THE UNIVERSITY OF CHICAGO

# Age and interest rates: Evidence from home mortgages in Illinois

By

Zhendong Chen

May 2023

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Pablo Peña
Preceptor: Pablo Peña

# Age and interest rates: Evidence from home mortgages in Illinois

Zhendong Chen*

May, 2023

**Abstract**

Using publicly available data from the Home Mortgage Disclosure Act, in this study I analyze the relationship between the age of the borrower and the interest rate paid on over 2 million loans in Illinois between 2018 and 2021. I estimate a hedonic price model using an econometric specification with fixed effects for age (in 10-year intervals), county, and other borrower or loan characteristics. I interpret differences in estimates across age groups as evidence of statistical discrimination. Additionally, I provide a simple theoretical model to explain why in equilibrium the relationship between age and interest rate may be positive, negative, or non-monotone. Last, I compare estimates across urban and rural counties to measure the interaction with age. The theoretical model sheds light on how to interpret urban versus rural differences.

**Keywords**: Mortgage lending, Interest rates, Geographic disparities, Statistical discrimination

*E-mail:zhendong1@uchicago.edu I am sincerely grateful to my advisor Professor Pablo Peña at the University of Chicago, for his invaluable advice and feedback. I am also truly thankful for Professor Stephen Ross at the University of Connecticut, for his guidance on developing my research interests.

# 1 Introduction

Studies have predominantly examined age discrimination in mortgage lending, focusing primarily on whether older people receive higher interest rates than young people. A major challenge in studying age discrimination in mortgage lending involves developing an appropriate test to identify the potential discrepancies in the mortgage rate pattern and addressing the lack of reliable recent data on loan applications. Although many studies have emphasized taste-based discrimination in mortgage lending, statistical discrimination is a more prevalent type of discrimination, given the assumption that the market is efficient. Compared to young people, lenders may perceive older people as having extra risks because of factors such as a higher mortality rate and being less favored by labor markets. Consequently, older people are more likely to be charged differently than young people on mortgage interest rates and similar borrowings.

Researchers have given substantive guidance to understanding age difference in mortgage lending. Dietrich and Johannsson (2005) used Home Mortgage Disclosure Act (HMDA) data to test age and gender discrimination by bivariate and multivariate analysis. Their study revealed that, in mortgage lending, applicants aged 55 and older are more likely to be denied than those aged 41 to 54 and younger. Additionally, age is highly correlated with economic factors that lenders consider in the underwriting process, such as collateral and credit score. Based on these findings, I hypothesized that lenders favor younger people in mortgage lending, and I anticipated that my research results, all else being equal, would demonstrate that people of different ages would receive different interest rates as evidence of statistical discrimination.

Ondrich, Ross, and Yinger (2001) proposed that discrimination varies geographically and is widespread across the United States. However, discriminatory behaviors are unique to specific areas, including its patterns and impact on the housing market. Ondrich, Ross, and Yinger (2001) believed that real estate professionals aim to minimize the time and effort spent on transactions while maximizing profit by considering the likelihood of transactions in specific locations. Applying this hypothesis to my research on rural and urban counties, geographic disparity will certainly result in varying social interactions between borrowers and lenders. As a result of these different

social interactions in housing prices and real estate market, individuals are more likely to experience varied treatment in real estate transactions. I also assume that such disparity in mortgage interest rates will be substantial because of the heterogeneous real estate market. Xu and Ting (2017) stated that the real estate market often experiences "hot" and "cold" periods in terms of timing and geographic locations. In hot areas, liquidity is good because average selling times are short, prices are higher than usual and generally display an upward trend, and the volume of transactions exceeds the norm. In contrast, cold areas have longer average selling times, lower and declining prices, and a low transaction volume. Compared to rural counties, urban counties have a hotter real estate market, which in turn indicates better real estate liquidity.

Studies exploring the interaction between age and geographic effect are virtually nonexistent. Considering geographic disparities, my research focuses on statistical discrimination in home mortgage borrowing and compares patterns in mortgage interest rates between urban and rural counties. I analyze whether the systematic or non-systematic relationship between age and mortgage interest rates differs between these two types of counties. I use two major economic drivers to explain potential differences in loan interest rate patterns between two types of counties: mortality rate and real estate market liquidity. In this thesis I investigate whether individuals of different ages receive different interest rates on loans because of different death rates. This relationship between age and interest rates could be positive, negative, or non-monotone. In addition, I study the factors that contribute to variations in different types of counties. In particular, I study the pattern of interest rate differences between urban and rural counties and identify the differences of interest rates in home mortgages. I answer two research questions: (1) Is there a systematic relationship between age and mortgage interest rate in home mortgages? and (2) Does the relationship between age and interest rate in home mortgages vary between urban and rural areas?

# 2 Institutional Background

## 2.1 Equal lending regulations

According to Fair Lending Laws and Regulations, the Equal Credit Opportunity Act (ECOA) is a law that prevents discrimination in credit transactions. The ECOA prohibits discrimination based on the following factors: (1) race or color; (2) religion; (3) national origin; (4) sex; (5) marital status; (6) age; (7) the applicant's receipt of income derived from any public assistance program; and (8) the applicant's exercise, in good faith, of any right under the Consumer Credit Protection Act.

Although the subject of discrimination can be applied to numerous topics such as education and labor markets, the most recent focus within the topic of discrimination has shifted to the credit markets. Journalists, businesspeople, politicians, and regulators are increasingly focused on lending discrimination, and many existing laws have defined such practice and regulations. Walter (1995) reviewed the Fair Lending Laws and Regulations and their enforcement and provided a detailed process and limitations of methods for law enforcement. The Fair Housing Act (FHAct) has provided guidance on the prohibition of discrimination in many activities related to the real estate industry besides lending, in addition to considering the factors of race, color, religion, sex, handicap, familial status, and national origin. First, any refusal to sell, rent, or negotiate for the property selling or rental of housing by discriminatory behaviors is forbidden. Second, the terms of sale or rental must not be revised in a discriminatory manner. Third, real estate brokerage organizations such as multiple-listing services (MLS)[1] are prohibited from discrimination in their terms of access to the organization. The downside of the FHAct is that it does not specify any actions or policies that are considered discrimination in the mortgage lending section. As a result, agencies might find it difficult to guide themselves in identifying potential lending discrimination and conducting further investigation because lenders are typically profit-based and avoid borrowers who are unlikely to repay.

[1] A multiple listing service, also referred to as a multiple listing system or multiple listings service, is an organization that provides a range of services for real estate brokers. These services help brokers establish contractual offers of cooperation and compensation with each other, as well as gather and share information to facilitate property appraisals.

Although both the FHAct and the ECOA have well defined the prohibition of discrimination by race, color, national origin, religion, sex, and other social factors in credit transactions, the laws do not provide clear, practical guidance for enforcement. The major methods for these law enforcements include the following: (1) complaints from aggrieved parties leading to investigations by enforcement agencies, (2) civil court actions, and (3) periodic fair lending examinations by the federal banking agencies. One of the biggest concerns in the efficiency of these enforcements is the striking difference between the examination of every institution and the practices of the federal agencies that are responsible for the antidiscrimination law enforcement. The common use of exams for banks is questioned to be the most efficient method for fair lending law because of the absence of periodic exams in some areas of antidiscrimination.

## 2.2 Statistical discrimination

Becker (2010) mentioned two concepts of discrimination: taste-based discrimination and statistical discrimination. Statistical discrimination occurs when people judge and make decisions based on statistical information rather than personal prejudice. In contrast, taste-based discrimination occurs when people have a bias against a particular group based on their personal preference. Taste-based discrimination might result in less efficiency in the market when compared to statistical discrimination.

Age discrimination can occur because of taste-based discrimination from people among social interactions. The fair lending regulations aim to eliminate taste-based discrimination, and this will largely reduce disparity of interest rates solely based of the borrower's age difference. In mortgage lending, a typical example of taste-based discrimination is the racial discrimination. For example, African Americans get lower loan approval rates than whites when other factors are equal, such as credit score and household income. Although lenders might view a borrower's age as a factor to determine the mortgage interest rate, the determination is based more on the disparity in credit risk between young people and older people, which leads to profit-based discrimination or statistical discrimination.

