Insurance Pricing and Market Structure: A Study of GSE-Securitized Mortgage Loans*

Hsin-Tien Tsai†

January 14, 2019

Link to the Latest Version

Abstract

This paper studies inefficiencies arising from the insurance pricing schedule of government-sponsored enterprises (GSEs) in the U.S. mortgage market, and how these inefficiencies interact with market structure and information asymmetry. I estimate an industry model of borrowers, lenders, and GSEs using loan-level data on repayment and pricing decisions. The estimation exploits a significant change to U.S. banking regulation that gives exogenous variation in credit supply. I find that GSE mortgage insurance pricing results in a redistribution of 56.82 billion dollars (1.60 percent of the mortgage interest) from lower-risk borrowers to higher-risk borrowers and leads to a welfare loss of 519.25 dollars per loan relative to a benchmark with full risk-based pricing. In my counterfactual analysis, I find that, under GSE pricing, a 50 percent decrease in market concentration reduces welfare by 171.77 dollars per loan, while under full risk-based pricing, the same decrease in market concentration improves welfare by 844.29 dollars per loan.

*I thank my dissertation co-chairs, Benjamin Handel and Kei Kawai for their guidance on this project, and my committee member, David Sraer, who provided invaluable advice. This work has benefited particularly from the comments of Matteo Benetton, Zarek Bort-Goldberg, Nan Chen, Przemyslaw Jeziorski, Rupal Kamdar, Amir Kermani, Chang Cheng Song, Katalin Springel, and Steve Tadelis.

†Department of Economics, UC Berkeley (hsintien@econ.berkeley.edu).

1. Introduction

Mispricing of credit risk in mortgage loans was found to be a key cause of the 2008 financial crisis.[1] Given that the Government Sponsored Enterprises (GSEs) Fannie Mae and Freddie Mac securitize the majority of the U.S. residential mortgage loans,[2] much of the credit mispricing may be attributable to the GSEs.[3] The GSEs operate with both implicit and explicit financial support from the government, with a mandate to provide affordable housing for lower-income borrowers.[4] As a result, their credit risk pricing is not entirely determined by the costs and risks of loans and may be influenced by the government's policy goals.

The GSEs may misprice in two ways (Federal Housing Finance Agency, 2009). First, they may underprice most of the mortgage loans to meet the affordability goal. Second, their pricing may involve substantial cross-subsidization from lower-risk borrowers to higher-risk borrowers: they overprice lower-risk borrowers to cover some of the costs from underpricing higher-risk borrowers.

These practices of underpricing and cross-subsidization could distort the market. Higher-risk borrowers are oversupplied with credit under lower prices. This overprovision could result in inefficiency by pushing excessive risks into the market, at the cost of the government and taxpayers. The concerns about overspending taxpayers' money motivate long-lasting policy debates on administrative reforms of the GSEs.[5]

The distortion created by credit risk mispricing may have been exacerbated by the increase in lender competition before the financial crisis. Increased competition among lenders may incentivize them to expand their credit supply and lower their interest rates. This can exacerbate the problem of overprovision to higher-risk borrowers (see Figure 1). Therefore, pro-competitive policies have ambiguous welfare impacts in this sector.

[1] See Financial Crisis Inquiry Commission (2011).

[2] Before the crisis, more than 50 percent of the mortgage loans were securitized by the GSEs. The statistic is calculated using Home Mortgage Disclosure Act data during the 2000 – 08 period.

[3] Federal Housing Finance Agency (2009) estimates an evident gap between the estimated fee revenue from mortgage insurance and the estimated cost and concludes that "[c]redit losses were at historic lows when house price appreciation accelerated rapidly in 2002 through 2005. However, it has become clear that the industry as a whole underpriced mortgage credit risk significantly in those years as well as in 2006 and 2007."

[4] Ambrose and Thibodeau (2004) and Bhutta (2012) provide estimates on the effect of the affordable housing goals on the mortgage rates and credit supply from the GSEs.

[5] The Trump Administration released recommendations for reforms in June 2018. The Obama Administration released a report of a reform plan in February 2011.

This paper empirically investigates inefficiencies arising from GSE pricing, and how they interact with market structure. Understanding the impact of insurance pricing and market structure has important implications on the design of financial regulation and competition policy. It can have a broader impact on the real estate market and overall economic growth.

I develop a rich industry model that quantifies borrower willingness to pay and the cost of a loan. My model features three types of agents: borrowers, lenders, and the GSEs. In the model, a borrower makes an initial purchase decision and subsequent dynamic repayment decisions conditional on the interest rate and her risk parameters. My model allows for unobserved heterogeneity in borrowers' risk types. This unobserved heterogeneity affects borrowers' purchasing and defaulting decisions.

Modeling unobserved heterogeneity is important for capturing potential advantageous selection in this market; a borrower who buys at a higher price has higher willingness to pay and lower risk. When price is lower, the marginal borrowers tend to have higher risks. This extensive margin selection is critical in understanding the economic impacts of pricing and market expansion.

I estimate the model using data that cover over 20 million fixed-rate mortgage loans guaranteed by the GSEs. The data contain detailed information on loan characteristics and long-run performances, both of which are crucial for this analysis. Borrowers' repayment decisions allow me to recover their underlying risk parameters. Under a framework of differentiated Bertrand competition, observed prices help me to identify lender markups and costs.

My estimation leverages exogenous variation in credit supply resulting from a regulatory change in 2004 instituted by the Office of the Comptroller of the Currency (OCC) that exempts national banks from state anti-predatory lending laws (APLs). This regulatory change alleviates the constraints in subprime lending, leading to a market expansion in the subprime market. Indirectly, it lowers the (fixed) cost of credit in the prime market and intensified competition for local lenders.

To test the impacts of the preemption on credit supply in the prime (GSE) market, I implement a difference-in-differences research design. My empirical evidence confirms that conventional mortgage lending is indeed significantly affected by the preemption. I find that the preemption decreases interest rates in markets with state APLs. Moreover, interest rates on average decrease more in markets with a large share of national banks, whereas default rates increase more in those markets. I use this variation to identify unobserved risk types.

The model is estimated using generalized method of moments (GMM). I use borrowers' dynamic repayment decisions as empirical moments. The default probability is driven by observed loan characteristics, such as credit score and income, as well as unobserved heterogeneity. My model picks up a meaningful degree of unobserved heterogeneity in risk types. It highlights the potential welfare impacts of market expansion through selection on unobservables.

I use the estimated model to perform two series of counterfactual analyses. In the first set of counterfactuals, I consider two alternative schemes for GSE pricing: (i) a full risk-based pricing scheme that charges the exact expected cost and (ii) a uniform pricing scheme that charges a flat rate to every loan.

Using full risk-based pricing as a benchmark, I find that the GSEs underprice most of their mortgage loans – the average interest rate is 25.98 basis points lower under GSE pricing. The average mortgage subsidy received per loan is about $2,235$ dollars.[6] GSE pricing increases demand for higher-risk loans, increasing default cost by 843.22 dollars per loan. Overall, it leads to a deadweight loss of 519.25 dollars per loan, or 17.38 percent of the average mortgage subsidy. In addition, I find that GSE pricing redistributes 56.82 billion dollars, or 1.60 percent of the mortgage interest from lower-risk borrowers to higher-risk borrowers.

Next, I consider the effects of uniform pricing. Uniform pricing is an extreme case of cross-subsidization between lower-risk borrowers and higher-risk borrowers. Higher-risk borrowers face a lower price under uniform pricing relative to GSE pricing. Consequently, uniform pricing induces a welfare loss of 33.89 dollars and a default cost of 43.41 dollars relative to GSE pricing. This finding suggests that an increase in the cross-subsidization of insurance pricing could result in a sizable distortionary cost.

In the second set of counterfactuals, I alter market concentration and examine the effects of market structure and information asymmetry under the following four different environments: (i) GSE pricing; (ii) full risk-based pricing; (iii) GSE pricing under symmetric information (i.e., a design that removes asymmetric information in unobserved risks types between borrowers and

[6] I measure the mortgage subsidy as the change in consumer welfare in dollar value from full risk-based pricing to GSE pricing (see Section 8.1).

lenders); and (iv) full risk-based pricing under symmetric information.

My analysis reveals that welfare decreases as the market becomes less concentrated under GSE pricing. A 50 percent decrease in market concentration leads to an additional default cost per loan of 186.35 dollars, or 14.15 percent of the average default cost. These results confirm the intuition that more competition can actually exacerbate the underpricing problem. Overall, it lowers welfare by 171.77 dollars per loan (5.75 percent of the average mortgage subsidy).

The results are quite different under full risk-based pricing. Under full risk-based pricing, the price for each loan is higher than or equal to its marginal cost. There is a deadweight loss associated with lender market power, so welfare increases as market concentration decreases. A 50 percent decrease in market concentration improves welfare substantially by 844.29 dollars per loan (28.27 percent of the average mortgage subsidy).

I also find that welfare impacts of insurance pricing and market structure depend on the degree of information asymmetry between borrowers and lenders. When lenders can price on individual unobserved risk types (i.e., symmetric information), GSE pricing versus full risk-based pricing have different welfare impacts.[7]

Under GSE pricing, symmetric information increases default cost by 5.48 dollars and results in a deadweight loss of 7.36 dollars per loan relative to asymmetric information. When GSE pricing does not price on the additional information, removing private information causes a greater price distortion for credit risks. Profit-maximizing lenders offer lower prices to borrowers who are higher-risk in unobserved types because they are more price sensitive. This price distortion brings higher-risk borrowers into the market and leads to a higher deadweight loss.

On the other hand, symmetric information under full risk-based pricing increases welfare by 132.10 dollars per loan relative to asymmetric information in the same pricing regime. Full risk-based pricing prices on borrowers' unobserved risk types, so that it limits the room for lender price discrimination and reduces demand from higher-risk borrowers (in unobserved types). Under symmetric information combined with full risk-based pricing, welfare increases with the level of competition.

My findings highlight that pricing, market structure, and information asymmetry in the

[7] I assume that GSE pricing does not depend on the information of individual unobserved risk types (i.e., same as the case of asymmetric information).

mortgage market can have important interactions with one another. They also shed some light on many reform scenarios that have been proposed, such as limiting the government's subsidies, restricting the GSEs' cross-subsidies, and shifting toward market risk-based pricing.

1.1. Related Literature

This study contributes to several strands of associated literature. It is closely related to existing empirical work that studies the impact of pricing regulation on efficiency and distributional consequence (Hurst et al., 2016; Bachas, 2017) and the welfare implication of sub-optimal pricing in a setting with selection (Einav et al., 2010; Bundorf et al., 2012; Handel et al., 2015). It is also related to numerical work that investigates the consequence of GSE subsidies (Jeske et al., 2013; Elenev et al., 2016).[8]

My analysis builds upon prior work that structurally estimates parameters on borrower repayment decisions to study policies in consumer credit markets (Bajari et al., 2008; Einav et al., 2012; Kawai et al., 2014) and other prior work that estimates the impact of market structure in financial markets (Hastings et al., 2013; Egan et al., 2017; Benetton, 2017).

This study is also related to papers that document incentive problems for lenders under mortgage securitization, including loan renegotiation (Piskorski et al., 2010; Agarwal et al., 2011), screening efforts (Keys et al., 2010), and a decline in loan quality (Keys et al., 2009; Agarwal et al., 2012; Krainer and Laderman, 2014).

This study adds to the literature on the interaction of horizontal market structure with vertical contracts. Theoretical work has shown that competition could distort the incentives of downstream firms in deposit markets with deposit insurance (Keeley, 1990; Hellmann et al., 2000; Allen and Gale, 2004) and in retail markets (Mathewson and Winter, 1984; Rey and Tirole, 1986; Winter, 1993). In the context of mortgage market, empirical knowledge is scant.

Lastly, this study is also related to a growing body of recent empirical work that examines how market structure in mortgage markets changes various market outcomes, including lender

[8]Both of them use simulation methods to quantify effects of GSE subsidies and find results similar to mine. I take an empirical approach and explore a wider set of questions, for example, the impact of price discrimination. Although my empirical approach allows me to construct a richer and more realistic demand model, it imposes more constraints on identification and computation. I do not incorporate endogenous house prices and allocation of capital as they do.

types and loan characteristics (Rosen, 2011), mortgage rates (Scharfstein and Sunderam, 2014), distributional consequences (Tewari, 2014), refinancing incentives (Agarwal et al., 2015), housing prices and credit supply (Favara and Imbs, 2015), and contract features and pricing strategies (Di Maggio et al., 2016; Agarwal et al., 2017).

The remainder of this paper is presented as follows. Section 2 provides an overview of the mortgage industry. Section 3 describes in detail the data sets used for the analysis. Section 4 describes my empirical model. Section 5 discusses the identification strategy and provides the reduced-form results that support my empirical setup. Section 6 presents my estimation procedure. Section 7 shows my parameter estimates and discusses model fit. Section 8 performs and discusses a series of counterfactual simulations. Section 9 discusses conclusions. Additional technical details and robustness checks are available in the appendices.

2. Industry Background

This section provides an overview of the mortgage market in the United States. The mortgage market is organized into two segments, a primary and secondary market. Most of the mortgage loans go through both primary and secondary markets and are funded by capital markets.

The primary mortgage market is where financial institutions provide mortgage loans to home buyers. Borrowers and lenders meet and negotiate lending terms to create a mortgage transaction. Lenders include mortgage brokers, mortgage bankers, and financial institutions such as commercial banks, credit unions, and savings and loan associations. The original cost of a loan includes the commission of a loan officer, expenses from loan processors, and the fees associated with an underwriter. Lenders make money on a mortgage through the origination fee, which is an upfront fee charged for processing a new loan application. The origination fee is quoted as a percentage of the total loan and is generally between 0.5 and 1 percent on mortgage loans. Lenders may also charge other fees, such as processing fees and application fees.

The secondary mortgage market trades mortgage loans and Mortgage Backed Securities (MBS). A mortgage usually comes with a very large loan balance; lenders cannot afford to keep every loan they provide without exhausting their funds. Most mortgages that originate in the primary mortgage market are sold[9] to mortgage securitizers, such as pension funds, insurance companies, and the federal government, in the secondary mortgage market. In most cases, a lender receives MBS or cash in exchange for the loans. How much money a lender receives varies depending on the actual interest rate on the loan. A loan with a higher rate is worth more as it produces more cash flow. The sold mortgage ends up as a part of a package of the pools of mortgages that, so bundled, turn into securities and bonds sold to investors in the capital market. This directly affects the amount and the cost of funds in the primary market. The secondary mortgage market is dominated by the government agency Ginnie Mae and the GSEs Fannie Mae and Freddie Mac.

2.1. Government-Sponsored Entities

Before the subprime mortgage crisis in 2008, the GSEs Fannie Mae and Freddie Mac guaranteed more than 50 percent of all U.S. mortgages. The GSEs have been regulated by the Federal Housing Finance Agency (FHFA). According to their financial statements, the total loans purchased per year are valued at approximately 588 billion dollars, on average, from 2000 to 2008.

These companies purchase conforming mortgages from lenders. Conforming mortgages are those that meet certain borrower quality characteristics, such as credit score, debt-to-income (DTI) ratio, and loan-to-value (LTV) ratio. The GSEs provide a guaranteed return on their MBS, which are backed with the principal and interest on the conforming mortgage loans that are packaged together in pools. The guarantee makes MBS a product with a safe return relative to some investment products, such as stocks.

2.1.1. GSE Mortgage Insurance Pricing

The GSEs charge a guarantee fee for providing a guarantee. The guarantee fee is meant to cover the projected credit losses from borrower defaults over the terms of the loans, administrative expenses, and a return on capital. The guarantee fee is an important determinant of the cost of

[9] A lender normally sells a mortgage with its service retained. This means that originating lender retains the services of the loan. Borrowers still make their monthly payment to that lender. The lenders accept the payment and make a servicing fee out of the payment (typically around 0.125 percent). The rest of the payment is passed along to the party that purchased the loan (e.g., Fannie Mae, Ginnie Mae, Freddie Mac, Farmer Mac, private securitization and commercial bank, savings bank, or savings association). The lender can also sell the loan service released. Lenders gain additional income by selling the service right.

mortgage credit for mortgage borrowers. A lender typically passes the cost of the guarantee fee to the borrower in the form of a higher interest rate on the mortgage.[10]

Fannie Mae and Freddie Mac establish prices on base rate changes according to the required return for MBS that the GSEs' investors demand. The guarantee fee pricing is based on a grid of specific risk attributes, including LTV ratio and credit score. The GSEs also provided a pricing discount to lenders that delivered a larger volume of loans.[11]

Many questioned whether the GSEs' credit risk pricing models adequately assessed the costs and risks of loans. Federal Housing Finance Agency (2009) provides evidence that the GSEs underpriced the mortgage loans before the mortgage crisis and set higher prices for lower-risk loans to subsidize lower prices for higher credit risk loans. After the crisis, the average guarantee fee was greatly increased as a surcharge for challenging market conditions.[12]

Hurst et al. (2016) show that GSE pricing does not involve regional risk-based pricing, whereas the interest rates on loans from private securitizers are positively related to ex-ante regional default risk. The political constraint is a plausible explanation for the fact that the GSEs do not price on local economic risk.

2.2. *Subprime Lending and Mortgage Crisis*

Subprime mortgages refer to mortgage loans issued to those who have weak credit histories and those with a greater risk of loan default than prime borrowers. Subprime lending expanded dramatically in the early 2000s and reached its peak from 2004 to 2006. Around 1995, the GSEs that, until then, were relatively conservative and stayed away from buying subprime loans directly, were pressured to take on risk to be profitable. As the secondary market shifted from a duopoly to a competitive market, the market of prime borrowers became limited. To compete with pri-

[10]A lender's guarantee fee payment generally takes the form of ongoing monthly payments and may also include an upfront payment at the time of loan acquisition. Whether the GSEs charge guarantee fees to lenders as ongoing fees or upfront fees typically makes no difference to borrowers because they generally are included in the interest rate charged to the borrowers.

[11]In August 2012, the FHFA took action to remove this pricing disparity. They directed the GSEs to raise the guarantee fee more for lenders who exchange loans for MBS and offer lower fees to lenders that sell loans for cash. This helps reduce the pricing disparity between large and small volume lenders because smaller lenders tend to sell loans for cash.

[12]The guarantee fee adjustments are often rigid and bureaucratic. According to the FHFA's report to Congress on guarantee fees in 2016, Fannie Mae and Freddie Mac have made six changes to the single-family guarantee fees with the guidance of their regulator in the past 10 years.

vate label securitizers, the GSEs encouraged lenders to loosen underwriting standards and allowed riskier mortgages to subprime borrowers. They also started to guarantee non-traditional products in response to the prevalence of these products in the primary market.

As home prices started to plummet after the collapse of the housing bubble beginning in 2007, many borrowers ended up owing more than their property was worth. Most of them chose to foreclose their houses. Research has shown that the subprime meltdown was the consequence of the decline in lending standards along with the increase in securitization of subprime mortgages (Dell'Ariccia et al., 2008; Mian and Sufi, 2009; Demyanyk and Van Hemert, 2009).

The GSEs were vulnerable during the mortgage crisis. Credit losses from their mortgage insurance activities had started rising up to roughly 7 billion dollars in 2007 and accelerated to about 44 billion dollars in 2008 according to their financial statements.[13] By the fourth quarter of 2008, both GSEs had negative core capital positions, triggering an insolvency concern, and were put under a conservatorship by their regulator, the FHFA. According to the U.S. Securities and Exchange Commission, Fannie Mae drew 116.1 billion dollars and Freddie Mac drew 71.3 billion dollars from the U.S. Treasury in conservatorship to cover their foreclosures and credit losses.

2.2.1. Anti Predatory Lending Laws

Predatory lending practices[14] were one of the factors contributing to the high mortgage default rates among subprime mortgages (Agarwal et al., 2014). As predatory lending grew in 1994, Congress enacted the Home Ownership and Equity Protection Act (HOEPA), which restricted

[13]For example, Fannie Mae's 10K filing in 2007 reported that "[w]e are experiencing high serious delinquency rates and credit losses across our conventional single-family mortgage credit book of business, especially for loans to borrowers with low credit scores and loans with high loan-to-value ratios. In addition, in 2007 we experienced particularly rapid increases in serious delinquency rates and credit losses in some higher risk loan categories, such as alt-A loans, adjustable rate loans, interest only loans, negative amortization loans, loans made for the purchase of condominiums, and loans with second liens. Many of these higher risk loans were originated in 2006 and the first half of 2007."

[14]Predatory lending practices have been prevalent in the subprime loan markets. According to the Federal Deposit Insurance Corporation (FDIC), illegal predatory lending typically involves (i) imposing unfair and abusive loan terms on borrowers, often through aggressive sales tactics, (ii) taking advantage of a borrower's lack of understanding of complicated transactions, and (iii) outright deception. In other words, predatory lending generally refers to lending practices in which lenders take advantage of borrowers' lack of understanding. Most predatory lending targets less sophisticated borrowers, usually those with a lower income and credit score.

the lending terms and practices for mortgages with either very high APR or restricted terms on total points and fees. HOEPA regulated only around 1 percent of subprime mortgages (Bostic et al., 2008). States also enacted stronger legislation to deal with the problem of predatory lending. North Carolina was the first state to pass an anti-predatory lending law (APL) in 1999, followed by 20 other states. Table 13 shows the states with anti-predatory lending laws and the laws' dates of implementation.

APLs are designed to protect consumers by restricting the origination of loans with predatory features, therefore they focus on the practices and loan terms rather than interest rates. APLs in most states prohibit the following activities in mortgage markets, though prohibited activity is not limited to the following.

First, they set minimum credit criteria that a borrower must meet for the credit to be considered. Second, they restrict pricing structure and the terms of credit, including repayment schedules, amortization, balance, payments due, minimum payments, and term to maturity. They also limit the use of balloon payments and prepayment penalties, which are used to make a loan more affordable by lowering monthly payments. Third, they establish maximum aggregate loan amounts that may be lent with the security of real estate. Fourth, they require licensing, registration, filings, and reports from mortgage lenders to local authorities. Lenders are also required to provide mandated statements and disclose certain information in the credit application forms. Lastly, they prohibit certain advertising activity to prevent lenders aggressively soliciting to vulnerable borrowers, for example, failing to explain the terms of the loan and dissuading the borrower from other lower cost options.[15] It is costly for lenders to comply with all state and local lending laws.

