Essays in Homeownership and Mortgage Finance

by

Nirupama R. Kulkarni

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Nancy Wallace, Co-chair
Professor Ulrike Malmendier, Co-chair
Associate Professor David Sraer
Associate Professor Danny Yagan

Summer 2016

Essays in Homeownership and Mortgage Finance

Abstract

Essays in Homeownership and Mortgage Finance

by

Nirupama R. Kulkarni

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Nancy Wallace, Co-chair

Professor Ulrike Malmendier, Co-chair

This dissertation consists of two chapters on mortgage finance and homeownership. Federal policy often institutes uniform pricing across regions in the name of fairness. I study the unintended consequences of such uniform pricing in the context of the residential mortgage market, which is heavily influenced by the securitization policies of the government sponsored enterprises (GSEs). I show that the regional uniformity of GSE-conforming mortgage rates leads to credit rationing. I develop three results by exploiting differences in the strength of lender rights — state laws that limit a lender's recourse and ability to foreclose on property — as a source of regional variation. First, controlling for borrower characteristics, I find that GSE-securitized mortgage rates *do not vary* across lender rights whereas those of privately securitized mortgages *do vary.* Second, the lack of regional variation in mortgage rates leads to the credit rationing of marginal borrowers in regions with borrower-friendly laws, whereas, regression discontinuity and bunching estimates show that the GSEs "cherry-pick" the better risks leading to greater credit access in lender-friendly areas. Finally, I find that the GSEs' cost of funds advantage distorts the pool of borrowers available to the private market and that only some of the GSE-rationed borrowers can access privately securitized mortgages. Overall, the results demonstrate how uniform regional pricing and cost of funds advantages of the GSEs distorts the competitive landscape of the US mortgage market.

The second chapter studies the impact of homeownership on intergenerational mobility. The benefits of homeownership feature prominently in the academic and policy discussions alike. Increasing homeownership has been a major policy goal for decades, especially in low-income areas. We show that the positive relationship between homeownership and intergenerational mobility is highly place-dependent. First, we link commuting zone-level homeownership rates to intergenerational mobility, and find a strong positive relationship. The relationship persists after instrumenting for ownership using housing supply and price shocks. Second, we show that the positive relation between of homeownership and upward mobility is significantly diminished or disappears in areas with high sprawl or segregation, whether we use income segregation, racial segregation, or a new measure of homeowner seg-

regation. These results, as well as additional findings on the formation of social capital and on school quality, suggest that homeownership may not benefit, or may even disadvantage children in segregated, poor areas, possibly through reduced residential mobility.

To my family

Contents

List of Figures

List of Tables

Acknowledgments

I want to thank the many who helped me throughout this dissertation. I am deeply indebted to my advisers Nancy Wallace, Ulrike Malmendier, David Sraer, Danny Yagan and Dwight Jaffee who encouraged me throughout the process and without whom this project would not have been possible.

My deepest gratitude is to my adviser, Nancy Wallace who provided invaluable advice. The countless hours she spent with me helped this project evolve into what it is today. I am also deeply indebted to the late Dwight Jaffee who encouraged me from the very beginning, forcing me to ask important questions. His institutional knowledge ensured that I thought long and hard about the mechanisms behind my findings. He will be fondly remembered.

Ulrike Malmendier provided invaluable perspective and has been encouraging and supportive throughout this process. David Sraer provided support and encouragement.

I am also grateful for deep discussions with Amir Kermani, Gustavo Manso, Christopher Palmer, Christine Parlour and Hoai-Luu Nguyen. I have learnt the most from my fellow PhD students who have taught me much: Carlos Avenancio, Sanket Korgaonkar, Aya Bellicha, Isaac Hacamo, Paulo Issler, Chris Lako, Haoyang Liu, Pratish Patel, Abhay Aneja and Jikakai Chen. I also want to acknowledge the support provided by the staff at the Fisher Center of Real Estate and Urban Economics and the Institute of Business and Economic Research. Many thanks to Thomas Chappelear, Kim Guilfoyle, Melissa Hacker, Bradley Jong and Charles Montague for their amazing administrative assistance over the past six years. I also want to thank Paul and Marsha at Masse's for their affection and their endless cups of coffee which helped me get through many long days.

Lastly, I would also like to thank my family for their love and kindness. My sincere thanks to my mom, dad, mother-in-law and father-in-law who have shown endless love and support. My sisters Tejaswini Kulkarni and Chitra Rentala have always been generous with their time and love. Lastly, I would also like to thank my husband, Vikram Rentala for always being there for me.

Chapter 1

Uniform Pricing and Credit Rationing

1.1 Introduction

Federal policy often institutes uniform or pooled pricing across regions in the name of fairness. The US Postal Service, for example, exhibits limited geographic variation in prices. The rationale behind these policies is that consumers in areas with low costs should cross-subsidize consumers in areas with high costs. Such policies, however, may also have the unintended consequence of rationing consumers out of the market in some regions but not in others. I analyze these trade-offs in the context of the US residential mortgage market, which represents nearly 30 percent of the nation's credit market. The Federal National Mortgage Association ("Fannie Mae") and the Federal Home Loan Mortgage Corporation ("Freddie Mac") with their special status as government sponsored enterprises (GSEs) heavily influence borrower access to credit through their role in the secondary market for residential mortgages. They fund originated mortgages by purchasing loans directly from the primary market mortgage originators and holding these loans in their portfolio; or alternatively, by acting as conduits and issuing mortgage-backed securities (MBS). While the GSEs do not directly lend to borrowers, they can discourage regional risk-based pricing in the mortgage market by only accepting mortgage loans that they evaluate to be GSE-conforming.[1] I exploit variation in state foreclosure law at state borders to examine the direct and spillover effects of such uniform regional pricing of GSE-conforming mortgages in the US residential mortgage market.[2]

State foreclosure law that limits lenders' recourse and ability to foreclose on property affect lenders' payoffs from the mortgages in case of borrower default. Motivated by this, I build a lender rights (henceforth, LR) index based on state foreclosure law and use it as a source of regional variation in my empirical research design. I focus on three specific types of foreclosure law: judicial foreclosure laws, deficiency judgement laws (or recourse) and state

[1] Origination of new mortgages is done by the primary market mortgage originators and not the GSEs. The GSEs' charters prohibit them from originating mortgages ([49]).

[2] I refer to loans that are guaranteed or purchased by GSEs in the secondary market as GSE-eligible, GSE-conforming or simply conforming. Loans which are not GSE-conforming are referred to as non-conforming.

redemption laws. I use variation in LR at state borders combined with plausibly exogenous variation from a policy regulation based on affordable housing goals, a novel source of quasi-experimental variation in mortgage interest rates, and unique rate sheet and loan level data to examine three sets of questions. First, I test whether mortgage interest rates of GSE-conforming loans vary across LR. Second, I examine whether the lack of regional variation in mortgage interest rates affects credit access (that is, the extensive margin) across LR. Additionally, in a regression discontinuity (RD) design focused on the marginal borrowers, I also test whether the GSEs "cherry-pick" the better risks leading to greater credit access to marginal borrowers in high LR areas. Finally, I examine the interaction between the GSE-conforming and the private market in a bunching design based on the regulatory cutoff on loan size for GSE-conforming loans. I conclude by examining the overall impact on credit access in the primary mortgage market.

I begin by examining the mortgage rates of GSE-conforming loans during the pre-crisis period from 2000 to 2005.[3] Specifically, I focus on the 38 cross-border MSAs that straddle the borders of two or more states with differing LR. I use unique rate sheet data on mortgage rates offered by lenders *before* screening and borrower sorting to show that — holding borrower and property characteristics constant — mortgage rates of GSE-conforming loans do not vary across LR.[4] Lenders do not use non-rate terms such as down payment and debt-to-income constraints, either, to adjust for this lack of variation in mortgage rates. The first innovation of this study is that it uses *regional* variation in LR across state borders combined with unique rate sheet data to credibly establish that — controlling for borrower characteristics — mortgage rates of the GSE-conforming loans do not vary by LR. Prior literature ([44]) has found consistent results but was constrained to ex-post mortgage level data which are confounded by lender screening and borrower sorting.[5]

The paper's second set of results examines whether this lack of regional variation in interest rates of GSE-conforming mortgages affects the extensive margin of credit, that is, whether borrowers are denied access to GSE-conforming loans across LR. I find that within the same MSA-year cell, the total number of GSE-conforming mortgages per housing stock was 17 percent higher for a 1 SD higher LR index.[6] That is, the GSEs credit ration borrowers in areas with low LR. One channel through which the GSEs can institute this credit rationing on the extensive margin is through their credit evaluation models which make a pass or fail decision on whether a mortgage is GSE-conforming. Importantly, credit rationing by the GSEs persist (and is statistically more significant) for the more credit-constrained borrowers

[3] I focus on the period between 2000 to 2005 to avoid the confounding effects of the crisis. The baseline results hold even when I look at all MSAs or also include 2006 and 2007 data.

[4] The rate sheet data from `RateWatch.com` surveys lenders on mortgage rates for a loan of a given size. Thus, I control for borrower and property level characteristics. Year-quarter fixed effects ensure I also control for macroeconomic variables.

[5] [44] show that mortgage rates (after controlling for borrower characteristics) do not vary with MSA-level variation in default rates.

[6] Specifically, I focus on the mortgages purchased or guaranteed by the GSEs in the secondary mortgage markets. Denial rates in the primary market show similar results.

such as minority, younger, single female borrowers and first-time home buyers. For minority borrowers, a 1 SD higher LR index is associated with a (higher) 22 percent total number of GSE-conforming mortgages per housing stock.

A simple credit rationing argument suggests that, given uniform regional pricing of mortgage rates, credit rationing should bind more for the marginal borrowers. To explicitly examine whether the GSEs "cherry-pick" loans with better risk, that is, those in high LR areas and especially so for the marginal borrowers, I exploit a regulation that obliges the GSEs to increase their business in underserved areas. Specifically, under the "Underserved Area Goals" (UAG), the GSEs are obliged to assist mortgage funding for low-income and minority neighborhoods in the secondary mortgage market. As [51] note, the underserved area goals were (in spirit) meant to unify the geographic distribution of mortgage funding.[7] Using an RD design based on the regulatory cutoff on the median family income at the census tract level, I show that number of GSE-conforming mortgages per housing stock in low-income neighborhoods was 54 percent *higher* in high LR (above median LR index) areas where the GSEs have higher payoffs. By contrast, the number of GSE-conforming mortgages per housing stock in low-income neighborhoods in low LR (below median LR index) areas did not increase discontinuously at the RD cutoff. The RD estimates show that there is a large impact of the GSE housing goals in high LR but not in low LR areas, indicating that marginal borrowers in low LR areas were especially impacted by the GSE credit rationing.

These findings on regional credit rationing (especially for marginal borrowers) due to the regional uniformity of mortgage rates of GSE-conforming loans is an important contribution of this paper. Regional uniformity of mortgage rates may result in borrowers inefficiently self-selecting into mortgages. On the other hand, it also helps regional risk-sharing among borrowers. Prior literature ([44]) recognizes this and has focused on the cross-subsidization of borrowers in high-risk areas by borrowers in low-risk areas, a theme prominent in policy and public debate. While the benefits of cross-subsidization hold for the high credit-quality borrowers who pass certain credit-quality thresholds and successfully obtain a mortgage, my results show that the GSEs regain some of the efficiency loss of uniform pricing by restricting credit access in the conforming market to the lowest credit-quality borrowers. This has the unintended consequence of cutting off credit-market access (in the conforming market) to the lowest quality borrowers. As the RD estimates show, these results are especially stronger for marginal borrowers, specifically for low-income neighborhoods targeted by the GSE housing goals.

The paper's third set of results analyzes how the uniform pricing of GSE-conforming mortgages and subsequent credit rationing impacts the behavior of the private securitizers and lenders that do not securitize through the GSEs. During the 2000 to 2005 sample period, private securitizers also rapidly increased their market share in the secondary mortgage

[7] In 1992 Congress passed the GSE Act to increase homeownership to low-income and minority neighborhoods. The 1992 GSE Act mandates that a certain percentage of GSE mortgage business be composed of loans targeted to underserved groups. The GSEs satisfy this mandate by either creating and guaranteed MBS; or by purchasing and holding whole mortgages and MBS in their on-balance retained-mortgage portfolios.

market. I exploit a regulatory cutoff on the size of loans (also known as the conforming loan limit, or CLL) above which the GSEs cannot purchase or guarantee loans and first establish that the comparative cost of funds advantage of the GSEs are passed onto borrowers.[8] Mortgage rates of GSE-conforming mortgages are lower by 23 basis point in high LR (above median LR) areas and by 40 basis points in low LR areas, consistent with prior literature ([55], [62] and [6]). This discontinuous increase in mortgage rates around the CLL creates a "notch" in the borrower budget constraint. Borrowers who in the absence of the regulatory cutoff would have chosen a loan above the CLL, now choose to reduce the size of the loan and locate exactly at the CLL. Given the higher notch in interest rates in low LR areas we would expect (in the absence of credit rationing) higher bunching of borrowers at the CLL. However, credit rationing of GSE-conforming loans implies that only the higher quality borrowers are able to access credit in the GSE-conforming market. The observed bunching is indeed lower (rather than higher) in low LR areas consistent with GSE credit rationing. Analysis of bunching around the CLL clearly shows how the comparative advantage of the GSEs combined with credit rationing of the lower quality borrowers helps the GSEs establish their dominance in the non-jumbo market.

Next, I examine whether borrowers rationed out of the conforming mortgage market are able to access credit in the non-conforming market.[9] I test whether the presence of private securitizers diminishes the impact of uniform pricing and subsequent GSE credit rationing. I find that, while some of the borrowers rationed out of the GSE-conforming market are able to access credit in the private market, the lowest quality borrowers in low LR areas are rationed out of the non-conforming market too.

I also document an effect on the design of contract terms. [7] and [61] point to the supply of complex mortgages — such as Interest Only (IO) loans and Negative Amortization (NEGAM) loans – as a reason for the rise in risky lending. Here, I find that 1 SD higher LR index implied borrowers were 3 percent more likely to have an adjustable rate mortgage (ARM) and 4.2 percent more likely to have complex mortgage compared to a fixed rate mortgage (FRM) among the GSEs. Private securitizers thus seem to have taken into account the higher payoffs (in case of borrower default) in high LR areas in their contract terms. I next examine the market segment above and below the CLL and include the GSE-conforming FRM mortgages. A graphical analysis shows that there were fewer complex mortgages or ARMs as a percentage of total loans in high LR for the market segment where the GSEs can participate, that is, loans with size below the CLL. By contrast, there were more complex mortgages or ARMs in high LR areas for loans above the CLL (the market segment where the GSEs cannot participate). I also find that 1.5 percent more likeley to be purchased

---

[8]See [52] for a survey of the literature estimating the effects of GSE funding advantages on mortgage rates.

[9]For example, Franklin Raines, the then Chief Executive Officer (CEO) of Fannie Mae was quoted in a New York Times Article saying "Fannie Mae has expanded home ownership for millions of families in the 1990's by reducing down payment requirements. Yet there remain too many borrowers whose credit is just a notch below what our underwriting has required who have been relegated to paying significantly higher mortgage rates in the so-called subprime market."([1])

in high LR areas. Additionally, GSE loans were more likely to have high FICO and low debt-to-income and especially so in high LR areas. Thus, possibly borrowers rationed out of the conforming mortgage market took on increasingly risky loans (in terms complexity of loans). My third contribution is highlighting this potential indirect spillover effect of credit rationing (because of the uniform mortgage rate policy) in the conforming market on the private market.

These results on spillover effects of the GSEs also shed new light on the recent literature on change in lending behavior during the pre-crisis period. [65] and [64] point out that supply of mortgages increased to low-income zipcodes during this period whereas [4] and [3] argue that it was the relatively high-income borrowers who received these loans. [71] and [56] point to expansion of the private securitization markets as the cause of the subprime crisis. Although, my study ends before the crisis period, my results highlight possibly the importance of spillover effects that the GSEs exert on the private market through their influence in the secondary mortgage market, especially through their uniform mortgage rate policy. Borrowers rationed out of the conforming mortgage market perhaps had to turn to the private securitizers who pushed them into increasingly risky loans. In other words, the private market competed with the GSEs possibly by moving down the credit curve of increasingly risky mortgage loans.

To summarize, this study's results document that the uniform mortgage policy of the GSEs has important consequences for regional differences in credit access, especially so for marginal borrowers. Additionally, the uniform pricing of GSE-conforming mortgages and subsequent credit rationing combined with the GSEs' comparative advantage also influences the private market.

The rest of the paper is organized as follows. Section 1.2 reviews the literature and the institutional details behind uniformity of mortgage rates of the GSE-conforming loans. Section 1.3 lays out the conceptual framework for the empirical analysis. Section 1.4 details the paper's empirical methodology and describes the data. Section 1.5 analyzes mortgage rates and the extensive margin of GSE-conforming mortgages. Section 1.6 studies the interaction of the GSE-conforming loans and the private market. Section 2.8 concludes.

1.2 Institutional Details

In this section, I briefly discuss the institutional details of the residential mortgage market and the reasons why there is uniformity in mortgage rates of the GSEs. The GSEsare the two largest sources of housing finance in the secondary mortgage market. The GSEs do not directly lend to borrowers and thus do not directly set mortgage rates. Instead, mortgage rates and other contract terms are determined by originators. The originators then enter mortgage level information into a GSE credit evaluation model which then determines whether a mortgage is a "conforming" mortgage. However, the GSEs determine the contract design and underwriting standards for the loans they acquire or guarantee. The credit evaluation process is automated through proprietary software, namely, "LoanProspector" for Freddie

Mac and "Desktop Originator and Underwriter" for Fannie Mae. The mortgage originator enters information about the mortgages and the GSE evaluates whether the mortgage passes the GSE-eligibility standards. This is one channel through which the GSEs could discourage regional risk-based pricing and make decisions to ration borrowers in certain areas.

Mortgages that do not pass the underwriting standards of the GSEs trade in the non-conforming market. One important criteria for a mortgage to be conforming is regulatory restrictions on loan size: the conforming loan limit or CLL. Conforming loan size is required to be lower than the CLL in order to be eligible to be held or guaranteed by the GSEs. Mortgages that do not meet this criteria can only be purchased by private securitizers. I use this loan size threshold in the empirical analysis on bunching, which will be described later.

One important feature of the mortgage market is that there exists an active liquid forward contract market for agency Mortgage-Backed-Securities (MBS). This is known as the To Be Announced (TBA) market. In the forward market, the seller of the MBS agrees to sell at a certain pre-specified price without fully specifying the details of all mortgages to be included in the MBS. The lender is thus able to lock in the price at which he sells even before he originates the loans. At a given time, TBA trading is in only about 10 to 20 different sets of contracts for each given maturity ([80]). The TBA-eligible securities are "pass-through" securities where principal and interest payments of the underlying mortgages are paid to investors on a pro-rata basis. Only a few terms such as the coupon rate and face value of the mortgages are agreed upon.[10] Thus, the most important characteristic of the TBA market is that trading in the TBA market is reduced from millions of mortgages to a few contracts, despite the large heterogeneity of the underlying securities.

Adverse selection between the TBA seller and the TBA buyer implies that the TBA buyer knows that the loans he receives will be of lower quality than the average security. The security will thus trade at a cheapest-to-deliver discount.[11] However, the lack of information at the time of transacting is *also* thought to increase liquidity in the TBA market. [37] present a theoretical model which suggests that limiting information may actually enhance liquidity despite the negative effects of adverse selection. Thus, the lemon's discount due to the limited information in the TBA market is outweighed by the liquidity benefits of the TBA-eligible securities.

As [80] note, investors' treatment of TBA pools as fungible is predicated on the assumption of homogeneity of the underlying securities. The underwriting and pooling process for conforming mortgages ensures that all conforming mortgages delivered in TBA-eligible contracts pass a minimum quality threshold. I hypothesize that the market convention of treating the MBS pools and the underlying mortgages as homogenous possibly results in regional uniformity of mortgage rates. Although it is difficult to establish just how much

[10] The six general parameters agreed upon are: issuer, maturity, coupon, price, par amount, and settlement date.

[11] [25] provide evidence that the TBA market functions as a market for lemons. They show that borrowers in the MBS class that trade on the TBA market exercised their prepayment options more efficiently — to the detriment of the MBS investors — compared to borrowers in pools that could have transacted on the TBA market but did not.

homogeneity of underlying mortgages is necessary for the smooth functioning of the TBA market, the homogeneity of mortgages in TBA-eligible pools is an important ingredient. I hypothesize that this homogeneity of mortgages in the TBA pool is also responsible for the limited regional variation of GSE mortgage rates that I observe in my empirical analysis.

At any given point there are only a handful of TBA pools with different coupon rates. Thus, one could argue that regional differences in mortgage rates could be accommodated with these differences in mortgage rates at the TBA level. However, the liquidity of the TBA pools comes from the lack of information of the underlying mortgages. As [37] note, when an asset is illiquid and in the presence of asymmetries of information bundling (TBA pools) reduced information disclosure is optimal. [37] note that we do not see separate pools for New York and Houston mortgages even if there are real reasons for Houston markets to be superior to New York markets. The liquidity of the TBA pools comes from restricting information about the underlying mortgages. Specifically, my hypothesis for regional uniformity of mortgage rates stems from the fact that allowing mortgage rates to vary by region may also result in certain TBA pools (for a given coupon) being concentrated in certain geographies. Potentially, the liquidity of the TBA markets comes from this lack of information of the geographic composition of the mortgages in the TBA pools. Hence, we see uniformity across regions —controlling for borrower characteristics — of all GSE-conforming mortgages. Without this uniformity across regions the liquidity benefits of the TBA market may, arguably, be lower.

One concern in interpreting the lack of regional variation in mortgage rates of the GSEs is that the GSEs can account for regional differences by varying guarantee fees (g-fees). For investor-held MBS, the GSEs guarantee timely payment of principal and interest to the investor. As compensation for providing this guarantee, the GSEs charge lenders g-fees. Thus, GSEs could potentially address the lack of regional variation in mortgage rates by varying the g-fees to account for differences in lender rights. However, during the period of analysis between 2000 to 2005, it seems there was no regional variation in g-fees.[12]

Existing literature has suggested other reasons for the lack of regional variation of mortgage rates of the GSE-conforming loans. [30] theoretically motivate the uniform pricing policy of the GSEs as a way to deter competitors from reverse-engineering their proprietary screening technology. [59] theoretically motivates how it is optimal for a monopoly intermediary to only reveal whether quality is above some minimum threshold. While it focuses on certification intermediaries, the [59] model can be used to motivate why the GSEs use a zero/one decision in the credit evaluation process and do not explicitly state the reasons for a mortgage failing the credit evaluation process. Political pressure may also result in regional uniformity of GSE-conforming mortgages ([44]). For example, in 2012 the Federal Housing Finance Authority (FHFA) announced that guarantee fees would vary by location. The new guarantee fees would have accounted for difficult market conditions and declining

[12]This document from the Federal Housing Finance Agency (FHFA) states that "The Enterprises traditionally have charged similar guarantee fees nationwide for securitizing single-family mortgages with similar characteristics. Those fees are intended to cover all costs associated with the guarantees." `http://www.fhfa.gov/AboutUs/Reports/ReportDocuments/20131209_StateLevelGfeeAnalysis_508.pdf`

house prices in some states, and would have possibly resulted in higher mortgage interest rates for borrowers in these states. However, the decision to vary guarantee fees for certain states was back-tracked after facing public opposition. It is plausible that a political economy story could explain the regional uniformity of mortgage rates of GSE-conforming loans.

1.3 Conceptual Framework

In this section, I present a conceptual framework to analyze the impact of a notched mortgage interest rate on the bunching of loans at the CLL in the presence of GSE credit rationing. To analyze the structure of underwriting and pricing policies observed in US mortgage markets, I also analyze lender outcomes under three different pricing regimes and the subsequent impact on credit rationing when there is regional variation in lender payoffs or more specifically in LR. Note, that I will refer to "lender" outcomes and lenders "'varying" mortgage rates and abstract away from the fact that the GSEs do not directly lend to borrowers. Instead, "lending" by the GSEs (or non-GSE securitizers) refers to the lenders (originators) in the private market that securitizer through the GSEs (or non-GSEs). The full detail of the pricing regimes of the lenders (GSEs and the private market) is provided in the Appendix and I briefly describe the intuition below in Section 1.3.

The conceptual framework in this section mainly focuses on the bunching analysis. Essentially, borrowers face a jump in mortgage rates at the CLL. The notch in interest rates — and subsequently in the borrower budget constraint — induces borrowers to bunch at the CLL. I use this notch in the borrower budget constraint to develop a bunching model (see [23], [27], [15] and [58]). Specifically, I exploit the regulatory requirement that GSEs only purchase or guarantee loans below the CLL.[13] This generates an exogenous variation in the relationship between loan size and interest rates which results in borrower response to bunch around the CLL. The observed bunching will, however, depend on the pricing regimes followed by the monopoly lender. That is, whether or not the GSEs credit ration borrowers on the extensive margin will impact the observed bunching of loans at the CLL. We will also look at the impact of the GSE funding advantage on the mortgage rates around the CLL for loans above the CLL (jumbo loans) and loans below the CLL (GSE-conforming loans) also referred to as the jumbo-conforming spread.

Model Setup

The residential mortgage market structure is as follows. Retail lenders originate loans for the GSEs. Retail loan originators apply underwriting criteria imposed by the GSEs in the secondary market and charge a loan rate determined by the secondary market. For the rest of the exposition I abstract away from this additional layer between retail originator and the GSEs. After originating the loan, an originator will feed the information into GSE credit

[13] [23] and [14] also look at notches in the household budget constraint to estimate the interest rate elasticity of mortgage demand. However, neither accounts for credit rationing by the lender.

evaluation software which then evaluates a pass or fail decision as to whether the mortgage is conforming. Thus, while the GSEs do not directly lend to borrowers, I will abstract away from this additional layer and refer to GSEs "lending" to borrowers and the GSEs "varying" mortgage rates to refer to this pass/fail decision as a channel through which the GSEs can institute their pricing decisions.

The borrower has credit quality summarized by θ with $\theta \in [0,1]$. $1-\theta$ is the probability that the lender receives zero payoff. θ can be thought of as a reduced form value capturing the payoff in case of borrower default. Alternatively, θ can be thought of as the output from a credit evaluation model which depends on borrower quality, macroeconomic conditions, state foreclosure laws and any other variables that may affect credit quality. θ can vary along many dimensions. For the purposes of my empirical analysis, θ also encapsulates differences in state foreclosure laws.[14]

Lenders are risk neutral. Loan screening in the pricing regimes with screening occurs as follows. The lender gets the borrower characteristics data and then combines it with other information — such as, macroeconomic variables and state foreclosure laws — as an input into a proprietary credit evaluation model to determine credit quality of the borrower, θ. Given θ which is determined from this screening process, the loan could be denied in which case the applicant (potential borrower) is rationed out of the credit market. Alternatively, the loan could be approved and credit is offered with the loan size determined by the demand schedule, $D(r)$ derived from the household optimization problem (described below).

Borrowers are price takers. Borrowers/households are credit constrained and borrow to finance the purchase of their homes. Households purchase quality-adjusted housing h at price p per unit. To buy a house of value ph, they borrow at an interest rate r (with $1+r=R$).[15] After making housing choices, households receive an income y, repay all their debts and consume their remaining income. The household interest rate (r) depends on its own credit quality (θ). The household cannot determine its own credit quality which depends on many factors including external macroeconomic conditions and state foreclosure laws. This is not an unreasonable assumption in the mortgage market. Borrowers do not play a repeated dynamic game and hence have limited ability in inferring their own credit risk. Thus, credit risk affects borrower demand only in the way it affects the mortgage interest rate. That is, I explicitly make the assumption that the borrower's own credit quality does not enter the household optimization problem except in the way it affects the mortgage interest rate r. [11] make a similar assumption on borrower inability to determine their own future payoffs, or in other words, borrowers do not know their own credit quality at the time of entering into a loan.

One of the implications of this assumption is that borrowers do not ex-ante take into account their ability to strategically default more in states with low LR where the lender

[14] For ease of exposition, I will later explicitly represent θ as a function of LR (H for high LR and L for low LR, $H > L$) and all other remaining variables are denoted by X.

[15] In reality borrowers need to make a down payment financed from their liquid assets. However, if we assume the standard down payment of 20 percent this only adds a constant factor to the analysis and I abstract away from this for now.

might not pursue foreclosure given the high costs to the lender. The empirical literature supports this assumption. [41] find that ex-ante borrowers do not know whether a lender can pursue recourse.[16] [66] find that default rates during the crisis were the same for judicial and non-judicial states implying borrowers do not take into account the longer foreclosure timelines in judicial states and hence do not strategically default more in judicial states. Additionally, [34] finds that delinquencies are not higher in non-recourse states compared to recourse states and differ only ex-post on whether the mortgage is underwater (negative equity).

Lenders and Credit Rationing

Before we turn to the household maximization problem, I briefly describe the intuition for why uniform pricing of mortgages by the GSEs would imply credit rationing. To analyze the structure of underwriting and pricing policies observed in US mortgage markets, I analyze lender outcomes under three different pricing regimes and subsequent impact on credit rationing when there is regional variation in lender payoffs or more specifically in LR. The full detail of the pricing regimes of the lenders (GSEs and non-GSEs) is provided in the Appendix A.2. I begin by considering a monopoly[17] lender (GSE) under three pricing regimes for mortgage rates. As previously mentioned although the GSEs do not originate mortgages, they directly influence the observed mortgage rates and originations by retail originators through their operations in the secondary mortgage markets. The GSEs evaluate whether a mortgage is GSE-conforming (a pass/fail decision in the credit evaluation model) and effectively the GSEs lend directly to consumers. This allows us to abstract away from the complications of the intermediate retail market. I look at three pricing regimes: a risk-based pricing regime with screening, pooled pricing regime with no screening, and pooled pricing regime with screening. Given a particular pricing regime and household demand, I then show the impact on credit rationing of borrowers.[18]

The intuition for the results from the pooled pricing regime with no screening is as follows. The lender simply charges the mortgage rate corresponding to the mortgage rate corresponding to all borrowers in the market. Thus, the lender will not credit ration borrowers. Low credit-quality borrowers pay a lower rate than what they would have paid in a risk-based regime. Thus, efficiency in this regime is the lowest compared to the other pricing regimes.

The intuition for the results from the pooled pricing regime with screening is as follows. In this pricing regime, the lender simultaneously solves for the minimum credit quality

[16] [41] survey borrowers on whether they think their bank can come after the borrowers assets (in addition to the property securing the loan). Around 50% of borrowers said yes irrespective of the state they stayed in.

[17] The assumptions on cost of funds ensures the GSEs are able to dominate non-GSEs in both the risk-based and pooling strategies. Hence, I use the term "monopoly" to refer to the GSEs.

[18] The basic structure of the pricing regimes draws on [31] and [30] who use a similar framework to study the uniform mortgage rate policy of GSEs. [31] and [30] hypothesize that the GSEs use uniform pricing to deter competitive entry.

threshold above which he is willing to lend and the pooled pricing for all borrowers who cross the minimum credit quality threshold. If the credit quality threshold is low, more borrowers get a loan from the lender. In this case the pooled rate for all borrowers who get a mortgage in the screening pooled regime is higher, reflecting the lower quality of the pool of borrowers. However, if the threshold is high then the average credit quality of borrowers that the lender lends to is higher and this implies the pooled mortgage rate of borrowers is lower. Thus, the lender trades off a lower optimal credit quality threshold with higher mortgage rates. For exposition, let the minimum credit quality threshold in the screening pooled pricing regime be denoted by $\underline{\theta}^*$.

In Appendix A.2 I also incorporate the risk-based pricing regime that we observe for the non-GSEs. The assumption on cost of funds for the lenders (GSEs and non-GSEs) ensures that mortgage rates are higher for the non-GSEs compared to the GSEs.[19] Additionally, in risk-based pricing mortgage rates are higher for low quality borrowers. Thus the difference in mortgage rates of the conforming mortgages (GSE mortgages which have pooled pricing) and non-conforming mortgages (non-GSEs which follow a risk-based pricing) is wider for lower quality borrowers. Since low LR also implies low θ the jumbo-conforming spread in low LR areas is wider. We use this "notch" or jump in mortgage rates as one switches from a GSE (conforming) mortgage to a non-GSE (non-conforming) mortgage in the bunching analysis. Let us denote this difference between the conforming and non-conforming mortgage rates by Δr. Specifically, the difference in mortgage rates around the CLL is referred to as the jumbo-conforming spread (and is assumed to be Δr).

I now turn to the household maximization problem which determines mortgage demand.

Household Problem

Households derive utility from consumption and housing. Let c denote units of numeraire consumption good and have the following utility:

$$U(c,h) = c + \frac{A}{1+1/\epsilon}\left(\frac{h}{A}\right)^{1+1/\epsilon}$$

where $A > 0$ and $\epsilon < 0$ are parameters characterizing housing preferences. ϵ is the unconstrained price elasticity of housing demand. As mentioned above, I explicitly make the assumption that the borrower's own credit quality (θ) does not enter the household optimization problem except in the way it affects the mortgage interest rate ($R = 1 + r$).

The household's maximization problem is:

$$\begin{aligned} \max_{c,h} \quad & U(c,h) \\ \text{subject to} \quad & c = y + ph - Rph \end{aligned} \tag{1.1}$$

[19] This assumption is motivated by the large literature documenting the empirically positive jumbo-conforming spread in the mortgage market ([55], [62] and [6]).

Heterogeneity of borrowers is incorporated into the parameter A which is smoothly distributed with density function $f(A)$. The budget constraint depends on the household's income (y), the price of the house (ph) and the mortgage rate ($R = 1 + r$) on the mortgage taken out to finance the purchase. y, ϵ and p are constant across households. [20]

Solving the maximization problem, we can solve explicitly for housing demand h^*.

$$h^* = \left(\frac{p^\epsilon}{A^{\epsilon+1}}\right)(R-1)^\epsilon \tag{1.2}$$

The corresponding mortgage taken out to finance the purchase is given by $D^* = ph^*$. The mortgage demand is then given by

$$D^* = \left(\frac{p}{A}\right)^{\epsilon+1}(R-1)^\epsilon \tag{1.3}$$

We have assumed that p, y and ϵ are constant across households. For a given A, there is a one-to-one mapping between mortgage rates (R) and optimal mortgage choice (D^*). To make the exposition clearer, we denote mortgage demand across households as $D(R)$.[21]

Bunching Analysis

I now motivate the bunching of borrowers at the CLL. I focus on the household choice faced by borrowers close to the CLL. Loans below the CLL are eligible to be held by the GSEs whereas loans above the CLL cannot be held by the GSEs.

Borrowers at the CLL face a notched interest rate schedule as mortgage rates jump from lower mortgage rates for conforming loans (GSE-eligible) to non-conforming loans above the CLL (jumbo loans). In the previous section we denoted this difference between the conforming and non-conforming mortgage rates by Δr. Specifically, the difference in mortgage rates around the CLL is referred to as the jumbo-conforming spread (and is assumed to be Δr). In Section 1.3 we derived the optimal mortgage choice for mortgagors which depends on mortgage interest rates and household preferences. In the model, heterogeneity of borrowers was incorporated into the parameter A which is smoothly distributed with density function $f(A)$. In Equation A.1 we showed that there is a one-to-one mapping between mortgage rates (R) and optimal mortgage choice ($D^*(R)$). If we had a smooth mortgage rate schedule, the one-to-one mapping implies a smooth distribution of the mortgages at different loan amounts. Thus, the mortgage demand for each borrower, $D^*(R)$, is determined by the loan interest rate, R. We assumed the borrower demand ($D^*(R)$) depends on his own credit quality (θ) only through the way it affects the mortgage rate at which the lender is willing to lend.

If there were no notch in the mortgage interest rate schedule, given the smooth distribution of delta, the distribution of mortgage amounts would also be smooth. Let us denote this

[20] This assumption can be easily relaxed without materially affecting the results. All we need is that y, ϵ and p are smoothly distributed.

[21] Note, the elasticity of mortgage demand $\eta = -dD/dRR/D > 1$ given that $\epsilon < 0$.

using the density function $g_0(m)$. Now consider a notched interest rate schedule as observed at the CLL. Let $\bar{D}$ represent the CLL. Loans above the CLL have a higher mortgage interest rate (jumbo-conforming spread). Thus, the new $r(D) = r + \Delta r \mathbb{1}_{D>\bar{D}}$. Δr is the difference between jumbo and conforming loans where $r = R - 1$.

This notch in the budget constraint results in borrowers bunching at the conforming limit. Figure 1.9a illustrates the implications of this notch in a budget set diagram (Panel A) and density distribution diagrams (Panel B). The budget set diagram (depicted in (consumption (C), loan size (D))-space) illustrates the borrower responses among individuals with heterogeneous housing preferences A, but a specific demand elasticity ϵ. Individual A has the highest preference parameter (A) among those who locate at the CLL. This household (individual A) chooses the same loan size $\bar{D}$ both in the notch and pre-notch case. Individual B has the highest preference parameter $A + \Delta A$ among those who locate at the CLL. In the pre-notch schedule this household would have borrowed a loan of size $\bar{D} + \Delta D$. With the new notched schedule, the household B is indifferent between $\bar{D}$ and the optimal interior point D' under the higher non-conforming interest rate schedule. Thus every individual between $\bar{D}$ and D' locates at the notch point and there is a hole in the density distribution with a notched interest rate schedule. Any household which had an optimal mortgage size in the range $(\bar{D} + D')$ in the pre-notch schedule would choose to locate to $\bar{D}$ in the notched interest rate schedule. Thus, all these borrowers between $(\bar{D} + D')$ bunch at the CLL. In other words, all individuals with preference parameters in the range $(A, A + \Delta A)$ choose to locate at the CLL. Thus, in a frictionless world with no credit-rationing, the mortgage density will have bunching (excess mass) at the CLL ($\bar{D}$) and a “hole” of missing mass just above the CLL. This corresponds to the borrower response to the notched interest rate schedule. Also, a higher notch (larger spread in the jumbo-conforming spread) would also necessarily imply more bunching of borrowers.

The borrower response by borrowers facing a notched interest rate schedule is given by:

$$B = \int_{\theta=0}^{1} \int_{\bar{D}}^{\bar{D}+\Delta\bar{D}} g_0(m, \theta) dD d\theta \approx g_0(\bar{D}) \Delta \bar{D}$$

$g_0(m)$ is the counterfactual distribution with no notch. The above expression does not account for credit rationing by the GSEs. If there were no credit rationing by the GSEs, we can back out $\Delta \bar{D}$ to give an estimate of the response of borrowers to a notched interest rate schedule.

In the absence of credit rationing by the GSEs, borrowers with lower credit quality will have higher bunching compared to borrowers with high credit quality. Additionally, borrowers in low LR areas will have a higher jumbo conforming spread compared to borrowers in high LR areas holding all else equal. This is because non-GSEs follow risk-based pricing regimes and in the risk-based pricing regime high LR areas have lower mortgage rates compared to the low LR areas. The GSEs follow a pooled pricing regime and mortgage rates are the same in high LR and low LR areas. Thus, given the bunching analysis above, based on purely response by the borrowers and no credit rationing by the GSEs, bunching would be *higher* in low LR areas compared to high LR areas.

Under a pooled pricing regime with screening, borrowers below a certain minimum credit quality threshold are rationed out of the conforming mortgage market. Thus, the resulting credit distribution in the presence of GSE screening is given by,

$$\begin{aligned} J &= \int_{\theta=\bar{\theta}}^{1} \int_{\bar{D}}^{\bar{D}+\Delta\bar{D}} g_o(m,\theta) dD d\theta \\ &= B - \int_{0}^{\theta=\bar{\theta}} \int_{\bar{D}}^{\bar{D}+\Delta\bar{D}} g_o(m,\theta) dD d\theta \\ &\approx g_0(\bar{D})\Delta\bar{D} - \int_{0}^{\theta=\bar{\theta}} \int_{\bar{D}}^{\bar{D}+\Delta\bar{D}} g_o(m,\theta) dD d\theta \end{aligned}$$

Thus, the observed bunching at the CLL is the bunching that would have occurred in the absence of any credit rationing (the first term which I call the pure borrower response) minus the bunching amount of the low credit quality borrowers who are rationed out of the conforming (GSE-eligible) market (second term).

Note that in the presence of a notched interest rate schedule, the borrower response to the notch is highest by the poorest credit quality borrowers because the jump in interest rates is the highest for these low quality borrowers. The notch around the CLL reflects the jump in mortgage rates from the pooled GSE mortgage rate to the risk-based mortgage rate for loans above the CLL. Since in the risk-based pricing regime, mortgage rates are higher for poorer quality borrowers, the jumbo-conforming spread is highest for the poorer quality borrowers. Thus, in the absence of credit rationing, bunching estimates should be highest for the lowest quality borrowers. In the presence of credit rationing, however, bunching should reflect credit rationing by the GSEs. The overall effect on the observed amount of bunching, given that the GSEs credit ration, is ambiguous depending on which of the two effects is stronger the response of the borrowers or credit rationing by lenders.

This is more clearly explained in Table 1.1. Holding all else equal, borrowers in low LR areas have lower credit quality (θ) compared to high LR areas. This would imply that for low LR the jumbo-conforming spread is higher compared to high LR areas. Under a purely borrower response as in Equation 1.3, bunching should be higher in low LR ares. However, there is also higher credit rationing by the GSEs in low LR areas. Thus, overall observed credit rationing is the difference between the borrower response and the credit rationing by the GSEs and hence the overall effect on bunching is ambiguous.

Empirical Hypothesis

I now turn to testable implications from the model described above. For ease of exposition, I will put some structure on the functional form of θ. For simplicity, let us also explicitly assume that $\theta(X, LR) = \phi(X) * LR$. Thus, $\theta(X, LR)$ can be decomposed into a borrower specific component combined with other information such as macroeconomic variables $\phi(X)$; and a regional adjustment due to differences in LR (L/H). Consider two borrowers with the same credit quality along all dimensions ($\phi(X)$) except that they reside in two states with different state foreclosure laws. Holding all else equal, the overall credit quality (θ) of the borrower in the state with high LR will be greater than the borrower in a state with low LR, given $H > L$. The hypotheses below will compare borrowers given $\theta(L, X) < \theta(H, X)$.

We have two sets of hypotheses. The first set of hypotheses is for the GSE (which we refer to as the monopoly lender in the conceptual framework in Appendix A.2). The second set of hypotheses is for the non-GSEs or private securitizers (which we refer to as the competitor lender in Appendix A.2). As previously mentioned, though the GSEs do not lend directly to borrowers, we will refer to the GSE pass or fail decision in the credit evaluation model as a channel through which they can institute their pricing policies. "Screening" regimes below also refer to the GSE perceived quality of a borrower from the credit evaluation model. In the following sections I will refer to the GSEs "varying rates" or "lending" as a shorthand. Similarly, the non-GSE mortgages refer to the secondary private market purchases. We also refer to the market for the GSE-eligible loans as the conforming market and the market for the non-GSE mortgages as the non-conforming market.

Mortgage rates to differentiate between pooled and risk-based regimes

Hypothesis 1 *Under a risk-based pricing regime, holding borrower characteristics constant, mortgage rates will vary across LR. Under a pooled pricing regime mortgage rates will not vary across LR. Under a risk-based pricing regime, identical borrowers (same X) in areas with high (low) LR areas will have lower (higher) mortgage rates since $\theta(L, X) < \theta(H, X)$.*

In risk-based pricing regime lenders incorporate LR into the mortgage terms. Thus, in risk-based pricing higher LR corresponds to lower mortgage rates. In a pooling regime (both screening and no screening) borrowers will not incorporate LR into the mortgage terms. Thus, in pooled pricing regimes there should be no variation across LR. Note that we test the regional risk-based pricing versus screening and pooled pricing separately for the GSE-conforming loans and loans purchased by the private sector.

Credit rationing to differentiate between screening and no screening regimes pooling regimes

Under a pooling regime with screening for the GSE-conforming loans, lower quality borrowers are rationed out of the conforming mortgage market. Under a pooling regime with no screening, lower quality borrowers are not rationed out of the conforming mortgage market. In Section 1.3, I briefly described the intuition for credit rationing when the lenders who securitize through the GSEs follow a pooled with screening pricing regime. The GSEs simultaneously solve for a minimum credit quality of borrowers that they are willing to lend to $\underline{\theta}^*$ and then charge a uniform pooled mortgage rate for all borrowers that pass this minimum credit quality threshold. Given similar distribution of X in H and L, $\phi(X_H)$ is less than $\phi(X_L)$. Thus, more borrowers fall below the credit quality threshold in low LR areas and are subsequently rationed out of the conforming mortgage market. Under a pooling without screening model, there is no credit rationing and hence no borrowers are rationed out of the mortgage market by the GSEs.

Hypothesis 2 *Given a similar distribution of borrower characteristics (X) in high and low LR areas (H and L), borrowers in areas with low LR are rationed out of the mortgage market under a pooling regime with screening. Under a pooling regime with no screening there is no credit rationing in areas with low LR.*

Interaction between the GSE-conforming and private market around CLL using bunching analysis

To analyze the interaction between the GSEs and non-GSEs or more specifically between the conforming and non-conforming market we explicitly look at loans close to the CLL.

Hypothesis 3a *The GSE cost of funds advantage is passed onto borrowers and we observe a positive jumbo-conforming spread (difference in mortgage rates for loans above the CLL and mortgage rates for loans below the CLL) for all borrowers.*

Hypothesis 3b *If the GSEs-conforming mortgages exhibit pooled regional pricing and the private markets follow a regional risk-based pricing, the jumbo-conforming spread is higher in low LR areas compared to high LR areas for all borrowers.*

This hypothesis is necessary to establish that the mortgages rates of the conforming loans are lower than the mortgage rates of non-conforming mortgages for all borrowers in both high and low LR areas. This will give clear predictions for the bunching analysis below.

Hypothesis 4 *Under a pooled no-screening regime for GSE-conforming mortgages (without credit rationing), borrowers will be able to sort freely around the CLL and bunching of mortgages will be higher for low LR areas. Under a pooled with screening regime, borrowers will not be able to sort freely around the CLL and the net impact on observed bunching by LR is ambiguous.*

Under a pooled no-screening regime of GSE-conforming mortgages, borrowers in low LR areas face a higher jumbo-conforming spread. Thus, borrowers in low LR areas have a higher incentive to bunch at the CLL. A pooled, no screening regime predicts no credit rationing and borrowers can bunch freely at the CLL. Additionally, bunching will be higher for low LR areas.

Under a screening regime, borrowers in low LR areas face a higher jumbo-conforming spread. Again, borrowers in low LR areas have a higher incentive to bunch at the CLL. However, a pooled with screening pricing regime also predicts that borrowers below a certain credit quality threshold will be rationed out of the conforming mortgage market. Thus, the observed bunching around the CLL will be equal to the difference in the response of borrowers facing a jump in the jumbo-conforming spread (increased bunching) and the low quality borrowers who are rationed out of the conforming market (decreased bunching). The net impact on bunching is therefore ambiguous. Thus, if the impact of credit rationing is extreme, then bunching for low quality borrowers will be very low. If credit rationing is low, the borrower response dominates and the net effect is higher bunching. See Table 1.1 for a summary of this hypothesis by credit quality θ.