### 2.3 Community Reinvestment Act

The Community Reinvestment Act (CRA) is a regulation enacted in 1977 to ensure that financial institutions fulfill the credit needs of the communities in low- and moderate-income neighborhoods. As a regulation designed to focus on geographic distribution of credit, the CRA has significantly affected low- and moderate-income neighborhoods.

Existing laws such as the CRA would influence a lender's behavior in certain types of counties, especially rural areas. Such exogenous variation can also be used to interpret my empirical findings on age and loan interest rate pattern. Agarwal, Benmelech, Bergman, and Seru (2012) mentioned that adherence to the CRA leads to riskier lending, which includes 5 percent leverage loan amount and 15 percent more loan default. In addition, this effect is strongest when the market for private securitization is booming. In Illinois, the CRA covers major low- and moderate-income neighborhoods that primarily spread among rural counties. In mortgage lending, a higher interest rate reflects a borrower's higher risk, and lenders charge that to compensate for the risk bearing. As lending activities become riskier in those rural counties, I expect to see interest rates grow for individuals.

## 3 Theoretical Model

I propose a simple theoretical model to explain the relationship between age and interest rates. In a competitive market, brokers present mortgage proposals to borrowers. Borrowers can choose from a variety of offers, typically[2] selecting the one with the lowest interest rate. This theoretical model is a profit-maximization function and is based on a lender's perspective. Lenders set the optimal interest rate based on the borrower's risk and potential revenue from loan repayments. A zero-interest rate would certainly maximize borrowing demand but eliminate all revenue.

Conversely, an exceedingly high interest rate might result in a lucrative monthly debt payment but also increase the risk of default. Another issue

[2]In some circumstances, borrowers may choose a loan offer with a slightly higher rate due to considerations such as lender-borrower relationships or geographic locations. However, I assume that borrowers are focused on profit maximization.

associated with high interest rates is adverse selection because those more likely to default may opt for high-cost loans without concern about high debt payments. Older people who have higher mortality rates might be more likely to accept high interest rate loans and default because they are less likely to be concerned about long-term creditworthiness compared to younger people who are active in the job market.

A lender will not offer a mortgage to borrowers if there is no profit to be gained. It will otherwise maintain its deposits and pay a savings rate as the opportunity cost of funds. Because of market competition, lenders offer lower interest rates to secure borrowers. Borrowers are able to accept the minimum interest rate given by the market. Thus, in a perfectly competitive market, a borrower's interest rate will be the optimum interest rate in a zero-profit condition. Therefore, I use a zero-profit condition to solve for the optimum interest rate. I propose a simple theoretical model to illustrate the relationship between a lender's profit, the borrower's age, and the interest rate:

$$\pi = \frac{r^* - r}{r} - LP(a)[1 + \theta_r(a)]$$

| | |
|---|---|
| $\pi$ | lender's profit |
| $r$ | interest rate as the opportunity cost of funds |
| $r^*$ | optimum interest rate |
| $L$ | percentage of mortgage loss due to borrower's death |
| $P(a)$ | probability of default for people of different ages |
| $\theta_r(a)$ | county factor, which increases costs of selling collateral |
| $a$ | age of borrower |

$\frac{r^*-r}{r}$ represents the net present value of mortgage. The expected mortgage loss is expressed by the term, $LP(a)[1 + \theta_r(a)]$. $L$ stands for the percentage of mortgage loss due to the borrower's death. $P(a)$ signifies the probability of default based on the borrower's age, ranging from 0 to 1. $\theta_r(a)$ represents the county type factor that contributes to the mortgage loss, ranging from 0 to 1, because different types of counties have a different premium on the costs of selling the collateral, which causes profit loss for lenders. The value

of $\theta_r(a)$ for urban counties is relatively small (equal or close to zero) because it is comparatively easy to liquidate the collateral there, as opposed to in a rural county. In contrast, $\theta_r(a)$ is relatively large for a rural county (equal or close to one).

### 3.1 Zero-profit condition

In a zero-profit condition, $\pi = 0$ and the interest rate at optimum, $r^*$ , is solved by the following equation:

$$r^* = r\{1 + LP(a)[1 + \theta_r(a)]\}$$

where $P'(a) > 0$ as the aging increases mortality rate and thus increases the default probability. $\theta'_r(a) \geq 0$, as it signifies that the rural county imposes more difficulty in selling the collateral, meaning that counties with more rural characteristics can increase costs of selling the collateral and reduce a lender's profit, which increases the $r^*$. Similarly, as older people exhibit a higher mortality rate compared to young people, $P(a)$ will be larger and therefore increase the $r^*$, other factors being equal.

In practice, lenders often charge a higher interest rate to older borrowers to compensate for the reduced profit due to the higher default rate, other factors being equal. Additionally, taking geographic disparities into account, liquidating collateral in the event of a loan default proves to be more challenging in rural counties. This leads to higher costs of selling the collateral as compared to their urban counterparts. Therefore, to offset the increase in the cost of selling collateral (or reduction of lender's profit), lenders will apply a higher interest rate to similar loans in rural counties, other factors being equal.

## 4 Data

### 4.1 Data description

The sample for this analysis is the Home Mortgage Disclosure Act

(HMDA) data in Illinois. HMDA is a data source regularly used in mortgage discrimination research. Economists and researchers have used it multiple times[3]. Because of time constraints, I used the most recent data available in the HMDA, ranging from 2018 to 2021, and concentrated on Illinois. My focus on Illinois will serve as a preliminary study, and there is a possibility of extending this research to the entire United States.

The HMDA was originally enacted by the US congress as a response to concern about discriminatory credit practices and redlining in the housing market. These data aim to promote transparency and fair lending practices by requiring financial institutions to report data on mortgage transactions. The sources of HMDA data primarily come from financial institutions, including banks, credit unions, mortgage firms, and the Consumer Financial Protection Bureau. These publicly available loan data contain information on individual home loan applications, including loan characteristics, individual-level age, race and ethnicity, and purchaser and denial information. The reliability of HMDA data depends on financial institutions' reporting practices and might contain data errors caused by misinterpretation of reporting requirements. In addition, HMDA data contains many missing values and demographic information and credit score, which limits the ability to fully understand the fairness of lending activities.

In the data sample I used, the key measurements include the loan-to-value ratio; the loan term; the total loan costs; the loan amount; the property value; and the income and interest rates, with the number of observations of 41,071, 41,066, 36,478, 41,071, 41,009, 39,302, and 41,071, respectively. The loan-to-value ratio represents the ratio of loan amount to the appraised value of the property, typically[4] varying from 0 percent to 100 percent. The loan term refers to the duration of a mortgage that takes a borrower to repay the mortgage, ranging from 1 month to 480 months. Total loan costs are numerical values of any fees associated with obtaining and closing on the new loan, with values from \$0 to \$25,693. Loan amount, the total sum of money that a borrower receives from a lender in the form of a loan, ranges from

[3]For example, Munnell, Tootell, Browne, and McEneaney (1996) analyzed mortgage lending practices in the Boston area using data from the Home Mortgage Disclosure Act (HMDA). Bayer, Ferreira, and Ross (2014) used and analyzed HMDA data to examine the variation in high-cost mortgage lending across different racial and ethnic groups.

[4]A total of 316,247 observations with loan-to-value ratios greater than 100 percent have been removed from the data set.

$5,000 to $260,000,000. Property value, indicating the value of mortgaged property, spans from $5,000 to $453,000,000. Income, the combined earnings of all household members, ranges from 0 to 164 million. The interest rate, which is defined as the percentage signifying the cost of borrowing money or the profit for the lender, generally[5] varies from 0 percent to 15 percent. The applicant's age is reported as an age group, at a 10-year interval, not as an integer. Additionally, age groups are created to categorize the applicant with different age intervals. The age groups are dummy variables that index with the following age intervals: smaller than 25 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 years to 64 years, 65 to 74 years, and more than 75 years.

