The Economics of Mortgage Lending Regulations

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA

BY

Konstantin Golyaev

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Advisor: Patrick L. Bajari

January 2012

Acknowledgements

I am indebted to my advisor Patrick Bajari for his constant support and encouragement. I would like to thank my committee members Amil Petrin, Thomas Holmes and Jeffrey McCullough for their suggestions that greatly improved the dissertation. Special thanks go to Richard Todd, without whom this project would never have been possible. I am grateful to David Wiczer for his C code for some of my computations and to Brian Adams for his help with the GIS software. I thank Michael Grover, Ron Feldman, Joseph Schneberger, Jason Schmidt and Daniel Rozycki for their assistance with the HMDA and CRA data, Paul Calem, Glenn Canner and Robert Avery for their suggestions, Harry Paarsch, Minjung Park, Kyoo Il Kim, Joel Waldfogel, Gautam Gowrisankaran, Robert Town, Jesse Shapiro, Christina Marsh, Andrea Szabo, Itai Sher, Thomas Youle, Nadezda Golyaeva, Michael Reszler, Kevin Williams, Alexander Stepanov, Sunil Kishore, Lydia Cheung, Le Chen, Mahmoud Elamin and all participants of the University of Minnesota Applied Microeconomics Workshop. The usual disclaimer applies.

To my family. All of you.

Abstract

In this dissertation, I consider various aspects of the U.S. residential mortgage lending market in 2005. In particular, I examine how existing regulations may have contributed to the mortgage default crisis that began in early 2007.

The first chapter of the dissertation is concerned with the Community Reinvestment Act (CRA). The CRA is a federal lending regulation that creates incentives for depository institutions to lend in low- and moderate-income areas. Only a subset of market areas is closely monitored by the regulators. I exploit this CRA enforcement mechanism to identify its effect on the banks' loan approval decisions. I employ a novel nonlinear Bayesian Instrumental Variables method to quantify the above effect while admitting unobserved heterogeneity among mortgage lenders. I find that, other things equal, loans in closely monitored areas have a 21.7 percent higher average chance of being approved. This implies that more than 327,000 extra loans originated in 2005 in California and suggests that banks' responses to the CRA enforcement mechanism sharply contradict the original CRA goals of providing credit in all eligible neighborhoods. Namely, CRA-induced incentives led banks to issue substantially more loans to marginal borrowers in monitored areas.

The second chapter of the dissertation explores the degree of strategic interactions among mortgage lenders and how these interactions differ depending on the regulatory agency. Conventional economic wisdom suggests that competition among mortgage lenders will result in overall welfare improvements. Recent theoretical research challenges this wisdom. Using the data concerning home mortgage loan applications, I test the "race-to-the-bottom" hypothesis that competition among lenders causes them to relax lending standards. I exploit the recently developed structural methods of estimating static games with incomplete information

to identify how lenders form beliefs about the actions of their competitors. I find strong evidence supporting the “race-to-the-bottom” story among all types of mortgage lenders, with the exception of those regulated by the Federal Reserve System. Thus, my results provide a partial explanation for the subprime mortgage collapse of the early 2007.

Contents

Front Matter **i**

Acknowledgements ii

Dedication ii

Abstract iv

List of Tables **viii**

List of Figures **x**

1 The Impact of the Community Reinvestment Act on the Home Mortgage Lending Industry **1**

1.1 Introduction 1

1.2 Background and Literature 9

1.2.1 The Community Reinvestment Act 9

1.2.2 Related Literature 14

1.3 Model 17

1.4 Estimation 22
1.4.1 Identification Strategy 22
1.4.2 Nonlinear Bayesian IV 33
1.5 Data 37
1.5.1 Data Sources 37
1.5.2 Dataset Construction 39
1.5.3 Preliminary Evidence 41
1.6 Estimation Results 48
1.6.1 Linear Probability Model Estimates 48
1.6.2 Bayesian IV 52
1.6.3 Discussion 57
1.7 Conclusion 60

2 Race-to-the-Bottom In Home Mortgage Lending 62

2.1 Introduction 62
2.2 Industry Overview 66
2.3 Estimation 69
2.3.1 Estimation Method 69
2.3.2 Implementation 73
2.4 Data 76
2.4.1 Data Sources 76
2.4.2 Dataset Construction 77

2.4.3 Preliminary Evidence 78
2.5 Estimation Results 81
2.6 Conclusion 85

Bibliography **86**

A Implementing Bayesian Instrumental Variables Estimation **93**
A.1 Step 1. Updating y^*. 94
A.2 Step 2. Updating β and γ. 95
A.3 Step 3. Updating δ. 96
A.4 Step 4. Updating $\mathbf{\Sigma}$. 98

B Sampling from the Truncated Normal Distribution **99**
B.1 Inverse CDF Sampling 100
B.2 Mixed Rejection Sampling 100
B.2.1 Exponential Rejection Sampling 102
B.3 Algorithm 103

C Supplementary Evidence **105**

List of Tables

1.1 Descriptive Statistics on Loan Applications 44
1.2 Lending Patterns and Branches by Lender, “Closest-Two Tracts” Sample 46
1.3 Decomposition of Selected Average Loan Characteristics By Lender, “Closest-Two Tracts” Sample 47
1.4 Linear Probability Model for Loan Approval 49
1.5 2SLS Linear Probability Model for Loan Approval 51

2.1 Top Lenders, By Regulatory Agency 79
2.2 Summary of Applications, By Regulatory Agency 80
2.3 Share of Applications Approved 83
2.4 Approximated Marginal Effects for Lenders’ Beliefs 84

C.1 Two Stage Least Squares: First Stage Results 106
C.2 Decomposition of Average Loan Characterstics By Lender, All Observations 109

C.3 Decomposition of Average Loan Characterstics By Lender, Lower-Income Tracts 110
C.4 Decomposition of Average Loan Characterstics By Lender, "Closest-Five Tracts" Sample 111
C.5 Decomposition of Average Loan Characterstics By Lender, "Closest-Two Tracts" Sample 112
C.6 Lending Performance Ratings 113

List of Figures

1.1 Model Illustration 20

1.2 An Illustration of Washington Mutual Bank's Assessment Area . . 27

1.3 Matching Algorithm: An Example from the Greater Los Angeles Area 28

1.4 Preliminary Evidence 41

1.5 Matching Results 43

1.6 MCMC Results 53

1.7 The CRA Marginal Effect 55

2.1 Mortgage Lending Industry: A Schematic Overview 66

C.1 MCMC Convergence for β 107

Chapter 1

The Impact of the Community Reinvestment Act on the Home Mortgage Lending Industry

1.1 Introduction

For the average American, a mortgage loan represents the only chance of purchasing a home. Thus, home mortgage lending is a sizeable industry which is subject to a number of federal regulations. The Community Reinvestment Act (CRA) is a mortgage lending regulation that encourages depository institutions[1] to

[1] The most common types of depository institutions are banks and thrifts. This paper uses the term "banks" for all of them.

expand what is considered "safe and sound" lending (including mortgage lending) in low- and moderate-income areas.[2] Recently the CRA has received considerable attention for its possible contribution to the ongoing mortgage default crisis.

The last decade saw an unprecedented increase in mortgage originations, particularly with the widespread proliferation of so-called subprime mortgages. Such loans were usually made at higher-than-normal rates to people with weak credit histories. In early 2007, the mortgage industry experienced an abrupt increase in delinquencies and, later, foreclosures, particularly in its subprime segment. The ensuing turbulence in the housing market triggered broader economic turmoil, which is linked to the latest worldwide recession.[3]

Some observers have speculated that CRA-induced incentives forced financial institutions to weaken loan evaluation standards, extending too much credit to high-risk individuals.[4] Several studies have sought to measure the causal relationship between the CRA and extra mortgage loan approvals, including Avery, Calem, and Canner (2003), Berry and Lee (2007), and Bhutta (2008). These researchers used regression discontinuity techniques to identify the banks' marginal willingness to approve loan applications from CRA-eligible areas as compared to applications from the CRA-ineligible census tracts. This research design is insightful, but sub-

[2] The CRA-eligible census tracts (also called "lower-income tracts") are those with median family income less than 80 percent of the median family income of the metropolitan statistical area in which this tract belongs.

[3] Demyanyk and Van Hemert (2009) and Bajari, Chu, and Park (2009) have considered these events.

[4] Liebowitz (2009) advances this view, while Bair (2008) provides a contrasting interpretation of events.

ject to a potential criticism.

The important feature, not discussed much in these studies, is the CRA enforcement mechanism: specifically, CRA requires its subjects to define *assessment areas* (AAs). These areas serve as proxies for markets in which banks undertake the majority of their business. When regulators evaluate a bank's operation under the CRA, a substantially higher weight is placed on the bank's performance within its assessment areas. At the same time, banks are expected to apply analogous evaluation standards to similar loan applications. Thus, two identical applications from the neighboring CRA-eligible census tracts, where one belongs to a bank's AA and the other does not, are expected to be processed in a unified manner.[5]

In this chapter, I improve on the existing studies by explicitly quantifying how banks react to the incentives that the CRA monitoring scheme creates. The main challenge that most researchers face is that the data provide no information concerning applicants' credit scores—the major determinants of loan approval. Failure to account for this key unobserved factor suggests an omitted variable problem. To address this problem, I employ two different identification strategies. The first strategy provides a lower bound on the effect of interest, while the second strategy yields the upper bound. Therefore, even though both approaches only address the problem partially, the true magnitude of the effect is somewhere between the two estimates that I obtain.

[5] 12 U.S.C. §2901(a)(1) explicitly states that "regulated financial institutions are required by law to demonstrate that their deposit facilities serve the convenience and needs of the communities in which they are chartered to do business," which says nothing about the assessment areas. Banks are usually charted to do business outside their AAs as well.

Unobservable credit scores pose a fundamental challenge to estimating the effect of being inside the CRA assessment area on loan approval decision. The observed definitions of AAs are the outcomes of a negotiation process between the banks and their regulators. This suggests that the estimate for the assessment area effect will be biased. In my application, predicting the sign of the bias is not trivial. Econometric theory suggests that the bias will be a product of two factors: the effect of credit scores on loan approval and the covariance between the credit scores and assessment area inclusions. It seems obvious that higher credit scores increase the chance of loan approvals. Thus, the sign of the bias would depend on whether applicants from within AAs have higher credit scores than those from outside AAs.

Ex ante, there are equally compelling arguments for the above effect being either positive or negative. Banks prefer their AAs to include only areas with the most creditworthy borrowers, while regulators try to ensure the definitions do not arbitrarily exclude low- and moderate-income tracts. If the regulators have more bargaining power, banks have to include a large number of lower-income tracts in their assessment area that they would have included otherwise. In this case, AAs are likely to include borrowers with weak credit scores, and the omitted variable will negatively bias the estimate of interest. The raw evidence from the data hints at this pattern: the difference in approval rates inside and outside AAs for the whole dataset is about -2 percentage points. A naïve interpretation of this number would be that the CRA monitoring rules force banks to approve fewer loans in the areas where they are closely watched; however, the omitted variable bias is likely driving this result.

My first solution borrows from the program evaluation literature: specifically, I exploit the data concerning assessment areas boundaries to obtain identification. This idea has previously been used by Holmes (1998) and Black (1999), among others. Thus, for every bank, I construct a matched sample of CRA-eligible census tracts that are in close proximity and are extremely similar to each other in terms of all observable socio-economic characteristics.[6] I then choose a subset of these similar census tracts such that roughly half of them belong to the assessment areas, and the others do not. Within this selected sample of observations, the assumption that the unobservable components that enter the loan approval decision are fairly homogenous across applicants seems reasonable, provided that all the observable components are selected to be quite similar.

By comparing the loan approval rates inside and outside the assessment areas in this matched sample, I am able to interpret the observable difference as the impact of the CRA monitoring rules. This number is a lower bound to the true size of the effect of interest. The data suggests the omitted variable bias is likely to be negative, and the matching procedure may fail to solve the problem completely.

My second approach to address the unobservable credit score problem is to employ an instrumental variables (IV) strategy. I construct a measure of distance from the collateral property to the nearest bank branch. The distances from customers to branches are quite small in my sample. It is also common for people to

[6] I do not require the tracts to be perfectly adjacent to each other, but distance between the tracts is one of the characteristics that determine how “similar” a pair of tracts actually is.

apply for loans at several banks simultaneously, so the information content of an application who is from further away is, for the bank, minuscule. In most cases, the closest branch is the one located within the assessment area, so the distance measure is correlated with the assessment area status. This makes it a valid and relevant instrumental variable.

If banks choose their branch locations strategically in the areas where people have higher credit scores, then the IV estimate will also be biased. It is possible, however, to predict the sign of the bias. The bias will be proportional to the ratio of two covariances: between the instrument and the credit score, and between the instrument and the endogenous variable. By construction, the latter covariance is negative: the smaller the distance from a borrower to a branch, the more likely this borrower comes from within the bank's AA. If the former covariance is negative as well—that is, branches are located in neighborhoods with more creditworthy borrowers—then the ratio of these effects will be positive. Consequently, the IV estimate will be biased upwards, overestimating the true size of the effect in question.

To implement the second approach, one could use the method of two-stage least squares (2SLS) applied to a linear probability model. However, a linear probability model is only an approximation of the true nonlinear model in case of a limited dependent variable. Blundell and Powell (2004) demonstrate how this approximation can sometimes be quite imprecise. This prompts me to employ the nonlinear Bayesian IV method that allows me to address the issue of endogeneity in a probit setting. Using the probit model avoids non-sensible predictions such as probabilities outside the unit interval. This also allows me to account for the heterogeneity across mortgage lenders by making model coefficients random. Finally, Bayesian methods have attractive computational properties.

I find that the CRA-induced incentives have a strong impact on banks' loan approval decisions. The matching estimator suggests that loans from within AAs have a 10.1 percent higher chance of getting approved, other things equal. The Bayesian IV estimator demonstrates that on average, there is a 21.7 percent higher chance that a loan application will be approved if the corresponding collateral property is located within a bank's assessment area. The above discussion indicates that the true magnitude of the CRA effect is bounded by these two numbers. These results imply that the CRA enforcement mechanism caused banks to approve between $102,000$ and $328,000$ extra mortgage loan applications in 2005 in California, and given the average loan size of $\$276,000$, this amounts to up to $\$90.5$ billion of additional lending. Moreover, to the extent that the regulators expect the banks to respond to CRA-induced incentives only in the areas that get monitored—i.e., assessment areas—these are the true estimates of the CRA effect.

Two interesting implications emerge from my findings. First, the CRA has created incentives for banks to move down along the demand curve for mortgage loans. Some of these CRA-induced loans seem to have been given out to borrowers who would not have been able to qualify for loans in the absence of this regulation. Second, the enforcement procedure used by the regulators for the assessment of banks' performance under the CRA gives rise to a large disparity in the way banks treat borrowers in lower-income areas. Specifically, borrowers from within the banks' assessment areas enjoy a substantially higher chance of obtaining mortgage credit.

One way to understand the last result is through a prism of the costly state verification model introduced in Townsend (1979) and discussed in detail in Laffont and Martimort (2002). In this model, a principal outsources production to the agent who has uncertain productivity, which can be revealed via a costly audit procedure. In equilibrium, the principal commits to a stochastic audit schedule, which induces the agent to reveal his true productivity. This situation is inefficient ex post, because the audit wastes resources, but without the ex ante commitment, contract compliance is not incentive compatible for the agent. Thinking of the regulator as the principal and the bank as being the agent, I see two possible interpretations. If regulators really want the CRA to have an impact only inside AAs, then the enforcement mechanism makes perfect sense because it induces banks to lend more inside their AAs. If, however, regulators see the CRA as an instrument that promotes access to credit in all lower-income areas, the existing enforcement mechanism may suggest regulators are unable to commit to the ex-post inefficient audit. The absence of a commitment device then renders the desired outcome infeasible, which is in perfect agreement with the costly state verification model.

The remainder of this chapter is organized as follows. Section 1.2 provides an overview of the mortgage origination industry and outlines the relevant regulations, notably the Community Reinvestment Act. Section 1.3 presents a stylized economic model that predicts that the CRA creates incentives for the banks to approve more loans within their assessment areas. I discuss my estimation strategy in Section 1.4. Section 1.5 goes over the data sources that I draw upon in the paper and details the construction of the final dataset. Results are presented in Section 1.6, and Section 1.7 concludes.

1.2 Background and Literature

1.2.1 The Community Reinvestment Act

The home mortgage loan process begins with a person's decision to purchase a home, be it a house or a condominium, and so forth. Few buyers pay the price in a single installment. Instead, they choose to borrow most of the money (usually at least eighty percent) from a financial institution using the home in question as collateral. I refer to this type of borrowing as a mortgage loan.

Mortgage lending is a huge industry with many participants. In 2005, some 35.5 million loan applications were recorded, and about 60 percent of those applications were approved. In California, the corresponding numbers were 5.45 million and 51.2 percent, respectively.[7] The average loan amount in 2005 was $\$183,000$ across the country, and $\$276,000$ in California. The complete structure of the industry is quite complex. I focus on the very first stage of the overall process, which

[7] As I discuss in Section 2.4.1, these numbers potentially include applications from the same people to multiple banks, so the total number of loan applicants was likely lower.

is the loan approval decision by lenders, and I treat the actions of all other industry actors as given.