Spillover Effects of the GSEs

Hypothesis 5 *If the GSEs cream-skim, non-GSEs lend less to marginal borrowers in high LR areas under the assumption that the GSEs follow a pooled screening pricing regime.*

The first hypothesis follows from the fact that if the GSEs cherry-pick mortgages then the cost of funds advantage of the GSEs implies that credit quality of borrowers left in the private market is much lower. As a result, mortgage rates in the private market directly depend on the number of GSE-conforming mortgages. Additionally, these effects should be stronger in high LR areas where the number and volume of GSE-conforming mortgages is higher.[22] If there is extreme sorting (cream-skimming) or credit rationing by the GSEs, then the pool of borrowers in high LR is of very low quality. If there is no extreme sorting (cream-skimming) or credit rationing by the GSEs, then the pool of borrowers is similar in high and low LR areas. Thus, credit rationing by the non-GSEs is higher for the non-marginal borrowers.[23]

1.4 Estimation Framework and Data

In this section, I describe the empirical methodology of this paper. This section also looks at the data used and provides validation of the cross-border MSA strategy used in my analysis.

Estimation Framework

This study uses state-level variation in LR to study the impact on lending. In this section, I describe the empirical methodologies I use to determine the impact on mortgage terms and lending.

Cross-Border MSA Selection

The central focus of this paper is to empirically determine how state foreclosure laws affects the GSE-conforming and private market. To empirically identify the effect of the state foreclosure laws, I focus on the variation in state foreclosure laws that occurs at state borders. Some states have lower LR in terms of ease of foreclosure whereas other states grant higher

[22] In Appendix A.2 I explicitly show that non-GSEs credit ration too, but this is affected by GSE credit rationing.

[23] Note, Fair Lending laws limit the maximum mortgage rates that the non-GSEs can charge. Thus, the non-GSEs credit ration too.

rights to the lender in case of foreclosure. However, one concern with using state-level variation is that outcomes can vary across states for reasons other than just state foreclosure laws. To overcome this difficulty, I will focus on geographical areas — Metropolitan Statistical Areas (MSAs) — which are economically connected but straddle state borders. To avoid the confounding effects if the crisis, I focus on the period between 2000 to 2005.

41 MSAs in the continental United States straddle the border of two or more states. Of these, I focus on the 38 MSAs that have differing LR index on either side of the state border. Figure 1.1, Panel B shows the 38 MSAs used in our analysis. The MSAs are distributed across the United States but there is some concentration in the Northeast.[24]

Establishing uniformity of mortgage interest rates

As the baseline analysis, I first establish that LR is not incorporated into mortgage interest rates controlling for borrower characteristics. The main concern is the multi-dimensional nature of mortgage terms. Mortgage terms such as loan amount, mortgage rate, loan-to-value, debt-to-income are all simultaneously determined.[25] Additionally, all terms depend non-linearly on each other.

To address these concerns I first use rate sheet data. Prior literature has used ex-post loan level data to look at mortgage interest rates. Rate sheet data provide mortgage rates prior to extensive lender screening and borrower sorting and are thus not confounded by these factors. `Ratewatch.com` surveys each branch of a lender (institution) for mortgage rates of 30 year Fixed Rate Mortgages (FRM). I use the rate sheet data corresponding to the non-jumbo loans for the GSE analysis and the rate sheet data corresponding to jumbo loans for the private market analysis.[26] The mortgage rate data from these rate sheets provide clean unbiased estimates of the impact of LR on mortgage rates. To directly estimate the impact of LR, I run the following regression

$$Mortgage\ Rates_{b,MSA,s,t} = \alpha + \beta * LR_s + \delta_{MSA,t} + \epsilon \tag{1.4}$$

where b is the branch in an MSA in state s at time t. LR_s varies at the state level. I include MSA-year-quarter fixed effects ($\delta_{MSA,t}$). Thus, using the cross-border strategy, I am essentially comparing mortgage rates within an MSA-year-quarter cell. Standard errors are clustered at the state level, that is, the aggregation level of our variable of interest LR_s.

I supplement the mortgage rate analysis by looking at ex-post loan level data. To address concerns of non-linearity and simultaneity discussed above, I use a matching procedure to

[24] My baseline results hold even when all MSAs are included and the years 2006 and 2007 are included.

[25] [69] also looks at impact of these LR on loan size using data from the 90's ignoring changes along other dimensions.

[26] Typically, a rate sheet will have the base mortgage rate. The mortgage rate is then adjusted depending on a grid of FICO, LTV and DTI cutoffs. `Ratewatch.com` does not survey for this additional adjustment. However, the base mortgage can be thought of as the mortgage rate holding all other borrower and loan characteristics constant. [77] use this data to look at how concentration of lenders in the primary market affects transmission of monetary policy.

first find comparable mortgages on either side of the state border. Traditional matching methods such as propensity score matching and nearest neighbor matching will not work in this case since they implicity need to account for the pricing of various mortgage terms. For example, the propensity score matching method uses a probit or a linear probability model to assign a propensity score to the probability of treatment or in this case being in a high LR area.

To overcome this, I use Coarsened Exact Matching (CEM) from [45] and [46]. In this matching procedure, I bin mortgage terms and then match mortgages exactly within those bins. Note, the advantage of this method comes not from the simple discretization (binning) but the fact that I do not need to reduce these binned values to a single dimension such as a propensity score or Mahalanobis distance. This allows us to abstract away from the way all the mortgage terms depend on one another. I exactly match mortgages within bins of loan-to-value (LTV), credit score (FICO) and debt-to-income (DTI) across the state-border within a particular MSA-year-quarter. One additional advantage of this method is that it corresponds to rate sheets and how originators screen borrowers while making the decision to loan. Thus, the matching method enables us to compare mortgage rates separately for GSE-conforming mortgages and private market mortgages in comparable loans across state borders (but within MSA-year-quarters) using ex-post data.

Direct effects of uniform pricing on credit rationing

One of the main implications of the model in Section 1.3 and Appendix A.2 is that the uniformity of mortgage rates can impact credit rationing. Hypothesis 2 states that regional uniformity of mortgage rates may imply credit rationing of borrowers in low LR areas. In Section 1.5 I turn to analysis of the extensive margin, that is, do the GSEs credit ration. I look at both number and volume of mortgages per housing stock.

In the first step of the analysis I look at the mortgages in the secondary mortgage market either purchased or guaranteed by the GSEs. Similarly for credit rationing analysis of the private market I look at the purchases by private securitizers. To round out the analysis, in Section 1.6 I also look at the total originations in the primary mortgage market.[27]

First, I aggregate the number of loans (or total volume of loans) to the lowest level of geo-coding available within the time period of analysis. In the baseline analysis I aggregate to the census tract(-state) level within each year. A census tract is within one state. To compare aggregate credit, I normalize the aggregate number (or volume) of mortgages by size of housing stock.[28] My baseline analysis includes all borrowers. I also for different sub-samples such as those of young borrowers (age < 45), single female borrowers, first time home-buyers and minority borrowers

My hypothesis is that the empirically observed lack of variation in mortgage rates across LR for GSE-conforming mortgages implies that there is credit rationing in areas with low

[27]Focusing on the originations in the primary market also helps address any concerns that the analysis in this paper does not include mortgages held by banks on their portfolio.

[28]Analysis using number of mortgages per capita instead of per housing stock yields similar results.

LR compared to high LR, thereby, leading to higher credit in high LR areas holding all else constant.

Our first set of regressions test whether total mortgage volume is higher in high LR areas.

$$\begin{aligned} Ln(Total\ No.\ of\ Mortgages/Housing\ Stock)_{ct_s,MSA,t} &= \alpha + \beta * LR_s \\ &+ \gamma * X_{ct_s,t} + \delta_{MSA,t} + \epsilon. \end{aligned} \tag{1.5}$$

In the above equations the dependent variable is the logarithm of the total mortgage purchases to the total housing stock in each of our sub-samples in MSA MSA, census tract ct_s in state s at time (year) t. Note that a census tract is within a state and hence represented as ct_s. LR_s is our measure of LR at the state level and constant over time. I add MSA-year fixed effects ($\delta_{MSA,t}$) so that comparison is within a MSA-year unit. The controls included are percentage with less than high school education, percentage Hispanic, percentage black, logarithm of median income and unemployment rate from the Census 2000 data and Census 1990 data. Also, all regressions are weighted by population. The coefficient β gives the percentage increase in either total number of loans or total volume of loans per housing stock. As before, standard errors are clustered at the state level.

Cherry-picking by the GSEs: Regression Discontinuity Design

In this section, I motivate a regression discontinuity (RD) approach to estimate the cherry-picking of mortgages by GSEs in high LR areas. The goal is to identify areas where the GSE credit rationing will be particularly evident. I exploit a policy regulation that requires that the GSEs increase their purchases to underserved groups such as borrowers in low-income and minority neighborhoods. Focusing on these marginal borrowers implies that the GSE credit rationing will be more pronounced in these areas. I then estimate using a regression discontinuity design whether the GSEs increased their purchases of these borrowers *particularly* in high LR areas where they have higher payoffs in case of borrower default. I will first briefly describe the policy regulation that I exploit in my analysis and then outline the empirical approach for estimating the direct effect of uniformity of GSE mortgage rates on cherry-picking by the GSEs. I analyze cross-sectional data at the census tract level from 2000 to 2005 (as before)[29] using standard nonparametric regression discontinuity methods as explained in [48].

The GSEs are required to devote a percentage of their business targeted to underserved groups. The Department of Housing and Urban Development (HUD) monitors whether the GSEs are meeting these policy goals. A loan purchased by the GSEs can be classified as "Underserved Areas Goal" (UAG) targeted if the property is in a census tract with median family income less than or equal to 90 percent of the MSA median family income.[30]

[29]Note, I retain the panel structure for the RD analysis because UAG thresholds are determined based on the 1990 Census median family income (MSA and census tract) for years 2000 to 2002. For 2003 to 2005 the UAG thresholds are determined on the 2000 Census MSA median family income.

[30]A census tract is a much smaller geographical unit than an MSA.

This study exploits the fact that the eligibility of census tracts under the UAG changes discontinuously at this threshold of 90 percent. Census tracts with income ratios far above this threshold (high income census tracts) are likely to be very different from census tracts far below this threshold (low income census tracts). However, when I narrow our focus to the set of census tracts very close to the 90 percent threshold it becomes plausible that the impact on lending outcomes is determined by idiosyncratic factors and not by systemic differences in census tract characteristics around the threshold. Thus, under certain conditions census tracts just above the 90 percent threshold become good counterfactuals for census tracts just below the 90 percent threshold.[31]

Comparing lending in census tracts just above the eligibility threshold to lending just below the eligibility threshold will enable us to estimate the impact of the UAG goals on GSE purchases. In my empirical research design, I want to examine whether the UAG goals *differentially* affected GSE purchases in high LR areas compared to low LR areas. Hence, I look at the RD estimate separately for above median LR census tracts and compare it with below median LR census tracts. This gives me an estimate of whether the GSEs — which were mandated to purchase loans targeted to underserved groups — cherry-picked loans in areas with higher payoffs (high LR).

The UAG threshold is based on MSA median family income (relative to the census tract median family income) and restricting to cross-border MSAs provides an ideal setting to analyze cherry-picking by the GSEs. Additionally, our hypothesis is that credit rationing will bind more for the marginal borrowers. Since the UAG targeted areas are the low-income neighborhoods, the regression discontinuity analysis focuses on these marginal borrowers.

In order to perform the regression discontinuity analysis, it is helpful to first specify the regression form. I then examine the plausibility of the RD identifying assumptions. I restrict the data to a small window (also called the bandwidth) around the UAG threshold. Thus, only census tracts in a narrow window contribute to the estimate of the discontinuity. I choose a bandwidth of 1 percent.[32] The RD analysis estimates the following regression model within a narrow window (bandwidth) around the eligibility threshold:

$$\begin{aligned} y_{c,t} &= \beta_0 + \beta_1 * Targeted_{c,t} + \beta_2 * Targeted_{c,t} * f(spread_{c,t}) \\ &+ \beta_3 * (1 - Targeted_{c,t}) * f(spread_{c,t}) + X_{c,t} + \delta_{MSA,t} + \epsilon_{c,t} \end{aligned} \tag{1.6}$$

where the dependent variable is the outcome of interest in census tract c at time t. $Targeted_{c,t}$ is an indicator equal to 1 if the census tract is eligible to be classified as UAG targeted, that is, if the tract to MSA median family income ratio is less than or equal to 90 percent. The variable $spread_{c,t}$ is the running variable and is the difference between the tract to

[31] [10] uses a similar regression discontinuity approach to estimate the impact of the 1992 GSE Act on GSE purchases and finds the regulation increase GSE-eligible originations.

[32] The results are robust to other bandwidths of 5 percent, 10 percent and to using the [47] bandwidth. I choose a narrow bandwidth of 1 percent so as not to confound my estimates with another common threshold of 80 percent used in other policy regulations such as the Community Re-investment Act (CRA). The CRA threshold affects both mortgage lending and small business lending and is thus not suited for my purposes.

MSA median family income ratio and 90 percent. The analysis will examine robustness to different functional forms, $f(.)$ for the RD polynomial and different bandwidths. As the above equation shows, the RD polynomial is estimated separately on either side of the RD threshold. The baseline specification will use a quadratic polynomial. [32] recommend using a lower order polynomial since higher order polynomials may lead to misleading results. The RD specification is repeated separately for above median and below median LR areas. All specifications include MSA-year fixed effects ($\delta_{MSA,t}$) and baseline controls ($X_{c,t}$). Although it is not necessary to include controls for identication, the precision of the RD estimates improves when they are included. As before, standard errors are clustered at the state level.[33]

For the identification assumption to hold, it must be true that all relevant factors besides treatment vary smoothly at the UAG eligibility threshold. Formally, let the potential outcome if a tract is UAG eligible be y_1 and the outcome if a tract is not UAG eligible be y_0. The identification assumption requires that $E[y_1|spread]$ and $E[y_0|spread]$ are continuous at the UAG eligibility threshold.

I assess the plausibility of the identifying assumption in Figure 1.3. Specifically, I examine demographic and economic covariates of census tracts just below the 90 percent threshold (normalized to a spread of 0 as described above) to those just above the threshold. More formally, I examine the estimates of β_1 in Equation 1.6 with the following dependent variables: percentage with less than high school education, Hispanic, percentage black and unemployment rate. This is done separately for above median LR and below median LR. In Figure 1.3 I plot these RD estimates for above median LR (top half of the plot in red) and below median LR (bottom half of the plot in blue). I find that all covariates are balanced across the UAG threshold. As the graph shows, there is no difference in any of these covariates around the cutoff threshold. The estimated RD coefficient is almost a precise zero for all the demographic and economic variables in our analysis. Thus census tracts below the 90 percent threshold are good counterfactuals for census tracts above the 90 percent threshold. This is true for both the above LR and below median LR.[34]

Bunching Estimates

In Section 1.3, I showed that given a notch in the interest rate schedule, borrowers will bunch at the conforming limit. However, if GSEs credit ration borrowers, then the observed bunching will reflect the difference between the response of borrowers and the borrowers rationed out of the conforming loan market. Another way of saying this is that observed bunching reflects the borrower response of the *successful* borrowers who do manage to get a loan from the lender. I now show how this resulting amount of bunching or observed bunching

[33] Results are robust to other levels of clustering.

[34] In the Online Appendix I also provide estimates of the difference in means estimates of the covariates around the RD threshold. While most covariates are balanced, UAG targeted areas have a higher percentage of Hispanic population. Note, however, that for the RD assumptions to hold I only need that all relevant factors besides the treatment vary smoothly around the RD cutoff.

is estimated. This section follows the calculations in [23], [58] and [16]. Note, however, my estimates will reflect the bunching response after accounting for credit rationing by the GSEs which prior literature has not accounted for. The standard approach is to fit a polynomial to the observed distribution of the data after excluding a region around the bunching cutoff (the CLL in our case). The bunching estimates are then determined based on the observed distribution relative to the counterfactual distribution determined by the fitted polynomial.

Each loan amount in our data is measured as a percentage of the conforming loan limit. For simplicity let the conforming loan limit be centered at zero. A value of zero represents a loan exactly equal to the conforming loan limit. A value of 0.1 will refer to a loan 1.1 times the conforming loan limit. I then divide the normalized loans into bins around the conforming loan limit. Let these bins be centered at values m_j with $j = -J, ...L, ..., 0, ..., U, ...J$ and the total count in each of these bins be given by n_j.

I am interested in determining the bunching as the excess mass relative to the density around the notch (b). To get an estimate of bunching, I first need to estimate the counterfactual loan size distribution. The amount of bunching is then determined relative to this counterfactual density. One difficulty in measuring the observed bunching is that noise in measuring the loan amount[35] leads to a diffuse excess mass around the conforming loan limit threshold, 0. To measure b in the presence of such a notch, we must exclude the region around the conforming loan limit defined by $[m_L, m_U]$ and then fit the following regressions to the count of loans in each bin:

$$n_j = \sum_{i=0}^{p} \beta_i^0 (m_j)^i + \sum_{k=L}^{U} \gamma_k^0 1(m_k = m_j) + \epsilon_j. \tag{1.7}$$

The first term on the right hand side is a p^{th} degree polynomial in loan size and the second term is a set of dummy variables for each bin in the excluded regions $[m_L, m_U]$. Thus the estimate of the counterfactual distribution is determined by the predicted values of this regression after omitting the effect of the dummies in the excluded region. Thus, letting $\hat{n}_j$ denote the estimated counterfactual number of loans in each bin j, we can then write

$$\hat{n}_j = \sum_{i=0}^{p} \hat{\beta}_i^0 (m_j)^i. \tag{1.8}$$

The initial estimate of the observed bunching is then estimated as the difference between the observed and counterfactual bin counts in the excluded region to the left of the conforming loan limit (centered on 0)

$$\hat{b}_n^0 = \sum_{j=L}^{0} (n_j - \hat{n}_j^0) = \sum_{j=L}^{0} \hat{\gamma}_j^0 \tag{1.9}$$

[35] For example, for the GSE single-family FRM data, the loan amount is rounded to the nearest thousand.

while the amount of missing mass due to bunching is then given by $\hat{b}_n^0 = \sum_{j>0}^{U}(n_j - \hat{n}_j) = \sum_{j>0}^{U} \hat{\gamma}_j$

This calculation, however, overestimates $\hat{b}_n^0$ because it does not account for the additional mass from the right of the threshold, that is, it does not satisfy the constraint that the area under the counterfactual must equal the area under the empirical distribution.[36] To account for this problem, the counterfactual distribution is shifted upward to the right of the threshold until it satisfies the integration constraint. The counterfactual distribution is then defined as $n_j = \beta_i(m_j)^i$ as the fitted values from the regression:

$$n_j(1 + 1[j >= U]\frac{\hat{b}_n}{\sum_{j=U}^{\inf}}) = \sum_{i=0}^{p} \beta_i(m_j)^i + \sum_{k=L}^{U} \beta_i(m_j)^i \gamma_k 1(m_k = m_j) + \epsilon_j \tag{1.10}$$

while the amount of missing mass due to bunching is then given by $\hat{b}_n = \sum_{j>0}^{U}(n_j - \hat{n}_j) = \sum_{j>0}^{U} \hat{\gamma}_j$

Finally, I define the empirical excess mass around the notch relative to the average counterfactual density earnings distribution between $[m_L, m_U]$ as:

$$\hat{b} = \frac{\hat{b}_b}{\sum_L^U \hat{n_j}/(U - L + 1)} \tag{1.11}$$

In the bunching analysis, I use a 5^{th} order polynomial to determine the counterfactual distribution. In my analysis I determine the bunching estimates in Equation 1.11 for above median LR and below median LR separately. I use 1 percent bins around the CLL. The iterative procedure ensures that the difference between the excluded region $[m_L, m_U]$ so chosen minimizes the excess mass below the notch and the missing mass above the notch. I next turn to the empirical analysis.

Data

I use four main data sources described in the Online Appendix. Here I briefly describe the data sources and the variables used. For the full data construction details, please see the Online Appendix. I restrict my analysis to the period between 2000 to 2005 to avoid the confounding effects of the crisis.[37]

First, I have access to unique rate sheet data from `RateWatch.com` which provides mortgage rates at the monthly level. Prior literature has used ex-post loan level information to analyze mortgage rates. However, ex-post data are confounded by the effects of massive

[36] The technical issue that arises here is in determining $[m_L, m_U]$. In the case of notched interest rate schedule (as opposed to kinks) there is a diffuse mass above the notch which makes it difficult to visually determine the excluded regions. Hence, I use the structure imposed in [15] and [57] and follow an iterative procedure that ensures bunching mass equals missing mass.

[37] The baseline results remain the same when I include 2006 and 2007.

lender screening and borrower sorting. The rate sheet data are not confounded by these factors and give an unbiased estimate of the impact of LR.

For credit rationing, I use the Department of Housing and Urban Development (HUD) data on mortgages purchased by the GSEs. I use this dataset rather than the more widely used Home Mortgage Disclosure Act (HMDA) data for the GSEs since HMDA underestimates the amount of loans purchased by the GSEs in a given year. Additionally, since my analysis focuses on the 38 MSAs that cross state borders, I need good geo-coding of data. Hence my baseline analysis focuses on the HUD data. In the Online Appendix I replicate my results using the other datasets described below and the qualitative results remain the same. The HUD data has geo-coding at the census tract level. Census tract data for the years 2000 to 2002 is based on the 1990 Census delineations and for the years 2003 to 2005 is based on the 2000 Census.

To supplement the rate sheet data, I also look at ex-post loan level data provided by Fannie Mae and Freddie Mac on single-family 30-year Fixed Rate Mortgages (FRM). This provides good information on contract terms such as mortgage rate, loan-to-value (LTV) and debt-to-income (DTI). However, this has limited geocoding at only the MSA and 3-digit zipcode level and hence is unsuited for the credit rationing analysis. However, the detail on the contract terms make it ideally suited for ex-post mortgage rate analysis.

For the analysis of the non-GSEs or private securitizers, I use data from ABSnet Letwan. For non-jumbo, non-GSE data I construct a set comparable to the GSE data by restricting to 30-year FRMs with FICO greater than 620 and LTV less than 1.00 and restricting to the loans below the conforming loan limit. For the jumbo non-GSE data, I restrict my analysis to 30 year fixed rate mortgages with FICO greater than 620 and loan amount above the conforming loan limit (jumbo) and less than 2 times the conforming loan limit. To ensure that I am comparing the same set of locations and time periods, I restrict my analysis to MSA-year-quarter cells for which both GSE and ABSnet data exist.

For the control variables percentage Hispanic, percentage black, log(median income), unemployment rate and housing stock, I use data from the 2000 Census and 1990 Census based on the delineations in the HUD data. Appendix A1 lists the data and variables used from the different sources.

The choice of datasets in the main analysis is based on the level of geocoding for the credit rationing analysis and on the detail of the contract terms for the loan level analysis.[38]

State foreclosure laws and the Lender Rights index

To test whether regional differences in lender rights are incorporated into mortgage terms by the GSEs, I use state-level variation in foreclosure laws. I focus on three specific types of foreclosure law: judicial foreclosure laws, deficiency judgement laws (or recourse) and state redemption laws. See Appendix A.1 for detail on how the index os constructed. Here I briefly describe the main components of the index.

[38]Baseline results remain the same when I use the different datasets. Results are available in the Online Appendix.

I look at a number of different sources to classify the foreclosure laws. For the classification of states as judicial or non-judicial, I follow [33]. I use [34] for the classification of states as recourse or non-recourse. For the classification of states as fair-market-value or for equity right of redemption, I follow [72].

In judicial foreclosure states, a lender needs to go to court to foreclose on a property. This implies lower creditor rights compared to non-judicial foreclosure states, where the lender does not need to go to court to start foreclosure proceedings. The right to redeem enables a borrower to redeem or make whole the amount due, even after the foreclosure process has started. The time allowed to redeem varies from state to state, however, ranging from a mere 10 days to a full year.

Lastly, I also look at whether a state allows deficiency judgement (also called recourse). In non-recourse states, the lender only has access to the property securing the loan. In recourse states, borrowers can also access borrower wages and personal property. Additionally, I make one more distinction within recourse states which the previous literature has ignored. In fair market value states, the amount a lender can access is limited by the fair market value of the property. Thus, after foreclosure, the amount that can be pursued in case of a deficiency judgment is given by the shortfall between the debt and the fair value of the property. In non-fair-market-value states, on the other hand, the amount that can be pursued in case of a deficiency judgment is given by the shortfall between the debt and the foreclosure sale price of the property, which tends to be lower than the fair market value of the property. Thus, in non-fair-market-value states the lender can sell the property at a depressed foreclosure price and thus can claim a greater deficiency judgment. Another complication is that the statute of limitations varies from state to state. For the 11 states with the lowest statute of limitations, I down-weight the amount of recourse available to the lender.

I assign a value of between 1 to 3 for each mortgage law described above with higher values corresponding to higher lender rights. Adding these values I then create a simple index called the "lender rights index" or the LR index. I standardize this and use this continuous measure in my analysis. See Appendix A.1 for details on how the lender rights index is calculated. Lastly, I retain the 38 cross-border MSAs which differ in the LR index across state borders. The MSAs retained are shown in Figure 1.1

Summary Statistics and Covariate Balance

Table 1.2 shows the summary statistics of the variables used in our analysis. The LR index which we use as our main source of variation has been z-scored. The LR index ranges from -1.73 to 2.58. LR value of 0 corresponds to the median LR and a 1 SD higher LR corresponds LR value of 1.

Average mortgage interest rates in the rate sheet data is 6.62. This is slightly higher compared to the GSE 30 year FRM loan-level data which has an average mortgage rate of 6.46 with average FICO score of 715. Our non-GSE data for 30 year FRM from private label securities is even higher at 6.96 reflecting the slightly riskier borrowers with average FICO of 706 (lower credit score). Debt-to-income of the GSE data is also much higher at 34 compared to the non-GSEs at 9.[39] Average loan size is also slightly larger for the non-GSEs ($250,015) compared to the GSEs ($166,727).

Turning to the HUD data which we use for our census tract analysis, we see that a census tract had 120 loans on average with only 12 loans made to minority borrowers, 13 to first-time home buyers and 29 to single-female borrowers.

I next provide validation for using the cross-border MSA strategy. In Table 1.3 I regress each of the covariates we use in our analysis against LR index. Each regression includes MSA fixed effects and is clustered at the state level. We see that the variables high school education, percentage Hispanic, percentage black, percentage with less than high school education and unemployment are not statistically different for high LR areas. The coefficient on logarithm of median income while slightly higher is still statistically insignificant. The regression discontinuity design exploits variation in neighborhood level income and will thus address any concerns of potential bias arising from a sightly higher income of the high LR census tracts.

1.5 GSE-conforming market, Mortgage Rates and Credit Rationing

I now turn to the empirical analysis of the GSE-conforming mortgage market. First, I test the how mortgage rates vary across LR. The goal is to see whether there is regional variation in mortgage rates of GSE-conforming mortgages. Second, I test whether this uniformity of mortgage rates is reflected on the extensive margin of credit.

Mortgage Rates Across LR

Table 1.4 shows the results of our analysis. All columns include MSA-year-quarter fixed effects. All regressions are clustered at the state level. As a first step, I compare mortgage rates within the MSA-year-quarter using the unique rate sheets data. This ex-ante mortgage rates data is not confounded by lender screening and borrower sorting, so it gives us unbiased estimates of the differences in mortgage rates across areas with different LR. In Column 1, I show the results of the regression of mortgage rates on LR using the rate sheet data from `RateWatch.com`. I find that mortgage rates do not vary by LR – the point estimate is almost a precise zero.

I supplement the analysis by looking at ex-post loan level data on mortgage rates. Using a simple OLS regression I find that mortgage rates do not vary across LR. Additionally, I do not find evidence that lenders are adjusting on non-rate terms either in the level of down-payment/LTV (column 4) or in the debt-to-income ratios (column 3). The problem with using ex-post loan level data is that all mortgage terms can vary simultaneously and

[39] ABSnet has many missing values for debt-to-income. In the matching exercise that we will see later, I match on the missing value. Baseline results remain the same even if we exclude these values.

these terms depend on each other non-linearly. To overcome this simultaneity and non-linearity of mortgage terms, I use Coarsened Exact Matching (CEM) from [45] as described in Section 1.4. I match along DTI, LTV and FICO bins in a particular MSA with differing LR variable on either side of the MSA. These cutoffs correspond to rate sheets and how originators screen borrowers and provide loans. I use Coarsened Exact Matching (CEM) to find comparable mortgages in areas with high and low creditor rights to address non-linearity. I match along DTI (bins with cutoffs 20, 40), LTV (bins with cutoffs 70, 75, 80, 85, 90 and 95) and FICO (cutoffs at 620, 660 and 720) in a particular MSA-year-quarter cell with differing LR measure on either side of the state border. In the baseline analysis, I only retain the important cutoffs in the matching procedure and analyze the outcome variable mortgage rates. The trade-off of including more bins is that while we find a better matched sample, the sample size reduces considerably. Controlling for FICO and LTV in Column 5 shows that mortgage rates are only marginally statistically significant (only 1.6 basis point) lower in high lender right areas. That the matched procedure results in lower mortgage rates shows that the lack of variation in mortgage rates is accounted for along other dimensions, namely FICO and LTV, though this is not economically (or even statistically) meaningful. Next we turn to the analysis on credit rationing by the GSEs.

Additionally, in columns 3 and 4, I test for non-rate rationing as suggested by [26]. [26] suggest that lenders can use non-rate rationing, that is, even if mortgage rates do not vary other terms such as debt-to-income or down payment (LTV) can be adjusted. Column 3 and 4 suggests that this is not the case. Similar to our findings for mortgage rates, LTV and DTI do not vary across LR.

Credit Rationing across LR

Now I turn to analysis of the extensive margin, that is, does the regional uniformity of mortgage rates of GSE-conforming mortgages imply an adjustment on the extensive margin?

In Table 1.5, I test whether the GSEs (through their secondary market activity) took into account the regional variation in LR. First, I look at the impact on aggregate volume and aggregate number of GSE-conforming mortgages at the census tract level. I use panel data from 2000 to 2005 and normalize by housing stock. In panel A, I look at the number of GSE-conforming loans per housing stock. I find that total number of GSE-conforming loans per housing stock was higher by 17 percent in areas with 1 SD higher LR. However, this effect is only marginally significant. In columns 2–5 I focus on the more credit-constrained borrowers and find that the effects are much stronger when we restrict to these borrowers. The number of GSE-conforming mortgages is almost 22 percent higher for minority borrowers when LR is 1 SD higher. Additionally, effects are much stronger for younger, single-female and first time home buyers. These results are consistent with the GSEs pooling mortgages and credit rationing all borrowers who do not meet a minimum credit quality threshold. The minimum credit quality threshold does not bind for the relatively less risky borrowers (column 1) but does bind for the riskier or more credit-constrained borrowers (columns 2–5). The results are very similar when we look at total volume of GSE-conforming mortgages. We hypothesized using a credit rationing argument that the empirically observed lack of regional variation in mortgage interest rates should be reflected on the extensive margin, that is, in the aggregate GSE activity in the secondary mortgage market. This is what we observe in Table 1.5. Additionally, the results are stronger for more credit-constrained borrowers. We next turn to an RD analysis to explicitly answer the question: were the GSEs cherry-picking mortgages in high LR areas?

Cherry-picking by the GSEs using an RD Design

I now turn to an RD framework to assess cherry picking by the GSEs. In Section 1.4 I established the plausibility of the identifying assumption. I found that all relevant factors besides the treatment varied smoothly across the UAG threshold.

Remember, a census tract is classified as UAG targeted under the housing goals if the property (securing a mortgage) is in a census tract with median family income less than or equal to 90 percent of the MSA median family income. I defined spread as the difference of this ratio of census tract with median family income to MSA median family income and 90 percent. A spread of 0 corresponds to a census tract whose median family income is 90 percent of the median family income of the MSA. A negative spread corresponds to a census tract whose median family income is below 90 percent of the median family income of the MSA and is thus UAG targeted. A positive spread corresponds to a census tract whose median family income is above 90 percent of the median family income of the MSA and is thus not UAG targeted.

The four panels in Figure 1.4 graphically analyzes the relationship between GSE activity and UAG goals separately for above and below median LR against the spread. Negative spreads indicate that the census tracts are UAG targeted. Each point represents the average value of the outcome in spread bins of width 0.001. The solid line plots predicted values from a local linear regression, with separate spread trends estimated on either side of the UAG cutoff (spread of 0) and includes MSA-year fixed effects and control variables. The dashed lines show 95% confidence intervals with standard errors clustered at the state level. I use a bandwidth of 1 percent.[40] In Panel A, I restrict to census tracts above the median LR and in panel B I restrict to census tracts below the median LR. The dependent variable in Panel A (plot(a)) is the average of the dependent variable — the logarithm of the number of loans to housing stock — in spread bins of width 0.001 in high LR areas. The number of GSE-conforming loans per housing stock is significantly higher in UAG targeted areas for high LR areas. The dependent variable in Panel A (plot(b)) is the average of the dependent variable — the logarithm of the volume of loans to housing stock — in spread bins of width 0.001 in high LR areas. The volume of GSE-conforming loans per housing stock is significantly higher in UAG targeted areas for high LR areas. Next, Panel B examines the impact of UAG classification in below median LR. As before, the dependent variable in Panel B (plot(c)) is

[40]Results are similar with bandwidths of 5 percent, 10 percent and to using the [47] bandwidth. See the Online Appendix

the average of the dependent variable — the logarithm of the number of loans to housing stock — in spread bins of width 0.001 in low LR areas. The number of GSE-conforming loans per housing stock is not statistically different from zero in UAG targeted areas for low LR areas. The dependent variable in Panel B (plot(d)) is the average of the dependent variable — the logarithm of the volume of loans to housing stock — in spread bins of width 0.001. The volume of GSE-conforming loans per housing stock is again not statistically different from zero in UAG targeted areas for low LR areas.[41]

Table 1.6 examines this more formally with RD specification given by Equation 1.6 and estimates the magnitudes and standard errors of the discontinuities plotted in Figure 1.4. As before, all regressions include MSA-year fixed effects and include the control variables percentage with less than high school education, Hispanic, percentage black and unemployment rate. The dependent variable in columns 1–2 in Panel A is the logarithm of number of loans to housing stock. The dependent variable in columns 3–4 in Panel A is the logarithm of volume of GSE-conforming loans to housing stock. Column 1 and 3 show that the RD coefficients for the above median LR areas and columns 2 and 4 show the results for the below median LR areas. We see that the total number of loans is higher in UAG targeted areas as shown by the RD coefficient for the high LR areas in column 1. "Targeted" refers to the discontinuity at the RD threshold with a value of 1 for "Targeted" corresponding to census tracts below the 90 percent tract to MSA median family income. The number of loans per housing stock is almost 55 (β_1=-0.55, s.e.= 0.15)[42] percent higher for high LR areas. In contrast, for the low LR census tracts, the RD coefficient is not statistically different from zero (β_1=-0.453, s.e.=0.328). Thus, the GSEs satisfied the UAG goals by cherry-picking mortgages in areas with high LR where lender payoffs are *higher* in case of borrower default. The total volume of GSE-conforming loans per housing stock is consistent with this result and the point estimates are of similar magnitude and significance. Total volume of GSE-conforming loans per housing stock is 53 percent (s.e.=13) higher in high LR areas but not significantly different from zero in low LR areas (β_1=-0.453, s.e.=0.328) consistent with our hypothesis that the GSEs were cherry-picking mortgages in high LR areas.

Although the RD estimates for the low LR areas are not significantly different from zero in low LR areas, the point estimates are negative and large. One hypothesis may be that the GSEs were shifting their activity from low LR areas to high LR areas. To test this hypothesis I run a difference-in-difference specification by restricting to UAG targeted areas in Panel B column 1. I use the same set of census tracts corresponding to bandwidth of 1 percent in Panel A, Table 1.6. However, I now subset to UAG targeted in columns 1 and 3 in Panel B

[41]The non-linearity of the fitted polynomials (to the left of spread of 0 or in the UAG targeted areas) far away from the RD cutoff for high LR areas may suggest there is some shifting GSE activity around the RD cutoff. Additionally, the negative (though insignificant) drop in GSE activity for the UAG areas may point to the same hypothesis. I examine this below and do not find a mere shifting of GSEs activity from low LR to high LR. Note, however, in the RD context I am interested in the discontinuity *at* the RD cutoff.

[42]Note, the dependent variable is in logarithm and hence I explicitly express in percentage terms as it is easier to interpret. That is, β_1=0.55 corresponds to 55 percent increase in number of loans per housing stock.

and to non-UAG targeted in columns 2 and 4.[43] I find that contrary to a shifting hypothesis, the GSEs increase their activity in high LR areas in the UAG targeted areas (column 1), but did not decrease their activity in the non-UAG targeted census tracts (for low LR areas). In "Targeted" (UAG-eligible) census tracts a 1 SD higher LR index corresponded to 18 percent higher number of GSE-conforming loans per housing stock. For "Not-Targeted" (not UAG-eligible) census tracts a 1 SD higher LR index did not result in statistically significant increase in number of GSE-conforming loans (coeff= 0.12, s.e.=0.08) as shown in column 2, Panel B. Similarly, a 1 SD higher LR index corresponds to an increase in volume of GSE-conforming loans per housing stock for both the "Targeted" (β_1=0.242, s.e.=0.118) and "Not-Targeted" (β_1=0.138, s.e.=0.071). Thus, this confirms that the discontinuous jump in RD estimates is being driven by the "Targeted" areas and is not a mere shifting of GSE activity from low LR to high LR areas within an MSA.

Figure 1.5 examines the heterogeneity of the RD estimates when I restrict to marginal borrowers. Plot(a) examines the RD discontinuity for above median LR census tracts where the dependent variable is each of the marginal group of borrowers I examined in Table 1.5. The RD estimates repeat the specification in Equation 1.6 by subsetting to young borrowers (defined as borrowers with age less than 45), single female borrowers, first time home-buyers and minority borrowers. Figure 1.5 also shows the point estimates from Panel A (column 1, 2) in Table 1.6 designated as "All" on the vertical axis. The grey lines show the 95% confidence intervals with standard errors clustered at the state level. As before, the point estimates correspond to the coefficient on "Targeted" in Equation 1.6 and refers to the discontinuity at the RD threshold with a value of 1 corresponds to census tracts below the 90 percent tract to MSA median family income. The top half of Plot(a), Figure 1.5 shows that GSE purchases marked discontinuity in high LR areas for young borrowers and single female borrowers. That is, the GSE purchases increased discontinuously in high LR areas for these marginal borrowers. Estimates on first-time home buyers and minority borrowers is noisy (but still positive).[44] In contrast, the RD estimates at the UAG threshold for the below median income census tracts is not significantly different from zero when I subset to young borrowers, first time home-buyers and minority borrowers.[45] Thus, for young borrowers, first time home-buyers and minority borrowers the GSEs did not increase their purchases in

[43] In other words, the census tracts included in columns 1 and 3 in Panel B, Table 1.6 correspond to the same census tracts to the left of 0 (negative spread) in the plots (a) *and* (c) in Figure 1.4. Columns 2 and 4 in Panel B, Table 1.6 correspond to the same census tracts to the right of 0 (positive spread) in the plots (a) *and* (c) in Figure 1.4.

[44] The point estimates, however, are lower for the marginal borrowers than for all borrowers (corresponding to "All" on the vertical axis). This is in contrast to Table 1.5 where the point estimates were slightly higher for minorities. In the RD design when I restrict to marginal borrowers, there are two margins of adjustment. First, moving from above the RD cutoff to below the RD cutoff implies we are moving to low income census tracts. Second, I subset to either young borrowers/single female borrowers/first time home-buyers/minority borrowers. The lower point estimates are being driven by these two margins of adjustment and consistent with the idea that the GSEs were cherry-picking mortgages in high LR areas.

[45] Similarly, for single female borrowers (Figure 1.5in low LR areas, however, the RD estimate is even significantly negative. I addressed the shifting hypothesis above.

the low LR areas to satisfy their UAG goals under the "GSE Act". Overall these results are consistent with cherry-picking of mortgages by the GSEs.

Thus, I show that uniformity of mortgage rates of GSE-conforming loans has the unintended consequence of rationing the *marginal* borrowers out of the mortgage market as indicated by the insignificant impact of UAG goals in low LR areas. Prior literature has also alluded to this cherry-picking of mortgages by the GSEs ([50], [79] and [51]). My results also shed light on prior literature on the housing goals which has found only small effects of these goals on mortgage lending ([10], [12]). I highlight the significant heterogeneity in the impact of these goals across geographies which the prior literature has missed. Strikingly, the mandate to unify geographic distribution of mortgage funding through the underserved areas goals combined with regional uniformity of mortgages inadvertently results in these very underserved borrowers being denied access to the conforming mortgage market.

This paper is most closely related to a recent paper by [44]. [44] relate regional variation in predictable default risk to mortgage rates of GSE-conforming loans. Their analysis specifically focuses on the redistribution across regions through the US mortgage market. I add to the [44] results by showing that while the regional redistribution *does* occur for high credit-quality borrowers who obtain a loan, the GSE undoes some of the inefficiency arising from borrowers self-selecting in certain regions by restricting access to GSE-conforming loans for the low credit-quality borrowers.

I next turn to examine the spillover effects of GSE credit rationing on the non-GSEs.

1.6 Interaction between the GSE-conforming and Private Markets

In this section I turn to the interaction between the GSE-conforming and private markets. First, I analyze the interaction between the GSE-conforming and the private market around the regulatory cutoff CLL. Second, I explicitly look at the spillover in the private mortgage market and the complexity of contract terms in the private mortgage market and examine all jumbo and non-jumbo loans in the private market.

Bunching Analysis

In this section, I look at the interaction between the GSE-conforming and private market. The identification strategy in this section exploits the discrete jump in mortgage rates created by the CLL. The discontinuity in mortgage rates creates a "notch" in the interest rate schedule. I empirically establish the link between changes in interest rate and the extent to which loans bunch at the CLL. I then establish how this bunching varies for high LR and low LR areas. I focus on the bunching that occurs at the CLL to infer whether GSEs are credit rationing.

Mortgage rates and sorting across the CLL

First, I establish how much mortgage rates change across the CLL in high and low LR areas. The goal is to first establish whether the GSE comparative advantage is passed onto borrowers in the form of lower mortgage rates. Figure 1.6, panel (a) plots the mortgage interest rates against the distance from the CLL. The horizontal axis represents the ratio of the loan amount to CLL minus one. Data for loans above the CLL is the non-GSE data from ABSnet for the same period. Data for loans below the CLL is the GSE data from the publicly available 30–year FRM single–family mortgage data provided by Fannie Mae and Freddie Mac. Each point represents the average value of the outcome (mortgage interest rate) in the 5% interval. The solid line plots the predicted values with separate quadratic distance from CLL trends on either side of the CLL. The dashed lines show the 95 percent confidence intervals. The plots use cross-border MSAs and include MSA-year-quarter fixed effects. The plots correspond to bandwidths of 30 percent around the cutoff. These figures are analogous to the figure in the RD analysis. However, we will later see that there is extreme sorting around the CLL and hence all the RD identification assumptions fail. Thus, we cannot interpret these differences as RD estimates. Additionally, since borrowers can choose loan size, they can freely sort around the CLL cutoff and the RD assumptions of being as good as randomly assigned around the CLL does not hold. Instead these plots should be interpreted as the amount of sorting that occurs at the CLL.

We see that for loans below the CLL — that is, those corresponding to GSE mortgages — mortgage rates are the same in both high and low LR areas. For loans above the CLL — that is, those corresponding to non-GSE mortgages — mortgage rates in high LR areas are significantly lower than mortgage rates in low LR areas. Note, in these graphs only the MSA-year-quarter fixed effects have been taken into account. Mortgage rates (Figure 1.6, panel(a)) for loans above the CLL are significantly higher than mortgage rates for loans below the CLL. This reflects the jumbo-conforming mortgage rate spread. In Panel (b) I also control for FICO, LTV and DTI and the results remain unchanged.

We now turn to examine this jump up in mortgage rates more formally. As a first step we wish to estimate the difference in mortgage rates at the CLL in both high LR areas and low LR areas. Looking at mortgage rates close to the bunching boundary, however, will be confounded by other borrower characteristics changing simultaneously at the CLL boundary. Thus, we use the same Coarsened Exact Matching procedure that we employed in Section 1.5. Table 1.8 shows the results of this analysis. We match mortgage loans close to the CLL cutoff. We restrict our analysis to the 10 percent band around the CLL cutoff. Then similar to the analysis in Table 1.4 we match mortgages bins across the CLL cutoff based on different bins in a given MSA in a given year-quarter. As before, for CEM I match mortgages exactly based on different bins in a given MSA-year-quarter. I use loan-to-value (LTV) bins with cutoffs 70, 75, 80, 85, 90 and 95,; credit score (FICO) bins with cutoffs of 620, 660 and 720; and debt-to-income (DTI) bins with cutoffs 20 and 40. We see that the unadjusted jumbo-conforming loan spread — controlling for MSA-year-quarter fixed effects — in high LR areas (column 1) is 38 basis points. However, when we use the CEM procedure and control for FICO, LTV and DTI we find that the jumbo-conforming loan spread reduces to 24 basis points. Similarly, the unadjusted jumbo-conforming loan spread in low LR areas (column 3) in 51 basis points. Using the CEM matched sample reduces the jumbo-conforming loan spread to 40 basis points. This establishes the first step in our analysis of the interaction between GSE and private market. The GSE-conforming loans enjoy lower mortgage rates compared to the private market mortgages and are thus are able to dominate the market in which they compete (loans below the CLL).

Bunching Estimates

We now turn to the bunching analysis. Figure 1.7, Panel A shows both the observed empirical distribution and the counterfactual distribution estimated from the bunching procedure using the 30 year fixed FRM loans. Panel A shows bunching of borrowers in areas with above median LR. Panel B shows bunching of borrowers in areas with below median LR. The x-axis represents the loan size expressed as the ratio of the loan amount to CLL in percentage. Thus, 100 on the horizontal axis corresponds to the CLL. 101 represents loans of size equal to 1.1 times the CLL. The vertical axis corresponds to the number of loans in each bin. Each dot corresponds to number of loans in the 1 percent bins around the CLL (100). The plots show loans between 70 percent to 130 percent of the CLL. The figure shows the actual distribution of loans (in blue dots) and the counterfactual distribution (smooth red line). The counterfactual loan distribution fits a 5^{th} order polynomial to the loan distribution excluding a region around the CLL. The region for exclusion is obtained using the methodology described in Section 1.4.

The empirical distribution denoted by the dotted blue line exhibits sharp bunching at the CLL at 100. The solid red line is the fitted 5^{th} degree polynomial showing the empirical counterfactual distribution. We see that for both high and low LR areas, there is sharp bunching at the CLL. The bunching estimate is based on the amount of loans in the empirical distribution relative to the smoothed counterfactual distribution. This estimate can be thought of as the summary measure of the relative amount of bunching that occurs at the CLL.