There are other categorical variables that provide information for the mortgage characteristics. Race and ethnicity, gender, and county code[6] are used to describe the borrower's characteristics and geographic distributions. Race and ethnicity can be classified into the following groups: white, black, Asian, and other. Gender is categorized as male, female, or not available.

Because my research focuses on the state of Illinois and uses data from all urban counties and rural counties within Illinois, I used the county code to identify different types of counties classified by rural and urban. According to the US Census Bureau, in 2010 rural counties were defined as counties with smaller populations ranging from 2,500 to 5,000 people and lower population densities, often fewer than 500 people per square mile or 200 people per square kilometer. These areas predominantly have agriculture-based economies and are characterized by small towns and villages. In contrast, urban counties are defined as counties with larger populations, often exceeding 5,000 people, and higher population densities, usually over 500 people per square mile or 200 people per square kilometer. These areas have diverse economies, encompassing a range of industries such as manufacturing, services, technology, and finance. Figure 1 shows a map of urban counties and rural counties. There are over 2 million loan data from 19 urban counties and 83 rural counties

[5]A total of 72,216 observations with an interest rate greater than 100 percent have been removed from the dataset.

[6]County code specifies which county the loan data originate from. County FIPS Codes are unique five-digit codes that represent specific US counties where the first two digits represent the county's state, and the last three digits represent the county. For example, "17031" is the FIPS Code for cook county, Illinois; "17" represents Illinois; and "031" represents Cook County.

between 2018 and 2021.

## 4.2 Descriptive statistics

Table 1 shows the descriptive statistics for gender and race and summarizes the gender and race characteristics of loan applicants who secured a loan. Urban counties display a greater diversity in racial composition among loan applicants because the proportion of minority applicants is considerably higher than that of rural counties. Table 2 shows the distribution of applicants' age. People of different ages are grouped with a 10-year interval, from below 25 years old to above 74 years old. Among 40,585 observations in rural counties and 429,522 observations in urban counties, the result shows that the population of younger people below 25 years old in rural counties is more than that of urban counties. The main loan applicants are between 25 and 54 years old.

The loan characteristics are summarized in Table 3, presenting the general loan characteristics, including loan amount, loan term, loan-to-value ratio, total loan costs, and applicant age. Most loans have an interest rate of 3 percent. As Figure 3 shows, the distribution of loan-to-value ratio is discontinuous, separated by 80 percent as the threshold. One possible explanation is the existence of private mortgage insurance. People try to borrow the maximum amount while avoiding paying private mortgage insurance, which is required when the loan-to-value ratio exceeds 80 percent. Therefore, the majority of approved loans are concentrated at 80 percent loan-to-value ratio, whereas the loans above 80 percent are less popular. Regarding loan term, the most popular approved loans are 30-year term loans. Using log value to exclude polarity, the distribution of property value and income are both in a normal distribution shape. I also see differences between urban counties and rural counties. Property values in urban counties are twice as much as those in rural counties, and the household income is even higher. However, the statistics of interest rate and loan-to-value ratio are similar.

# 5 Empirical Strategy

To test whether there are different effects of age on interest rates between rural and urban counties, I used a linear version of the hedonic price model, presented in Table 4. The exact fixed-effect regression design is with the following specification:

$$Y_i = \beta X_i + \eta_c + \lambda_a + \kappa_{a,u} + \epsilon_i$$

where $Y_i$ is the loan interest rate. $X_i$ are the independent variables, which are total loan costs, loan amount, property value, loan-to-value ratio, and loan term. The applicant's age is represented by dummy variables that correspond to 10-year age intervals, as detailed in Table 2. These dummies are combined as age group. To capture the county effect, $\eta_c$ is the parameter of fixed effect for county, controlling the difference for each county. $\lambda_a$ is the parameter for the interaction between age and interest rates in rural counties, which is the coefficient for age group in Table 4. $\lambda_a + \kappa_{a,u}$ is the parameter for interaction between age and mortgage rates in urban counties. $\kappa_{a,u}$ is the parameter for the difference between rural and urban counties for the interaction of age and county, and this parameter will be captured by the coefficient of urban dummy, an added coefficient for estimating the difference of rural counties and urban counties.

As robustness checks, I considered an alternative assumption and a new aggregation of the loan data set. Because of the potential impact of COVID-19 on mortgage lending since 2020, I divided the data from 2018 to 2021 into two parts: loan transactions in 2018 and 2019 are treated as pre-COVID-19 loan data, and those in 2020 and 2021 are treated as post-COVID-19 loan data. I compared the relationship between interest rate and age for pre- and post-COVID-19 loan data on an annual basis and found similar patterns in 2018, 2019, 2020, and 2021. Figures 9 to 11 illustrate the patterns for each year. The relationship holds, although there is a varying degree of impact from age on interest rate. Although I expected the pattern in 2020 to differ from other years, given that COVID-19 emerged as an initial exogenous shock and significantly affected lender's perceptions of older people's default risks, such as a rapid increase in their mortality rates, the results indicate that the pattern does not deviate from other years. Therefore, I conclude that the

monotonically increasing trend in interest rate and age is robust over recent years.

# 6 Results

Table 4 presents a fixed-effect estimation with robust standard errors and a comparison between rural and urban counties. The coefficient of the urban dummies captures the differences in interaction between rural and urban counties, which is the interaction term using age group dummies multiplied by urban county. The coefficient of the rural county estimates the interaction between the rural county and age group, whereas the coefficient of the urban county is the sum of the rural county coefficient plus urban dummies. All the results are relative to the age group of 35 to 45. Figure 7 illustrates the trend of coefficients for rural and urban counties, representing the effect of age on interest rates for both types of counties, respectively. The patterns for age groups in urban and rural counties are similar. Both counties exhibit a monotonically increasing relationship between age and interest rate. Below age 35, people are more likely to receive a negative interest rate premium based on their age. After age 55, people are more likely to receive a positive interest rate premium based on their age. The coefficient for an urban county is consistently higher than that for a rural county within the same age group, all else being equal.

Column 7 of Table 5 presents the fixed-effect regression results for counties, along with the regression results for other covariates, one by one. To manage the data with heavily skewed distributions, I take the logarithm transformations on the following covariates in my regression: property value, income, and total loan costs. All the results are relative to the age group of 35 to 45. In my model, I control counties and other loan characteristic covariates, such as property value, loan-to-value ratio, income, loan term, and total loan costs. Given the adequately large data set, the findings reveal significant differences across all covariates. Among the major controlling variables, property value and income show a strong positive correlation with interest rate, whereas the loan-to-value ratio demonstrates a negative correlation. Figure 8 illustrates the trend of coefficients representing the effect of age on interest rates for both types of counties. Considering the monotonic

upward trend, the effect of age on interest rate is expected to be higher for older people, all else being equal.

# 7 Discussion

In this thesis I empirically investigate the systematic relationship between age and interest rates in Illinois and whether interactions based on county type (rural or urban) and age of borrower affect the loan interest rates charged to individuals. I detect a strong relationship between interest rate and age. Regardless of geographic disparities, people of different ages are charged different loan interest rates, all else being equal. The pattern from 2018 to 2021 shows a systematic relationship between the borrower's age and loan interest rate, which remains consistent over these years. People receive higher interest rates as their age increases. Although this relationship is robust, exogenous variations such as policy reforms can lead to disparity in lending behaviors and outcomes.

People of different ages have a self-selection process, where older people might consider employment opportunities as less of a concern. The moral hazard and adverse selection can cause complications for lenders on charging the optimum interest rate. It is likely that older people are more inclined to default on loan repayment because they are less likely to be concerned about their credit in the long term. Meanwhile, people who are more likely to default would borrow high-interest rate loans and further hurt the lender's profits. As the needs of different aged groups vary, the effect of age on interest rates would vary.