Empirical work has shown that anti predatory laws significantly affected the credit supply in the mortgage market. Pennington-Cross and Ho (2008) show that the APLs increased the cost of credit. The loans originating in states with APLs had higher interest rates for fixed-rate loans than in unregulated states. White et al. (2011) find that state APLs were associated with a reduction in default probability from riskier borrowers. Agarwal et al. (2014) provide evidence that restrictions from APLs significantly reduced the number of loans that were originated.

[15] See Ho and Pennington-Cross (2005) for a comprehensive summary of various local predatory lending laws.

2.2.2. OCC Preemption

In January 2004, the Office of the Comptroller of the Currency, a division of the Treasury Department, issued federal regulations that preempted the ability of state attorneys general to enforce state APLs against the lenders they supervise[16] (*i.e.*, national banks and their operating subsidiaries).[17] Specifically, the OCC waived state APLs and enforcement of loan terms (e.g., LTV requirements and provisions for prepayment penalties) for national banks and their operating subsidiaries. This means that in states with APLs, two different regulatory regimes were operative in the same markets during the 2004 – 08 period covered in this paper. National banks became substantially less constrained by APLs in the states outlined in Table 13.

Recent empirical evidence suggests that the preemption rule had an important impact on the mortgage market. Di Maggio et al. (2016) find that national banks significantly increased origination of loans with prepayment penalties with the preemption. Facing more competition, local lenders increased the origination of riskier loans that were not regulated by the state predatory-lending laws. Di Maggio and Kermani (2017) show that OCC preemption resulted in an 18 percent increase in annual loan issuance in states with local APLs. The effects on credit expansion are stronger in counties with a higher fraction of national banks.

I follow this literature and exploit the OCC preemption in 2004 as an exogenous shock to the credit supply.[18] The APLs have no direct effect on lending in the conventional loan market. However, I find empirical evidence of an indirect credit expansion effect in the GSE market.[19] A possible explanation is that the subprime market expansion led to a lower fixed operating cost for lenders. Credit supply for GSE loans was indirectly affected by the preemption because of

[16] Mortgage lenders are supervised by different government agencies. National banks and Federal thrift institutions are regulated by the OCC or the Office of Thrift Supervision (OTS). State banks and state-chartered thrift institutions are supervised by either the Federal Reserve System, the Federal Deposit Insurance Corporation (FDIC), or by their chartering state. Credit unions are supervised by the National Credit Union Administration (NCUA), while nondepository mortgage companies are regulated by the Department of Housing and Urban Development (HUD) and the Federal Trade Commission (FTC).

[17] 12 C.F.R. §§7.4007-7.4009 (2004); 12 C.F.R. §34.4 (2004)

[18] Many papers have explored the exogenous sources of variation in credit supply from financial deregulation. A widely used source, for example, is the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 (Jayaratne and Strahan, 1996; Rice and Strahan, 2010; Tewari, 2014; Favara and Imbs, 2015; Agarwal et al., 2017). Other papers have used bank mergers (Scharfstein and Sunderam, 2014; Nguyen, 2014) and partial deregulation on state anti-predatory laws (Di Maggio et al., 2016; Di Maggio and Kermani, 2017).

[19] White et al. (2011) mention that "[p]reemption may have had subtle effects as well, such as pushing the market toward looser underwriting standards overall. The 2004 OCC preemption may have been seen by many lenders as a tacit endorsement of loosened underwriting guidelines and regulations."

the lower (fixed) cost of credit.

3. Data

The data used in this analysis come from three sources. The first is a loan-level credit performance data set from Fannie Mae and Freddie Mac. The second is the loan application data required by the Home Mortgage Disclosure Act (HMDA) of 1975. The third is the housing price index which measures price movement of single-family houses. I discuss each aspect of the data used in this analysis below.

3.1. Loan Performance Data

To support the risk sharing and transparency encouraged by the FHFA, Fannie Mae and Freddie Mac made available a portion of single-family fixed-rate mortgages that they purchased or guaranteed from 2000 – 2017.[20] The data consist of two parts: acquisitions and performance files.

The acquisitions file provides characteristics of loans that are acquired by Fannie Mae and Freddie Mac at the loan origination level. Loan characteristics include credit score, DTI ratio, LTV ratio,[21] name of lending institution, unpaid principal balance, property zip code, loan purpose (e.g., home purchase, no cash-out refinance, cash-out refinance),[22] occupancy status (e.g., primary homeowner, second homeowner, investor), and property type (e.g., single-family, condo, co-op, manufactured housing, planned unit development).

The performance file provides monthly credit performance and actual loss information. Credit performance information includes monthly loan balance, delinquency status (up to the earliest of the following termination events: prepaid or matured/voluntary payoff), foreclosure

[20] The following types of mortgages in their portfolio were excluded from the data: adjustable rate mortgages (ARMs), initial interest mortgages, balloon mortgages, government-insured mortgages, Home Possible/Home Possible Neighborhood Solution mortgages, along with other affordable mortgages (including lender branded affordable loan products), and the mortgages delivered under alternate agreements.

[21] DTI and LTV are the common measures for the GSEs to determine the risk of a loan. DTI is the ratio between the amount of recurring debt and a borrower's gross income. LTV is the ratio between the amount of money borrowed and the value of the property.

[22] A home purchase loan is any loan secured by or made to purchase a dwelling. Refinancing is any property-secured loan that replaces and satisfies another property-secured loan to the same borrower.

alternative group (short sale, third party sale, charge off, or note sale), and repurchase prior to property disposition. The credit loss associated with a terminated loan can be calculated using the actual loss information disclosed.

3.2. HMDA Data

Loan performance data contain only a portion of the loans that originated in the primary market. To supplement the loan performance data, I use HMDA data collected by the Federal Financial Institutions Examination Council (FFIEC) to construct a measure of market structure. HMDA data is released on an annual basis. The data contain the information on every application that was made since 1990 in the U.S., including loan characteristics, applicant demographic (e.g., income, geographic area, ethnicity, race, and sex), lender identifier,[23] supervisory/regulatory agency, the reporting institution, whether the loan application was granted and sold, and the types of purchaser if the loan was sold. The HMDA provides links to lender identifiers consistently starting in 2010.

3.3. Housing Price Index

I collect the housing price index (HPI) from the Federal Housing Finance Agency. The HPI measures average price changes in repeat sales or refinancing on the same properties. This information is obtained by reviewing repeat mortgage transactions on single-family properties whose mortgages were purchased or securitized by Fannie Mae and Freddie Mac. The HPI serves as a timely, accurate indicator of house price trends at various geographic levels.

3.4. Descriptive Statistics

The sample period for this analysis is from 2000 – 2008.[24] I link these mortgages with performance data using a loan identifier that allows me to track the post origination performance of these loans for at least 10 years. I only consider fixed-rate mortgage loans with 30-year maturity

[23]The agency code along with the lender identifier is a unique combination that represents a specific institution. I am not able to link specific institutions loan performance data to HMDA because the identifiers are not consistent before 2010.

[24]I performed my reduced-form analysis using different time periods, for example, from 2000 – 2011. The results are not qualitatively or quantitatively different in a meaningful way.

in the structural estimation to simplify the borrower's maturity choice.[25] To be consistent with the structural estimation, I use the same sample in estimating reduced form evidence. However, the reduced-form results do not change qualitatively or quantitatively in a meaningful way when using all the fixed-rate loans in the data with different maturities.

I present the loan-level statistics of my sample in Table 1. My final sample covers approximately 22.73 million loans. The average contractual interest rate of a loan is 6.42 percent. The credit score is reported using the Equifax score, which ranges from 300 to 850. 95.57 percent of the loans cover prime borrowers.[26] Income states the monthly income of the primary borrower. The DTI and LTV ratios are 34.96 percent and 73.27 percent on average.[27] Refinancing loans account for 69.91 percent of loans, while 40.84 percent are cash-out refinance, and 29.08 percent are no cash-out refinance. 91.07 percent of borrowers are principal homeowners. Most property is planned unit development (78.82 percent) or manufactured housing (12.36 percent). I assign a loan as a defaulted loan if the GSEs incurred any loss after the loan is terminated. The default rate is 3.60 percent on average.

I calculate the market share at 3-digit zip code level using HMDA data.[28] I report the statistics of the market-level measure in Table 2. Each market has 55.31 lenders on average. National banks are usually larger lenders; they account for 22.36 percent of the number of lenders but account for 39 percent of market share on average. I also report the Herfindahl-Hirschman Index (HHI), which I calculate by squaring the market share of each lender competing in a market. The mortgage market is moderately competitive during my sample period, as the HHI is 0.16 points on average.[29]

[25] A 30-year fixed-rate mortgage is the most common form in the data, accounting for around 70 percent of the loans.

[26] Equifax scores above 720 are generally considered to be prime borrowers or super-prime borrowers. Equifax scores below 620 are generally considered to be subprime borrowers.

[27] 43 percent DTI and 80 percent LTV are used in the industry as an important threshold for first-lien mortgages. Above the threshold, the pricing and availability of loans change very significantly. Having less than 36 percent DTI and 75 percent LTV presents a lower risk to lenders.

[28] I only consider the loans that were approved and originated when I calculate the market share.

[29] Agencies generally consider markets in which the HHI is between 0.15 and 0.25 to be moderately concentrated and consider markets in which the HHI is excess of 0.25 to be highly concentrated.

4. Model

My model starts with loan origination. A borrower i enters with a property valued v_{i0}. Given the interest rate offered, a borrower first chooses whether or not to purchase a fixed-rate mortgage with a 30-year maturity. If she decides to purchase, the borrower then chooses a down payment size D_i and borrows a loan size equal to $v_{i0} - D_i$. The borrower then repays the loan every period or chooses to default on the loan.

A lender j offers an interest rate r_j in view of the loan characteristic $\boldsymbol{x}$, such as the borrower's income, credit score, down payment, property value, and loan type. The interest rate is chosen based on the parameters of the demand system, competition from other lenders, and marginal cost. Marginal cost captures the guarantee fees charged by the GSEs. I begin my discussion of the borrower's decisions in reverse order.

4.1. *Mortgage Borrowers*

Without mortgage repayment, in a given period t, the borrower i's baseline period utility is given by the following CRRA utility function

$$\underline{u}_i(\boldsymbol{x}, t) = \frac{I_{it}^{(1-\gamma_i)} - 1}{1 - \gamma_i}.$$

The expression is the indirect utility function, which gives the utility level a borrower can achieve by spending the income I on the consumption goods, including housing rental. Since I do not observe an individual's saving and wealth, I simplify the model by assuming that borrowers do not save. The parameter γ measures the degree of relative risk aversion that reflects the intertemporal substitution factor in consumption.[30] I provide more discussion on this assumption in Section 4.4.

4.1.1. *Utility: Repaying a Mortgage*

Suppose a borrower i took a loan of size $v_{i0} - D_i(\boldsymbol{x}, p_{ij})$ with maturity T months; she consumes her monthly income I minus the required mortgage repayment m every period. Additionally,

[30] Estimating a model with saving may require assumptions on initial wealth and saving returns. It adds a continuous and unobserved state variable and increases computational complexity substantially.

the borrower derives utility from owning a house each period. This value from home ownership can be interpreted as a monetary gain from leasing the house. When a borrower chooses to live in the house (i.e., leases the house to herself), the monetary gain can be seen as saving rental costs.

The borrower i's period utility when the loan was originated for t period is given by

$$u_i(\boldsymbol{x}, p_{ij}, t) + \epsilon_{it} = \frac{\left(I_{it} - m(\boldsymbol{x}, p_{ij}, t)\right)^{(1-\gamma_i)} - 1}{1-\gamma_i} + \lambda_i \frac{v_{it}^{(1-\gamma_i)} - 1}{1-\gamma_i} + \epsilon_{it}, \tag{1}$$

where ϵ is a period-specific unobserved component in the utility function. λ is an individual-specific parameter that captures the unobserved heterogeneity in the utility derived from home ownership relative to the consumption. v_{it} is the nominal rental income from leasing, which depends on 0.8% of real-time housing prices to capture the market rental value.

The specification is similar to Campbell and Cocco (2015). Borrower preferences are separable in consumption and housing preference weighted by λ. In their paper, the parameter reflects the relative importance of terminal wealth from the housing (a function of housing prices) and other financial sources. The λ here measures the heterogeneity in housing preference (i.e., some people value home ownership more than others) and may control a few more things. For example, borrowers can lease the house and may have different leasing costs (i.e., maintenance expenses, hassle costs, and tax rates). Borrowers can have heterogeneous (expected) returns on other investment options such as buying stocks and bonds, so they may have different opportunity costs of housing. The same rental income may be valued differently for borrowers due to these explicit and implicit costs.

m is the monthly required payment. If offered interest rate p_{ij} from lender j, a borrower is required to make the monthly required payment corresponding to her choice of down payment D_i in each period t:

$$m(\boldsymbol{x}, p_{ij}, t) = \begin{cases} \left(\frac{f^{-1}(p_{ij})}{1-(1+f^{-1}(p_{ij})^{-T}}\right) \times \left(v_{i0} - D_i(\boldsymbol{x}, p_{ij})\right) & \text{, if } t \leq T, \\ 0 & \text{, otherwise.} \end{cases}$$

For simplicity of expression, the interest rate p is defined as the summation of future repayment for each dollar of the loan balance. The relationship between p and the contractual interest rate r follows $p = f(r) = \sum_{t=1}^{T} \frac{r}{1-(1+r)^{-T}}$.

4.1.2. Utility: Defaulting on a Mortgage

If a borrower defaults, she loses the mortgage collateral (*i.e.*, home ownership). The borrower is exempted from the future payments, but she no longer receives the utility gain from home ownership. In this case, the borrower gets the baseline utility each period in Equation 4.1. The post-default value function is written as

$$\underline{U}_i(\boldsymbol{x}, t) = \sum_{k=1}^{\Upsilon-t} \beta^k \underline{u}_i(\boldsymbol{x}, t) = \sum_{k=1}^{\Upsilon-t} \beta^k \frac{I_{ik}^{(1-\gamma_i)} - 1}{1-\gamma_i},$$

where β is the discount factor for each period.

4.1.3. Repayment Decision

A borrower makes a sequence of repayment decisions to maximize her intertemporal utility. In period t, it is optimal for a borrower to repay as long as:

$$d_i(\boldsymbol{x}, p_{ij}, t, \epsilon_{it}) = \begin{cases} 0 & \text{, if } \epsilon_{it} \geq \underline{U}_i(\boldsymbol{x}, t) - u_i(\boldsymbol{x}, p_{ij}, t) - \beta V_i(\boldsymbol{x}, p_{ij}, t). \\ 1 & \text{, otherwise.} \end{cases}$$

Default choice at period t is equal to 0 if the borrower repays the loan; otherwise, it is equal to 1. V is the borrower's value function, which is written as:

$$V_i(\boldsymbol{x}, p_{ij}, t) = \max\left\{u_i(\boldsymbol{x}, p_{ij}, t) + \beta \mathbb{E}_{\{\boldsymbol{v}, \boldsymbol{\epsilon}\}}\left[V_i(\boldsymbol{x}, p_{ij}, t+1) + \epsilon_{it}\right], \underline{U}_i(\boldsymbol{x}, t)\right\}.$$

The operator $\mathbb{E}$ is the expectation taken over future property value and repayment shocks. I assume the borrower's belief in future property value follows a random walk with drift, such that

$$\begin{aligned} \Delta v_t = v_t - v_{t-1} = g + \iota_t \\ \iota \sim N(0, \sigma_v^2). \end{aligned} \tag{2}$$

My model captures the major factors indicated to influence a borrower's repayment decisions, including borrower heterogeneity (Deng et al., 2000), changes in property values (Bajari et al., 2008), and loan size (Adams et al., 2009). Borrowers' optimal repayment decisions in my model depend on the following factors.

First, the model decisions must capture the importance of several vital loan characteristics used in the industry, such as income relative to monthly debt obligations and loan size relative to collateral value. Suppose a borrower's income is close to the mortgage payment or the loan's size is close to the property value; my model predicts the borrower will have a higher probability to default. Second, a borrower's repayment decision is modeled as a dynamic programming problem. When the loan balance is nearly paid off, the borrower's continuation value of repaying the loan is high and therefore likelihood of defaulting is low. This is consistent with what I observe in the data that the default probability decreases over time. Lastly, the default decision depends on changes in property value. A borrower is likely to foreclose on her house when the value of the property falls.

4.1.4. Down Payment Decision

When a borrower purchases a mortgage, the borrower chooses an optimal down payment D at loan origination. The optimal down payment is chosen to equalize the marginal utility of increased consumption at period zero and the marginal disutility of reduced consumption in the future due to larger repayments. The down payment is chosen by a borrower to satisfy the following optimization problem,

$$\max_{D_i} V_i(\boldsymbol{x}, p_{ij}, 0) - \delta_i D_i(\boldsymbol{x}, p_{ij}),$$

subject to $D_i > \underline{D}(v_{i0})$. $\underline{D}$ is the minimum down payment required, which usually is $3-5$ percent of property value. The associated first-order condition gives:

$$\sum_{t=1}^{T} \beta^{t-1} \mathbb{E}_{\{\epsilon\}} \left[d(\boldsymbol{x}, p_{ij}, t, \epsilon) \right] \times \left[\frac{\partial p_{ij}}{\partial D_i} \frac{C_{it}(\boldsymbol{x}, p_{ij}, t)^{(1-\gamma_i)}}{(1-\gamma_i)} + p_{ij} C_{it}(\boldsymbol{x}, p_{ij}, t)^{-\gamma_i} \right] - \delta_i = 0. \tag{3}$$

The down payment cost is modeled separately as a disutility source from reducing consumption. This is because the down payment is usually very large and comes from other sources such as saving rather than non-durable consumption. I simplify the saving process explicitly; δ captures the effect of paying D down payment up front relative to consumption in a reduced-form way. The optimal down payment choice can be written as a function of observable and price, $D^*_{ij} = D_i(\boldsymbol{x}, p_{ij})$.

Equation 3 describes the trade-off between current and future consumption. A borrower who values current consumption at a greater value relative to future consumption (e.g., high γ or low δ) chooses a lower down payment.

4.1.5. Purchase Decision

At loan origination, given a down payment size D and optimal repayment decisions d, a borrower i's expected utility of purchasing a mortgage from lender j is

$$U_{ij}(\boldsymbol{x}) + \varepsilon_{ij} = \alpha_j + V_i(\boldsymbol{x}, p_{ij}, 0) - \delta_i D_i(\boldsymbol{x}, p_{ij}) + \varepsilon_{ij},$$

where $\boldsymbol{\alpha}$ is the lender-specific preference. α_j is interpreted as the incremental utility that a borrower is willing to pay in order to choose a preferred lender. δ measures the cost of paying a down payment upfront. $\boldsymbol{\varepsilon}$ is the lender-purchase-specific unobserved component of utility for borrowers. The unobserved component is known to each borrower but not to lenders and researchers.

One could think of mortgage products as being nearly identical across lenders, but lenders are differentiated by agent, customer service, payment system, and brand loyalty. The lender-specific preference can rationalize a borrower's decision to purchase a mortgage with higher interest rates due to brand preference.

The borrower would choose to purchase from lender j who provides her with the highest utility

$$j^*_i = \max_j U_{ij}(\boldsymbol{x}, 0) + \varepsilon_{ij}.$$

If a borrower decides not to purchase a mortgage, I assume that she receives the same utility as the post-default value function. That is, a borrower would purchase the mortgage by comparing the utility of purchasing or not,

$$\max_j U_{ij}(\boldsymbol{x}) + \varepsilon_{ij} \geq \underline{U}_i(\boldsymbol{x}, 0) + \varepsilon_{i0}.$$

4.1.6. Selection on Unobserved Heterogeneity

Recall that λ is an individual-specific parameter that captures the unobserved heterogeneity in housing preference (risk types). Modeling unobserved heterogeneity is important for allowing potential advantageous selection in this market.

Conditional observed characteristics $\boldsymbol{x}$, borrowers differ in parameter λ and error terms on repayment shocks and lender preference. The probability that borrower i will purchase the mortgage is

$$\varphi(\boldsymbol{x}, p_{ij}, \lambda_i) = \int \Pr\left(\max_j U_{ij}(\boldsymbol{x}) + \varepsilon_{ij} \geq \underline{U}_i(\boldsymbol{x}, 0) + \varepsilon_{i0}\right) f(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon},$$

The likelihood of default for borrower with observed characteristics $\boldsymbol{x}$ (integrated over the unobserved heterogeneity) is therefore[31]

$$d(\boldsymbol{x}, p_{ij}, t) = \int d(\boldsymbol{x}, t, p_{ij}, \lambda_i) \frac{\varphi(\boldsymbol{x}, p_{ij}, \lambda_i) f(\lambda_i)}{\int \varphi(\boldsymbol{x}, p_{ij}, \lambda_i) f(\lambda_i) d\lambda_i} f(\lambda_i) d\lambda_i.$$

Supposing there is an exogenous decrease in interest rates, my model predicts that marginal borrowers (with lower λ) purchase the mortgages. The marginal borrowers tend to have higher default risks related to their lower housing preference. Therefore, the average default rate increases, with this increase depending on the degree of unobserved heterogeneity in λ. This extensive margin selection is critical in understanding the economic impacts of pricing and market expansion.