The industry is subject to a number of federal regulations. In the past, a major concern of lawmakers had been to fight discrimination practices, especially along racial and ethnic dimensions. With this goal in mind, the Home Mortgage Disclosure Act (HMDA) and the CRA were passed.

The HMDA is implemented under the U.S. Federal Reserve Board's Regulation C. Initially passed in 1975, its main purpose was to fight discrimination in mortgage lending. Regulation C requires almost every application for a home mortgage loan to be recorded, and reported to the Federal Financial Institutions Examination Council (FFIEC) at the end of the year. This is the main source of data for this study; I provide a detailed description of it in Section 2.4.1.

The CRA was passed in 1977. Its initial goal was to discourage "redlining" practices that had previously been in place at many depositary institutions. Redlining amounted to blanket refusals by the banks to lend in certain areas. This practice originated with the Federal Housing Administration (FHA) in the 1930s. The Home Owners' Loan Corporation (HOLC) created the "residential security maps" for the FHA. These maps were used by lenders for years afterwards to withhold mortgage loans from neighborhoods that were perceived as "unsafe". (Avery, Bostic, and Canner, 2000) note that the act "encouraged commercial banks and savings associations to meet the needs of borrowers in all segments of their communities, including low- and moderate-income neighborhoods".

Two key points must be understood concerning the CRA. First, every census tract can be either CRA-eligible or CRA-ineligible. The act spells out the exact criterion that defines eligibility status: the median family income of people living in a given census tract has to be less than 0.8 of the median family income in the greater area that the tract belongs to (in the vast majority of cases, this is the Metropolitan Statistical Area, or the MSA). Only depository institutions are subject to the CRA. The original text of the act is that "regulated financial institutions are required ... to demonstrate that their deposit facilities serve the ... communities in which they are chartered to do business."[8]

Second, for the purposes of monitoring banks' performance under the act, all respondents are required to define their *assessment areas*. A bank's CRA assessment area (AA) has to be a geographic area that is delineated by the bank. The delineation has to be approved by its regulatory agency, which will later use it in evaluating the bank's record of meeting the credit needs of its community.[9] This is in contrast with the excerpt cited above, because banks are, as a rule, charted to do business outside their assessment areas as well as inside.

When a bank proposes an assessment area that looks like it may violate some of the conditions listed below, regulators usually choose to talk to the bank and get a justification for any irregularities in the proposed definition, as opposed to citing it with a violation. An assessment area consists of one or more contiguous political

[8] 12 U.S.C. §2901(a)(1).

[9] Every depositary institution is supervised by one of four federal regulatory agencies. These are the Office of the Comptroller of the Currency (OCC), the Federal Reserve System (FRS), the Federal Deposit Insurance Corporation (FDIC) and the Office of Thrift Supervision (OTS).

subdivisions, such as counties or cities. It must include neighborhoods in which the bank has its main office and branches, as well as the surrounding census tracts in which it originates a substantial portion of its loans. An AA must consist only of whole census tracts, may not reflect illegal discrimination, and may not arbitrarily exclude low- and moderate-income tracts. A bank may adjust the boundaries of its AA to include only the portion of a political subdivision that it can reasonably expect to serve (CRA Reference, 2005).

At the end of each year, a bank's performance under the CRA is evaluated, and the bank receives a CRA compliance rating. The possible ratings range from "Outstanding" (the best) to "Substantial Noncompliance" (the worst), as Table C.6 in Appendix C shows. The CRA compliance record is used by regulators when a bank seeks to expand through merger or acquisition or when opening of a new branch. Also, the CRA ratings are public information, which can be accessed by community activist groups. Therefore, banks can be expected to place considerable value on having a good CRA record, and the majority of banks end up receiving "Outstanding" or "Satisfactory" ratings.

A bank and the regulators have conflicting incentives on the assessment area composition. An ideal situation from the bank's perspective is to construct an assessment area that includes the most creditworthy lower-income borrowers. The bank would then be able to originate a fair amount of loans in lower-income areas, thus complying with the CRA, and yet be fairly certain about the quality of those loans. However, assuming that the regulators would like to see more lending in all lower-income areas, it is reasonable to expect that regulators would like to have the bank include as many of these areas in its AA as possible. The resulting composition of the AA that emerges from the bank-regulator negotiations is therefore a function of the relative bargaining powers of the participating sides. When the regulator has considerably more bargaining power, I would expect the final AA definitions to include a fair number of tracts that the banks might not have included on their own.

Several other regulations affect the loan approval decisions made by banks. The Fair Housing Act (FHA) outlaws the explicit usage of the following factors as inputs into the decision-making process: race, color, national origin, religion, sex, familial status, or disability.

Mortgage brokers are separate, but important, party that participates in the lending process. Mortgage brokers are the intermediaries between a borrower and a lender who facilitate both the loan application and the origination processes. According to Kleiner and Todd (2007), in 2004, over $53,000$ mortgage broker firms were operating in the U.S., and they were partially involved in the origination of more than two-thirds of mortgages in that year. As my discussions with the industry experts suggest, banks tend to rely more on mortgage brokers in marketing their services outside their assessment areas.

The interesting feature of all CRA compliance tests is that there is no predetermined quotas in the rules. Given the bank size, the definitions of its assessment areas, and its activity during the year, a number of criteria are used to evaluate a bank's performance under the CRA. The lending test, which is of primary interest to me, involves the total number and amount of approved loans by a bank. It primarily rewards the banks for lending more in the lower-income parts of their AAs. Table C.6 in Appendix C, adapted from the CRA Reference (2005), provides other criteria relevant to the bank examiners and details how the bank's performance is evaluated. At the end of each fiscal year, regulators look at the number of CRA-eligible loans the bank had originated and compare it with a number of benchmarks. These include the number of loans this bank originated in previous years and the average number of loans originated by other banks this year in nearby areas. Thus, a bank can never be sure that it has originated "enough" loans to be perceived as compliant. Other tests are also applied, but the lending test carries the most weight in determining a bank's final CRA rating, which then becomes part of the supervisory record for that bank.[10]

1.2.2 Related Literature

This dissertations builds on the findings from several different fields. First, the question of the existence of a causal relationship between the CRA and mortgage loan originations has been addressed previously by several researchers. Most of them employ a form of the regression discontinuity approach to identify the above effect. Only the lower-income census-tracts are CRA-eligible, which allows for a

[10] As noted by Berry and Lee (2007), the lending test accounts for 50 percent of the overall rating, and "no bank can receive an overall rating of "Satisfactory" ... if it does not receive a rating of at least "Satisfactory" on its lending test".

quasi-random experiment. These studies consider a pair of census tracts that are identical except for the fact that the income ratio is 79.9 percent in one of those tracts and 80.1 percent in the other. The argument is that the difference in the number of loans approved between these tracts has to be driven by the CRA.

Avery, Calem, and Canner (2003) were the first to use a variation of this identification strategy. They estimated regressions that explained the changes in the amount of lending that occurred in census tracts just above the threshold of CRA-eligibility in 1990, and then used the estimated equations to predict changes in outcomes for a cohort of census tracts that were just below the same threshold in 1990. Their findings are inconclusive because the results do not pass several falsification tests. Bhutta (2008) sets up a regression discontinuity around the same cutoff and finds a moderately-sized effect of about 4-5 percent higher approval volumes between 1994 and 2005. The question of CRA enforcement mechanism, however, was not the primary focus of these authors.

Joint Center for Housing Studies (2002) were the first study that incorporated information on banks' assessment areas in its estimation procedures. Three types of lending were examined: loans made by the institutions that are subject to the CRA inside and outside their assessment areas, plus the loans made by nondepository institutions (i.e., institutions that are not targeted by the CRA). The main drawback of this study is that it defines assessment areas as the whole county if the lender happens to have a branch within that county. The actual definitions of the CRA assessment areas are considerably more convoluted, and this fact potentially subjects the above results to the attenuation bias problem. Berry and Lee (2007) used an approach quite similar to that of Bhutta (2008) in trying to quantify the causal impact of the CRA. They exploited the same income ratio CRA-eligibility cutoff and they also look at the approval differences inside and outside the assessment areas. The main conclusion of their study is that it fails to uncover any significant effect; however, Berry and Lee (2007) adopt the definition of assessment areas employed by the Joint Center for Housing Studies (2002). This subjects the results of their study to the same potential criticism.

This paper also contributes to the nonlinear IV literature. The approaches to dealing with endogeneity in a nonlinear model have been previously addressed by various studies including Berry, Levinsohn, and Pakes (1995), Geweke, Gowrisankaran, and Town (2003), Blundell and Powell (2003), and, more recently, Gandhi, Kim, and Petrin (2010). My approach is most similar to that of Berry (1994): I take a nonlinear model with endogeneity and demonstrate how it can be converted into a linear one. The method builds on the Bayesian technique of *data augmentation*, a term originated by Tanner and Wong (1987). In the marketing literature, extensive use had been made of Bayesian methods, see Rossi, Allenby, and McCulloch (2006), the seminal source. These tools are quite attractive from a computational standpoint and allow for more flexibility in estimation under the same identifying assumptions, which motivates my choice of methodology.

Finally, several other researchers have considered the potentially unexpected consequences of regulations. Health economists have extensively studied the phenomenon of Medicare "upcoding": a situation when hospitals manipulate the diagnoses of their Medicare patients. The Prospective Payment System, introduced in 1988, tied the amounts of reimbursements hospitals get for treating Medicare patients to the severity of the patients' sicknesses. As Silverman and Skinner (2004) noted, the not-for-profit hospitals reported 10 percent more severe (which in this case means more generously reimbursed) cases of pneumonia and respiratory infections between 1989 and 1996. The similar number among for-profit hospitals during the same time was 23 percent. Dafny (2005) demonstrated that the problem has been exacerbated by the subsequent changes in policy that allowed for easier "upcoding."

More generally, empirical studies that quantify how people respond to incentives are numerous and include **?**paarsch-shearer-1999,paarsch-shearer-2000,paarsch-shearer-2007,paarsch-shearer-2009), Copeland and Monnet (2009), Misra and Nair (2009), Baker and Hubbard (2004), and many more.

1.3 Model

Below, I present a stylized model of the effect the CRA enforcement rules on a bank's loan approval decisions. Consider an economy made up of a single CRA-eligible census tract populated by a unit continuum of loan applicants. These applicants differ with respect to their type, which I call "credit quality," denoted by $\theta \in [0, 1]$. The values of θ can be interpreted as the probability that a given applicant will default on a loan, so a higher θ corresponds to less reliable applicants. I assume that applicant types θ are publicly observable.

Consider a single bank in this economy and where person applies for a mortgage loan at this bank. Assuming the total profit function from approving loans to borrowers in $[0, \theta]$ for the bank is

$$\pi(\theta) = \alpha\theta - \frac{1}{2}\theta^2, \tag{1.3.1}$$

where $\alpha \in (0, 1)$. This function implies that the profit from lending to a single person of type θ for the bank is $r(\theta) = \alpha - \theta$. Thus, α has the intuitive interpretation as the fraction (or amount) of borrowers that are profitable for the bank.[11] To keep the model as simple as possible, I specify the closed-form expression for the profit function. The model is capable of explaining how a risk-neutral bank without capital constraints decides on how many loans to approve.

The standard behavioral assumption of profit maximization implies

$$\theta_1 = \arg\max_{\theta \in [0,1]} \alpha\theta - \frac{1}{2}\theta^2,$$

and the solution is

$$\theta_1 = \alpha \tag{1.3.2}$$

[11] This is merely a convenient assumption; the implications of the model would not change were α not constant.

so, in equilibrium, the bank will approve applications from all borrowers with $\theta \leq \alpha$. A total of α loans will be approved, and the bank's profit will be

$$\pi\left(\theta_1\right) = \frac{1}{2}\alpha^2. \tag{1.3.3}$$

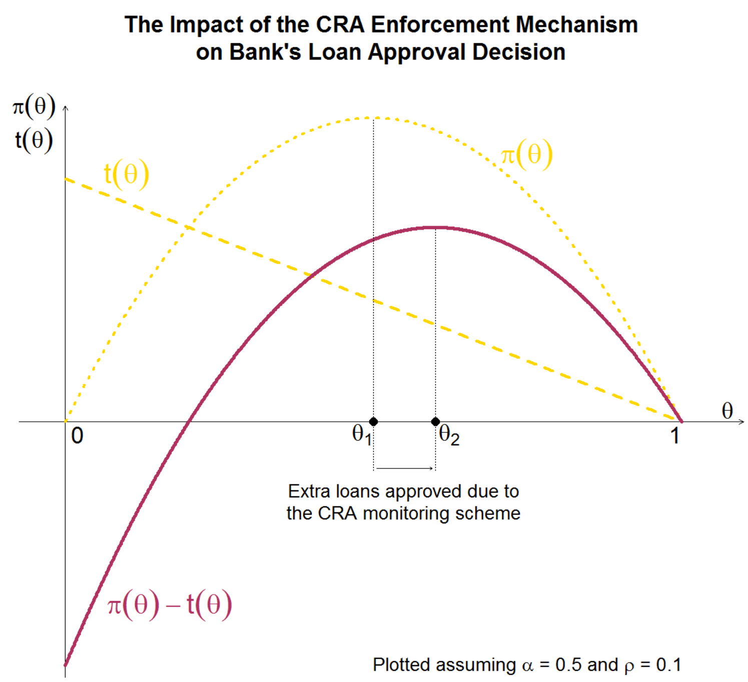
Now suppose that this census tract becomes the bank's assessment area. Under the CRA, the bank's regulator examines how many loans have been approved in the lower-income parts of the assessment area. As Section 1.2.1 describes, the Act does not set quotas on lending volume. However, if the number is deemed too low, then the bank may face restrictions on its future ability to expand. A lawsuit from an activist group is also possible.

In terms of the model, I assume the following CRA non-compliance cost function:

$$t\left(\theta\right) = \rho\left(1 - \theta\right). \tag{1.3.4}$$

This functional form is intuitive: if the bank approves applications from all borrowers, then the cost of non-compliance is zero. Otherwise the cost is strictly decreasing in the amount of loan approvals. The parameter ρ is a measure of the regulator's negotiating power. If ρ is high, then the regulator can punish non-complying banks severely. I assume $\rho \in (0, 1)$.

Figure 1.1: Model Illustration



The new profit-maximization problem for the bank is now

$$\max_{\theta \in [0,1]} \pi(\theta) - t(\theta),$$

and the solution is

$$\theta_2 = \alpha + \rho, \tag{1.3.5}$$

which implies

$$\theta_1 < \theta_2. \tag{1.3.6}$$

In other words, the need for compliance with the CRA forces the bank to approve more loan applications within its assessment area. Now the bank's profit is

$$\pi(\theta_2) - t(\theta_2) = \frac{1}{2}\alpha^2 + \frac{1}{2}\rho^2 - \rho + \alpha\rho.$$

It is easy to show that

$$\pi(\theta_2) - t(\theta_2) < \pi(\theta_1), \tag{1.3.7}$$

i.e. the regulation lowers the bank's profit.[12]

Figure 1.1 depicts the profit function when $\alpha = 0.5$ and $\rho = 0.1$. The dotted golden curve is $\pi(\theta)$ and the dashed golden line is $t(\theta)$.[13] The solid maroon curve is the difference of the two, $\pi(\theta) - t(\theta)$. The optimal increase in lending from θ_1 to θ_2 in response to the need to comply with the CRA is explicitly illustrated.

Since $\theta_1 = \alpha$ and $t(0) = \rho$, the comparative statics of the model are apparent: a decrease in ρ, which I interpret as a decrease in the negotiating power of the regulator, leads the lower maroon curve to approach the top golden one, as $\rho \longrightarrow 0$, θ_2 approaches θ_1 from above. In the next section, I discuss how I can estimate the average difference between θ_1 and θ_2.

[12] Technically, this can be violated if $\frac{1}{2}\rho > 1 - \alpha$, but this inequality cannot hold for $(\alpha + \rho) \in [0, 1]$.

[13] In greyscale, the golden-colored objects will looked lighter than maroon-colored ones.

1.4 Estimation

1.4.1 Identification Strategy

I build on the predictions of the model developed in Section 1.3.[14] Consider the following equation:

$$y_i^* = AA_i \cdot \beta_j + x_i' \gamma_j + cs_i \cdot \lambda_{j,1} + u_{i,1}, \tag{1.4.1}$$

where $i = 1, \ldots, I$ indexes loan applications, $j = 1, \ldots, J$ indexes banks, y_i^* is the unobservable loan approval "score," x_i is a collection of observable covariates that impact a bank's decision to approve the loan (borrower's income and demographics, loan amount), AA_i is an indicator for the mortgaged property being located inside a bank's assessment area. Instead of y_i^*, I observe the loan approval indicator $y_i = \mathbb{1}\{y_i^* \geq 0\}$. The two remaining terms in equation (1.4.1) are unobservable to me: cs_i is the applicant's credit score, and $u_{i,1}$ is the error term. It includes all other factors that are independent from AA_i, x_i and cs_i.

At the same time, assessment areas are not drawn at random, so there is the second equation:

$$AA_i = z_i' \delta_j + cs_i \cdot \lambda_{j,2} + u_{i,2}, \tag{1.4.2}$$

where z_i are some exogenous factors that determine AA assignment, which include

[14] Given the functional form assumptions in the model, direct analogy between parameter ρ and estimate of β_j can be drawn.

x_i. The credit scores also enter equation (1.4.2) because I expect this to be a major factor which banks and regulators negotiate about when the AA boundaries are determined. I expect the banks to apply different application evaluation standards and allow the coefficients to differ across banks.