Regarding geographic disparities, the age effect demonstrates similar effects on both urban and rural counties; however, the interest rate in urban counties is always higher than that in rural counties for the same age group. In light of the differences, there are several reasons to explain the disparities in interest rates between rural and urban counties from the labor market and real estate market perspectives. First, urban counties may have more commercial activities and employment opportunities, which makes people better off in the labor market. Then, another economic driver explaining this difference is the disparity in real estate markets. Urban counties have a more liquid real estate market with higher average housing prices compared

to rural counties. Conversely, rural counties have lower living costs and property values on average, which benefits people in terms of living expenses. These geographic disparities likely shape people's preferences for obtaining mortgages because of variations in budgets and preferences for goods caused by costs. Furthermore, rural counties have smaller populations and lower population density, potentially leading to a diffusion of loan demand and future property resales.

The limitation of this study is the lack of geographic diversity. Because this study focuses only on one state, the results are suggestive rather than conclusive. A better extension of the study would involve using data from all US states to identify similarities in the findings.

# 8 Implications

The research implications are twofold. Policymakers should focus more on lending policies regarding age and should adjust policies on aging and mortgage lending to protect potential weak groups such as aging people. As the vulnerable group, older people need policy attention in fair lending.

In terms of borrowers and lenders, the risk for people in different age groups differ, which might affect the rate of default and repayments, so lenders might adjust loan rates for people in different age groups. For borrowers, people of different ages can consider the trade-offs they face while choosing to borrow home mortgages. Mortgage rates vary by geographic locations; therefore, home buyers should learn how to choose proper lenders based on geographic location. Because different types of counties might have different effects on loan premium based on real estate market efficiency, borrowers with similar credit scores should decide which type of county to live in and borrow mortgages because a lower interest rate alleviates financial pressure from less debt payment and also creates a higher return.

**Table 1**

*Descriptive statistics of gender and race or ethnicity (HMDA data 2021, Illinois)*

| Rural Counties | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Gender** | Male | Female | N/A | **Race** | White | Black | Asian | Other |
| number | 13047 | 7906 | 20118 | number | 36090 | 430 | 201 | 4350 |
| percent | 31.8% | 19.2% | 49.0% | percent | 87.9% | 1.0% | 0.5% | 10.6% |

| Urban Counties | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Gender** | Male | Female | N/A | **Race** | White | Black | Asian | Other |
| number | 145217 | 97886 | 191562 | number | 299907 | 28692 | 35751 | 70315 |
| percent | 33.4% | 22.5% | 44.1% | percent | 69.0% | 6.6% | 8.2% | 16.2% |

Notes: This table summarizes the gender and race distribution of rural counties and urban counties in Illinois based on HMDA county-level data of Illinois in 2021. 49% of the data in rural counties and 44.1% of the data in urban counties miss information about applicant's gender. The racial makeup of rural and urban counties includes white, black, Asian, and other races, which encompass American Indian, Native Hawaiian, or Other Pacific Islander populations.

**Table 2**

*Distirbution of age group (HMDA data 2021, Illinois)*

| | Rural Counties | | Urban Counties | |
|---|---|---|---|---|
| Age | N | *Percent* | N | *Percent* |
| 0-25 | 2,227 | 5.4% | 8,534 | 2.0% |
| 25-34 | 8,872 | 21.6% | 86,757 | 20.0% |
| 35-44 | 9,948 | 24.2% | 123,603 | 28.4% |
| 45-54 | 8,174 | 19.9% | 100,112 | 23.0% |
| 55-64 | 6,599 | 16.1% | 66,737 | 15.4% |
| 65-74 | 3,679 | 9.0% | 33,645 | 7.7% |
| 74+ | 1,086 | 2.6% | 10,134 | 2.3% |

Notes: This table summarizes the age distribution of rural counties and urban counties in Illinois based on HMDA county-level data of Illinois in 2021. The loan applicant's age divides into seven age groups: below 25 years, 25 to 34 years, 35 to 44 years, 45 to 55 years, 55 to 64 years, 65 to 74 years, and above 74 years. Excluding data missing age, the total number of observations is 40,585 in rural counties and 429,522. In both rural and urban counties, the distribution of borrowers among various age groups is similar. However, the proportion of young people under the age of 25 in rural counties exceeds twice that of urban counties.

**Table 3**

*Descriptive statistics of loan characteristics (HMDA data 2021, Illinois)*

| | Rural Counties | | | | | |
|---|---|---|---|---|---|---|
| | N | Min | Max | Mean | Median | Std. |
| Interest rate (percent) | 41071 | 0.00 | 14.99 | 3.07 | 2.99 | 0.71 |
| Loan to value ratio (percent) | 41071 | 0.63 | 100.00 | 77.13 | 80.00 | 18.22 |
| Total loan costs | 36478 | 0.00 | 25693 | 2942 | 2363 | 2039 |
| Property value | 41009 | 5000 | 17500000 | 201443 | 165000 | 220238 |
| Income (thousand) | 39302 | 0 | 4559 | 90 | 74 | 81 |
| Loan amount | 41071 | 5000 | 10500000 | 144329 | 125000 | 133365 |
| Loan term (month) | 41066 | 1 | 480 | 287 | 360 | 97 |
| | Urban Counties | | | | | |
| | N | Min | Max | Mean | Median | Std. |
| Interest rate (percent) | 434664 | 0.00 | 15.99 | 3.02 | 2.98 | 0.74 |
| Loan to value ratio (percent) | 434664 | 0.01 | 100.00 | 74.08 | 76.69 | 17.54 |
| Total loan costs | 384162 | 0 | 188510 | 3942 | 2996 | 2923 |
| Property value | 433661 | 5000 | 453000000 | 410042 | 315000 | 1302456 |
| Income (thousand) | 417897 | 0 | 164000 | 134 | 100 | 328 |
| Loan amount | 434665 | 5000 | 260000000 | 277764 | 225000 | 742441 |
| Loan term (month) | 434549 | 1 | 480 | 306 | 360 | 87 |

Notes: This table summarizes the descriptive statistics of loan characteristics in rural counties and urban counties in Illinois based on HMDA county-level data of Illinois in 2021. The loan is characterized by a few measurements, including interest rates, loan-to-value ratio, total loan costs property value, income, loan amount and loan terms. The number of observations vary between these measurements. The maximum interest rate is near 15 percent, but most interest rates are concentrated between 2 and 4 percent, with little difference between the two types of counties. Loan-to-value data ranges from a minimum of 0 percent to a maximum of 100 percent, but most loan-to-value data are concentrated around 80 percent. In urban counties, total loan costs are significantly higher compared to rural counties, and both property values and related loan amounts are larger in urban areas as well. In terms of household income, urban counties are higher than rural counties. Loan terms in both types of counties are very similar, with most concentrated at 360 months and the longest loan term being 40 years.

**Table 4**

*Fixed effect estimation to capture difference in rural and urban county interaction with age*

| | Coefficient | Robust Standard Error | 95% Confidence Interval | |
|---|---|---|---|---|
| ln(Property value) | -0.206 | 0.002 | [ -0.210 | -0.201 ] |
| Loan-to-value ratio | 0.002 | 0.000 | [ 0.001 | 0.002 ] |
| ln(Income) | 0.044 | 0.002 | [ 0.041 | 0.047 ] |
| Loan term | 0.003 | 0.000 | [ 0.002 | 0.003 ] |
| ln(Total loan costs) | -0.034 | 0.002 | [ -0.037 | -0.031 ] |
| **Age group dummies (omitted age is 35-44)** | | | | |
| Age (0-25) | -0.070 | 0.013 | [ -0.096 | -0.043 ] |
| Age (25-34) | -0.048 | 0.009 | [ -0.065 | -0.032 ] |
| Age (45-54) | 0.019 | 0.009 | [ 0.008 | 0.046 ] |
| Age (55-64) | 0.027 | 0.010 | [ 0.017 | 0.064 ] |
| Age (65-74) | 0.041 | 0.012 | [ 0.052 | 0.137 ] |
| Age (74+) | 0.094 | 0.022 | [ -0.315 | -0.275 ] |
| *Age group dummies×(urban=1)* [a] | | | | |
| Age (0-25) | 0.030 | 0.015 | [ 0.001 | 0.058 ] |
| Age (25-34) | 0.015 | 0.009 | [ -0.002 | 0.033 ] |
| Age (45-54) | 0.013 | 0.010 | [ -0.005 | 0.032 ] |
| Age (55-64) | 0.013 | 0.010 | [ -0.007 | 0.032 ] |
| Age (65-74) | -0.002 | 0.012 | [ -0.026 | 0.023 ] |
| Age (74+) | -0.018 | 0.022 | [ -0.062 | 0.026 ] |
| $R^2$ | 0.19860 | | | |
| N | 405879 | | | |

Notes: This table displays the fixed effect estimation with robust standard error for two types of counties, according to econometric specification with fixed effect for age (in ten-year intervals) and other borrower and loan characteristics. It estimates the interaction between age and county type. The control variables include log value of property value, loan-to-value ratio, log value of income, loan term, and log value of total loan costs. The coefficient of age group dummies is the estimation of interaction between rural county and age group. In this table, all the results from different age groups are relative to age group from 35 to 45 years.