4.2. Mortgage Lenders

In the model, lenders are assumed to engage in Bertrand competition with differentiated products, meaning that they compete on price with horizontal product differentiation. I assume

[31] Optimal repayment policy also depends on repayment shocks. For simplicity of expression, I integrate the policy function over the shocks in the following equation: $d(\boldsymbol{x}, t, p_{ij}, \lambda_i) = \int d(\boldsymbol{x}, t, p_{ij}, \lambda_i, \boldsymbol{\epsilon}_i) f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon}$.

that loans are securitized through the GSEs.[32] For a given loan, a lender prices an interest rate at origination to maximize the expected profit. Given the demand system discussed previously, lender j's profit for a loan is given by

$$\pi_j(p;\boldsymbol{x}) = s_j(p;\boldsymbol{x})\left(p - c_j(\boldsymbol{x})\right) \times \left(v_{i0} - D_i(\boldsymbol{x}, p)\right),$$

where s is the market-share function, i.e., the probability that borrower i would purchase. I assume that the lender has knowledge of the distribution of $\boldsymbol{\lambda}$ and $\boldsymbol{\varepsilon}$, but they do not observe the λ_i and $\boldsymbol{\varepsilon_i}$ for each individual. Conditional on the same loan characteristics $\boldsymbol{x}$, each loan generates the same expected profit for a given lender. c denotes the marginal cost, coming from the guarantee fee paid to the GSEs when a lender sells the mortgage. The pricing on guarantee fees is what the GSEs have relied on to interpret differences in predicted risk across loans.[33]

4.2.1. *Interest Rate Pricing*

Lender j offers price p that maximizes the expected profit at loan origination. Specifically, the lender solves the following first-order condition with respect to price:

$$\begin{aligned}\frac{\partial \pi_j(p;\boldsymbol{x})}{\partial p} &= s_j(\boldsymbol{x}, p) \times \left(v_{i0} - D_i(\boldsymbol{x}, p) - \left(p - c_j(\boldsymbol{x})\right)\frac{\partial D_i(\boldsymbol{x}, p)}{\partial p}\right) \\ &+ \frac{\partial s_j(\boldsymbol{x}, p)}{\partial p}\left(p - c_j(\boldsymbol{x})\right) \times \left(v_{i0} - D_i(\boldsymbol{x}, p)\right) = 0. \end{aligned} \tag{4}$$

This condition reveals a lender's trade-off when choosing a higher price. A higher interest rate has a direct impact on profit by increasing the value of the loan. On the other hand, a higher interest rate results in a lower expected market share, as lenders are competing against prices.

[32] I assume no private market for lenders. That is, lenders do not have the choice of whether to securitize a loan or to which securitizers to sell.

[33] For simplicity, I assume that agreements between lenders and the GSEs are represented in a simple fee in ongoing monthly payments until the life of a loan expires. In reality, guarantee fee payments generally take the form of an ongoing monthly payment and an upfront payment. The upfront fee is a one-time payment made by lenders when a loan is acquired.

Solving Equation 4 yields

$$
\begin{aligned}
p_j(\boldsymbol{x}) &= c_j(\boldsymbol{x}) + \Delta_j(\boldsymbol{x}), \\
\Delta_j(\boldsymbol{x}) &= \left(\frac{s_j(\boldsymbol{x}, p)}{v_{i0} - D_i(\boldsymbol{x}, p)} \frac{\partial D_i(\boldsymbol{x}, p)}{\partial p} - \frac{\partial s_j(\boldsymbol{x}, p)}{\partial p} \right)^{-1} s_j(\boldsymbol{x}, p),
\end{aligned} \tag{5}
$$

where Δ denotes lender markup, which represents a fixed charge over the marginal cost for a lender with market power. The optimal interest rate in Equation 5 is written as additively separable in marginal cost c and markup Δ.

4.3. GSEs

To complete my model, I define a profit function of the GSEs in the secondary mortgage markets. When a lender j sells a loan with characteristics $\boldsymbol{x}$ to the GSEs, the GSEs' expected profit from purchasing the loan is given by

$$
\Pi_j(\boldsymbol{x}, p_{ij}) = \left(c_j(\boldsymbol{x}) - mc - UST10Y_t\right) \times \left(v_{i0} - D_i(\boldsymbol{x}, p_{ij})\right) - L(\boldsymbol{x}, p_{ij}), \tag{6}
$$

where $c_j(\boldsymbol{x})$ denotes the lender marginal cost. The marginal cost largely depends on the guarantee fee paid to the GSEs. mc is the time-invariant funding cost, which is the cost for the GSEs to fund each dollar of mortgage in the primary market. $UST10Y$ is the U.S. Treasury 10-year rate.[34] $L(\boldsymbol{x}, p)$ is the expected default cost of the loan given by

$$
L(\boldsymbol{x}, p_{ij}) = \sum_{t=1}^{T} \mathbb{E}_{\{\epsilon\}} \left[d(\boldsymbol{x}, p_{ij}, t, \epsilon) \right] \times \left[p_{ij} \left(v_{i0} - D_i(\boldsymbol{x}, p_{ij}) \right) - tm_i(\boldsymbol{x}, p_{ij}, t) - H(v_{it}) \right],
$$

where $H(.)$ is the net proceeds from foreclosing a property in a defaulted mortgage. The expenses and credits associated with foreclosure include the cash received from the sale of the property, incomes, and costs associated with holding the property post-foreclosure (e.g., rental income and title insurance costs), selling expenses (e.g., fees and commissions), and foreclo-

[34] The 10-year Treasury rate is a debt obligation issued by the United States government with a maturity of 10 years. This is viewed as an approximate solution for funding costs for the GSEs. Another measure is MBS yield, which is the required return for MBS that GSEs' investors demand. Both measures do not change results in a meaningful way. I choose the 10-year Treasury rate as the primary specification because the interest rates offered in many cases are below MBS yield, but not below the 10-year Treasury rate.

sure costs (e.g., fees associated with bankruptcy and foreclosure).

I consider a perfect competitive capital market for investors. The capital market has many sellers of investment products not limited to an MBS. The GSEs can only act as price-takers and earns an economic profit of zero (or less). All the profit goes to investors in the capital markets. This assumption has no meaningful implication on market efficiency because the funding cost is a transfer from the institution to investors.[35] Because the GSEs receive government subsidies, the GSE market sector can withstand even when the GSEs earn negative economic profits.

4.4. Discussion

I discuss here some important assumptions in the model that simplify a complex reality.

First, I assume that choice of the maturity is exogenous. I only consider the mortgage loans given 30-year maturity in my model. One major difficulty in modeling maturity choice is that it typically requires a fixed loan size because a borrower can choose a shorter maturity or higher down payment in exchange for a lower interest rate. To avoid adding another complication to the model, I endogenize down payment instead of maturity choice. I do this because down payment determining LTV is an important element in this market. Loans come with a 15-year amortization schedule that requires monthly repayments twice as large as those for loans with 30-year schedules (excluding the mortgage interest of approximately an additional 1,000-dollar monthly payment on average). It is reasonable to suppose that borrowers who chose a 30-year maturity would not take a 15-year maturity.[36]

Second, I take the choice of housing as given. It is possible that housing choice is not exogenous in reality. A lower interest rate may cause a borrower to purchase a more expensive property. It is also possible that a lower interest rate causes borrowers to choose a lower down payment (*i.e.*, the higher borrowing amount) due to less expensive future consumption. Because the two choices both determine the borrowing amount, allowing endogenous property value is the observational equivalent to allowing endogenous down payment. Without having

[35] Similarly, I assume no wedge between lender cost and GSE guarantee income. The wedge is also a transfer and does not create welfare provision. In reality, some of these costs are administrative and may not go entirely to the GSEs.

[36] 15-year maturity is the second most common type of loan, reasonably accounting for 21 percent of the sample. 10-year and 20-year maturities are also common, accounting for 3 percent and 4 percent of the data respectively.

information on prices and properties borrowers were considering, including endogenous property choices would require adding even more assumptions and structures to the model.

Third, saving is not taken into account in the model. This is because individual income trends and savings are not observed in the data. The life-cycle model of mortgage default, as in Campbell and Cocco (2015), requires additional assumptions and simplifications on unobserved initial wealth, tax rate, and saving returns. Instead, I use a flexible parameter λ to embed these unobserved factors in borrowers' default decisions. In the demand system, the main outcome of interest is the default risk of each loan. The predicted default risk as a function of observed characteristics is like a non-linear prediction. The prediction is not sensitive to the choice of allowing saving. Accordingly, the predicted market outcomes under different pricing regimes may not change much in a more sophisticated saving model, except for the measure of borrower surplus. If most borrowers have income growth through saving, my estimates for λ parameter could be biased upwards. This means that I estimate a lower bound of deadweight loss resulting from mortgage overprovision. The deadweight loss from GSE pricing and its interactions could be even larger if saving is taken into consideration. In addition, I conduct a robustness check to account for the potential bias estimates in the absence of individual-specific income trends. I show that the main takeaways from the counterfactual exercises remain robust. The quantitative changes in the welfare results are discussed in Appendix A.

Lastly, borrowers in my model do not have the option to prepay or refinance the loans. Although these are important elements to study in the mortgage markets, the focus of this paper is on borrowers' default and foreclosure, which bring direct loss to the GSEs. Prepayment and refinancing options could be related to a borrower's choice in terms of their purchase and default decisions, but the effects are indirect. Since I do not observe the refinancing loans or the reasons for prepayment when a loan is terminated, I abstract from these options of borrowers to simplify the model.

5. Identification

The primary identification concerns of this model are (i) to separate the housing preference, λ, from the risk aversion parameter, γ, and (ii) to trace out the distribution of unobserved heterogeneity λ. I discuss two key sources of variation below.

5.1. Time-Varying Housing Prices

The default probability predicted by the model is sensitive to a borrower's risk aversion, γ, and housing preference, λ. To separately identify the level of γ and λ, I leverage time-varying housing prices and their corresponding changes in default rates.[37] Since a borrower's utility of repaying in the model is a function of housing prices, the variations in housing prices over time help to identify the mean value of housing preference.

I assume that the variations in housing prices are orthogonal to unobserved repayment shocks. One might be concerned that the housing prices are correlated with local economic conditions. The correlation could result in overestimating housing preference. This implies that I estimate a lower bound of deadweight loss resulting from mortgage overprovision.

5.2. Exogenous Shock to Credit Supply

Given that mean level of λ and risk aversion parameter γ are identified, the parameter left to be identified is the degree of unobserved heterogeneity in λ. My model predicts that marginal borrowers purchase mortgages when interest rates decrease. Accordingly, average default rate across borrowers changes through selection on unobserved heterogeneity; if there is a larger degree of unobserved heterogeneity, average default rate increases more when interest rates decrease.

Suppose I observe an exogenous variation in interest rates. Given a distributional assumption on λ, the variance of λ is identified to explain how much average default rate changes with the decrease in interest rates.

Following this logic, I leverage a regulatory variation in credit supply caused by the pre-

[37] I examine the correlation between housing price index and *ex post* loan performance using the following specification:

$$D_{ht} = constant + \mathbb{I}_{\mathbb{h}\mathbb{t}} + \boldsymbol{x}_{ht} + \boldsymbol{\alpha}_t + \boldsymbol{\kappa}_h + \epsilon_{ht}, \tag{7}$$

where $\mathbb{I}_{\mathbb{h}\mathbb{t}}$ is the log of housing price index for market h at period t. D_{ht} is the fraction of loans that went into defaults for market h at period t. The coefficient on $\mathbb{I}_{\mathbb{h}\mathbb{t}}$ is the main parameter of interest. It measures the correlation between the housing price index and the *ex post* performance of a loan. I report the results from Equation 7 in Table 3. I find a negative correlation between the housing price index and the share of default loans with a coefficient of −0.18. The coefficient is significant at 99 percent confidence level.

emption of national banks from state laws against predatory lending in 2004. Although the preemption has the greatest impact in the subprime mortgage market, I also find that it led to a moderate credit expansion in the GSE market. The preemption exempts national banks from state anti-predatory lending laws. This regulatory change alleviates the constraints in subprime lending, leading to a market expansion in the subprime market. Indirectly, it lowers the (fixed) cost of credit in the GSE market and intensified competition for local lenders.

I find that the preemption decreases interest rates in markets with state APLs. Moreover, average interest rate decreases more in markets with a large share of national banks, whereas average default rate increases more in those markets. The variance of λ is identified through this variation.

To see how the preemption drives the variations in the data, I conduct a series of reduced-form analyses below.

5.2.1. *The Impact of the OCC Preemption on Credit Supply*

First, I visually inspect the average market share of national banks for loans purchased by the GSEs over time, separately for APL markets and non-APL markets in Figure 2. I also plot the average number of GSE loans over time, separately for APL markets and non-APL markets in Figure 3.

The difference of market share between the APL and non-APL markets is roughly constant before the OCC preemption. The average market share for APL markets increases and eventually surpasses non-APL markets after 2004. National banks significantly expand their market share in GSE loans following the preemption. I observe the same pattern for the number of GSE loans; lenders in APL markets originate more GSE loans relative to non-APL markets after the preemption. The figures confirms that GSE market is indeed significantly affected by the preemption.

5.2.2. *The Impact of the Preemption on Interest Rates*

To examine the impact of the preemption on interest rates, I employ a difference-in-differences research design. The following specification compares the loans originated in APL markets relative to the loans originated in non-APL markets:

$$\boldsymbol{Y}_{ht} = constant + APL_{ht} \times POST_t + \boldsymbol{x}_{ht} + \boldsymbol{\alpha}_t + \boldsymbol{\kappa}_h \times APL_{ht} + \epsilon_{sht}, \tag{8}$$

where Y_{sht} is the outcome variable originated in market h at year t, and $POST_t$ is an indicator of whether year t is after 2004. APL_{ht} is a time-varying indicator of whether the market h is being regulated by APL in the year t. $\boldsymbol{x}$ is a vector of individual loan characteristics including income, credit score, loan balance, down payment, mortgage insurance percentage, loan purpose (purchase, no cash-out refinance, and cash-out refinance), occupancy status (primary home, secondary home, and investor), and property type (single-family, condominium, co-op, planned unit development [PUD], and manufactured housing). $\boldsymbol{\alpha}_t$ is quarterly time fixed effects. I include market-APL fixed effects to control for variations that differ across markets but are constant before and after the APLs. The term APL_{ht} is absorbed by the fixed effects.

I report the results from using contractual interest rates as outcome variables in Equation 8 in column (1) of Table 4. The coefficient on $APL_{ht} \times POST_t$ measures the impact of the preemption on interest rates. I find the estimated coefficient is negative and significant (a coefficient of −2.03 percent); there is a small average decrease in interest rates for APL markets after the preemption.

I report the results from using an indicator of whether the loan defaulted during my sample period as outcome variables in Equation 8 in column (3) of Table 4. The estimated coefficient for default rates is not significant. However, it is important to address the potential heterogeneity between APL states and non-APL states (see Appendix B). Therefore, I further exploit that the preemption has differential impacts on markets with differing presences of national banks within APL states.

5.2.3. Triple Difference-in-Differences

To account for state-specific time trends, I take advantage of the heterogeneity in the presence of national banks in 2003.[38] The market share of national banks in both subprime and prime loans is a proxy of the treatment intensity on credit supply from the preemption. Di Maggio and

[38] Following Di Maggio et al. (2016) and Di Maggio and Kermani (2017), this measure is based on the fact that the share of national banks in 2003 is a good predictor of market concentration in subsequent years.

Kermani (2017) show that the preemption has stronger effects in a market with a stronger presence of national banks (i.e., with a higher share of national banks). The following specification compares the loans originated in APL markets with different shares of national banks:

$$\begin{aligned} \boldsymbol{Y}_{ht} = constant + POST_t \times Share_h + APL_{ht} \times POST_t \times Share_h \\ + \boldsymbol{x}_{ht} + \boldsymbol{\alpha}_t + \boldsymbol{\kappa}_h + \boldsymbol{\kappa}_h \times APL_{ht} + \boldsymbol{S}_{ht} + \epsilon_{sht}, \end{aligned} \tag{9}$$

where $Share$ denotes the market share of national banks in both subprime and prime markets in 2003. I include state-time fixed effects to control for state-specific trends. The term $APL_{ht} \times Share_h$ is absorbed by the fixed effects.

Column (2) of Table 4 reports the results from using contractual interest rates as outcome variables in Equation 9. The coefficient on $APL_{ht} \times POST_t \times Share_h$ measures the impact of the preemption on interest rates depending on the heterogeneity in presence of national banks. I find the estimated coefficient is negative and significant (a coefficient of −7.12 percent); interest rates on average decrease more in APL markets with a higher presence of national banks after the preemption.

I report the results from using an indicator of whether the loan defaulted during my sample period as outcome variables in Equation 9 in column (4) of Table 4. I find that the estimated coefficient is positive and significant (a coefficient of 3.76 percent); there is a larger average increase in default risks for markets with a higher share of national banks relative to markets with a lower share of national banks. My estimation exploits this variation to identify the degree of unobserved heterogeneity.

6. Estimation

In this section, I describe the empirical methods to estimate structural parameters in the model. I first estimate demand parameters by matching default probabilities. Given the estimated demand parameters, I next estimate other supply parameters. I start my discussion with the estimation on the borrower's side.

6.1. Demand Estimation

I consider a stationary income $I_{it} = I_i$. The variations in income are captured by repayment shocks $\boldsymbol{\epsilon}$. I assume that repayment shocks, $\boldsymbol{\epsilon}$, follow a normal distribution with mean 0 and standard deviation σ_ϵ. The borrower primitives in the model are risk aversion, γ; housing preference, λ; variance of repayment shocks, σ_ϵ; cost of upfront payment, δ; and lender preference, α.

λ is assumed to follow a log-normal distribution with mean, μ, and standard deviation, σ. I allow heterogeneity in μ, σ, and γ. The heterogeneity is modeled as a function of a vector of observed loan characteristics $\boldsymbol{x}$:

$$\mu(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta_\mu},$$

$$\sigma(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta_\sigma},$$

$$\gamma(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta_\gamma}.$$

$\boldsymbol{x}$ is a vector that includes a constant, the log of income, the log of credit score and the log of initial property value, and a set of dummies indicating home purchase and no-cash refinance. I collapse the model decisions to a quarterly level.[39] I set the quarterly discount factor β to 0.992 and the terminal age Υ to 60 years past loan origination.

To capture the fact that borrowers tend not to default in the first few years after loan origination, I add two dummies on the period utility in Equation 1 indicating whether a loan originated 2 years or 4 years ago.

6.1.1. Inner loop

Given a vector of primitives, $\boldsymbol{\Theta_0} = \{\boldsymbol{\beta_\mu}, \boldsymbol{\beta_\sigma}, \boldsymbol{\beta_\gamma}, \sigma_\epsilon\}$, for each loan i, I draw $N = 7$ sample points of λ_{in} from $LN(\mu(x), \sigma(x))$ using Hermite-Gauss quadrature nodes for $n = 1, 2, ..., N$.[40]

For each draw of λ_{in}, I solve for the optimal repayment policy, $d_i(\boldsymbol{\Theta_0}; \lambda_{in}, p_{ij}, t)$, and the value function, $V_i(\boldsymbol{\Theta_0}; \lambda_{in}, p_{ij}, t)$, for each period. In addition, I derive marginal cost of down payment, δ, from the first-order condition in Equation 3. I also solve for lender preference, $\boldsymbol{\alpha}$,

[39] Specifically, t is a counter on the quarter level. A borrower's period utility depends on consumption in a quarter level, and an optimal repayment decision is whether to repay or not in each quarter.

[40] The objective function value is quite stable for choices of nodes $N \geq 7$.

using observed market shares in the data.[41] I normalize the preference for small lenders[42] to zero, so the lender preference measures the incremental utility of each lender relative to the small lenders. I discuss the estimation procedure in greater detail in Appendix C.

Finally, to obtain empirical moments, average default rate over the draws of λ_{in} predicted by the model is approximated as

$$\begin{aligned} d_i(\boldsymbol{\Theta_0}; \boldsymbol{x}, \boldsymbol{p_i}, t) &= \int d_i(\boldsymbol{\Theta_0}; \lambda_{in}, \boldsymbol{x}, \boldsymbol{p_i}, t) \rho(\boldsymbol{\Theta_0}; \lambda_{in}, \boldsymbol{x}, \boldsymbol{p_i}) d\lambda_{in} \\ &\approx \sum_{n=1}^{N} \omega_n d_i(\boldsymbol{\Theta_0}; \lambda_{in}, \boldsymbol{x}, \boldsymbol{p_i}, t) \rho(\boldsymbol{\Theta_0}; \lambda_{in}, \boldsymbol{x}, \boldsymbol{p_i}), \end{aligned} \tag{10}$$

where $\rho_i(\boldsymbol{\Theta_0}; \lambda_{in}, \boldsymbol{x}, \boldsymbol{p_i})$ is the probability distribution of λ among the borrowers who purchase mortgages (see Appendix C), ω is the Hermite-Gauss quadrature weight associated with each node n.[43]

6.1.2. Outer Loop

I adopt method of moments estimator by minimizing the differences between predicted default probabilities (from Equation 10), $\boldsymbol{d}$, and observed default probabilities, $\hat{\boldsymbol{d}}$. The moment restrictions used in the estimation are

$$\boldsymbol{g}(\boldsymbol{\Theta_0}) = \mathbb{E}_{\{i\}} \left[d_i(\boldsymbol{\Theta_0}; \boldsymbol{p_i}; t) - \hat{d}_{it} | \boldsymbol{z_{it}} \right] = \boldsymbol{0}, \forall t = 1, , \ldots, T.$$

$\boldsymbol{z_{it}}$ is a covariate vector given by

$$\boldsymbol{z_{it}} = \begin{bmatrix} \text{APL}_i \times \text{POST}_i \times \text{HIGH}_i \\ \text{APL}_i \times \text{POST}_i \times (1 - \text{HIGH}_i) \\ 1 - \text{APL}_i \times \text{POST}_i \end{bmatrix}^{\intercal} \otimes \begin{bmatrix} \boldsymbol{x_i}^{\intercal} \\ v_{it} \cdot \boldsymbol{x_i}^{\intercal} \end{bmatrix},$$

where POST is an indicator of whether the loan originated before or after 2004, APL is an indica-

[41] I follow the assumption made in Scharfstein and Sunderam (2014). They define markets locally at the geographic level. Market share and brand preference are measured at 3-digit zip code by year level.

[42] I define small lenders to be a lender with market share no more than 1 percent.