The main object of interest is the estimate of β_j. If β_j is positive (as equation (1.3.6) predicts), then one can argue that the loan has a higher chance of getting approved if it is within bank j's AA.

From an econometrician's perspective, the error term in equation (1.4.1) is $\varepsilon_{i,1} \equiv cs_i \cdot \lambda_{j,1} + u_{i,1}$. In the data I have, there exists no information concerning applicants' credit scores cs_i. This is an extremely important factor which banks consider during the loan approval process. When an important covariate is missing from the regression, most standard methods (least squares regression, the method of maximum likelihood, and so on) are inconsistent due to the omitted variable bias. It is straightforward to show that

$$\text{Bias}\, \hat{\beta}_j = \operatorname*{plim}_{n \to \infty} \hat{\beta}_j - \beta_j = \lambda_{j,1} \cdot \lambda_{j,2}.$$

The bias will be a product of two terms, where the first term, $\lambda_{j,1}$, represents the effect of the credit score on loan approval, and is most likely positive—banks are more willing to approve loan applications to creditworthy borrowers, other things equal. The second term, $\lambda_{j,2}$, reflects the difference in credit scores of applicants across AA boundaries. If $\lambda_{j,2}$ is negative, then applicants from within AAs have lower credit scores on average than applicants from outside AAs. Such an outcome is likely when regulators have more bargaining power than banks. In this case, banks would have to include some census tracts in their AAs that they would not have included otherwise. A bank on its own is probably willing to include a tract with creditworthy borrowers and to exclude a tract populated with people who have weak credit scores. If regulators can have things their way, however, then one would expect $\lambda_{j,2}$ to be negative and $\hat{\beta}_j$ to be downward biased.

I employ two different identification strategies to reduce the size of the omitted variable bias. The first strategy is likely to produce an estimate that is biased downwards, which allows me to interpret its results as a lower bound on β_j. The second strategy may result in an upward biased estimate, suggesting an upper bound interpretation. Together, these estimates provide an interval that should contain the true value of β_j.

The first approach treats assessment area boundaries as predetermined, i.e. I do not attempt to estimate equation (1.4.2). Moreover, I use a linear probability model which is an approximation to the true nonlinear model that predicts the loan approval probability. This forces all coefficients in (1.4.1) to be the same across banks.[15] The equation I estimate is

$$y_i = AA_i \cdot \beta + x_i' \gamma + \varepsilon_{i,1}, \tag{1.4.3}$$

[15] Alternatively, I could estimate a separate equation for each bank j, but this would require "enough" observations from every bank in the sample.

and the essence of this identification strategy is to find a subset of data where AA boundaries appear to be drawn almost randomly. Specifically, for every bank I put together a collection of census tracts around the borders of AAs that appear to be quite similar in all observable characteristics. The only major observable difference is that roughly half of those tracts falls within the bank's assessment area, whereas the other half does not. In essence, I find pairs of tracts that look like clones of one another.

I construct the matching function at the census tract level; it computes a "similarity" measure for each census tract in the sample to every other tract. It then selects a predetermined number of neighboring tracts in terms of this "similarity" measure, and by changing the number of neighbors I am able to look as closely to the assessment area boundary as I wish. This measure for tracts A and B is a weighted average of

1. the geographic (also known as the "great circle") distance from the center of tract A to the center of tract B,
2. and Euclidean distance between standardized values of all socio-economic characteristics.[16]

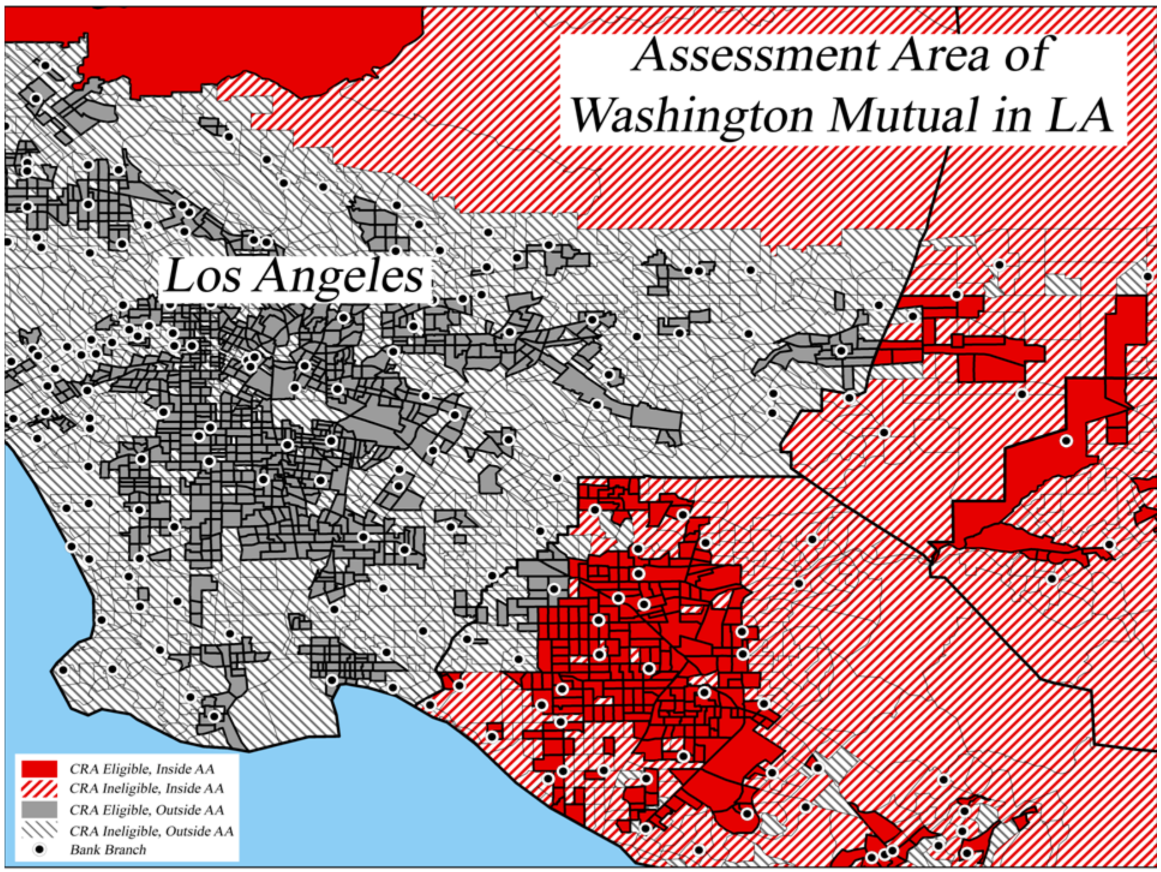
I use an extensive collection of socio-economic variables for the matching procedure. From the 2000 Census, I construct measures of median incomes, racial

[16] Standardization involves subtracting the mean of each variable from every value it takes and dividing the result by its standard deviation. This ensures that the units of different variables are irrelevant, otherwise differences in income would swamp differences in, say, poverty rates. My procedure explicitly guarantees that every covariate will receive the same weight in the overall criterion.

and ethnic composition, house values, home ownership costs and poverty levels for every census tract. I also construct an index of the annual crime rates (on the county level) using the data on the number of various crimes from the California Attorney General's website. Finally, I use 2000–2004 HMDA data to construct a county-level measure of average credit scores. When a loan application is denied, most lenders list up to three reasons for denial, and "poor credit history" is one possible reason. I compute the average number of loans that were denied due to poor credit history and use this number as a proxy for the true distribution of credit scores. Other studies have followed this approach before; see Ergungor (2007) and the references therein.

In constructing the matching algorithm, I make an implicit assumption that most census tract characteristics change smoothly from one adjacent tract to another. This condition is similar to the standard continuity condition made in most regression discontinuity studies; see, for example, Imbens and Lemieux (2008), Assumption 2.2. Thus, the only factor that changes discontinuously is thus the status of being in a bank's assessment area.

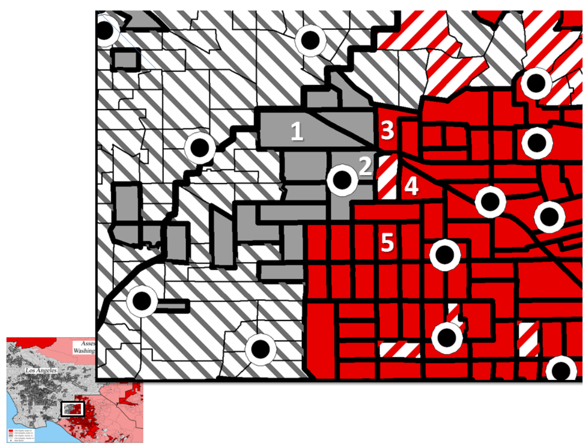
Figure 1.2: An Illustration of Washington Mutual Bank's Assessment Area



The figure uses the map of Washington Mutual Bank's assessment area around Los Angeles. Plotted using the ArcGIS software.

Figures 1.2 and 1.3 depict the application of the matching algorithm. I present the map of the Washington Mutual Bank's assessment area around Greater Los Angeles. The solid-colored red areas are lower-income tracts located inside the AA, and the solid-grey ones are lower-income tracts outside the AA.[17] Similarly, areas colored with red stripes are CRA-ineligible tracts inside AAs, and the areas with grey stripes correspond to medium- and high-income areas outside the AA. Black dots with white halos around them indicate the branch offices of Washington Mutual. Figure 1.2 provides a "helicopter view" of the AA, while Figure 1.3 illustrates the matching process.[18]

Figure 1.3: Matching Algorithm: An Example from the Greater Los Angeles Area



A close-up illustration for the matching algorithm, showing a part of the Washington Mutual Bank's assessment area around Los Angeles, depicted in Figure 1.2

Consider the tract labeled "1". The matching function computes the "similar-

[17] In greyscale, the red areas should look somewhat darker than the grey ones.

[18] The tracts need not be adjacent to one another. Rather, the distance between the tracts is one of the inputs in the matching function. Sometimes, observable tract characteristics change very abruptly as one moves from one adjacent tract to the other.

ity" measure between it and every other tract. Suppose that the tract labeled "3" is found to be the best match for "1." Since "1" and "3" are on different sides of the assessment area boundary, this pair would be included in the matched sample. If "2" was found to be the best match for "1," the pair of "1" and "2" would not be included because they are both in the "control" group. My small matched sample, which I call "Closest-Two Tracts," is constructed as a collection of pairs of tracts that are best matches for each other and have different treatment status, like "1" and "3" above. Again, only the solid-colored tracts are given consideration.

I also construct a larger matched sample, which I call "Closest-Five Tracts", as follows. Take the same tract "1" and find the best four matches for it; these would be tracts "2", "3", "4"and "5"in Figure 1.3. I would require that not all of these five tracts are in the same group (treatment or control). I use both of these samples in most of my estimation as a robustness check because the "Closest-Five Tracts" sample has four times as many observations in it. A common problem with all local identification sources is that one usually does not have enough data.

By estimating equation (1.4.3) on the samples obtained from the matching process, I am partially controlling for cs_i. Even though I still do not observe individual credit scores, by looking at very similar applicants, the confounding effect is reduced: similar applicants are likely to have similar credit scores. In case some omitted variable bias remains, if $\lambda_{j,2} < 0$ for all j, then the estimation results from the matching procedure would provide a lower bound on the effect of interest.

The second strategy that I employ amounts to using an appropriate instrumental variable (IV) for AA_i. I use the location of bank branches to construct the instrument. I geocode the locations of all branches and construct a measure of distance from every census tract to each branch and record the minimum distance. My solution is to estimate equations (1.4.1) and (1.4.2) jointly and use this distance measure (which I call $dist_i$) as an instrument for the assessment area "treatment." In this case equation (1.4.2) becomes

$$AA_i = dist_i \cdot \delta_{j,1} + x_i' \delta_{j,2} + cs_i \cdot \lambda_{j,2} + u_{i,2}.$$

A usable instrument has to satisfy two conditions: relevance and validity. The relevance condition states that the instrument has to be correlated with the endogenous variable. This condition can be tested directly, but it is intuitive to see why this holds in my application. For some banks, all their branches are located inside AAs. For other banks, this is not the case, but in a matched sample of tracts along the AA borders the closest branch is in many cases the one inside the AA. That is why people applying from within the bank's assessment area will necessarily be closer to the branch than those who apply from the outside. The validity condition states that distance should only affect the outcome of the loan approval process indirectly, otherwise it should be included among the other explanatory variables. Formally, I require the instrument to be uncorrelated with cs_i.

I have reasons to believe that the validity condition holds. Banks get no explicit credit for approving loans that are closer to their branches. It is quite common among the mortgage applicants to apply for loans with several different banks at once. Some of these banks may have branches near the house the applicant wants to purchase, and some may not: people may visit branches located next to their offices that are somewhat removed from the house in question. Given the distances in my sample (see Table 1.2), the assumption that distance from the house to the branch does not matter for the loan approval decision by itself seems tenable. The banks might, however, get credit from the regulators for approving "nearby loans" because they fall into the bank's assessment area and hence help the bank to comply with the CRA.

Banks probably decide strategically where to locate their branches as well. This has been a subject of numerous studies including Ho and Ishii (2010). For the purposes of my identification strategy, however, it is only necessary that CRA compliance not be the primary reason for branch locations. The strategy would break down if banks choose to locate their branches in subsets of AAs that are populated by people with better credit scores.[19] In this case, the IV estimate will be biased, and the sign of the bias is predictable:

$$\text{Bias}\,\hat{\beta}_j = \text{plim}\,\hat{\beta}_j^{IV} - \beta_j = \frac{\text{Cov}\,(dist_i, cs_i)}{\delta_{j,1}}.$$

By construction, $\delta_{j,1}$ in the denominator is negative – larger distance to a

[19] In general, if applicants that are closer to branches have better credit histories. This can be true if, for example, borrowers that come from distant places were previously rejected at some other bank. Or perhaps borrowers from outside AAs are more likely to use mortgage brokers, and banks do not trust these applications as much.

branch located inside AA is negatively related with the chance of being inside AA. The results of the first stage regressions in Table C.1 in Appendix C show that this is indeed the case. The covariance in the numerator cannot be measured from the data, but if it is not zero, I would rather expect it to be negative as well. This would mean banks try to locate their branches next to more reliable borrowers. In other words, the larger the distance to a branch, the lower a person's credit score would be. Since a ratio of two negative numbers is positive, if the IV estimates are biased, then the bias is likely positive. The true magnitude of the CRA monitoring impact is then somewhat lower than the IV estimates suggest. In this sense, my second approach provides an upper bound on the effect of interest.

Dealing with an endogenous variable in a nonlinear model is, however, a complicated econometric problem with no "one-size-fits-all" solution. One feasible solution is to use the two-stage least squares (2SLS) method on a linear probability model:

$$y_i = AA_i \cdot \beta + x_i'\gamma + \varepsilon_{i,1}, \tag{1.4.4}$$

where the notation is exactly the same as in (1.4.1), and the instrument equation is

$$AA_i = dist_i \cdot \delta_1 + x_i'\delta_2 + \varepsilon_{i,2}, \tag{1.4.5}$$

where $dist_i$ is the measure of distance from the assessment area boundary to the nearest branch inside AA, and it is bank-specific. I define $\varepsilon_{i,1} \equiv cs_i \cdot \lambda_1 + u_{i,1}$, and $\varepsilon_{i,2} = cs_i \cdot \lambda_2 + u_{i,2}$ so the error terms are correlated through cs_i. Hence these two equations must be estimated jointly.

This model, however, will only be an approximation to the nonlinear model with heterogeneous coefficients that I really seek to estimate. Also, linear probability model does not account for the fact that predicted values have to be within the unit interval. I hence turn to the nonlinear Bayesian IV model to address these concerns.

1.4.2 Nonlinear Bayesian IV

The full model is

$$y_i^* = AA_i \cdot \beta_j + x_i' \gamma_j + \varepsilon_{i,1}, \tag{1.4.6}$$

where AA_i is endogenous. The instrument equation is

$$AA_i = z_i' \delta_j + \varepsilon_{i,2}, \tag{1.4.7}$$

where $z_i = (dist_i, x_i)$, $\delta = (\delta_1, \delta_2)$, and where y^* is unobserved. Instead, I observe $y_i = \mathbb{1}\{y_i^* \geq 0\}$. The fact that y^* is unobservable implies the need to apply a limited dependent variable technique to (1.4.6), most commonly a probit model. And since ε_1 and ε_2 are correlated through cs_i which is included in them, (1.4.6) and (1.4.7) must be estimated simultaneously. I apply techniques developed in the Bayesian literature to estimate this model.

Specifically, I supplement the standard linear Bayesian IV method with the data augmentation step. What makes this nonlinear model with endogeneity hard to estimate is the fact that y^* is not observed, otherwise two-stage least squares method would be perfectly applicable. The data augmentation procedure is a vehicle that converts a nonlinear model into a linear one. If one makes a distributional assumption on the error terms ε, it becomes possible to determine the distribution of y^*. After that, the values of y^* can be simulated and, conditional on these simulated values, the whole model becomes linear. One can then apply standard linear IV methods to the transformed model. The general idea of the Bayesian approach to inference follows below.[20]

I start with the data at hand and a prior distribution on parameters. This prior distribution represents my initial beliefs about the parameters' distributions and is, in practice, mostly chosen for computational convenience. The data provide the likelihood function, and together the product of the prior and the likelihood yields the posterior distribution for parameters. The Bernstein von-Mises theorem ensures that means of posterior distributions are asymptotically equivalent to maximum likelihood estimates (MLE). In finite samples, one can select a fairly diffuse prior (e.g., a normal distribution with a large variance), so that it would contribute almost nothing to the posterior, which would then be mostly driven by the likelihood.