[a] Age group dummies×(urban=1) is the coefficient of the interaction term of urban dummies and age group, which captures the differences of the urban county and rural county for the age effect on interest rates.

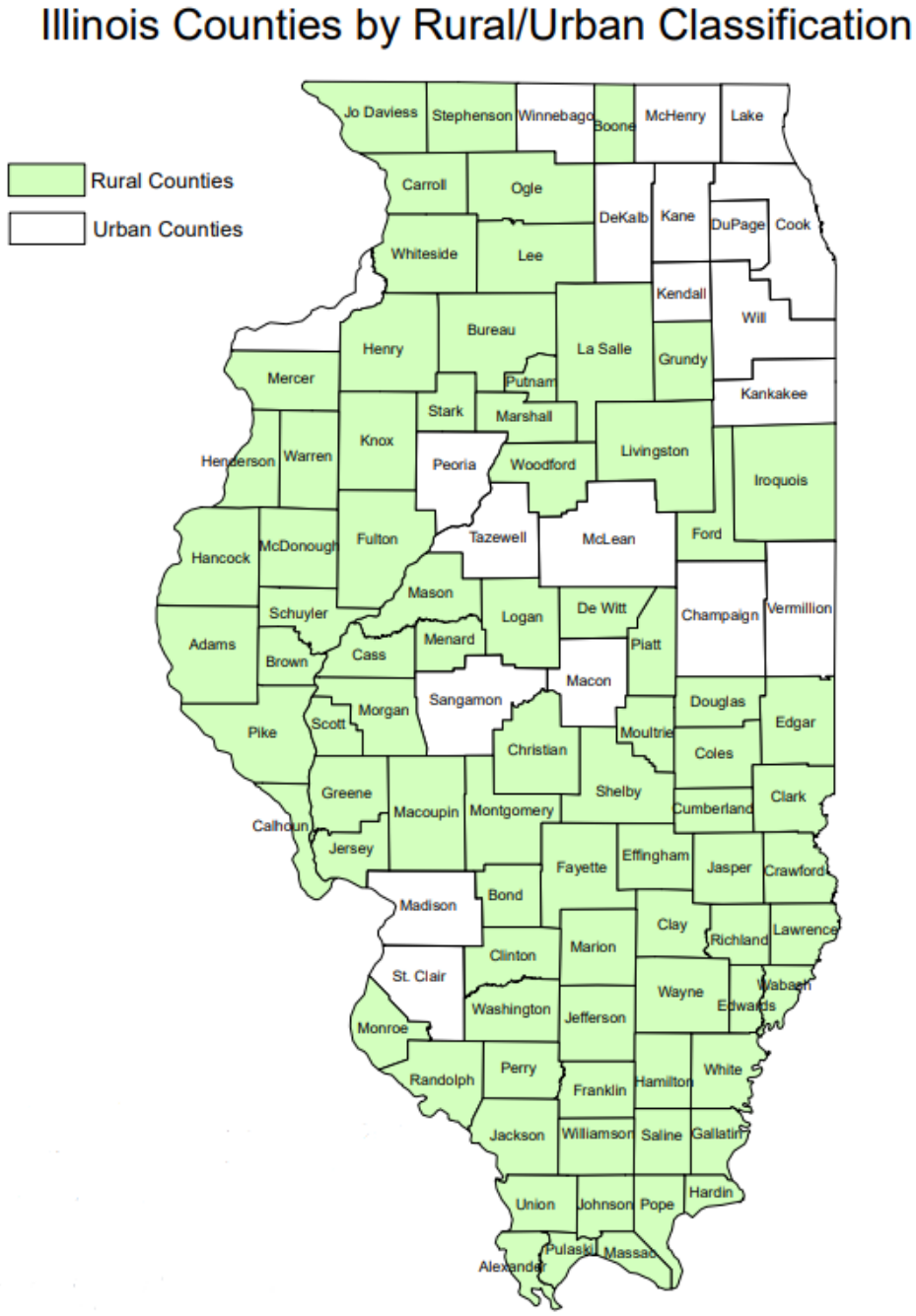
**Table 5**

*Fixed effect estimation of individual variables*

| | Dependent Variable, Interest Rate | | | | | | |
|---|---|---|---|---|---|---|---|
| Explanatory Variable | [1] | [2] | [3] | [4] | [5] | [6] | [7] |
| ln(Property value) | -0.147 | | | | | | -0.206 |
| | (0.002) | | | | | | (0.002) |
| Loan to value ratio | | 0.005 | | | | | 0.002 |
| | | (0.000) | | | | | (0.000) |
| ln(Income) | | | -0.076 | | | | 0.044 |
| | | | (0.001) | | | | (0.002) |
| Loan term (in year) | | | | 0.001 | | | 0.003 |
| | | | | (0.000) | | | (0.000) |
| ln(Total loan costs) | | | | | 0.006 | | -0.034 |
| | | | | | (0.001) | | (0.002) |
| Age (0-25) | | | | | | 0.088 | -0.045 |
| | | | | | | (0.006) | (0.005) |
| Age (25-34) | | | | | | 0.025 | -0.035 |
| | | | | | | (0.003) | (0.002) |
| Age (45-54) | | | | | | -0.007 | 0.031 |
| | | | | | | (0.003) | (0.002) |
| Age (55-64) | | | | | | 0.008 | 0.039 |
| | | | | | | (0.003) | (0.003) |
| Age (65-74) | | | | | | 0.033 | 0.039 |
| | | | | | | (0.004) | (0.003) |
| Age (74+) | | | | | | 0.106 | 0.078 |
| | | | | | | (0.007) | (0.006) |
| $R^2$ | 0.016 | 0.015 | 0.006 | 0.028 | 0.013 | 0.079 | 0.199 |
| N | 474670 | 475735 | 452255 | 475615 | 420640 | 475735 | 405879 |

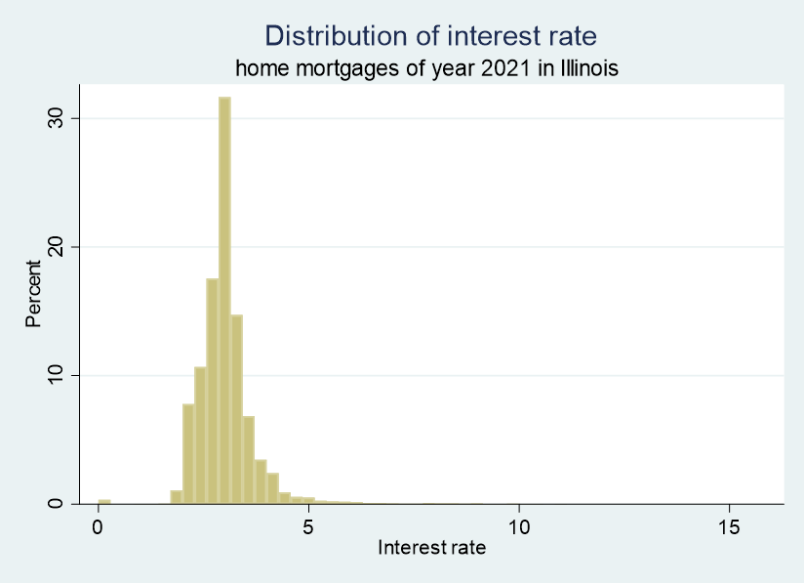
Notes: This table presents the fixed effect estimation with robust standard error for different controlling variables. Column 6 displays the fixed effect regression estimation results, controlling for age, and shows that the coefficient between age and interest rates, which captures the likelihood of positive or negative pricing premium on loan rates people of different age group expected to receive, declines from 0-25 years to 45-54 years in the age groups but increases since then. Column 7 shows the fixed effect regression estimation results, controlling for county type (rural or urban). The results show a different pattern as column 6, where the coefficient between age groups and interest rates are monotonically increasing. All the results from different age groups are relative to age group from 35 to 45 years. Column 1 to 6 shows the fixed effect regression results controlling for log value of property value, loan to value ratio, log value of income, loan term, and log value of total loan costs, respectively.