[43] I approximate the expectation over future housing prices in the same way. I first estimate g and σ_v^2 in Equation 2 for each market. I obtain the expected value function by integrating over three draws of housing prices from a normal distribution with standard deviation σ_v.

tor of whether the loan originated in markets with local APLs. HIGH is an indicator of whether the market share of national banks in 2003 was above its median. v_{it} denotes the property value of borrower i at period t.[44]

In practice, I implement two-step generalized method of moments and match quarterly default rates up to 30 quarters past origination. In the first step, I estimate the model with identity weighting matrix and obtain an estimator $\hat{\boldsymbol{\Theta}}_1$. Using this estimator, I obtain the optimal weighting matrix, $\hat{W}(\hat{\boldsymbol{\Theta}}_{\mathbf{1}})$, and solve for $\hat{\boldsymbol{\Theta}}$ in the outer loop:

$$\hat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}}{\operatorname{argmin}}\, \bar{\boldsymbol{g}}(\boldsymbol{\Theta})^\top \hat{W}(\hat{\boldsymbol{\Theta}}_{\mathbf{1}})\bar{\boldsymbol{g}}(\boldsymbol{\Theta}).$$

In total, I have 36 covariate vectors and $36 \times 30 = 1,080$ conditional moment restrictions to estimate 21 parameters. I summarize the moment restrictions in Appendix D.

6.2. *Cost Estimation*

Given the estimated demand parameters $\hat{\boldsymbol{\Theta}}$, I now estimate marginal cost of lenders. I parameterize the marginal cost into a vector of $\boldsymbol{x}$, market-lender fixed effect, origination quarter-year fixed effects, and an unobserved component, ν, so that

$$c_{ij}(\boldsymbol{\vartheta}) = \boldsymbol{x}^\top \boldsymbol{\beta_c} + \beta_{APL} APL_i + \beta_{OCC} OCC_i \times APL_i \times POST_i + \boldsymbol{\eta}_{hj} + \boldsymbol{\kappa}_t + \nu_i,$$

where $\boldsymbol{\vartheta} = \{\boldsymbol{\beta_c}, \boldsymbol{\eta}, \beta_{APL}, \beta_{OCC}, \boldsymbol{\kappa}\}$ denotes a vector of parameters to be estimated. OCC is an indicator of whether the lender is regulated by the OCC (i.e., a national lender). The parameter $OCC_i \times APL_i \times POST_i$ captures the impact of preemption on costs for national lenders.

I derive lender markup $\Delta_j(\hat{\boldsymbol{\Theta}}; \boldsymbol{x})$ for each loan according to Equation 5. The pricing equation gives

$$\nu_i(\boldsymbol{\vartheta_0}) = p_{ij} - \Delta_j(\hat{\boldsymbol{\Theta}}; \boldsymbol{x}) - \boldsymbol{x}^\top \boldsymbol{\beta_c} + \boldsymbol{\eta}_{hj} + \boldsymbol{\kappa}_t.$$

[44] The future property value in period t is obtained using the house price index $\mathbb{I}$ from the Federal Housing Finance Agency as follows, $v_t = v_0 \times \frac{\mathbb{I}_t}{\mathbb{I}_0}$.

I obtain cost estimates using linear least squares, as in

$$\hat{\boldsymbol{\vartheta}} = \underset{\boldsymbol{\vartheta}}{\operatorname{argmin}}\, \boldsymbol{\nu}(\boldsymbol{\vartheta})^{\top} \boldsymbol{\nu}(\boldsymbol{\vartheta}).$$

Finally, I assume that the GSEs earn an economic profit of zero (a perfect competitive capital market) during the 2000 – 01 period. I do not assume the zero profit condition in the later 2002 – 08 period, because evidence suggests that the GSEs mispriced (underpriced) most of their loans during this period (Federal Housing Finance Agency, 2009; Lucas and McDonald, 2010). Another interpretation of this assumption is that I examine GSE pricing during the 2002 – 08 period using the 2000 – 01 period as an optimal benchmark.

Using credit loss information in the data, I estimate expected net proceeds H from a sale of the property post-foreclosure. The average net proceeds of a property are approximately 33.90 percent of its initial value.

In Equation 6, $UST10Y_t$ is a known index, mc is estimated by equating the average of funding costs to the average of estimated marginal costs, $c_{ij}(\hat{\boldsymbol{\vartheta}})$ during the first two years. Under the assumption that mc is time-invariant, I calculate the funding costs in the later periods using the estimated mc.

7. Results

In this section, I first discuss my parameter estimates. I then provide several pieces of evidence to evaluate the model fit. I report the parameter estimates in Table 5. Standard errors are calculated using the delta method.

7.1. Parameter Estimates: Demand

The main parameter of interest is the distribution of housing preference, λ, which is specified as a log-normal distribution with mean μ and standard deviation σ. Both are specified as a function of observed loan characteristics. The average values of μ and σ across loans in the sample are approximately 1.48 and 0.08, respectively. My model estimates 25.17 percent of potential borrowers (i.e., those who did not purchase mortgages under GSE pricing). The estimated standard deviation of repayment shocks, σ_ϵ, is 0.99.

The baseline estimate of μ is -1.09. The estimated coefficient on the log of credit score is especially large compared with other coefficients, suggesting that credit score may be a meaningful characteristic in determining borrowers' default risks. A 1 percent increase in credit score increases the mean of λ by 0.51. I estimate a 0.37 increase in standard deviation of λ when credit score increases by 100 (13.87 percent of the credit score). The coefficient on the log of income is relatively smaller. I estimate a 0.06 increase in standard deviation of λ when income increases by $1,000$ dollars (28.8 percent of the average income).

Home purchase loans and no cash-out refinance loans also relate to μ positively with coefficients of 0.10 and 0.06. The estimated coefficient on the log of property value is negative (a coefficient of -0.09). This is consistent with my reduced-form evidence that a loan with higher property value (i.e., a higher loan balance) is more likely to default.

The σ parameter shows the degree of unobserved heterogeneity. The baseline estimate of σ is 0.24. The estimated coefficient on unobserved heterogeneity is negative for credit score and income, suggesting a lower degree of unobserved heterogeneity for borrowers with higher credit score and income. On the other hand, the coefficient on the log of property value is positive (a coefficient of 0.02). My model allows for heterogeneity in risk aversion parameter γ, which is modeled as a function of observed loan characteristics. The average value of γ across loans in the sample is about 1.16. The baseline estimate of risk aversion parameter, γ, is 1.01. The coefficient on the log of property value is 0.09. Similarly, the coefficient on the log of credit score is 0.01; a borrower with higher property value and higher credit score is less sensitive to interest rates. On the contrary, the coefficients on income and no cash-out refinancing are negative (coefficients of -0.06 percent and -0.03, respectively).

Cost of up-front payment, δ, and lender preference, α, are estimated non-parametrically. I plot the distributions of these two variables in Figure 4. The average of δ and α are 0.49 basis points and -0.58, respectively. To clarify the meaning of these numbers, borrowers paying an additional $10,000$ dollars of down payment would decrease their utilities by 0.49 on average. Choosing any mortgage lender would decrease borrowers' utilities by -0.58 on average relative to small lenders due to brand preference.[45]

7.2. Parameter Estimates: Cost

Parameter c represents lender marginal cost for originating an additional 1 dollar of a mortgage loan. The marginal cost also represents guarantee fee income for the GSEs in the model. The mean of marginal cost is roughly 1.37 dollars. The baseline estimate of c is 1.33. The marginal cost decreases with the log of credit score, the log of income, and the log of property value, each with coefficients of -0.11, -0.01, and -0.04, respectively. I also include the log of monthly required repayment. The estimated coefficient on the log of monthly required repayment is 0.04.

The cost estimates reveal that lender marginal cost largely depends on the loan characteristics related to expected default probabilities. This is consistent with the risk-based features of GSE pricing; guarantee fee increases with default risks.

In addition, the parameters on APL_h and $OCC_j \times APL_h \times POST_t$ are meant to capture the increased and reduced fixed costs from implementing and preempting the APLs. The estimated coefficients are 0.25 percent and -0.46 percent, respectively. These numbers translate to changes in contractual interest rates by 1.00 and -1.84 basis points. Finally, the average of GSE funding costs (i.e., the cost for funding an additional 1 dollar of a mortgage loan) is around 1.37.

7.3. Model Fit

In Figure 5 I show fitting errors for the moments used in GMM estimation. Most of the moments are fitted with negligible errors. My empirical moments can be explained reasonably well by the model.

In order to further assess the model fit, I compare a simulated default path with the observed default path from the data in the upper panel of Figure 6. They align reasonably well. My model is able to closely trace the downward-sloping trend for default probabilities over time.

I also perform an out-of-sample validation. I estimate the model using 50 percent of the

[45]In the data large lenders usually offer lower interests compared with small lenders, so that the parameter α rationalizes the behavior of borrowers who choose small lenders because of brand preference that offsets the utility loss from higher interest rates.

loans (random selection) and use the estimated parameters to predict default probabilities for the remaining 50 percent of the loans. I compare the out-of-sample prediction with the corresponding actual default path in the lower panel of Figure 6. The out-of-sample prediction is not as accurate as the in-sample prediction, but overall my model has a satisfactory prediction on default probabilities for the sample that is not used in the estimation.

8. Implications for Pricing, Competition Policy, and Market Design

In this section, I use the estimated model to analyze the impact of different insurance pricing schemes and a competition policy that changes market concentration. I also examine how these results could depend on information asymmetry between lenders and borrowers.

For a given market equilibrium, $\boldsymbol{p}$,[46] I calculate borrower surplus, CS, lender profit, LS, and GSE profit, GS, as

$$\begin{aligned}
CS_i(\boldsymbol{p}) &= \int \left(V_i(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p},0) - \underline{U}_i(\hat{\boldsymbol{\Theta}})\right)\varphi(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},0)\,d\lambda_{in},\\
LS_i(\boldsymbol{p}) &= \int \left(\boldsymbol{p} - c_i\right)\left(v_0 - D_i(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p})\right)\varphi(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p})\,d\lambda_{in},\\
GS_i(\boldsymbol{p}) &= \int (c_i - mc)\times\left(v_0 - D_i(\hat{\boldsymbol{\Theta}};\boldsymbol{x},\boldsymbol{p}) - L_i(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p})\right)\varphi(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p})\,d\lambda_{in},
\end{aligned}$$

where $L_i(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p})$, the loss from loan defaults, follows

$$L_i(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p}) = \sum_{t=1}^{T}\mathbb{E}_{\{\epsilon\}}\left[d_i(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p},t,\epsilon)\right]\times\left(v_{i0} - D_i(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},\boldsymbol{p}) - tm_i(\boldsymbol{x},\boldsymbol{p},t)\right).$$

I calculate borrower welfare using a certainty equivalent approach, which equates a borrower's value function to the utility of consuming CE dollars each period. Given this, social surplus in dollar value, TS, equals the summation in surplus in dollar value of the three market participants.

$$TS(\boldsymbol{p}) = \sum_i\left[LS_i(\boldsymbol{p}) + GS_i(\boldsymbol{p}) + \sum_{t=1}^{\Upsilon}\beta^{t-1}CE_i(\boldsymbol{p})\right].$$

I discuss the results of my counterfactual simulations in detail below.

[46] For each market, I calculate the lenders' optimal pricing by iterating their first-order conditions (Equation 5) associated with the counterfactual policy and competitors' pricing. I stop at iteration n when the lenders' pricing in this iteration, $\boldsymbol{p^n}$, satisfies the following condition: $|\boldsymbol{p}^n - \boldsymbol{p}^{n-1}| < \mathbf{0.1}$ percent.

8.1. *Mortgage Insurance Pricing: The Impact of Mispricing*

To quantify the impact of mispricing from the GSEs. I examine the counterfactuals of two alternative price schemes: (i) a full risk-based pricing scheme that charges exact expected cost and (ii) a uniform pricing scheme that charges a flat rate to every loan. To make a fair comparison, I set the uniform price to be the average of GSE prices.

Under full risk-based pricing, the GSEs charge guarantee fees that fully reflect the expected costs and default risks. Specifically, lenders' pricing equations are captured as

$$p_j^c(\boldsymbol{x}) = mc + UST10Y_t + \boldsymbol{\Delta}_j(\hat{\boldsymbol{\Theta}};\boldsymbol{x})(1+L_i(\hat{\boldsymbol{\Theta}};\boldsymbol{x},p_{ij})),$$
$$L_i(\hat{\boldsymbol{\Theta}};\boldsymbol{x},p_{ij}) = \int L_i(\hat{\boldsymbol{\Theta}};\lambda_{in},\boldsymbol{x},p_{ij})d\lambda_{in}, \tag{11}$$

where $L_i(\hat{\boldsymbol{\Theta}};\boldsymbol{x},p_{ij})$ denotes the expected default cost of borrower i.

8.1.1. *Evidence on Cross-Subsidization*

Using full risk-based pricing as a benchmark, I find that GSE pricing involves a substantial cross-subsidization. Figure 7 compares borrowers' average prices under different pricing schemes (on the y-axis) with their average full risk-based prices (on the x-axis). For lower-risk borrowers (i.e., those with lower full risk-based prices), their GSE prices are higher than their full risk-based prices. On the other hand, for medium-risk and high-risk borrowers (i.e., those with higher full risk-based prices), their GSE prices are lower than their full risk-based prices.

I estimate that the average mortgage subsidy received per loan is about 2,235 dollars.[47] In addition, GSE pricing redistributes 56.82 billion dollars from lower-risk borrowers to higher-risk borrowers; the number translates to approximately 1.60 percent of the mortgage interest.

Furthermore, I compare mortgage subsidies across markets. I plot the mean deviation of mortgage subsidy for each market in the upper panel of Figure 8. The mean deviation of mortgage subsidy is negatively correlated with the average income in each market (see the lower panel of Figure 8). Borrowers from a lower-income region on average benefit more from GSE pricing. This shows another piece of evidence on the cross-subsidization from GSE pricing.

[47] Mortgage subsidy is defined as the change in consumer welfare in dollar value from full risk-based pricing to GSE pricing.

As for the case of uniform pricing, borrowers all get the same uniform price regardless of their risks and costs. Uniform pricing is an extreme case of cross-subsidization between lower-risk borrowers and higher-risk borrowers. Higher-risk borrowers receive an even lower price relative to GSE pricing.

8.1.2. Uniform Pricing

I first consider the effects of uniform pricing (column (1) of Table 6). I find that, under uniform pricing, default cost increases by 43.41 dollars (3.30 percent of the average default cost). Lender profit increases by 413.08 dollars but GSE profit decreases by 14.63 dollars. Overall, deadweight loss per loan increases by 33.89 dollars relative to GSE pricing (1.13 percent of the average mortgage subsidy).

Uniform pricing increases the degree of cross-subsidization from lower-risk borrowers to higher-risk borrowers. However, uniform pricing exacerbates the problem of overprovision to higher-risk borrowers, as higher-risk borrowers face a lower price under uniform pricing relative to GSE pricing. This finding suggests that an increase in the cross-subsidization of insurance pricing could result in a sizable distortionary cost.

8.1.3. Full Risk-Based Pricing

Next, the welfare results of full risk-based pricing are reported in column (2) of Table 6. Full risk-based pricing does not involve cross-subsidization and underpricing. Equation 11 shows that the interest rate is always higher than or equal to the cost of each loan under full risk-based pricing.

I find that the average interest rate is 6.50 percent (25.98 basis points in contractual interest rates) higher under full risk-based pricing than under GSE pricing. This implies that the GSEs underprice most of their mortgage loans. The estimated magnitude of underpricing is in line with the results from Lucas and McDonald (2010); they suggest that the GSEs underpriced contractual interest rates by around 20 to 30 basis points in 2005.

Under full risk-based pricing, default cost decreases significantly by 843.22 dollars (64.01 percent of the average default cost). GSE profit increases by 2,199.33 dollars. Borrowers bear an average decrease in their surplus due to the higher full risk-based prices. Overall, welfare improves substantially when the mispricing is eliminated; deadweight loss per loan decreases by 519.25 dollars (17.38 percent of the average mortgage subsidy) relative to GSE pricing.

8.2. The Impact of Competition Policy under GSE Pricing

I evaluate the impact of a competition policy that potentially reduces entry costs of new entrants and decreases market concentration in each market. Some such policies might include things like: removing entry barriers, liberalizing product restrictions, and so on. I simply assume that such policies increase the number of entrants in each market. This counterfactual analysis is intended to provide a generalizable result that the welfare impacts of GSE pricing could interact with market structure.

To simulate a market with different concentration levels, given the same distribution of marginal cost, I scale market shares of entrants in each market by a fraction of the existing firms. Market concentration in each market is decreasing with N. For example, N is the current market structure, which has 5 lenders of 20 percent market share each, the counterfactual of $2N$ means that the counterfactual market has 5 lenders and 5 new entrants of 10 percent market share each (a 50 percent decrease in market concentration). The counterfactual of $1.5N$ means that the counterfactual market has 5 lenders of 13.33 percent market share each, and 5 new entrants of 6.67 percent market share each (i.e., market share is 0.5 of each existing firm). $0.75N$ means that 4 lenders have 15 percent market share each (i.e., market share is 0.75 of each existing firm) and 1 lender has the rest of market share of 40 percent. Perfect competition is when $N \to \infty$.

In Table 7, I report the results for various instances of market structure from $0.75N$ to $2N$, and perfect competition. I compare surplus of each market participant to the baseline, which is the case of N under GSE pricing.

For the case of $2N$, my model predicts a decrease in interest rates by 4.07 percent as lender markup decreases in a less concentrated market. Lender profit per loan decreases by $7,663.55$ dollars. Borrower surplus increases by $7,859.92$ dollars because of lower interest rates. The decrease in market concentration leads to an additional default cost per loan of 186.35 dollars (14.15 percent of the average default cost) and an additional deadweight loss of 171.77 dollars (5.75 percent of the average mortgage subsidy).

For the less competitive case of $0.75N$, my model predicts an increase in interest rates by 2.32 percent as lender markup increases. Default cost decreases by 106.98 dollars per loan, or 8.12 percent of the average default cost. The surplus goes from borrowers to lenders, as borrower surplus per loan becomes $4,365.54$ dollars lower. Overall, deadweight loss decreases by 17.78 per loan relative to the baseline.

For the cases of $1.25N$ and $1.5N$, the results are qualitatively similar. Deadweight loss is the largest in a perfectly competitive market when the market bears 544.67 dollars welfare loss per loan, or 18.23 percent of the average mortgage subsidy.

In Figure 9, I plot changes in deadweight loss as a function of market concentration under GSE pricing. I find that deadweight loss is lower than the baseline between $0.65N$ and N. A moderate increase in market concentration is socially beneficial as default cost goes down. The increased market power helps to correct some of the underpricing. However, welfare decreases substantially when the market becomes much more concentrated. Excessive market power is not socially desirable, as the inefficiency from increased market power outweighs the underpricing problem. Additionally, a decrease in market concentration exacerbates the underpricing problem, so deadweight loss increases as market concentration decreases.

8.3. The Impact of Competition Policy under Full Risk-Based Pricing

I examine the same competition policies discussed in Section 8.2 under full risk-based pricing. In Table 8 I present the results of this counterfactual for various instances of market structure from $0.75N$ to $2N$, and perfect competition.

Interestingly, I find different welfare effects of competition under GSE pricing and full risk-based pricing. Under full risk-based pricing, deadweight loss decreases as the market becomes less concentrated. A higher market power is less desirable, as lenders charge a higher interest rate that leads to a lower mortgage demand. Competition is always socially beneficial and provides more loans to borrowers whose willingness to pay is higher than their costs.

For the case of $2N$, borrower surplus increases, partially offsetting the decreased lender profit from lower market concentration. The lower market concentration decreases deadweight loss by 844.29 dollars (28.27 percent of the average mortgage subsidy) relative to the baseline. For the cases of $1.25N$ and $1.5N$, the results are qualitatively similar. Welfare gain is the largest in a perfectly competitive market; deadweight loss decreases by 965.42 dollars per loan (32.32 percent of the average mortgage subsidy) relative to the baseline.

8.4. Financial Innovation: Removing Information Asymmetry

Another potential inefficiency is that unobserved heterogeneity in λ_i is private information known to each borrower but not to lenders and the GSEs. Financial technology (fintech) advances may help lenders to learn borrowers' private information and engage in personalized pricing, which can be seen as perfect price discrimination on borrowers' unobserved risks.[48]

I examine the same competition policies discussed in Section 8.2 under a counterfactual that lenders observe the λ_i for individual borrowers. Lenders are able to price discriminate against λ_i. I assume that GSE pricing does not depend on the information of individual unobserved risk types (i.e., same as the case of asymmetric information). I also consider the case when the GSEs charge guarantee fees according to borrowers' unobserved risks (i.e., symmetric information with full risk-based pricing) later in this section.

In Table 9, I find that symmetric information under GSE pricing leads to a higher default cost and a higher welfare loss relative to the baseline. For the case of N, because lenders are able to extract more borrower surplus, their profits increase by 121.84 dollars. Default cost increases by 5.48 dollars on average. In total, deadweight loss increases by 7.36 dollars per loan.

In general, price discrimination increases market efficiency by allowing the market to capture more consumer surplus. However, removing information asymmetry under a sub-optimal pricing may actually reduce efficiency when it further distorts price. This is exactly what I find in this counterfactual exercise. Profit-maximizing lenders charge lower prices on borrowers who are private low types (i.e., with lower λ) since they are more price sensitive. At the same time, lenders raise the prices for borrowers who are private high types (i.e., with higher λ). This worsens the overprovision of mortgages to higher-risk borrowers and the underprovision of mortgages to lower-risk borrowers.

[48] One may think financial technology also helps lenders to learn a borrower's unobserved lender preference, α_i. My simulation predicts that prices will be close to the marginal costs when lenders are allowed to perfectly price discriminate on both unobserved risks and lender preference (see theoretical predictions in Stole (2007)). The market structure changes substantially; lender profit is close to zero and borrower surplus increases substantially. To study the equilibrium close to the current market structure, I still allow unobserved lender preference in this counterfactual exercise.

Under symmetric information, lender market power (price discrimination) leads to more price distortion, so the deadweight loss resulting from symmetric information decreases as market concentration decreases. In the extreme case of perfect competition, when price equals marginal cost, symmetric information achieves the same market outcomes as the asymmetric information.