With the exception of several simple models, the posterior distributions are,

[20] A complete exposition of the Bayesian approach is beyond the scope of this paper. Some excellent sources on this question include Rossi, Allenby, and McCulloch (2006) as well as Geweke (2005).

in general, too complicated to be analyzed directly. The standard approach is to use Markov chain Monte Carlo (MCMC) methods to obtain a random sample of draws from the posterior, and then get the required information from these draws. The Gibbs sampling technique simplifies the process of making such draws by splitting the whole parameter space into nonoverlapping blocks. I choose blocks so that, conditional on all other blocks being held constant, the posterior of any given block is of a known form. It is then straightforward to make random draws from this block. By alternating between blocks, I obtain a sequence of draws that converges to the draws from the joint posterior for all parameters.[21]

Apart from assuming priors, I impose a distributional assumption on the error terms:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right],$$

where $\sigma_{12} \neq 0$ and which implies that $\varepsilon_1 \mid \varepsilon_2$ and $\varepsilon_2 \mid \varepsilon_1$ are also normal.

Given that y^* is a linear function of ε_1, it also follows the normal distribution (and this stays true if one conditions on ε_2). I treat $\{y_i^*\}_{i=1}^I$ as an extra set of parameters that can be simulated. The simulation must incorporate the observed information on the sign of each y^* (i.e. y), so draws must be made from the truncated normal distribution. I use the mixed rejection algorithm developed in Geweke

[21] A vivid example is given in Rossi, Allenby, and McCulloch (2006): suppose I want to draw from a bivariate normal distribution, but I can only make draws from univariate normals. For a bivariate normal, the conditional distribution of any component given the other is also normal. The Gibbs sampling approach would imply making alternating draws from two univariate conditionals, which eventually produces a sequence of pairs of draws. This sequence in turn is asymptotically equivalent to a sequence of random draws from a bivariate normal distribution, which is precisely what I needed in the first place.

(1991), the relevant part of which is given in Appendix B.

The entire Gibbs sampling scheme can be summarized as (omitting subscript j for brevity)

$$\begin{aligned}
\{y_i^*\}_{i=1}^I &\mid \beta, \gamma, \delta, \mathbf{\Sigma}, y, AA, z && (1.4.8)\\
\beta, \gamma &\mid \delta, \mathbf{\Sigma}, \{y_i^*\}_{i=1}^I, AA, z && (1.4.9)\\
\delta &\mid \beta, \gamma, \mathbf{\Sigma}, \{y_i^*\}_{i=1}^I, AA, z && (1.4.10)\\
\mathbf{\Sigma} &\mid \beta, \gamma, \delta, \{y_i^*\}_{i=1}^I, AA, z, && (1.4.11)
\end{aligned}$$

and I detail each of those four steps in Appendix A.

The model is complete after I specify the priors:

$$\begin{aligned}
\begin{pmatrix} \beta \\ \gamma \end{pmatrix} &\sim N\left[\begin{pmatrix} \overline{\beta} \\ \overline{\gamma} \end{pmatrix}, \mathbf{A}_{\beta\gamma}^{\mathbf{-1}}\right], \\
\delta &\sim N\left(\overline{\delta}, \mathbf{A}_{\delta}^{\mathbf{-1}}\right), \\
\mathbf{\Sigma} &\sim IW\left(\upsilon_0, \mathbf{V_0}\right),
\end{aligned}$$

where $\left(\overline{\beta}, \overline{\gamma}, \overline{\delta}\right)$ are priors' means, $\mathbf{A}_{\beta,\gamma}^{\mathbf{-1}}$ and $\mathbf{A}_{\delta}^{\mathbf{-1}}$ are prior variance matrices. In the Bayesian literature, it is customary to work with the inverses of the variance matrices, which are referred to as precision matrices, and hence $\mathbf{A}_{\beta,\gamma}$ and $\mathbf{A}_{\delta}$ are prior precisions. IW is the inverse Wishart distribution, which is essentially a generalization of χ^2-distribution to the space of positive-definite matrices (instead of positive integers), υ_0 is its scale parameter, and $\mathbf{V_0}$ is its location parameter.

I chose very uninformative priors and set $\bar{\beta}, \bar{\gamma}, \bar{\delta}$ to vectors of zeros, $\mathbf{A}_{\beta,\gamma} = \mathbf{A}_{\delta} = 0.01\mathbf{I_{k+1}}$ (where $k = \dim[\delta]$), $\upsilon_0 = 3$ and $\mathbf{V_0} = 0.01\mathbf{I_2}$ (I is the identity matrix). This ensures the priors are considerably "spread out", so that the shape of the posterior is mostly driven by the likelihood function. The choice of functional forms for the priors is motivated by computational considerations: given the Gibbs sampler blocks in (1.4.11), posteriors on β, γ and δ will also be normal, and posterior on $\mathbf{\Sigma}$ will also be inverse Wishart. See Appendix A for details.

1.5 Data

1.5.1 Data Sources

I use several data sources for my estimation procedure. The most important source is the HMDA data set (pronounced "humda"), made available by the Federal Financial Institutions Examination Council (FFIEC). It contains a majority of all home mortgage loan applications in the U.S.. The potential mortgage originators (called "respondents" in the HMDA language) are the main subjects of the HMDA.

An observation in the dataset is a loan application, and several important characteristics of the application are available. These can be divided into three major groups:

1. Borrower's characteristics, such as race, gender, ethnicity and income.

2. Respondent's characteristics, such as name, type, address, parent company name (if applicable), supervisor's identity.

3. Loan characteristics, such as amount, property address (aggregated up to a census tract), various type measures (single or multi-family; conventional, FHA or VA;[22] owner-occupied or not; new or refinanced loan, etc). The important loan characteristics are the decision taken on the application, and the reasons for denial (if applicable).

The FFIEC aims to preserve the anonymity of mortgage applicants by not disclosing the application date, which is rounded to a calendar year. A few potentially interesting covariates are absent in the HMDA data: no information exists concerning the term structure of the loan (fifteen-year mortgage or thirty-year mortgage), as well as very limited information concerning the loan interest rate. HMDA respondents must report the difference between the loan annual percentage rate (APR) and the rate on Treasury securities of comparable maturity, as long as the spread is above the designated threshold. It is also impossible distinguish between the fixed rate and adjustable rate mortgages. On average, there were 31 million loan applications per year in the HMDA data (during 2000–2005). In Section 2.4.2, I describe the way the final dataset was constructed. Avery, Brevoort, and Canner (2006, 2007) are the two best sources in which these data are discussed extensively.

The FFIEC also provides the CRA-related information. Each year, every fi-

[22] FHA stands for "Federal Housing Administration," and VA is an acronym for "Veteran Affairs." Loan applications with such labels are usually at least partially subsidized by the federal government.

nancial institution that is subject to the CRA must file a number of reports to the FFIEC. I use the definitions of institutions' assessment areas from the CRA disclosure reports. These are reported for each institution on a census tract level and can be linked to the HMDA loan applications directly by institution.

I also draw on several supplemental data sources. First, I use the Census 2000 data on racial and ethnic composition, home ownership costs, median family incomes, poverty rates and house values for every census tract in California. The Census Bureau also provides information on latitude and longitude for each census tract. Next, the 2005 FDIC Summary of Deposits data contains a complete list of all bank branches with their addresses. I take the crime data from the Attorney General of California website and construct an annual crime rate index for the years 1999-2005. Finally, the Bureau of Labor Statistics makes available the CPI data, which allows me to express all nominal variables (annual income and loan amount) in 1999 dollars.

1.5.2 Dataset Construction

I chose to look at the loan applications in California in 2005. By examining a single state, I abstract from inter-state variations in laws that regulate banks' operations. California is one of the largest states, and its home mortgage lending market makes up 14.8 percent of the overall U.S. market (on average between 2000 and 2005). By examining 2005 data, I concentrate on one of the last years before the mortgage crisis started to unfold.

I focus on a subset of mortgage lenders that are explicitly subject to the CRA. In the middle of the 2000s, only one out of three loan applications were reported by such lenders. Thus, excluding the other two-thirds of the mortgage applications would permit me to isolate the possible effects of the CRA enforcement mechanism.

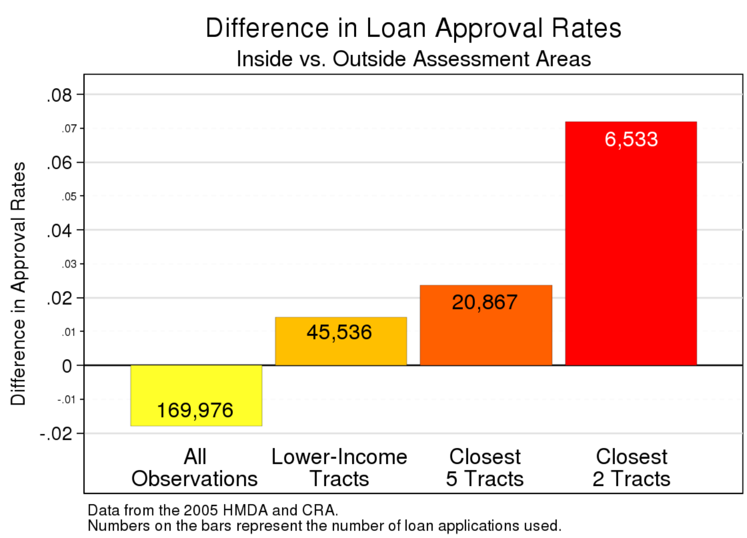
I keep only conventional loans (no FHA or VA applications). I exclude loans that are not for single family owner-occupied homes, and those that are secured by anything other than a primary lien. Finally, I only keep home purchase loans. Non-primary liens and non-single family loans, in practice, are more likely to be associated with real estate speculative purchases, especially during the time of interest. There are also reasons to believe that borrowers' characteristics are more accurately reported for home purchase loans; see Bhutta (2008).

In the HMDA data, the decision taken on each application is reported, ten different decisions are provided. I exclude all loans that were purchased from another institution, because the decision on those had been taken by some other entity. I exclude applications with decision reported as "application withdrawn," which are thought to be associated with indirect lending through mortgage brokers rather that directly through banks of interest. I then construct the loan approval indicator which equals one if loan was approved (whether originated or not). Borrowers can apply for loans with different banks at the same time, and there is no way to identify two different loan applications with a single borrower in the data. My primary interest, however, is in banks' approval decisions, and it is quite possible that different banks assess the same person's application differently.

1.5.3 Preliminary Evidence

Figure 1.4 illustrates the difference in loan approval rates inside and outside the assessment areas in the four samples considered above. To compare loan approval rates using the whole dataset would be misleading. If anything, the first bar in the chart suggests that, overall, slightly fewer loans get approved within the assessment areas. This result is in perfect agreement with the negative omitted variable bias interpretation. It suggests that banks end up having to include a number of tracts into their assessment areas that are populated with borrowers that have weak credit scores.

Figure 1.4: Preliminary Evidence



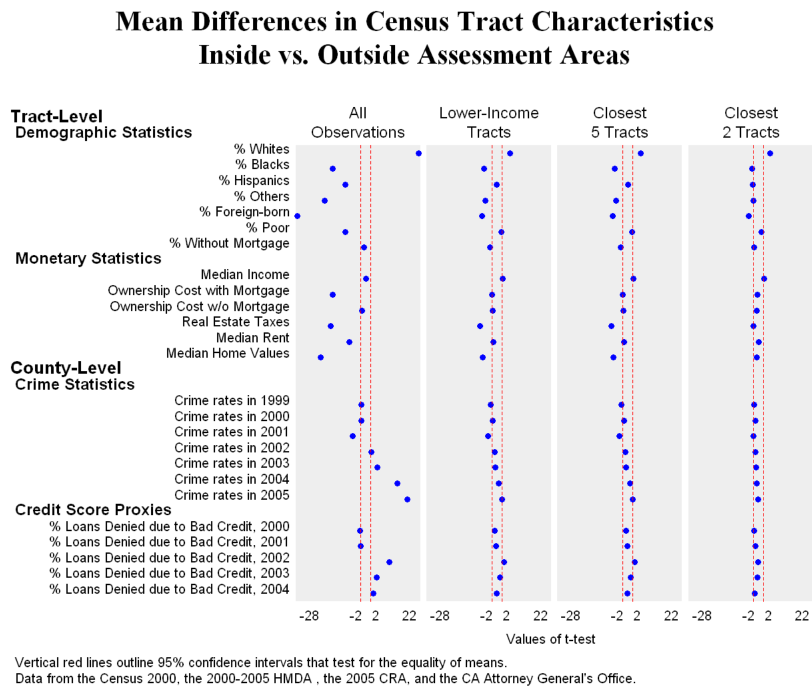
The second bar indicates that when attention is restricted to only CRA-eligible tracts, a small positive effect exists: there is a 1.5 percentage point higher chance that a bank would approve the loan within its assessment area. The full magnitude of the AA "treatment" does not reveal itself before I use the matched samples for the comparison. When I look at the sample of the most informative loans ("Closest-Two Tracts") which is by construction extremely homogenous as illustrated by Figure 1.5 below, I find a 7.19 percent higher chance of loan approval in the "treated" areas. This is a striking observation given that it is not controling for loan-level observables, it relies only on the matching which is done at the census tract level.

Figure 1.5 illustrates the results of the matching procedure outlined in Section 1.4.1. The complete list of matching variables appears on the vertical axis. Every dot represents the value of the t-statistic that tests for the equality of means inside and outside assessment areas. The vertical dashed bands set the boundaries for the 95 percent confidence intervals. If a given dot is within the band, then the null of means being equal cannot be rejected.[23]

It is clear that before any matching algorithm is applied, the "treatment" and "control" census tracts are quite different on average. Notably, the first column provides more evidence in support of the negative selection bias conjecture. The matching variables at the bottom are the proxies for the distribution of credit scores.

[23] The number of degrees of freedom for each t-statistic is at least 300 (much higher in most cases), and at this point there is little difference between the t-distribution and the standard normal distribution.

Figure 1.5: Matching Results



It appears that more loans were historically denied inside AAs due to poor credit histories of borrowers as compared with the applications from outside AAs. I interpret this as indication that people inside AAs indeed have weaker credit histories.

Some of the variation in observables smoothes out once I look only at the CRA-eligible tracts (second column), but a number of differences are still fairly significant. Once I get to the “Closest-Two Tracts” matched sample, however, virtually all the differences in means appear to be small enough to claim that they do not matter. An ideal outcome of the matching procedure would have all the blue dots line up vertically at zero, and I get fairly close to that in the small matched sample.

Under the standard assumption that "the unobservables are just like the observables," I can conclude that cs_i will also vary little across the treatment and control groups in the "Closest-Two Tracts" sample.

Perhaps the largest differences in means across the treatment status can be observed in the tract-level demographic characteristics (the top several dots in each column). I would argue, however, that it is more important to control for these characteristics at the individual loan level. Most of these demographic covariates are available at precisely this level of disaggregation.

Table 1.1: Descriptive Statistics on Loan Applications

	All Observations		Lower-Income Tracts		Closest-Five Tracts		Closest-Two Tracts	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Loan Amount, \$1000	399.21	299.56	299.24	150.31	279.48	150.66	277.80	146.91
Annual Income, \$1000	132.13	174.67	96.89	89.03	92.64	94.14	93.00	99.03
Applicant Female	0.348	0.476	0.377	0.485	0.351	0.477	0.341	0.473
Applicant Not White	0.372	0.483	0.373	0.484	0.364	0.481	0.368	0.482
Applicant Hispanic	0.338	0.473	0.435	0.496	0.442	0.497	0.440	0.496
Has a Co-Applicant	0.481	0.500	0.399	0.490	0.411	0.492	0.411	0.492
# of Counties	56		53		42		16	
# of Census Tracts	6,638		2,943		1,269		390	
# of Loan Applications	169,964		44,546		20,867		6,533	
	Breakdown of Means by Assessment Area ("Treatment" Status)							
	Inside	Outside	Inside	Outside	Inside	Outside	Inside	Outside
Loan Amount, \$1000	353.70	455.51	277.26	328.75	267.76	316.03	276.65	281.96
Annual Income, \$1000	119.79	147.41	93.00	102.16	89.36	102.89	92.99	93.02
Applicant Female	0.342	0.355	0.335	0.433	0.338	0.391	0.327	0.389
Applicant Not White	0.335	0.419	0.363	0.386	0.361	0.374	0.366	0.376
Applicant Hispanic	0.355	0.316	0.442	0.426	0.457	0.393	0.434	0.464
Has a Co-Applicant	0.491	0.468	0.423	0.366	0.422	0.377	0.428	0.352
# of Loan Applications	93,994	75,970	25,527	19,019	15,802	5,065	5,114	1,419

Data from the 2005 HMDA and CRA.
Included are the conventional, owner-occupied 1-4 family home purchase loans secured by primary liens.