**Figure 1**



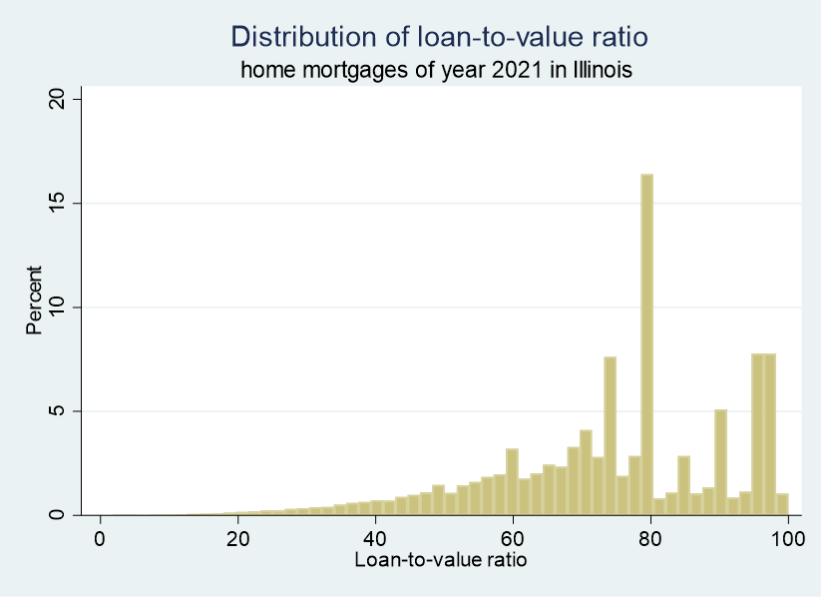

Notes: This figure shows the classification of urban counties and rural counties in Illinois and their geographic location in the state map. This map is derived from the county map provided by the Illinois Primary Health Care Association in December, 2020. In Illinois, there are 83 counties categorized as rural area, while 19 counties are identified as urban region.

**Figure 2** *Distribution of interest rate (HMDA data 2021, Illinois)*



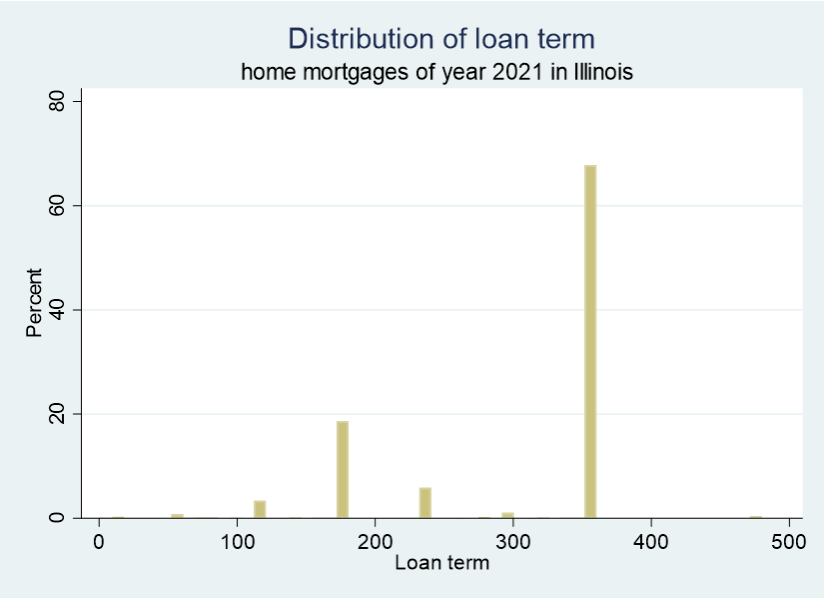

Notes: This figure shows the distribution of interest rate of home mortgages in Illinois based on HMDA data in Illinois in 2021. The chosen range is from 0 percent minimum to 15 percent, excluding outliers.

**Figure 3** *Distribution of loan-to-value ratio (HMDA data 2021, Illinois)*



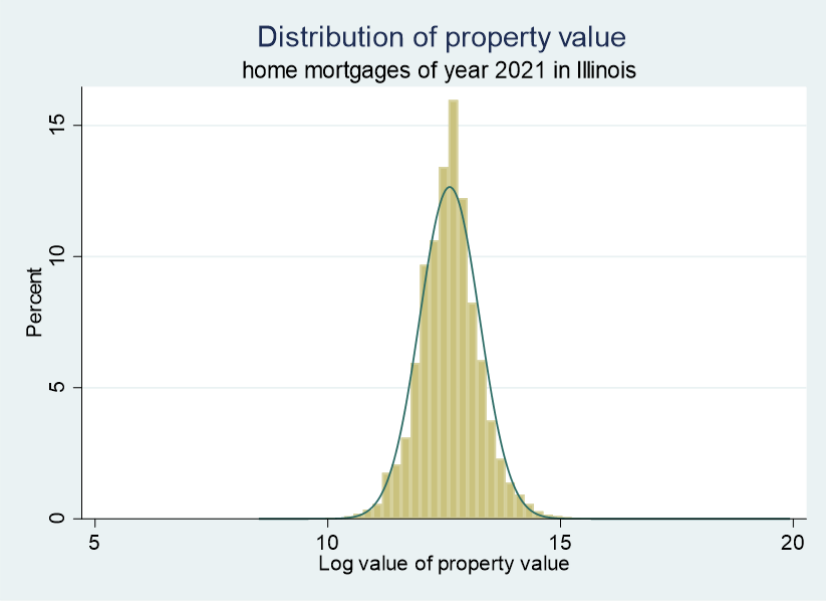

Notes: This figure shows the distribution of loan-to-value ratio of home mortgages in Illinois based on HMDA data in Illinois in 2021. The chosen range is from 0 percent to 100 percent at maximum. The highest frequency of mortgage's loan-to-value ratio is at 80 percent.

**Figure 4** *Distribution of loan term (HMDA data 2021, Illinois)*



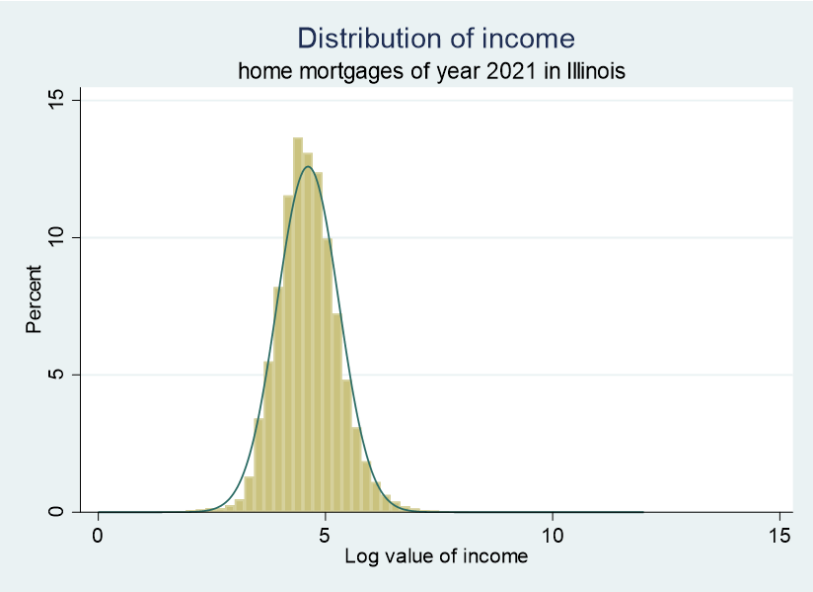

Notes: This figure shows the distribution of loan term of home mortgages in Illinois based on HMDA data in Illinois in 2021. The chosen range is from 1 month to 480 months at maximum. The highest frequency of mortgage loan term is at 360 months.

