(R,s,d) &= \int_{\theta_{\min}(s,d)}^{\theta_{\text{upper}}(s,d)} [-1+\theta R] \times f_{\theta|s;d}(\theta|s;d)d\theta + \int_{\theta_{\text{upper}}(s,d)}^{\theta_{\max}(s,d)} 0 \times f_{\theta|s;d}(\theta|s;d)d\theta \\
&= \frac{1}{\theta_{\max}(s,d)-\theta_{\min}(s,d)} \int_{\theta_{\min}(s,d)}^{\theta_{\text{upper}}(s,d)} (-1+\theta R)d\theta \\
&= -\frac{\theta_{\text{upper}}(s,d)-\theta_{\min}(s,d)}{\theta_{\max}(s,d)-\theta_{\min}(s,d)} + \frac{\theta^2_{\text{upper}}(s,d)-\theta^2_{\min}(s,d)}{2(\theta_{\max}(s,d)-\theta_{\min}(s,d))}R
\end{aligned}$$

Suppose first that $\theta_{\max}(s,d) \leq \bar{u}/R$. Equivalently, we assume $R < \bar{u}/\theta_{\max}(s,d)$. In this branch, it follows that $\theta_{\text{upper}}(s,d) = \theta_{\max}(s,d)$. The function becomes

$$-1 + \frac{\theta^2_{\max}(s,d)-\theta^2_{\min}(s,d)}{2(\theta_{\max}(s,d)-\theta_{\min}(s,d))}R = -1 + \frac{\theta_{\max}(s,d)+\theta_{\min}(s,d)}{2}R,$$

which is a linear function with a positive slope. Therefore, the maximum in this branch is given at $R_1 = \bar{u}/\theta_{\max}(s,d)$.

Alternatively, suppose that $\theta_{\max}(s,d) > \bar{u}/R$. Equivalently, we assume $R \geq \bar{u}/\theta_{\max}(s,d)$.

In this case, the function becomes trickier

$$\begin{aligned}
\text{Payoff}(R,s,d) &= -\frac{\frac{\bar{u}}{R} - \theta_{\min}(s,d)}{\theta_{\max}(s,d) - \theta_{\min}(s,d)} + \frac{\frac{\bar{u}^2}{R^2} - \theta^2_{\min}(s,d)}{2(\theta_{\max}(s,d) - \theta_{\min}(s,d))} R \\
&= \frac{\theta_{\min}(s,d)}{\theta_{\max}(s,d) - \theta_{\min}(s,d)} - \frac{\bar{u}(1-0.5\bar{u})}{\theta_{\max}(s,d) - \theta_{\min}(s,d)} \frac{1}{R} - \frac{\theta^2_{\min}(s,d)}{2(\theta_{\max}(s,d) - \theta_{\min}(s,d))} R
\end{aligned}$$

which is basically a quadratic function. The first derivative of the function is

$$\frac{d\text{Payoff}(R,s,d)}{dR} = -\frac{\theta^2_{\min}(s,d)}{2(\theta_{\max}(s,d) - \theta_{\min}(s,d))} + \frac{\bar{u}(1-0.5\bar{u})}{\theta_{\max}(s,d) - \theta_{\min}(s,d)} \frac{1}{R^2}$$

which is equal to zero when

$$\begin{aligned}
&\Leftrightarrow \frac{\theta^2_{\min}(s,d)}{2(\theta_{\max}(s,d) - \theta_{\min}(s,d))} = \frac{\bar{u}(1-0.5\bar{u})}{\theta_{\max}(s,d) - \theta_{\min}(s,d)} \frac{1}{R^2} \\
&\Leftrightarrow \frac{\theta^2_{\min}(s,d)}{2} = \bar{u}(1-0.5\bar{u}) \frac{1}{R^2} \\
&\Leftrightarrow \frac{\theta^2_{\min}(s,d)}{2\bar{u}(1-0.5\bar{u})} = \frac{1}{R^2} \\
&\Leftrightarrow \frac{2\bar{u}(1-0.5\bar{u})}{\theta^2_{\min}(s,d)} = R^2 \\
&\Leftrightarrow R_2 = \frac{\sqrt{2\bar{u}(1-0.5\bar{u})}}{\theta_{\min}(s,d)}.
\end{aligned}$$

The second derivative is

$$\frac{d^2\text{Payoff}(R,s,d)}{dR^2} = -2\frac{\bar{u}(1-0.5\bar{u})}{\theta_{\max}(s,d) - \theta_{\min}(s,d)} \frac{1}{R^3} < 0$$

and so R_2 is a local maximum. It may happen that $R_2 > R_1$. In this case, the function is increasing after R_1, so the payoff is maximized at R_2. It may also happen that $R_1 > R_2$, and so the optimum is R_1. Therefore, the maximum interest rate is

$$R^\star(s,d) = \max\left\{\frac{\bar{u}}{\theta_{\max}(s,d)}, \frac{\sqrt{2\bar{u}(1-0.5\bar{u})}}{\theta_{\min}(s,d)}\right\}$$

which concludes our proof.