We saw in Table 1.8 that the jumbo-conforming spread was 24 basis points higher in high LR and 40 basis points higher in low LR. This is consistent with our hypothesis that since lender payoffs are lower in low LR areas (see Table 1.1), the jumbo-conforming spread should be higher in low LR areas. We also hypothesized that a purely borrower response would imply bunching should be higher in low LR areas as borrowers face a higher notch in their budget constraint. However, in the presence of credit rationing by the GSEs, the lower quality borrowers get rationed out of the mortgage market and observed bunching (corresponding to the empirical distribution) should be lower. Figure 1.7 examines this hypothesis. Panel (a) in Figure 1.7 shows that areas with high LR had a higher amount of bunching at the CLL with a bunching estimate of 7.36 (s.e.=2.3) compared to a bunching estimate of 6.34 (s.e.=2.08) for low LR areas. This is inconsistent with a purely borrower response but consistent with credit rationing of the lowest quality borrowers by the GSEs.

Next, I supplement our analysis by examining heterogeneity across borrowers. In Figure 1.7, I look at two sets of borrowers, the low-risk (or unconstrained) and the high-risk (or constrained) borrowers. I define low-risk borrowers as those with credit scores (FICO) above 720 and with loan-to-value (LTV) below 80. The remaining are classified as (relatively) high-risk. We hypothesized credit rationing by the GSEs should bind *more* for credit-constrained or marginal borrowers. Indeed, bunching estimates for the more pronounced for the credit constrained borrowers. For credit constrained borrowers the bunching estimate is higher at 9.185 (s.e.=4.09) for high LR areas compared to 7.71 (s.e.=2.65) for low LR areas. In contrast, for the less credit constrained borrowers the bunching estimates are very similar with a bunching estimate of 7.92 (s.e.=2.55) for high LR areas and 7.94 (s.e.=2.39) for low LR areas. These results parallel our analysis in Table 1.5 where we found that the minimum credit quality threshold above which the GSEs are willing to lend binds only for the low credit quality borrowers.

Spillovers in the non-GSE mortgage market

In the previous section we saw the strong spillover effects of credit rationing of GSE-conforming loans around the CLL. We now turn to look explicitly at the spillover effects of a uniform pricing policy of the GSEs and how the subsequent credit rationing by the GSEs distorts the private market. In Table 1.9, Panel A, I look at credit rationing by the private market. Since in the period of analysis the private market was increasing its market share, one could argue that the credit rationing effects of the GSE-conforming loans were not as stark since the private market purchased these rationed-out loans. I do find some evidence of this. However, marginal borrowers still get rationed out.[46] [47]

Complexity of loans

[7] and [61] point to the supply of complex mortgages — such as Interest Only (IO) loans and Negative Amortization (NEGAM) loans – as a reason for the rise in risky lending during the pre-crisis period. I explicitly analyze what the GSE credit rationing meant for borrowers in low LR areas. In Figure 1.9, Panel A, I look at the percentage of complex mortgages and ARMs as a percentage of the total mortgages (IO/NEGAM/ARM/FRMs). As in the previous section I restrict to loans between 70 percent and 130 percent of the CLL. I find that for loans below the CLL, the percentage of complex mortgages is *higher* for low LR areas compared to the high LR areas. Thus, borrowers in low LR areas credit rationed by the GSEs were able to access credit from the non-GSEs but of (arguably) risky quality. In contrast,

[46] I only focus on the minority and single-family borrowers for private securitizers. The data for private securitizers comes from a different dataset (HMDA) which does not provide information on first time home-buyers and age of borrowers. The GSE data is based on a different data source (HUD data). See Appendixapp:data and the Online Appendix for details.

[47] See recent literature which has also used the CLL as an exogenous source of variation ([60], [60], [29] and [2]).

for loans above the CLL, high LR areas had higher percentage of complex mortgages and ARMs. Figure 1.9, Panel B, shows that when we restrict to non-GSE mortgages, the high LR areas have higher ARMs/complex mortgages. The graphical analysis seems to suggest that despite the fact that the private market did take the higher payoffs in low lender rights areas into account (Panel b), but the low LR borrowers rationed out of the GSE-conforming market accessed the private securitizers who then pushed them into ARMs and complex mortgages.[48]

In Table 1.9 I analyze this more formally. Column 1 looks at whether a loan was more likely to be purchased by the GSEs in the non-jumbo conforming market.[49] The dependent variable in column 1 is an indicator equal to 1 if a loan is purchased by the GSEs. In column 1, I look at all loans (FRM/IO/NEGAM/ARM) between 70 percent to the CLL and test whether a loan was more likely to be purchased by the GSEs. Column 1 shows that there was a 1.5 percent higher probability that loans in high lender rights were more likely to be purchased by the GSEs. Additionally, they were 11 percent more likely to have a FICO above 720. Additionally, GSE puehcases they were 12 percent more likely to have a debt-to-income below 36 (which can be considered less risky). In column 2, I interact this term with the LR index and find that in high LR areas, GSE purchases were 20 percent less likely to have a FICO above 720 and 2 percent less likely to have DTI below 36. In other words, in exactly the low LR areas, the non-GSEs had higher purchases. The lower quality of loans of the non-GSEs (seen in column 1), especially so for the high LR areas (seen in column 2) points to the distortion of the market (with lower quality loans) in the non-GSE market possibly due to the GSEs' funding advantage. Next, I turn to whether the non-GSEs pushed borrowers rationed out of the mortgage market into riskier loans. When I turn to the analysis of the non-GSE loans in columnc 3–4, I find that the non-GSEs do take the higher payoffs in high lender rights into account. I find that mortgages were 3 percent more likely to be ARMs (relative to FRM) and 4 percent more likely to be complex mortgages (relative to FRM) in high LR areas. Additionally, relative to a (non-GSE) FRM, ARMs were less likely to have FICOs above 720. Relative to a (non-GSE) FRM complex mortgages were less likely to have FICOs above 720.

Potentially, these results shed new light on the recent literature on the change in lending behavior during the pre-crisis period. [65] and [64] point out that supply of mortgages increased to low-income zipcodes during this period whereas [4] and [3] argue that it was the relatively high income borrowers who received these loans. [5], [9], [18], [53] and [42] blame the relaxation of lending standards for the subprime crisis. Others ([71], [70], [67], [56] and [61]) point to expansion of the private securitization markets as the cause of the subprime crisis. While my results do not extend beyond the crisis period, they highlight the possible spillover effects that the GSEs exert on the private market through their influence in the secondary mortgage market. Borrowers rationed out of the conforming mortgage market had

[48] Arguably ARMs need not be strictly considered risky, but have been included here for the sake of completeness ([7]).

[49] In the previous section, we looked at the interaction of the GSEs with the jumbo non-GSE market. In columns 1 and 2, I focus on the interaction between the GSE and non-GSEs in the non-jumbo market.

to turn to the private securitizers who possibly gave them riskier loans. My results point to the role of the GSEs in pushing these private securitizers further down the credit curve. I provide here only suggestive evidence and leave the full analysis for future research.

Overall Mortgage Market

Now I turn to the overall impact of lender rights in the private market and the originations in the primary market. To round out the analysis, I first focus on all private sector purchases in the secondary market for all loans below 2 times the CLL. Similarly, I look at the total originations in the primary market for all loans below 2 times the CLL.

Table 1.10, Panel A looks at total number of loans per housing stock in the private market (private sector purchases). I find that a 1 SD higher LR index implies 20 percent higher non-conforming loans per housing stock in the private market. Further, in contrast to our findings in Section 1.5 there is no credit rationing of the marginal borrowers whereas in Section 1.5 I found that minority borrowers had a higher number of GSE-conforming loans in high LR areas. For the single-female borrowers who were rationed out, the lowest quality borrowers are still rationed out of the mortgage market in the high LR areas. Thus, for the single-female borrowers it seems that the private market did not cater to those borrowers who were rationed out of the conforming (GSE) mortgage market. To summarize, our results show that while some of the borrowers credit rationed by the GSEs in low LR areas were able to access credit from the private market, the borrowers with the lowest quality were rationed out of the private mortgage market too.

To the analyze the overall effect of credit rationing, I turn to the originations in the primary market. Until now I have abstracted away from the additional layer between the borrowers and secondary market activity of the GSEs and private securitizers. However, I now look explicitly at direct lending by the originators. This helps me analyze the overall impact of credit rationing of the GSEs and private securitizers. In Table 1.10, Panel B, I find that the credit rationing by the GSEs persisted for all borrowers. Thus, number of originations per housing stock is 17 percent in areas with 1 SD higher LR index. The effects are even stronger for minority borrowers though slightly lower for the single-female borrowers.

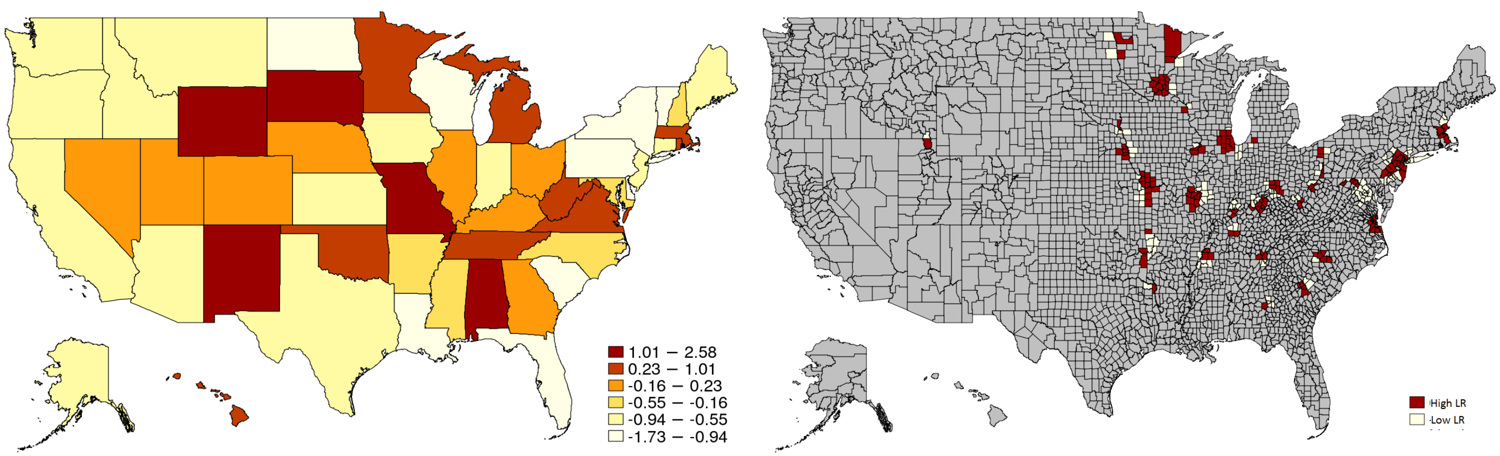
In Table 1.10, Panel C, we look at denials by the originators. We find that denials were lower for mortgages in high LR areas. Although denials were lower by 1.184 percent when LR Index is higher by 1 SD for the full population of loans, this effect is only marginally significant. However, when we restrict to minority borrowers the effects are much stronger and higher in magnitude with a 1 SD high LR index implying a 2.4 percent denial rate. Similarly, for single-female borrowers the denial rates were higher at 2.14 percent. One caveat in using denial rate data is that there is already significant sorting of borrowers and not all borrowers who seek loans and are denied will show up on the database. This explains the lower point estimates of these denial rates compared to the effects in Panel B.

1.7 Conclusion

In this paper, I used one aspect of regional variation, namely state foreclosure law that affect mortgage lending. Utilizing this sharp variation in foreclosure law at state borders combined with plausibly exogenous variation from a policy regulation based on affordable housing goals, a novel source of quasi-experimental variation in mortgage interest rates, and unique rate sheet and loan level data, I show three sets of results. First, I show that mortgage interest rates of GSE-conforming loans vary across LR. Second, I show that this lack of regional variation in mortgage interest rates affects credit access (that is, the extensive margin) across LR. Using a regression discontinuity (RD) design focused on the marginal borrowers, I also show that the GSEs "cherry-pick" the better risks leading to greater credit access to marginal borrowers in high LR areas. Finally, using bunching analysis based on the regulatory cutoff on loan size for GSE-conforming loans, I show that the GSE's cost of funds advantage which are passed onto borrowers in the form of lower mortgage rates, implies that it gets the better quality borrowers. Additionally, borrowers who are pushed out of the GSE-conforming market are pushed into increasingly risky loans.

There are two important policy implications. First, while uniform mortgage rates do imply that borrowers who get mortgages cross-subsidize each other, one unintended consequence may be that marginal borrowers get completely rationed out of the conforming mortgage market. Second, this credit rationing combined with the GSEs' comparative advantage implies that the private market faces a heavily sorted market. Thus, while the private market was able to pick up the slack and purchase mortgages of the borrowers rationed out of the private market, the private market pushed these borrowers into increasingly risky mortgages. Importantly, my results show this may disproportionately affect marginal borrowers.

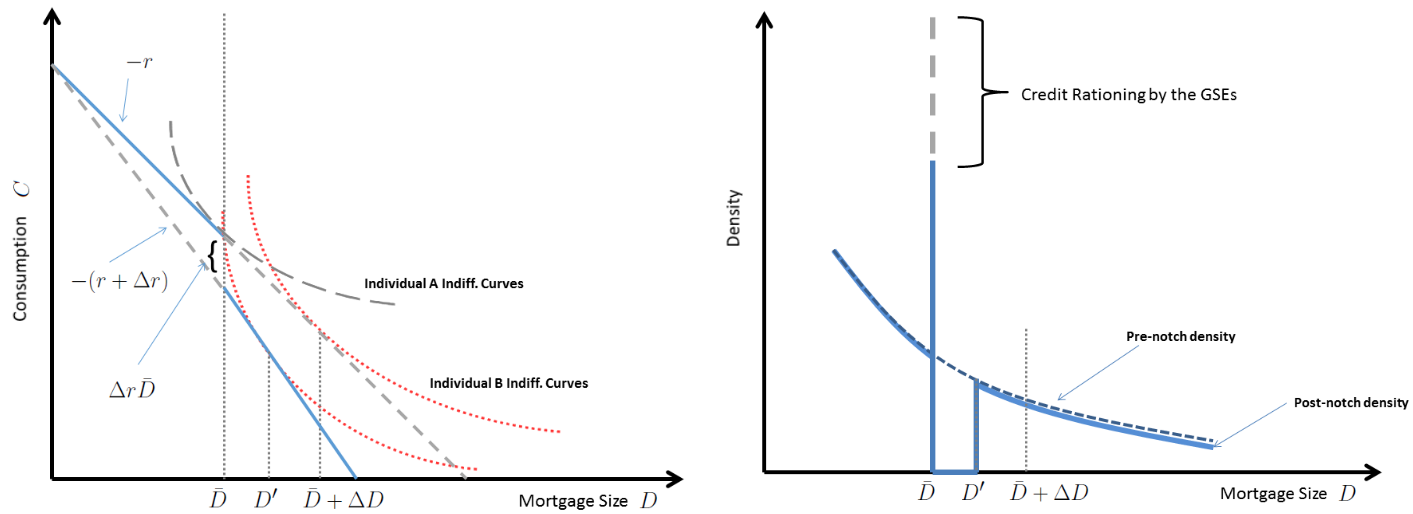
Figure 1.1: Lender Rights Index and Cross-border Sample MSAs retained



(a) **LR Index** (b) **Cross-Border Sample**

Panel (a) shows the geographic variation of lender rights (LR) index by state. Panel (b) shows the cross-border sample retained in my analysis. The LR index is a continuous variable and has been z-scored, with a value of 0 corresponding to the median value of the index. The LR index is calculated by assigning a value between 1-3 for each of the lender rights non-judicial, recourse (fair-market-value and non-fair-market-value) and right-to-redeem for each state. See Appendix A.1 for further details. I retain 38 Metropolitan Statistical Areas (MSAs) that cross state borders and for which the LR index varies within the MSA. Panel (b) shows the cross-border MSAs retained in my sample. The map highlights in darker maroon the side of the MSA corresponding to higher LR and in lighter yellow the side of the MSA corresponding to lower LR. The areas shaded in grey are not used in the analysis.

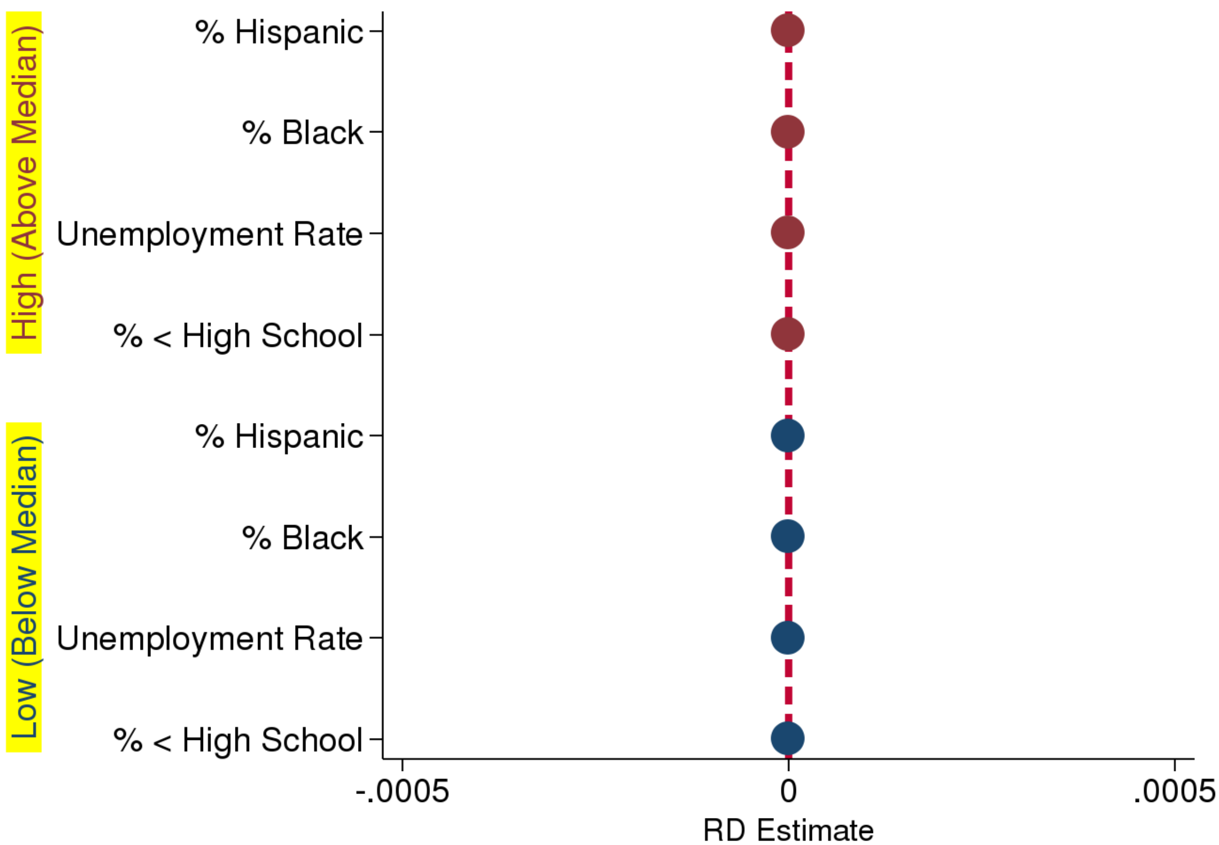
Figure 1.2: Marginal Bunching Households and GSE Credit Rationing



(a) **Household Budget Constraint** (b) **With Credit Rationing**

This figure shows the effects of a notch, that is, a discrete increase in the mortgage interest rate from r to $r + \Delta r$ at the conforming loan limit (CLL). Panel (a) shows the effects of the notch in a budget set diagram and Panel (b) shows the corresponding density distribution diagrams in the presence of credit rationing by the GSEs. Panel (a) shows the choice faced by a household (borrower) given a notched interest rate schedule at the CLL in a budget set diagram. The budget set diagram (Panel (a)) is depicted in ((consumption (C), loan size (D))-space) and illustrates the behavioral responses among individuals with heterogeneous housing preferences A, but with a specific demand elasticity ϵ. The bunching cutoff CLL is depicted by $\bar{D}$. The solid blue line depicts the budget constraint with slope $-r$ below the CLL and $-(r + \Delta r)$ above the CLL. The indifference curves of Individual A shown correspond to the individual who would choose a loan of size $\bar{D}$ in both the linear (corresponding to slope of $-r$) and the notched interest rate schedule (corresponding to the solid blue line). The indifference curves of Individual B shown correspond to the individual who when facing the notched interest rate schedule is indifferent between choosing a loan of size $\bar{D}$ and D'. In the linear interest rate schedule the individual B would have chosen a loan of size $\bar{D} + \Delta D$. Panel (b) shows the post-notch density distribution, with sharp bunching at $\bar{D}$. The amount of bunching determined by the empty region between $\bar{D}$ and D' is equal to the sharp bunching denoted by the blue and grey line at $\bar{D}$. The dashed grey line denotes credit rationing by the GSEs. Observed bunching is thus equal to the solid blue line at $\bar{D}$.

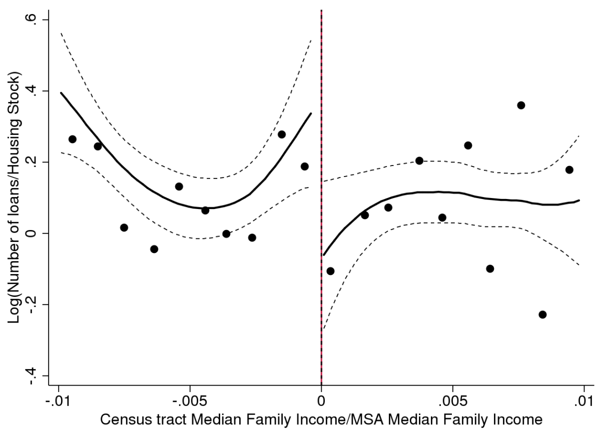
Figure 1.3: Cherry-picking by the GSEs (RD Analysis): Covariate Balance



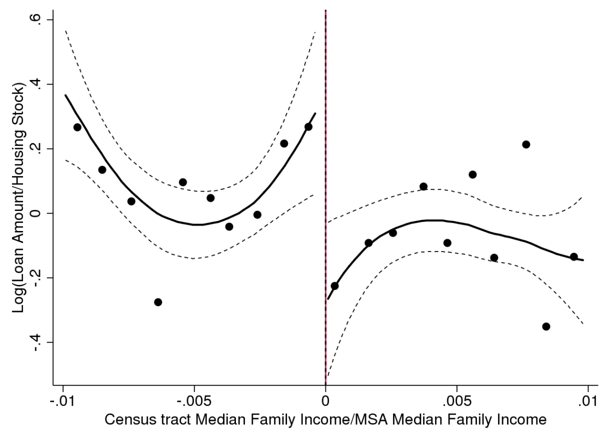
This figure shows the covariate balance for the regression discontinuity (RD) using the Underserved Area Goals (UAG). The above graph shows the RD estimates — corresponding to the coefficient on "Targeted" — when the respective characteristic (on the vertical axis) is used as the dependent variable in Equation 1.6. A tract is considered eligible under UAG goals if the ratio ($\frac{Census\ Tract\ Median\ Family\ Income}{MSA\ Median\ Family\ Income}$) is less than 90 percent. I show the RD estimate of each covariate separately for census tracts in the above median lender rights (LR) (top half of graph) and below median LR (bottom half of graph). Error bands corresponding to 95 percent confidence interval are depicted in grey around the RD estimate and are very small. Covariates shown are percentage with less than high-school education, percentage Hispanic, percentage black and unemployment rate from the 2000 Census. All regressions are weighted by total number of households from the 2000 Census. Standard errors are clustered at the state level. See Appendix A.1 for details on construction of the LR index. See the Online Appendix for details on variable construction.

Figure 1.4: Cherry-picking by the GSEs (RD Analysis): Baseline Results

Panel A: Above Median LR

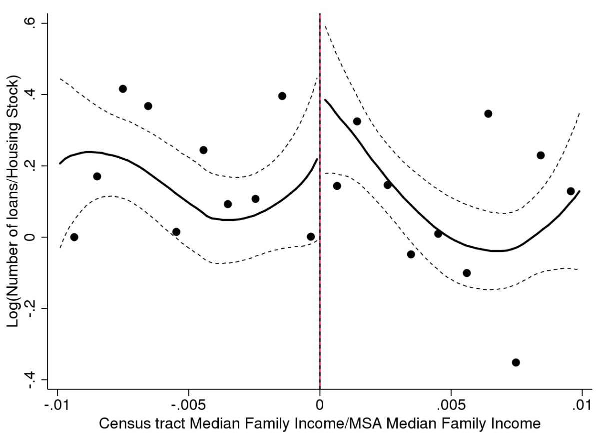


(a) **Number of loans**

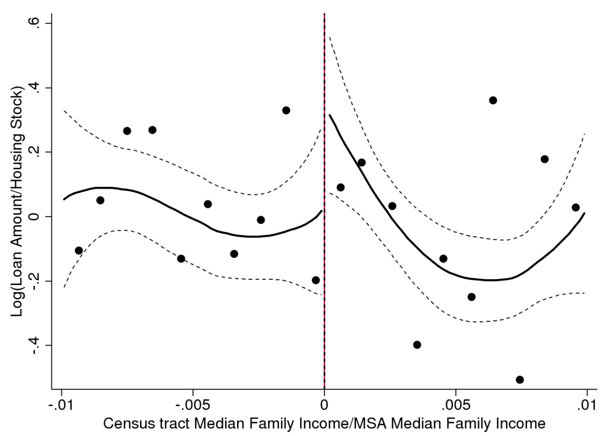


(b) **Loan Amount**

Panel B: Below Median LR



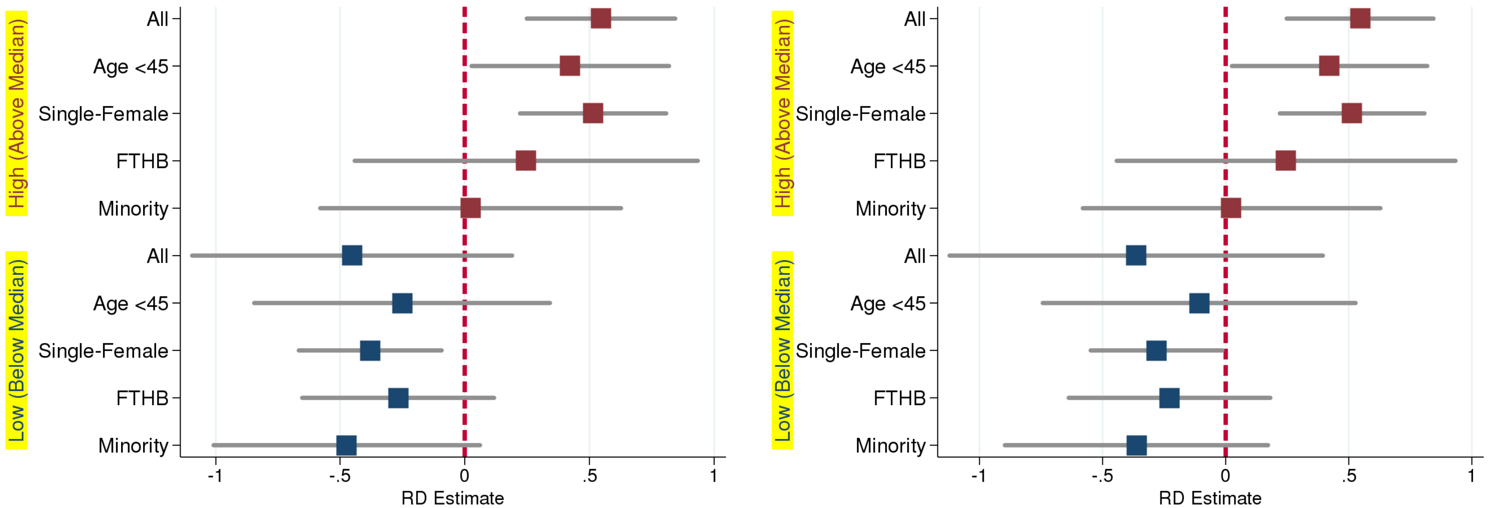
(c) **Number of loans**



(d) **Loan Amount**

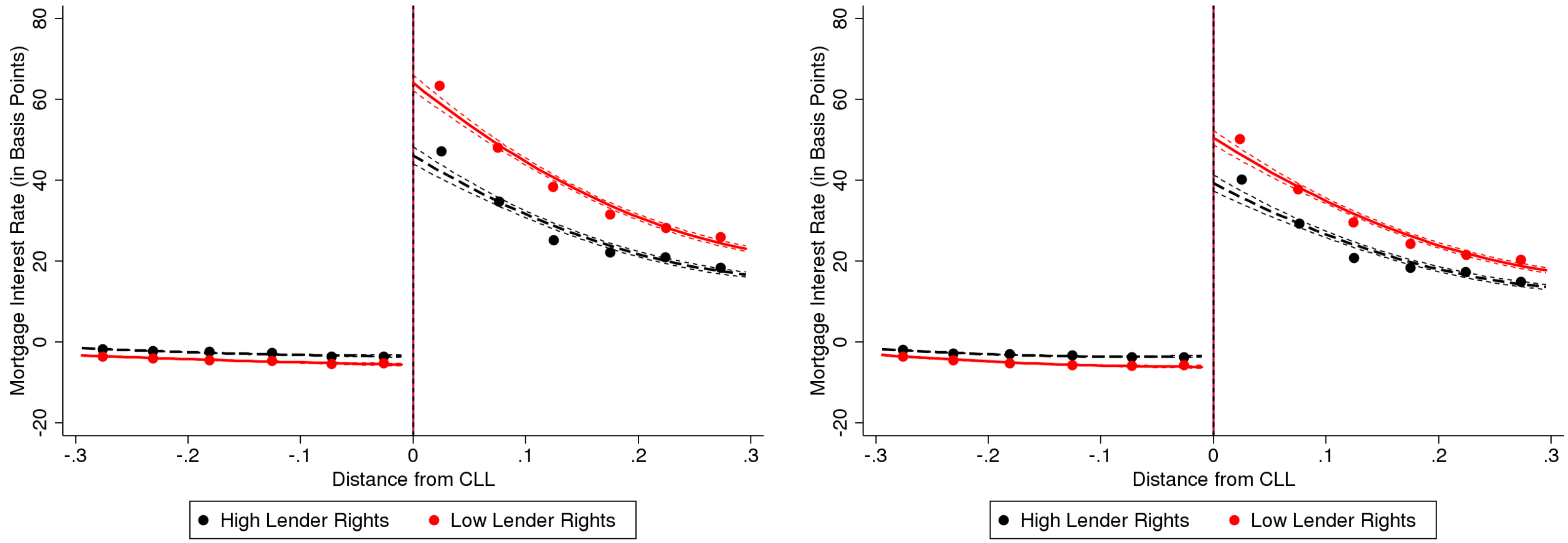
This figure shows the regression discontinuity (RD) plots using Underserved Area Goals (UAG). The dependent variable in panels (a) and (c) is the logarithm of the total number of loans per housing stock. The dependent variable in panels (b) and (d) is the total volume of loans per housing stock. A tract is considered eligible under UAG goals if the ratio ($\frac{Census\ Tract\ Median\ Family\ Income}{MSA\ Median\ Family\ Income}$) is less than 90 percent. 0 on the horizontal axis corresponds to this 90 percent RD threshold. Thus, negative values on the x-axis correspond to UAG eligible or "Targeted" census tracts. I use a bandwidths of 1 percent around the RD cutoff. Each point on the plot represents the average value of the dependent variable (y-axis) in bins of one-tenth of a percentage point of the ratio ($\frac{censustractmedianfamilyincome}{theMSAmedianfamilyincome}$) (x-axis). Panels (a) and (b) corresponds to census tracts with above median lender rights (LR) index. Panels (c) and (d) corresponds to census tracts with below median LR. The solid line plots the predicted values with separate quadratic polynomials fitted on either side of the RD cutoff. The dashed lines show the 95 percent confidence intervals. Data on number and volume of loans aggregated to the census tract level is from Housing and Urban Development (HUD) for the years 2000 to 2005. All plots include control variables and MSA-year fixed effects. Control variables percentage with less than high-school education, percentage Hispanic, percentage black and unemployment rate are from the 2000 Census and 1990 Census.

Figure 1.5: Cherry-picking by the GSEs (RD Analysis): Heterogeneity Results



(a) **Log(Total Number of Loans/Housing Stock)** (b) **Log(Total Loan Volume/Housing Stock)** This figure shows the heterogeneity of the regression discontinuity (RD) estimates using Underserved Area Goals (UAG) for each subset of marginal borrowers. The dependent variable in panel (a) is the logarithm of the total number of loans per housing stock using all borrowers, young borrowers (age < 45), single female borrowers, first time home-buyers and minority borrowers for above median lender rights (LR) index (top half of graph) and below median LR (bottom half of graph). The dependent variable in panel (b) is the logarithm of the total volume of loans per housing stock using all borrowers, young borrowers (age < 45), single female borrowers, first time home-buyers and minority borrowers for above median lender rights (LR) index (top half of graph) and below median LR (bottom half of graph). Panel (a) shows the RD estimates — corresponding to the coefficient on "Targeted" — for the respective subset of borrowers (all/age < 55/single female/first time home-buyers/minority) using the dependent variable logarithm of the total number of loans per housing stock in Equation 1.6. Panel (b) shows the RD estimates — corresponding to the coefficient on "Targeted" — for the respective subset of borrowers (all/age < 55/single female/first time home-buyers/minority) using the dependent variable logarithm of the total volume of loans per housing stock in Equation 1.6. Error bands corresponding to 95 percent confidence interval are depicted in grey around the RD estimate. I use a bandwidth of 1 percent around the RD cutoff. Data on number and volume of loans aggregated to the census tract level is from Housing and Urban Development (HUD) for the years 2000 to 2005. All plots include control variables and MSA-year fixed effects. Control variables percentage with less than high-school education, percentage Hispanic, percentage black and unemployment rate are from the 2000 Census and 1990 Census. All regressions are weighted by total number of households from the 1990 Census for the years 2000–2002 and the 2000 Census for the years 2003–2005. Standard errors are clustered at the state level. See Appendix A.1 for details on LR index construction. See the Online Appendix for details on variable construction.

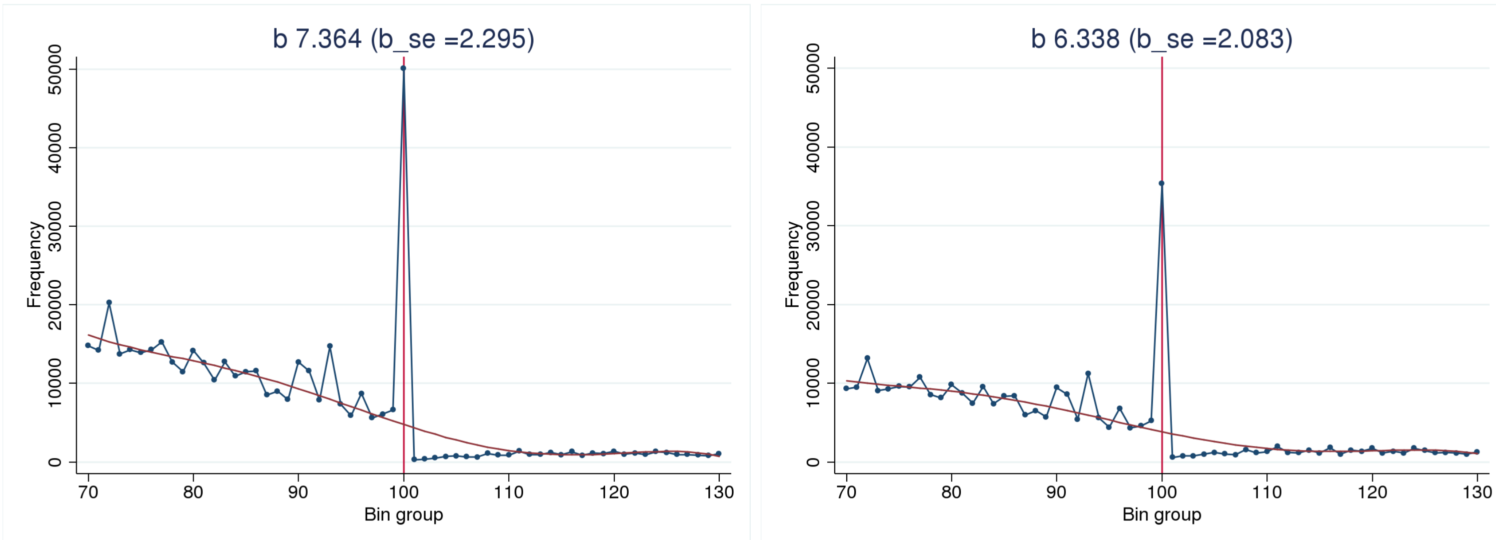
Figure 1.6: Spillovers in the Private Market (Bunching Analysis): Mortgage Rates



(a) **With MSA-Year-Qtr FE** (b) **With MSA-Year-Qtr FE and Controls**

This figure plots the mortgage interest rates across the conforming loan limit (CLL) ny lender rights (LR) index. Panel (a) shows the mortgage rates of loans with year-quarter fixed effects and panel (b) shows mortgage rates with controls and MSA-year-quarter fixed effects. The x-axis represents the loan size expressed as the ratio of the loan amount to conforming loan limit (CLL) minus one. Thus, 0 on the horizontal axis corresponds to the CLL. 0.1 represents loan sizes of 1.1 times the CLL. The plots correspond to bandwidths of .3 — that is, all loans of loan size between 0.3*CLL and 1.3*CLL — around the cutoff. Each point on the plot represents the average value of the dependent variable mortgage interest rate (y-axis) in bins of 5 percent of the normalized distance from the CLL (x-axis). The black dots correspond to mortgages in areas with above median lender rights (LR) and the red dots correspond to mortgages in areas with below LR. In Panel (b) also controls for bins of loan-to-value (LTV), credit scores (FICO) and debt-to-income. We use the same bins with cutoffs as in Table 1.4. The solid line plots the predicted values with separate quadratic polynomials fitted on either side of the CLL. The dashed lines show the 95 percent confidence intervals. Data is for the period 2000 to 2005. Data for loans below the CLL is the GSE data from publicly available 30-year Fixed Rate Mortgage (FRM) single-family mortgage data provided by Fannie Mae and Freddie Mac. Data for loans above the CLL is the non-GSE data from ABSnet for the same period. Standard errors are clustered at the state level. See Appendix A.1 for details on LR index construction. See the Online Appendix for details on variable construction.

Figure 1.7: Spillovers in the Private Market (Bunching Analysis): Loans around the CLL

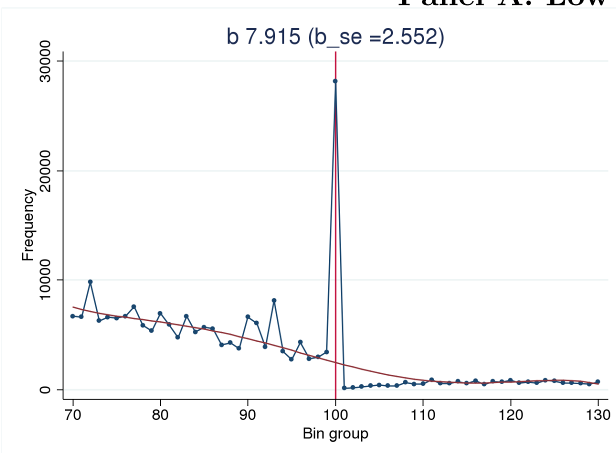


(a) **High LR** (b) **Low LR**

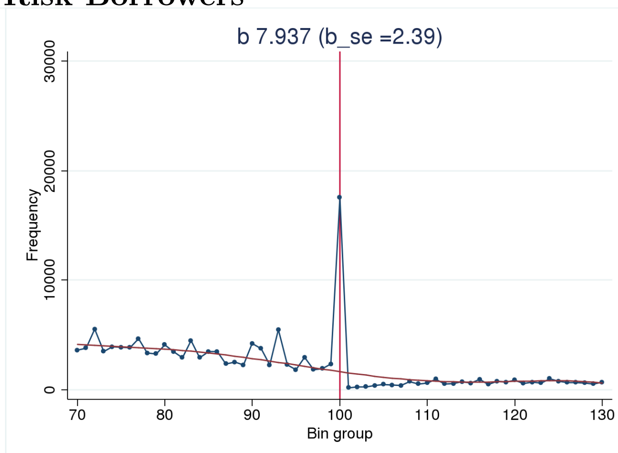
This figure estimates the bunching of loans around the conforming loan limit (CLL) across the lender rights (LR) index. Panel A shows bunching of loans in areas with above median LR. Panel B shows bunching of loans in areas with below median LR. The x-axis represents the loan size expressed as the ratio of the loan amount to conforming loan limit (CLL) in percentage. Thus, 100 on the horizontal axis corresponds to the CLL. 101 represents loans of size equal to 1.1 times the CLL. The vertical axis corresponds to the number of loans in each bin. Each dot corresponds to number of loans in the 1 percent bins around the CLL (100). The plots show loans between 70 percent to 130 percent of the CLL. The plots show the actual distribution of loans (in blue dots) and the counterfactual distribution (smooth red line). The counterfactual loan distribution fits a 5^{th} order polynomial to the loan distribution excluding a region around the CLL. The region for exclusion is obtained using the methodology described in Section 1.4. Data is for the period 2000 to 2005. Data for loans below the CLL is the GSE data from publicly available 30-year Fixed Rate Mortgage (FRM) single-family mortgage data provided by Fannie Mae and Freddie Mac. Data for loans above the CLL is the non-GSE data from ABSnet. Standard errors are clustered at the state level. See Appendix A.1 for details on LR index construction. See the Online Appendix for details on variable construction.

Figure 1.8: Spillovers in the Private Market (Bunching Analysis): Loans around the CLL and Heterogeneity

Panel A: Low-Risk Borrowers

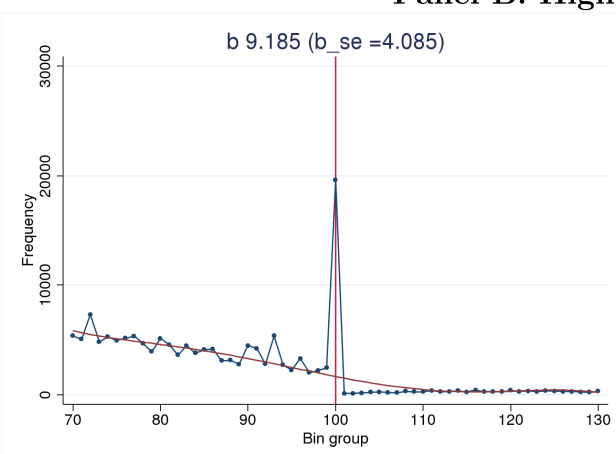


(a) **High LR**

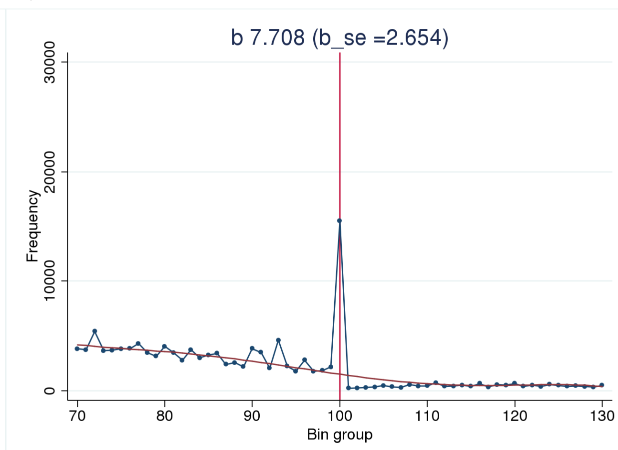


(b) **Low LR**

Panel B: High-Risk Borrowers



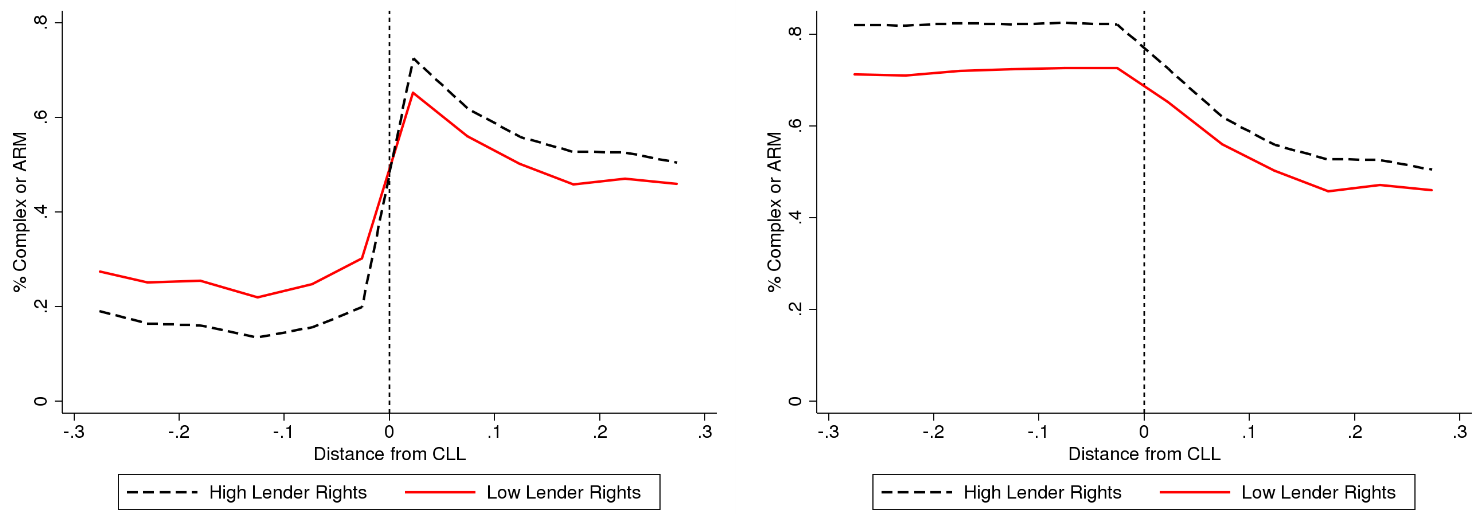
(c) **High LR**



(d) **Low LR**

This figure estimates the heterogeneity in the bunching of loans around the conforming loan limit (CLL) across lender rights (LR) index. Panel A shows the low-risk borrowers, that is, those with credit scores (FICO) above 720 and with loan-to-value (LTV) below 80. Panel B shows the remaining borrowers ("high-risk"). Plots (a) and (c) show bunching of borrowers in areas with above median LR. Plots (b) and (d) show bunching of loans in areas with below median LR. The x-axis represents the loan size expressed as the ratio of the loan amount to conforming loan limit (CLL) in percentage. Thus, 100 on the horizontal axis corresponds to the CLL. 101 represents loans of size equal to 1.1 times the CLL. The vertical axis corresponds to the number of loans in each bin. Each dot corresponds to number of loans in the 1 percent bins around the CLL (100). The plots show loans between 70 percent to 130 percent of the CLL. The figure shows the actual distribution of loans (in blue dots) and the counterfactual distribution (smooth red line). The counterfactual loan distribution fits a 5^{th} order polynomial to the loan distribution excluding a region around the CLL. The region for exclusion is obtained using the methodology described in Section 1.4. Data is for the period 2000 to 2005.