Table 1.1 provides some basic summary statistics on the loan applications in the four samples discussed above. The rows correspond to the borrower characteristics that I observe on the loan level. The top part of the table contains the averages and the standard deviations for each sample, and the bottom part displays the breakdown of means by the treatment status. Many census tracts belong to the AA of more than one bank, and about 20 percent of them do not belong to any. The pattern from Figure 1.5 persists: larger samples suggest that loan applications are not quite similar on average from the treated and the untreated tracts. At the same time, the differences virtually disappear in the “Closest-Two Tracts” sample. The largest difference is in the proportion of applicants that have a co-applicant; the applicants from inside the banks’ AAs are around eight percent more likely to bring a co-applicant along. The only other discrepancy is that I have more than twice as many applications in the “treatment” group as I have in the “control” group, but the absolute numbers are large enough (5114 and 1419 in the smallest sample), so that this is unlikely to be a problem.

Table 1.2 provides a general overview of the market shares of different mortgage lenders that are explicitly subject to the CRA. Wells Fargo Bank is the largest; effectively every fourth loan application in my data had been handled by them. At that time, Wachovia had not yet merged with Wells Fargo. Washington Mutual is the second largest, and Bank of America is effectively tied for third place with Fremont Investment & Loan (which is, as of 2010, a part of the CapitalSource Bank). These banks were the top twelve lenders in California in 2005 that are subject to the CRA in terms of the number of loan applications received. Including other lenders does not seem possible because I would then risk not having enough observations in both the treatment and the control group.

Table 1.2: Lending Patterns and Branches by Lender, "Closest-Two Tracts" Sample

	Number of Applications	% From Within AA	% Approved		Number of Branches[a]		Median Distance To Branch[b]	
			Inside AA	Outside AA	Inside AA	Outside AA	Inside AA	Outside AA
Bank Of America	875	90.17	76.68	87.21	514	472	1.93	16.96
Downey Savings & Loan	452	67.69	73.86	67.12	94	74	5.88	98.63
First Bank	166	39.76	60.61	57.00	24	24	39.07	67.68
First Federal Bank of CA	88	17.05	86.67	65.75	1	28	33.75	142.24
Fremont Investment & Loan	939	65.92	63.49	61.88	10	10	37.53	146.43
Guaranty Bank	62	48.39	70.00	62.50	41	9	22.86	40.56
Provident Savings Bank	164	32.93	77.78	80.00	12	0	7.08	54.43
US Bank	25	84.00	95.24	50.00	178	137	1.47	90.22
Union Bank of CA	39	92.31	86.11	100.0	223	110	1.88	30.24
Wachovia	733	72.17	70.32	65.19	68	55	7.14	49.78
Washington Mutual Bank	1230	86.50	76.41	73.49	314	284	2.14	20.56
Wells Fargo Bank	1760	90.06	89.27	90.29	469	404	1.85	14.86

[a] In these two columns I display the number of all branches in California, not just in the "Closest 2 Tracts" sample.

[b] Distances measured in miles.

Data from the 2005 HMDA, CRA and the Summary of Deposits.

The last four columns contain information on banks' branches with breakdown by treatment status. For example, Bank of America had 514 branches inside its assessment areas in California in 2005, and 472 outside. The last two columns tell us that the median distance from a given census tract in the "Closest-Two Tracts" sample to the nearest branch. Because some banks have a lot more branches than others (Bank of America tops the list with 953, whereas Provident Savings Bank is the last with only twelve), there is considerable variation in median distances. I do not present the average distances since there are several census tracts that are really far away from some of the branches, and these distances inflate the averages.[24]

On average, distances from a tract inside the assessment area to a branch are lower than the same distances for loans that are not in the assessment areas. The large number of branches outside AAs for many banks, together with the historic data on branches' locations, suggest that the banks' decision to open a new branch can be thought of as not being driven *primarily* by the need to comply with the CRA. The vast majority of branches have been in place for several years, and over 75 percent have been in place for over five years.

Table 1.3: Decomposition of Selected Average Loan Characteristics By Lender, "Closest-Two Tracts" Sample

	Loan Amount		Annual Income		Applicant Not White		Applicant Hispanic	
	Inside AA	Outside AA	Inside AA	Outside AA	Inside AA	Outside AA	Inside AA	Outside AA
Bank Of America	240.95	327.22	86.40	112.43	0.329	0.500	0.406	0.233
Downey Savings & Loan	282.87	305.69	92.59	100.43	0.484	0.547	0.588	0.581
First Bank	249.77	213.57	79.11	68.53	0.167	0.149	0.591	0.772
First Federal Bank of CA	334.26	276.32	111.20	96.37	0.368	0.409	0.632	0.511
Fremont Investment & Loan	236.39	254.32	87.93	85.14	0.284	0.325	0.659	0.491
Guaranty Bank	264.37	273.31	87.50	86.53	0.233	0.375	0.300	0.500
Provident Savings Bank	185.43	294.98	70.83	97.02	0.167	0.209	0.407	0.464
US Bank	226.88	277.50	91.91	91.75	0.333	0.500	0.250	1.000
Union Bank of CA	409.10	425.33	152.78	134.33	0.250	0.667	0.250	1.000
Wachovia	264.95	258.06	89.25	93.92	0.320	0.385	0.556	0.561
Washington Mutual Bank	327.86	319.10	101.22	95.68	0.372	0.512	0.414	0.339
Wells Fargo Bank	277.96	310.23	93.97	98.69	0.305	0.449	0.314	0.222

Data from the 2005 HMDA and CRA.
Every cell contains the average value of the corresponding loan-level observable for a given bank and treatment status.

Table 1.3 presents evidence for a closer look at the primary sample of interest, "Closest-Two Tracts". It breaks down the selected loan-level observables by bank and by treatment status. For example, it suggests that the average loan application size at Wells Fargo was almost $\$278,000$ inside its assessment areas, and a little over $\$310,000$ outside the assessment areas. The largest average loan sizes were

[24] The outliers inflate the medians as well, but not as much. I tried estimating the models without the observations that have distance outliers, and the results virtually did not change at all.

observed at the Union Bank of California (both in the treatment and the control groups). Provident Savings Bank and the First Bank were the two lenders with the lowest percentage of non-white applicants (16.7 percent). Overall, there was a certain degree of choice on the consumer side in terms of which bank to apply to. Taken together, Tables 1.2 and 1.3 suggest that accounting for heterogeneity among lenders is important. This motivates my usage of the Bayesian IV method. The complete picture of differences in loan-level observables across four different samples that I use can be found in the Appendix C (Tables C.2, C.3, C.4, and C.5).

1.6 Estimation Results

1.6.1 Linear Probability Model Estimates

Table 1.4 presents the estimation results for the OLS linear probability model defined in (1.4.1). The dependent variable is the loan approval indicator and the key regressor of interest is the CRA assessment area dummy. I shall refer to this estimate as "the AA effect".

In the first column are reported results using all available observations, i.e. estimating equation (1.4.1) using all the data on hand. No attempt is made to control for the unobservable cs_i. Hence, the estimate of β should not be taken literally. The fact that the estimate is negative again suggests that the omitted variable bias is large and negative. Some of the bias can be accounted for by only considering observations from CRA-eligible tracts only, which is what the second

Table 1.4: Linear Probability Model for Loan Approval

	Column 1	Column 2	Column 3	Column 4	Column 5
	All Observations	All	Lower-Income Tracts Closest-Five	Lower-Income Tracts Closest-Two	Closest-Two
Loan in Assessment Area (β)	-0.015*	0.013†	0.030*	0.068*	**0.101***
	(0.002)	(0.004)	(0.007)	(0.013)	(0.020)
Loan Size, \$100k (γ_1)	-0.001 ⋄	-0.003 ⋄	-0.002	0.002	-0.018*
	(0.000)	(0.002)	(0.002)	(0.004)	(0.005)
Annual Income, \$100k (γ_2)	0.001	0.010*	0.007	0.011†	0.011†
	(0.001)	(0.003)	(0.004)	(0.004)	(0.004)
Applicant Female (γ_3)	-0.006 ⋄	0.004	0.004	-0.000	-0.011
	(0.002)	(0.004)	(0.006)	(0.011)	(0.012)
Applicant Not White (γ_4)	-0.024*	-0.023*	-0.021*	-0.017	-0.018
	(0.002)	(0.004)	(0.006)	(0.011)	(0.012)
Applicant Hispanic (γ_5)	-0.088*	-0.092*	-0.088*	-0.085*	-0.054*
	(0.002)	(0.004)	(0.006)	(0.011)	(0.013)
Has a Co-Applicant (γ_6)	0.052*	0.048*	0.044*	0.035†	0.032†
	(0.002)	(0.004)	(0.006)	(0.011)	(0.011)
Constant (γ_7)	0.799*	0.785*	0.769*	0.724*	0.746*
	(0.003)	(0.006)	(0.010)	(0.017)	(0.022)
Tract Fixed Effects	No	No	No	No	Yes
Number of observations	169,964	44,546	20,867	6,533	6,533

Dependent variable: loan approval indicator. Data from the 2005 HMDA and CRA.
Standard errors in parentheses. ⋄ $p < 0.05$, † $p < 0.01$, * $p < 0.001$

column shows.

In columns 3 through 5, I am only using observations from census tracts identified as similar by the criterion described in Section 1.4.1. As the number of neighboring tracts decreases, I am left with fewer observations (sample size drops from 20,867 to 6,533), but in return I have tracts that are "more similar" to each other. This allows me to partially account for the unobservable cs_i: since I only use applications from very similar people, their credit scores should be fairly homogenous. The model in column 5 also has the census tract fixed effects.

The estimates of β are positive and very significant. The model in column

5 suggests that loan applicants from inside the banks' assessment areas have on average 10.1 percent higher chance of getting approved compared to applications from identical people from outside the assessment areas, other things equal. This result is consistent with the model from Section 1.3, and, given the discussion in Section 1.4.1, is likely to provide the lower bound for the AA effect. The matching procedure may fail to purge all the omitted variable bias completely, but the bias appears to be negative, so the true value of the AA effect might be higher.

Table 1.4 also illustrates that, all else equal, banks tend to approve loan applications from Hispanics less willingly. The magnitude of the effect really does not change much no matter how finely I "slice" the data. Namely, there is at least five percent lower chance that a loan will be approved if it is coming from a Hispanic applicant, controlling for all observable differences and using the aforementioned identification strategy.

I have also implemented an intuitively appealing robustness check. I took the data from the "rich" census tracts (defined as tracts with median income of at least 1.2 of the corresponding MSA median income), and applied the same matching function described in Section 1.4.1 to them to obtain a matched subsample. After that, I estimated equation (1.4.1) on this subsample, and found the AA effect among those loans to be statistically undistinguishable from zero. This result was intuitive: since the CRA does not explicitly reward banks for extra lending in the high-income areas, this suggests that the estimation procedure is indeed identifying the effect of interest on the lower-income tracts.

I next turn to the second estimation approach, modeling the AA effect as a function of distance to the nearest bank branch. I first present the estimation results from the linear probability approximation, since this can be easily done using the two-stage least squares method. In Table 1.5, I present the estimation results.

Table 1.5: 2SLS Linear Probability Model for Loan Approval

	Column 1	Column 2	Column 3	Column 4	Column 5
	All Observations	Lower-Income Tracts			
		All	Closest-Five	Closest-Two	Closest-Two 2
Loan in Assessment Area (β)	0.163*	0.237*	0.209*	0.214*	**0.251***
	(0.006)	(0.015)	(0.020)	(0.032)	(0.036)
Loan Size, \$100k (γ_1)	0.00004*	0.011*	0.005 ⋄	0.003	−0.017†
	(0.000004)	(0.002)	(0.002)	(0.004)	(0.006)
Annual Income, \$100k (γ_2)	0.0000003	0.005	0.007	0.011†	0.012 ⋄
	(0.000007)	(0.003)	(0.004)	(0.004)	(0.006)
Applicant Female (γ_3)	−0.005 ⋄	0.004	0.004	−0.003	−0.012
	(0.002)	(0.004)	(0.006)	(0.012)	(0.012)
Applicant Not White (γ_4)	−0.008†	−0.002	−0.015 ⋄	−0.010	−0.020
	(0.002)	(0.005)	(0.006)	(0.012)	(0.012)
Applicant Hispanic (γ_5)	−0.094*	−0.091*	−0.095*	−0.082*	−0.050*
	(0.002)	(0.004)	(0.006)	(0.011)	(0.012)
Has a Co-Applicant (γ_6)	0.045*	0.033*	0.036*	0.026 ⋄	0.025 ⋄
	(0.002)	(0.004)	(0.006)	(0.011)	(0.012)
Constant (γ_7)	0.68*	0.618*	0.618*	0.608*	0.627*
	(0.005)	(0.013)	(0.019)	(0.029)	(0.034)
Tract Fixed Effects	No	No	No	No	Yes
Number of observations	169,859	44,546	20,867	6,533	6,533

Dependent variable: loan approval indicator. Data from the 2005 HMDA, CRA and the Summary of Deposits.
Distance to the nearest bank branch used as an instrument for loan being in the AA.
Standard errors in parentheses. ⋄ $p < 0.05$, † $p < 0.01$, * $p < 0.001$

The results are qualitatively quite similar to the ones in Table 1.4. However, the AA effect is more than twice as large now, depending on which sample is used for comparison. In Section 1.4.1, I discussed that the 2SLS estimate, if biased, is likely to be biased upwards. Hence, the estimates of β from Table 1.5 should provide the upper bound for the AA effect. It appears that regulators have more bargaining power than banks, given the evidence from the data. Surprisingly, the distance instrument seems to work in a similar way in both matched samples, as well as with all CRA-eligible observations (the fact the estimates of β are virtually the same across columns suggests this). It also seems to deal with the bias in the whole dataset to an extent, as column 1 indicates.

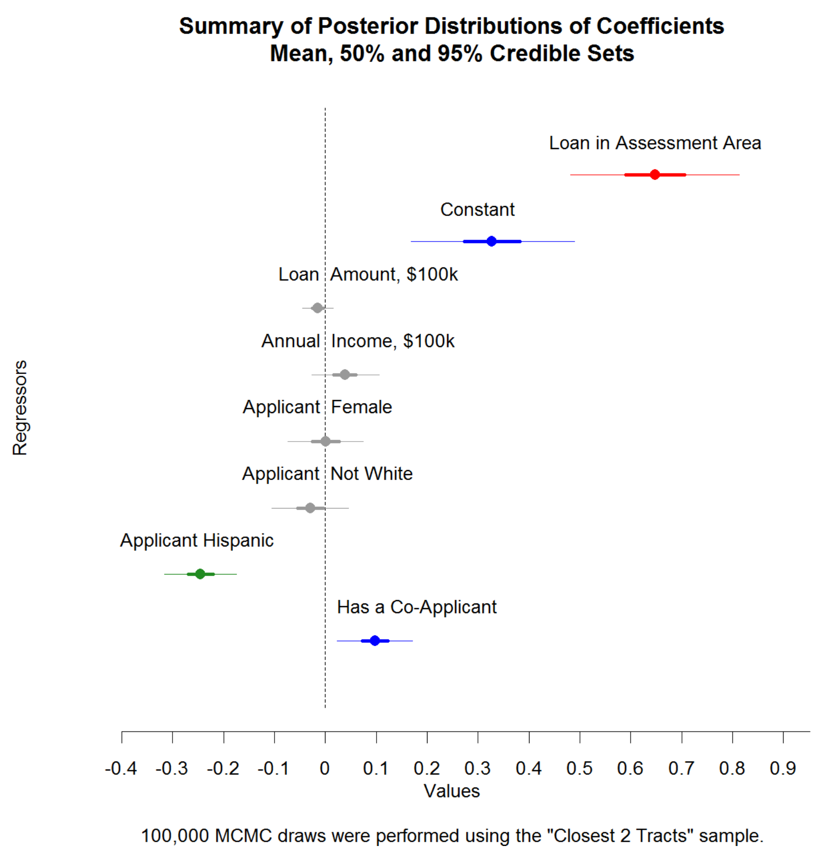
Most of the qualitative patterns from Table 1.4 persist. It is still less likely that a loan from a Hispanic applicant will be approved. It is still more likely that an application will go through if a co-applicant is present. The results suggest that, after controlling for endogeneity in AA_i, the applicants from inside and outside of the assessment areas are treated differently by the banks. Namely, the applicants from inside AAs are 25.1 percent more likely to get their loan applications approved. The results of the first stage estimation, available in Table C.1 in Appendix C indicate that there is no evidence of the weak instrument problem, as defined by Staiger and Stock (1997).

The 2SLS model, however, cannot account for heterogeneity across banks, and is necessarily only an approximation to the true nonlinear model of interest. I present the results of applying Bayesian IV in the next subsection.

1.6.2 Bayesian IV

Because of the model's structure, all the posteriors on the coefficients of the main equation (1.4.6) will be normally distributed. Figure 1.6 below summarizes the results of the MCMC procedure. I use the smaller of the matched samples, the one called "Closest-Two Tracts", and I take 100, 000 MCMC draws. Of these, the first 10 percent are discarded as "burn-in": the information in those draws can be strongly influenced by the choice of the initial conditions. I then "thinned" the sample by keeping only every ninth of the remaining 90, 000 draws. This breaks the serial dependence in the chain that gets introduced naturally by the nature of the Gibbs sampling procedure.[25]

Figure 1.6: MCMC Results



[25] Figure C.1 in Appendix C illustrates that the MCMC procedure had successfully converged to the stationary distribution of the underlying Markov chain.