Additionally, I investigate a counterfactual of symmetric information combined with full risk-based pricing (Table 10). The changes in welfare are reversed directionally as opposed to GSE pricing. Welfare increases as market concentration decreases under full risk-based pricing.

For the case of N, deadweight loss decreases by 651.35 dollars, or 21.81 percent of the average mortgage subsidy. Relative to the case of asymmetric information under full risk-based pricing, default cost is 30.25 dollars lower and deadweight loss is 132.10 dollars lower. Welfare is maximized in a perfectly competitive market. When inefficiencies resulting from both pricing and information asymmetry are removed, perfect competition achieves the first-best outcome.

8.5. *Differential Effects*

In addition to studying the effects on efficiency, I examine differential effects across different groups of borrowers. I split borrowers into subgroups based on the following three categorizations: (i) an indicator of whether or not the borrower's income is greater than the 50th percentile; (ii) an indicator of whether or not the borrower's credit score is greater than the 50th percentile; and (iii) an indicator of whether or not the λ_i is greater than the 50th percentile. I focus on changes in borrower surplus to study the distributional implications.

The differential effects under GSE pricing are reported in column (2) of Table 11. The lower-risk groups (i.e., with higher income, higher credit score, and higher λ) benefit more from lower market concentration. Since lower-risk borrowers are more likely to purchase a mortgage, they gain more from lower interest rates through more competition.

The differential effects of full risk-based pricing are reported in column (3) – (4) of Table 11. Lenders raise interest rates under full risk-based pricing, so most borrowers are worse off, especially for lower-risk borrowers who are more likely to purchase mortgages.

The differential effects of symmetric information are presented in Table 12. I also plot average prices separately for private high types and private low types under different pricing regimes and information environment in Figure 10.

Under GSE pricing with symmetric information, Lenders will price on borrower willingness to pay. Private low types with lower willingness to pay are offered lower prices than the case of asymmetric information. Therefore, private low types are better off from the lower prices, and private high types are worse off from the higher prices.

Alternatively, symmetric information under full risk-based pricing hurts private high types. Asymmetric information with full risk-based pricing implies that unobserved riskier borrowers are pooling with others who have lower default risks. When full risk-based pricing takes account for unobserved default risks under symmetric information, lenders charge private low types higher prices to cover their higher guarantee fees. As a result, private high types are worse off from the higher full risk-based prices.

8.6. *Equilibrium Prices*

I compare equilibrium interest rates of counterfactual policies to GSE pricing in Figure 11. The upper left panel shows the distribution of interest rates under full risk-based pricing. Full risk-based pricing corrects the underpricing problem, so most interest rates become higher under full risk-based pricing than GSE pricing.

The upper right panel of Figure 11 shows the distribution of interest rates in the counterfactual that removes information asymmetry between lenders and borrowers under GSE pricing. Lenders charge higher interest rates to private high types who value mortgage more and charge lower interest rates to others to increase sales. Accordingly, most of interest rates shift lower and the rest of interest rates shift higher relative to the case of asymmetric information.

The lower left panel shows the distribution of interest rates with symmetric information under full risk-based pricing. Lenders charge higher interest rates for those who are likely to default (i.e., private low types) and charge lower interest rates to other borrowers. As a result, interest rates become more dispersed. Many interest rates shift much higher to account for higher unobserved risks.

9. Conclusion

Although the GSEs may have a positive influence in the mortgage market, their mispricing could give rise to market inefficiency. There are diverging opinions among policymakers and academics on GSE reforms, to some extent because the proposed reform scenarios have not been implemented on a large enough scale to provide conclusive answers. On the other hand, market structure is becoming more and more important in mortgage lending. Though a few recent papers point out the effects of competition among lenders in the mortgage market, the interaction effects of competition and mispricing in the GSE market sector are unexplored.

This paper provides a framework to analyze the welfare impacts of insurance pricing and market structure. I estimate an industry model of borrowers, lenders, and GSEs. My estimation leverages a natural experiment that gives cost advantages for certain lenders. The reduced-form analyses show that the regulatory changes had significant impacts on average interest rate and average default rate in the GSE market.

I use the estimated model to study two counterfactual pricing schemes for mortgage insurance. I find that GSE pricing suffers substantial welfare loss from mispricing (underpricing). An alternative full risk-based pricing reduces welfare loss. GSE pricing also implies a redistribution from lower-risk borrowers to higher-risk borrowers. A uniform pricing scheme that increases the degree of cross-subsidization leads to a larger welfare loss relative to GSE pricing.

I study how inefficiencies under GSE pricing interact with market structure. Under GSE pricing, a decrease in market concentration leads to a higher default cost to the GSEs and a higher welfare loss to the society. On the other hand, under full risk-based pricing, welfare increases as market concentration decreases. It is important to note that my analysis does not account for some potential factors, such as the competition effects in other market sectors and the equilibrium effects on housing prices in the real estate markets. These factors are difficult to quantify. However, they are important for policymakers to consider when determining a competition policy.

The welfare impacts of insurance pricing and market structure also depend on the degree of asymmetric information on borrower risks. Under GSE pricing, symmetric information on borrowers' private risk types increases deadweight loss, whereas under full risk-based pricing, it reduces deadweight loss substantially.

The empirical findings in this paper highlight the importance of considering the welfare impacts of mortgage insurance pricing and its interaction with market structure and information asymmetry. In this market sector, competition policy and financial innovation could have very different welfare impacts under different pricing regimes.

This study focuses on a specific consumer lending market, but the policy takeaways could apply broadly to other risk-sharing markets where the government plays a role through regulations, subsidies, tax policy, etc. For instance, the pricing for the federal health insurance program could have impacts on medical spending. The impacts could interact with the competition among health care providers. It will be intriguing to study other markets that share similar features and to extend the framework to study the private securitization markets.

References

Adams, W., Einav, L., and Levin, J. (2009). Liquidity Constraints and Imperfect Information in Subprime Lending. *American Economic Review*, 99(1):49–84.

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., and Evanoff, D. D. (2011). The Role of Securitization in Mortgage Renegotiation. *Journal of Financial Economics*, 102(3):559–578.

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., and Evanoff, D. D. (2014). Predatory Lending and the Subprime Crisis. *Journal of Financial Economics*, 113(1):29–52.

Agarwal, S., Amromin, G., Chomsisengphet, S., Landvoigt, T., Piskorski, T., Seru, A., and Yao, V. (2015). Mortgage Refinancing, Consumer Spending, and Competition: Evidence from the Home Affordable Refinancing Program.

Agarwal, S., Chang, Y., and Yavas, A. (2012). Adverse selection in mortgage securitization. *Journal of Financial Economics*, 105(3):640–660.

Agarwal, S., Song, C., and Yao, V. W. (2017). Banking Competition and Shrouded Attributes: Evidence from the US Mortgage Market.

Allen, F. and Gale, D. (2004). Competition and Financial Stability. *Journal of Money, Credit, and Banking*, 36(3):453–480.

Ambrose, B. W. and Thibodeau, T. G. (2004). Have the gse affordable housing goals increased the supply of mortgage credit? *Regional Science and Urban Economics*, 34(3):263–273.

Bachas, N. (2017). The Impact of Risk-Based Pricing in the Student Loan Market: Evidence from Borrower Repayment Decisions.

Bajari, P., Chu, C. S., and Park, M. (2008). An Empirical Model of Subprime Mortgage Default from 2000 to 2007.

Benetton, M. (2017). Leverage Regulation and Market Dtructure: An Rmpirical Model of the UK Mortgage Market.

Bhutta, N. (2012). Gse Activity and Mortgage Supply in Lower-Income and Minority Neighborhoods: The Effect of the Affordable Housing Goals. *The Journal of Real Estate Finance and Economics*, 45(1):238–261.

Bostic, R. W., Engel, K. C., McCoy, P. A., Pennington-Cross, A., and Wachter, S. M. (2008). State and Local Anti-Predatory Lending Laws: The Effect of Legal Enforcement Mechanisms. *Journal of Economics and Business*, 60(1-2):47–66.

Bundorf, M. K., Levin, J., and Mahoney, N. (2012). Pricing and Welfare in Health Plan Choice. *American Economic Review*, 102(7):3214–48.

Campbell, J. Y. and Cocco, J. F. (2015). A Model of Mortgage Default. *The Journal of Finance*, 70(4):1495–1554.

Dell'Ariccia, G., Igan, D., and Laeven, L. (2008). *Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market*. Number 2008-2106. International Monetary Fund.

Demyanyk, Y. and Van Hemert, O. (2009). Understanding the Subprime Mortgage Crisis. *The Review of Financial Studies*, 24(6):1848–1880.

Deng, Y., Quigley, J. M., and Order, R. (2000). Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. *Econometrica*, 68(2):275–307.

Di Maggio, M. and Kermani, A. (2017). Credit-Induced Boom and Bust. *The Review of Financial Studies*, 30(11):3711–3758.

Di Maggio, M., Kermani, A., and Korgaonkar, S. (2016). Partial Deregulation and Competition: Effects on Risky Mortgage Origination.

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

Einav, L., Finkelstein, A., and Cullen, M. R. (2010). Estimating Welfare in Insurance Markets using Variation in Prices. *The Quarterly Journal of Economics*, 125(3):877–921.

Einav, L., Jenkins, M., and Levin, J. (2012). Contract Pricing in Consumer Credit Markets. *Econometrica*, 80(4):1387–1432.

Elenev, V., Landvoigt, T., and Van Nieuwerburgh, S. (2016). Phasing Out the GSEs. *Journal of Monetary Economics*, 81:111–132.

Favara, G. and Imbs, J. (2015). Credit Supply and the Price of Housing. *American Economic Review*, 105(3):958–92.

Federal Housing Finance Agency (2009). Fannie Mae and Freddie Mac Single-Family Guarantee Fees in 2007 and 2008.

Financial Crisis Inquiry Commission (2011). *The financial crisis inquiry report, authorized edition: Final report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*. Public Affairs.

Handel, B., Hendel, I., and Whinston, M. D. (2015). Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk. *Econometrica*, 83(4):1261–1313.

Hastings, J. S., Hortaçsu, A., and Syverson, C. (2013). Advertising and competition in privatized social security: The case of mexico. *NBER Working Paper Series*, page 18881.

Hellmann, T. F., Murdock, K. C., and Stiglitz, J. E. (2000). Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough? *American Economic Review*, 90(1):147–165.

Ho, G. and Pennington-Cross, A. (2005). The Impact of Local Predatory Lending Laws.

Hurst, E., Keys, B. J., Seru, A., and Vavra, J. (2016). Regional Redistribution through the US Mortgage Market. *American Economic Review*, 106(10):2982–3028.

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

Jeske, K., Krueger, D., and Mitman, K. (2013). Housing, Mortgage Bailout Guarantees and the Macro Economy. *Journal of Monetary Economics*, 60(8):917–935.

Kawai, K., Onishi, K., and Uetake, K. (2014). Signaling in Online Credit Markets.

Keeley, M. C. (1990). Deposit Insurance, Risk, and Market Power in Banking. *The American Economic Review*, pages 1183–1200.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2009). Financial Regulation and Securitization: Evidence from Subprime Loans. *Journal of Monetary Economics*, 56(5):700–720.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2010). Did Securitization Lead to Lax Screening? Evidence from Subprime Loans. *The Quarterly Journal of Economics*, 125(1):307–362.

Krainer, J. and Laderman, E. (2014). Mortgage Loan Securitization and Relative Loan Performance. *Journal of Financial Services Research*, 45(1):39–66.

Lucas, D. and McDonald, R. (2010). Valuing Government Guarantees: Fannie and Freddie Revisited. In *Measuring and Managing Federal Financial Risk*, pages 131–154. University of Chicago Press.

Mathewson, G. F. and Winter, R. A. (1984). An Economic Theory of Vertical Restraints. *The RAND Journal of Economics*, pages 27–38.

Mian, A. and Sufi, A. (2009). The Consequences of Mortgage Credit Expansion: Evidence from the Us Mortgage Default Crisis. *The Quarterly Journal of Economics*, 124(4):1449–1496.

Nguyen, H.-L. Q. (2014). Do Bank Branches Still Matter? the Effect of Closings on Local Economic Outcomes.

Pennington-Cross, A. and Ho, G. (2008). Predatory Lending Laws and the Cost of Credit. *Real Estate Economics*, 36(2):175–211.

Piskorski, T., Seru, A., and Vig, V. (2010). Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis. *Journal of Financial Economics*, 97(3):369–397.

Rey, P. and Tirole, J. (1986). The Logic of Vertical Restraints. *The American Economic Review*, pages 921–939.

Rice, T. and Strahan, P. E. (2010). Does Credit Competition Affect Small-Firm Finance? *The Journal of Finance*, 65(3):861–889.

Rosen, R. (2011). Competition in Mortgage Markets: The Effect of Lender Type on Loan Characteristics.

Scharfstein, D. and Sunderam, A. (2014). Market Power in Mortgage Lending and the Transmission of Monetary Policy.

Stole, L. A. (2007). Price Discrimination and Competition. *Handbook of Industrial Organization*, 3:2221–2299.

Tewari, I. (2014). The Distributive Impacts of Financial Development: Evidence from Mortgage Markets during Us Bank Branch Deregulation. *American Economic Journal: Applied Economics*, 6(4):175–96.

White, A., Reid, C., Ding, L., and Quercia, R. G. (2011). The Impact of State Anti-Predatory Lending Laws on the Foreclosure Crisis. *Cornell JL & Pub. Pol'y*, 21:247.

Winter, R. A. (1993). Vertical Control and Price versus Nonprice Competition. *The Quarterly Journal of Economics*, 108(1):61–76.

Table 1: Loan-Level Descriptive Statistics

	Mean	Standard Deviation	Minimum	Maximum
Interest Rate (%)	6.418	0.773	2.99	12.63
Credit Score (Equifax)	721.554	56.002	300	850
Income ($)	3,472.799	2,861.42	56.382	316,034
Property Value ($)	245,608.816	158,952.993	6,666.667	54,900,000
Down Payment ($)	77,076.587	109,590.926	0	54,351,000
Insurance Percentage (%)	4.798	10.089	0	55
DTI (%)	34.961	12.342	1	65
LTV (%)	73.269	15.78	1	100
Default Rate	0.036	0.186	0	1
Loan Purpose - Purchase	0.301	0.459	0	1
Loan Purpose - Cash-Put Refinance	0.408	0.492	0	1
Loan Purpose - No Cash-Put Refinance	0.291	0.454	0	1
Property Type - Single-Family	0.078	0.267	0	1
Property Type - Condo	0.003	0.057	0	1
Property Type - Co-Op	0.007	0.085	0	1
Property Type - Manufactured Housing	0.124	0.329	0	1
Property Type - Planned Unit Development	0.788	0.409	0	1
Occupancy Status - Principal	0.905	0.293	0	1
Occupancy Status - Second	0.039	0.193	0	1
Occupancy Status - Investor	0.056	0.23	0	1
# of Observations		22,726,545		

Note: Table 1 presents a summary of loan characteristic statistics for key variables in the sample during the 2000 – 08 period.

Table 2: Market-Level Descriptive Statistics

	Mean	Standard Deviation	Minimum	Maximum
Share of National Banks – GSE Loans	0.386	0.159	0.000	0.899
Share of National Banks – All Loans	0.311	0.116	0.000	0.833
# of Lenders	55.306	36.114	1.000	256.000
# of OCC Lenders	12.373	6.703	0.000	54.000
HHI	0.164	0.180	0.033	1.000
# of Observations		7,426		

Note: Table 2 shows descriptive statistics of market level competition measures. The statistics are calculated using all market-year observations during the 2000 – 08 period.

Table 3: Correlation between Housing Price Index and *Ex Post* Loan Performance

Housing Price Index	-0.176***
	(0.0339)
log(Credit Score)	-0.00888***
	(0.00157)
log(Income)	0.0533***
	(0.00481)
log(Loan Amount)	0.0250***
	(0.00278)
log(Down Payment)	0.00598***
	(0.000937)
Insurance Percentage	-0.000412***
	(0.0000602)
Loan Purpose - Home Purchase	0.00210
	(0.00163)
Loan Purpose - No Cash-Out Refinance	-0.00356***
	(0.00113)
Property Type - Condo	0.0144***
	(0.00509)
Property Type - Co-Op	-0.0168***
	(0.00349)
Property Type - Manufactured Housing	0.00725**
	(0.00352)
Property Type - Planned Unit Development	-0.0130***
	(0.00267)
Occupancy Status - Second	0.0145***
	(0.00282)
Occupancy Status - Investor	0.0264***
	(0.00290)
Quarterly Time Fixed Efects	yes
Market Fixed Effects	yes
# of Observations	43,418
Adjusted R-squared	0.364

Note: Table 3 shows regression results from Equation 7 using the loans from the 2000 – 17 period. The sample is collapsed by market-quarter level. Robust standard errors in parentheses are clustered at the market level. Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

Table 4: The Impact of Preemption on Market Outcomes

	Interest Rate		Default Rate	
	(1)	(2)	(3)	(4)
APL × POST	-0.0203***		-0.00108	
	(0.00693)		(0.00926)	
APL × POST × Share		-0.0712***		0.0376*
		(0.0255)		(0.0223)
POST × Share		0.0390**		0.00361
		(0.0162)		(0.0102)
log(Credit Score)	-1.241***	-0.974***	-0.429***	-0.429***
	(0.184)	(0.0315)	(0.0274)	(0.0274)
log(Income)	0.0962***	0.0517***	-0.0181***	-0.0181***
	(0.0229)	(0.00570)	(0.00348)	(0.00348)
log(Loan Amount)	-0.282***	-0.177***	0.0702***	0.0702***
	(0.0358)	(0.00843)	(0.00819)	(0.00819)
log(Down Payment)	-0.0393	-0.0575***	-0.00461	-0.00461
	(0.0259)	(0.00582)	(0.00524)	(0.00524)
Insurance Percentage	-0.000314	-0.000467	-0.000158	-0.000158
	(0.000939)	(0.000372)	(0.000393)	(0.000393)
Loan Purpose - Home Purchase	-0.0662*	-0.0656***	-0.0736***	-0.0736***
	(0.0333)	(0.00777)	(0.00619)	(0.00619)
Loan Purpose - No Cash-Out Refinance	-0.139***	-0.0692***	-0.0214***	-0.0214***
	(0.0240)	(0.00732)	(0.00731)	(0.00731)
Property Type - Condo	0.0382	0.0367	-0.00765	-0.00765
	(0.0421)	(0.0736)	(0.0188)	(0.0188)
Property Type - Co-Op	0.365***	0.285***	0.0503*	0.0503*
	(0.0657)	(0.0235)	(0.0284)	(0.0284)
Property Type - Manufactured Housing	0.170***	0.0236	-0.00937	-0.00937
	(0.0457)	(0.0201)	(0.0180)	(0.0180)
Property Type - Planned Unit Development	0.0943*	0.0799***	-0.0126	-0.0126
	(0.0509)	(0.0170)	(0.0161)	(0.0161)
Occupancy Status - Second	0.109***	0.0613***	-0.0219	-0.0219
	(0.0226)	(0.0142)	(0.0146)	(0.0146)
Occupancy Status - Investor	0.487***	0.479***	0.0884***	0.0884***
	(0.0281)	(0.0129)	(0.0159)	(0.0159)
Time Fixed Effects	yes		yes	
Market-APL Fixed Effects	yes	yes	yes	yes
State -Time Fixed Effects		yes		yes
# of Observations	93,507	93,507	93,507	93,507
Adjusted R-squared	0.996	0.998	0.700	0.869

Note: Table 4 reports coefficient estimates of weighted least square regressions from Equation 8 and Equation 9. The sample is collapsed by market-month level with weights equal to number of loans purchased by the GSEs in each market. Robust standard errors in parentheses are clustered at the state level in column (1) and (3), and at market level in column (2) and (4). Significance levels: *(p<0.10), **(p<0.05), ***(p<0.01).

Table 5: Parameter Estimates

	β_μ	β_σ	γ	β_c
Constant	-1.094	0.236	1.006	1.331
	(0.252)	(0.200)	(0.075)	(0.254)
log(Credit Score)	0.512	-0.006	0.013	-0.111
	(0.034)	(0.053)	(0.111)	(0.021)
log(Income)	0.028	-0.015	-0.001	-0.010
	(0.040)	(0.014)	(0.011)	(0.002)
log(Property Value)	-0.087	0.020	0.089	-0.044
	(0.032)	(0.014)	(0.001)	(0.008)
Loan Purpose - Purchase	0.103	-0.003	0.001	-0.012
	(0.008)	(0.035)	(0.004)	(0.002)
Loan Purpose - No Cash-Out Refinance	0.056	-0.033	-0.034	-0.013
	(0.010)	(0.028)	(0.005)	(0.002)
Origination ≤ 2 years	0.089	(–)	(–)	(–)
	(0.034)	(–)	(–)	(–)
2 years < Origination ≤ 4 years	0.038	(–)	(–)	(–)
	(0.015)	(–)	(–)	(–)
log(Monthly Required Payment)	(–)	(–)	(–)	0.041
	(–)	(–)	(–)	(0.008)
APL	(–)	(–)	(–)	0.002
	(–)	(–)	(–)	(0.001)
OCC × APL × POST	(–)	(–)	(–)	-0.005
	(–)	(–)	(–)	(0.001)

Note: Table 5 shows structural parameter estimates. The standard error is in parentheses. Market-seller and origination year fixed effects in the cost estimation are omitted for brevity.

Table 6: Welfare Impact of Mortgage Insurance Pricing

	Uniform Pricing	Full Risk-Based Pricing
Panel A: Welfare Change to Baseline		
Mean Δ Interest Rate	0.000	0.065
Mean Δ Default Loss ($)	43.406	-843.218
Mean Δ Lender Profit ($)	413.079	8,068.516
Mean Δ Borrower Surplus ($)	-432.345	-9,748.600
Mean Δ GSE Profit ($)	-14.625	2,199.333
Mean Δ Deadweight Loss ($)	33.892	-519.249
Δ # of Accounts (10k)	5.284	-76.842
Panel B: Percentage Change to Baseline		
Mean Δ Interest Rate	0.000	4.463
Mean Δ Default Loss	3.295	-64.014
Mean Δ Lender Profit	2.563	50.071
Mean Δ Borrower Surplus	-0.352	-7.930
Mean Δ GSE Profit	0.133	-20.009
Mean Δ Deadweight Loss	3.132	-47.980
Δ # of Accounts	0.241	-3.509

Note: Table 6 presents the welfare results of counterfactuals of uniform pricing and perfect risk-based pricing. Mean changes in the statistics are calculated using total changes in the statistics divided by sample size (22 million). Panel A presents the welfare changes in dollar value relative to the baseline (GSE pricing). Panel B presents the welfare changes in percentage relative to the baseline.