Figure 1.9: Spillovers in the Private Market (Bunching Analysis): ARMS and Complex mortgages around the CLL



(a) **Including GSE loans** (b) **Excluding GSE loans**

This figure shows the complexity of mortgages across lender rights (LR) index around the conforming loan limit (CLL) for all loans in panel (a) and for loans excluding GSE loans in panel (b). The x-axis represents the loan size expressed as the ratio of the loan amount to conforming loan limit (CLL) in percentage. Thus, 100 on the horizontal axis corresponds to the CLL. 101 represents loans of size equal to 1.1 times the CLL. In the above plots I retain loans between 70 percent to 130 percent of the CLL. The vertical axis corresponds to the percentage of loans within each bin that is either Interest Only (IO), Negative Amortization (NEGAM) or Adjustable Rate Mortgages (ARM). The denominator contains all the loans and also includes Fixed Rate Mortgages (FRM). IO and NEGAM are together termed as Complex Mortgages (CM). Panel (a) includes all GSE (FRM) and non-GSE loans. Panel (b) includes only the non-GSE loans. Thus the loans above 0 on the x-axis in panel (a) and panel(b) show the same set of mortgages. Data is for the period 2000 to 2005. Data for loans below the CLL is the GSE data from publicly available 30-year Fixed Rate Mortgage (FRM) single-family mortgage data provided by Fannie Mae and Freddie Mac. Data for loans above the CLL is the non-GSE data from ABSnet for the same period. See Appendix A.1 for details on LR index construction. See the Online Appendix for details on variable construction.

Table 1.1: Bunching Predictions

	Low LR	High LR
Jumbo-Conforming Spread	Higher	Lower
Borrower Response (1)	Higher	Lower
GSE Credit Rationing (2)	Higher	Lower
Observed Impact (1) - (2)	?	?

This table shows the predictions for the jumbo conforming spread and the resulting borrower response, credit rationing by the GSEs and observed bunching estimates for low and high lender rights (LR) areas. See Section 1.3 for further details on conceptual framework and bunching predictions.

Table 1.2: Summary Statistics

	Mean	SD	Min	Max
LR Index (Standardized)	0.00	1.00	-1.73	2.58
Observations	51			
RateWatch Data (Rate Sheets) – Branch Level				
Interest Rate: Rate Sheet	6.62	0.84	0.62	11.07
Observations	502765			
HUD Variables – Census Tract level				
Total # loans	120	162	1	7809
# loans to Minority	12	22	0	451
# loans FTHB	13	15	0	497
# loans Single- female	29	35	0	1006
# loans with No Co-borrower	69	111	0	6195
# loans with Age<45	60	93	0	5455
Observations	120207			
GSE 30 year FRM – Loan level				
Original Interest Rate	6.46	0.86	2.99	12.12
FICO	715	56	300	850
LTV	73	16	1	103
Debt-to-Income	34	12	0	65
Loan Amount	166727	73476	5000	692000
Observations	2892199			
ABSnet 30 year FRM – Loan level				
Original Interest Rate	6.96	1.41	1.00	17.10
FICO	706	49	620	837
LTV	76	15	3	99
Debt-to-Income	9	17	0	92
Loan Amount	254015	144381	10000	1000000
Observations	304966			

This table shows the summary statistics of the variables used in our analysis. The lender rights (LR) index is standardized. See Appendix A.1 for details on LR. Rate sheet data on mortgage rates at branch level is from `RateWatch.com`. Data on number of loans aggregated to the census tract level is from Housing and Urban Development (HUD). Loan level GSE data is from publicly available 30-year Fixed Rate Mortgage (FRM) single-family mortgage data provided by Fannie Mae and Freddie Mac. Loan level non-GSE (private securitizer) data is from ABSnet. All data is for the period 2000 to 2005 for the 38 MSAs in our analysis. See Online Appendix for details on variable construction and frequency used.

Table 1.3: Covariate Balance

	Coefficient	t-stat
Census 2000 – Census Tract Level		
% < High School 2000	-0.00975	-0.978
Unemployment Rate 2000	-0.00281	-1.252
Log(Median Income) 2000	0.0609	1.536
% Black 2000	-0.0266	-1.018
% Hispanic 2000	-0.00460	-0.516
Census 1990 – Census Tract Level		
% < High School 1990	-0.0166	-1.550
Unemployment Rate 1990	-0.00157	-1.190
Log(Median Income) 1990	0.0508	1.542
% Black 1990	-0.0272	-1.120
% Hispanic 1990	-0.00530	-0.666

This table shows the covariate balance of characteristics of census tracts in the 38 MSAs retained in our sample. I regress the respective characteristic in the first column on lender rights (LR) index. All columns include MSA-year fixed effects and standard errors are clustered at the state level. The coefficient on LR and the corresponding t-statistics is reported in the second and third column respectively. Covariates used in the analysis are percentage Hispanic, percentage black, Ln(median income), percentage with less than high school education and unemployment rate from the 2000 Census and 1990 Census at the census tract level. See Appendix A.1 for details on LR. See the Online Appendix for details on variable construction.

Table 1.4: GSE: Mortgage Interest Rates and Non-rate Rationing

	(1)	(2)	(3)	(4)	(5)
	Rate Sheet	Ex-post: Loan Level			
	Rates	Rate	DTI	LTV	Rate
LR Index	-0.00425	-0.00977	-0.159	-0.690	-0.0166*
	(0.00781)	(0.00989)	(0.107)	(0.524)	(0.00871)
Observations	414625	245239	236258	245236	76055
MSA-Yr FE	X	X	X	X	X
R^2	0.917	0.810	0.0314	0.101	0.818
Regression Type	OLS	OLS	OLS	OLS	CEM

Standard errors in parentheses, clustered by state.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the variation of mortgage rates and non-price terms — debt-to-income (DTI) and loan to value (LTV) — with the lender rights (LR) index for GSE mortgages. Columns 1–4 show the OLS estimates with no controls. Column 5, shows the estimates from the coarsened exact matching (CEM, see [45] and [46]) procedure using a 10 percent sample. In column 5, I use the CEM procedure to match mortgages exactly based on different bins in a given MSA-year-quarter. I use LTV bins with cut-offs 70, 75, 80, 85, 90 and 95; credit score (FICO) bins with cutoffs of 620, 660 and 720; and DTI bins with cutoffs 20 and 30. Data is for the period 2000 to 2005. The dependent variable in columns 1, 2 and 4 is mortgage rates. The dependent variables in columns 3 and 4 are DTI and LTV respectively. Rate sheet data on mortgage rates in column 1 is from `RateWatch.com`. Loan level GSE data in columns 2–5 are from the publicly available 30-year Fixed Rate Mortgage (FRM) single-family mortgage data provided by Fannie Mae and Freddie Mac. All columns include MSA-year-quarter fixed effects and standard errors are clustered at the state level. See Appendix A.1 for details on LR. See the Online Appendix for details on variable construction.

Table 1.5: GSE: Credit Rationing and Heterogeneity across Borrowers

Panel A: Dependent variable Ln(Total Number of Loans/Housing Stock)

	(1) All	(2) Age < 45	(3) Single Female Borrower	(4) First-time Home Borrower	(5) Minority Borrowers
LR Index	0.169*	0.150**	0.149***	0.124***	0.224***
	(0.0958)	(0.0683)	(0.0457)	(0.0418)	(0.0729)
MSA-Yr FE	X	X	X	X	X
Controls	X	X	X	X	X
Observations	89373	85870	83824	80385	75169
R^2	0.415	0.599	0.502	0.294	0.385

Panel B: Dependent variable Ln(Total Loan Volume/Housing Stock)

	(1) All	(2) Age < 45	(3) Single Female Borrower	(4) First-time Home Borrower	(5) Minority Borrowers
LR Index	0.178*	0.177*	0.181***	0.144**	0.239**
	(0.0950)	(0.0993)	(0.0644)	(0.0685)	(0.0894)
MSA-Yr FE	X	X	X	X	X
Controls	X	X	X	X	X
Observations	89373	85157	82790	76747	72464
R^2	0.439	0.563	0.533	0.369	0.401

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the results for credit rationing by the GSEs across lender rights (LR). The dependent variable in panel (A) is the logarithm of the total number of loans per housing stock for each subset of borrowers. The dependent variable in panel (b) is the logarithm of the total volume of loans per housing stock for each subset of borrowers. Column 1 in both panels includes all borrowers. Columns 2, 3, 4 and 5 in both panels restrict to only young borrowers (age < 45), single female borrowers, first time home-buyers and minority borrowers respectively. Data on number and volume of loans aggregated to the census tract level are from Housing and Urban Development (HUD) for the years 2000 to 2005. All regressions include control variables and MSA-year fixed effects. Control variables percentage with less than high-school education, percentage Hispanic, percentage black and unemployment rate are from the 2000 Census and 1990 Census. All regressions are weighted by total number of households from the 2000 Census for the years 2000–2002 and the 1990 Census for the years 2003–2005. Standard errors are clustered at the state level. See Appendix A.1 for details on LR index construction. See the Online Appendix for details on variable construction.

Table 1.6: Cherry-picking by the GSEs (RD Analysis): Regression Estimates

Panel A: Regression Discontinuity Estimates

	(1)	(2)	(3)	(4)
	Ln(No. of Loans/HS)		Ln(Vol. of Loans/HS)	
	Above Median LR	Below Median LR	Above Median LR	Below Median LR
Targeted	0.546***	-0.453	0.532***	-0.364
	(0.152)	(0.328)	(0.133)	(0.387)
MSA-Yr FE	X	X	X	X
Controls	X	X	X	X
Observations	975	1004	975	1004
R^2	0.369	0.228	0.470	0.266

Panel B: Robustness within targeted areas

	(1)	(2)	(3)	(4)
	Ln(No. of Loans/HS)		Ln(Vol. of Loans/HS)	
	Targeted	Not-Targeted	Targeted	Not-Targeted
LR Index	0.183**	0.118	0.242*	0.138*
	(0.0802)	(0.0763)	(0.118)	(0.0711)
MSA-Yr FE	X	X	X	X
Controls	X	X	X	X
Observations	968	1011	968	1011
R^2	0.299	0.308	0.347	0.382

This table is based on the Underserved Area Goals (UAG) and shows the regression discontinuity (RD) in panel A and the difference-in-difference estimates in panel B across the lender rights (LR) index. The dependent variable in columns 1–2 of both panels is the logarithm of the total number of loans per housing stock. The dependent variable in columns 3–4 of both panels is the logarithm of the total volume of loans per housing stock. Panel (A) shows the RD estimates — corresponding to the coefficient on "Targeted" — using the respective dependent variables for above median LR (columns 1 and 3) and below median LR (columns 2 and 4) in Equation 1.6. Columns 1 and 3 in panel (A) restrict to census tracts with above median LR. Columns 2 and 4 in panel (A) restrict to census tracts with below median LR. A tract is considered eligible under UAG if the ratio ($\frac{Census\ Tract\ Median\ Family\ Income}{MSA\ Median\ Family\ Income}$) is less than 90 percent. "Targeted" refers to UAG-eligible census tracts. I use a bandwidth of 1 percent around the RD cutoff and fit quadratic polynomials on either side of the RD cutoff to determine the RD estimate. Panel B uses the same set of census tracts used in Panel A and shows the difference-in-difference by restricting to the "Targeted" (or UAG eligible) census tracts (columns 1 and 3) and to the "Not-Targeted" (or UAG ineligible) census tracts (columns 2 and 4). Panel (B) shows the difference-in-difference estimates — corresponding to the coefficient on "LR Index" — with the respective dependent variables as in Equation 1.5. Data on number and volume of loans aggregated to the census tract level is from Housing and Urban Development (HUD) for the years 2000 to 2005. All plots include control variables and MSA-year fixed effects. Control variables percentage with less than high-school education, percentage Hispanic, percentage black and unemployment rate are from the 2000 Census and 1990 Census. All regressions are weighted by total number of households from the 2000 Census for the years 2000–2002 and the 1990 Census for the years 2003–2005. Standard errors are clustered at the state level. See Appendix A.1 for details on LR index construction. See the Online Appendix for details on variable construction.

Table 1.7: Private Market: Mortgage Interest Rates and Non-rate Rationing

	(1)	(2)	(3)	(4)	(5)
	Rate Sheet	Ex-post: Loan Level			
	Rates	Rate	DTI	LTV	Rate
LR Index	-0.0520***	-0.0774**	-1.241	-0.00934	-0.0551**
	(0.0288)	(0.0288)	(0.905)	(0.395)	(0.0234)
Observations	112456	135101	19541	135133	65034
MSA-Yr FE	X	X	X	X	X
R^2	0.818	0.639	0.407	0.0343	0.649
Regression Type	OLS	OLS	OLS	OLS	CEM

This table shows the variation of mortgage rates and non-price terms — debt-to-income (DTI) and loan to value (LTV) — with the lender rights (LR) index for non-GSEs (private securitizers). Columns 1–4 show the OLS estimates with no controls. In column 5, I show the estimates from the coarsened exact matching (CEM, see [45] and [46]) procedure. In column 5, I use the CEM procedure to match mortgages exactly based on different bins in a given MSA-year-quarter. The bins for LTV, credit score (FICO) and DTI are the same as in Table 1.4. Data is for the period 2000 to 2005. The dependent variable in columns 1, 2 and 4 is mortgage rates. The dependent variable in columns 3 and 4 is DTI and LTV respectively. Rate sheet data on mortgage rates in column 1 is from `RateWatch.com`. Loan level non-GSE data in columns 2–5 for loans above the conforming loan limit (CLL) are the 30-year Fixed Rate Mortgage (FRM) single-family mortgage data from ABSnet. All columns include MSA-year-quarter fixed effects and standard errors are clustered at the state level. See Appendix A.1 for details on LR. See the Online Appendix for details on variable construction.

Table 1.8: Interaction with the Private Market: Mortgage Interest Rates

	(1)	(2)	(3)	(4)
	Above Median LR		Below Median LR	
$\mathbb{1}_{Jumbo}$	0.377***	0.237***	0.506***	0.399***
	(0.0224)	(0.0572)	(0.0522)	(0.0104)
Observations	135906	4272	104937	6475
MSA-Yr FE	X	X	X	X
R^2	0.837	0.697	0.801	0.619
Regression Type	OLS	CEM	OLS	CEM

Standard errors in parentheses, clustered by state.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the differences in mortgage rates across the conforming loan limit (CLL) for above median and below median lender rights (LR). A loan is classified as jumbo if the loan amount is greater than the CLL. Loans below the CLL are GSE loans and all loans above the CLL are the non-GSE loans. The coefficient on $\mathbb{1}_{Jumbo}$ is the jumbo-conforming spread. Columns 1 and 3 show the OLS estimates. In columns 2 and 4, I show the estimates from the coarsened exact matching (CEM, see [45] and [46]) procedure. For CEM, I match mortgages exactly based on different bins in a given MSA-year-quarter. The bins for loan-to-value (LTV), credit score (FICO) and debt-to-income (DTI) are the same as in Table 1.4. Data is for the period 2000 to 2005. The dependent variable in all columns is mortgage rates. Loan level GSE data are the 30-year Fixed Rate Mortgage (FRM) single-family mortgage data provided by Fannie Mae and Freddie Mac. Loan level non-GSE data are the 30-year Fixed Rate Mortgage (FRM) single-family mortgage data from ABSnet. All columns include MSA-year-quarter fixed effects and standard errors are clustered at the state level. See Appendix A.1 for details on LR. See the Online Appendix for details on variable construction.

Table 1.9: Spillovers in the Private Mortgage Market

	(1)	(2)	(3)	(4)
	Including GSE loans (Non-Jumbo)		Only Non-GSE loans (Non-Jumbo/Jumbo)	
	$\mathbb{1}_{GSE}$	$\mathbb{1}_{GSE}$	$\mathbb{1}_{ARM}$	$\mathbb{1}_{CM}$
LR Index	0.0145**	0.0218***	0.0423*	0.0729***
	(0.00658)	(0.00577)	(0.0237)	(0.0219)
FICO> 720	0.110***	0.0747***	-0.321***	-0.0475**
	(0.00833)	(0.00408)	(0.0102)	(0.0173)
LR Index*FICO> 720		-0.0153**		
		(0.00632)		
DTI<= 36	0.120***	0.220***	-0.144***	-0.00698
	(0.0138)	(0.0149)	(0.0169)	(0.0149)
LR Index*DTI<= 36		-0.0209*		
		(0.0119)		
LTV	-0.0000891***	-0.0000745***	0.0000257***	0.0000124***
	(0.00000994)	(0.00000374)	(0.00000747)	(0.00000414)
LR Index*LTV		-0.0000158***		
		(0.00000333)		
MSA-Yr-Qtr FE	X	X	X	X
Controls	X	X	X	X
Observations	770432	927506	278162	205057
R^2	0.604	0.547	0.160	0.302

This table shows the spillover effects of GSE credit rationing on the complexity of loans. This table shows the impact of LR on the intensive margin using a Linear Probability Model (LPM) of whether a loan is more likely to be purchased by the GSEs (column 1–2) or purchased by a private securitizer. Column 3 looks at whether a loan is more likely to be an adjustable rate mortgages (ARMs) compared to fixed rate mortgages (FRMs). Columns 4 looks at whether loans are more likely to be interest only (IO) or negative amortization (NEGAM) compared to FRMs. IO and NEGAM mortgages are together termed as Complex Mortgages (CM). In columns 2 the dependent variable is an indicator for a complex mortgage (IO/NEGAM) or an FRM. To look at heterogeneity, I also look at an indicator for whether the FICO is above 720 and an indicator for whether the debt-to-income is below 36. As in the previous tables, loans between 70 percent and 130 percent of the CLL are retained. Column 1–2 retain only loans below CLL and columns 3–4 retain only loans above CLL. Data is for the period 2000 to 2005 and from ABSnet. Controls included are whether a loan is full documentation loan, whether the property is owner occupied, and whether it is a refinance loan. Column 2 looks at the interaction between LR Index and the heterogeneity terms. All columns include MSA-year fixed effects. In all regressions both jumbo and non-jumbo loans are included. See Appendix A.1 for details on LR index construction. See the Online Appendix for details on variable construction.

Table 1.10: Overall Impact: Private Market and Originations in the Primary Mortgage Market

Panel A: Credit Rationing in the non-GSE (private) market

	(1)	(2)	(3)
	All	Minority	Single-female
LR Index	0.203***	0.0690	0.161***
	(0.0561)	(0.0473)	(0.0424)
MSA-Year FE	X	X	X
Controls	X	X	X
Observations	84116	39568	66942
R^2	0.619	0.329	0.459

Panel B: Ln(Originations/Housing Stock)

	(1)	(2)	(3)
	All	Minority	Single-female
LR Index	0.167**	0.263***	0.152***
	(0.0747)	(0.0758)	(0.0492)
MSA-Year FE	X	X	X
Controls	X	X	X
Observations	86048	77417	85346
R^2	0.631	0.454	0.555

Panel C: Denials

	(1)	(2)	(3)
	All	Minority	Single-Female
LR Index	-0.0184*	-0.0236**	-0.0214**
	(0.0102)	(0.00958)	(0.0104)
MSA-Year FE	X	X	X
Controls	X	X	X
Observations	7690532	1068134	2366237
R^2	0.0162	0.0289	0.0182

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table analyses the overall impact of credit rationing in the non-GSE market (Panel A), total originations (panel B) and denial rates (panel C) in the primary mortgage market. The dependent variable in panel (A) and (B) is the logarithm of the total number of loans per housing stock for each subset of borrowers aggregated to the census tract level. In panel (A) only the non-GSE (private securtizer) purchases are included and in panel (B) originations in the primary market are included. In all panels, Column 1 includes all borrowers. Columns 2 and 3 (in all panels) restrict to only minority and single female borrowers respectively. Data on number originations aggregated to the census tract level are from Home Mortgage Disclosure Act (HMDA) for the years 2000 to 2005. Panel C examines denial rates at the loan level. The dependent variable in panel (C) is an indicator variable for whether a loan was denied. Columns 2 and 3 restrict to only minority and single female borrowers respectively. All regressions include control variables and MSA-year fixed effects.

Chapter 2

Homeownership and the American Dream

2.1 Introduction

> *" No person, even the President, can ... guarantee you ... [that] you will always have the job you have today ... But we can guarantee to people that we're going to empower them to help themselves. We'll make home ownership more accessible."*
>
> – President Bill Clinton (1995)

> *"We can put light where there's darkness, and hope where there's despondency in this country. And part of it is working together as a nation to encourage folks to own their own home."*
>
> — President George W. Bush (2002)

Owning a home has long been considered an integral part of achieving the American Dream and ensuring upward mobility for ones' family and children. As the above quotes indicate, the belief in owning your home has been strongly held across the political spectrum.[1] Numerous policy measures since the 1930s, especially with the establishment of Fannie Mae, have aimed at increasing homeownership rates. More recently, policy measures have been enacted to ensure lending to underserved groups. The Federal Housing Enterprise Financial Safety and Soundness Act ("GSE Act") encourages lending by the Government Sponsored Enterprises (GSEs) — such as Fannie Mae and Freddie Mac — to low-income and minority families. The older Community Reinvestment Act ("CRA Act") of 1977 has a similar mandate and applies to all banking institutions that receive the Federal Deposit Insurance (FDIC), not just the GSEs.

[1]See below for articles discussing policies on homeownership of President Bill Clinton and President George Bush during their respective presidential terms in the period leading up to the crisis. `http://www.nytimes.com/2008/12/21/business/worldbusiness/21iht-admin.4.18853088.html?pagewanted=all` `http://spectator.org/articles/42211/true-origins-financial-crisis`

Much of the existing research on the effect of homeownership confirms this notion. For example, [40] show that children of homeowners are less likely to drop out of high school and have lower rates of teenage pregnancies. [19] find that homeowners are less likely to be unemployed (though they also have lower wages). A large older strand of literature has examined the impact of homeownership rates on various other outcomes. Homeownership is one of the main sources of wealth accumulation and provides insurance against rising housing costs ([68]). Homeownership is also associated with better housing quality and satisfaction ([74]).

In addition to these individual (direct) benefits of homeownership, a large body of research points to indirect benefits and positive externalities. [24], for example, find that homeowners are more likely to be involved in local government and, thus, areas with high homeownership have higher social capital. They argue that home-owning gives individuals the incentive to invest in the community. High homeownership rates have also been related to increased housing prices, possibly through the channel of higher maintenance ([20], [39]).

The recent crisis, however, has challenged the rationale behind this drive to increase homeownership and has drawn attention to harmful lock-in effects. Recent homeowners have witnessed plummeting house prices and increased foreclosures ([66], [63]). Negative equity, in turn, reduces household mobility or household ability to migrate, which was especially detrimental during the Great Recession ([28], [28],[81]).

Of course, such lock-in effects of homeownership are always present, not only during a crisis, and they are in fact the mechanism behind some of its merits, such as improved social capital: homeownership creates barriers to residential mobility, thereby encouraging individuals to invest more in their community ([24]). If we take this logic seriously, it implies that the encouragement of homeownership in areas that provide for little upward mobility is counterproductive as it ties family and children to those disadvantaged areas. For example, in areas with segregated living, high homeownership rates may exacerbate the effect of living in a bad neighborhood. Policies that encourage homeownership based on aggregate characteristics should take this heterogeneity into account.

In this paper, we test whether homeownership is related to intergenerational mobility and, if so, whether this relationship persists in areas associated with fewer opportunities, such as high-sprawl and high-segregation areas. Both the 1992 GSE Act and the 1977 CRA Act encourage lending to low income individuals. In addition, these policies also explicitly encourage lending in low income *neighborhoods*.[2] Given the emphasis on encouraging homeownership in underserved low-income areas, its benefits ought to be as strong if not stronger.

We use the comprehensive data on intergenerational mobility provided by [15] and [17] which covers the entire US as well as the corresponding homeownership rates, which we

[2]Specifically, The 1992 GSE Act designates census tracts where median family income is less than 90% of the median family income of the MSA as underserved. The GSE 1992 Act mandates that a certain portion of the GSE lending be targeted to these underserved areas. Similarly, the CRA Act defines areas as underserved if the median family income is less than 80% of the median family income of the MSA.

obtain from the 2000 Census. Previously, data limitations on homeownership and measures of children's outcomes have made the analysis of homeownership effects and location-based heterogeneity challenging. Prior work has used survey data such as the Panel Study of Income Dynamics (PSID) and National Longitudinal Surveys (NLSY) ([40], [43]). While one advantage of survey data is that they provide outcomes at the individual level, they have very limited geo-coded data (such as at the MSA-level for the public use data) and are based on only a subsample of individuals.

The data provided on intergenerational mobility by [17] – and in particular its causal component provided by [17] – are ideally suited for our purposes. The measure is based on confidential individual level federal income tax records of nearly 40 million children and their parents. The intergenerational mobility measure links children's income to parents' income. This intergenerational mobility measure of the permanent residents is comprised of both the sorting and the causal component. rents' income. [15] estimate the impact of neighborhoods on intergenerational mobility. This causal estimate focuses on the subset of 5 million families who move across neighborhoods in the US.[3] The paper estimates the causal effect of growing up in a commuting zone (CZ) by using a fixed effects model identified from the families that move. [15] also decompose overall intergenerational mobility measures of neighborhoods into two components, the causal component and the sorting component. The causal component of intergenerational mobility measures the causal impact of growing up in a neighborhood. The sorting component measures the intergenerational mobility of children in a given neighborhood whose outcomes would have been the same regardless of where they grow up. We focus on the causal component of the impact of neighborhoods on intergenerational mobility. We look at the across-CZ effect of homeownership rates and intergenerational mobility. We use the causal component in the [15] data at the commuting zone (CZ) level. To capture heterogeneity of children from different income backgrounds, we analyze two groups of children, namely those with below-median income parents and those with above-median income parents. In our baseline results, we first look at across-CZ variation and find that higher homeownership rates in 2000 is associated with higher intergenerational mobility of children. The effect is driven by the causal component of intergenerational mobility. Specifically, homeownership rates are higher where the causal effect of growing up in a neighborhood is higher. For children growing up in families at the 25^{th} percentile a 1 standard deviation higher homeownership rate is also associated with a 0.37 percentile increase in income rank. We use an instrumental variables strategy to instrument for homeownership rates and confirm our findings.

[3]This is a follow-up to a previous paper, [17]. [17] track children born in 1980–82 (1980–82 birth cohorts) and measure the parent income in 1996–2000 when the children are from 14–16 years old. All children are ranked at the national level based on their income in 2011–2012. Similarly parents are ranked at the national level based on their mean income between 1996–2000. The intergenerational mobility measure is the rank-rank relationship between children's income and parents' income. [17] find that this rank-rank relationship between mean child ranks and parent ranks is almost perfectly linear. These estimates in [17] merely represent the intergenerational mobility measure for all children at the given levels of geography and do not attempt to disentangle the causal effect of growing up in a neighborhood from sorting effects.

First, we instrument for homeownership rates in 2000 using the stock of single family detached homes in 1990 as a percentage of all housing structures in 1990. [39] find that the structure of housing, specifically, single family detached dwellings are a good measure of owner occupied housing. [39] use the MSA level of single family detached homes in 1980 to instrument for homeownership at the individual level.[4] Following [39] we use the stock of single family detached homes in 1990 as a percentage of all housing structures in 1990 as an instrument for homeownership rates in 2000. We find that a 1 standard deviation higher percentage single family detached homes in 1990 is associated with nearly 0.0782 standard deviation higher homeownership rates in 2000. Additionally, 20 years of exposure to a CZ with 1 standard deviation higher instrumented homeownership rate in 2000 increases a child's income rank by 0.707 percentile for those in below-median income families. This translates to almost 2.4 percent higher earnings. For above-median income individuals the effect is almost twice as high. A 1 standard deviation higher instrumented homeownership rate in 2000 increases child's rank by 1.410 percentile for above-median income families.

Second, we supplement the analysis by using the median house price shock between 1980 to 1990 as an instrument for homeownership rates in 2000. Higher house prices are associated with lower homeownership rates. We use the difference in the median house price in 1990 and 1980 as a measure of the affordability of owning a home in 2000. Our results are similar to using the other instrument, and children's ranks in CZs with one standard deviation higher homeownership rates results in a 0.812 percentile higher income rank. The effect for children with above-median income parents is higher at 0.915 percentile increase in child income rank.

Our main goal is to examine the large geographic heterogeneity in the impact of higher aggregate homeownership rates on children's outcomes across the US.

First, we examine how the impact of homeownership rate on intergenerational mobility varies by segregation and by sprawl. [36] notes that policies that encourage homeownership implicitly encourage people to move away from higher density living towards areas with more sprawl. Areas with high sprawl, however, might be associated reduced positive effects of homeownership on social capital. For example, high voter turnouts and involvement in local communities may be diminished in sprawling areas due to the higher costs of social interaction. Sprawl may make it harder to access jobs and to experience income mobility.

Sprawl may also be associated with more segregated living as well as with difficulties in accessing to grocery stores, retail and schools. Additionally, homeownership results in reduced household mobility and homeownership may exacerbates the impact of living in bad neighborhoods especially in highly segregated areas.

For our analysis we use the measure of sprawl that [15] use. Sprawl is measured as the fraction of people (not working from home) with greater than 15 minutes of commute time to work. Since the focus of our analysis is to capture a measure of sprawl more closely linked to segregated living, the commuting distance based sprawl measure is ideally suited for our

[4] Specifically, they examine the benefits of the home mortgage interest deduction and find that the deduction is particularly poor instrument for encouraging homeownership.

purposes. We find that the positive impact of homeownership on intergenerational mobility is diminished in areas with high sprawl. Possibly, the positive spillovers associated with high social capital and high homeownership rates on children is diminished in more sprawling and segregated areas where there are higher costs to interacting with people. We find that for children from below-median income families the negative effect of homeownership in areas with high sprawl dominates, with the positive effect of homeownership disappearing almost completely in some specifications. For children from above-income families, the positive impact of homeownership on intergenerational mobility persists across all specifications and is diminished by 37–40 percent in areas with a 1 SD higher sprawl.

To examine what aspect of sprawling areas is driving this heterogeneity in impact of homeownership rates on intergenerational mobility, we turn to two distinct measures of segregation, namely racial segregation and segregation by income. For the racial segregation measure we use the [78] measure which captures how different on average is the racial composition of each census tract within a CZ compared to the racial composition of the entire CZ. Growing up in a neighborhood with 1 SD higher racial segregation causes a 37 percent reduction in the positive impact of homeownership rates on intergenerational mobility of below-median income families (when we instrument using the house price shocks between 1980–90). The effect for children from above-median income families is a similar 35 percent reduction for children from above-median income families.

Our measure of income segregation based on [73]. This measure of segregation captures the uneven distribution of income levels within a CZ. Intuitively, this measure captures how different the income distribution in each census tract is on average from the income distribution of the entire CZ. We find that in areas with high level of income segregation, the positive impact of homeownership rates is diminished. That is, for children from below-median income families, growing up in neighborhoods with 1 SD higher segregation diminishes the positive impact of homeownership rates by 40 to 55 percent (for the instrumented regressions). For children from above-median income families, living in a neighborhood with 1 SD higher segregation of income reduces the positive impact of homeownership rate by a lower 32 to 38 percent. For the lower income families, the negative impact of sprawl seems to be driven by segregation of *income* rather than racial segregation.[5]

We then examine the channels through which homeownership leads to higher intergenerational mobility. [24] find that homeowners invest more in social capital and homeownership may encourage higher investment in local amenities. [38] find that there are social benefits to homeownership. Motivated by this literature, we look at the impact of homeownership rates on social capital. To proxy for social capital we use the index from [75] also used in [15]. This index is constructed using the response rate to the Decennial Census, the voter turnout rates in the presidential elections and number of tax-exempt non-profit organizations (representing community involvement). We find that a 1 SD higher homeownership rate is also associated with a 0.220 SD higher value of the social capital index. To examine

[5] [15] also look at sprawl and segregation of income measures and find that high sprawl and high segregation is associated with lower intergenerational mobility.

why high sprawl and segregation leads to reduced positive effects of homeownership at the CZ level, we look at heterogeneity of the impact of homeownership rate on social capital. A 1 SD higher homeownership rate in an area with a 1 SD higher sprawl results in a 20 percent reduction in the social capital index. Similarly, segregation of income reduce the social capital index by 15 percent respectively. Consistent with our results on heterogeneity for sprawl the negative impact of sprawl on social capital is driven by segregation of income rather than *racial* segregation.

Our paper is organized as follows. Section 2.2 explains the data used in our analysis. Section 2.4 shows our baseline estimates estimating the impact of homeownership rates on intergenerational mobility. Section 2.5 looks at the cross-sectional heterogeneity across areas with differing sprawl. Section 2.6 shows the cross-sectional heterogeneity across areas with differing segregation of income. Section 2.7 looks at the impact of homeownership rates on social capital. Section 2.8 concludes.

2.2 Data and Summary Statistics

Data

We use data at the commuting zone (CZ) level mainly from the data provided by [15] and from the Census 2000. Additional data used and their sources are described below. For our analysis we focus on the CZ level analysis since it covers the entire US as opposed to Metropolitan Statistical Areas (MSAs) which cover only urban areas. Additionally, there is less sorting across CZs than at the more granular county level. Hence, we focus mainly on CZ level analysis.

Intergenerational Mobility measure

In an earlier paper, [17] use administrative records on the incomes of around 40 million children and their parents to describe features of intergenerational mobility in the United States. The main focus of their paper is on the geographical or spatial variation in intergenerational mobility. In the subsequent paper by [15], the authors build on this measure and provide causal estimates of growing up in a neighborhood. We use these causal estimates of intergenerational mobility of a CZ from [15] in our analysis.

The mobility measures in [15] track children born between 1980–91 (1980–91 birth cohorts). Parent income is measured as the average family income from 1996 to 2000. For the estimates we use, cohort (children) income is recorded when the child is 26 years old.[6] The children's age when the parents' income is measured will thus vary across cohorts. [15] then rank parents based on their position in the *national* income distribution. Similarly, they rank children — within a cohort — at the national level. They find that the rank-rank

[6] [15] use this as the baseline measure. However, they also provide estimates for other ages of outcome measurement and find that all yield very similar estimates.

relationship between parents' income rank and children's income rank to be almost perfectly linear.[7]

To get an estimate of the causal effect of growing up in a neighborhood, [15] focus on the subset of families that move. We describe their estimation procedure below. The following discussion closely follows Section VII in the [15] paper. To get an estimate of the causal effect of growing up in a neighborhood, first [15] subset to the families that move. Let T_C represent the age at which children enter the labor market. Let y_i be the outcome of the child when adult. In our estimates this is the child's income rank at age 26. Children's outcome is a function of family input, neighborhood characteristics and the disruption costs of moving. Let μ_{pc} denote the causal effect of growing up in a neighborhood. Let the mean level of parental inputs to child i be $\bar{\theta}_i$.[8] First, [15] make the simplifying assumption that disruption costs do not vary across neighborhoods. Let $\bar{\kappa}_0$ be the disruption costs of moving.[9]

Second, they assume that neighborhood effects are additive. Focusing only on the first-time movers, who move from origin o to destination d at age m, [15] model the child's outcome as a simple linear exposure time specification as below:

$$y_i = (T_C - m)(\mu_{pd}) + m\mu_{po} + \bar{\theta}_i + \bar{\kappa}_0 \tag{2.1}$$

where μ_{pd} is the causal effect of growing up in the destination d with parental income at percentile p. Analogously, μ_{po} is the causal effect of growing up in the destination o with parental income at percentile p.

They make a third assumption that for all origin-destination pairs the choice of when to move is independent of other inputs $\bar{\theta}_i$ conditional on origin and destination. Intuitively, this says that there is no sorting for any origin-destination pair.

The parental inputs $\bar{\theta}_i$ can be decomposed into a component which is origin-destination pair specific and a residual as follows:

$$\bar{\theta}_i = \alpha_{odps} + \eta_{1i} \tag{2.2}$$

where η_{1i} is independent of exposure time to the origination and destination and α_{odps} captures variation in outcomes across parent income (p), cohort (s), origin (o) and destination (d). In their empirical specification they parameterize separate controls for each origin-destination pair with a linear control for income and a quadratic term for cohort. Adding the cohort controls ensures that they are controlling for the fact that outcome for different cohorts is measured at different years.

This motivates their empirical model as follows:

$$y_i = (T_C - m)[(\mu_d^0 + \mu_d^P p)1\{d(i) = d\} - (\mu_o^0 + \mu_o^P p)1\{o(i) = o\}] + \alpha_{odps} + \eta_{1i} \tag{2.3}$$

[7]Specifically, [17] first showed that the rank-rank is almost perfectly linear. [15] builds on this analysis.

[8]This is the average parental input across the entire childhood.

[9]All [15] need is that the disruption costs do not vary in a differentially age-dependent manner across neighborhoods. For heterogeneous disruption costs, one can think of $\bar{\theta}_i$ as incorporating these disruption costs.

Thus, for every origin-destination pair, [15] estimate a regression of child outcomes on exposure time to the destination $T_C - m$,

$$y_i = (T_C - m)(\mu^0_{od} + \mu^1_{od} p) + \alpha_{odps} + \eta_{2i} \tag{2.4}$$

where $\mu^0_{od} + \mu^1_{od} p$ gives the estimate of spending an additional year of childhood in destination d relative to origin o. α_{odps} includes controls for parental income and cohort described above.

Let $\mu^p_{od} = \mu^0_{od} + \mu^1_{od} p$ for each origin-destination pair at percentile p. Them the causal effect of each place μ_{pc} can be estimated from the regression of

$$\mu^p_{od} = G\mu_{pc} + \eta_{3od} \tag{2.5}$$

where G is a matrix with the rows representing origin-destination pairs and columns representing the unique places (N_c). For each row the origin column is coded as a -1 and the destination is coded as $+1$. With this, we get the estimates of causal effect of growing up in a neighborhood as μ_{pc}. Note, each row sums to zero, since each entry will have a $+1$ for destination and -1 for origin. Since, the matrix G does not have full rank, the impact of exposure to places is measured relative to one omitted place. μ_{pc} is normalized to have a population-weighted value of zero. Then, μ_{pc} can be interpreted as the effect of exposure to a place (CZ) c relative to where the average population lives.

Intuitively, the procedure can be described as follows. Specifically, they first focus on the population of residents who move across CZs to determine μ_{pc}. Second, they use a exposure-time identification strategy to identify the fixed effects using the movers in the sample. The intuition of how the estimates are constructed is clearer from the following example. Consider families who move from Phoenix to Oklahoma. If children of families who moved at younger ages had higher outcomes when adult compared to children who moved later, then one can posit that this is due to the causal effect of growing up in Oklahoma is higher relative to Phoenix. To claim that the effect if causal they need the assumption that the timing of the moves is orthogonal to the children's potential outcomes.

The above procedure gives the causal effect of growing up in a neighborhood (μ_{pc}) which is the main focus of our analysis. We also supplement the analysis by looking at the sorting component of intergenerational mobility.

To get the sorting component, they focus on the permanent residents, that is, families that never move. The intergenerational mobility measures for the permanent residents represents both the causal effect of growing in a neighborhood and a sorting component, that is, differences in the characteristics of the families that reside in these CZs.

To determine the intergenerational mobility measures for the permanent residents, y_{pc}, first, they rank at national level child i (in cohort s) based on their income, y_i. Similarly, they rank at national level parents of these children based on their incomes, p_i. The intergenerational mobility measure is then rank-rank relationship between parents' income rank and children's income rank for each CZ

Thus, they estimate the relationship between child rank (y_i) and parents' rank (p_i) as:

$$y_i = \alpha_{cs} + \psi_{cs} p_i + \epsilon_i \tag{2.6}$$

They find that this rank-rank relationship is almost perfectly linear in all CZ's c. Expected rank of a child in cohort s whose parents' national income rank is p and are permanent residents of CZ c is then given by:

$$\hat{y}_{pcs} = \hat{\alpha_c} + \hat{\psi}_{cs} p \tag{2.7}$$

Thus, the above gives an estimate of the intergenerational mobility for permanent residents which comprises of both the sorting and causal effect of growing up in a CZ. To decompose the observed outcome of permanent residents into a sorting and causal component, we need to make an assumption of the total relevant exposure time, T_C. The selection component of the permanent residents is then $\hat{\theta}_{pc} = \bar{y}_{pc} - T_C * \hat{\mu}_{pc}$. The mean selection effect depends on the assumption about T_C. We use $T_C = 20$ year exposure as in [15]. In our analysis, we focus on the intergenerational mobility measure for children of parents at the 25^{th} and 75^{th} percentile for which the causal component measure os available. Additionally, focusing on both these percentiles allows us to look at the heterogeneity of the effects of homeownership we observe for both the low and high income families. Note that given the linearity of the rank-rank relationship, 25^{th} and 75^{th} percentile measures correspond to the average outcomes of children from below-median and above-median income families.

Instrument 1: Single Family Detached Homes

In our analysis we two different instruments to instrument for homeownership rates in 2000. The first instrument we use is the fraction of single family detached homes to the total housing units in 1990. This instrument has been used in prior literature to instrument for individual level of homeownership rates. [39] instrument for individual homeownership in 1993 using MSA level % of single family detached homes in 1980. Following this idea, we instrument for CZ-level homeownership rates in 2000 using CZ-level percentage of single-family detached homes to the total housing units in 1990. We use the data from the 1990 Census to calculate the fraction of single family detached homes to the total housing units in a CZ.

Instrument 2: House price shock 1980–1990

The second instrument we use in our analysis is the median house price shock between 1980 to 1990. 1990 approximately corresponds to the affordability of homeownership when parents (mothers) are around 35 years of age.[10] The average age of first-time home buyers is between 31 (National Association of Realtors) to 34 (2009 American Housing Survey). Thus, the house price shock corresponds to roughly when the parents of the children in our

[10] This estimate of the mother's age is based on the [17] sample.

analysis become homeowners. We estimate the house price shock as the difference in the median value of the house in 1980 to 1990. Median house price data is from the 1990 Census and 1980 Census.

Measure of Sprawl

For our cross-sectional heterogeneity results we use the measure described in [15]. We use the fraction of people not working from home with greater than 15 minutes of commute time to work. [35] use sprawl to describe cities where people need to drive large distances to conduct their daily lives. Sprawl in this case is higher wherein people need to drive large distances for employment, or in other words, cities in which employment is very decentralized. The commuting time based sprawl measure can be thought of as capturing this version of sprawl. The advantage of using this sprawl measure is that it is constructed using the 2000 Census and thus has the most extensive geographic coverage. Additionally, we are interested in a measure of sprawl that more closely captures the effect of living in more segregated areas and the commuting time based measure of sprawl more accurately captures this.

Measures of Segregation

We use the same measures of segregation as in [15]. For racial segregation, we use the [78] measure and for segregation of income we use the measure from [73].

The [78] measure of segregation at the CZ level uses the census tract level data from the 2000 Census. Let $\phi(r)$ be the fraction of individuals of a race r in a CZ. In the analysis, we consider the following racial groups: black, white, Hispanic and others.

At the CZ level, the racial diversity is given by the entropy index

$$E = \sum_r \phi_r log_2 \frac{1}{\phi_r} \tag{2.8}$$

For each tract j, across race r, the level of racial diversity is given by the entropy index:

$$E_j = \sum_r \phi_{rj} log_2 \frac{1}{\phi_{rj}} \tag{2.9}$$

The degree of racial segregation at the CZ level is then given by

$$H = \sum_j \frac{population_j}{population_{CZ}} \frac{E - E_j}{E} \tag{2.10}$$

where $population_j$ and $population_{CZ}$ respectively refer to the tract and CZ level population. Intuitively the segregation measure here measures how different the racial distribution of each census tract is from the CZ. $H = 1$ corresponds to the highest level of segregation and $H = 0$ corresponds to when there is no racial segregation at all.

For segregation of income, we use the measure in [73]. The segregation of income uses a measure analogous to the one above. The idea is to look at the population in different percentiles of income as opposed to the different racial groups in the [78] index. We measure the degree to which the population below the p^{th} percentile is segregated from the population above the p^{th} percentile. Let p denote the fraction below the p^{th} percentile.

The two-group entropy index is then given by:

$$E(p) = plog_2\frac{1}{p} + (1-p)log_2\frac{1}{1-p} \tag{2.11}$$

The index $H(p)$ at the CZ level for each percentile p is then given by

$$H(p) = \sum_j \frac{population_j}{population_{CZ}} \frac{E(p) - E(p)_j}{E(p)} \tag{2.12}$$

The overall income segregation is then given by:

$$Income\ Segregation_p = 2log(2)\int_p E(p)H(p)dp \tag{2.13}$$

This measure is also provided by [15] and they use the 2000 Census data income data to get a measure of the segregation of income.

In our analysis we use segregation of homeowners. This is similar to the [78] index, except we consider only two groups the homeowners and renters at the census tract level. At the CZ level, the entropy index for each tenure (homeowners or renters) is

$$E = \sum_t \phi_t log_2\frac{1}{\phi_t} \tag{2.14}$$

For each tract j, across tenure t (which is either homeownership rate of renters), the level of tenure diversity is given by the entropy index:

$$E_j = \sum_t \phi_{tj} log_2\frac{1}{\phi_{tj}} \tag{2.15}$$

The degree of segregation of homeowners at the CZ level is then given by

$$Segregation\,of\ Homeowners = \sum_j \frac{population_j}{population_{CZ}} \frac{E - E_j}{E} \tag{2.16}$$

where $population_j$ and $population_{CZ}$ respectively refer to the tract and CZ level population. Intuitively the segregation measure here measures how different the tenure (homeowner versus renter) distribution of each census tract is from the CZ. A measure of 1 corresponds to the highest level of segregation of homeownership. Intuitively, our measure of segregation of homeownership is analogous to the [78] measure described above but with only two races. The only difference is we look at the fraction of homeowners (versus renters) in a given census tract.

Other variables

Our main independent variable of interest is the homeownership rate. We use the Census 2000 to measure the homeownership rate at the CZ level. The other control variables included in our analysis are percentage of population below the poverty level, percentage female, percentage divorced and percentage black. All control variables data is from the 2000 Census. For weighting the data we use the number of housing units in each CZ from the Census 2000.

We also use a social capital index at the county level which is provided by [15] and is from [75]. The social capital index is constructed based on voter turnout rates, fraction of people who return their census forms, and other measures of participation in community organizations at the county level. This measure is then aggregated up to the CZ level.

CZs for which all the above data is available were used in our analysis. We look at 588 CZ is our final analysis. Since most of the data is available from the Census 2000, most of the data limitation is imposed by the number of CZs for which the causal effect of intergenerational mobility measure from [15] is available.