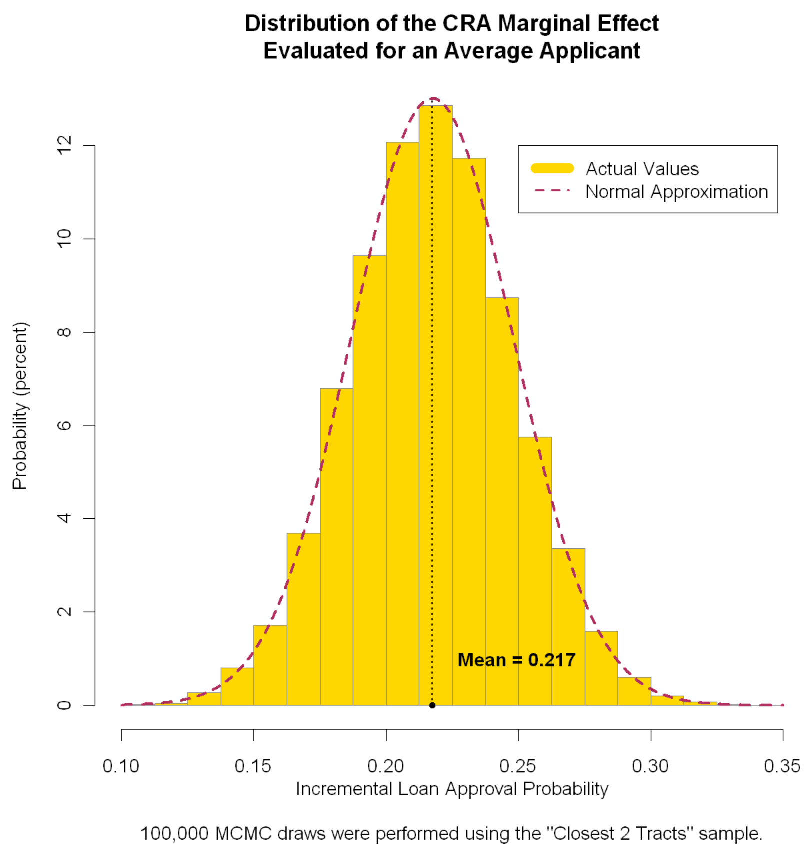
I plot credible sets (which is the Bayesian term that corresponds to that of classical confidence intervals) at different heights so that elements would not overlap. The thin vertical line indicates the zero value: if a credible set contains zero, one could think of a corresponding coefficient as being insignificant in the classical sense. Thicker portions of lines indicate 50 percent credible sets, thinner portions represent the 95 percent sets and the thickest dots represent posterior means. For a normally distributed random variable, a 50 percent confidence interval is roughly equal to a ± 1 standard deviation bound, and a 95 percent confidence interval is almost the same as the ± 2 standard deviations bound.

Notice that the credible sets for the AA effect are positive and far from zero. In fact, the 95 percent credible set for the AA effect is $[0.482, 0.813]$, with the mean at 0.648. These numbers are not directly comparable with the ones from the linear probability models, but it is still instructive to see that the posterior is fairly tight.

A credible set that lies to the right of the dashed vertical line indicates that the corresponding factor contributes positively to the loan approval score in the equation (1.4.6). Thus, the fact that the credible set for the Hispanic borrower indicator contains only negative values suggests again that, other things equal, Hispanic borrowers have a harder time getting their applications approved. Overall, the qualitative results from the Bayesian IV model are quite similar to those from its linear probability approximation discussed in Section 1.6.1.

Figure 1.7 depicts the AA marginal effect as predicted by the Bayesian IV model. The marginal effect should be thought of as the incremental probability of loan approval. If I take two identical loans such that their only difference is that one is subject to the CRA and the other is not, then the AA marginal effect demonstrates the difference between the chances each of these loans will be approved.

Figure 1.7: The CRA Marginal Effect



Because the model is nonlinear, the picture would look slightly different for every observation. Moreover, since β is stochastic, it also looks different for every MCMC draw. I preserve the stochastic part of β and compute the marginal effect for an average loan in the sample; that is, I replace the values of x_i with its column means.

The dashed maroon line represents the normal approximation of the histogram, using the mean and the standard deviation of the underlying actual values. The mean marginal effect is 0.217, which is quite similar with the values predicted by the 2SLS linear probability model from Section 1.6.1. Thus, the linear approximation was in fact quite accurate.

A back-of-the-envelope calculation of the overall magnitude of the CRA enforcement mechanism suggests that the average loan approval rate in the sample is about 0.76, which, when multiplied by the average CRA marginal effect of 0.217, yields 0.165. That is, almost every sixth approved loan had been approved due to the CRA-induced incentives. In a sample of $6,533$ loans, this translates into approximately 1078 extra loan approvals. Given that the average loan amount is $\$276,000$, this amounts to $\$297$ million in extra loans given out.

I take these numbers one step further and get an estimate of the AA effect on lending in the whole state of California. I assume that the marginal effect stays at 0.217, which may not be completely justifiable, but is still instructive. Carrying out the same steps as in the previous paragraph suggests that a total of $327,946$ loans were generated in the whole state of California in 2005 by the CRA incentives. This translates into more than $\$90.5$ billion of additional loans. Generalizing this calculation for the rest of the country does not seem possible given various unobservable state-level specific factors.

1.6.3 Discussion

Two points warrant extra discussion at this stage. First, there is vast theoretical literature concerning mechanism design and how to design a contract that aligns the incentives of the other party. The model of costly state verification, first proposed by Townsend (1979), and later adapted by Laffont and Martimort, 2002, appears to be a natural reference point. It is a standard principal-agent model in which an agent has private information on his productivity level. What makes this model different is that principal has the ability to obtain this hidden information, albeit at a cost. The optimal contract is ex post inefficient: the principal commits ex ante to random verification and does just that. Inefficiency is generated because once the principal commits, it is no longer optimal for the agent to misrepresent information. Hence verification is a waste of resources ex post. However, the contract is perfectly efficient ex ante; without the ability of principal to commit to audit, this contract is not incentive compatible for the agent.

Suppose that regulator is the principal, banks are the agents, and that the former wants to ensure compliance with the CRA. Existing mechanism that employs assessment areas can be rationalized using the costly state verification model as follows. If what regulators truly aim to achieve is just more lending in lower-income areas inside AAs, then the current enforcement scheme works well. Banks know that in some areas they will get audited and in some areas they will not, and respond to incentives by lending more inside assessment areas. However, if regulators instead view the CRA as a vehicle to bring more credit to all lower-income areas, existing monitoring mechanism may be used because regulators cannot credibly commit to an ex-post inefficient audit schedule. Whether the inability to commit is a result of regulatory capture seems to be a promising research question in itself.

Second, the results in the previous subsections establish that the incentives created by the CRA monitoring mechanism induced banks to approve a substantial number of mortgage loans that would otherwise not have been approved. Anyone who would like to draw policy implications from this finding would be interested in the post-origination performance of these extra CRA-induced loans. Unfortunately, the HMDA data contains no such information.

Recently, Bajari, Chu, and Park (2009) quantified the importance of different factors that induce subprime borrowers to default on their loans. Using a bivariate probit model with partial observability, they computed how default probabilities respond, on the margin, to changes in different loan-level observables. For example, one of the estimated specifications predicts that a one standard deviation increase in the borrower's monthly payment-to-income ratio raises the probability of default by 17.15 percent, according to Table 5 in that paper.

While the HMDA data has no information concerning the monthly payments, I can approximate these quite well since in the vast majority of cases they are a fixed fraction of the loan size. To do this, I first assume that all the loans in my data are for thirty years and have fixed APRs equal to the average fixed annual mortgage rate for 2005, which was 5.87 percent.[26] The fixed rate assumption has some support in the data: as Section 2.4.1 details, the HMDA respondents must report the interest rate spread on the loans if it is too high. Almost 86 percent of the loans in the data do not report the spread, suggesting the rates were somewhat conventional. I also presume that the applicants have to pay 20 percent of their annual incomes in combined taxes. Finally, I put the total monthly mortgage payment to be the sum of $1/360$th of the loan amount (12 months for 30 years) and the interest payment for the rate specified above.

For the average applicant in the sample, with a loan amount of $\$278,000$ and an annual income of $\$98,000$, this translates into an average payment-to-income ratio of 0.299 with the standard deviation of 0.105. The corresponding numbers for the whole sample of loans used in Bajari, Chu, and Park (2009) are 0.312 and 0.135, which seem pretty close. The fact that the average payment-to-income ratio is somewhat higher is likely due to the primary focus of the study on the subprime loans. I interpret this result as indirect evidence that the CRA-induced loans, on average, are not systematically different from other existing loans. Therefore, it seems natural to expect the holders of those loans to exhibit the same qualitative responsiveness to exogenous market conditions as the more general population of borrowers. Thus, the indirect evidence suggests that the CRA monitoring mechanism created a considerable number of loans to borrowers that do not appear to be systematically more reliable than average subprime borrowers at that time. How-

[26] Source: http://www.mortgage-x.com/.

ever, a more detailed analysis of loan post-origination performance data is due before any definitive conclusions can be drawn.

1.7 Conclusion

In this chapter I have considered the impact of the Community Reinvestment Act monitoring mechanism on the residential mortgage lending market. The Act stimulates depository institutions to lend more in low- and moderate-income census tracts. However, only a certain subset of those tracts matter for the evaluation of the institution's performance under the CRA. A lack of application-level credit score data complicates the process of estimating how CRA enforcement affects banks' decisions to approve mortgage loans.

Using two empirical strategies, I have estimated the effect of interest. The first strategy provides a lower bound on the effect of CRA enforcement rules, and the second strategy yields the upper bound. Therefore, even if both approaches do not address the problem of omitted credit scores completely, the true magnitude of the effect will be bounded between the two estimates.

First, I used techniques developed by the program evaluation literature to construct a sample of observations that would allow me to identify the causal effect of the CRA enforcement scheme on loan approval decisions. The CRA makes its subjects define assessment areas, which are proxies for their primary markets of operations. I have examined the small subsample of loans located in census tracts along the boundaries of assessment areas for identification. This allows me to reduce the effect of unobservable credit scores, and come up with an estimate that may be biased downwards, a fact that allows me to interpret it as a lower bound.

I have used the instrumental variables approach for my second strategy. Specifically, I used the distance from the boundary of the assessment area to the nearest bank branch as an instrument for the AA "treatment." I employed a nonlinear Bayesian IV model to address the endogeneity concern in a probit setting and to account for unobservable heterogeneity among mortgage lenders.

I have found that the need to comply with the CRA has a strong impact on banks' loan approval decisions. Specifically, a CRA-eligible loan has, on average, a 21.7 percent chance of getting approved, other things equal. This suggests that about $95 billion of mortgage lending in California in 2005 was inspired by CRA incentives. The findings indicate that CRA-induced incentives led the banks to originate substantially more loans than they might have otherwise wanted. I interpret these results using the Townsend (1979) costly state verification model to conclude that either regulators have reasons to favor tracts inside AAs more than those outside, or that they are unable to credibly commit to an ex-post wasteful audit mechanism. The question of the post-origination performance of CRA-induced loans seems to be a promising topic for further research.

Chapter 2

Race-to-the-Bottom In Home Mortgage Lending

2.1 Introduction

The first half of the last decade had experienced an across-the-board upward spike in home mortgage loan approvals. There were 19.2 million applications for home mortgage loans in the U.S. in 2000, and 51 percent of those were approved. In contrast, in 2005 the corresponding numbers were 35.4 million and 49.9 percent, respectively. Several competing explanations for this phenomenon were proposed in the literature. For example, Mian and Sufi (2009) suggest that the growth in the supply of mortgage credit by lenders was the primary driver. On the other hand, Demyanyk and Van Hemert (2009) attribute this to the widespread proliferation of securitization practices. Golyaev (2010) comes up with a partial explanation for higher approval rates among the depository institutions: the regulatory pressure of compliance with the Community Reinvestment Act induced its subjects to approve loans more aggressively in closely monitored areas.

The effects of competition among mortgage lenders on their loan approval decisions have not yet been examined in the literature, and I believe this is an important oversight. Recently there had been a considerable amount of research interest in the economic role of market competition in the banking industry. Conventional wisdom would suggest that restraining competitive forces should produce welfare losses. Lenders with high market power are likely to exercise their ability to extract rents by charging higher loan interest rates. These higher rates, in turn, would distort applicants' incentives along the lines of the standard adverse selection argument, thus weakening the stability of mortgage credit markets. Lower supply of credit, associated with higher rates, would also be reflected in a slower process of equity accumulation and, therefore, in lower levels of income per capita. Pagano (1993) and Guzman (2000) develop theoretic underpinnings for these arguments, and Long and Vittas (1991) provide an extensive overview.

Despite the effects outlined above, a number of recent studies pointed out the potential detrimental welfare effects of competition in the financial sector. For example, Dell'Ariccia (2000) explores a model of bank screening and demonstrates that as the number of competing banks increases, the likelihood that banks will actually screen loan applicants decreases. The argument goes that during recessions, screening may be the optimal strategy, since there is a high probability that applicants that demand credit may be of low quality and have already been rejected by other lenders. However, in periods of economic expansion, when there is a higher proportion of new, untested borrowers, lenders competing for market share may choose to offer lending contracts involving no screening. Other studies outlining potential downsides from competition include Petersen and Rajan (1995), Shaffer (1998), and Cao and Shi (2001), among others.

Evaluating the competitive effects on loan approvals correctly should be done empirically using relevant data. To this end, I bring to bear the latest advances in econometric methods of estimating strategic interactions between agents. I write down and estimate the structural model that describes the behavior of a typical mortgage lender. Unlike the models commonly used in the literature, I explicitly allow lenders' decisions to depend on actions of their competitors. This enables me to quantify the optimal response of a lender to the behavior of its competitors. Using the static games estimator developed by Bajari, Hong, Krainer, and Nekipelov (2010), I recover the structural parameters of the model.

I find that it is optimal for mortgage lenders to approve more loans in a given market if they believe their competitors are doing the same. This finding is consistent with the "race-to-the-bottom" story: approving more loans generally translates into lowering loan approval standards. Suppose, for example, that Wells Fargo believes that all its competitors are willing to approve applications from borrowers with credit scores, say, above 700. In this case, the model suggests it is optimal for Wells Fargo to undercut its competition by slightly lowering loan acceptance thresholds. Continuing with the example above, this may translate into Wells Fargo being willing to approve all applicants with credit scores above 690. However, similar argument suggests that Wells Fargo's competitors now may find it optimal to approve applications from people with credit scores as low as 680, and so on. Such reinforcing feedback loop is generally referred to as the "race-to-the-bottom."

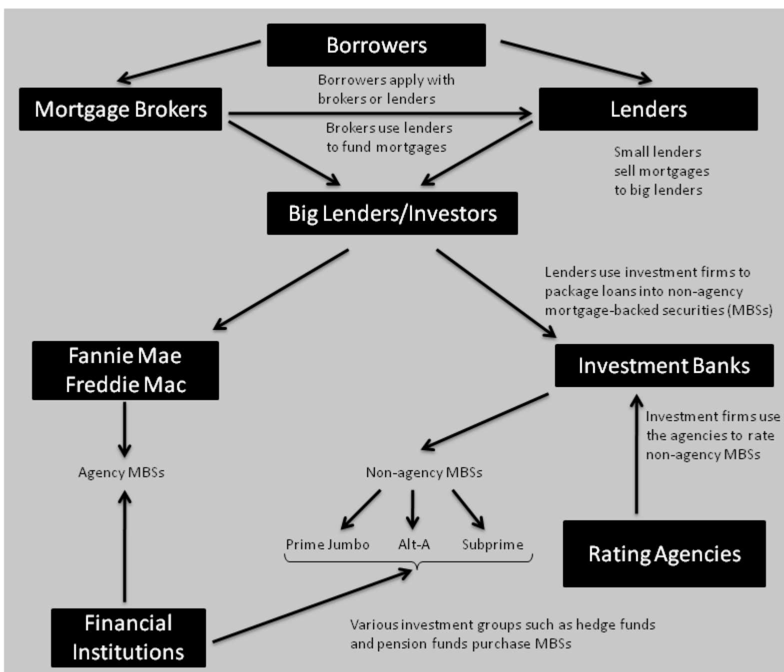
My results also indicate that, depending on the type of lender, it may have a different reaction to the expected actions of its competitors. Lender types are closely related to the identity of the federal regulatory agency that they report to. I demonstrate that lenders regulated by the Federal Reserve behave contrarily to the rest of California lenders. A financial institution supervised by the FRS actually is more likely to approve *fewer* loans if it believe the competitors are expanding the supply of mortgage credit. While I do not have sufficiently reach data to identify the source of such a contrast in optimal reactions, it certainly warrants further inquiry.

The remainder of this paper is organized as follows. Section 2.2 provides the background information about the market for home mortgage loans, and reviews the relevant existing literature. Section 2.3 introduces the model and overviews the estimation algorithm. Section 2.4 describes the data and details the construction of the final dataset. Section 2.5 presents and discusses the results, and Section 2.6 concludes.