**Figure 5** *Distribution of property value (HMDA data 2021, Illinois)*



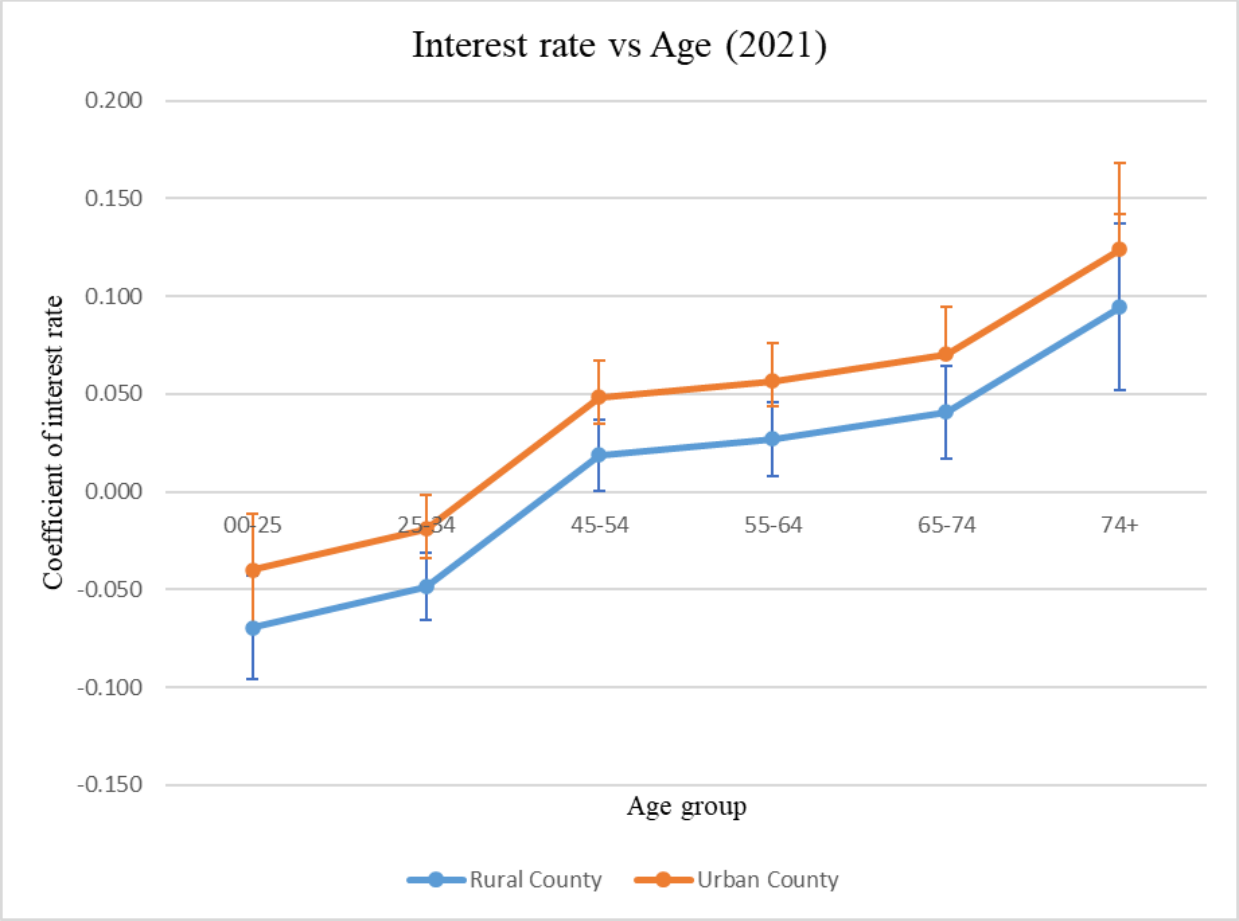

Notes: This figure shows the distribution of property value in Illinois based on HMDA data in Illinois in 2021. The log value of the property value ranges from 5 to 20. This distribution follows a normal distribution shape.

**Figure 6** *Distribution of income (HMDA data 2021, Illinois)*



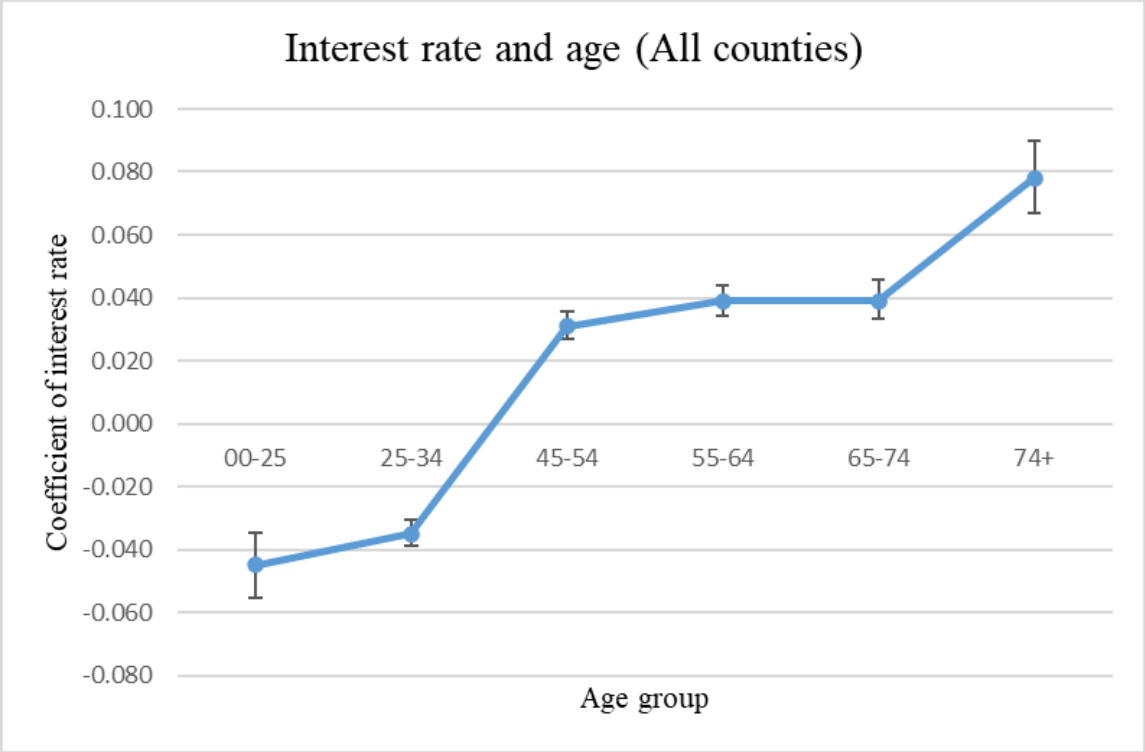

Notes: This figure shows the distribution of applicant's income in Illinois based on HMDA data in Illinois in 2021. This distribution follows a normal distribution shape.