Summary Statistics

Table 2.1 gives the summary statistics of the variables used in our analysis. Data are at the county level and there are 588 CZs for which all data is available.

The causal component of intergenerational mobility is the income rank of the children in percentiles — relative to the mean across all CZ — of the children of parents at the 25^{th} and 75^{th}. The causal effect of growing up in a neighborhood for children of parents at the 25^{th} percentile is 3.69 percentiles. For children at the 75^{th} percentile this measure is 2.45 percentiles. Figure 2.1, Panel A shows the spatial variation of the data. We see that there is substantial regional variation in intergenerational mobility.

Average homeownership rate in 2000 was at 71.31 percent with a standard deviation of 5.51 percent. However, the minimum and maximum homeownership rates are between 43.53 percent to 84.41 percent displaying a wide range of variation across US states similar to the intergenerational mobility measure. Figure 2.1, Panel B shows the spatial variation of homeownership rates in 2000. Again, we see that there is substantial regional variation in homeownership rates.

On average CZs have around 14.26 percent of population with people below the poverty line, 9.8 percent divorced, 9.4 percent black, 21 percent single mothers and 26 percent with age above 55.

We use the fraction of all housing structures which are single family detached units in 1990 as the first instrument for homeownership rates. On average, around 68 percent of all housing structures are single family detached units in CZ. The second instrument, the difference in median hose prices between 1980-1990 ranges from a decline of \$15,400 to an increase of \$167,070 from 1980 to 1990. On average, median house prices increased \$17, 670 between 1980 to 1990.

To analyze the cross-sectional heterogeneity of impact of homeownership against intergenerational mobility, we look sprawl. We use the measure of sprawl from [15], the fraction of people not working from home with more than 15 minutes of commute time to work. On average around 59 percent of the population lives more than 15 minutes of their place of work. However, there is a wide range from a low as 24 percent of the population to a high of 84 percent of the population living at large commuting distance.

We look at three additional measures of segregation, namely, segregation of homeowners, racial segregation and segregation of income. The segregation of homeowners ranges from 0 to 0.31 percent. These low values of segregation of homeowners imply that homeownership on average does not seem to be segregated at the CZ level. Our racial segregation measure based on the [78] index was on average 14 percent. There was a wide range for this index too from 1 percent to 48 percent. Segregation of income was on average 4.61 percent. Income on the other hand tends to be much more segregated compared to segregation of homeowners.

The social capital index ranges from -3.2 to 3.07. This measure from [75] is constructed based on voter turnout rates, fraction of people who return their census forms and other measures of participation in community organizations. Low values of the index correspond to low social capital.

We weight all our regressions using the total number of housing units in 2000. On average the CZs in our analysis had 177,226 housing units. The size of the counties captured in our analysis varies widely as can be seen from fact that total number of housing units in the CZ varied from 8166 housing units to CZs with more than 5 million housing units.

2.3 Empirical Methodology

Our main regression specifications test for the link between intergenerational mobility and homeownership. All regression specifications are at the commuting zone level. Intergenerational mobility is calculated from [15]. The mobility measures track children born between 1980–91 (1980–91 birth cohorts). Parent income is the average family income from 1996 to 2000. The children's age when the income is measured will vary across cohorts. Cohort (children) income is recorded at age 26. The causal component of growing up in a neighborhood for 20 years is measured for children from below-median income families, that is at the 25^{th} percentile and for above-median income families, that is, at the 75^{th} percentile. Homeownership data is from the US Census Bureau and is as of 2000.

Baseline Specification

The baseline empirical specification is as follows:

$$Intergenerational\ Mobility_c = \beta_0 + \beta_1 * Homeownership\ Rate2000_c + \gamma X_c + \epsilon \tag{2.17}$$

All data is at the CZ level c. We repeat this analysis for each of our mobility measures, that is for children from below-median income families, that is at the 25^{th} percentile and for above-median income families, that is, at the 75^{th} percentile. For ease of interpretation we standardize the homeownership rate variable. The controls included are percentage below poverty level, percentage female, percentage divorced and percentage black in the CZ. All regressions are clustered at the state level. All regressions are weighted by the number of housing units in a county in 2000 to get representative estimates of the US population.[11] In all our specifications we show the weighted least squared regressions. While there is a loss of efficiency using the weighted estimators ([22], [13], [8]), this criticism only applies when the treatment effect is homogenous. Since the treatment effect of the homeownership rates on intergenerational mobility is heterogeneous — as we will also empirically establish later — we show the weighted estimate results. All results remain qualitatively the same in the unweighted estimates.

The above regression specification, however, only establishes causality. In the subsection below we describe the instruments we use for homeownership rates and provide some justification for their validity.

Instrumenting for homeownership

Our second set of specifications instrument for homeownership using the single family detached homes in 1990. [38] use the stock of single family detached homes at the MSA level in 1980 as an instrument for homeownership at the individual level in 1990.[12] The idea is that the housing structure is generally a good predictor of homeownership.

Following the same logic, we use the single family detached homes in 1990 to instrument for homeownership rates in 2000. Our regression specification is as follows.

The first stage:

$$Homeownership\ Rate2000_c = \delta X_c + \rho * Fraction\ of\ Single\ family\ detached\ homes\ 1990_c + \epsilon_c \tag{2.18}$$

The second stage instruments for homeownership:

$$Intergenerational\ Mobility_c = \theta X_c + \beta * \widehat{Homeownership}_{2000_c} + \eta_c \tag{2.19}$$

$Homeownership\ Rate2000_c$ represents the homeownership rate in 2000 at the CZ level. Equation 2.18 represents the first stage, where the instrument is the CZ-level fraction of single family detached homes. All standard errors are clustered at the state level. Equation 2.19 represents the second stage using the instrumented homeownership rate. We include CZ level controls.

[11] Note, we also used the number of children in the [15] sample for weighting and results remain quantitatively and qualitatively the same.

[12] [19] use a similar instrument to test the impact of housing tenure on labor market outcomes.

Figure 2.4, Panel A graphically shows the binned scatter plots of fraction of single family detached house in 1990 against homeownership rates in 2000. This is analogous to the first stage of the instrumented regression, except without the controls. High fraction of single family detached homes in 1990 also predict high homeownership rates in 2000.

We also instrument for homeownership rates using the median house price shock in 1980 to 1990 as an instrument for homeownership rates in 2000. We use the median house price shock between 1980–1990 as a measure of the affordability of owning a home in 2000. The average age of the mothers in the [17] is 41 in 1996. Thus, the 1990 median house price value corresponds to the affordability of the house when parents (mothers) are around 35 years of age. According to the 2009 American Housing Survey data the average age of the first-time home buyers was 34. Another survey conducted recently by the National Association of Realtors also estimates the average age of the first-time home buyers to be 31 years. Thus, using the median house price shock between 1980 to 1990 as a measure of affordability of owning a home seems reasonable. This instrument aims to capture the effect of owning a home. However, note the effect of homeownership that we capture will include both the individual impact of homeownership and the aggregate impact of homeownership rate on intergenerational mobility. Figure 2.4, Panel B shows the first stage results. Higher house price shocks are associated with lower homeownership rates. The first stage and second stage specification is similar to Equation 2.18 and Equation 2.19.

We also tried instrumenting for homeownership rates using the [76] measure. The [76] instrument has been recently used to instrument for housing prices ([66], [63]). The [76] measure calculates the fraction of land unavailable for development due to steep slopes and bodies of water. The hypothesis is that single family detached homes may be easier to build compared to multi-family structures where land availability is higher. The first stage results are robust, that is, unavailability of land is inversely correlated with single family detached homes. However, given the large cross-sectional heterogeneity that we find across sprawl the [76] instrument was particularly bad at predicting the impact on intergenerational mobility. Another way to say this is that the exclusion restriction is violated because sprawl (loosely, the inverse of the unavailability measure) also affects intergenerational mobility. For a recent critique of using the [76] measure as an instrument for house prices, see [21]

Cross-sectional Heterogeneity: Difference-in-difference specification

In Section 2.5 and Section 2.6 we look at the cross-sectional heterogeneity of the effect of homeownership rates on intergenerational mobility. In Section 2.5 we examine how the impact of homeownership rate on intergenerational mobility varies by sprawl or the spread of cities. In Section 2.6 we also examine cross-sectional heterogeneity with segregation of homeowners, racial segregation and segregation of income. We explicitly show the empirical specifications below for the heterogeneity with the sprawl measure. The other empirical specifications simply replace the sprawl measure with the respective interaction terms namely segregation of homeowners, racial segregation and segregation of income.

The specification for the instrumented cross-sectional heterogeneity using the fraction of single family detached homes in 1990 is as follows.

The first stage:

$$\begin{aligned} Homeownership\,Rate2000_c = \delta X_c + \rho * Fraction\,Single\,family\,detached\,homes\,1990_c + \\ \omega * Fraction\,Single\,family\,detached\,homes\,1990_c * Sprawl + \epsilon_c \end{aligned} \tag{2.20}$$

The second stage instruments for homeownership:

$$\begin{aligned} Intergenerational\,Mobility_c = \theta X_c + \beta * \widehat{Homeownership}_{2000_c} \\ + \tau * \widehat{Homeownership}_{2000_c} * Sprawl + \eta_c \end{aligned} \tag{2.21}$$

For ease of interpretation we standardize the homeownership rate and interaction variable. As before the above specifications include CZ-level controls and are weighted at the state level. The specification for house price shock is similar to the above.

2.4 The link between homeownership and intergenerational mobility

We now turn to our main empirical analysis and examine the relationship between homeownership and intergenerational mobility. As a first step of our analysis, we wish to link homeownership to intergenerational mobility. Prior literature has found that owning a home leads to better outcomes for children ([40]. We first present the baseline estimates of the link between homeownership rates and the casual impact of living in a neighborhood.

Figure 2.5 shows the relationship between average intergenerational mobility and the homeownership rate in 2000 weighted by the population in each CZ. The dependent variable is the causal component of intergenerational mobility measure for children of parents from the 25^{th} percentile (panel (a)) and 75^{th} percentile (panel (b)) from [15]. Due to the linearity of the rank-rank relationship between parents' incomes and children's incomes, this corresponds to the intergenerational mobility measure of the children with parents below the median income and of parents above the median income. Higher values of intergenerational mobility correspond to higher intergenerational mobility. Figure 2.5 shows that there is a strong positive relationship between the two variables for children from below-median income families and children from above-median income families.

Table 2.2 looks at this relationship more formally. To get good estimates of heterogeneity across groups, we look at two different measures of intergenerational mobility.[13] We look

[13] [15] provides these two measures of intergenerational mobility for childhood exposure effects of living in a CZ.

at the impact of homeownership rates on the below-median income backgrounds (columns 1–4) and on children with above-median income backgrounds (columns 5–8). The variables for homeownership rate has been standardized for ease of interpretation. All columns are weighted by the number of housing units in each CZ in 2000 and are clustered at the state level. Except for columns 1 and 5, all specifications include CZ-level controls.

In Panel A we focus on the causal effect of growing up in a CZ. Specifically, the dependent variable in Panel A is the causal effect of growing up in a CZ for twenty years. Twenty years of exposure to a CZ with 1 standard deviation higher homeownership rate is associated with a 0.728 increase in the child's income rank for families with below-median income. Including the controls percentage with age above 55, percentage single working mothers,percentage below poverty level, percentage divorced, percentage with less than High school education, unemployment rate, percentage black in the CZ and an indicator for whether the CZ is an urban area reduces the impact of homeownership rate to a 0.368 increase in income rank. A 0.601 percentile increase in income translates to a roughly 1.16 percent increase in earnings. For above-income families, children growing up in areas with a 1 standard deviation higher homeownership rate causes the child's income rank to increase by 0.909 percentiles. Including the controls, however, makes this statistically insignificant.

In columns 3 and 7 we instrument for homeownership rate in 2000 using the stock of single family detached homes in 1990 as a percentage of all housing structures in 1990. Instrumenting for homeownership rates in column 3 shows that a one standard deviation higher instrumented homeownership rate in 2000 results in 0.707 percentile increase in children's rank from below-median income families. This is similar in magnitude to the results from the OLS regressions in columns 1–2, though slightly higher. For above median-income families (column 7), children growing up in CZs with a 1 standard deviation higher homeownership rate causes children's income to increase by 1.410 percentiles.

In columns 4 we use the median house price shock between 1980–1990 as an instrument for homeownership. Twenty years of exposure to a CZ with a 1 standard deviation higher homeownership rate increases the children's rank by 0.812 percentiles for below-median income families and by 0.914 percentiles for above median income families.

Our hypothesis is that higher homeownership rates affect intergenerational mobility through two channels. First, homeownership directly leads to better outcomes for children by providing higher stability ([40]. Second, homeownership are better citizens and this provides positive externalities which in turn affect intergenerational mobility ([38]). While it is difficult to completely disentangle the two effects, in Table 2.2 Panel B we try to explore whether the second indirect channel is at play. Specifically, we construct a measure of segregation of homeownership (an entropy based measure). We find that higher the segregation of homeownership, it lower is the intergenerational mobility. Additionally, this negative effect is stronger for the below median income families. For the above median income families this effect is insignificant when we add CZ level controls. This suggests that the second indirect channel through the positive externalities of homeownership is important for intergenerational mobility and more so for poorer families.

2.5 Does the impact of homeownership on intergenerational mobility vary by sprawl?

In the previous section, we saw that higher homeownership rates is associated with higher intergenerational mobility consistent with the findings in prior literature. We next look at whether there are place-based differences in the impact of homeownership rates on intergenerational mobility. In this section we explore the cross-sectional heterogeneity of the effect of homeownership rate on intergenerational mobility. We examine how the impact of homeownership rates on intergenerational mobility varies by sprawl or the spread of cities. [36] notes that policies that encourage home-owning implicitly encourage people to move away from higher density living. Thus, sprawl is intricately linked with homeownership. The hypothesis is that areas with high sprawl also diminish the positive effects associated with homeownership. Many of the positive effects of homeownership such as the high social capital — for examples, more investment in local amenities and higher involvement in local communities — may be more diminished in more sprawling areas. Sprawl may also be associated with that more segregated living. Additionally, homeownership results in reduced household mobility and homeownership exacerbates the impact of living in bad neighborhoods.

We use the measures of sprawl that [15] use.[14] Sprawl is measured as the fraction of people — not working from home — with more than 15 minutes of commute time to work.[15] We use this measure of sprawl since it is based on the 2000 Census has the most expansive coverage across the US. This index implicitly measures the version of sprawl considered in [35]. Sprawl in this case is higher wherein people need to drive large distances for employment, or in other words, cities in which employment is very decentralized. Additionally, we are interested in measure of sprawl that more closely captures the effect of living in more segregated areas.

Figure 2.6 examines the heterogeneity of the effect of homeownership rate on intergenerational mobility for CZs across areas with differing sprawl. We split the CZs into terciles based on the sprawl measure. The top tercile corresponds to “high sprawl” and “low sprawl” corresponds to the bottom tercile. In Figure 2.6 we restrict to CZs with homeownership rates in 2000 above the first percentile and below the 99^{th}. In panel (a) we see that consistent with the previous findings, high homeownership rates is associated with high intergenerational mobility for children of parents with income at the 25^{th} percentile in low sprawl areas. However, in high sprawl areas this relationship is reversed. Higher homeownership rate is in fact associated with lower intergenerational mobility. For children of parents from above-median income backgrounds, we see that homeownership is associated with higher causal

[14] Sprawl is not the main focus of the [15] paper. They club it together with the segregation measure and find that high sprawl is associated with lower intergenerational mobility.

[15] For some recent coverage of the relationship between intergenerational mobility and sprawl see: `http://www.nytimes.com/2013/07/29/opinion/krugman-stranded-by-sprawl.html?_r=0)` `http://bettercities.net/article/intergenerational-mobility-vs-sprawl-there-connection-20382` `http://www.newgeography.com/content/003868-distortions-and-reality-about-income-mobility` `http://realestateresearch.frbatlanta.org/rer/2013/08/does-sprawl-really-limit-income-mobility.html`

impact of growing up in a neighborhood for both high sprawl and low sprawl areas.

In Table 2.3, Panel A, we examine this relationship more formally. Columns 1–3 show the cross-sectional heterogeneity across areas with sprawl for children of below-median income parents. The variables homeownership rates and sprawl measure have all been standardized for ease of interpretation. Column 1 shows the simple OLS results. For the below median income families, the positive coefficient on homeownership disappear (columns 1–2) and is strengthened for the above median income families indicating there is important cross-sectional heterogeneity in the impact of homeownership. Accounting for the heterogeneity with respect to sprawl increases the coefficient on homeownership rates across all specifications for the above median income families. Growing up in a CZ with one standard deviation higher homeownership rate is associated with no impact on the child's rank for below-median income families (column 1). The interaction term with the measure of sprawl, our coefficient of interest, is also negative. This indicates that for low income families, the impact of homeownership is negative in CZs with high sprawl. That is, a one standard deviation higher homeownership rates of a CZ in more sprawled cities leads to a 0.56 percentile lower income rank for children from below-median income families. In column 3 when we instrument for homeownership rates using the house price shock, we find that the positive effect of higher homeownership rates on intergenerational mobility for below-median income children is diminished by 65 percent in areas with 1 SD higher sprawl. The coefficient on the sprawl measure is also negative indicating that high sprawl cities are in general associated with low intergenerational mobility which is consistent with the findings in [15]. The direct impact of living in areas with high commute times (high sprawl) decreases the causal impact of living in a neighborhood on children's incomes by 2.6 percentiles. As [15] note, that this is the impact on the children's outcomes while they are growing up and hence does not directly correspond to their commute times when adult. Thus, commute times are capturing some characteristic of the CZ that is driving this relationship.

The impact on the children for the above-median income families is starkly different. Growing up in a CZ with one standard deviation higher homeownership rate is associated with a 1.639 percentile increase in the child's rank for above-median income families. Looking at the interaction term, the positive impact of homeownership rates diminishes in CZs with high sprawl. That is, a one standard deviation higher homeownership rates of a CZ in more sprawled cities leads to a 0.671 percentile lower income rank for children from above-median income families. For children from above-median income families, the positive effect of higher homeownership rates on intergenerational mobility is diminished by 41 percent (compared to only a negative effect of homeownership for below-median income families as seen in column 1) in areas with 1 SD higher sprawl. As before, the coefficient on the sprawl measure is also negative a 1 SD higher sprawl measure decreasing the causal impact of living in a neighborhood on children's incomes by 1.291 percentiles (column 6) and insignificant in some specifications (columns 4 and 5).

We next explore the idea of disentangling the direct and indirect effects of homeownership. Figure 2.7 examines the heterogeneity of the effect of homeownership rate on intergenerational mobility for CZs across areas with segregation of homeowners. This segregation of homeowners measure can be thought of as capturing the externalities of homeownership. Though it is not completely possible to disentangle the pure externalities of homeownership from the direct effect, the segregation of homeownership captures the idea that being surrounded by *more* homeowners should amplify (or exacerbate) the effects of homeownership. In Figure 2.7 we split the CZs into terciles based on the segregation of homeowners measure. The top tercile corresponds to "high segregation of homeowners" and "low segregation of homeowners" corresponds to the bottom tercile. In panel (a) we see that consistent with the previous findings, high homeownership rates is associated with no impact on intergenerational mobility for children of parents with income at the 25^{th} percentile in low segregation of homeowners areas. However, in areas with high segregation of homeowners) this relationship is slightly reversed. For children of parents from above-median income backgrounds, we see that homeownership is associated with higher causal impact of growing up in a neighborhood for both high and low segregation of homeowners.

In Panel B, we further explore the idea of disentangling the direct and indirect effects of homeownership. As in Panel A, we find that the direct effect of homeownership is insignificant in the un-instrumented regression for below median income families in Column 1. In fact the negative effect of segregation of homeownership dominates. Instrumenting with single family detached homes shows the same results. When we instrument using house prices, we find that there is both a direct positive effect on intergenerational mobility through higher homeownership rates. As before there is a negative effect of segregation of homeownership. Additionally, there is also heterogeneity across segregation (of homeownership). Thus, if homeownership is more geographically segregated, then the externalities of homeownership have a much lower impact on intergenerational mobility.

In contrast, for the above median families the direct effect of homeownership rates dominates. However, this effect is diminished in areas with high segregation of homeownership. Surprisingly, there is no negative effect of segregated homeownership for these above median income families across all specifications. This result is consistent with Table 2.2 Panel B. Thus, for the above median income families the direct effect of homeowning dominates and for the poor families the indirect effect of homeownership dominates.

2.6 Does the impact of homeownership on intergenerational mobility vary by segregation?

In the previous section, we documented large place-based heterogeneity across areas with varying sprawl. High sprawl areas may also be associated with high segregation. To augment the analysis in the previous section, we look at place-based heterogeneity across segregation. We examine both racial segregation and segregation of income.

We first graphically examine the relationship between segregation and the effect of homeownership rates on intergenerational mobility. Analogous to our analysis in Figure 2.6, we split the CZs into terciles based on the racial segregation. The top tercile corresponds to "high segregation" and "low segregation" corresponds to the bottom tercile. In Figure 2.8 we restrict to CZs with homeownership rates in 2000 above the first percentile and below the 99^{th}. In panel (a) the relationship between high homeownership rates and intergenerational mobility for children of parents with income at the 25^{th} percentile is weak in both the high and low segregation areas. For children of parents from above-median income backgrounds, we see that homeownership is associated with higher causal impact of growing up in a neighborhood especially in the low segregation areas. In high segregation areas, homeownership rates still have a positive effect on intergenerational mobility, though the effect is slightly lower. Looking at the segregation of income in Figure 2.9 shows very similar effects.

In Table 2.4, Panel A we explore the heterogeneity across areas with varying racial segregation. High homeownership rate is associated with high intergenerational mobility as can be seen from the coefficient on homeownership rate. The direct impact of growing up in a neighborhood with 1 standard deviation higher homeownership rates for the children from below-median income families ranges from 1.716–2.359 percentiles. For children from above-median income families, the effect on children's outcomes is between 1.714–2.699 percentiles.

The direct effect of living in a racially segregated neighborhoods zero across all specifications. This is consistent with [15] who find racial segregation to have a lower impact on intergenerational mobility. The interaction term of the [78] measure of racial segregation and homeownership rates has a negative coefficient though significant only when we instrument using the house price shocks between 1980–1990. Thus, neighborhoods with 1 standard deviation higher racial segregation and 1 standard deviation higher homeownership rate results in a reduction in intergenerational mobility of 0.892 percentiles for children from the below-median income families. Thus, the overall impact of 1 standard deviation higher homeownership rate in a neighborhood with 1 standard deviation higher racial segregation neighborhood is a 1.467 percentile lower income for the children from the below-median income families. For children from the above-median income families, we find that the positive effects of homeownership is significant across specifications. A 1 standard deviation higher racial segregation in a neighborhood with 1 standard deviation higher homeownership reduces the positive impact of homeownership rate by 0.606 percentiles. Instrumenting for homeownership rates yields very similar results though the effect is slightly higher ranging from 1.019 to 1.136 percentile reduction in above-median income children.

In Table 2.4, Panel B we explore the heterogeneity across areas with differing segregation of income. First, the causal effect of living in CZs with higher homeownership rates is associated with higher income of children as can be seen from the coefficient on homeownership rates. Twenty years of exposure to CZs with 1 standard deviation higher homeownership rates causes child's income rank to increase between 2.73 and 4.42 percentile for children from below-median income families. For above-median income families this ranges from 1.64 to 5.94 percentiles.

Second, the causal effect of living in CZs with higher segregation of income is lower in areas with higher segregation of income as can be seen from the coefficient of segregation of income. This is consistent with [15].

Third, we see that 20 years of exposure to CZs with 1 standard deviation higher segregation of income and 1 standard deviation higher homeownership rates decreases a child's income rank by 1.158 to 1.781 percentiles for below-median income families (though it is insignificant for the un-instrumented regression). This corresponds to 40 to 49 percent reduction in child's income. For above-median income families, children's income rank increases by 0.53 percentiles.

2.7 Potential channels

In this section we examine the potential channels through which homeownership can impact intergenerational mobility. We examine two potential channels. First we look at the impact of homeownership on social capital. Second, we also look at whether higher school quality explains the positive impact on intergenerational mobility of children.

Social Capital

We now supplement our empirical analysis by examining the channels through which homeownership impacts intergenerational mobility. Homeownership is associated with higher outcomes for the children of homeowners. In our analysis, we focused on place-based differences that can operate through the aggregate impact of homeownership. [38] find that there are large social benefits to homeownership. [24] find that homeowners are more likely to be involved in local government and areas with high homeownership have higher social capital. To test this channel of the impact of homeownership we analyze whether areas with high homeownership rates are also associated with high social capital. In our setting, we proxy for social capital using the social capital index constructed by [75] (as in [15] and [17]). The index is constructed using voter turnout rates, the fraction of people who return their census forms, and other measures of participation in community organizations. The CZ level social capital index measure is constructed by population weighting the county level measures provided by [75]. This measure of social capital index is similar in spirit to the involvement in local government and community explored in [24].

Table 2.5 shows the results of this analysis. In columns 1–2, we first relate homeownership rates to social capital. For ease of interpretation, we standardize both homeownership rates and the social capital index. Column 1 shows that a 1 standard deviation higher CZ level homeownership rate is associated with a 0.220 standard deviation increase in the social capital index. All columns are weighted by the number of housing units in each CZ in 2000 and are clustered at the state level. Adding CZ-level controls yields very similar results. A 1 standard deviation higher homeownership rates is associated with a 0.281 standard deviation increase in the social capital index measure.

In columns 3 we instrument for homeownership rate in 2000 using the stock of single family detached homes in 1990 as a percentage of all housing structures in 1990. The second stage results in column 4 shows the estimates using the instrumented homeownership rates. A 1 standard deviation higher instrumented homeownership rate in 2000 results in 0.540

standard deviation increase in the social capital index. This is consistent with the results from the OLS regressions in columns 1–2 though higher in magnitude.

In columns 4 we use the increase in median house price between 1980–1990 as an instrument for homeownership. 1990 corresponds to when the parents buy houses. The increase in median house price between 1980–1990 can be thought of as a measure of how affordable homeownership is in a particular area. Column 4 shows the second stage results. The second stage is similar to the previous results. A one standard deviation increase in homeownership rate is associated with a higher social capital index of 0.355 standard deviation.

In this subsection we confirmed the positive link between homeownership rates and the social capital index. Advocates of pro-homeownership policies cite the positive externalities of homeownership as a reason for encouraging homeownership. Next, we examine whether the heterogeneity of the impact of homeownership rates on social capital. Note, in Section 2.6 and 2.5 we established that the impact of homeownership rates on intergenerational mobility diminishes in areas with high segregation and high sprawl. In Table 2.5 we look at the heterogeneity of the relationship between social capital and homeownership rates. The hypothesis is that segregation and sprawl decrease the impact of homeownership rates on social capital. In panel A, we examine the heterogeneity with respect to sprawl as measured by the fraction of people with less than 15 minutes commuting time to work. As before, higher homeownership rate is associated with higher value of the social capital index. All variables have been standardized for ease of interpretation. Neighborhoods with high fraction of people with commuting times greater than 15 minutes is associated with higher social capital. Higher homeownership rates in areas with high sprawl *decreases* social capital. A 1 standard deviation higher homeownership rate in a a neighborhood with a one standard deviation higher sprawl is associated with 0.056 standard deviation reduction in the social capital index. Instrumenting for homeownership rates yields very similar results.

In Panel B, we look at heterogeneity across the segregation of income. Consistent with the results in Panel A, we find that areas with high homeownership rates and high segregation of income have lower social capital though the results are noisy. Similarly, in panel C, we look at heterogeneity across racial segregation using the [78] index and find that high homeownership rate in highly racially segregated areas is associated with low social capital.

School Quality

Another potential channel through which homeownership can impact intergenerational mobility is through higher quality schools. [40] find that children of homeowners have lower dropout rates. However, the causality in this case is not very clear. It is possible that homeownership also implies greater stability due to lower residential mobility. Additionally, homeowners can also influence school funding through voting. However, it a reverse causality implies that parents of children move to CZs with higher school quality (and subsequently higher intergenerational mobility).

In table 2.6 we explicitly analyze the link between schools quality and homeownership. The dependent variable is a binary variable qual to one if the CZ is of above median (better quality) quality based on an income-residualized measure of test scores provided by [15]. Higher test scores correspond to better quality schools.[16] Panel A shows that high homeownership is indeed associated with higher quality schools. Thus, the positive impact of homeowners may be driven by the parents moving to CZs with better schools.

In Panel B we now turn to look at the cross-sectional heterogeneity of homeownership with our sprawl and segregation measures. However, across all specifications, we find no effect of higher segregation and higher homeownership on school quality. Thus, the cross-sectional heterogeneity in Section 2.5 and Section 2.6 cannot be explained by higher quality of schools and high homeownership rates. Thus, these results also address concerns that a reverse causality — wherein parents of children move to CZs with higher school quality — may be driving our results.

2.8 Conclusion

In this paper we relate homeownership to children's upward mobility. We establish a positive relationship, on average, but also significant cross-sectional heterogeneity depending on sprawl. We find that in areas with higher sprawl there is a lower impact of homeownership on intergenerational mobility. We also find that in neighborhoods with high segregation, higher homeownership is associated with lower intergenerational mobility, possibly through reduced residential mobility of households. Our results caution against encouraging homeownership based on prior evidence of benefits of homeownership. Instead, policies aimed at encouraging homeownership should take into account the important place-based heterogeneity across the US.

[16] We focus on this measure of school quality because it has the widest coverage. Our results remain qualitatively the same even when we use the two other measures of school quality provided by [15], namely, dropout rates and ratio of number of pupils to teachers.

Table 2.1: Summary Statistics

We present the summary statistics of all variables used in our analysis. Intergenerational mobility measures from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile and 75^{th} percentile of the parents' income distribution. Variables percentage with homeownership rate, all housing units, age above 55, percentage single working mothers, percentage below poverty level, percentage divorced, percentage with less than High school education, unemployment rate, percentage black in the CZ and an indicator for urban area are from the U.S. Census in 2000. Percentage single family detached units is from 1990 Census. Difference in median house price value between 1980 to 1990 is from 1980 and 1990 Census respectively. Sprawl is defined as the fraction of people not working from home with greater than 15 minutes of commute time to work and from Census 2000. Segregation of homeowners is an entropy-based measure calculated at the CZ level using Census 2000 data. Racial segregation is measured using [78] Index. Income segregation is measured as in [73], calculated at the at the CZ level with Census 2000 data provided by [15]. Social capital index is from [75] and provided by [15]. School quality is an indicator equal to one if the CZ is above median (better quality) for an income-residualized measure of test scores at CZ level provided by [15]. Urban area and school quality are binary variables.

	Mean	SD	Min	Max
Intergenerational mobility measure (25^{th} percentile)	3.69	12.29	-48.32	66.81
Intergenerational mobility measure (75^{th} percentile)	2.45	13.83	-89.60	69.03
Homeownership Rate 2000	71.31	5.51	43.53	84.41
% Age above 55	26.42	3.31	16.61	37.84
% Singe mothers	20.93	4.94	8.21	43.37
% Divorced	9.83	1.55	4.23	14.57
% < High School Graduates	21.81	7.44	6.98	57.87
% below Poverty Level	14.26	5.24	5.51	35.68
Unemployment Rate	5.02	1.55	2.36	17.70
% Black	9.47	13.12	0.01	66.36
% Urban (Binary)	54.42	49.85	0	100
% Single family detached units (1990)	68.36	8.40	26.90	86.18
Difference in median HP $(80-90) in'000s$	17.67	19.11	-15.40	167.07
Sprawl(% with Commute > 15 min.)	58.79	11.06	24.48	84.39
Segregation of homeonwers	0.09	0.06	0.00	0.31
Theil Index	14.34	9.00	0.67	47.63
Segregation of Income	4.61	3.15	0.38	13.79
Social Capital Index	-0.07	1.13	-3.20	3.07
Top Tercile School Quality (Binary)	76.19	42.63	0	100
All Housing Units 2000	177226	412976	8166	5355469
Observations	588			

Table 2.2: Intergenerational Mobility and Homeownership

We present the OLS and IV estimates of the effects of homeownership on intergenerational mobility. The instruments are CZ-level fraction of single family detached homes in 1990 (IV1) and the difference in median house price value between 1980 to 1990 (IV2). All columns include controls percentage with age above 55, percentage single working mothers, percentage below poverty level, percentage divorced, percentage with less than High school education, unemployment rate, percentage black in the CZ and an indicator for urban area. The dependent variable from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (columns 1–5) and 75^{th} percentile (columns 6–10) of the parents' income distribution. In columns 5 and 10 we show the relationship between segregation of homeownership and intergenerational mobility. Columns 5 and 10 also include the homeownership rate as a control. Segregation of homeowners is an entropy-based measure calculated at the CZ level using Census 2000 data. Columns 1–2 and 6–7 show the OLS estimates. Columns 3–4 and columns 8–9 show IV estimates. All other data are from the US Decennial Census. All columns are weighted by the number of housing units in each CZ in 2000. The homeownership rate and segregation of homeownership variables has been standardized for ease of interpretation. Standard errors are clustered by state.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	(25^{th} Percentile)					(75^{th} Percentile)				
	OLS	OLS	IV1	IV2	OLS	OLS	OLS	IV1	IV2	OLS
Homeownership Rate 2000	0.728^{***}	0.368^{*}	0.707^{*}	0.812^{***}	-0.234	0.909^{**}	0.622	1.410^{**}	0.914^{*}	0.614
	(0.230)	(0.208)	(0.409)	(0.263)	(0.294)	(0.424)	(0.395)	(0.715)	(0.475)	(0.432)
Segregation of homeowners					-1.189^{***}					-0.0159
					(0.391)					(0.503)
Number of Observations	588	588	588	588	588	588	588	588	588	588
R squared	0.0313	0.206	0.203	0.200	0.221	0.0390	0.0856	0.0715	0.0836	0.0856
Controls		X	X	X	X		X	X	X	X
Number of Clusters	49	49	49	49	49	49	49	49	49	49

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3: Intergenerational Mobility and Homeownership: Sprawl and Segregation of Homeowners

We present estimates of the cross-sectional heterogeneity in the effect of homeownership on intergenerational mobility across sprawl and segregation of homeowners. Columns 1 and 4 show the OLS results. Columns 2–3 and columns 5–6 show the IV results. The instruments are CZ-level fraction of single family detached homes in 1990 (IV1) and the difference in median house price value between 1980 to 1990 (IV2). The dependent variable from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (columns 1–3) and 75^{th} percentile (columns 4–6) of the parents' income distribution. All other data are from the US Decennial Census.

Panel A: Sprawl

	(1)	(2)	(3)	(4)	(5)	(6)
		(25^{th} Percentile)			(75^{th} Percentile)	
	OLS	IV1	IV2	OLS	IV1	IV2
Homeownership Rate 2000	0.675	0.515	2.771***	1.639**	4.341***	3.415***
	(0.505)	(1.221)	(0.750)	(0.648)	(1.343)	(0.962)
(Sprawl) * (Homeownership Rate 2000)	-0.555**	-0.868	-1.661***	-0.671***	-1.649***	-1.492***
	(0.218)	(0.569)	(0.415)	(0.244)	(0.480)	(0.406)
Sprawl	-2.591***	-3.369***	-3.306***	-0.929	-0.912	-1.291**
	(0.595)	(0.706)	(0.661)	(0.681)	(0.669)	(0.641)
Number of Observations	588	588	588	588	588	588
R squared	0.253	0.240	0.217	0.0968	0.0555	0.0788
Controls	X	X	X	X	X	X
Number of Clusters	49	49	49	49	49	49

Panel B: Segregation of Homeownership

	(1)	(2)	(3)	(4)	(5)	(6)
		(25^{th} Percentile)			(75^{th} Percentile)	
	OLS	IV1	IV2	OLS	IV1	IV2
Homeownership Rate 2000	0.243	0.751	2.207***	1.861***	6.175***	3.568***
	(0.499)	(1.343)	(0.663)	(0.638)	(1.975)	(0.668)
(Segregation of Homeowners) * (Homeownership Rate 2000)	-0.201	-0.452	-0.763***	-0.524***	-1.510***	-1.018***
	(0.131)	(0.307)	(0.186)	(0.173)	(0.445)	(0.239)
Segregation of Homeowners	-1.391***	-1.715***	-1.459***	-0.542	0.0288	-0.616
	(0.427)	(0.540)	(0.518)	(0.538)	(0.683)	(0.557)
Number of Observations	588	588	588	588	588	588
R squared	0.224	0.219	0.194	0.102	.	0.0836
Controls	X	X	X	X	X	X
Number of Clusters	49	49	49	49	49	49

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: Intergenerational Mobility and Homeownership: Racial and Income Segregation

We present estimates of the cross-sectional heterogeneity in the effect of homeownership on intergenerational mobility across racial and income segregation. Columns 1 and 4 show the OLS results. Columns 2–3 and columns 5–6 show the IV results. The instruments are CZ-level fraction of single family detached homes in 1990 (IV1) and the difference in median house price value between 1980 to 1990 (IV2). The dependent variable from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (columns 1–3) and 75^{th} percentile (columns 4–6) of the parents' income distribution.

Panel A: Racial Segregation

	(1)	(2)	(3)	(4)	(5)	(6)
	(25^{th} Percentile)			(75^{th} Percentile)		
	OLS	IV1	IV2	OLS	IV1	IV2
Homeownership Rate 2000	0.395	1.716*	2.359**	1.714***	3.748***	2.699***
	(0.452)	(1.040)	(0.943)	(0.486)	(1.165)	(0.516)
(Racial Segregation) * (Homeownership Rate 2000)	-0.0139	-0.516	-0.892**	-0.606***	-1.136***	-1.019***
	(0.177)	(0.376)	(0.367)	(0.167)	(0.371)	(0.243)
Racial Segregation	0.0359	-0.171	-0.351	0.208	0.0369	0.0310
	(0.282)	(0.273)	(0.328)	(0.259)	(0.308)	(0.293)
Number of Observations	588	588	588	588	588	588
R squared	0.206	0.188	0.162	0.105	0.0661	0.0966
Controls	X	X	X	X	X	X
Number of Clusters	49	49	49	49	49	49

Panel B: Income Segregation

	(1)	(2)	(3)	(4)	(5)	(6)
	(25^{th} Percentile)			(75^{th} Percentile)		
	OLS	IV1	IV2	OLS	IV1	IV2
Homeownership Rate 2000	0.160	2.739**	4.421***	1.643**	5.942***	4.572***
	(0.463)	(1.360)	(1.456)	(0.625)	(1.888)	(1.104)
(Income Segregation) * (Homeownership Rate 2000)	-0.0257	-1.158**	-1.781***	-0.530**	-1.991***	-1.732***
	(0.152)	(0.503)	(0.560)	(0.218)	(0.631)	(0.469)
Income Segregation	-0.847**	-1.274***	-1.345***	-0.578	-0.525	-0.912*
	(0.375)	(0.448)	(0.485)	(0.440)	(0.529)	(0.469)
Number of Observations	588	588	588	588	588	588
R squared	0.214	0.163	0.0877	0.0953	0.00263	0.0478
Controls	X	X	X	X	X	X
Number of Clusters	49	49	49	49	49	49

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: Potential Channels: Social Capital and Homeownership

Panel A present the OLS and IV estimates of the effects of homeownership on social capital. Panel B presents estimates of the cross-sectional heterogeneity in the effect of homeownership on social capital across sprawl and segregation.

Panel A

	(1) OLS	(2) OLS	(3) IV1	(4) IV2
Homeownership Rate 2000	0.220***	0.281***	0.540***	0.355***
	(0.0537)	(0.0399)	(0.141)	(0.0790)
Number of Observations	588	588	588	588
R squared	0.135	0.649	0.559	0.642
Controls		X	X	X
Number of Clusters	49	49	49	49

Panel B: Sprawl and Segregation

	(1) OLS	(2) IV1	(3) IV2
Sprawl			
HO Rate 2000	0.315***	0.910***	0.486**
	(0.0632)	(0.199)	(0.197)
(Sprawl) * (HO Rate 2000)	-0.0815***	-0.289***	-0.162**
	(0.0254)	(0.0681)	(0.0701)
Sprawl	-0.420***	-0.399***	-0.457***
	(0.0867)	(0.0680)	(0.0756)
R squared	0.708	0.587	0.698
Segregation of Homeowners			
HO Rate 2000	0.272***	1.199***	0.537***
	(0.0574)	(0.321)	(0.200)
(Segregation of Homeowners) * (HO Rate 2000)	-0.0563***	-0.256***	-0.130***
	(0.0133)	(0.0724)	(0.0415)
Segregation of Homeowners	-0.340***	-0.182**	-0.343***
	(0.0737)	(0.0909)	(0.0850)
R squared	0.701	0.397	0.675
Racial Segregation			
HO Rate 2000	0.350***	1.037***	0.537**
	(0.0681)	(0.280)	(0.235)
(Racial Segregation) * (HO Rate 2000)	-0.0380	-0.251***	-0.104
	(0.0231)	(0.0896)	(0.0882)
Racial Segregation	0.0169	-0.0615	-0.00954
	(0.0385)	(0.0572)	(0.0510)
R squared	0.654	0.413	0.636
Income Segregation			
HO Rate 2000	0.322***	1.429***	0.850***
	(0.0678)	(0.366)	(0.330)
(Income Segregation) * (HO Rate 2000)	-0.0469**	-0.407***	-0.256**
	(0.0207)	(0.118)	(0.110)
Income Segregation	-0.209***	-0.166*	-0.256***
	(0.0692)	(0.0914)	(0.0866)
R squared	0.672	0.304	0.583
Number of Observations	588	588	588
Controls	X	X	X
Number of Clusters	49	49	49

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Potential Channel: School Quality and Homeownership

Panel A present the OLS and IV estimates of the effects of homeownership on school quality. Panel B presents estimates of the cross-sectional heterogeneity in the effect of homeownership on social capital across sprawl and segregation.

Panel A

	(1) OLS	(2) OLS	(3) IV1	(4) IV2
Homeownership Rate 2000	0.105***	0.0861**	0.225***	0.171***
	(0.0243)	(0.0329)	(0.0651)	(0.0378)
Number of Observations	588	588	588	588
R squared	0.101	0.304	0.219	0.273
Controls		X	X	X
Number of Clusters	49	49	49	49

Panel B: Sprawl and Segregation

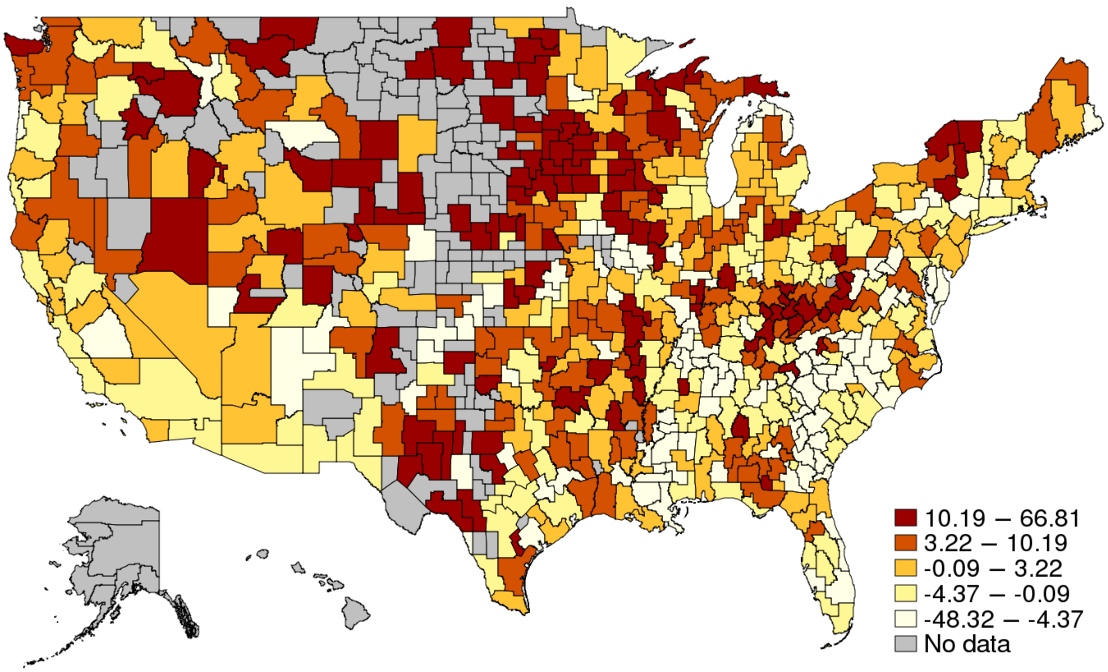
	(1) OLS	(2) IV1	(3) IV2
Sprawl			
HO Rate 2000	0.0467	0.278*	0.228*
	(0.0626)	(0.150)	(0.131)
(Sprawl) * (HO Rate 2000)	-0.0101	-0.0643	-0.0665
	(0.0255)	(0.0495)	(0.0527)
Sprawl	-0.201***	-0.138***	-0.180***
	(0.0419)	(0.0500)	(0.0459)
R squared	0.354	0.273	0.314
Segregation of Homeowners			
HO Rate 2000	0.0115	0.421*	0.211*
	(0.0630)	(0.254)	(0.112)
(Segregation of Homeowners) * (HO Rate 2000)	-0.00422	-0.0767	-0.0323
	(0.0145)	(0.0504)	(0.0248)
Segregation of Homeowners	-0.171***	-0.0559	-0.0946**
	(0.0386)	(0.0724)	(0.0471)
R squared	0.352	0.142	0.296
Racial Segregation			
HO Rate 2000	0.0862	0.269**	0.181***
	(0.0574)	(0.108)	(0.0638)
(Racial Segregation) * (HO Rate 2000)	-0.00225	-0.0306	-0.00705
	(0.0211)	(0.0379)	(0.0234)
Racial Segregation	-0.0647	-0.0685	-0.0607
	(0.0439)	(0.0480)	(0.0471)
R squared	0.319	0.235	0.286
Income Segregation			
HO Rate 2000	0.0239	0.234	0.115
	(0.0632)	(0.173)	(0.127)
(Income Segregation) * (HO Rate 2000)	0.00343	-0.0298	0.00626
	(0.0204)	(0.0524)	(0.0415)
Income Segregation	-0.173***	-0.102**	-0.112**
	(0.0411)	(0.0479)	(0.0459)
R squared	0.359	0.283	0.326
Number of Observations	588	588	588
Controls	X	X	X
Number of Clusters	49	49	49

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.1: Map of Intergenerational Mobility

The figures below show the heat maps for causal component of intergenerational mobility at the 25^{th} and 75^{th} percentile (Panel A and Panel B) at the CZ level. Data are divided into 5 quintiles are shown. Intergenerational mobility measure from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (Panel A) and 75^{th} percentile (Panel B) of the parents' income distribution.