Table 7: Welfare Impact of Market Structure - GSE Pricing

	N	$0.75N$	$1.25N$	$1.5N$	$2N$	$N \to \infty$
Panel A: Welfare Change to Baseline						
Mean Δ Interest Rate	(–)	0.023	-0.016	-0.027	-0.041	-0.085
Mean Δ Default Loss ($)	(–)	-106.981	71.080	120.902	186.349	401.294
Mean Δ Lender Profit ($)	(–)	4367.090	-2,935.142	-4,998.073	-7,663.553	-16,114.002
Mean Δ Borrower Surplus ($)	(–)	-4565.536	3,031.216	5,146.138	7,859.923	16,328.823
Mean Δ GSE Profit ($)	(–)	216.230	-142.219	-241.180	-368.143	-759.490
Mean Δ Deadweight Loss ($)	(–)	-17.784	46.145	93.116	171.773	544.669
Δ # of Accounts (10k)	(–)	-27.103	17.208	28.980	43.816	87.836
Panel B: Percentage Change to Baseline						
Mean Δ Interest Rate	(–)	1.592	-1.072	-1.826	-2.796	-5.831
Mean Δ Default Loss	(–)	-8.122	5.396	9.178	14.147	30.465
Mean Δ Lender Profit	(–)	27.101	-18.215	-31.017	-47.558	-100.000
Mean Δ Borrower Surplus	(–)	-3.714	2.466	4.186	6.393	13.282
Mean Δ GSE Profit	(–)	-1.967	1.294	2.194	3.349	6.909
Mean Δ Deadweight Loss	(–)	-1.643	4.264	8.604	15.872	50.329
Δ # of Accounts	(–)	-1.238	0.786	1.323	2.001	4.011

Note: Table 7 presents the welfare results of counterfactuals that change the number of lenders from $0.75N$ to $2N$ and perfect competition. Mean changes in the statistics are calculated using total changes in the statistics divided by sample size (22 million). Panel A presents the welfare changes in dollar value relative to the baseline, which is the case of N under GSE pricing. Panel B presents the welfare changes in percentage relative to the baseline.

Table 8: Welfare Impact of Market Structure - Full Risk-Based Pricing

	N	$0.75N$	$1.25N$	$1.5N$	$2N$	$N \to \infty$
Panel A: Welfare Change to Baseline						
Mean Δ Interest Rate	0.065	0.088	0.049	0.038	0.024	-0.020
Mean Δ Default Loss ($)	-843.218	-889.285	-811.637	-788.553	-758.258	-652.063
Mean Δ Lender Profit ($)	8,068.516	11,968.503	5,450.174	3,611.062	1,228.934	-6,388.580
Mean Δ Borrower Surplus ($)	-9,748.600	-14,127.147	-6,843.098	-4,817.043	-2,217.208	5,906.097
Mean Δ GSE Profit ($)	2,199.333	2,401.445	2,059.945	1,960.041	1,832.566	1,447.906
Mean Δ Deadweight Loss ($)	-519.249	-242.800	-667.021	-754.061	-844.291	-965.423
Δ # of Accounts (10k)	-76.842	-105.846	-57.732	-44.359	-27.404	22.636
Panel B: Percentage Change to Baseline						
Mean Δ Interest Rate	4.463	6.061	3.388	2.632	1.659	-1.378
Mean Δ Default Loss	-64.014	-67.511	-61.616	-59.864	-57.564	-49.502
Mean Δ Lender Profit	50.071	74.274	33.823	22.409	7.626	-39.646
Mean Δ Borrower Surplus	-7.930	-11.491	-5.566	-3.918	-1.803	4.804
Mean Δ GSEs Profit	-20.009	-21.847	-18.740	-17.832	-16.672	-13.172
Mean Δ Deadweight Loss	-47.980	-22.435	-61.634	-69.677	-78.015	-89.207
Δ # of Accounts	-3.509	-4.834	-2.637	-2.026	-1.252	1.034
Panel C: Welfare Change to GSE Pricing with Same Market Concentration						
Mean Δ Interest Rate	0.065	0.065	0.065	0.065	0.065	0.065
Mean Δ Default Loss ($)	-843.218	-782.304	-882.717	-909.455	-944.607	-1,053.357
Mean Δ Lender Profit ($)	8,068.516	7,601.413	8,385.317	8,609.135	8,892.487	9,725.421
Mean Δ Borrower Surplus ($)	-9,748.600	-9,561.611	-9,874.314	-9,963.181	-10,077.132	-10,422.726
Mean Δ GSE Profit ($)	2,199.333	2,185.214	2,202.163	2,201.222	2,200.709	2,207.396
Mean Δ Deadweight Loss ($)	-519.249	-225.017	-713.166	-847.176	-1,016.064	-1,510.091
Δ # of Accounts (10k)	-76.842	-78.743	-74.939	-73.339	-71.220	-65.199

Note: Table 8 presents the welfare results under full risk-based pricing for the cases from $0.75N$ to $2N$ and perfect competition. Mean changes in the statistics are calculated using total changes in the statistics divided by sample size (22 million). Panel A presents the welfare changes in dollar value relative to the baseline, which is the case of N under GSE pricing. Panel B presents the welfare changes in percentage relative to the baseline. Panel C presents the welfare changes in dollar value relative to the case under GSE pricing with the same competition level.

Table 9: Welfare Impact of Symmetric Information - GSE Pricing

	N	$0.75N$	$1.25N$	$1.5N$	$2N$	$N \to \infty$
Panel A: Welfare Change to Baseline						
Mean Δ Interest Rate	0.000	0.023	-0.016	-0.027	-0.041	-0.085
Mean Δ Default Loss ($)	5.475	-101.869	75.407	124.347	188.572	401.294
Mean Δ Lender Profit ($)	112.837	4,509.973	-2,849.710	-4,932.952	-7,622.870	-16,114.002
Mean Δ Borrower Surplus ($)	-116.540	-4,726.801	2,945.473	5,081.180	7,819.195	16,328.823
Mean Δ GSE Profit ($)	-3.655	221.549	-147.247	-245.748	-370.973	-759.490
Mean Δ Deadweight Loss ($)	7.357	-4.721	51.485	97.519	174.648	544.669
Δ # of Accounts (10k)	-1.220	-30.153	16.612	28.605	43.577	87.836
Panel B: Percentage Change to Baseline						
Mean Δ Interest Rate	0.000	1.592	-1.072	-1.826	-2.796	-5.831
Mean Δ Default Loss	0.416	-7.734	5.725	9.440	14.316	30.465
Mean Δ Lender Profit	0.700	27.988	-17.685	-30.613	-47.306	-100.000
Mean Δ Borrower Surplus	-0.095	-3.845	2.396	4.133	6.360	13.282
Mean Δ GSE Profit	0.033	-2.016	1.340	2.236	3.375	6.909
Mean Δ Deadweight Loss	0.680	-0.436	4.757	9.011	16.138	50.329
Δ # of Accounts	-0.056	-1.377	0.759	1.306	1.990	4.011
Panel C: Welfare Change to Asymmetric Information Case with Same Market Concentration						
Mean Δ Interest Rate	0.000	0.000	0.000	0.000	0.000	0.000
Mean Δ Default Loss ($)	5.475	5.112	4.327	3.445	2.223	0.000
Mean Δ Lender Profit ($)	112.837	142.884	85.432	65.122	40.683	0.000
Mean Δ Borrower Surplus ($)	-116.540	-161.264	-85.743	-64.957	-40.728	0.000
Mean Δ GSE Profit ($)	-3.655	5.318	-5.029	-4.568	-2.830	0.000
Mean Δ Deadweight Loss ($)	7.357	13.062	5.340	4.404	2.875	0.000
Δ # of Accounts (10k)	-0.913	-2.282	-0.445	-0.281	-0.179	0.000

Note: Table 9 presents the welfare results of symmetric information under GSE pricing for the cases from $0.75N$ to $2N$ and perfect competition. Mean changes in the statistics are calculated using total changes in the statistics divided by sample size (22 million). Panel A presents the welfare changes in dollar value relative to the baseline, which is the case of asymmetric information under GSE pricing when $N = 1$. Panel B presents the welfare changes in percentage relative to the baseline. Panel C presents the welfare changes in dollar value relative to the asymmetric information case with the same competition level and market design.

Table 10: Welfare Impact of Symmetric Information - Full Risk-Based Pricing

	N	$0.75N$	$1.25N$	$1.5N$	$2N$	$N\to\infty$
Panel A: Welfare Change to Baseline						
Mean Δ Interest Rate	0.065	0.088	0.049	0.038	0.024	-0.020
Mean Δ Default Loss ($)	-873.466	-913.523	-846.219	-826.683	-801.388	-712.195
Mean Δ Lender Profit ($)	8,053.664	12,001.376	5,394.835	3,524.420	1,103.114	-6,616.933
Mean Δ Borrower Surplus ($)	-9,628.946	-14,061.527	-6,684.221	-4,630.223	-1,995.929	6,204.245
Mean Δ GSE Profit ($)	2,226.633	2,429.450	2,088.277	1,990.299	1,865.193	1,494.911
Mean Δ Deadweight Loss ($)	-651.350	-369.299	-798.891	-884.496	-972.379	-1,082.223
Δ # of Accounts (10k)	-67.673	-98.047	-47.909	-34.230	-16.925	33.520
Panel B: Percentage Change to Baseline						
Mean Δ Interest Rate	4.463	6.061	3.388	2.632	1.659	-1.378
Mean Δ Default Loss	-66.310	-69.351	-64.242	-62.759	-60.838	-54.067
Mean Δ Lender Profit	49.979	74.478	33.479	21.872	6.846	-41.063
Mean Δ Borrower Surplus	-7.832	-11.438	-5.437	-3.766	-1.623	5.047
Mean Δ GSE Profit	-20.257	-22.102	-18.998	-18.107	-16.969	-13.600
Mean Δ Deadweight Loss	-60.186	-34.124	-73.819	-81.730	-89.850	-100.000
Δ # of Accounts	-3.091	-4.478	-2.188	-1.563	-0.773	1.531
Panel C: Welfare Change to Asymmetric Information Case with Same Market Concentration						
Mean Δ Interest Rate	0.000	0.000	0.000	0.000	0.000	0.000
Mean Δ Default Loss ($)	-30.248	-24.238	-34.582	-38.130	-43.130	-60.132
Mean Δ Lender Profit ($)	-14.852	32.873	-55.339	-86.642	-125.820	-228.353
Mean Δ Borrower Surplus ($)	119.654	65.620	158.877	186.820	221.280	298.148
Mean Δ GSE Profit ($)	27.300	28.006	28.332	30.257	32.628	47.005
Mean Δ Deadweight Loss ($)	-132.102	-126.498	-131.870	-130.435	-128.088	-116.800
Δ # of Accounts (10k)	6.862	5.836	7.350	7.580	7.842	8.144

Note: Table 10 presents the welfare results of symmetric information under fully risk-based pricing for the cases from $0.75N$ to $2N$ and perfect competition. Mean changes in the statistics are calculated using total changes in the statistics divided by sample size (22 million). Panel A presents the welfare changes in dollar value relative to the baseline, which is the case of asymmetric information with N under GSE pricing. Panel B presents the welfare changes in percentage relative to the baseline. Panel C presents the welfare changes in dollar value relative to the asymmetric information case with the same competition level and market design.

Table 11: Differential Welfare Impact of Mortgage Insurance Pricing and Market Structure

	GSE Pricing		Full Risk-Based Pricing	
	N	$2N$	N	$2N$
Mean Δ Borrower Surplus ($)				
Income – High	(–)	5,136.338	-6,860.261	-1,931.337
Income – Low	(–)	2,723.616	-2,888.389	-285.892
Credit Score – High	(–)	4,352.715	-5,974.797	-1,758.461
Credit Score – Low	(–)	3,514.065	-3,791.653	-469.288
λ – High	(–)	4,164.116	-5,270.880	-1,240.688
λ – Low	(–)	3,695.807	-4,477.720	-976.521
Mean Δ Deadweight Loss ($)				
Income – High	(–)	-629.488	-454.899	-629.488
Income – Low	(–)	-214.808	-64.355	-214.808
Credit Score – High	(–)	-268.240	-126.152	-268.240
Credit Score – Low	(–)	-573.555	-390.932	-573.555
λ – High	(–)	-243.202	-104.168	-243.202
λ – Low	(–)	-601.089	-415.081	-601.089

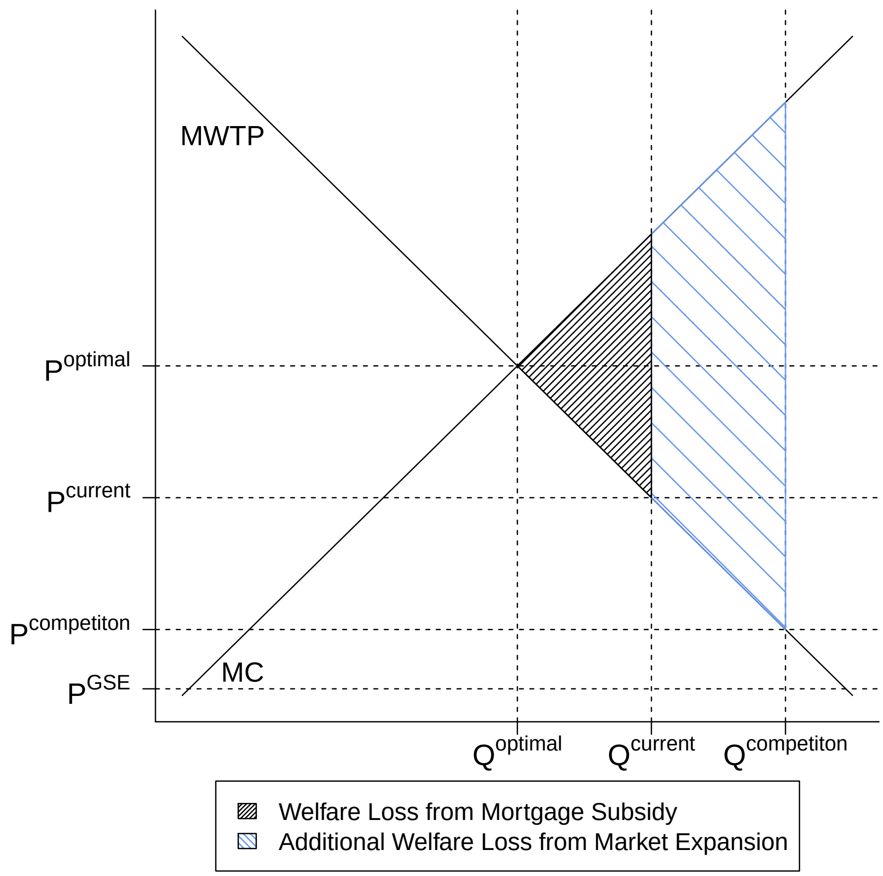
Note: Table 11 studies the differential effects on borrower surplus and total surplus based on three categorizations: (i) an indicator of whether or not the borrower's income is greater than the 50th percentile; (ii) an indicator of whether or not the borrower's credit score is greater than the 50th percentile; (iii) an indicator of whether or not λ_{in} is greater than the 50th percentile. The numbers show borrower surplus and total surplus change in dollar value under different competition levels and market designs relative to the baseline, which is the case of N under GSE pricing.

Table 12: Differential Welfare Impact of Symmetric Information

	GSE Pricing		Full Risk-Based Pricing	
	N	$2N$	N	$2N$
Panel A: Welfare Change to Baseline				
Mean Δ Borrower Surplus ($)				
Income – High	-47.931	5,121.311	-6,772.285	-1,795.859
Income – Low	-68.609	2,697.915	-2,856.711	-200.090
Credit Score – High	-45.220	4,337.586	-5,937.579	-1,683.769
Credit Score – Low	-71.108	3,488.551	-3,709.583	-323.283
λ – High	-174.691	4,103.951	-5,145.555	-973.631
λ – Low	58.151	3,715.244	-4,483.392	-1,022.297
Mean Δ Deadweight Loss ($)				
Income – High	2.992	147.787	-532.442	-704.535
Income – Low	4.366	26.862	-118.914	-267.850
Credit Score – High	1.718	42.479	-167.483	-307.154
Credit Score – Low	5.608	131.442	-481.301	-662.321
λ – High	8.407	42.049	-214.573	-341.836
λ – Low	-1.050	132.599	-436.778	-630.543
Panel B: Welfare Change to Asymmetric Information Case with Same Market Concentration				
Mean Δ Borrower Surplus ($)				
Income – High	-47.931	-15.027	87.976	135.478
Income – Low	-68.609	-25.701	31.678	85.802
Credit Score – High	-45.220	-15.129	37.217	74.692
Credit Score – Low	-71.108	-25.514	82.070	146.004
λ – High	-174.691	-60.165	125.325	267.056
λ – Low	58.151	19.437	-5.671	-45.776
Mean Δ Deadweight Loss ($)				
Income – High	2.992	1.096	-77.543	-75.047
Income – Low	4.366	1.779	-54.559	-53.042
Credit Score – High	1.718	0.990	-41.332	-38.914
Credit Score – Low	5.608	1.878	-90.369	-88.766
λ – High	8.407	-0.898	-110.405	-98.634
λ – Low	-1.050	3.774	-21.697	-29.454

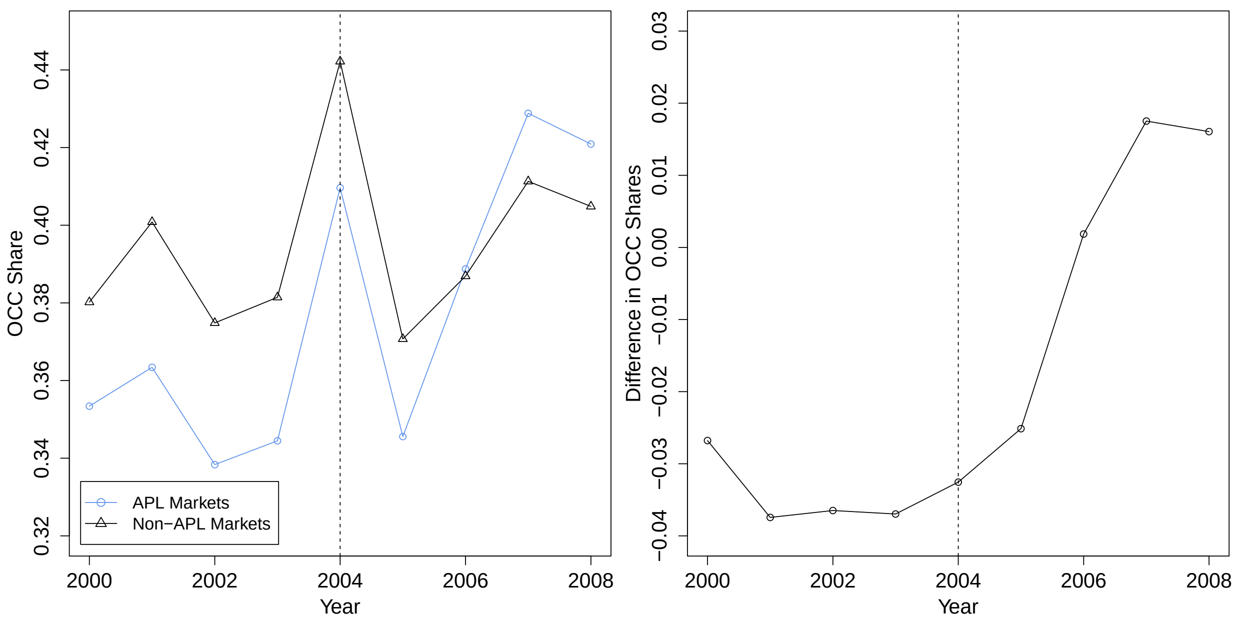
Note: Table 12 studies the differential effects on borrower surplus and total surplus based on three categorizations: (i) an indicator of whether or not the borrower's income is greater than the 50th percentile; (ii) an indicator of whether or not the borrower's credit score is greater than the 50th percentile; (iii) an indicator of whether or not λ_{in} is greater than the 50th percentile. Panel A presents the borrower surplus and total surplus changes in dollar value relative to the baseline, which is the case of N under GSE pricing. Panel B presents the borrower surplus and total surplus changes in dollar value relative to the asymmetric information case with the same competition level and market design.