2.2 Industry Overview

Mortgage lending is a busy industry with many participating sides. In 2005, some 35.5 million loan applications were recorded, and about 49.9% of those applications were approved. In California, the corresponding numbers were 5.45 million and 51.2%, respectively[1]. The average loan amount in 2005 was $\$183,000$ across the country, and $\$276,000$ in California. Figure 2.1, adapted from Bitner (2008), provides an schematic overview of the industry.

Figure 2.1: Mortgage Lending Industry: A Schematic Overview



Source: Bitner (2008)

[1] As I discuss in Section 2.4.1, these numbers potentially include applications from the same people to multiple banks, so the total number of loan applicants was likely lower.

There are three major types of mortgage lenders:

1. Depository institutions, i.e. banks, savings institutions, thrifts, and credit unions;
2. Mortgage banking subsidiaries of depository institutions (or of bank holding companies);
3. Independent mortgage banks.

There are two primary differences between lender types, and the first one concerns their regulators. State member banks and mortgage subsidiaries of financial holding companies are regulated by the Federal Reserve System (FRS). The large national banks are regulated by the Office of the Comptroller of the Currency (OCC). The Office of Thrift Supervision (OTS) handles thrift and thrift holding companies, whereas state banks that are not members of the FRS fall into the jurisdiction of the Federal Deposit Insurance Corporation (FDIC). The National Credit Union Association (NCUA) oversees credit unions, and the independent mortgage banks have to report to the Department of Housing and Urban Development (HUD).

The other major difference between lender types is in their source of funding. Depository institutions (i.e. "banks") attract money in the form of deposits from their customers, and can use these funds for financing originated mortgage loans. Nondepository institutions (subsidiaries and independent mortgage banks) primarily turn to securitization, which is the process of creating liquid securities out of a collection of illiquid assets.

All the originated mortgage loans that were intended for securitization can be divided in two groups: conforming and non-conforming. Conforming loans are such that satisfy certain criteria established by the GSE (government-sponsored enterprizes, i.e. "Fannie Mae" and "Freddie Mac"). These criteria include loan amount ceilings, lenders credit score floors and other factors, outlined in Keys, Mukherjee, Seru, and Vig (2010). All the other loans fall into the non-conforming category. GSEs turned conforming loans into agency MBSs (mortgage-backed securities), whereas non-conforming loans were turned into non-agency MBSs by other securitizers (usually mortgage banks).

Agency MBSs were considered to be extremely low-risk investments, since GSEs stood behind them and these institutions in turn had implicit guarantees from the U.S. government. These guarantees became explicit on September 7, 2008, when U.S. Treasury placed Fannie Mae and Freddie Mac into conservatorship.

The non-agency MBSs can be divided into three groups:

1. Prime jumbo – loans made to people with good credit ("prime" borrowers) but such that the loan amounts were too large to be conforming;
2. Subprime – risky loans made to people with lower credit scores below the cutoffs established by the GSEs;
3. Alt-A – loans made to borrowers with decent credit but who usually lacked complete documentation, moderately risky loans (more risky than prime but

less risky than subprime).

Because non-agency MBSs did not entertain the implicit guarantees of their agency counterparts, securitizers at the investment banks like Bear Sterns and Lehman Brothers had to turn to rating agencies like Fitch and Moody's. The job of a rating agency was to assess the inherent risk of these new securities, and to assign them a rating based on their own internal criteria. This process was extremely important for the makers of the MBSs because many large investors had explicit guidelines that do not allow investing into very risky assets. To make MBSs marketable to a general pool of investors they had to be rated as sufficiently low-risk investments. Benmelech and Dlugosz (2009) is an excellent source on this matter.

The complete structure of the mortgage origination industry is therefore quite complex. I focus on the very first stage of the overall process, which is the loan approval decision by lenders. I treat the actions of all other industry actors as given.

2.3 Estimation

2.3.1 Estimation Method

I model the interaction between mortgage lenders as a static game with incomplete information. The following exposition is based on Bajari, Hong, Krainer, and Nekipelov (2010). Consider a finite number of players, $i = 1, \ldots, I$; each player can simultaneously choose an action $a_i \in \{1, \ldots, K\}$ out of a finite set.

In my application, players are the mortgage lenders, and the actions are defined as follows:

- $a_i = 1 \iff$ lender i approves less than 10 percent of the mortgage loan applications received;
- $a_i = 2 \iff$ lender i approves less than 20 percent (but at least 10 percent) of the mortgage loan applications received;

  $\vdots$
- $a_i = 10 \iff$ lender i approves less than 100 percent (but at least 90 percent) of the mortgage loan applications received;
- $a_i = 11 \iff$ lender i approves all the mortgage loan applications received (so $K = 11$).[2]

Clearly, players have the same set of actions in this setting. Let $A = \{1, \ldots, K\}^I$ denote the vector of possible actions for all players. I will use $a = (a_1, \ldots a_I)$ to denote a typical element of A, and I let $a_{-i} = (a_1, \ldots a_{i-1}, a_{i+1}, \ldots, a_I)$ denote a vector of strategies for all players with the exception of player i.

Let $s_i \in S_i$ denote the state variable for player i. Let $S = \prod_{i=1}^{I} S_i$ and let $s = (s_1, \ldots, s_I) \in S$ denote a vector of state variables for all I players. I assume s is common knowledge to all players and that s is observable to me.

[2] I have chosen to put the continuous variable "share of received applications that got approved" into 11 discrete bins. This way the framework from Bajari, Hong, Krainer, and Nekipelov (2010) can be applied directly. The last bin was motivated by the considerable number of observations with approval share of 100 percent.

For each player, there are also K state variables labeled as $\varepsilon_i(a_i)$ which are private information to player i. These are independent and identically distributed across i and a_i. Let ε_i denote the $1 \times K$ vector of the individual $\varepsilon_i(a_i)$, and $f(\varepsilon_i(a_i))$ be the density of $\varepsilon_i(a_i)$.

The period utility function for player i is

$$U_i(a, s, \varepsilon_i, \theta) = v_i(a_i, a_{-i}, s, \theta) + \varepsilon_i(a_i) \qquad (2.3.1)$$

This utility specification in this model is quite similar to the one that arises in discrete-choice demand estimation problems. Each player i receives a stochastic preference shock, $\varepsilon_i(a_i)$, for each possible action a_i. In many applications, this will be drawn from an extreme value distribution as in the multinomial logit model. Starting with Rust (1994), $\varepsilon_i(a_i)$ is sometimes alternatively interpreted as an unobserved state variable. The first term in equation (2.3.1) that depends on the vector of state variables s and actions a, is commonly assumed to be a linear function of actions and states.

The key distinction from a standard random utility model is that the actions a_{-i} of other players in the game enter into i's utility. A standard discrete choice model assumes that agents act in isolation in the sense that a_{-i} is omitted from the utility function of i. In my application this assumption can be quite restrictive.

Player i's decision rule is a function $a_i = \delta_i(s, \varepsilon_i)$. Since ε_{-i} are private information to the other $-i$ players in the game and are unobservable to i, i's decisions do not depend on them. Define $\sigma_i(a_i \mid s)$ as

$$\sigma_i(a_i = k \mid s) = \int 1\{\delta_i(s, \varepsilon_i) = k\} f(\varepsilon_i(a_i))\ d\varepsilon_i(a_i) \tag{2.3.2}$$

In Equation (2.3.2), $1\{\delta_i(s, \varepsilon_i) = k\}$ is the indicator function that player i's action is k given the vector of state variables (s, ε_i). Therefore, $\sigma_i(a_i = k \mid s)$ is the probability that i chooses action k conditional on the state variables s that are public information. I define the distribution of a given s as $\sigma(a \mid s) = \prod_{i=1}^{I} \sigma_i(a_i \mid s)$.

Next, I rewrite $U_i(a_i, s, \varepsilon_i, \theta)$ as

$$U_i(a_i, s, \varepsilon_i, \theta) = V_i(a_i, s, \theta) + \varepsilon_i(a_i) \tag{2.3.3}$$

where $V_i(a_i, s, \theta)$ is the deterministic part of the expected payoff, defined as

$$V_i(a_i, s, \theta) = \sum_{a_{-i}} \upsilon_i(a_i, a_{-i}, s, \theta)\, \sigma_{-i}(a_{-i} \mid s)\,. \tag{2.3.4}$$

and where, in turn,

$$\sigma_{-i}(a_{-i} \mid s) = \prod_{h \neq i} \sigma_l(a_h \mid s)\,. \tag{2.3.5}$$

In Equation (2.3.3) above, $U_i(a_i, s, \varepsilon_i, \theta)$ is player i's expected utility from choosing a_i given parameters θ. Since i does not know the private information shocks, ε_j for the other players, i's beliefs about their actions are given by $\sigma_{-i}(a_{-i} \mid s)$ defined in (2.3.5). The term $V_i(a_i, s, \theta)$ is the expected value of $V_i(a_i, s, \varepsilon_i, \theta)$, marginalizing out the strategies of the other players using $\sigma_{-i}(a_{-i} \mid s)$. The structure of payoffs in (2.3.3) is quite similar to discrete-choice demand models, except that the probability distribution over other agents' actions enter into the formula for agent i's utility.

Note that if the error term ε_i has an atomless distribution for all $i = 1, \ldots, I$, then player i's optimal action is unique with probability one. This conveniently eliminates the need to consider mixed strategies. It follows immediately that the optimal action a_i^* for player i satisfies

$$\sigma_i(a_i^* \mid s) = \Pr\{\varepsilon_i \mid V_i(a_i^*, s, \theta) + \varepsilon_i(a_i^*) > V_i(a_h, s, \theta) + \varepsilon_i(a_h), \quad \forall h \neq i\}. \tag{2.3.6}$$

The next step is to define

$$\Gamma_i(\theta, \sigma_h(k \mid s), \forall h = 1, \ldots, I, \ \forall k = 1, \ldots, K) = \sigma_i(a_i \mid s) \tag{2.3.7}$$

and note that Equation (2.3.7) defines a fixed point equation in $\sigma_i(\cdot)$. The Brouwer's fixed point theorem guarantees the solution exists.

2.3.2 Implementation

In the data, I observe $j = 1, \ldots, J$ repetitions of the game across markets. Let $a_{i,j}$ denote the action of player i in market j and let s_j be the state in that market. By observing players' behavior across markets, I can form a consistent estimate

$\hat{\sigma}_i(a_i \mid s)$ of $\sigma_i(a_i \mid s)$ for $i = 1, \ldots, I$. In my application, this simply involves flexibly estimating $\Pr\{a_i = k\}$, conditional on a set of covariates, something that could be implemented using a number of standard techniques. Given first-stage estimates of $\hat{\sigma}_i(a_i \mid s)$, I could then estimate the structural parameters of the payoff, θ, by inverting the functions $\Gamma_i(\cdot)$ defined in Equation (2.3.7).

The key problem with identification in this approach is that both the first-stage estimates $\hat{\sigma}_i(a_i \mid s)$ and the term $\Pi_i(a_i, s, \theta)$ depend on the vector of state variables s. This will introduce a collinearity problem when I attempt to separately identify the effect of θ on the observed choices. The standard solution to this type of problem is to impose an exclusion restriction. Suppose that in my application s_i represents how much assets lender i has at its disposal. The exclusion restriction would require me to assume that when lender i decides on how many loans to approve, it does not take assets of other $-i$ lenders into account.

An observation is the activity of lender in a given market. I will use $i = 1, \ldots I$ to denote a lender and $j = 1, \ldots, J$ to denote a market. We will denote a particular approval decision by $a_{i,j}$. Since the dependent variable can be naturally ranked from highest to lowest, I assume that the utilities follow an ordered logit model. Let $s(i, j)$ denote a set of covariates that determine the actions of lender i in market j. Let $s(j)$ denote a vector of $(s(i, j))$ of payoff relevant covariates that enter into the utility of all the lenders that serve market j. Let $z(j)$ denote a set of covariates that shift the equilibrium, but which do not influence payoffs.

Define the utility of lender i from operating in market j to be,

$$v_{i,j} = \beta' s\left(i,j\right) + \eta E\left[a \mid s\left(j\right), z\left(j\right)\right] + \varepsilon_{i,j} \tag{2.3.8}$$

In Equation (2.3.8), the term $E\left[a \mid s\left(j\right), z\left(j\right)\right]$ is the expected loan approvals in market j and $\varepsilon_{i,j}$ is an error term drawn from an logistic distribution. The model is the well-studies ordered logit, where the probability that a particular approval ratio is observed is determined as follows, where I define $\mu_0 = 0$

$$\Pr\left\{a = 1\right\} = \Lambda\left(-\beta' s\left(i,j\right) - \eta E\left[a \mid s\left(j\right), z\left(j\right)\right]\right)$$

$$\begin{aligned} \Pr\left\{a = k\right\} &= \Lambda\left(\mu_{k-1} - \beta' s\left(i,j\right) - \eta E\left[a \mid s\left(j\right), z\left(j\right)\right]\right) \\ &\quad -\Lambda\left(\mu_{k-2} - \beta' s\left(i,j\right) - \eta E\left[a \mid s\left(j\right), z\left(j\right)\right]\right), \quad k = 2, \ldots, 10 \end{aligned} \tag{2.3.9}$$

$$\Pr\left\{a = 11\right\} = 1 - \Lambda\left(\mu_9 - \beta' s\left(i,j\right) - \eta E\left[a \mid s\left(j\right), z\left(j\right)\right]\right)$$

In Equation (2.3.9), the likelihood that determines the probability that the approval ratio is a depends on the latent estimated covariates β and η along with the cut points μ. Identification of this model depends crucially on having appropriate exclusion restrictions. I assume that the approval decisions of lender i are independent of characteristics of applicants that come to other lenders. For example, Wells Fargo does not care what kind of applicants turn to Washington Mutual; instead, it only cares about its own applicants. I believe that this is a reasonable assumption.

2.4 Data

2.4.1 Data Sources

The primary data source for this study is the Home Mortgage Disclosure Act data set (HMDA, pronounced "humda"), made available by the Federal Financial Institutions Examination Council (FFIEC). It contains a majority of all home mortgage loan applications in the U.S. The potential mortgage originators (called "respondents" in the HMDA language) are the main subjects of the HMDA.

An observation in the dataset is a loan application, and several important characteristics of the application are available. These can be divided into 3 major groups:

1. Borrower's characteristics, such as race, gender, ethnicity and income.
2. Respondent's characteristics, such as name, type, address, parent company name (if applicable), supervisor's identity.
3. Loan characteristics, such as amount, property address (aggregated up to a census tract), various type measures (single or multi-family; conventional, FHA or VA[3]; owner-occupied or not; new or refinanced loan, etc). The important loan characteristics are the decision taken on the application, and

[3] FHA stands for "Federal Housing Administration", and VA is an acronym for "Veteran Affairs". Loan applications with such labels are usually at least partially subsidized by the federal government.

the reasons for denial (if applicable).

The FFIEC aims to preserve the anonymity of mortgage applicants by not disclosing the application date, which is rounded to a calendar year. A few potentially interesting covariates are not present in the HMDA data. There is no information on the term structure of the loan (15-year mortgage or 30-year mortgage), as well as very limited information on the loan interest rate. HMDA respondents must report the difference between the loan APR (annual percentage rate) and the rate on Treasury securities of comparable maturity, as long as the spread is above the designated threshold. It is also not possible to tell apart the fixed rate and adjustable rate mortgages in the data. On average, there were 31 million loan applications per year in the HMDA data (during 2000-2005). Avery, Brevoort, and Canner (2006, 2007) are the two best sources that discuss this data extensively.

2.4.2 Dataset Construction

I choose to look at the loan applications in California in 2005. By looking within a single state, I abstract from inter-state variations in laws that regulate banks' operations. California is one of the largest states, and its home mortgage lending market makes up 14.8 percent of the overall U.S. market (on average between 2000 and 2005). By looking at 2005, I concentrate on one of the last years before the mortgage crisis started to unfold.

I keep only conventional loans (no FHA or VA applications). I exclude loans that are not for single family owner-occupied homes, and those that are secured by anything other than a primary lien. Non-primary liens and non-single family loans, in practice, are more likely to be associated with real estate speculative purchases, especially during the time of interest. I drop all applications to credit unions, since there are usually only a few applications per credit union in a year, and together these comprise less than one percent of the data.

The HMDA data reports the decision taken on each application, and there are ten different decisions that can be reported. I exclude all loans that were purchased from another institution, because the decision on those had been taken by some other entity. I drop applications with decision reported as "application withdrawn", which are thought to be associated with indirect lending through mortgage brokers rather that directly through lenders of interest. For every lender, I construct a census-tract level aggregate measure of the loan-level characteristics of the applicants. Borrowers can apply for loans with different banks at the same time, and there is no way to identify two different loan applications with a single borrower in the data. My primary interest, however, is in banks' approval decisions, and it is quite possible that different banks assess the same person's application differently.

2.4.3 Preliminary Evidence

The final data set contains information on 459 competing lenders that operate in $6,569$ distinct markets. A market is defined as a census tract, and I exclude six markets that only had applications to a single lender: it would not be possible to identify competitive forces in those markets. I also exclude markets where I observe more than 40 loan applications; 99 percent of the data falls under this cutoff.[4]

Table 2.1 presents a summary of the data used. I tabulate the top ten lenders split by their regulatory agency, and I list top twenty HUD-regulated lenders. For every lender, I present the total number of loan applications, the number of approved applications, and the approval ratio.