Panel A: 25^{th} **Percentile**



Panel B: 75^{th} **Percentile**

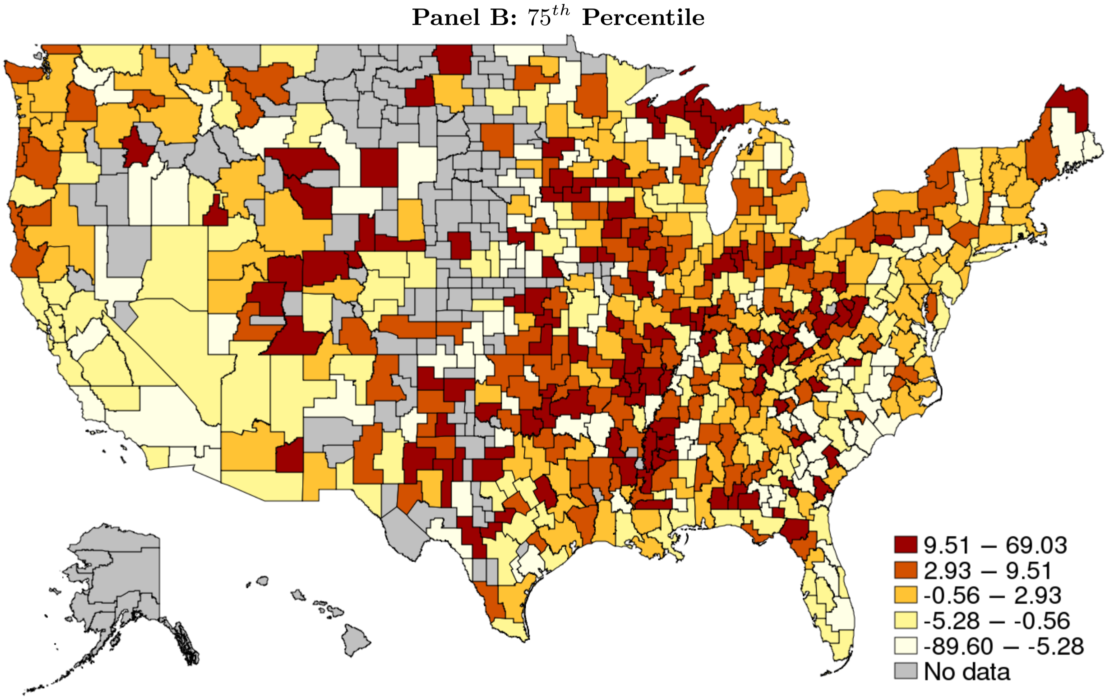


Figure 2.2: Map of Homeownership Rates 2000

The figures below show the heat maps for homeownership rate in 2000 at the CZ level. Data are divided into 5 quintiles are shown. Homeownership rates in 2000 are from the 2000 Census.

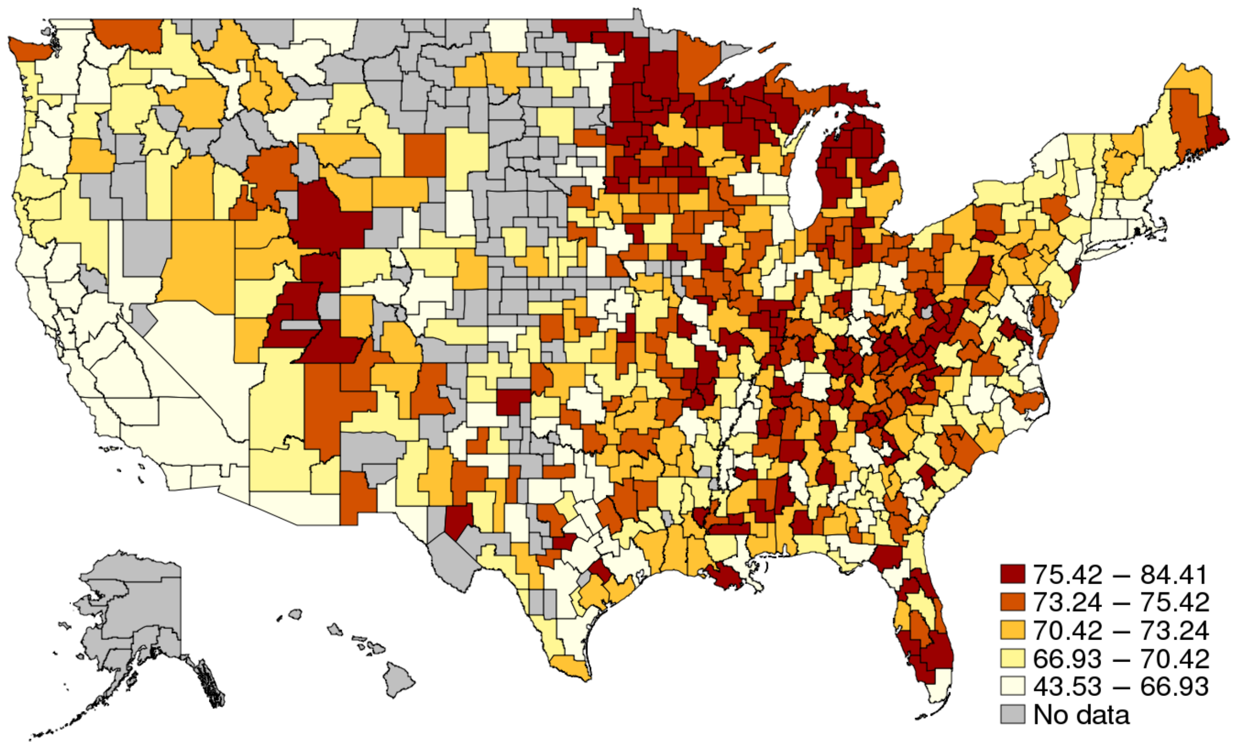
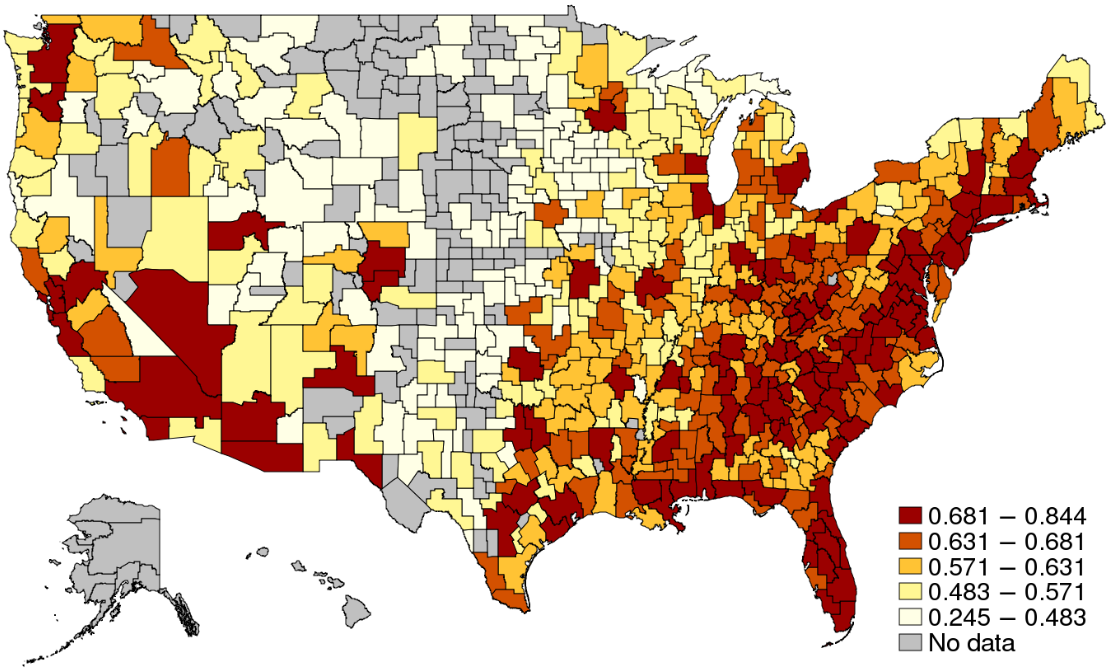


Figure 2.3: Map of Sprawl and Segregation

The figures below show the heat maps for the sprawl measure (Panel A) and segregation of income measure (Panel B) at the CZ level. Data are divided into 5 quintiles are shown. Sprawl is defined as the fraction of people not working from home with greater than 15 minutes of commute time to work and is from the 2000 Census. Income segregation is measured as in [73], calculated at the at the CZ level with Census 2000 data provided by [15].

Panel A: Sprawl measure (Fraction of population with more than 15 minutes commuting time)



Panel B: Segregation of Income

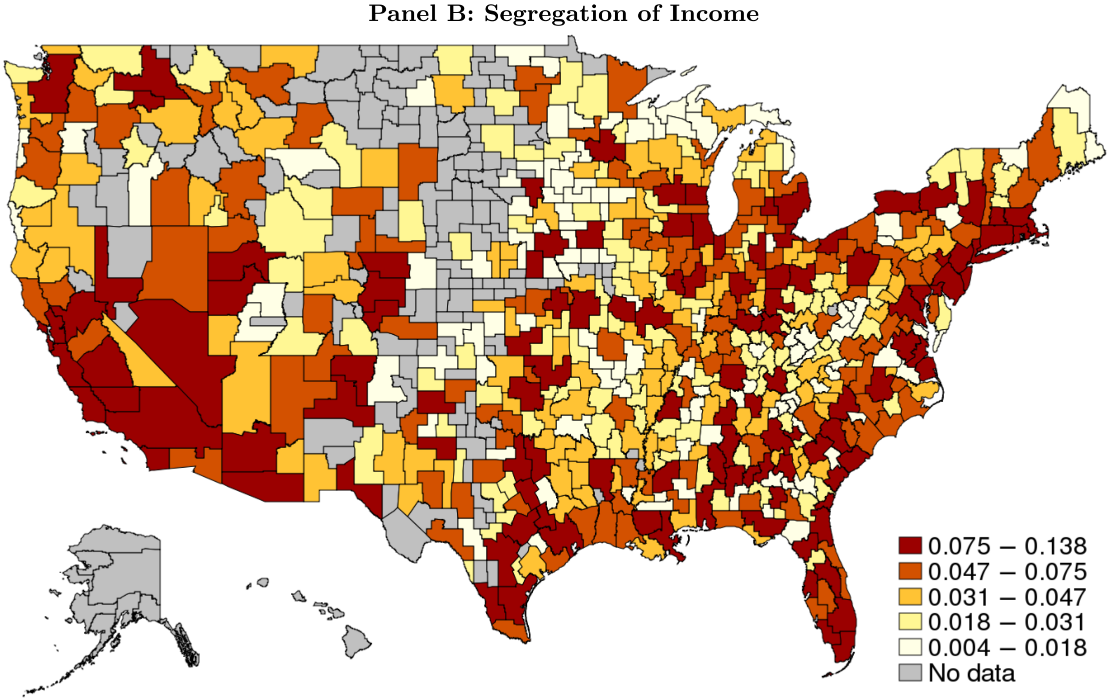
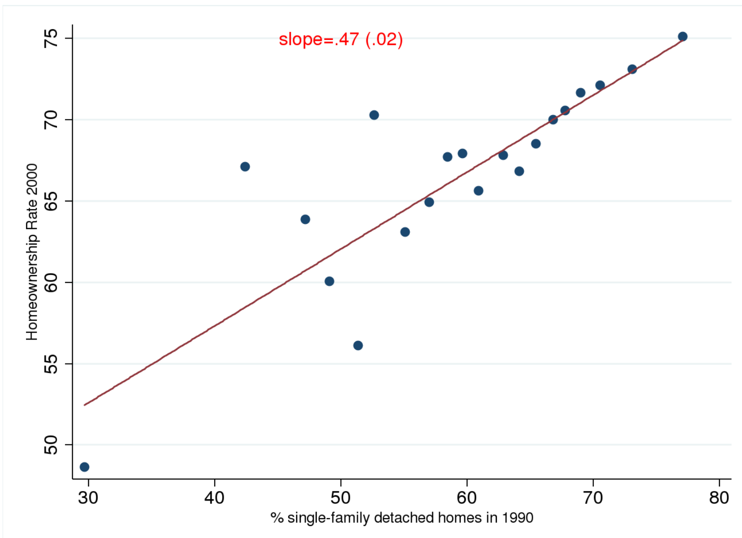
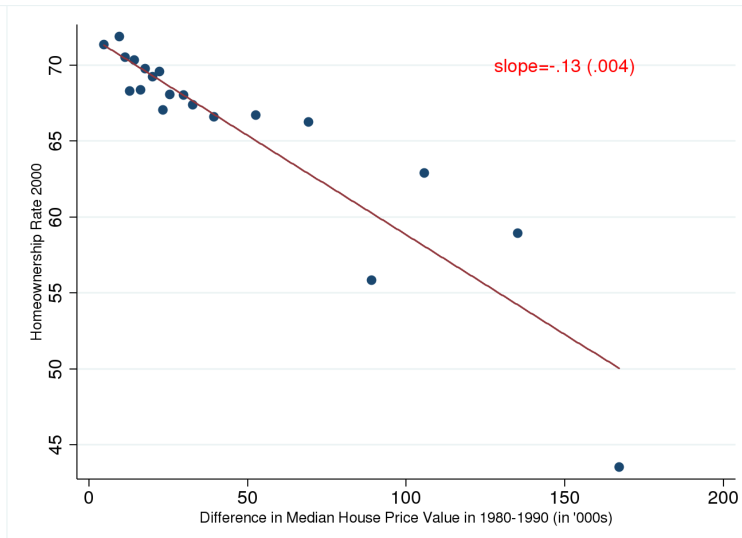


Figure 2.4: IV First Stage: Fraction of Single family detached homes 1990 and homeownership rates in 2000

The figures below graphically show the first stage of the two instruments we use — CZ-level fraction of single family detached homes in 1990 (IV1) and the difference in median house price value between 1980 to 1990 (IV2) — to instrument for homeownership rates in 2000. Panel (a) shows the binned scatter plots of fraction of single family detached homes in 1990 against homeownership rates in 2000 weighted by total number of housing units in 2000. The binned scatter plots divide the variable along the x-axis (single family detached homes in 1990) into 5 percentile bins and plot the mean of the x-axis and corresponding mean of the y-axis (homeownership rates in 2000) respectively. Panel (b) shows the first stage for our second instrument, the difference in median house price between 1980 and 1990. Homeownership rates, total number of housing units, fraction of single family detached homes and median house price are from the Decennial Census.



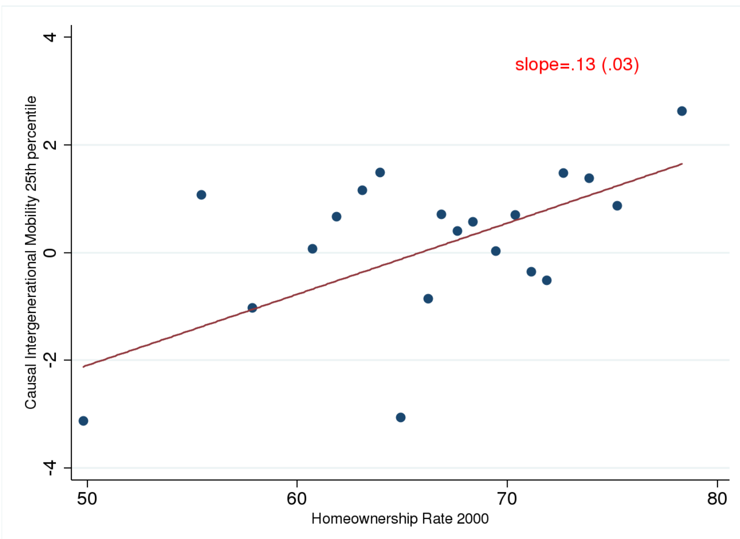
(a) **Single Family detached Homes 1990**



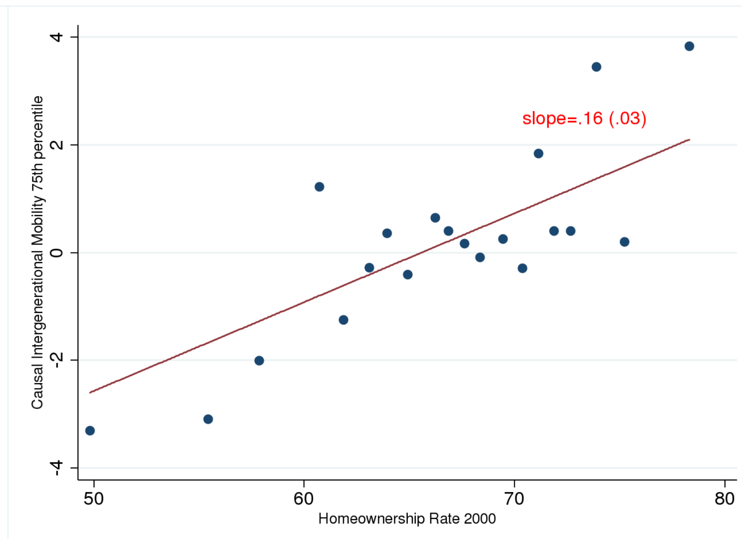
(b) **House Price Difference 1980–90**

Figure 2.5: Homeownership rates and Intergenerational Mobility

The figures below show the relationship between intergenerational mobility and homeownership rates using CZ level data. The vertical axis variable from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (panel(a)) and 75^{th} percentile (panel(b)) of the parents' income distribution. Homeownership rate data on the horizontal axis is from the US 2000 Census. Data are weighted by the number of housing units in each CZ in 2000.



(a) 25^{th} **percentile**

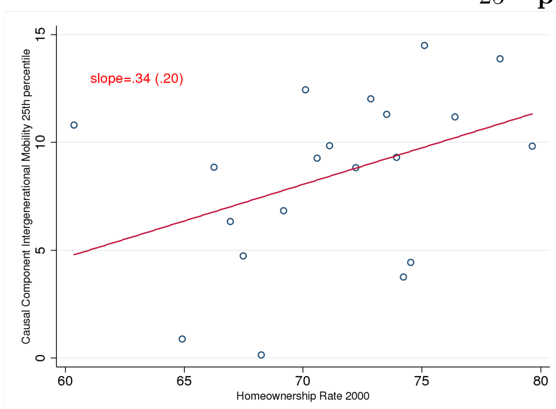


(b) 75^{th} **percentile**

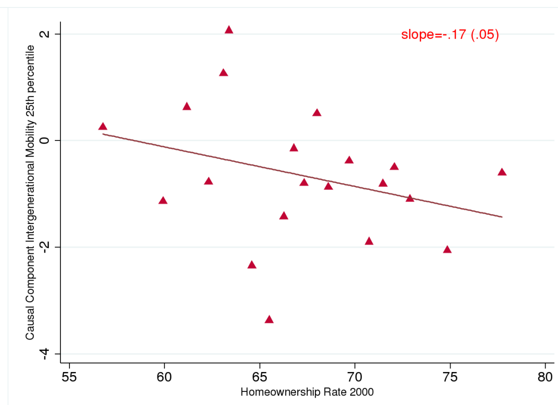
Figure 2.6: Heterogeneity across Sprawl

We present cross-sectional heterogeneity in the effect of homeownership on intergenerational mobility across sprawl at the CZ level. Sprawl is defined as the fraction of people not working from home with greater than 15 minutes of commute time to work and is from Census 2000. The sample is divided into terciles of sprawl. The high sprawl CZs refer to the top tercile (panel (b) and (d)) and the low sprawl CZs refer to the bottom tercile (panel (a) and (c)). The outliers corresponding to the top 1 percentile and bottom 1 percentile of homeownership rates have been dropped in this figure. The vertical axis variable from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (top panel) and 75^{th} percentile (bottom panel) of the parents' income distribution. Homeownership rate data on the horizontal axis is from the US 2000 Census. Data are weighted by the number of housing units in each CZ in 2000.

25^{th} percentile

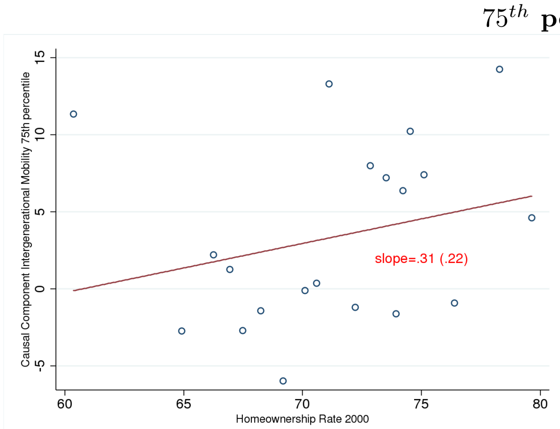


(a) **Low Sprawl**

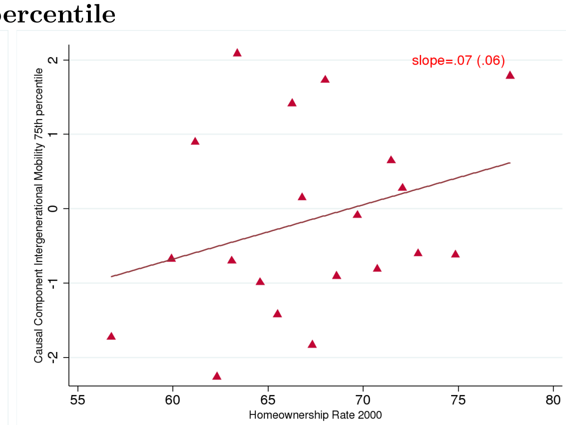


(b) **High Sprawl**

75^{th} percentile



(c) **Low Sprawl**

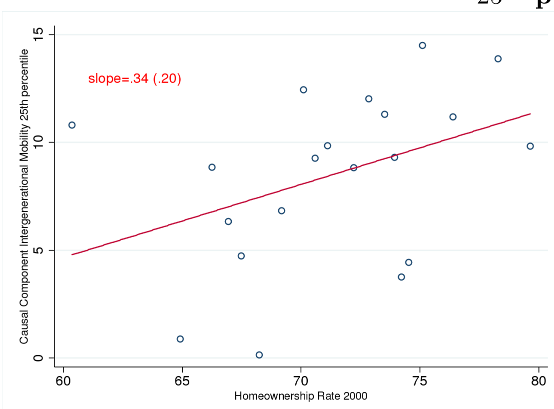


(d) **High Sprawl**

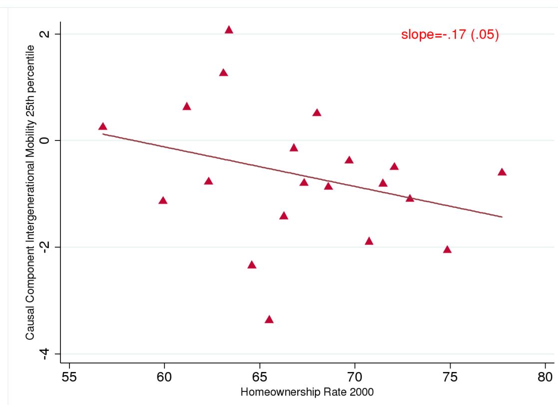
Figure 2.7: Heterogeneity across Segregation of homeowners

We present cross-sectional heterogeneity in the effect of homeownership on intergenerational mobility across segregation of homeowners at the CZ level. Segregation of homeowners is an entropy-based measure calculated at the CZ level using Census 2000 data. The sample is divided into terciles of segregation of homeowners. The high segregation CZs refer to the top tercile (panel (b) and (d)) and the low segregation CZs refer to the bottom tercile (panel (a) and (c)). The outliers corresponding to the top 1 percentile and bottom 1 percentile of homeownership rates have been dropped in this figure. The vertical axis variable from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (top panel) and 75^{th} percentile (bottom panel) of the parents' income distribution. Homeownership rate data on the horizontal axis is from the US 2000 Census. Data are weighted by the number of housing units in each CZ in 2000.

25^{th} percentile

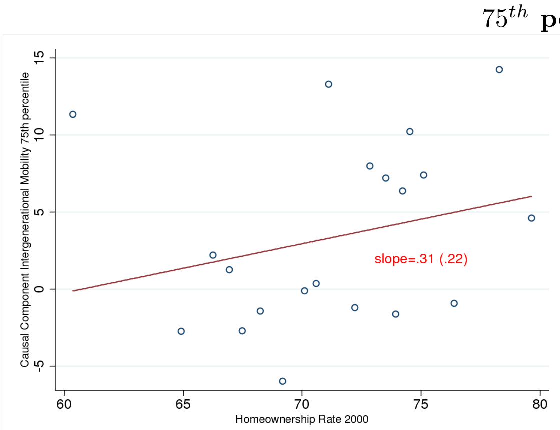


(a) **Low Segregation**

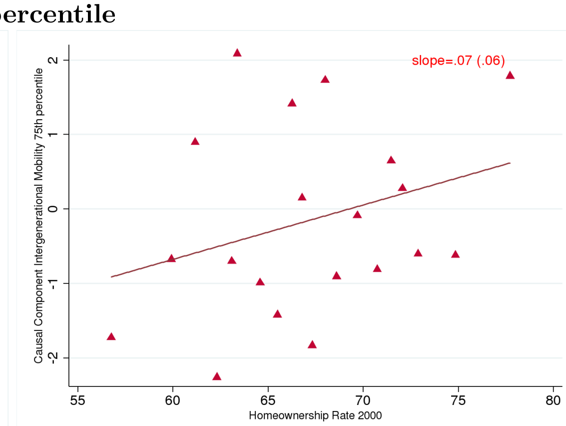


(b) **High Segregation**

75^{th} percentile



(c) **Low Sprawl**

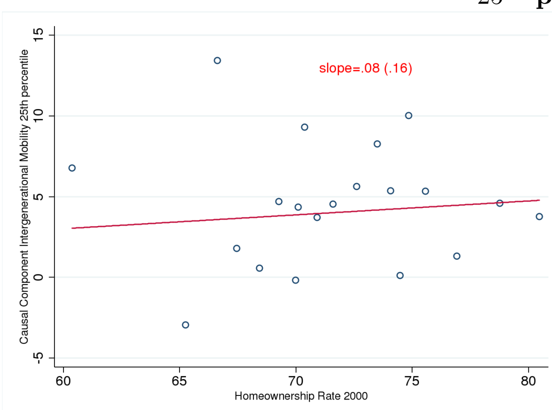


(d) **High Sprawl**

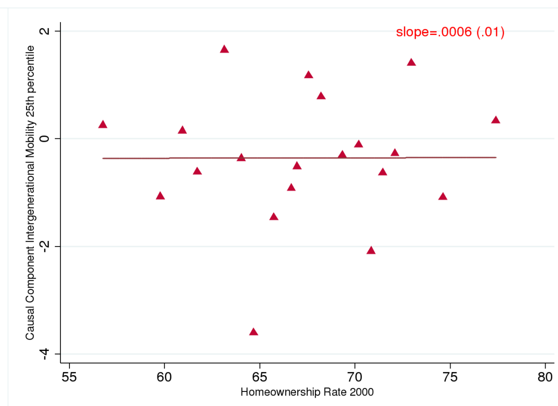
Figure 2.8: Heterogeneity across Racial Segregation

We present cross-sectional heterogeneity in the effect of homeownership on intergenerational mobility across racial segregation at the CZ level. Racial segregation is measured using [78] Index. The sample is divided into terciles of racial segregation index. The high segregation CZs refer to the top tercile (panel (b) and (d)) and the low segregation CZs refer to the bottom tercile (panel (a) and (c)). The outliers corresponding to the top 1 percentile and bottom 1 percentile of homeownership rates have been dropped in this figure. The vertical axis variable from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (top panel) and 75^{th} percentile (bottom panel) of the parents' income distribution. Homeownership rate data on the horizontal axis is from the US 2000 Census. Data are weighted by the number of housing units in each CZ in 2000.

25^{th} percentile

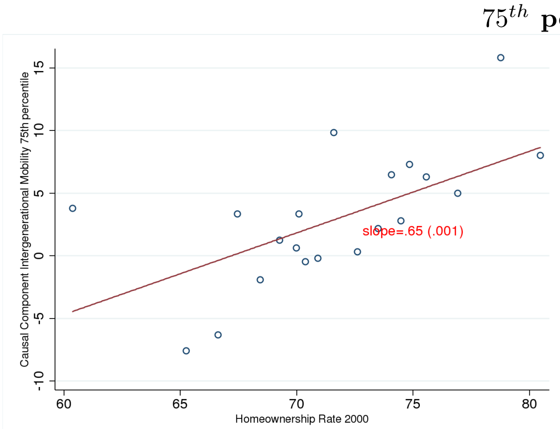


(a) **Low Segregation**

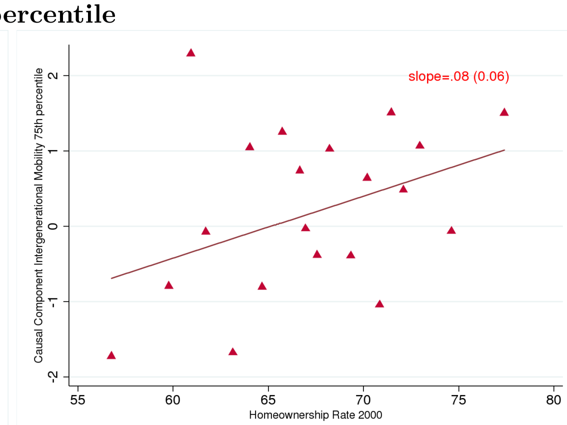


(b) **High Segregation**

75^{th} percentile



(c) **Low Segregation**

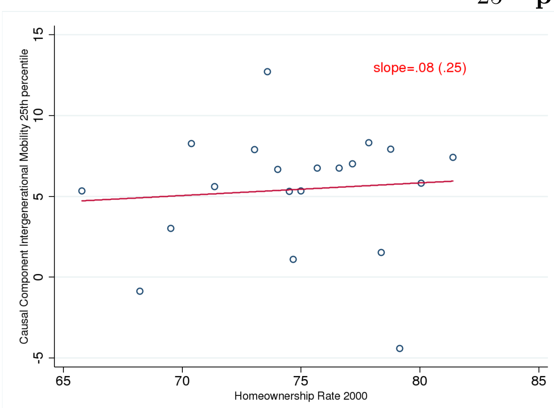


(d) **High Segregation**

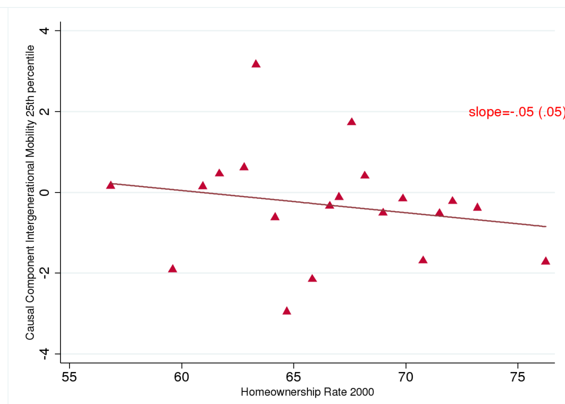
Figure 2.9: Heterogeneity across Segregation of Income

We present cross-sectional heterogeneity in the effect of homeownership on intergenerational mobility across income segregation at the CZ level. Income segregation is measured as in [73], calculated at the at the CZ level with Census 2000 data provided by [15]. The sample is divided into terciles of income segregation. The high income segregation CZs refer to the top tercile (panel (b) and (d)) and the low income segregation CZs refer to the bottom tercile (panel (a) and (c)). The outliers corresponding to the top 1 percentile and bottom 1 percentile of homeownership rates have been dropped in this figure. The vertical axis variable from [15] is the causal component of growing up in a neighborhood for 20 years on intergenerational mobility for the 25^{th} percentile (top panel) and 75^{th} percentile (bottom panel) of the parents' income distribution. Homeownership rate data on the horizontal axis is from the US 2000 Census. Data are weighted by the number of housing units in each CZ in 2000.

25^{th} percentile

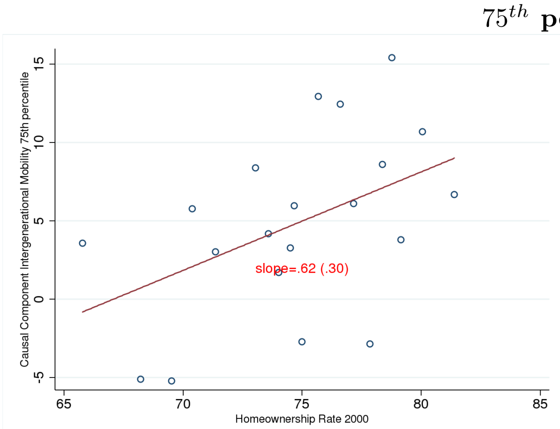


(a) **Low Segregation**

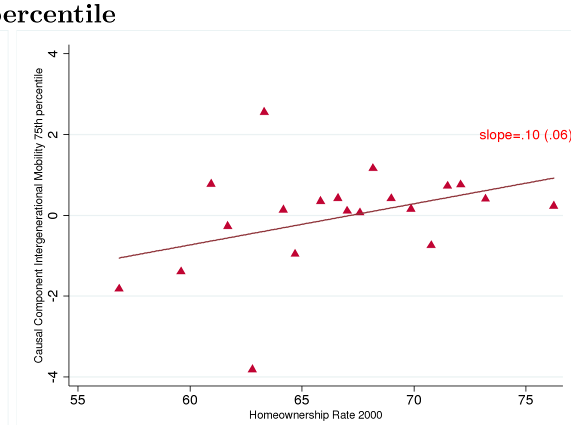


(b) **High Segregation**

75^{th} percentile



(c) **Low Segregation**



(d) **High Segregation**

Bibliography

[1] Viral V. Acharya et al. "Guaranteed to Fail: Fannie, Freddie, and the Debacle of Mortgage Finance". Princeton University Press. 2011.

[2] Manuel Adelino, Antoinette Schoar, and Felipe Severino. "Credit Supply and House Prices: Evidence from Mortgage Market Segmentation". Working Paper. 2014.

[3] Manuel Adelino, Antoinette Schoar, and Felipe Severino. "Loan Originations and Defaults in the Mortgage Crisis: Further Evidence". Working paper. 2015.

[4] Manuel Adelino, Antoinette Schoar, and Felipe Severino. "Loan Originations and Defaults in the Mortgage Crisis: The Role of the Middle Class". Working Paper. 2015.

[5] Sumit Agarwal et al. "Predatory lending and the subprime crisis". In: *Journal of Financial Economics* 113.1 (2014), pp. 29 –52. ISSN: 0304-405X. DOI: `http://dx.doi.org/10.1016/j.jfineco.2014.02.008`. URL: `http://www.sciencedirect.com/science/article/pii/S0304405X14000397`.

[6] Brent W. Ambrose et al. "The Effect of Conforming Loan Status on Mortgage Yield Spreads: A Loan Level Analysis". In: *Real Estate Economics* (2004), pp. 541–569.

[7] G. Amromin et al. "Complex mortgages". NBER Working paper. 2013.

[8] Joshua D. Angrist and Jörn-Steffen Pischke. "Mostly Harmless Econometrics: An Empiricist's Companion". In: *Princeton University Press* (2008).

[9] G. Barlevy and J. D. Fisher. "Mortgage choices and housing speculation". Technical report, Working Paper, Federal Reserve Bank of Chicago. 2010.

[10] Neil Bhutta. "GSE Activity and Mortgage Supply in Lower-Income and Minority Neighborhoods: The Effect of the Affordable Housing Goals". English. In: *The Journal of Real Estate Finance and Economics* 45.1 (2012), pp. 238–261. ISSN: 0895-5638. DOI: `10.1007/s11146-010-9258-z`. URL: `http://dx.doi.org/10.1007/s11146-010-9258-z`.

[11] Philip Bond, David K. Musto, and Bilge Yilmaz. "Predatory mortgage lending". In: *Journal of Financial Economics* 94.3 (2009), pp. 412 –427. ISSN: 0304-405X. DOI: `http://dx.doi.org/10.1016/j.jfineco.2008.09.011`. URL: `http://www.sciencedirect.com/science/article/pii/S0304405X09001469`.

[12] Raphael W. Bostic and Stuart A. Gabriel. “Do the {GSEs} matter to low-income housing markets? An assessment of the effects of the {GSE} loan purchase goals on California housing outcomes”. In: *Journal of Urban Economics* 59.3 (2006), pp. 458 –475. ISSN: 0094-1190. DOI: `http://dx.doi.org/10.1016/j.jue.2005.12.004`. URL: `http://www.sciencedirect.com/science/article/pii/S0094119005001051`.

[13] A. Colin Cameron and Pravin Trivedi. “Microeconometrics: Methods and Applications”. In: *New York: CambridgeUniversity Press* (2005).

[14] Micahel Carlos Best and Henrik Jacobsen Kleven. “Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK”. Working paper. 2010.

[15] Raj Chetty and Nathaniel Hendren. “The impacts of neighborhoods on Intergenerational Mobility: Childhood exposure effects and County-level estimates”. In: *Working paper* (2015).

[16] Raj Chetty et al. “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records”. In: *The Quarterly Journal of Economics* 126.2 (2011), pp. 749–804. DOI: `10.1093/qje/qjr013`. eprint: `http://qje.oxfordjournals.org/content/126/2/749.full.pdf+html`. URL: `http://qje.oxfordjournals.org/content/126/2/749.abstract`.

[17] Raj Chetty et al. “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”. In: *Quarterly Journal of Economics* (2015).

[18] A. Chinco and C. Mayer. “Misinformed Speculators and Mispricing in the Housing Market”. Technical report, National Bureau of Economic Research. 2014.

[19] Edward Coulson and Lynn M. Fisher. “Housing tenure and labor market impacts: The search goes on”. In: *Journal of Urban Economics* (2009).

[20] Edward Coulson and Herman Li. “Measuring the External Benefits of Homeownership”. In: *Paper presented at the 46th Annual AREUEA Conference, Denver, Colorado.* (2010).

[21] Thomas Davidoff. “Supply Constraints are Not Valid Instrumental Variables for Home Prices Because They Are Correlated with Many Demand Factors”. In: *Working Paper* (2014).

[22] Angus Deaton. “The Analysis of Household Surveys: A Microeconometric Approach to Development Policy”. In: *Baltimore, MD: Johns Hopkins University Press* (1997).

[23] Anthony DeFusco and Andrew Paciorek. “The interest rate elasticity of mortgage demand: Evidence from bunching at the conforming loan limit”. Working paper. 2014.

[24] Denise DiPasquale and Edward L. Glaeser. “Incentives and Social Capital: Are Homeowners Better Citizens?” In: *Journal of Urban Economics* (1999).

[25] Chris Downing, Dwight Jaffee, and Nancy Wallace. "Is the Market for Mortgage-Backed Securities a Market for Lemons?" In: *Review of Financial Studies* 22.7 (2009), pp. 2457–2494. DOI: 10.1093/rfs/hhn114. eprint: http://rfs.oxfordjournals.org/content/22/7/2457.full.pdf+html. URL: http://rfs.oxfordjournals.org/content/22/7/2457.abstract.

[26] JohnV. Duca and StuartS. Rosenthal. "Do mortgage rates vary based on household default characteristics? Evidence on rate sorting and credit rationing". English. In: *The Journal of Real Estate Finance and Economics* 8.2 (1994), pp. 99–113. ISSN: 0895-5638. DOI: 10.1007/BF01097032. URL: http://dx.doi.org/10.1007/BF01097032.

[27] Pascaline Dupas and Jonathan Robinson. "Why Don't the Poor Save More? Evidence from Health Savings Experiments". In: *American Economic Review* 103.4 (2013), pp. 1138–71. DOI: 10.1257/aer.103.4.1138. URL: http://www.aeaweb.org/articles.php?doi=10.1257/aer.103.4.1138.

[28] Fernando Ferreira, Joseph Gyourko, and Joseph Tracy. "Housing Busts and Household Mobility: An Update". In: *Working Paper, National Bureau of Economic Research* (2011).

[29] Andreas Fuster and James Vickery. "Securitization and the Fixed-Rate Mortgage". In: *Review of Financial Studies* (2014). DOI: 10.1093/rfs/hhu060. eprint: http://rfs.oxfordjournals.org/content/early/2014/11/04/rfs.hhu060.full.pdf+html. URL: http://rfs.oxfordjournals.org/content/early/2014/11/04/rfs.hhu060.abstract.

[30] Jie Gan and Timothy J. Riddiough. "Monopoly and Information Advantage in the Residential Mortgage Market". In: *Review of Financial Studies* 21.6 (2008), pp. 2677–2703. DOI: 10.1093/revfin/hhm005. eprint: http://rfs.oxfordjournals.org/content/21/6/2677.full.pdf+html. URL: http://rfs.oxfordjournals.org/content/21/6/2677.abstract.

[31] Jie M. Gan and Timothy Riddiough. "Piercing and then Papering over the Vail of Ignorance: GSE Monopoly and Informational Advantage in the Market for Residential Mortgages". Working Paper. 2011.

[32] Andrew Gelman and Guido Imbens. "Why High-order Polynomials Should not be Used in Regression Discontinuity Designs". NBER working paper. 2014.

[33] Kristopher Gerardi, Lauren Lambie-Hanson, and Paul S. Willen. "Do borrower rights improve borrower outcomes? Evidence from the foreclosure process". In: *Journal of Urban Economics* 73.1 (2013), pp. 1 –17. ISSN: 0094-1190. DOI: http://dx.doi.org/10.1016/j.jue.2012.06.003. URL: http://www.sciencedirect.com/science/article/pii/S0094119012000423.

[34] Andra C. Ghent and Marianna Kudlyak. "Recourse and Residential Mortgage Default: Evidence from US States". In: *Review of Financial Studies* 24.9 (2011), pp. 3139–3186. DOI: 10.1093/rfs/hhr055. eprint: http://rfs.oxfordjournals.org/content/24/9/3139.full.pdf+html. URL: http://rfs.oxfordjournals.org/content/24/9/3139.abstract.

[35] Edward Glaeser and Matthew E. Kahn. "Sprawl and Urban Growth". In: *Handbook of Regional and Urban Economics* (2004).

[36] Edward L. Glaeser. "Rethinking the Federal Bias Toward Homeownership". In: *Cityscape: A Journal of Policy Development and Research* (2011).

[37] Edward L. Glaeser and Hdi D. Kallal. "Thin Markets, Asymmetric Information, and Mortgage-Backed Securities". In: *Journal of Financial Intermediation* 6.1 (1997), pp. 64 –86. ISSN: 1042-9573. DOI: http://dx.doi.org/10.1006/jfin.1997.0209. URL: http://www.sciencedirect.com/science/article/pii/S1042957397902096.

[38] Edward L. Glaeser and Bruce Sacerdote. "The Social Consequences of Housing". In: *Journal of Housing Economics* (2000).

[39] Edward L. Glaeser and Jesse M. Shapiro. "The Benefits of the Home Mortgage Interest Deduction". In: *Working paper, Harvard* (2003).

[40] R. K. Green and M. White. "Measuring the Benefits of Homeowning: Effects on Children". In: *Journal of Urban Economics* (1997).

[41] LUIGI GUISO, PAOLA SAPIENZA, and LUIGI ZINGALES. "The Determinants of Attitudes toward Strategic Default on Mortgages". In: *The Journal of Finance* 68.4 (2013), pp. 1473–1515. ISSN: 1540-6261. DOI: 10.1111/jofi.12044. URL: http://dx.doi.org/10.1111/jofi.12044.

[42] A. Haughwout et al. "Real estate investors, the leverage cycle, and the housing market crisis". Staff Reports, Federal Reserve Bank of New York. 2011.

[43] Donald R. Haurin, Toby L. Parcel, and R. Jean Haurin. "The Impact of Homeownership on Child Outcomes". In: *Joint Cneter of Housing Studies of Harvard University Working Paper Series* (2001).

[44] Eric Hurst et al. "Regional Risk Sharing through the U.S. Mortgage Market". Working Paper. 2015.

[45] Stefano M. Iacus, Gary King, and Giuseppe Porro. "Causal Inference Without Balance Checking: Coarsened Exact Matching". In: *Political Analysis* (2011).

[46] Stefano M. Iacus, Gary King, and Giuseppe Porro. "Multivariate Matching Methods That are Monotonic Imbalance Bounding". In: *Journal of the American Statistical Association* 106 (2011), pp. 345–361.

[47] Guido Imbens and Karthik Kalyanaraman. “Optimal Bandwidth Choice for the Regression Discontinuity Estimator”. In: *The Review of Economic Studies* (2011). DOI: 10.1093/restud/rdr043. eprint: http://restud.oxfordjournals.org/content/early/2012/01/07/restud.rdr043.full.pdf+html. URL: http://restud.oxfordjournals.org/content/early/2012/01/07/restud.rdr043.abstract.

[48] Guido W. Imbens and Thomas Lemieux. “Regression discontinuity designs: A guide to practice”. In: *Journal of Econometrics* 142.2 (2008). The regression discontinuity design: Theory and applications, pp. 615 –635. ISSN: 0304-4076. DOI: http://dx.doi.org/10.1016/j.jeconom.2007.05.001. URL: http://www.sciencedirect.com/science/article/pii/S0304407607001091.

[49] Dwight Jaffee. “Reforming The U.S. Mortgage Market Through Private Market Incentives”. House of Cards: Reforming Americas Housing Finance System, George Mason University. 2012.

[50] Dwight Jaffee. “The Role of the GSEs and Housing Policy in the Financial Crisis”. Financial Crisis Inquiry Commission. 2010.

[51] Dwight Jaffee and John Quigley. “Housing Subsidies and Homeowners: What Role for Government-Sponsored Enterprises?” Brookings-Wharton Papers on Urban Economics. 2007.

[52] Dwight Jaffee and John M. Quigley. “The Future of the Government Sponsored Enterprises The Role for Government in the US Mortgage Market”. NBER Working Paper No. 17685. 2011.

[53] W. Jiang, A. Nelson, and E. Vytlacil. “Liar’s loan? Effects of origination channel and information falsification on mortgage delinquency”. In: *Review of Economics and Statistics* 96.18 (2013), pp. 0034–6535. DOI: http://dx.doi.org/10.1162/REST_a_00387.

[54] Andrew Weiss Joseph E. Stiglitz. “Credit Rationing in Markets with Imperfect Information”. In: *The American Economic Review* 71.3 (1981), pp. 393–410. ISSN: 00028282. URL: http://www.jstor.org/stable/1802787.

[55] A. Kaufman. “What do Fannie and Freddie do?” Unpublished Manuscript. 2012.

[56] Benjamin J. Keys et al. “Did Securitization Lead to Lax Screening? Evidence from Subprime Loans”. In: *The Quarterly Journal of Economics* 125.1 (2010), pp. 307–362. DOI: 10.1162/qjec.2010.125.1.307. eprint: http://qje.oxfordjournals.org/content/125/1/307.full.pdf+html. URL: http://qje.oxfordjournals.org/content/125/1/307.abstract.

[57] Henrik Kleven. “Bunching”. Annual Review of Economics. Forthcoming.

[58] Henrik J. Kleven and Mazhar Waseem. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan". In: *The Quarterly Journal of Economics* 128.2 (2013), pp. 669–723. DOI: 10.1093/qje/qjt004. eprint: http://qje.oxfordjournals.org/content/128/2/669.full.pdf+html. URL: http://qje.oxfordjournals.org/content/128/2/669.abstract.