Figure 1. The Welfare Impact of GSE Pricing and Market Structure



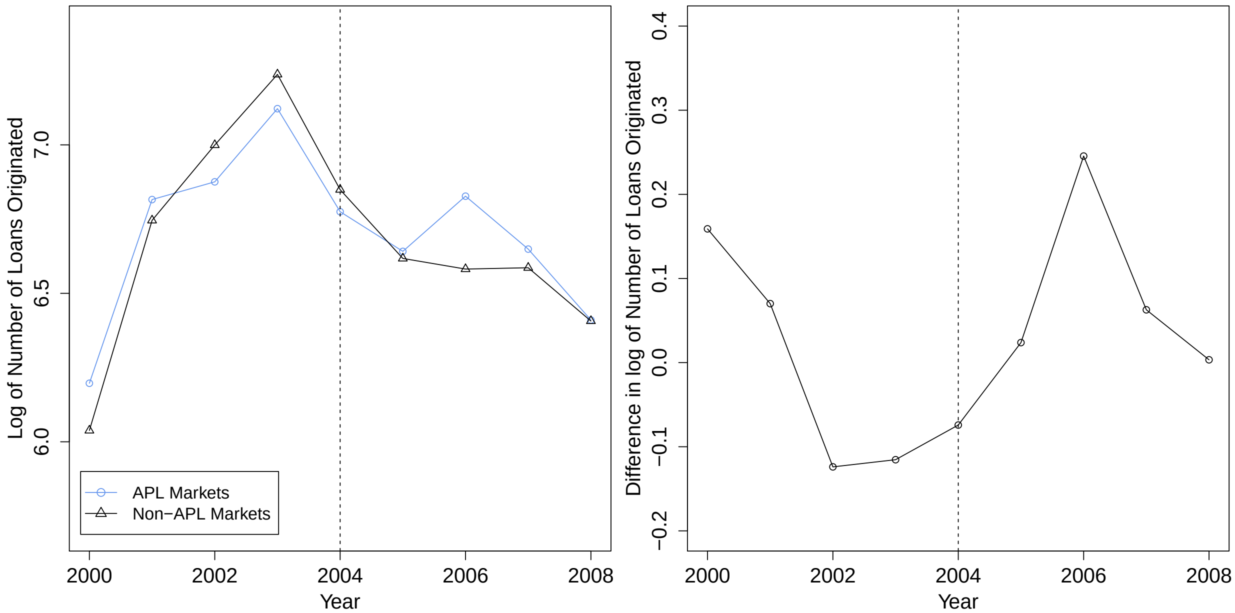
Note: Figure 1 illustrates deadweight loss from GSE pricing and how it interacts with the competition among lenders. The downward sloping line is the marginal willingness to pay (MWTP). The upward sloping line is the marginal cost (MC) for the GSEs. The GSEs underprice guarantee fee at P^{GSE}. Competition decreases equilibrium mortgage price from $\mathrm{P}^{\mathrm{current}}$ to $\mathrm{P}^{\mathrm{competition}}$ and increases equilibrium mortgage provision from $\mathrm{Q}^{\mathrm{current}}$ to $\mathrm{Q}^{\mathrm{competition}}$. Competition results in an additional deadweight loss of the blue shaded area.

Figure 2. Share of National Banks Trends



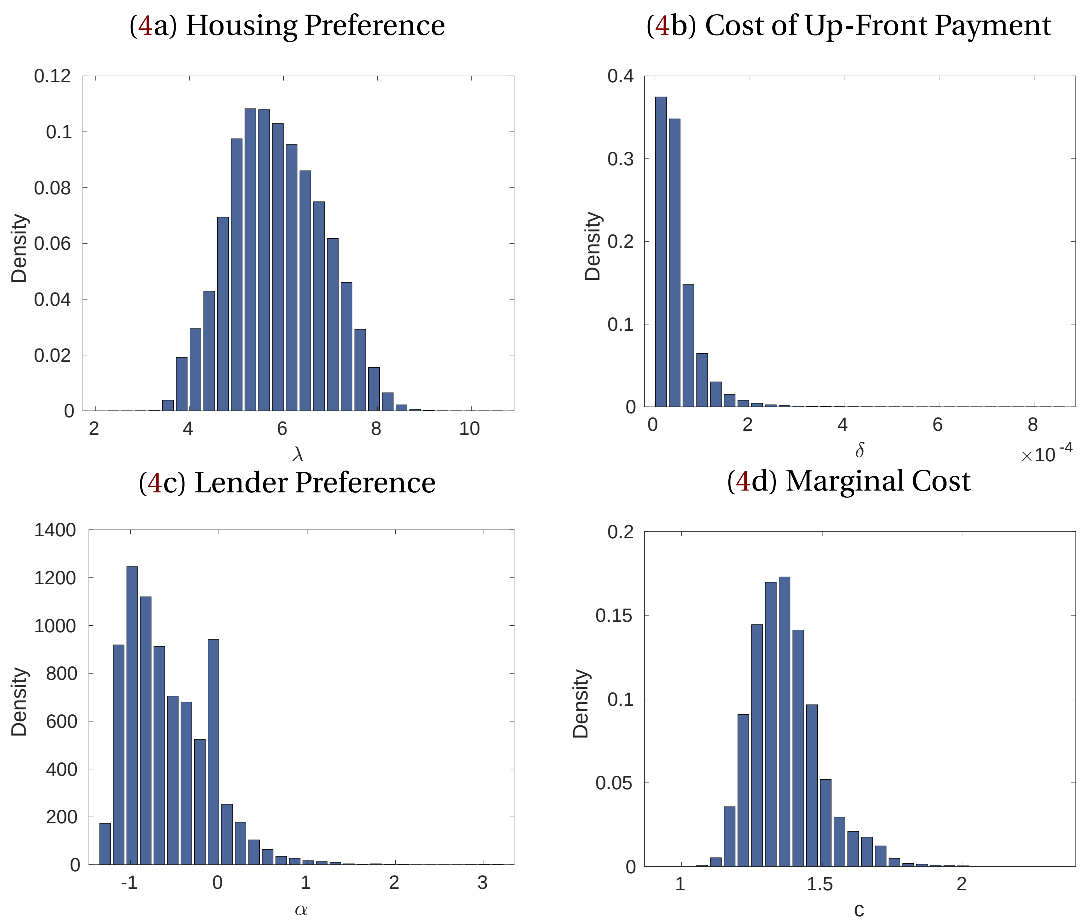
Note: Figure 2 plots the average share of national banks in the GSE market from 2000 – 08. The left panel shows the average share of national banks separately for APL market and non-APL market. The right panel shows the average difference between APL market and non-APL market. The dotted vertical line indicates the year when the OCC preemption rule was implemented in 2004.

Figure 3. Number of Loan Trends



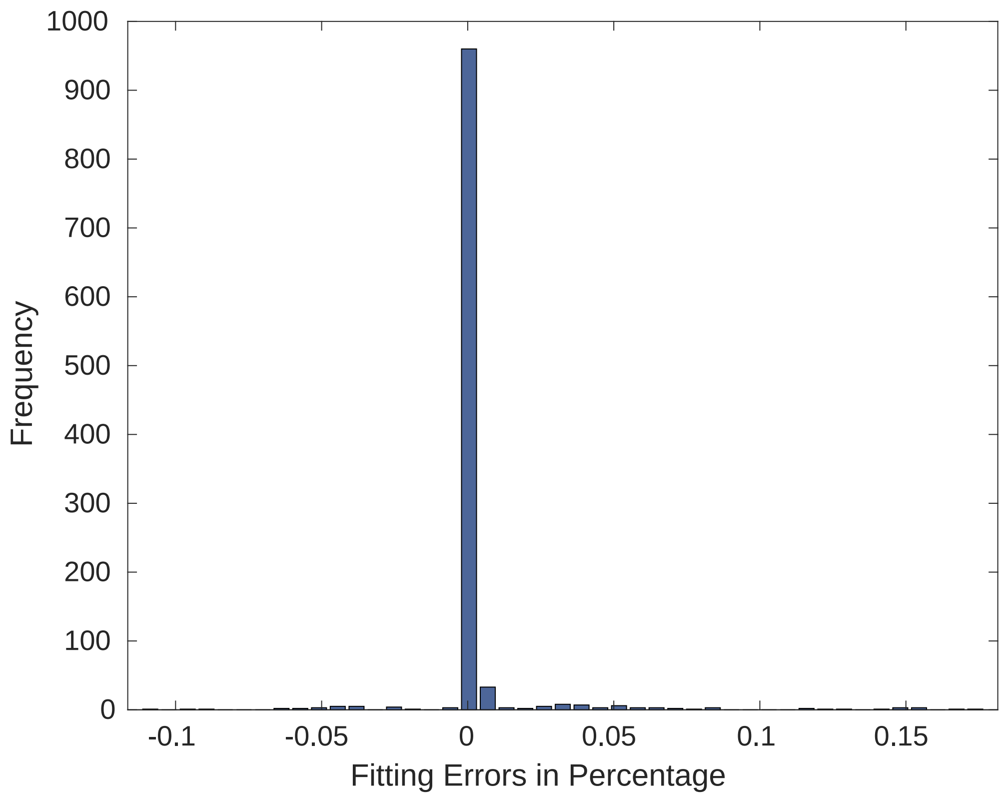
Note: Figure 3 plots the average number of loans (in log) purchased by the GSEs from 2000 – 08. The left panel shows the average number of loans (in log) separately for APL market and non-APL market. The right panel shows the average difference between APL market and non-APL market. The dotted vertical line indicates the year when the OCC preemption rule was implemented in 2004.

Figure 4. Distribution of Estimates



Note: Figure 4 plots the probability distribution of the parameter estimates from structural estimation.

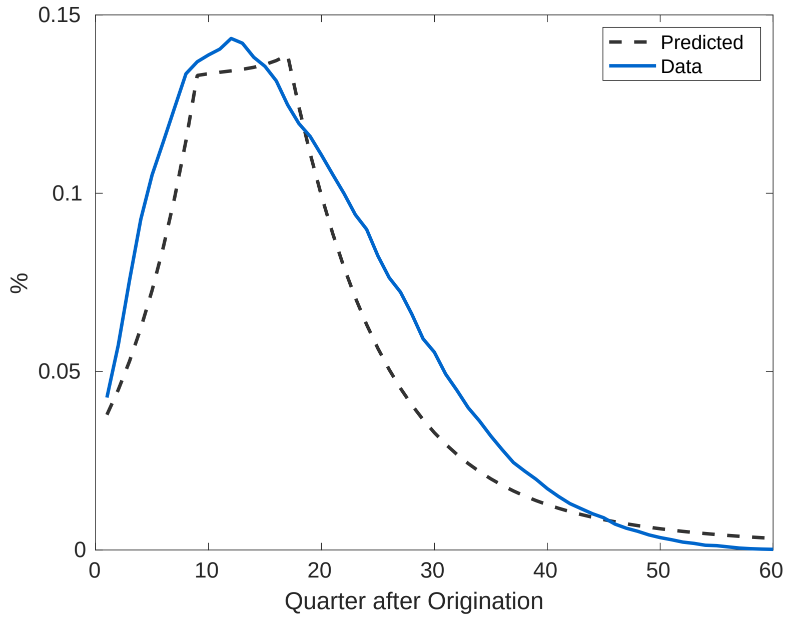
Figure 5. Fitting Errors



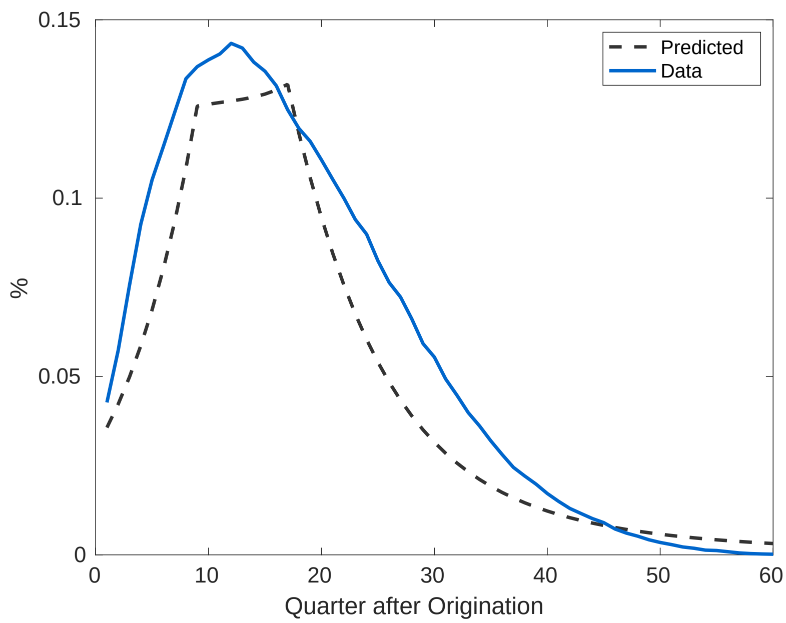
Note: Figure 5 plots the fitting errors for all 1,080 moments used in the GMM estimation. It is calculated as the absolute value of the residuals divided by the empirical means.

Figure 6. Fitted Default Path

(6a) In-Sample Fit

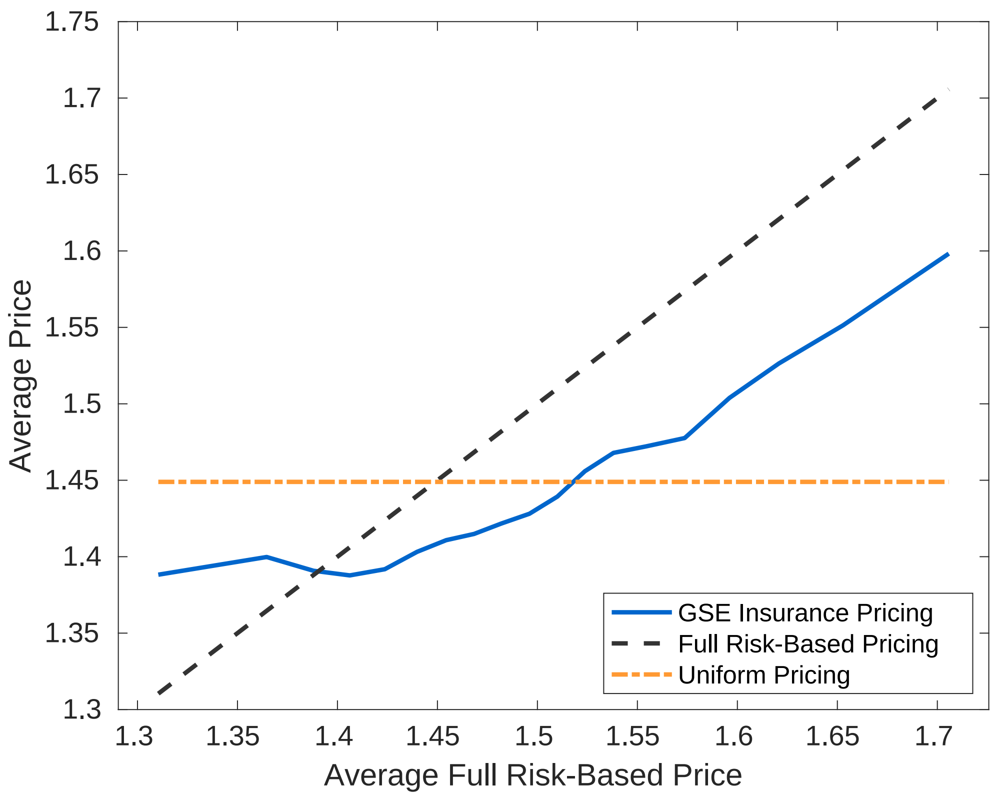


(6b) Out-of-Sample Fit



Note: Figure 6 compares the predicted hazard rate of default with the data. The dashed lines are the model prediction and the solid lines are the observed data. The upper panel is the in-sample prediction. The lower panel is the out-of-sample fit that uses the parameter estimates from the 50 percent of the loan (random selection) to predict the default path for the remaining 50 percent of the loans.

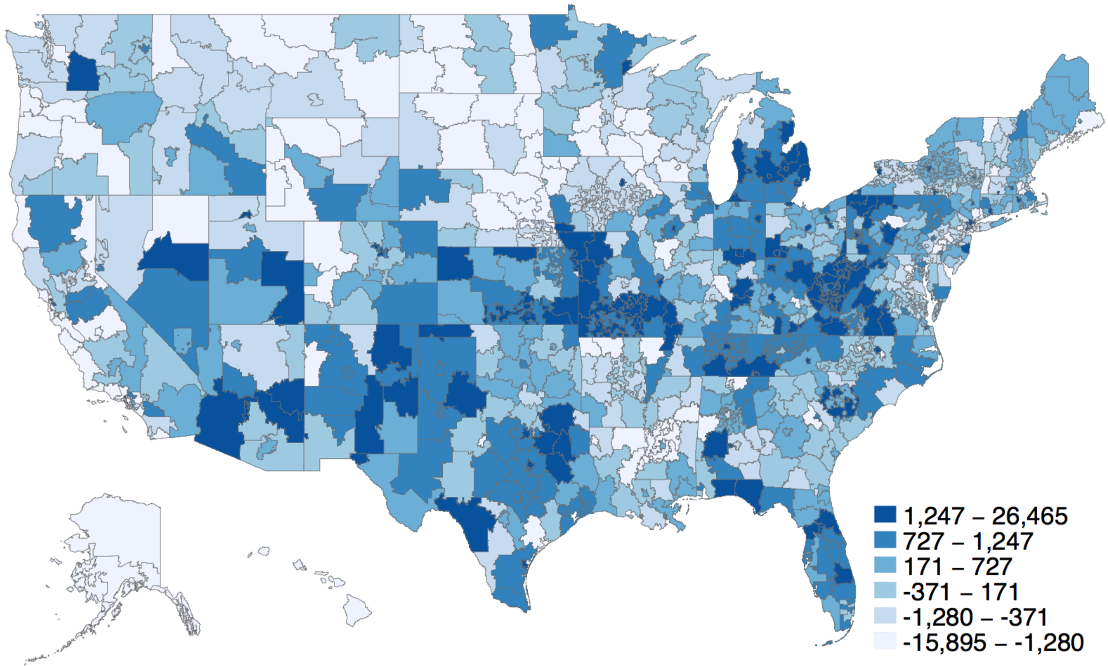
Figure 7. Comparison between GSE Pricing, Uniform Pricing, and Full Risk-Based Pricing



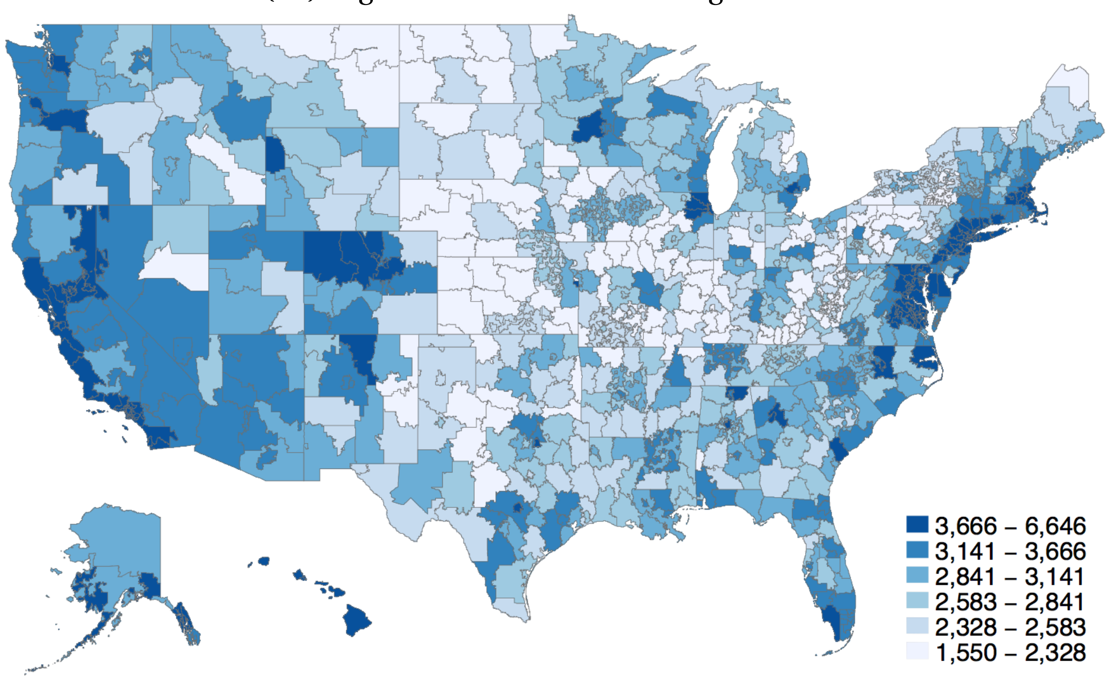
Note: Figure 7 compares borrowers' average prices under different pricing schemes (on the y-axis) with their average full risk-based prices (on the x-axis). The dashed line is a 45-degree line.