C.1.2 Density function of signal

The density of the signal is given by

$$f_s(s;d) = \frac{\theta_{\max}(s,d) - \theta_{\min}(s,d)}{e^{-(1/\beta)d}(\theta_1 - \theta_0)^2}$$

for $s \in [\theta_0, \theta_1]$, and zero otherwise. However, this may not integrate to one. Therefore, we need to scale the distribution by a constant, which we define as

$$\Omega(d) \equiv \int_{\theta_0}^{\theta_1} \frac{\theta_{\max}(s,d) - \theta_{\min}(s,d)}{e^{-(1/\beta)d}(\theta_1 - \theta_0)^2} ds$$

and so the distribution of the signal is given by

$$f_s(s;d) = \begin{cases} \frac{1}{\Omega(d)} \frac{\theta_{\max}(s,d) - \theta_{\min}(s,d)}{e^{-(1/\beta)d}(\theta_1 - \theta_0)^2}, & \text{if } s \in [\theta_0, \theta_1] \\ 0, \text{ if otherwise.} \end{cases} \tag{C.1}$$

C.2 Additional Results

C.2.1 Elasticity of the optimal interest rate with respect to distance

Proposition 2 *The elasticity of the optimal interest rate with respect to distance is given by*

$$\frac{d \log R^\star(s,d)}{d \log d} = \begin{cases} 0, & \textit{if } R_1(s,d) > R_2(s,d) \textit{ and } \theta_1 < G_0(s,d) \\ -\frac{d}{\beta}\left(1 - \frac{\theta_0}{G_0(s,d)}\right), & \textit{if } R_1(s,d) > R_2(s,d) \textit{ and } \theta_1 > G_0(s,d) \\ 0, & \textit{if } R_1(s,d) < R_2(s,d) \textit{ and } \theta_0 > G_1(s,d) \\ -\frac{d}{\beta}\left(1 - \frac{\theta_1}{G_1(s,d)}\right), & \textit{if } R_1(s,d) < R_2(s,d) \textit{ and } \theta_0 < G_1(s,d) \\ \textit{undefined}, & \textit{otherwise,} \end{cases}$$

where

$$\begin{aligned}
R_1(s,d) &\equiv \frac{\bar{u}}{\theta_{max}(s,d)},\\
R_2(s,d) &\equiv \frac{\sqrt{2\bar{u}(1-0.5\bar{u})}}{\theta_{min}(s,d)},\\
G_0(s,d) &\equiv \frac{s-(1-e^{-(1/\beta)d})\theta_0}{e^{-(1/\beta)d}},\\
G_1(s,d) &\equiv \frac{s-(1-e^{-(1/\beta)d})\theta_1}{e^{-(1/\beta)d}}.
\end{aligned}$$

Proof. The optimal interest rate is given by

$$R^\star(s,d) = \max\{R_1(s,d), R_2(s,d)\}$$

and

$$\begin{aligned}
\theta_{\min}(s,d) &= \max\{\theta_0, G_1(s,d)\}\\
\theta_{\max}(s,d) &= \min\{\theta_1, G_0(s,d)\}
\end{aligned}$$

Suppose first that $R^\star(s,d) = R_1(s,d)$. Then, we want to compute

$$\frac{d\log R^\star(s,d)}{d\log d} = \frac{d\log R_1(s,d)}{d\log d} = -\frac{d\log\theta_{\max}(s,d)}{d\log d}$$

If $\theta_{\max}(s,d) = \theta_1$, the derivative is zero. If $\theta_{\max}(s,d) \neq \theta_1$, then

$$\frac{d\log G_0(s,d)}{d\log d} = \frac{d}{\beta}\left(1-\frac{\theta_0}{G_0(s,d)}\right).$$

Therefore, we have that

$$\frac{d\log R_1(s,d)}{d\log d} = \begin{cases} 0, & \text{if } \theta_1 < G_0(s,d) \\ -\frac{d}{\beta}\left(1-\frac{\theta_0}{G_0(s,d)}\right), & \text{if } \theta_1 > G_0(s,d) \end{cases}$$

The second case where $R^\star(s,d) = R_2(s,d)$ is identical and so

$$\frac{d\log R_2(s,d)}{d\log d} = \begin{cases} 0, & \text{if } \theta_0 > G_1(s,d) \\ -\frac{d}{\beta}\left(1-\frac{\theta_1}{G_1(s,d)}\right), & \text{if } \theta_0 < G_!(s,d). \end{cases}$$

Putting everything together, we have that

$$\frac{d \log R^{\star}(s,d)}{d \log d} = \begin{cases} \frac{d \log R_1(s,d)}{d \log d}, & \text{if } R_1(s,d) > R_2(s,d) \\ \frac{d \log R_2(s,d)}{d \log d}, & \text{if } R_1(s,d) < R_2(s,d). \end{cases}$$

As the function is discontinuous at the kinks, the derivative is not well defined. This concludes the proof. ■