[59] Alessandro Lizzeri. "Information Revelation and Certification Intermediaries". In: *The RAND Journal of Economics* 30.2 (1999), pp. 214–231. ISSN: 07416261. URL: http://www.jstor.org/stable/2556078.

[60] Elena Loutskina and Philip E. Strahan. "Securitization and the Declining Impact of Bank Finance on Loan Supply: Evidence from Mortgage Originations". In: *The Journal of Finance* 64.2 (2009), pp. 861–889. ISSN: 1540-6261. DOI: 10.1111/j.1540-6261.2009.01451.x. URL: http://dx.doi.org/10.1111/j.1540-6261.2009.01451.x.

[61] Marco Di Maggio, Amir Kermani, and Sanket Korgaonkar. "Deregulation, Competition and the Race to the Bottom". Working paper. 2010.

[62] JosephA. McKenzie. "A Reconsideration of the Jumbo/Non-Jumbo Mortgage Rate Differential". English. In: *The Journal of Real Estate Finance and Economics* 25.2-3 (2002), pp. 197–213. ISSN: 0895-5638. DOI: 10.1023/A:1016525725585. URL: http://dx.doi.org/10.1023/A%3A1016525725585.

[63] Atif Mian, Kamalesh Rao, and Amir Sufi. "Household Balance Sheets, Consumption, and the Economic Slump*". In: *The Quarterly Journal of Economics* (2013). DOI: 10.1093/qje/qjt020. eprint: http://qje.oxfordjournals.org/content/early/2013/07/25/qje.qjt020.full.pdf+html. URL: http://qje.oxfordjournals.org/content/early/2013/07/25/qje.qjt020.abstract.

[64] Atif Mian and Amir Sufi. "Fraudulent Income Overstatment on Mortgage Applications During the Credit Expansion of 2002 to 2005". Working Paper. 2015.

[65] Atif Mian and Amir Sufi. "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis". In: *The Quarterly Journal of Economics* 124.4 (2009), pp. 1449–1496. DOI: 10.1162/qjec.2009.124.4.1449. eprint: http://qje.oxfordjournals.org/content/124/4/1449.full.pdf+html. URL: http://qje.oxfordjournals.org/content/124/4/1449.abstract.

[66] Atif Mian, Amir Sufi, and Francesco Trebbi. "Foreclosures, House Prices, and the Real Economy". In: *The Journal of Finance* 70.6 (2015), pp. 2587–2634. ISSN: 1540-6261. DOI: 10.1111/jofi.12310. URL: http://dx.doi.org/10.1111/jofi.12310.

[67] Taylor D. Nadauld and Shane M. Sherlund. "The impact of securitization on the expansion of subprime credit". In: *Journal of Financial Economics* 107.2 (2013), pp. 454 –476. ISSN: 0304-405X. DOI: http://dx.doi.org/10.1016/j.jfineco.2012.09.002. URL: http://www.sciencedirect.com/science/article/pii/S0304405X12001870.

[68] Shawna Orzechowski and Peter Sepiella. "New Worth and Asset Ownership of Households: 1998 and 2000". In: *Current Population Reports, P70-88, May 2003. Washington, DC: U.S. Commerce Department, Bureau of the Census.* (2003).

[69] Karen M. Pence. "Foreclosing on Opportunity: State Laws and Mortgage Credit". In: *Review of Economics and Statistics* 88.1 (2006), pp. 177–182.

[70] Amiyatosh Purnanandam. "Originate-to-distribute Model and the Subprime Mortgage Crisis". In: *Review of Financial Studies* 24.6 (2011), pp. 1881–1915. DOI: `10.1093/rfs/hhq106`. eprint: `http://rfs.oxfordjournals.org/content/24/6/1881.full.pdf+html`. URL: `http://rfs.oxfordjournals.org/content/24/6/1881.abstract`.

[71] U. Rajan, A. Seru, and V. Vig. "The failure of models that predict failure: distance, incentives and defaults". Chicago GSB Research Paper. 2010.

[72] John Rao and Geoff Walsh. "Foreclosing a dream". National Consumer Law Center. 2009.

[73] Sean Reardon. "Measures of income segregation." In: *CEPA Working Papers. Stanford, CA: Stanford Center for Education Policy Analysis.* (2011).

[74] William M. Rohe and Michael A. Stegman. "The Impact of Home Ownership on the Social and Political Involvement of Low-Income People." In: *Urban Affairs Quarterly* (1994).

[75] Anil Rupasingha and Stephan Goetz. "US County-level Social Capital Data, 1990–2005". In: *Northeast Regional Center for Rural Development, Penn State University* (2008).

[76] Albert Saiz. "The Geographic Determinants of Housing Supply". In: *Quarterly Journal of Economcis* (2010).

[77] David Scharfstein and Adi Sunderam. "Market Power in Mortgage Lending and the Transmission of Monetary Policy". Working Paper. 2015.

[78] Henri Theil. "Statistical decomposition analysis. With applications in the social and administrative sciences number v. 14. In Studies in mathematical and managerial economics.". In: *Amsterdam, New York: North-Holland Pub. Co.; American Elsevier Pub. Co.* (1972).

[79] Jason Thomas and Robert Van Order. "Housing policy, Subprime Markets and Faniie Mae and Freddie Mac: What we know, What we think we know and What we don't know". Working Paper. 2011.

[80] James Vickery and Joshua Wright. "TBA Trading and Liquidity in the Agency MBS Market". New York Fed Working Paper. 2010.

[81] Danny Yagan. "The Home Mortgage Interest Deduction and Migratory Insurance over the Great Recession". In: *Working paper, University of California, Berkeley* (2013).

Appendix A

Appendix Tables

Table A1: Data Used

	Data Source	Main Variables Used	Frequency used	Geo-coding
1	RateWatch.com	Mortgage rates on rate sheets.	Monthly	Latitude & Longitude
2	GSE HUD Data	Minority, Income, single female, FTHB flag, Co-borrowers, Age, loan amount.	Annual	Census-tract
3	GSE 30 year FRM	Mortgage rates, LTV, DTI, FICO, FTHB, Co-borrowers flag, loan amount.	Quarterly	Zipcode 3-digit, State and MSA
4	ABSnet (Non-GSE)	Mortgage rates, LTV, DTI, FICO, loan amount.	Quarterly	Zip 3-digit, State and MSA
5	HMDA (Both)	Minority, Income, single female flag, co-borrowers flag, loan amount.	Annual	Census-tract
6	Decennial Census	% Hispanic, % black, log(median income), % High School and unemployment rate	1990, 2000	Census-tract

This table lists the data used, the main variables used in the analysis, the frequency of data and the most granular geo-coding available for the data. Items 1/2/3 for GSE data; 4 for Non-GSE and 5 with both GSE and non-GSE data. All results hold within each dataset. See Section 2.2 on how the data is cleaned.

A.1 Lender Rights Index

For the classification of state foreclosure laws I look at a number of different sources. For the classification of states as judicial or non-judicial states, I follow [33]. I use [34] for the classification of states as recourse or non-recourse. For the classification of states as fair-market-value and for equity right of redemption, I follow [72].

In judicial foreclosure states, the lender needs to go to court to foreclose on a property. This implies lower creditor rights compared to non-judicial foreclosure states where the lender does not need to go to court to start foreclosure proceedings. Judicial states are assigned a value of 1 and non-judicial states are assigned a value of 3. Table A2 shows the values used.

The right to redeem enables borrower the right to redeem or make whole the amount due even after the foreclosure process has started. The time allowed to redeem, however, varies from state to state ranging from a mere 10 days to a full year. The degree of right to redemption is from [72] who assign a value of 0, 0.5 or 1 depending on the degree of redemption. Thus, a greater right to redemption corresponds to greater borrower rights and analogously lower lender rights. Thus, a right to redeem of 1 in [72] is assigned a 1, 0.5 is assigned a 2 and 0 is assigned a 3 in my index.

Lastly, I also classify whether a state allows deficiency judgement (also called recourse states). The process for calculating the degree of recourse available to the lender is complicated by the fact that the amount of recourse available varies based on "fair-market-value" laws, and additionally some states have a very limited statute of limitations which makes recourse difficult. In non-recourse states, the lender only has access to the property securing the loan. In recourse states, lenders can also access borrower wages and personal property. I make one more distinction within recourse states which the previous literature has largely ignored. In fair market value states, the amount a lender can access is limited by the fair market value of the property. Thus, after foreclosure, the amount that can be pursued in case of a deficiency judgment is given by the shortfall between the debt and the fair value of the property. In non-fair market value states on the other hand, the amount that can be pursued in case of a deficiency judgment is given by the shortfall between the debt and the foreclosure sale price of the property. The foreclosure sale price of the property tends to be lower than the fair market value of the property. Thus, in non-fair market value states the lender can sell the property at a depressed foreclosure price and thus can claim a greater deficiency judgment. Another complication is that the statute of limitations varies from state to state. For states with the lowest statute of limitations, I down-weight the amount of recourse available to the lender. The degree of recourse is calculated as follows. I assign a value of 1 for non-recourse, a value of 2 for recourse and fair-market-value states and a value of 3 for recourse and non-fair market value states. Let us call this the degree of recourseness. Note that these correspond to increasing ability of the lender to access borrower's assets. For the 10 states with limited statute of limitations, I weight this degree of recourseness by half.

Thus from the above procedure, I assign a value between 1 to 3 for each mortgage law described above with higher values corresponding to higher lender rights. I then create a simple index adding up an indicator between 1 to 3 for each mortgage law between a hypothetical 3 to 9. This is called the "lender rights index". I standardize this and use this continuous measure in my analysis. See Appendix A2 for the lender rights index calculated.

Table A2: Lender Rights Index

State	Recourse Non-recourse	Judicial Non-judicial	Fair Market Value	Statute of limitations	Right to Redeem	Lender Rights Index (Std)
Alabama	R	NJ	0	0	0.5	-0.08
Alaska	NR	NJ	0	0	1	0.32
Arizona	NR	NJ	0	0	1	-0.08
Arkansas	R	NJ	0	1	1	0.73
California	NR	NJ	0	0	1	0.32
Colorado	R	NJ	1	0	1	0.73
Connecticut	R	J	1	0	0.5	-0.08
Delaware	R	J	0	1	1	-0.08
DC	R	NJ	0	1	1	1.14
Florida	R	NJ	0	1	0.5	-0.08
Georgia	R	J	1	0	1	0.73
Hawaii	R	J	0	0	1	-0.08
Idaho	R	NJ	1	1	1	0.73
Illinois	R	J	0	0	0.5	-1.30
Indiana	R	J	0	0	1	-0.08
Iowa	NR	J	0	0	0	-1.30
Kansas	R	J	1	0	0.5	-1.70
Kentucky	R	J	0	0	0.5	-0.08
Louisiana	R	J	1	0	1	-1.30
Maine	R	J	1	0	0.5	-1.70
Maryland	R	J	0	1	1	1.14
Massachusetts	R	J	0	0	1	1.14
Michigan	R	NJ	1	0	0.5	-0.48
Minnesota	NR	NJ	0	0	0	-0.90
Mississippi	R	NJ	0	1	1	1.14
Missouri	R	NJ	0	0	0.5	1.14
Montana	NR	NJ	0	0	1	0.32
Nebraska	R	J	1	0	1	0.73
Nevada	R	NJ	1	0	1	0.73
New Hampshire	R	NJ	0	1	1	1.14
New Jersey	R	J	1	0	0.5	-0.48

Table A3: Lender Rights Index

State	Recourse Non-recourse	Judicial Non-judidical	Fair Market Value	Statute of limitations	Right to Redeem	Lender Rights Index (Std)
New Mexico	R	J	0	0	0	-0.08
New York	R	J	1	0	1	-0.48
North Carolina	NR	NJ	1	1	0.5	0.73
North Dakota	NR	J	0	0	0.5	-2.51
Ohio	R	J	0	0	0.5	-0.08
Oklahoma	R	NJ	1	0	0.5	-0.48
Oregon	NR	NJ	0	0	1	-0.08
Pennsylvania	R	J	1	0	1	-1.70
Rhode island	R	NJ	0	0	1	1.14
South Carolina	R	J	1	0	1	-0.48
South Dakota	R	J	1	0	0	0.73
Tennessee	R	NJ	0	0	1	1.14
Texas	R	NJ	1	1	1	0.73
Utah	R	NJ	1	0	1	0.73
Vermont	R	J	1	0	1	-1.70
Virginia	R	NJ	0	0	1	1.14
Washington	NR	NJ	0	0	1	0.32
West Virginia	R	J	0	0	1	1.14
Wisconsin	NR	NJ	1	0	0.5	-2.51
Wyoming	R	NJ	0	0	0.5	-0.08

A.2 Conceptual Framework Incorporating lender Pricing Regimes

In Section 1.3 we derived mortgage demand which depends on mortgage rates set by the lender and we also explicitly looked at bunching around the CLL. In this Appendix, I explicitly show the implication for credit rationing given lending pricing policies. To analyze the structure of underwriting and pricing policies observed in US mortgage markets, this section presents a stylized model of credit rationing when there is regional variation in lender payoffs or more specifically in LR. I begin by considering a monopoly[1] lender (GSE) under three pricing regimes for mortgage rates. As previously mentioned, although the GSEs do not originate mortgages they directly influence the observed mortgage rates and originations by retail originators through their operations in the secondary mortgage markets. The GSEs evaluate whether a mortgage is GSE-conforming (a pass/fail decision in the credit evaluation model) and effectively the GSEs lend directly to consumers. This allows us to abstract away from the complications of the intermediate retail market.

In the first part of the model, I look at three pricing regimes: a risk-based pricing regime with screening, pooled pricing regime with no screening, and pooled pricing regime with screening. Given a particular pricing regime and household demand, I then show the impact on the credit rationing of borrowers.[2] The aim in analyzing these different pricing regimes is twofold. First, the risk-based pricing regime with screening provides a useful benchmark to evaluate the efficiency of the two pooled pricing regimes that we consider. The pooled pricing regimes help analyze implications for both borrower sorting and credit rationing by a monopoly lender. The second goal of these pricing frameworks is to help me compare my hypothesis with the pricing regimes implicitly assumed in prior empirical literature. Using the simple model sketches under the three pricing regimes, I derive empirically testable hypotheses and implications for credit rationing.

In the second part of the model, I motivate the bunching analysis. This has already been shown in Section 1.3.

Model Setup

Here I repeat the assumptions used in our model. The residential mortgage market structure is as follows. Retail lenders originate loans for the GSEs. Retail loan originators apply underwriting criteria imposed by the GSEs in the secondary market and charge a loan rate determined by the secondary market. As mentioned previously, for the rest of the exposition I abstract away from this additional layer between retail originator and GSEs.

[1]The assumptions on cost of funds ensures the GSEs are able to dominate non-GSEs in both the risk-based and pooling strategies. Hence, I use the term "monopoly" to refer to the GSEs.

[2]The basic structure of the pricing regimes draws on [31] and [30] who use a similar framework to study the uniform mortgage rate policy of GSEs. [31] and [30] hypothesize that the GSEs use uniform pricing to deter competitive entry.

The borrower has credit quality summarized by θ with $\theta \in [0,1]$. $1-\theta$ is the probability that the lender receives zero payoff. θ can be thought of as a reduced form value capturing payoff in case of borrower default. Alternatively, θ can be thought of as the output from a credit evaluation model which depends on borrower quality, macroeconomic conditions, state foreclosure laws and any other variables that may affect credit quality. θ can vary along many dimensions. For the purposes of my empirical analysis, θ also encapsulates differences in state foreclosure laws. I explicitly represent θ as a function of LR (H for high LR and L for low LR, $H > L$) and all other remaining variables are denoted by X. Lenders have higher recovery in states with high LR compared to low LR and thus $\theta(X, H)$ should be higher than $\theta(X, L)$. For ease of exposition, I explicitly assume $\theta(X, LR) = \phi(X) * LR$. Thus, $\theta(X, LR)$ can be decomposed into a borrower specific component combined with other information such as macroeconomic variables $\phi(X)$ and a regional adjustment due to differences in LR (L/H). Consider two borrowers with the same credit quality along all dimensions ($\phi(X)$) except that they reside in two states with different state foreclosure laws. Holding all else equal, the overall credit quality (θ) of the borrower in the state with high LR will have greater credit quality than that of the borrower in a state with low LR given $H > L$. Without loss of generality, let the total size of the population be normalized to 1.

Lenders are risk neutral. Loan production has a marginal cost of capital (κ^C).[3] In the pricing regimes with screening, there is an additional fixed cost of loan screening (κ^S). Loan screening in the pricing regimes with screening occurs as follows. The monopoly lender gets the borrower characteristics data and then combines it with other information — such as macroeconomic variables and state foreclosure laws — as an input into a proprietary credit evaluation model to determine credit quality of the borrower, θ. Given θ which is determined from this screening process, either the loan is denied in which case the applicant (potential borrower) is completely rationed out of the credit market. Alternatively, the loan is approved and credit is offered and loan size is determined by the demand schedule, $D(r)$ derived from the household optimization problem (described below).

I also analyze the spillover effects of the GSE credit rationing on the non-GSEs. For this, I make three additional assumptions. First, I assume the competitor lender has a cost of capital disadvantage.[4] The competitor lender competes directly with the monopoly lender but faces higher marginal costs of capital ($\kappa_{comp} > \kappa_C/\overline{\theta}$) where $\overline{\theta}$ is defined as the average credit quality distribution, $\overline{\theta} = \int_0^1 \theta f(\theta) d\theta$.[5] As we will see later, this cost of funds assumption also implies that the GSEs dominate the non-GSEs in the risk-based pricing

[3]Note, I make this simplifying assumption to abstract from economies of scale in this simple model.

[4]The monopoly lender enjoys lower costs of capital possibly though the uniformity or homogeneity of loans and subsequent liquidity in the TBA forward markets. Specifically, I postulated in Section 1.2 that uniform prices provide liquidity in the secondary mortgage market for the GSEs and this is incorporated into the lower marginal costs of capital for the monopoly lender (GSEs).

[5]As we will see later, this assumption ensures that the highest rate that can be charged to all borrowers in the conforming mortgage market is higher than the lowest risk-based price that can be charged by the competitor lender. This is consistent with the large jumbo conforming spreads that we empirically observe across all quality of borrowers.

regime and the pooled pricing regimes (both screening and no screening). Hence, I use the term "monopoly" lender to refer to the GSEs.

Second, I assume that the competitor is not able to precisely determine the credit quality of the borrower. The competitor's estimate of the credit quality of the borrower is based on both its own screening and the assumption of heavy sorting (cream-skimming) by the monopoly lender. To formalize this, I make the assumption that credit quality of the borrower depends on both $\theta_{comp} = \theta'/\lambda(Q)$ where θ' is the output from the screening technology of the competitor. Q is the total lending by the monopoly lender which the competitor can observe.[6] This is reduced by a factor $\lambda(Q)$ which depends on cream-skimming by the monopoly lender (Q). Without loss of generality let us normalize λ to 1. If lending by the monopoly lender is high, the competitor knows it is facing a heavily sorted market of poorer credit quality. Higher lending by the monopoly lender implies higher λ and lower θ_{comp}. Thus, when the presence of the monopoly lender is higher, the competitor lender perceives the credit quality of the borrower (θ_{comp}) to be lower.

Third, I assume that regulatory constraints (such as Fair Lending Laws) prohibit competitor lenders from charging mortgage rates exceeding an upper bound threshold $\overline{R}$.

Borrowers are price takers. Borrowers/households are credit constrained and borrow to finance the purchase of their homes. Households purchase quality-adjusted housing h at price p per unit. To buy a house of value ph, they borrow at an interest rate r (with $1 + r = R$).[7] After making housing choices, households receive an income y, repay all their debts and consume their remaining income. The household interest rate (r) depends on its own credit quality (θ). The household cannot determine its own credit quality which depends on many factors including external macroeconomic conditions and state foreclosure laws. This is not an unreasonable assumption in the mortgage market. Borrowers do not play a repeated dynamic game and hence have limited ability in inferring their own credit risk. Thus, credit risk affects borrower demand only in the way it affects the mortgage interest rate. That is, I explicitly make the assumption that the borrower's own credit quality does not enter the household optimization problem except in the way it affects the mortgage interest rate r. [11] make a similar assumption on borrower inability to determine their own future payoffs, that is, borrowers do not know their own credit quality at the time of entering into a loan.

One of the implications of this assumption is that borrowers do not ex-ante take into account their ability to strategically default more in states with low LR where the lender might not pursue foreclosure given the high costs to the lender. The empirical literature supports this assumption. [41] find that ex-ante borrowers do not know whether a lender can pursue recourse.[8] [66] find that default rates during the crisis were the same for judicial and

[6]The competitor also observes the pooled mortgage rate. However, since mortgage rates are pooled (as we show empirically later) the competitor is not able to reverse-engineer the perceived credit quality from the mortgage rates of the monopoly lender.

[7]In reality borrowers need to make a down payment financed from their liquid assets. However, this only adds a constant factor to the analysis and I abstract away from this for now.

[8][41] survey borrowers on whether they think their bank can come after the borrowers assets (in addition to the property securing the loan). Around 50% of borrowers said yes irrespective of the state they stayed

non-judicial states implying that borrowers do not take into account the longer foreclosure timelines in judicial states and hence do not strategically default more in judicial states. Additionally, [34] finds that delinquencies are not higher in non-recourse states compared to recourse states and differ only ex-post on whether the mortgage is underwater (negative equity).

The household maximization problem in Section 1.3 determines the mortgage demand which is then given by

$$D^* = \left(\frac{p}{A}\right)^{\epsilon+1} (R-1)^{\epsilon}. \tag{A.1}$$

We have assumed that p, y and ϵ are constant across households. For a given A, there is a one-to-one mapping between mortgage rates (R) and optimal mortgage choice (D^*). To make the exposition clearer, we denote mortgage demand across households as $D(R)$. Note, the elasticity of mortgage demand $\eta = -dD/dRR/D > 1$ given that $\epsilon < 0$.

Credit rationing and the lenders

Now I turn to the monopoly lender decision to lend. For my stylized model sketch, I will consider only pure pooling and pure separating strategies.

The three possibilities considered are:

- Risk-based loan pricing with applicant screening.
- Uniform loan pricing without applicant screening.
- Uniform loan pricing with applicant screening.

In pure separating strategies, loan rate (R) will vary with credit quality. In uniform loan pricing policies, loan rates do not vary with credit quality. Additionally, in the screening regimes the lender incurs a fixed sunk cost of screening (κ_C).

Screening Risk-based (S-RB) loan pricing regime

The lender maximizes expected profits with respect to the loan rate, R, for each borrower with credit quality θ:

$$\max_R \pi_{S-RB}(\theta) = \max_R [\theta R D(R) - \kappa^C D(R) - \kappa^s] \tag{A.2}$$

where $\theta R D(R)$ is the expected revenue as a function of credit quality θ and loan size $D(R)$. $\kappa^C D(R)$ is the variable cost of capital as a function of loan size. The lender incurs κ^s , the fixed cost of loan screening in order to determine the exact credit quality θ of the borrower. The profits for each individual borrower is given by $\pi_{S-RB}(\theta)$ in the risk-based pricing regime. I will use π to represent individual profits and Π to represent total profits. The subscript $S - RB$ refers to screening risk-based pricing policy.

---

in.

Solving the above gives the optimal mortgage rate for a given θ to be:

$$R^*_{S-RB}(\theta) = \frac{\kappa^C}{\theta}\frac{\eta}{\eta - 1} \tag{A.3}$$

where $\eta = -(\frac{dD}{dR})(\frac{R}{D})$ is the price elasticity of mortgage demand and is greater than 1.[9] Mortgage rate is simply marginal costs adjusted by the price elasticity of demand representing the monopoly lender's markup. Importantly, in the risk-based pricing regime the loan rate varies with θ. A borrower with a higher credit risk thus has a higher mortgage rate. Then, under the risk-based pricing regime for borrowers who are the same on all characteristics, except reside in states with different foreclosure laws, we have $\theta(H, X) > \theta(L, X)$ which implies interest rates for the borrower in the low LR (denoted by R^*_L) is higher than interest rates for the borrower in high LR (denoted by R^*_H). Note that screening here implies that the lender knows the credit quality of the borrower and hence is able to vary mortgage rates based on the credit quality θ.

Profits for each borrower of credit quality θ in the screening risk-based loan pricing is given by

$$\pi^*_{S-RB}(\theta) = \frac{\kappa^C D(R^*_{S-RB}(\theta))}{\eta - 1} - \kappa_s. \tag{A.4}$$

The lender payoffs are represented in Figure A1 assuming low costs of screening. Panel (a) shows the lender payoffs in the risk-based pricing regime with screening. The blue line shows the payoffs for each θ. We see that as θ increases, individual profits are increasing. One, profits increase due to higher lender payoffs (θ) which would be linear. Second, higher θ implies lower mortgage rates in the risk-based pricing regime and subsequently higher mortgage demand due to lower mortgage rates. Thus, profits are increasing at higher θ at an increasing rate because of the direct impact of higher credit quality on lender payoffs *and* the indirect effect through the higher demand of high θ borrowers in the risk-based pricing regime.

Over the entire distribution of credit quality (normalized to 1), profits in the screening risk-based loan pricing is given by

$$\Pi^*_{S-RB} = \int_0^1 \frac{\kappa^C D(R^*_{S-RB}(\theta))}{\eta - 1} f(\theta) d\theta - \kappa_s. \tag{A.5}$$

In the screening risk-based loan pricing framework, everyone who applies for a loan gets a loan even if individual profits are less than zero since the initial fixed costs, κ_s, are required to determine credit quality across all borrowers. Note that the goal in laying out the risk-based pricing framework is to explicitly compare efficiency of the pooled pricing regimes.[10] If the

[9] Elasticity of mortgage demand $\eta = -dD/dRR/D > 1$ given $\epsilon < 0$.

[10] In reality regulatory constraints (such as Fair Lending Laws) prohibit lenders from charging mortgage rates exceeding an upper bound threshold $\overline{R}$. We explicitly incorporate this assumption for the competitor lender and see that there is also an upper bound beyond which there is no lending.

expected profit (Π^*_{S-RB}) from above is greater than 0, then the lender invests in a screening technology and lends to everyone who applies for a mortgage at the risk-based price.[11] The red line in Panel (a) in Figure A1 shows the total profits across all the entire credit quality distribution given by Π^*_{S-RB}. This will be a useful benchmark to compare other pricing regimes with. Next, we consider the no screening pricing regime.

No Screening Pooled (NS-P) loan pricing regime

Under a no screening regime, the lender does not know the credit quality (θ) of the borrower. Thus, the lender charges a fixed pooled mortgage rate based on the average credit quality distribution, $\overline{\theta} = \int_0^1 \theta f(\theta) d\theta$.

The optimal mortgage rate is then given by:

$$R^*_{NS-P} = \frac{\kappa^C}{\overline{\theta}} \frac{\eta}{\eta - 1}. \tag{A.6}$$

There are no sunk costs of screening, κ_s, since the monopoly lender does not need to know the exact credit quality of the borrower and hence has not invested in the screening technology.

Profits for each borrower of credit quality θ in the screening risk-based loan pricing is given by

$$\pi^{NS-P}(\theta) = \theta R^*_{NSP} D(R^*_{NSP}) - \kappa^C D(R^*_{NSP}). \tag{A.7}$$

In the no-screening pooled pricing regime, high (low) quality borrowers pay higher (lower) mortgage rates than they would otherwise have paid in the screening risk-based pricing regime. Low quality borrowers are cross-subsidized by the high quality borrowers. Additionally, total loans demanded by the low quality borrowers is higher than in the risk-based pricing regime. To see this, note that demand only depends on the pooled mortgage rate charged (R^*_{NSP}). Since mortgage rates in the pooled pricing regimes are lower than the risk-based pricing regimes for the low quality borrowers, demand is higher. The higher demand by the low quality borrowers is reflected in lender profits. Lender payoffs are shown in Panel (b), Figure A1. The blue line shows the payoffs for each θ. The lenders (inefficiently) make mortgages at the pooled mortgage rates to very low quality borrowers even when lenders make negative profits on these low quality borrowers. As θ increases, individual profits are increasing since profits increase due to the direct impact on higher lender payoffs. However, the profits at high credit quality are not increasing at an increasing rate due to the absence on the indirect effect through demand (present in risk-based screened pricing regime previously discussed). Thus, total profits across all borrowers are much lower than the benchmark risk-based pricing as shown below.

[11] Of course, in a more realistic scenario a cap on mortgage rates due to Fair Lending laws will imply that the lowest credit quality borrowers may still get rationed out.

Profits across the full population (which has been normalized to 1) in the no screening pooled loan pricing is given by

$$\Pi^{NS-P} = \frac{\kappa^C D(R^*_{NSP})}{\eta - 1}. \tag{A.8}$$

In the no screening pooled loan pricing framework, everyone who applies for a loan gets a loan at the pooled rate since the lender does not know the exact credit quality of each borrower. The risk-based pricing regime dominates the no screening pooled pricing regime if screening costs are low.

Screening Pooled (S-P) loan pricing Regime

In the screening pooled (S-P) pricing regime the lender first screens borrowers and then determines both the minimum credit quality ($\underline{\theta}$) above which he is willing to lend and the pooled mortgage rate for all loans that cross this minimum threshold. Thus, in this pricing regime the lender simultaneously solves for the uniform (pooled) mortgage rate and the minimum credit quality boundary that maximizes his total expected profits. The maximization problem is given by:

$$\max_{\underline{\theta},R} \Pi^{S-P} = \max_{\underline{\theta},R}[\int_{\underline{\theta}}^{1} (\theta RD - \kappa^C D) f(\theta) d\theta - \kappa_S, 0]. \tag{A.9}$$

All consumers need to be screened regardless of their quality and κ_S is the sunk cost invested in the screening technology.

Solving the above gives two F.O.Cs in $\underline{\theta}$ and R. In order to determine the optimal points, the lender simultaneously solves for the optimal pooled mortgage rate and the optimal cut-off point. The two F.O.Cs give:

$$R^*_{S-P} = \frac{\kappa^C}{\underline{\theta}^*} \tag{A.10}$$

$$R^*_{S-P} = \frac{\kappa^C}{\overline{\overline{\theta}}^*} \frac{\eta}{\eta - 1} \tag{A.11}$$

where $\overline{\overline{\theta}}^* = \frac{\int_{\underline{\theta}^*}^{1} \theta f(\theta) d\theta}{1 - F(\underline{\theta}^*)}$ is the average quality of people who qualify for the loan. The F.O.C.s imply that the lender solves simultaneously for the minimum credit quality and the optimal pooled mortgage rate. As we saw before, for two borrowers who are the same on all credit dimensions (all X) except that they reside in states with different state foreclosure laws, the credit risk of the borrower in the higher LR area is lower than credit risk of the borrower in the low LR area ($\theta(H, X) > \theta(L, X)$). The overall minimum credit quality threshold ($\underline{\theta}^*$)

above which the lender is wiling to lend is the same in high LR and low LR. However, given $H > L$, the minimum quality threshold for borrower characteristics in high LR ($\phi^*(X_H)$) is lower than the minimum credit quality threshold in low LR ($\phi^*(X_L)$) since $H > L$.

Borrowers with a credit quality (θ) above the minimum credit quality threshold $\underline{\theta}^*$ get a loan. The profits for each of these borrowers is given by:

$$\pi^{S-P}(\theta) = \theta R^*_{SP} D(R^*_{SP}) - \kappa^C D(R^*_{SP}) \tag{A.12}$$

In the screening pooled pricing regime, high (low) quality borrowers still pay higher (lower) mortgage rates than they would otherwise have paid in the screening risk-based pricing regime. However, since the lowest quality borrowers are rationed out, it is only the relatively high credit quality borrowers who are cross-subsidized by the high quality borrowers compared to the no-screening pooled regime. Additionally, since the average credit quality of the borrowers who do get loans is higher compared to the no-screening pooled regime, the pooled mortgage rate is also lower reflecting the higher average credit quality of borrowers. This credit rationing has the direct impact on payoffs as well as an indirect effect on the increased demand at the lower mortgage rates.

Lender payoffs are shown in Panel (c), Figure A1. The blue line shows the payoffs for each θ. Lenders only lend mortgages to borrowers with positive profits (in contrast to the no screening regime). As θ increases, individual profits are increasing since profits increase due to the direct impact on higher lender payoffs. Additionally, (assuming low screening costs) compared to the no screening regime, mortgage rates are lower (reflecting higher average quality of borrowers) and demand is also higher.

Compared to the risk-based pricing regime, however, the profits at high credit quality are not increasing at an increasing rate due to the absence of the indirect effect through demand, since mortgage rates do not change for the highest quality borrowers. Thus, total profits across the entire region are lower than the risk-based pricing regime but higher than the no screening pooled pricing regime.[12]

Additionally, total loans demanded by low quality borrowers are higher than in the risk-based pricing regime. To see this, note that demand only depends on the pooled mortgage rate charged (R^*_{NSP}). Since mortgage rates in the pooled pricing regimes are lower than the risk-based pricing regimes for the low quality borrowers, demand is higher. The higher demand by the low quality borrowers is reflected in lender profits.

Total profits over the entire population for the lender in this pricing regime is given by:

$$\Pi^{S-P} = \left[\kappa^C D(R^*_{SP})(1 - F(\underline{\theta}^*)) \left(\frac{\overline{\overline{\theta}}^*}{\underline{\theta}^*} - 1\right) - \kappa^s\right]. \tag{A.13}$$

This summarizes our discussion above. The first term, $\kappa^C D(R^*_{S-P})(1 - F(\underline{\theta}^*))$ is the average expected net revenue of all borrowers who pass the minimum quality threshold.

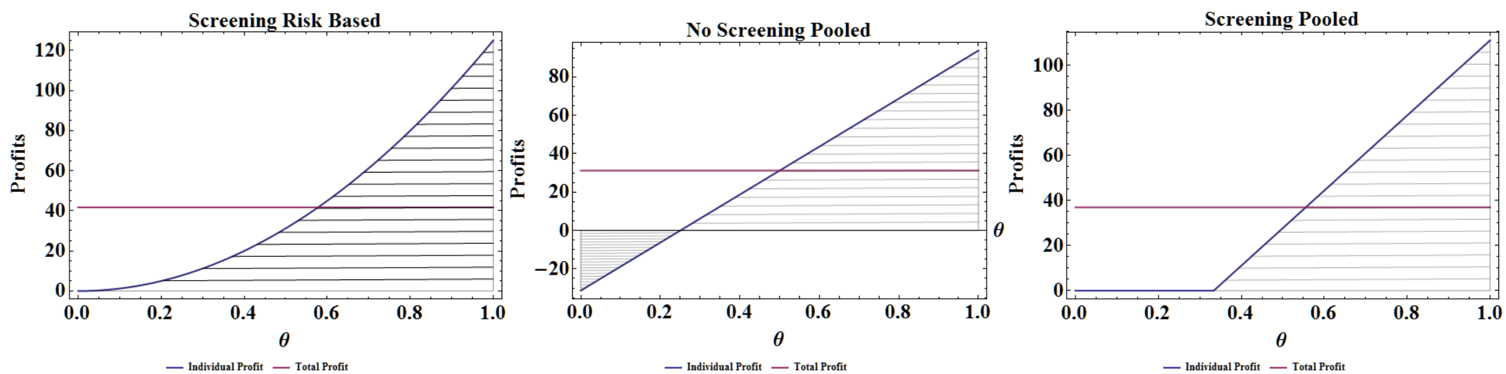
[12]Another way to see that within the pooled regimes the screening regime dominates the no-screening regime is directly through the maximization problem where (assuming low screening costs) no screening regimes is included in the optimization problem.

This is then adjusted by $\left(\frac{\bar{\bar{\theta}}^*}{\underline{\theta}^*}\right) > 1$ which is the ratio of the average credit quality to the marginal loan (that is, at the boundary, $\underline{\theta}^*$). This is the monopoly lender markup coming from screening out the borrowers with negative profits.

Assuming extremely low costs of screening (κ_s), the screening with pooling regime should produce higher profits. This is because the optimization also includes lending to all borrowers with minimum credit quality threshold equal to zero and then charging a pooled mortgage rate to all borrowers. Thus, profits in the no screening pricing regime for very low screening costs should be lower than the screening and pooled regime. The blue line in Panel (c), Figure A1 shows the total lender profits in this regime. Total profits are much higher compared to the no-screening regime though still lower than the risk-based regime consistent with our discussion above. Note, the credit rationing in this framework occurs due to the monopoly lender following a pooled pricing regiome.[13] This is different from credit rationing in [54] which arises out of asymmetries of information between borrower and lender. In the [54] model, the borrower has more information about her own credit quality and thus loan demand directly depends on credit quality. In the [54] increasing interest rates alter the distribution of potential borrowers and credit is rationed to avoid these high-risk borrowers.

[13] I motivate the pooling regime as providing liquidity benefits in the TBA market in Section 1.2.

Figure A1: Lender Payoffs in 3 pricing regimes with low screening costs



(a) **Screening Risk-based** (b) **No-Screening Pooled** (c) **Screening Pooled**

This figure represents the lender payoffs in the 3 pricing regimes considered in my conceptual framework. Panel (a), (b) and (c) shows the payoffs for risk-based pricing, no screening pooled and screening pooled regimes respectively. The horizontal axis represents credit quality of the borrower (θ). The blue line in each panel corresponds to the lender profits for each θ. The total area under the blue line (meshed region) is then plotted as the red line. The red line makes it easier to compare total profits across all borrowers in each of the pricing regimes.

Table A4: Credit Rationing: 30-year Fixed Rate Mortgages

Panel A: All loans

	(1) Total Volume	(2) Total Volume	(3) No. of loans	(4) No. of loans
LR Index	0.356**	0.244**	0.337*	0.245*
	(0.152)	(0.111)	(0.181)	(0.127)
MSA-Yr-Qtr FE	X	X	X	X
Controls		X		X
No. of Observations	3671	3671	3671	3671
Number of Clusters	25	25	25	25
Adj.R^2	0.469	0.687	0.525	0.720

Panel B: Dependent variable Log(Total Number of Loans/Housing Stock)

	(1) FTHB	(2) FICO< 640	(3) FICO< 720	(4) No Co-borrower
LR Index	0.193*	0.275	0.260*	0.247**
	(0.103)	(0.161)	(0.147)	(0.108)
CBSA-Yr-Qtr FE	X	X	X	X
Controls	X	X	X	X
No. of Observations	3267	3325	3579	3519
Number of Clusters	25	25	25	25
Adj.R^2	0.518	0.570	0.648	0.720

Panel C: Dependent variable Log(Total Loan Volume/Housing Stock)

	(1) FTHB	(2) FICO< 640	(3) FICO< 720	(4) No Co-borrower
LR Index	0.199**	0.265*	0.248*	0.243**
	(0.0900)	(0.144)	(0.125)	(0.0896)
CBSA-Yr-Qtr FE	X	X	X	X
Controls	X	X	X	X
No. of Observations	3267	3325	3579	3519
Number of Clusters	25	25	25	25
Adj.R^2	0.411	0.536	0.618	0.669

This table shows the weighted least squared (WLS) regressions of the aggregated credit against lender rights for 30-year Fixed Rate Mortgages (FRMs) purchased by the GSEs. GSE data is for the period 2000 to 2005 from the publicly available for 30-year FRM single-family mortgage data provided by Fannie Mae and Freddie Mac. The dependent variable in columns 1 and 2 in Panel A and in all columns in Panel B is the logarithm of the total volume of mortgages aggregated to the lowest level of geocoding available in each year-quarter normalized by the total housing stock. The dependent variable in columns 3 and 4 in Panel A and in all columns in Panel C is the logarithm of the total number of mortgages aggregated to the lowest level of geocoding available in each year-quarter normalized by the total housing stock.

Table A5: Regression Discontinuity using 1992 GSE Act: Covariate Balance

Panel A: RD Estimates

	High					Low				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	< High School Educ.	Unempl. Rate	% Black	%Hispanic	Pop.	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.
UAG	-0.00780	-0.00353	-0.0310	0.0205	-0.0155	0.000133	-0.000834	0.00862	0.0446	0.00746
	(-0.39)	(-1.42)	(-1.09)	(0.60)	(-1.32)	(0.01)	(-0.32)	(0.24)	(1.22)	(0.42)
No. of Obs.	160	160	160	160	160	165	165	165	165	165

Panel B: OLS

	High					Low				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.
UAG Targeted	0.0149	0.00213	-0.0303	0.102**	-0.0754*	0.0284	0.0178	0.0619	-0.0229	0.0780
	(0.41)	(0.50)	(-0.61)	(2.29)	(-2.06)	(0.82)	(1.71)	(0.38)	(-0.31)	(1.25)
No. of Observations	160	160	160	160	160	165	165	165	165	165

This table shows the covariate balance for the regression discontinuity design using the 1992 GSE Act for one percent bandwidths around the RD cutoff. Data is for the period 2000 to 2005 and from the Housing and Urban Development (HUD) U.S. department. The HUD provides data on mortgages purchased by the GSEs (Fannie Mae and Freddie Mac) as part of the 1992 GSE Act. Panel A shows the regression discontinuity (RD) estimates for the covariates percentage Hispanic, percentage black, population and unemployment rate from the Census 2000 data at the 0.90 threshold. Panel B shows the difference in means of the covariates. In each panel, columns 1–5 correspond to census tracts with above median lender rights index. In each panel, columns 5–10 correspond to census tracts with below median lender rights index.

Table A6: Non-GSE Credit Rationing: 30-year Fixed Rate Mortgages

Panel A: Private sector purchases – Non-Jumbo

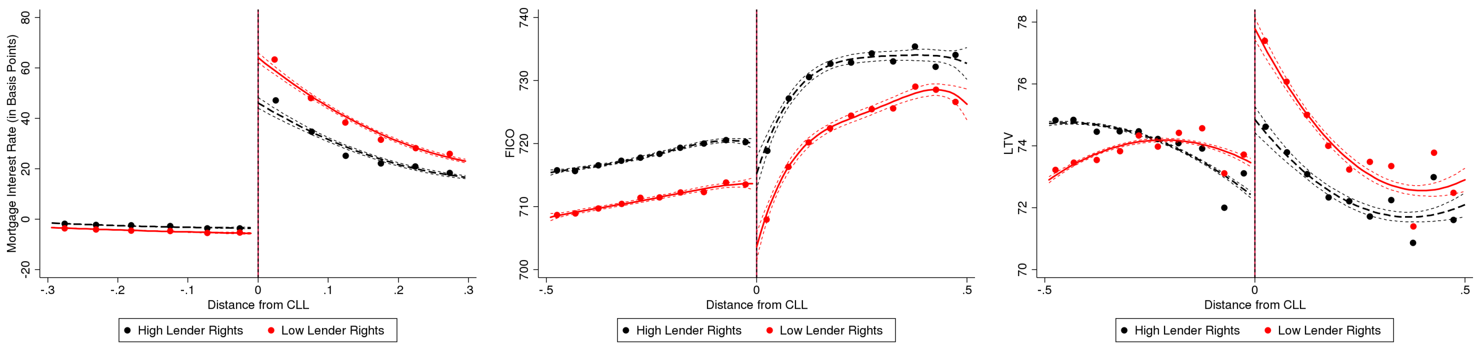
	(1) Total Volume	(2) Total Volume	(3) No. of loans	(4) No. of loans
High Lender Rights	0.252**	0.243***	0.265*	0.294***
	(0.0954)	(0.0748)	(0.135)	(0.0969)
CBSA-Yr-Qtr FE	X	X	X	X
Controls		X		X
No. of Observations	3036	3036	3036	3036
Number of Clusters	25	25	25	25
Adj.R^2	0.623	0.680	0.577	0.653

Panel B: Private sector purchases – Jumbo

	(1) Total Volume	(2) Total Volume	(3) No. of loans	(4) No. of loans
High Lender Rights	0.579***	0.354**	0.573***	0.348**
	(0.153)	(0.154)	(0.154)	(0.150)
CBSA-Yr-Qtr FE	X	X	X	X
Controls		X		X
No. of Observations	2667	2667	2667	2667
Number of Clusters	25	25	25	25
Adj.R^2	0.340	0.709	0.307	0.689

This table repeats the regressions in Table A4 for the non-GSEs (purchases by the private securitizers). This table shows the weighted least squared (WLS) regressions of the aggregated credit against lender rights for 30-year Fixed Rate Mortgages (FRMs) purchased by the private securitizers separately for jumbo and non-jumbo loans. Loans are classified as jumbo (non-jumbo) if the loan amount is greater (less) than the conforming loan limit. Panel A shows the regression results for Non-jumbo loans and Panel B shows the results for jumbo loans. Non-GSE/private securitizer data is from the ABSnet. I use the single-family 30-year Fixed Rate Mortgages for the period 2000 to 2005. The dependent variable in columns 1 and 2 in both panels is the logarithm of the total volume of mortgages aggregated to the lowest level of geocoding available in each year-quarter normalized by the total housing stock. The dependent variable in columns 3 and 4 in both panels the logarithm of the total number of mortgages aggregated to the lowest level of geocoding available in each year-quarter normalized by the total housing stock. For the publicly available 30-year FRM single-family GSE data, I aggregate to the zipcode-MSA-state level (similar to the analysis of Table A4). Columns 2 and 4 in both panels also include the controls percentage Hispanic, percentage black, log(median income) and unemployment rate from the Census 2000 data at the zipcode-MSA-state level. Total housing stock is also from the 2000 Census. All columns include MSA-year-quarter fixed effects and are clustered at the state level. In my analysis, only the 38 MSAs which cross state-borders and have different values of the lender rights index are retained. The lender rights index is calculated based on whether a state follows judicial procedure, whether lenders have recourse to the borrowers assets and on the borrower right-to-redeem. See Appendix A.1 for details on how the lender rights index is calculated. High lender rights in the table is this measure of the lender rights index.

Figure A2: Bunching Analysis: Mortgage Rates



(a) **Mortgage Rates** (b) **FICO** (c) **LTV**

This figure plots the mortgage interest rates (Panel (a)), FICO (Panel (b)) and LTV (Panel (c))against the distance from the conforming loan limit. The horizontal axis represents the ratio of the loan amount to conforming loan limit minus one. Data is for the period 2000 to 2005. Data for loans above the conforming loan limit is the non-GSE data from ABSnet for the same period. Loans below the conforming loan limit is the GSE data from the publicly available 30–year Fixed Rate Mortgage (FRM) single–family mortgage data provided by Fannie Mae and Freddie Mac. High (low) lender rights correspond to above (below) median value of the lender rights index. Each point represents the average value of the outcome (mortgage interest rate) in the 5% interval. The solid line plots the predicted values with separate quadratic distance from conforming loan limit trends on either side of the conforming loan limit. The dashed lines show the 95 percent confidence intervals. The plots use cross-border MSAs and include MSA-year-quarter fixed effects. The plots correspond to bandwidths of .5 around the cutoff.