Figure 8. Regional Distribution of Mortgage Subsidy and Average Income

(6a) Regional Distribution - Mean Deviation of Subsidy

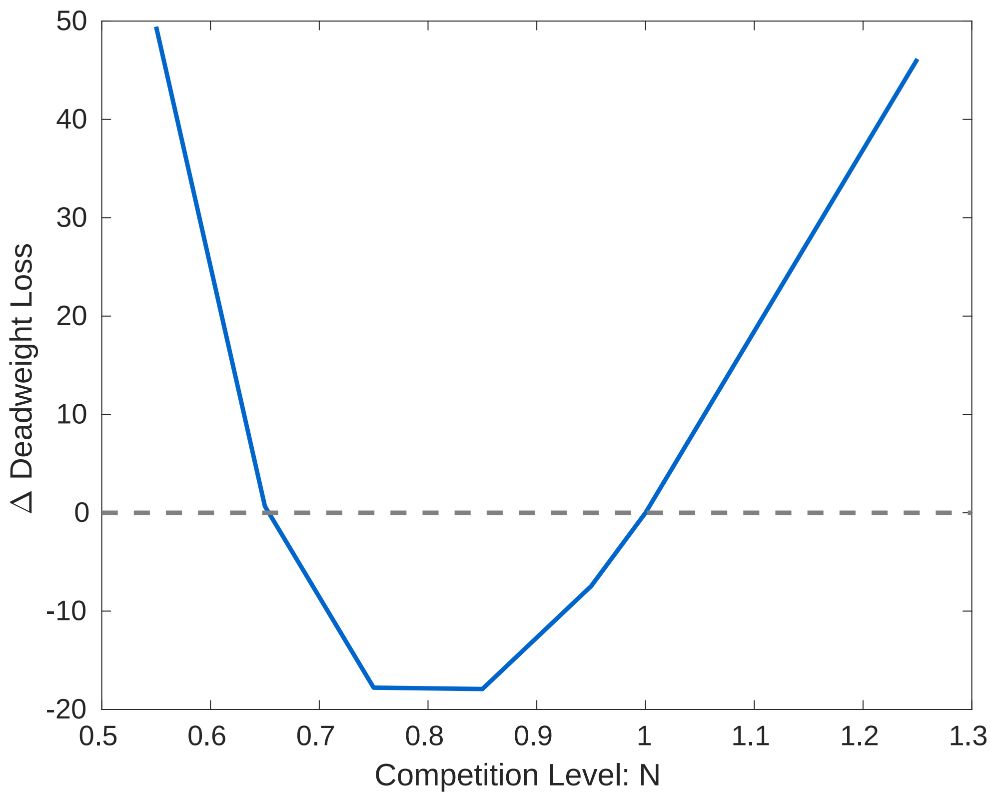


(6b) Regional Distribution - Average Income



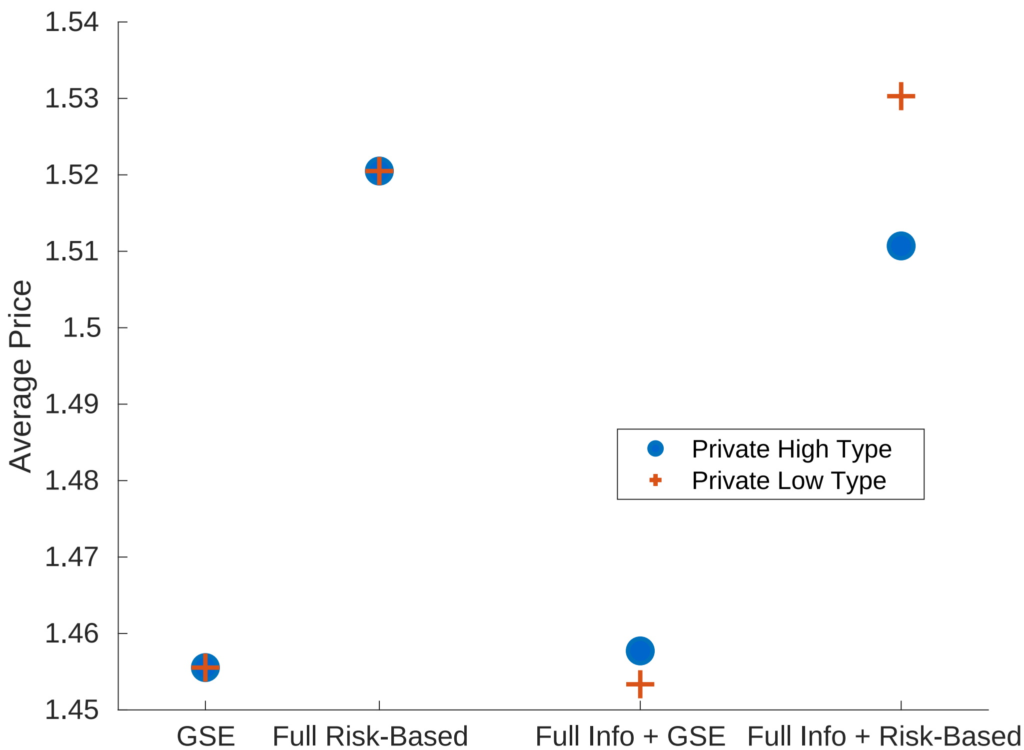
Note: Figure 8 plots the heat map charts for the distribution of outcomes of interest by 3-digit zip code level. The upper panel displays the mean deviation of mortgage subsidy for each region. The lower panel shows average income for each region. The mortgage subsidy is defined as the change in consumer welfare in dollar value from full risk-based pricing to GSE pricing.

Figure 9. The Changes in Deadweight Loss along with the Level of Competition



Note: Figure 9 plots the changes in deadweight loss relative to the baseline for various cases from $0.55N$ to $1.25N$. The baseline is the current market concentration N. The competition level is increasing with N, the case of N less than 1 implies lower competition than the baseline.

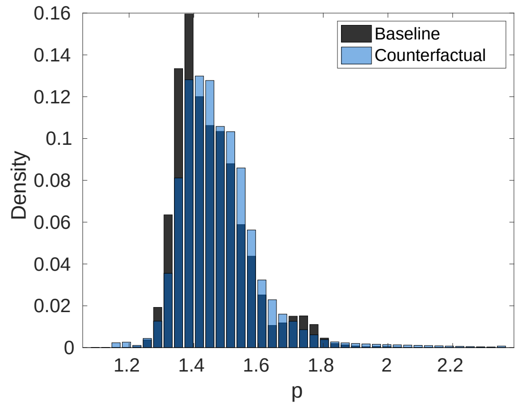
Figure 10. Differential Effects of Symmetric Information across Private Types



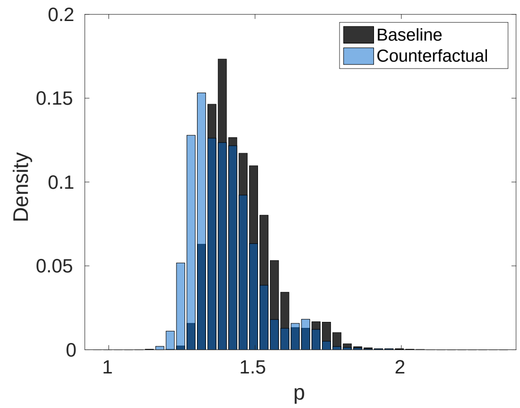
Note: Figure 10 plots average prices separately for private high type (i.e., with higher λ) and private low type (i.e., with lower λ) under the following four cases: (i) GSE pricing with asymmetric information; (ii) full risk-based pricing with asymmetric information; (iii) GSE pricing with symmetric information; (iv) full risk-based pricing with symmetric information. I categorize the private type based on an indicator of whether or not the λ_{in} is greater than the 50th percentile.

Figure 11. Distribution of Interest Rates in Counterfactual

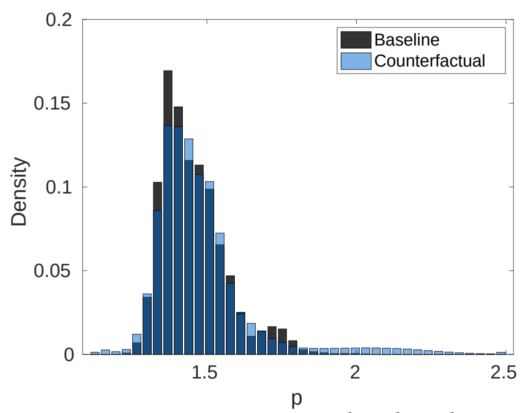
(11a) Full Risk-Based Pricing



(11b) GSE Pricing and Symmetric Information



(11c) Full Risk-Based Pricing and Symmetric Information



Note: Figure 11 compares the distribution of interest rates under different pricing regimes and economic environments. The baseline is GSE pricing with asymmetric information on risk types between lenders and borrowers.

Appendix A Homogeneous Housing Preference

The independent and identically distributed repayment shocks in my model are designed to capture the unobserved income shocks. However, a borrower's default and purchase decisions could depend on individual-specific income trends, which are unavailable to researchers. The non-stationary income could come from saving, wage growth, or unemployment. Under these circumstances, the unobserved income process may give rise to bias estimates in the housing preference λ. For example, a lower default rate is driven by income growth through saving rather than a higher housing preference. A borrower default may not be attributed to lower housing preference; it may instead be due to unemployment.

To account for the potential bias, I conduct analyses to ensure my welfare results are robust in the absence of the individual-specific trend in liquidity. I consider homogeneous housing preferences with each borrower's housing preference being the average of my parameter estimates. This measure assumes that the housing preference average does not change. The borrower's repayment and purchase decisions remain heterogeneous. I assume the heterogeneous default risks are driven by borrowers' financial sophistication (i.e., propensity for saving or employment risks), which is orthogonal to the housing preference. I calculate the welfare results from counterfactuals using this alternative measure of housing preference (see Table 14).

In reality, the housing preference should be heterogeneous to a certain degree, so this extreme case gives a conservative measure (a lower bound) of welfare changes. The welfare gain stemming from full risk-based pricing amounts to 82.10 percent smaller than the baseline results. The welfare loss due to uniform pricing amounts to 63.57 percent smaller than the baseline results. The welfare effects are directionally similar even under the restrictive assumption of homogeneous housing preference.

Under GSE pricing, welfare increases as the level of competition up to the case of $2N$. Welfare decreases when the market is perfectly competitive. Though the quantitative magnitude change is non-trivial, the interaction effect is qualitatively similar.

Appendix B Two-Sample t-Test on Loan Characteristics

Table 15 shows the two-sample t-test of loan characteristics. Column (1) compares the markets with and without local APLs before the preemption in 2004. Many loan characteristics are statically different between markets with and without local APLs.

Column (2) compares the markets that have higher and lower shares of national banks in APL states before the preemption in 2004. I split the markets into two subgroups based on an indicator of whether the share of national banks in 2003 is above its median. The no cash-out refinance indicator is statistically different. I find no significant differences between the two samples in the available statistics for other loan characteristics.

Appendix C Estimation Details

I derive δ for each draw of λ_{in} using the first-order condition on a borrower's down payment decision in Equation 3,

$$\boldsymbol{\delta}_i(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x},p_{ij}) = \left(p_{\boldsymbol{ij}}\sum_{t=1}^{T}\beta^{t-1}\mathbb{E}\left[\frac{\partial V_i(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x},p_{ij},t)}{\partial D}\right]\right)\times\left(D_i(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x},p_{ij})\right)^{-1}.$$

I assume that lender-purchase-specific shocks, $\boldsymbol{\varepsilon}$, follow type 1 extreme value distribution. Given a vector of lender preference, $\boldsymbol{\alpha}$, the fraction of borrowers who purchase the mortgage from lender j for each draw of λ_{in} is a closed-form solution:[49]

$$\varphi_i(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x},\boldsymbol{p_i}) = \frac{\exp\left(\max_j U_{ij}(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x})\right)}{\sum_{j=1}^{\mathbb{J}}\exp(U_{ij}(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x}))+\exp\left(\underline{U}_i(\boldsymbol{\Theta_0};\boldsymbol{x},0)\right)}.$$

The probability distribution of λ among the borrowers who purchase mortgages is given by

$$\rho_i(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x},\boldsymbol{p_i}) = \frac{\boldsymbol{\varphi}_i(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x},\boldsymbol{p_i})\boldsymbol{f}(\lambda_{in})}{\int\boldsymbol{\varphi}_i(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x},\boldsymbol{p_i})\boldsymbol{f}(\lambda_{in})d\lambda_{in}}.$$

[49] I estimate the following regression to predict the counter-offers of other lenders:

$$p_{ij} = \boldsymbol{x}^\top\boldsymbol{\beta_p}+\boldsymbol{\eta}_{hj}^{\boldsymbol{p}}+\boldsymbol{\kappa}_t^{\boldsymbol{p}}+e_i^p,$$

where η is market-lender fixed effects, κ is year fixed effects, and e is an error term.

The lender-specific preference is a function of observed market share, $\boldsymbol{\tau}$,

$$\alpha_j = \log(\tau_j) - \log(\tau_{\mathbb{J}}) + \mathbb{E}_{\{i,\lambda_{in}\}}\left[\Big(U_{ij}(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x}) - U_{i\mathbb{J}}(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x})\Big)\rho_i(\boldsymbol{\Theta_0};\lambda_{in},\boldsymbol{x},p_{ij})\right], \tag{12}$$

where $\mathbb{J}$ denotes small lenders. The preference for small lenders is normalized to zero. I then solve for $\boldsymbol{\alpha}$ by iterating Equation 12.[50]

Appendix D Parameters and Moment Restrictions

Table 16 outlines the moment restrictions used in the structural estimation. I discuss here the importance of the moment conditions to the identification of the relevant parameters. I group the moment conditions into four groups.

Group 1 includes the default probabilities for the loans that were not affected by the preemption (before 2004 or in the states without local APLs). I present the variation in the moment conditions of group 1 (average across the 30 periods) along with parameter γ in panel (a) of Figure 12, holding other parameters fixed.

Group 2 contains the interactions between default probability and time-varying housing prices. I plot the variation in the moment conditions of group 2 (average across the 30 periods) along with parameter μ in panel (b) of Figure 12, holding other parameters fixed. The first two groups' moments help to identify γ and μ.

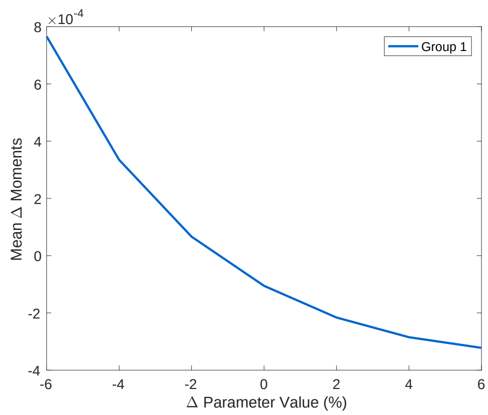
Group 3 covers the default choices for the loans that were affected by the preemption (after 2004 in the states with local APLs). I separately plot the variation in the moment conditions of group 3 and group 1 (average across the 30 periods) along with parameter σ in panel (c) of Figure 12, holding other parameters fixed. Changes in moments of group 1 and group 3 with respect to parameter σ have different slopes. σ helps to explain the increase in default rates associated with the regulatory decrease in interest rates.

Group 4 covers the interactions between the loan characteristics and group 1 – 3. This group identifies the parameters on loan characteristics in γ, λ, and σ with the same identification logic.

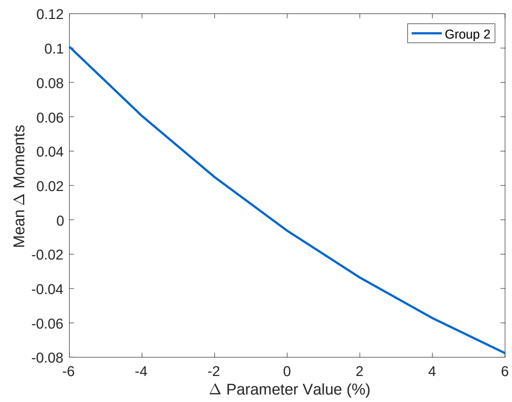
[50] The tolerance level is specified as $|\big(\log(\tau_j) - \log(\tau_{\mathbb{J}})\big) \times 10^{-5}|$.

Figure 12. The Relationship between Parameters and Moment Restrictions

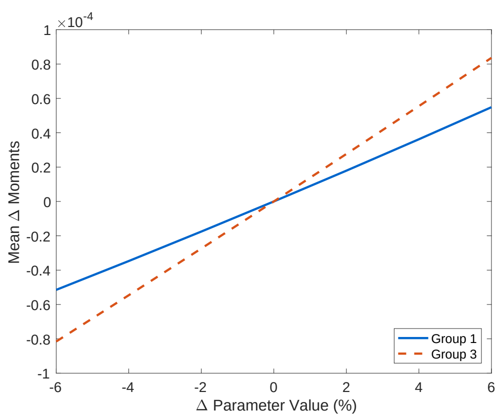
(11a) Parameter γ



(11b) Parameter μ



(11c) Parameter σ



Note: Figure 12 plots the average change of the moments relative to a particular parameter, holding other parameter estimates fixed.

Table 13: List of APL States

State	Date
Arkansas	July 16, 2003
California	July 1, 2002
Colorado	July 1, 2003
Connecticut	January 1, 2002
District of Columbia	May 7, 2002
Georgia	March 7, 2003
Illinois	January 1, 2004
Indiana	January 1, 2005
Maryland	October 1, 2002
Massachusetts	November 7, 2004
Michigan	December 23, 2002
Minnesota	January 1, 2003
New Jersey	November 27, 2003
New Mexico	January 1, 2004
New York	April 1, 2003
North Carolina	July 1, 2000
Rhode Island	December 31, 2006
South Carolina	January 1, 2004
Texas	September 1, 2001
West Virginia	June 8, 2000
Wisconsin	February 1, 2005

Note: Table 13 presents the list of states that implemented anti-predatory lending laws. The left column lists the state names, the right column lists the corresponding implementation date.

Table 14: Robustness Analysis - Homogeneous Housing Preference

	Uniform Pricing	Risk-Based Pricing			GSE Pricing	
	$N = 1$	$N = 1$	$N = 2$	$N \to \infty$	$N = 2$	$N \to \infty$
Panel A: Welfare Change to Baseline						
Mean Δ Interest Rate	0.000	0.065	0.024	-0.020	-0.041	-0.085
Mean Δ Default Loss ($)	43.406	-843.218	-758.258	-652.063	186.349	401.294
Mean Δ Lender Profit ($)	413.079	8,068.516	1,228.934	-6,388.580	-7,663.553	-16,114.002
Mean Δ Borrower Surplus ($)	-412.100	-10,174.920	-2,332.001	6,106.992	8,171.818	16,950.444
Mean Δ GSE Profit ($)	-14.625	2,199.333	1,832.566	1,447.906	-368.143	-759.490
Mean Δ Deadweight Loss ($)	13.647	-92.928	-729.498	-1,166.318	-140.122	-76.953
Δ # of Accounts (10k)	5.284	-76.842	-27.404	22.636	43.816	87.836
Panel B: Percentage Change to Baseline						
Mean Δ Interest Rate	0.000	4.463	1.659	-1.378	-2.796	-5.831
Mean Δ Default Loss	3.295	-64.014	-57.564	-49.502	14.147	30.465
Mean Δ Lender Profit	2.563	50.071	7.626	-39.646	-47.558	-100.000
Mean Δ Borrower Surplus	-0.336	-8.294	-1.901	4.978	6.662	13.818
Mean Δ GSE Profit	0.133	-20.009	-16.672	-13.172	3.349	6.909
Mean Δ Deadweight Loss	1.202	-8.183	-64.238	-102.703	-12.339	-6.776
Δ # of Accounts	0.241	-3.509	-1.252	1.034	2.001	4.011

Note: Table 14 presents the welfare results of different pricing regimes and various cases of market concentration. Mean changes in the statistics are calculated using total changes in the statistics divided by sample size (22 million). Panel A presents the welfare changes in dollar value relative to the baseline under homogeneous housing preference. Panel B presents the welfare changes in percentage relative to the baseline under homogeneous housing preference.

Table 15: Two-Sample t-Test on Loan Characteristics

	(1)			(2)		
	Mean		P-Value	Mean		P-Value
	APL = 0	APL = 1		HIGH = 0	HIGH = 1	
log(Credit Score)	6.574	6.573	0.128	6.543	6.575	0.320
log(Income)	7.709	7.838	0.000	7.814	7.824	0.827
log(Loan Amount)	11.554	11.697	0.000	11.664	11.674	0.882
log(Down Payment)	10.112	10.410	0.000	10.386	10.383	0.961
Insurance Percentage	8.012	6.635	0.000	6.665	6.560	0.723
Loan Purpose - Home Purchase	0.423	0.405	0.011	0.408	0.400	0.524
Loan Purpose - No Cash-Out Refinance	0.340	0.341	0.881	0.327	0.357	0.000
Property Type - Condo	0.000	0.007	0.005	0.005	0.008	0.532
Property Type - Co-Op	0.015	0.011	0.006	0.010	0.012	0.496
Property Type - Manufactured Housing	0.058	0.056	0.697	0.053	0.059	0.430
Property Type - Planned Unit Development	0.898	0.875	0.006	0.871	0.876	0.711
Occupancy Status - Second	0.042	0.043	0.791	0.041	0.047	0.380
Occupancy Status - Investor	0.042	0.046	0.070	0.045	0.046	0.980

Note: Table 15 compares the means of two samples. Column (1) compares the markets with and without local APLs before the preemption in 2004. Column (2) compares the markets with higher and lower market share of national banks in APL states before the preemption in 2004. High is an indicator of whether the share of national banks in 2003 was above its median.

Table 16: Summary of Moment Restrictions

Group	Moment Restrictions	Dimension
1	$\mathbb{E}_{\{i\}}[\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i \mid \text{APL}_i \times \text{POST}_i = 0]$	1×30
2	$\mathbb{E}_{\{i\}}[\boldsymbol{V} \circ (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i) \mid \text{APL}_i \times \text{POST}_i = 0]$	1×30
2	$\mathbb{E}_{\{i\}}[\boldsymbol{V} \circ (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i) \mid \text{APL}_i \times \text{POST}_i \times \text{HIGH}_i = 1]$	1×30
2	$\mathbb{E}_{\{i\}}[\boldsymbol{V} \circ (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i) \mid \text{APL}_i \times \text{POST}_i \times (1 - \text{HIGH}_i) = 1]$	1×30
3	$\mathbb{E}_{\{i\}}[\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i \mid \text{APL}_i \times \text{POST}_i \times \text{HIGH}_i = 1]$	1×30
3	$\mathbb{E}_{\{i\}}[\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i \mid \text{APL}_i \times \text{POST}_i \times (1 - \text{HIGH}_i) = 1]$	1×30
4	$\mathbb{E}_{\{i\}}[\boldsymbol{x} \otimes (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i) \mid \text{APL}_i \times \text{POST}_i = 0]$	5×30
4	$\mathbb{E}_{\{i\}}[\boldsymbol{x} \otimes (\boldsymbol{V} \circ (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i)) \mid \text{APL}_i \times \text{POST}_i = 0]$	5×30
4	$\mathbb{E}_{\{i\}}[\boldsymbol{x} \otimes (\boldsymbol{V} \circ (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i)) \mid \text{APL}_i \times \text{POST}_i \times \text{HIGH}_i = 1]$	5×30
4	$\mathbb{E}_{\{i\}}[\boldsymbol{x} \otimes (\boldsymbol{V} \circ (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i)) \mid \text{APL}_i \times \text{POST}_i \times (1 - \text{HIGH}_i) = 1]$	5×30
4	$\mathbb{E}_{\{i\}}[\boldsymbol{x} \otimes (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i) \mid \text{APL}_i \times \text{POST}_i \times \text{HIGH}_i = 1]$	5×30
4	$\mathbb{E}_{\{i\}}[\boldsymbol{x} \otimes (\boldsymbol{D}_i - \hat{\boldsymbol{D}}_i) \mid \text{APL}_i \times \text{POST}_i \times (1 - \text{HIGH}_i) = 1]$	5×30

Note: Table 16 summarizes the moment conditions used in the structural estimation. $\boldsymbol{D}$ is a sample size-by-30 matrix with the repayment decisions of each loan from 1 to 30 quarters after loan origination. $\boldsymbol{V}$ is a sample size-by-30 matrix with the (time-varying) housing prices of each loan from 1 to 30 quarters after loan origination. HIGH is an indicator of whether the share of national banks in the year 2003 was above its median.