**Figure 7** *Pattern of interest rate and age coefficient with county interaction (HMDA data 2021, Illinois)*




Notes: This figure plots the trend of coefficient of interest rate and age group for rural county and urban county, respectively, based on HMDA county-level data in Illinois 2021. The coefficient of rural county is based on the fixed effect regression with robust standard error, controlling for rural county and other variables including log value of property value, loan-to-value ratio, log value of income, loan term, total loan costs. The coefficient of urban county is based on the fixed effect regression with robust standard error, controlling for urban county and other variables including log value of property value, loan-to-value ratio, log value of income, loan term, total loan costs. This pattern separately shows the interaction between age and interest rates for rural counties and urban counties. For each age group, a confidence interval is provided.

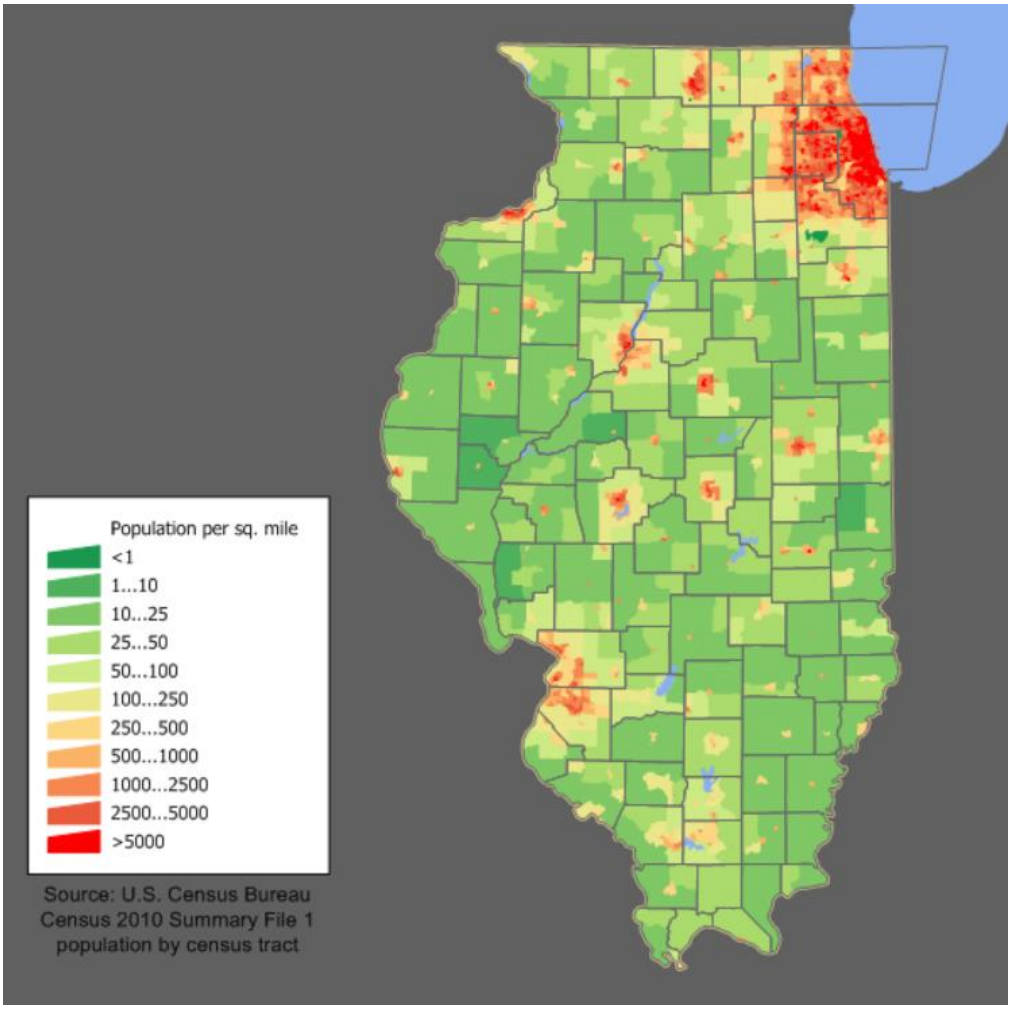
**Figure 8** *Pattern of interest rate and age coefficient (HMDA data 2021, Illinois)*




Notes: This figure plots the trend of coefficient of interest rate and age group for all counties in Illinois 2021, based on HMDA county-level data in Illinois 2021. This coefficient is based on the fixed effect regression with robust standard error, according to econometric specification with fixed effects for age (in ten-year intervals), county, and other borrower or loan characteristics. This pattern shows the interaction between age and interest rates for all types of counties. For each age group, a 95% confidence interval is provided.

## Appendices

### A Illinois demographic – population density map




Notes: This map shows the population density in Illinois, derived from the US Census Bureau 2010 summary.

**B Pattern of interest rate and age coefficient with county interaction from 2018 to 2020**

**Figure 9** *Pattern of interest rate and age coefficient with county interaction (HMDA data 2018, Illinois)*


Notes: This figure plots the trend of coefficient of interest rate and age group for rural county and urban county, respectively, based on HMDA county-level data in Illinois 2018. The coefficient of rural county is based on the fixed effect regression with robust standard error, controlling for rural county and other variables including log value of property value, loan-to-value ratio, log value of income, loan term, total loan costs. The coefficient of urban county is based on the fixed effect regression with robust standard error, controlling for urban county and other variables including log value of property value, loan-to-value ratio, log value of income, loan term, total loan costs. This pattern separately shows the interaction between age and interest rates for rural counties and urban counties. For each age group, a confidence interval is provided.

**Figure 10** *Pattern of interest rate and age coefficient with county interaction (HMDA data 2019, Illinois)*


Notes: This figure plots the trend of coefficient of interest rate and age group for rural county and urban county, respectively, based on HMDA county-level data in Illinois 2019. The coefficient of rural county is based on the fixed effect regression with robust standard error, controlling for rural county and other variables including log value of property value, loan-to-value ratio, log value of income, loan term, total loan costs. The coefficient of urban county is based on the fixed effect regression with robust standard error, controlling for urban county and other variables including log value of property value, loan-to-value ratio, log value of income, loan term, total loan costs. This pattern separately shows the interaction between age and interest rates for rural counties and urban counties. For each age group, a confidence interval is provided.

**Figure 11** *Pattern of interest rate and age coefficient with county interaction (HMDA data 2020, Illinois)*


Notes: This figure plots the trend of coefficient of interest rate and age group for rural county and urban county, respectively, based on HMDA county-level data in Illinois 2020. The coefficient of rural county is based on the fixed effect regression with robust standard error, controlling for rural county and other variables including log value of property value, loan-to-value ratio, log value of income, loan term, total loan costs. The coefficient of urban county is based on the fixed effect regression with robust standard error, controlling for urban county and other variables including log value of property value, loan-to-value ratio, log value of income, loan term, total loan costs. This pattern separately shows the interaction between age and interest rates for rural counties and urban counties. For each age group, a confidence interval is provided.

## References


Agarwal, Sumit, et al. Did the Community Reinvestment Act (CRA) lead to risky lending? No. w18609. *National Bureau of Economic Research*, 2012.

Bayer, Patrick, Fernando Ferreira, and Stephen L. Ross. Race, ethnicity and high-cost mortgage lending. No. w20762. *National Bureau of Economic Research*, 2014.

Becker, Gary S. The economics of discrimination. *University of Chicago press*, 2010.

Dietrich, Jason. Searching for age and gender discrimination in mortgage lending. *Comptroller of the Currency*, 2005.

Munnell, Alicia H., et al. "Mortgage lending in Boston: Interpreting HMDA data." *The American Economic Review* (1996): 25-53.

Ondrich, Jan, Stephen Ross, and John Yinger. "Geography of housing discrimination." *Journal of Housing Research* (2001): 217-238.

Walter, John R. "The fair lending laws and their enforcement." *FRB Richmond Economic Quarterly* 81.4 (1995): 61-77.

Xu, Ting. "The relationship between interest rates, income, GDP growth and house prices." *Research in Economics and Management* 2.1 (2017): 30-37.