Table A7: Mortgage Interest Rates across the Conforming Loan Limit

	(1) Bet. 0 to 0.5	(2) Bet. 0 to 0.5	(3) Bet. 0.5 to 1	(4) Bet. 0.5 to 1	(5) Bet. 1 to 1.5	(6) Bet. 1 to 1.5	(7) Bet. 1.5 to 2	(8) Bet. 1.5 to 2
LR Index	0.0612*	0.0228	-0.00261	-0.00619	-0.0847***	-0.0566**	-0.0530**	-0.0484*
	(0.0344)	(0.0153)	(0.0291)	(0.00993)	(0.0293)	(0.0221)	(0.0241)	(0.0244)
No. of Obs.	114517	31445	74752	27990	99061	45729	38369	16888
R squared	0.249	0.256	0.416	0.386	0.631	0.638	0.696	0.696
Type	OLS	CEM	OLS	CEM	OLS	CEM	OLS	CEM

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the differences in mortgage rates across areas with lender rights. In column 1,3,5 and 7 I compare mortgages interest rates within a MSA-year-quarter. In column 2 I use Coarsened Exact Matching (CEM, see [45] and [46]) to match mortgages exactly based on different bins in a given MSA in a given year-quarter. I use loan-to-value (LTV) bins with cut-offs 70, 75, 80, 85, 90 and 95 and FICO (credit score) bins of 620, 660 and 720. Loans are classified as jumbo (non-jumbo) if the loan amount is greater (less) than the conforming loan limit. In this panel, all loans below the conforming loan limit are GSE loans and all loans above the conforming limit are the non-GSE loans. Data is for the period 2000 to 2005. Data for loans above the conforming loan limit is the non-GSE data from ABSnet for the same period. GSE data is from the publicly available 30-year Fixed Rate Mortgage (FRM) single-family mortgage data provided by Fannie Mae and Freddie Mac. Non-GSE is from ABSnet. All columns include MSA-year-quarter fixed effects and are clustered at the state level. In my analysis, only the 38 MSAs which cross state-borders and have different values of the lender rights index are retained. The lender rights index is calculated based on whether a state follows judicial procedure, whether lenders have recourse to the borrowers assets and on the borrower right-to-redeem. See Appendix A.1 for details on how the lender rights index is calculated. High lender rights in the table is this measure of the lender rights index.

Table A8: Complex Mortgages: Aggregate Lending

Panel A: All

	(1) Total Volume: ARMs	(2) Total Volume: Complex	(3) No. of loans: ARMs	(4) No. of loans: Complex
LR Index	0.128***	0.137***	0.0861***	0.139***
	(0.0221)	(0.0360)	(0.0174)	(0.0305)
Region-Yr-Qtr FE	X	X	X	X
Controls	X	X	X	X
No. of Observations	9810	6256	9810	6256
Number of Clusters	96	96	96	96
Adj.R^2	0.856	0.888	0.813	0.897

Panel B: Non-Jumbo loans

	(1) Total Volume: ARMs	(2) Total Volume: Complex	(3) No. of loans: ARMs	(4) No. of loans: Complex
LR Index	0.125***	0.160***	0.0873***	0.158***
	(0.0218)	(0.0343)	(0.0176)	(0.0315)
Region-Yr-Qtr FE	X	X	X	X
Controls	X	X	X	X
No. of Observations	9807	5986	9807	5986
Number of Clusters	96	96	96	96
Adj.R^2	0.800	0.903	0.753	0.889

Panel A: Jumbo Loans

	(1) Total Volume: ARMs	(2) Total Volume: Complex	(3) No. of loans: ARMs	(4) No. of loans: Complex
LR Index	0.172***	0.152***	0.180***	0.155***
	(0.0286)	(0.0359)	(0.0284)	(0.0353)
Region-Yr-Qtr FE	X	X	X	X
Controls	X	X	X	X
No. of Observations	6028	3958	6028	3958
Number of Clusters	96	95	96	95
Adj.R^2	0.819	0.865	0.808	0.861

This table shows the impact of lender rights aggregate lending of ARMs and Complex Mortgages. The non-GSE data is from ABSnet data for the period 2000 to 2005. Columns 1 and 3 in each panel show aggregate lending per housing stock for adjustable rate mortgages. Columns 1 and 3 in each panel show aggregate lending per housing stock for complex mortgages. IO and NEGAM mortgages are together termed as complex mortgages (CM). Panel A shows the results for all loans. Panel B shows results for loans below the conforming loan limit (non-jumbo loans) and Panel C for loans above the conforming loan limit (jumbo loans). All regressions are clustered at the CBSA year-quarter. In each panel, columns 1 and 2 correspond to the logarithm of the total volume per housing stock. Column 3 and 4 correspond to the logarithm of the total number of loans per housing stock.

Table A9: Credit Rationing: HMDA data

Panel A: GSE Purchases (Non-Jumbo)

	(1) Total Volume	(2) No. of loans	(3) Total Volume	(4) No. of loans
LR Index	0.102***	0.0712**	0.112***	0.104***
	(0.0311)	(0.0254)	(0.0314)	(0.0250)
Region-Year FE	X	X		
CBSA-Year FE			X	X
Controls	X	X	X	X
Number of Observations	2484	2484	528	528
Number of Clusters	24	24	234	234
Adj.R^2	0.803	0.790	0.950	0.952

Panel B: Private sector purchases – Non-Jumbo

	(1) Total Volume	(2) No. of loans	(3) Total Volume	(4) No. of loans
LR Index	0.132***	0.0993**	0.102*	0.0979**
	(0.0425)	(0.0362)	(0.0536)	(0.0462)
Region-Year FE	X	X		
CBSA-Year FE			X	X
Controls	X	X	X	X
Number of Observations	2480	2480	526	526
Number of Clusters	24	24	234	234
Adj.R^2	0.763	0.738	0.941	0.933

Panel C: Private sector purchases – Jumbo

	(1) Total Volume	(2) No. of loans	(3) Total Volume	(4) No. of loans
LR Index	0.214***	0.217***	0.322**	0.309**
	(0.0558)	(0.0555)	(0.141)	(0.137)
Region-Year FE	X	X		
CBSA-Year FE			X	X
Controls	X	X	X	X
Number of Observations	2096	2096	372	372
Number of Clusters	24	24	203	203
Adj.R^2	0.778	0.765	0.896	0.891

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the weighted least squared (WLS) regressions of the logarithm at the aggregate CBSA-state-year-quarter level on lender rights for 30 year fixed rate mortgages. Panel B shows results for non-GSE loans below the conforming loan limit (non-jumbo loans) and Panel C for non-GSE loans above the conforming loan limit (jumbo loans). In each panel, columns 1 and 3 correspond to the logarithm of the total volume per housing stock. Column 2 and 4 correspond to the logarithm of the total number of loans per housing stock.

Table A10: Regression Discontinuity using 1992 GSE Act (HUD data): Mean Difference for Covariate Balance

Panel A: 10% Bandwidth

	High					Low				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.
UAG Targeted	0.00606	-0.00288***	-0.0586**	0.0138	-0.00645**	0.00543	-0.00219	0.0205	0.0535***	-0.0210
	(0.49)	(-3.25)	(-2.83)	(0.94)	(-2.48)	(0.56)	(-0.97)	(0.66)	(3.04)	(-1.60)
No. of Observations	1649	1649	1649	1649	1649	1883	1883	1883	1883	1883

Panel B: 5% Bandwidth

	High					Low				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.
UAG Targeted	0.0131	-0.00198	-0.0378	0.0164	-0.0135	-0.00673	-0.00532*	-0.0205	0.104**	-0.0188
	(0.89)	(-1.05)	(-1.71)	(0.92)	(-1.56)	(-0.56)	(-2.11)	(-0.50)	(2.38)	(-0.68)
No. of Observations	811	811	811	811	811	936	936	936	936	936

Panel C: [47] Bandwidth

	High					Low				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.
UAG Targeted	0.0149	0.00213	-0.0303	0.102**	-0.0754*	0.0284	0.0178	0.0619	-0.0229	0.0780
	(0.41)	(0.50)	(-0.61)	(2.29)	(-2.06)	(0.82)	(1.71)	(0.38)	(-0.31)	(1.25)
No. of Observations	160	160	160	160	160	165	165	165	165	165

This table shows the covariate balance for the regression discontinuity design using the 1992 GSE Act.

Table A11: Regression Discontinuity using 1992 GSE Act: RD estimates for Covariate Balance

Panel A: 10% Bandwidth

	High					Low				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	< High School Educ.	Unempl. Rate	% Black	%Hispanic	Pop.	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.
UAG	0.0361***	0.00642***	0.0286***	-0.121***	0.00742**	0.0351***	0.00684***	0.0545***	-0.124***	0.0324***
	(5.38)	(7.32)	(2.74)	(-10.73)	(2.33)	(6.55)	(7.17)	(4.72)	(-11.18)	(5.59)
No. of Obs.	1649	1649	1649	1649	1649	1883	1883	1883	1883	1883

Panel B: 5% Bandwidth

	High					Low				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	< High School Educ.	Unempl. Rate	% Black	%Hispanic	Pop.	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.
UAG	0.0154*	0.00234*	-0.00340	-0.0394**	0.00517	0.0209***	0.00288**	0.0257	-0.0614***	0.0188**
	(1.66)	(1.87)	(-0.24)	(-2.56)	(1.07)	(2.70)	(2.05)	(1.63)	(-3.90)	(2.36)
No. of Obs.	811	811	811	811	811	936	936	936	936	936

Panel C: [47] Bandwidth

	High					Low				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	< High School Educ.	Unempl. Rate	% Black	%Hispanic	Pop.	< High School Educ.	Unemp. Rate	% Black	%Hispanic	Pop.
UAG	-0.00780	-0.00353	-0.0310	0.0205	-0.0155	0.000133	-0.000834	0.00862	0.0446	0.00746
	(-0.39)	(-1.42)	(-1.09)	(0.60)	(-1.32)	(0.01)	(-0.32)	(0.24)	(1.22)	(0.42)
No. of Obs.	160	160	160	160	160	165	165	165	165	165

This table shows the covariate balance for the regression discontinuity design using the 1992 GSE Act for different bandwidths.

Table A12: Regression Discontinuity using 1992 GSE Act: Credit Rationing Estimates

Panel A: Dependent variable Log(Total Number of Loans/Housing Stock)

	(1) High	(2) Low	(3) High	(4) Low	(5) High	(6) Low
UAG Targeted	0.0913*	0.0349	0.157***	0.124	0.467***	-0.446
	(0.0502)	(0.0605)	(0.0514)	(0.110)	(0.127)	(0.326)
CBSA-Yr FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X
No. of Observations	10447	11422	5179	5715	977	1006
Number of Clusters	16	18	16	18	13	17
Adj.R^2	0.413	0.337	0.400	0.313	0.418	0.265

Panel B: Dependent variable Log(Total Loan Volume/Housing Stock)

	(1) High	(2) Low	(3) High	(4) Low	(5) High	(6) Low
UAG Targeted	0.120*	0.0589	0.196***	0.172	0.532***	-0.364
	(0.0588)	(0.0810)	(0.0640)	(0.131)	(0.133)	(0.387)
CBSA-Yr FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X
No. of Observations	10443	11388	5177	5699	975	1004
Number of Clusters	16	18	16	18	13	17
Adj.R^2	0.443	0.323	0.428	0.305	0.470	0.266

This table shows the estimates of credit rationing using a regression discontinuity design based on the 1992 GSE Act for different bandwidths. Data is for the period 2000 to 2005 and from the Housing and Urban Development (HUD) U.S. department. The HUD provides data on mortgages purchased by the GSEs (Fannie Mae and Freddie Mac) as part of the 1992 GSE Act. The dependent variable in Panel A is the logarithm of the total volume of mortgages aggregated to the census tract level in each year normalized by the total housing stock. The dependent variable in Panel B is the logarithm of the total number of mortgages aggregated to the census tract level in each year normalized by the total housing stock. Columns 1–2, columns 3–4 columns 5–6 correspond to bandwidths of 10 percent, 5 percent and 1 percent respectively around the RD cutoff. In each panel, the odd numbered columns (1, 3 and 5) correspond to census tracts with above median lender rights index. In each panel, the even numbered columns (2, 4 and 6) correspond to census tracts with below median lender rights index. Standard errors are included in parenthesis and are clustered at the state level and include MSA-year fixed effects.

Table A13: 10% Bandwidth Regression Discontinuity using 1992 GSE Act (HUD data): Heterogeneity

Panel A: Dependent variable Log(Total Number of Loans/Housing Stock)

	(1) First-time Home Borrower	(2) Age < 45	(3) Minority Borrowers	(4) Single Female Borrower	(5) No Co-Borrower	(6) First-time Home Borrower	(7) Age < 45	(8) Minority Borrowers	(9) Single Female Borrower	(10) No Co-Borrower
UAG	0.0688	0.0817*	0.0663	0.0128	0.100*	-0.0424	0.0505	-0.139	0.0109	0.0651
	(0.0604)	(0.0388)	(0.0386)	(0.0391)	(0.0564)	(0.0737)	(0.0661)	(0.0914)	(0.0544)	(0.0676)
CBSA-Yr FE	X	X	X	X	X	X	X	X	X	X
Controls	X	X	X	X	X	X	X	X	X	X
No. of Obs.	9759	10097	8563	10050	10163	10673	11057	9562	10946	11100
Adj.R^2	0.417	0.655	0.589	0.634	0.561	0.264	0.478	0.393	0.459	0.434

Panel B: Dependent variable Log(Total Loan Volume/Housing Stock)

	(1) First-time Home Borrower	(2) Age < 45	(3) Minority Borrowers	(4) Single Female Borrower	(5) No Co-Borrower	(6) First-time Home Borrower	(7) Age < 45	(8) Minority Borrowers	(9) Single Female Borrower	(10) No Co-Borrower
UAG	0.0931	0.109**	0.0722	0.0209	0.126**	-0.00217	0.0746	-0.135	0.0276	0.0906
	(0.0684)	(0.0445)	(0.0446)	(0.0498)	(0.0570)	(0.0633)	(0.0857)	(0.0929)	(0.0586)	(0.0874)
CBSA-Yr FE	X	X	X	X	X	X	X	X	X	X
Controls	X	X	X	X	X	X	X	X	X	X
No. of Obs.	9672	10090	8427	10044	10157	10121	10974	9273	10803	11030
Adj.R^2	0.532	0.660	0.600	0.697	0.567	0.356	0.445	0.426	0.490	0.403

Columns 1–2 repeats the weighted least squared (WLS) regressions of the logarithm at the aggregate CBSA-state-year-quarter level on lender rights for all mortgages using data provided by the Housing and Urban Development (HUD) U.S. department. Data is for the period 2000 to 2005 provided by Fannie Mae and Freddie Mac. Columns 3–10 show the heterogeneity results. In column 3 "Top tercile minority" is an indicator for whether the average minority share of borrowers falls in the top tercile within each year. Similarly, the remaining columns correspond to whether share of mortgages with co-applicants, share of first time home buyers, and borrower income. Columns 1, 3, 5, 7 and 9 correspond to the logarithm of the total volume per housing stock. Columns 2, 4, 6, 8 and 10 correspond to the logarithm of the total number of loans per housing stock.

Table A14: 5% Bandwidth Regression Discontinuity using 1992 GSE Act (HUD data): Heterogeneity

Panel A: Dependent variable Log(Total Number of Loans/Housing Stock)

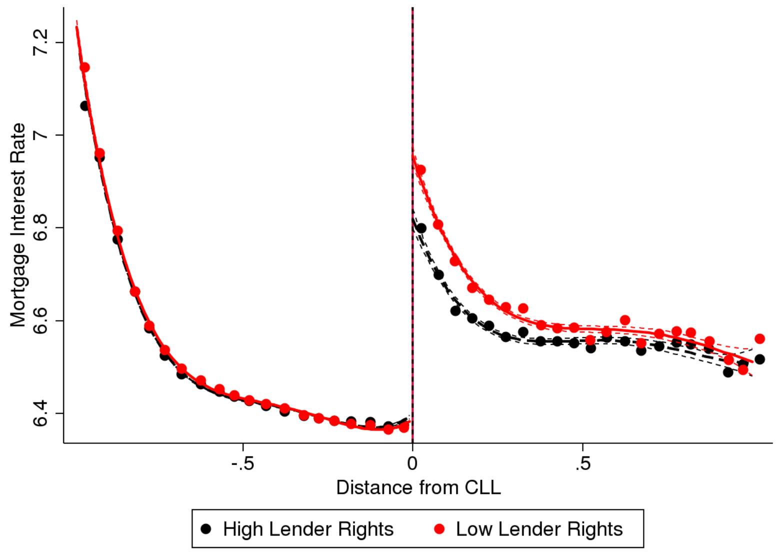
	(1) First-time Home Borrower	(2) Age < 45	(3) Minority Borrowers	(4) Single Female Borrower	(5) No Co-Borrower	(6) First-time Home Borrower	(7) Age < 45	(8) Minority Borrowers	(9) Single Female Borrower	(10) No Co-Borrower
UAG	0.123	0.180***	0.0436	0.140	0.160**	0.0467	0.194*	-0.0616	0.0669	0.209**
	(0.102)	(0.0550)	(0.116)	(0.0891)	(0.0577)	(0.0922)	(0.0983)	(0.132)	(0.0844)	(0.0867)
CBSA-Yr FE	X	X	X	X	X	X	X	X	X	X
Controls	X	X	X	X	X	X	X	X	X	X
No. of Obs.	4858	5010	4258	4991	5043	5338	5529	4796	5466	5550
Adj.R^2	0.423	0.647	0.595	0.629	0.541	0.278	0.458	0.400	0.458	0.413

Panel B: Dependent variable Log(Total Loan Volume/Housing Stock)

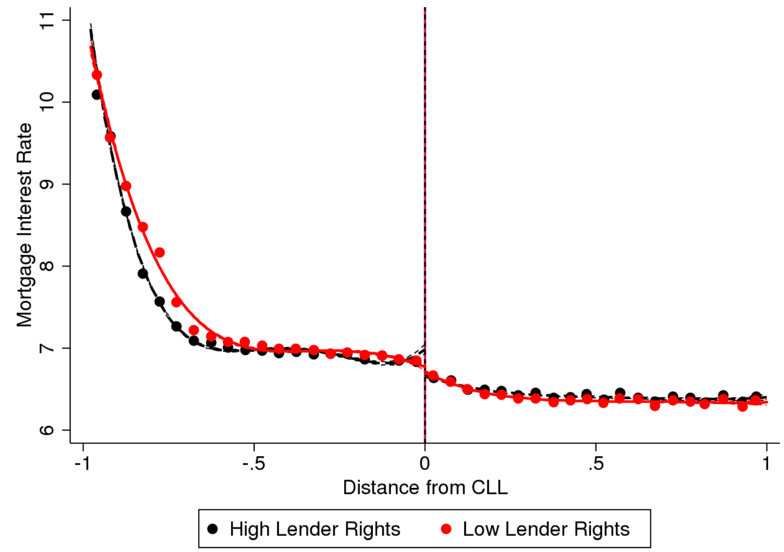
	(1) First-time Home Borrower	(2) Age < 45	(3) Minority Borrowers	(4) Single Female Borrower	(5) No Co-Borrower	(6) First-time Home Borrower	(7) Age < 45	(8) Minority Borrowers	(9) Single Female Borrower	(10) No Co-Borrower
UAG	0.160	0.230***	0.0322	0.148	0.199***	0.100	0.221	-0.136	0.112	0.233**
	(0.108)	(0.0682)	(0.0896)	(0.0972)	(0.0665)	(0.0960)	(0.130)	(0.167)	(0.101)	(0.0996)
CBSA-Yr FE	X	X	X	X	X	X	X	X	X	X
Controls	X	X	X	X	X	X	X	X	X	X
No. of Obs.	4813	5006	4190	4988	5038	5074	5488	4674	5403	5518
Adj.R^2	0.533	0.652	0.606	0.694	0.545	0.370	0.433	0.440	0.491	0.399

Columns 1–2 repeats the weighted least squared (WLS) regressions of the logarithm at the aggregate CBSA-state-year-quarter level on lender rights for all mortgages using data provided by the Housing and Urban Development (HUD) U.S. department. Data is for the period 2000 to 2005 provided by Fannie Mae and Freddie Mac. Columns 3–10 show the heterogeneity results. In column 3 "Top tercile minority" is an indicator for whether the average minority share of borrowers falls in the top tercile within each year. Similarly, the remaining columns correspond to whether share of mortgages with co-applicants, share of first time home buyers, and borrower income. Columns 1, 3, 5, 7 and 9 correspond to the logarithm of the total volume per housing stock. Columns 2, 4, 6, 8 and 10 correspond to the logarithm of the total number of loans per housing stock.

Figure A3: Bunching Analysis: Mortgage Rates



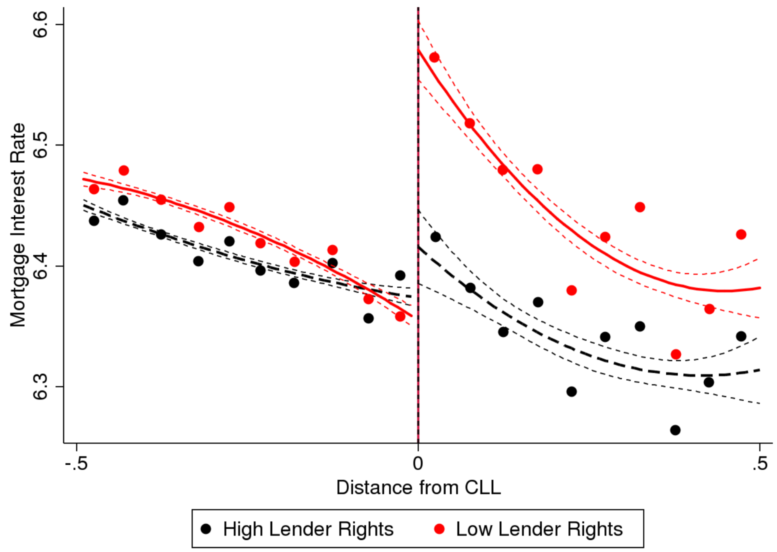
(a) **GSE and Jumbo 30 year FRMs**



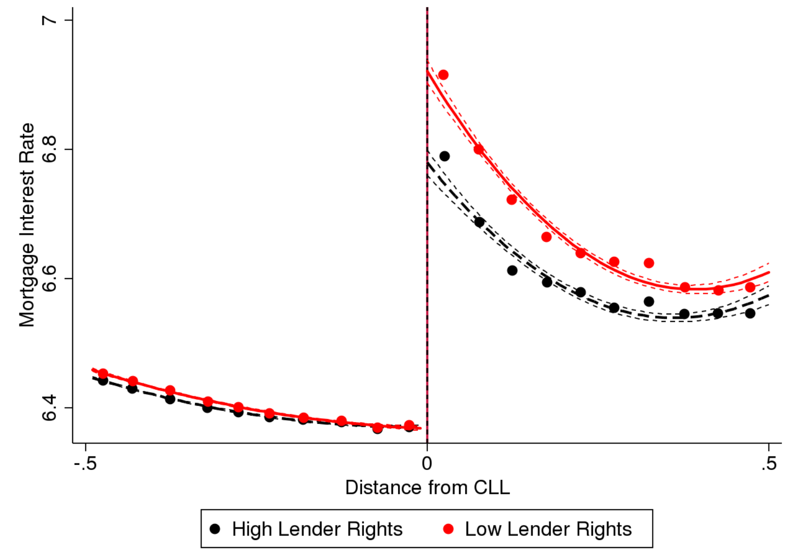
(b) **Only Non-GSE 30 year FRMs**

This figure plots the mortgage interest rates against the distance from the conforming loan limit. The horizontal axis represents the ratio of the loan amount to conforming loan limit minus one. Data is for the period 2000 to 2005. Data for loans above the conforming loan limit is the non-GSE data from ABSnet for the same period. In panel A data for loans below the conforming loan limit is the GSE data from the publicly available 30–year Fixed Rate Mortgage (FRM) single–family mortgage data provided by Fannie Mae and Freddie Mac. In panel B loans below the conforming loan limit only include the non-GSE ABSnet data. High (low) lender rights correspond to above (below) median value of the lender rights index. Each point represents the average value of the outcome (mortgage interest rate) in the 5% interval. The solid line plots the predicted values with separate fourth order polynomial distance from conforming loan limit trends on either side of the conforming loan limit. The dashed lines show the 95 percent confidence intervals. The plots use cross-border MSAs and include CBSA-year-quarter fixed effects. The plots correspond to bandwidths of 100 percent around the cutoff.

Figure A4: Bunching Analysis: Mortgage Rates



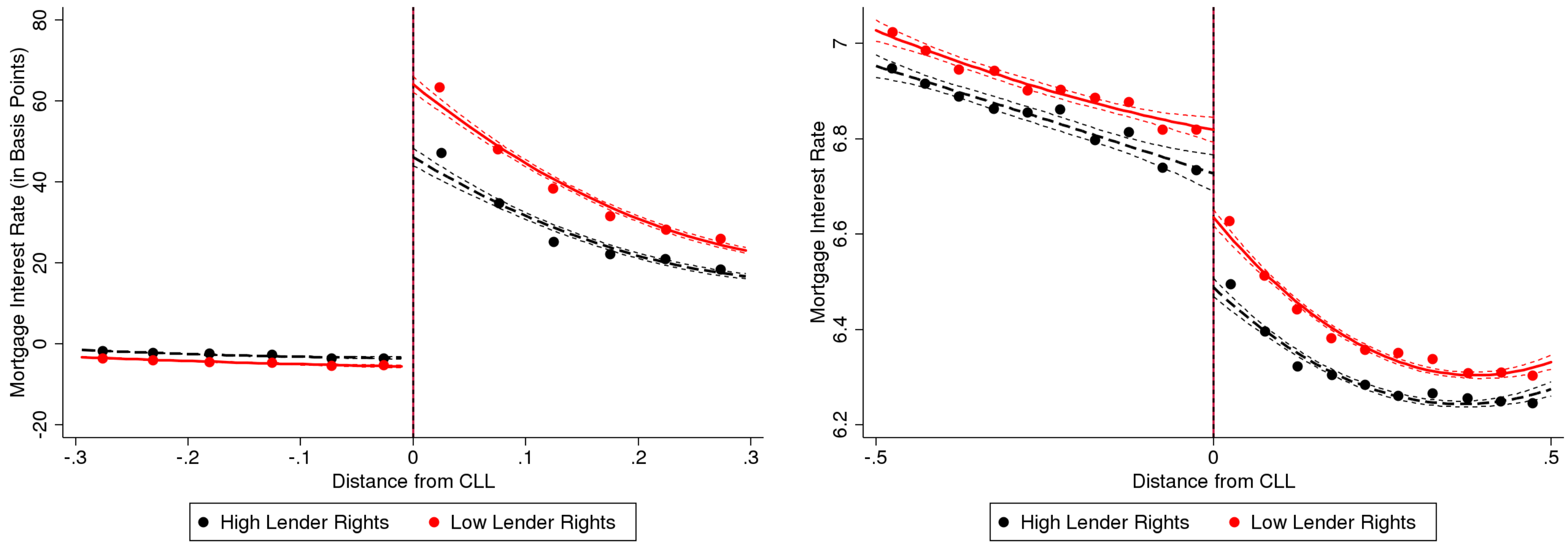
(a) **No Fixed Effects**



(b) **MSA-Year-Quarter Fixed Effects**

This figure plots the mortgage interest rates against the distance from the conforming loan limit. The horizontal axis represents the ratio of the loan amount to conforming loan limit minus one. Data is for the period 2000 to 2005. Data for loans above the conforming loan limit is the non-GSE data from ABSnet for the same period. In panel A data for loans below the conforming loan limit is the GSE data from the publicly available 30–year Fixed Rate Mortgage (FRM) single–family mortgage data provided by Fannie Mae and Freddie Mac. In panel B loans below the conforming loan limit only include the non-GSE ABSnet data. High (low) lender rights correspond to above (below) median value of the lender rights index. Each point represents the average value of the outcome (mortgage interest rate) in the 5% interval. The solid line plots the predicted values with separate fourth order polynomial distance from conforming loan limit trends on either side of the conforming loan limit. The dashed lines show the 95 percent confidence intervals. The plots use cross-border MSAs and include CBSA-year-quarter fixed effects. The plots correspond to bandwidths of 100 percent around the cutoff.

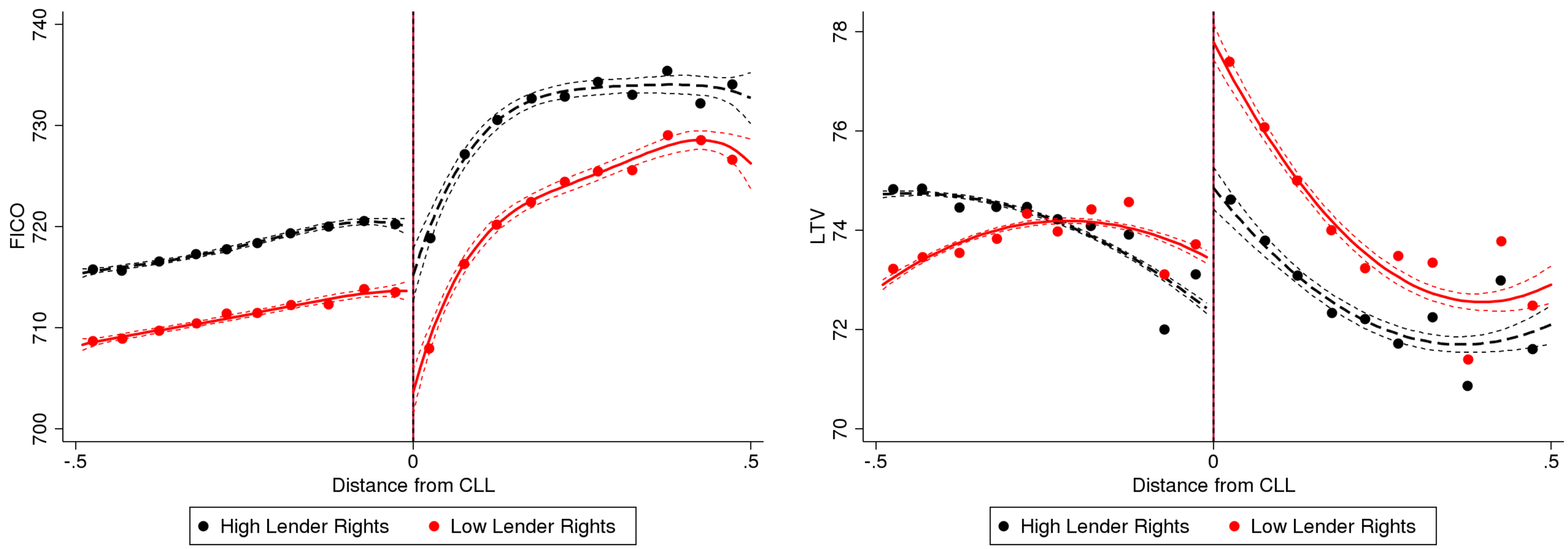
Figure A5: Bunching Analysis: Mortgage Rates



(a) **GSE and Jumbo 30 year FRMs** (b) **Only Non-GSE 30 year FRMs**

This figure plots the mortgage interest rates against the distance from the conforming loan limit. The horizontal axis represents the ratio of the loan amount to conforming loan limit minus one. Data is for the period 2000 to 2005. Data for loans above the conforming loan limit is the non-GSE data from ABSnet for the same period. In panel A data for loans below the conforming loan limit is the GSE data from the publicly available 30–year Fixed Rate Mortgage (FRM) single–family mortgage data provided by Fannie Mae and Freddie Mac. In panel B loans below the conforming loan limit only include the non-GSE ABSnet data. High (low) lender rights correspond to above (below) median value of the lender rights index. Each point represents the average value of the outcome (mortgage interest rate) in the 5% interval. The solid line plots the predicted values with separate quadratic distance from conforming loan limit trends on either side of the conforming loan limit. The dashed lines show the 95 percent confidence intervals. The plots use cross-border MSAs and include MSA-year-quarter fixed effects. The plots correspond to bandwidths of .5 around the cutoff.

Figure A6: Bunching Analysis: FICO and LTV

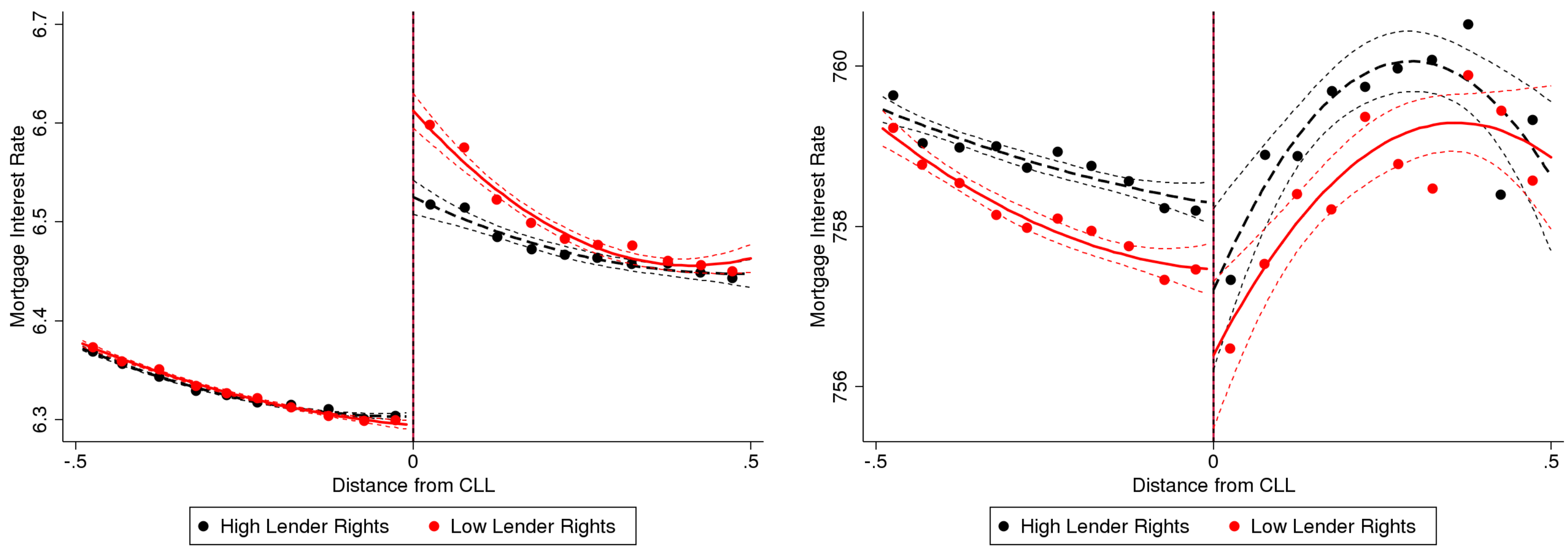


(a) **FICO**

(b) **LTV**

This figure plots the FICO and LTV values against the distance from the conforming loan limit. The horizontal axis represents the ratio of the loan amount to conforming loan limit minus one. Data is for the period 2000 to 2005. Data for loans above the conforming loan limit is the non-GSE data from ABSnet for the same period. In panel A data for loans below the conforming loan limit is the GSE data from the publicly available 30–year Fixed Rate Mortgage (FRM) single–family mortgage data provided by Fannie Mae and Freddie Mac. In panel B loans below the conforming loan limit only include the non-GSE ABSnet data. High (low) lender rights correspond to above (below) median value of the lender rights index. Each point represents the average value of the outcome — FICO in panel (a) and LTV in panel (b) — in the 5% interval. The solid line plots the predicted values with separate quadratic distance from conforming loan limit trends on either side of the conforming loan limit. The dashed lines show the 95 percent confidence intervals. The plots use cross-border MSAs and include MSA-year-quarter fixed effects. The plots correspond to bandwidths of .5 around the cutoff.

Figure A7: Bunching Analysis: FICO Above 720

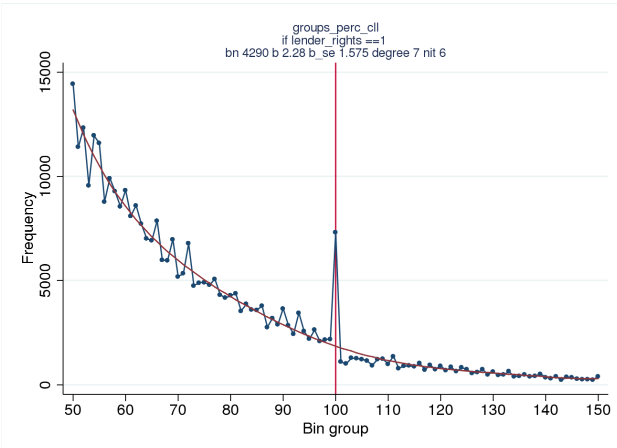


(a) **FICO** (b) **LTV**

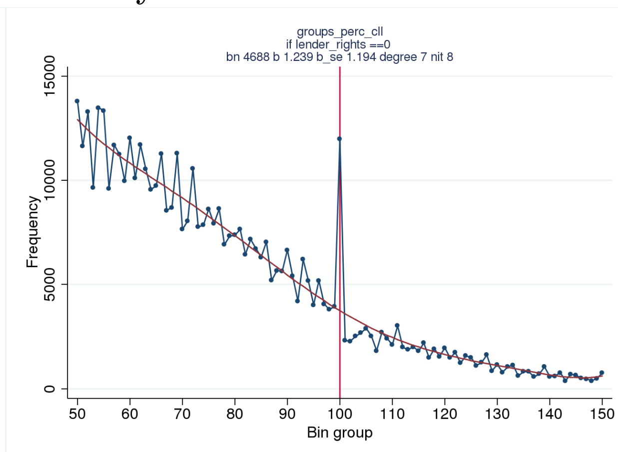
This figure plots the FICO and LTV values against the distance from the conforming loan limit. The horizontal axis represents the ratio of the loan amount to conforming loan limit minus one. Data is for the period 2000 to 2005. Data for loans above the conforming loan limit is the non-GSE data from ABSnet for the same period. In panel A data for loans below the conforming loan limit is the GSE data from the publicly available 30–year Fixed Rate Mortgage (FRM) single–family mortgage data provided by Fannie Mae and Freddie Mac. In panel B loans below the conforming loan limit only include the non-GSE ABSnet data. High (low) lender rights correspond to above (below) median value of the lender rights index. Each point represents the average value of the outcome — FICO in panel (a) and LTV in panel (b) — in the 5% interval. The solid line plots the predicted values with separate quadratic distance from conforming loan limit trends on either side of the conforming loan limit. The dashed lines show the 95 percent confidence intervals. The plots use cross-border MSAs and include CBSA-year-quarter fixed effects. The plots correspond to bandwidths of .5 around the cutoff.

Figure A8: Bunching Estimates: Minority Borrowers

Panel A: Minority

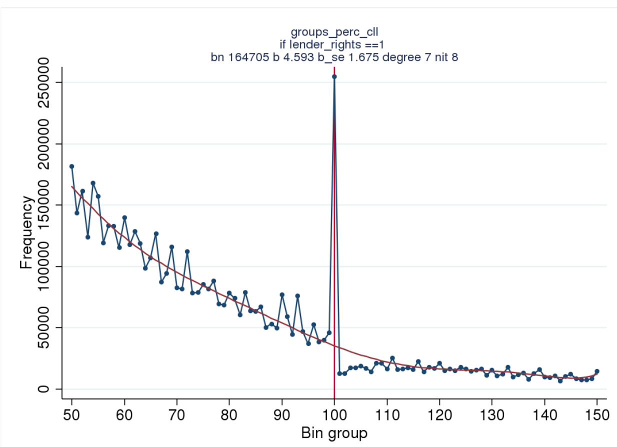


(a) **High Lender Rights**

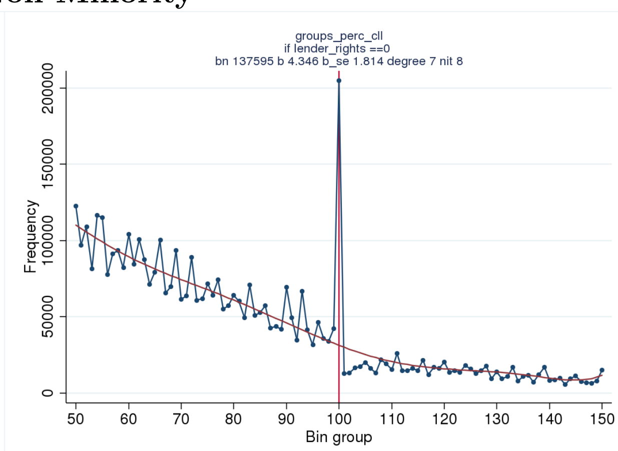


(b) **Low Lender Rights**

Panel A: Non-Minority



(c) **High Lender Rights**

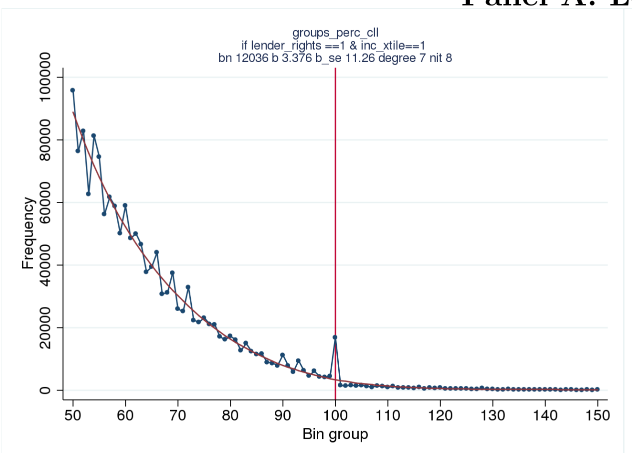


(d) **Low Lender Rights**

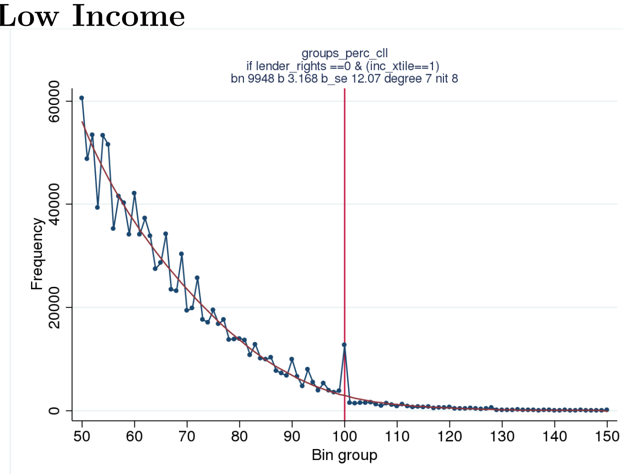
This figure estimates the amount of sorting that occurs around the conforming loan limit for minority (Hispanic or black) and non-minority borrowers. I use the [15] method to estimate the bunching around the conforming loan limit. The left panel shows the graph for mortgages with high lender rights and the right panel corresponds to the mortgages with low lending rights. Data is for the period 2000 to 2005. Data is from HMDA. High (low) lender rights correspond to above (below) median value of the lender rights index. Each dot corresponds to number of loans in the 1 percent bins around the conforming loan limit (100). The red line is the estimated counterfactual density obtained by fitting a 5^{th} order polynomial to the bin counts, omitting the contribution of the bins close to the cutoff.

Figure A9: Bunching Heterogeneity: By Income

Panel A: Low Income

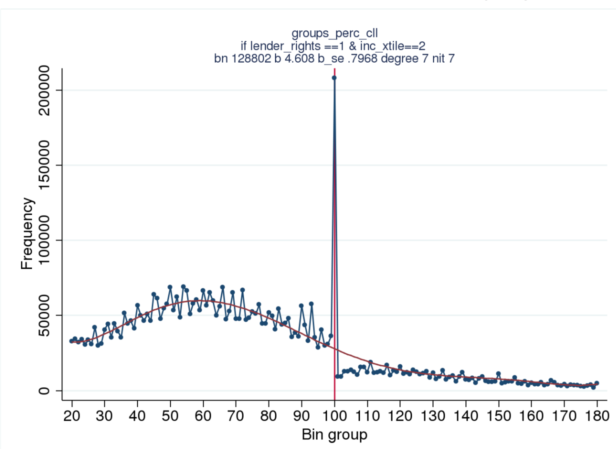


(a) **High Lender Rights**

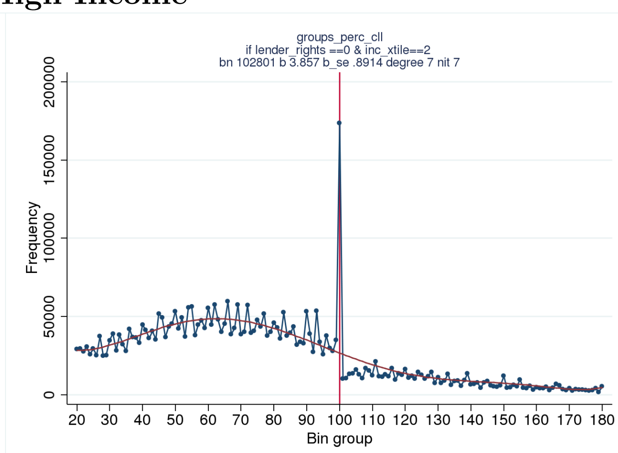


(b) **Low Lender Rights**

Panel A: High Income



(c) **High Lender Rights**



(d) **Low Lender Rights**

This figure estimates the amount of sorting that occurs around the conforming loan limit for high and low income borrowers. High income borrowers refer to the above median income borrowers and low income refer to the below median income borrowers. I use the [15] method to estimate the bunching around the conforming loan limit. Data is from HMDA for the years 2002 to 2005. The left panel shows the graphs for mortgages with high lender rights and the right panel corresponds to the mortgages with low lending rights. Lender rights are calculated by assigning a value of 1-3 for each of the lender rights non-judicial, recourse (fair-market value and non-fair market value) and right-to-redeem. High lender rights correspond to higher index value (easier to foreclose). The binary measure is 0 for states above the median index value and 1 otherwise.

Table A15: Homeownership and House Prices

	(1) % Home-Owners 00-05	(2) HP 2005	(3) HP 2000	(4) HP growth
LR Index	0.0230**	0.0673***	0.0505**	0.0332**
	(0.0104)	(0.0221)	(0.0196)	(0.0161)
CBSA-FE	X	X	X	X
Controls	X	X	X	X
Number of Observations	151	151	151	151
Adj.R^2	0.449	0.764	0.568	0.824

This table shows the impact of credit rationing on homeownership and house prices at the county level. In columns 1 the dependent variable is the percentage of all homeowners who moved in 2000 or after. The dependent variable in column 2 and 3 is the log of the median house value in the year 2000 and year 2005. The dependent variable in column 4 is the house price growth. Data for 2000 is at the county level from the 2000 Census. Data for 2005 is at the county level from the 2005 American Community Survey. All regressions include control variables and MSA fixed effects. Control variables percentage with less than high-school education, percentage Hispanic, percentage black and unemployment rate are from the 2000 Census. All regressions are weighted by total number of households from the 2000 Census. Standard errors are included in parenthesis and are clustered at the state level. In my analysis, only the 38 MSAs which cross state-borders and have different values of the lender rights index are retained. The lender rights index is calculated based on whether a state follows judicial procedure, whether lenders have recourse to the borrowers assets and on the borrower right-to-redeem. See Appendix A.1 for details on how the lender rights index is calculated. High lender rights in the table is this measure of the lender rights index.