Table 2.1: Top Lenders, By Regulatory Agency

Lender Name	Applications	Loans Approvals	Ratio	Lender Name	Applications	Loans Approvals	Ratio
Regulated by the Office of the Comptroller of the Currency (OCC)				Regulated by the Office of Thrift Supervision (OTS)			
Wells Fargo Bank	114,960	100,010	0.870	Washington Mutual Bank	131,277	98,770	0.752
National City Bank of Indiana	66,163	54,781	0.828	World Savings Bank	114,560	76,536	0.668
Bank of America	59,345	49,833	0.840	Downey Savings and Loan	61,612	44,024	0.715
Countrywide Bank	54,391	23,321	0.429	BNC Mortgage	41,538	24,758	0.596
JPMorgan Chase Bank	32,594	27,242	0.836	IndyMac Bank	29,990	21,252	0.709
ABN Amro Mortgage Group	19,698	14,473	0.735	Citimortgage	21,615	15,713	0.727
First Horizon Home Loan Corp	9,928	7,690	0.775	Citicorp Trust Bank	16,229	6,446	0.397
Chase Manhattan Bank	8,123	3,875	0.477	Finance America	14,717	6,877	0.467
HSBC Mortgage Corporation	7,353	6,339	0.862	AIG Federal Savings Bank	13,094	10,715	0.818
First National Bank of Arizona	5,214	3,128	0.600	First Federal Bank of CA	13,051	9,543	0.731
Regulated by the Federal Reserve System (FRS)				Regulated by the Department of Housing and Urban Development (HUD)			
Countrywide Home Loans	180,473	122,197	0.677	Ameriquest Mortgage Company	142,252	17,009	0.120
Beneficial Homeowners Service	31,852	4,313	0.135	Argent Mortgage Company	69,096	33,700	0.488
HFC Company LLC	27,692	3,557	0.128	New Century Mortgage Corp	65,832	48,694	0.740
Wells Fargo Financial	21,281	5,499	0.258	WMC Mortgage Corp	58,743	31,616	0.538
Suntrust Mortgage	12,411	11,623	0.937	GMAC Mortgage Corporation	57,628	24,653	0.428
Decision One Mortgage	11,335	8,727	0.770	Town & Country Credit Corp	47,333	4,003	0.085
Equifirst Corporation	7,202	3,379	0.469	Long Beach Mortgage Comp	38,317	25,012	0.653
RBC Mortgage	5,502	4,511	0.820	Centex Home Equity Comp	37,723	2,346	0.062
First Bank	5,309	2,881	0.543	Encore Credit Corp	32,572	17,598	0.540
Countrywide Mtg. Ventures	4,039	2,672	0.662	MortgageIt	29,329	25,243	0.861
Regulated by the Federal Deposit Insurance Corporation (FDIC)				⋮	⋮	⋮	⋮
Greenpoint Mortgage Funding	36,415	27,427	0.753	Option One Mortgage Corp	29,314	22,880	0.781
Fremont Investment & Loan	34,210	21,716	0.635	American Home Mortgage Corp	21,411	19,049	0.890
Gateway Business Bank	3,418	2,538	0.743	Scme Mortgage Bankers, Inc	20,052	15,978	0.797
Fremont Bank	2,994	2,447	0.817	Accredited Home Lenders, Inc	15,744	8,207	0.521
Merrill Lynch Credit Corp	1,221	946	0.775	American Mortgage Network	14,950	11,632	0.778
Bank of the West	708	517	0.730	First NLC Financial Services	14,897	8,335	0.560
Franklin Bank	701	483	0.689	Sierra Pacific Mortgage	14,126	10,739	0.760
California Bank & Trust	627	488	0.778	First Magnus Financial Corp	13,126	10,962	0.835
First Republic Bank	410	361	0.880	People's Choice Financial Corp	12,873	8,928	0.694
Tri Counties Bank	398	314	0.789	Aames Funding Corporation	12,630	5,175	0.410

Source: 2005 HMDA data.

[4] I have tested the robustness of the results to this cutoff, and the qualitative results are not sensitive to this truncation.

Table 2.1 clearly indicates that there is plenty of variation in the data that can be exploited towards identifying the parameters of interest. What is also needed, however, is variation in lender-specific state variables, and Table 2.2 summarizes those.

Table 2.2: Summary of Applications, By Regulatory Agency

	Regulatory Agency					
	OCC	FRS	FDIC	OTS	HUD[a]	Overall
Average Loan Size, $1000	339.58	321.06	362.56	342.02	332.91	335.19
Average Applicants Income, $1000	105.38	93.01	113.59	103.59	99.10	100.53
Share of White Applicants	62.03	63.89	63.65	59.80	56.55	58.48
Share of Female Applicants	37.30	37.74	36.94	38.81	39.33	38.82
Share of Applications with Co-Applicants	47.39	46.13	43.63	43.94	37.37	40.44
Share of Lenders That are Subsidiary	26.93	92.28	37.39	19.83	5.76	18.04
Share of Lenders With Assets Over $7.5 billion	64.39	67.03	36.38	61.43	1.96	24.89
Number of Lenders	31	28	21	52	327	459
Largest 10 by Application Count, % of Total	94.18	61.49	80.24	84.23	59.41	38.60
Average Approval Ratio Across Largest 10	72.51	53.99	75.90	65.81	57.69	66.97
Average Approval Ratio Across All Lenders	70.70	61.49	80.24	68.78	69.40	69.43

[a] For HUD-regulated lenders, I compute statistics over top 20 lenders, instead of over top 10.

Source: 2005 HMDA data. Except for rows 1, 2, and 8, all other rows are in percent.

The first seven rows summarize lender-specific state variables that were aggregated up from the level of individual applications to lender-market level averages. I chose to coarsen the lenders' assets information to a binary variable of whether a given lender is "large" in a sense that its assets exceed 7.5 billion dollars. This number represents the 75th percentile of assets distribution, and this choice allows me to avoid the effect of extreme outliers on my estimates – some of the lenders have a lot more assets than the average lender. In fact, the average assets are equal to the 90th percentile of the assets distribution, suggesting extreme right skewedness.

Some preliminary results can be seen from Table 2.2. Lenders regulated by the FRS appear, on average, to have lower approval ratios than all other lenders, and the pattern is even stronger if I focus only on the largest ten institutions. FRS respondents also have the highest fraction of "large" lenders among them, as measured by their assets. At this point, however, it is not possible to conclude that these lenders are more "prudent": lower approval ratios can also be caused by unobservable selection of applicants when they decide which lender to use. While accounting for this selection is beyond the scope of this paper, the raw data suggests that FRS-regulated lenders are somehow distinct from the rest of the competition.

2.5 Estimation Results

Table 2.3 presents the results of the second stage estimation, as discussed in Section 2.3. The dependent variable is a_{ij} – approval ratio bin for lender i in market j, and it takes the values from 1 to 11. The key explanatory variable is the average approval decisions by competitors $-i$. The first estimated equation "pools" all lenders, whereas the second allows for heterogeneity based on the regulator's identity.

The pooled model suggests that the "race-to-the-bottom" story is consistent with the data. For a given lender, there is a clear positive relation between the number of loans it approves and the expected approvals of its competitors. The next rows contain lender-specific information about its pool of applicants. For example, the share of applications with co-applicants is the fraction of number of applications that had a co-applicant listed to the total number of applications. As one would expect, such applications tend to be more reliable, and higher percentage of those translates into more approvals. Interestingly, "larger" lenders appear to have lower approval ratios than "smaller" lenders, where "size" of a lender is measured by its assets. However, in most cases larger institutions also receive more loan applications. As such, it is not possible to assign a causal interpretation to these estimates.

The last column of Table 2.3 presents estimation results for the model where I allow the lenders' beliefs to vary across lender types. It is instructive to compare the five estimates from the last column with the single estimate from the pooled model. The pooled estimate is closest to those for the HUD-regulated lenders, which is perhaps not surprising, given that they comprise over 60 percent of the data.

Table 2.3: Share of Applications Approved

	Pooled Model	Split by Regulator
Average Approvals by Competitors	0.022	
	(0.008)	
Average Approvals by Competitors, OCC Lenders		0.032
		(0.008)
Average Approvals by Competitors, FRS Lenders		−0.045
		(0.008)
Average Approvals by Competitors, FDIC Lenders		0.038
		(0.008)
Average Approvals by Competitors, OTS Lenders		0.033
		(0.008)
Average Approvals by Competitors, HUD Lenders		0.020
		(0.008)
Mean Loan Amount, $1000	−0.001	−0.001
	(0.00003)	(0.00003)
Mean Applicants' Income, $1000	0.004	0.004
	(0.00008)	(0.00008)
Share of Female Applicants	−0.163	−0.164
	(0.007)	(0.007)
Share of White Applicants	0.516	0.522
	(0.006)	(0.006)
Share of Applications with Co-Applicants	0.291	0.291
	(0.007)	(0.007)
Lender Is a Subsidiary	−0.309	−0.146
	(0.006)	(0.008)
Lender Has Assets Over $7.5 billion	−0.073	−0.064
	(0.005)	(0.008)
Number of Observations	578,040	578,040
Auxiliary Parameters	Yes	Yes

Dependent variable is the approval bin, see Section 2.3. Standard errors in parentheses.
The estimated equation is ordered logistic regression, auxiliary parameters include the cutoff estimates $\hat{\mu}$.

Remarkably, the competitive effects among lenders regulated by the FRS are entirely different from those among the remaining lenders. Specifically, when a FRS-regulated institution believes that its competitors will approve more applications, its optimal response is to tighten its loan evaluation criteria, in contrast to others. This finding is consistent with my preliminary data exploration discussed in Section 2.4.3.

Table 2.4: Approximated Marginal Effects for Lenders' Beliefs

	Pooled Model	Split by Regulator
Average Approvals by Competitors	0.016 (0.016)	
Average Approvals by Competitors, OCC Lenders		0.063 (0.016)
Average Approvals by Competitors, FRS Lenders		-0.172 (0.016)
Average Approvals by Competitors, FDIC Lenders		0.084 (0.016)
Average Approvals by Competitors, OTS Lenders		0.042 (0.016)
Average Approvals by Competitors, HUD Lenders		0.012 (0.016)
Number of Observations	$578,040$	$578,040$
Auxiliary Parameters	Yes	Yes

Dependent variable is the approval bin, see Section 2.3. Standard errors in parentheses.
The estimated equation is simple linear regression, auxiliary parameters include those featured in Table 2.3.

Because the main model is nonlinear, it is difficult to interpret its estimates directly. For this reason, I present Table 2.4, where I replace the ordered logistic regression with the simple linear regression. This way the estimates can be interpreted as incremental changes in approval ratios. The corresponding table for the nonlinear model would necessarily include multiple rows and columns, whereas Table 2.4 is effectively a first-order approximation of it. It is clear that the patterns suggested by the coefficients persist for the marginal effects.

It would be extremely insightful to dig deeper into what makes FRS respondents behave differently from others. Unfortunately, there is not enough richness in the data to address this question: only the identities of the regulators are known. This seems to be a promising topic for further research.

2.6 Conclusion

Conventional wisdom suggesting that limiting competition among lenders should produce welfare losses had recently been reconsidered in the literature. Several recent studies outlined the possible negative welfare effects of competition in the financial sector. The effects of competition among mortgage lenders on their loan approval decisions have not yet been studied, and I address this important oversight in this paper.

I employ the recently developed econometric methods of estimating strategic interactions between agents from Bajari, Hong, Krainer, and Nekipelov (2010). I find that it is optimal for mortgage lenders to approve more loans in a given market if they believe their competitors are doing the same. I label this finding as the "race-to-the-bottom" story: approving more loans translates into lowering loan approval standards. I also demonstrate that lenders regulated by the Federal Reserve behave quite differently than others: such lenders are actually more likely to approve *fewer* loans if they expect the competitors to relax their approval standards. While I do not have sufficiently reach data to identify the source of such a contrast in optimal reactions, it remains to be a promising topic for further research.

Bibliography

AVERY, R. B., R. W. BOSTIC, AND G. B. CANNER (2000): "The Performance and Profitability of CRA-Related Lending," *Economic Commentary*.

AVERY, R. B., K. P. BREVOORT, AND G. B. CANNER (2006): "Higher-Priced Home Lending and the 2005 HMDA Data," *Federal Reserve Bulletin*, 92.

——— (2007): "Opportunities and Issues in Using HMDA Data," *Journal of Real Estate Research*, 29(4), 351–380.

AVERY, R. B., P. S. CALEM, AND G. B. CANNER (2003): "The Effects of the Community Reinvestment Act on Local Communities," *Unpublished paper. Board of Governors of the Federal Reserve System*.

BAIR, S. (2008): "Did Low-income Homeownership Go Too Far?," in *Prepared remarks for Conference before the New America Foundation.*

BAJARI, P. L., S. CHU, AND M. PARK (2009): "An Empirical Model of Subprime Mortgage Default from 2000-2007," .

BAJARI, P. L., H. HONG, J. KRAINER, AND D. NEKIPELOV (2010): "Estimating Static Models of Strategic Interaction," *Journal of Business and Economic Statistics*, 28(3), 469–482.

BAKER, G. P., AND T. N. HUBBARD (2004): "Contractibility And Asset Ownership: On-Board Computers and Governance In US Trucking," *Quarterly Journal of Economics*, 119(4), 1443–1479.

BENMELECH, E., AND J. DLUGOSZ (2009): "The Alchemy of CDO Credit Ratings," *Journal of Monetary Economics*, 56(5), 617–634.

BERRY, C. R., AND S. L. LEE (2007): "The Community Reinvestment Act: A Regression Discontinuity Analysis," .

BERRY, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, pp. 242–262.

BERRY, S. T., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

BHUTTA, N. (2008): "Giving Credit where Credit is Due? The Community Reinvestment Act and Mortgage Lending in Lower Income Neighborhoods," *Board of Governors of the Federal Reserve Finance and Economics Discussion Paper*, 61.

BITNER, R. (2008): *Confessions of a Subprime Lender: An Insider's Tale of Greed, Fraud, and Ignorance*. Wiley.

BLACK, S. E. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *The Quarterly Journal of Economics*, 114(2), 577–599.

BLUNDELL, R. W., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, vol. 2, pp. 655–679.

——— (2004): "Endogeneity in Semiparametric Binary Response Models," *The Review of Economic Studies*, 71, 655–679.

CAO, M., AND S. SHI (2001): "Screening , Bidding , and the Loan Market Tightness," *European Finance Review*, 5, 21–61.

COPELAND, A. M., AND C. MONNET (2009): "The Welfare Effects of Incentive Schemes," *The Review of Economic Studies*, 76(1), 93–113.

CRA Reference (2005): "A Bankers Quick Reference Guide to CRA," *Federal Reserve Bank of Dallas*, Available at http://www.dallasfed.org/htm/pubs/pdfs/ca/quickref.pdf.

DAFNY, L. S. (2005): "How Do Hospitals Respond to Price Changes?," *The American Economic Review*, 95(5), 1525–1547.

DELL'ARICCIA, G. (2000): "Learning by Lending, Competition, and Screening Incentives in the Banking Industry," *Working Paper, IMF*.

DEMYANYK, Y., AND O. VAN HEMERT (2009): "Understanding the Subprime Mortgage Crisis," *The Review of Financial Studies*.

ERGUNGOR, O. E. (2007): "Foreclosures in Ohio: Does Lender Type Matter?," *Federal Reserve Bank of Cleveland, Working Paper 0724*.

GANDHI, A., K. I. KIM, AND A. PETRIN (2010): "The Interaction of Observed and Unobserved Factors in Non-Linear Demand Models," *Working Paper, University of Minnesota*.

GEWEKE, J. (1991): "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities," in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 571–578. Citeseer.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. Wiley-Interscience.

GEWEKE, J., G. GOWRISANKARAN, AND R. J. TOWN (2003): "Bayesian Inference for Hospital Quality in a Selection Model," *Econometrica*, 71(4), 1215–1238.

GOLYAEV, K. (2010): "The Impact of the Community Reinvestment Act on the Home Mortgage Lending Industry in California," *Working Paper, University of Minnesota*, pp. 1–55.

GUZMAN, M. G. (2000): "Bank Structure, Capital Accumulation and Growth: A Simple Macroeconomic Model," *Economic Theory*, 16(2), 421–455.

HO, K., AND J. ISHII (2010): "Location and Competition in Retail Banking," .

HOLMES, T. J. (1998): "The Effect of State Policies on the Location of Manufacturing: Evidence From State Borders," *The Journal of Political Economy*, 106(4), 667–705.

IMBENS, G. W., AND T. LEMIEUX (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142(2), 615–635.

Joint Center for Housing Studies (2002): "The 25th Anniversary of the Community Reinvestment Act: Access to Capital in the Evolving Financial Services System," .

KEYS, B. J., T. MUKHERJEE, A. SERU, AND V. VIG (2010): "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *The Quarterly Journal of Economics*, 125(1), 307–362.

KLEINER, M. M., AND R. M. TODD (2007): "Mortgage Broker Regulations That Matter: Analyzing Earnings, Employment, and Outcomes for Consumers," .

LAFFONT, J.-J., AND D. MARTIMORT (2002): *The Theory of Incentives: the Principal-Agent Model*. Princeton University Press.

LIEBOWITZ, S. J. (2009): "Anatomy of a Train Wreck: Causes of the Mortgage Meltdown," .

LONG, M., AND D. VITTAS (1991): "Financial Regulation Changing the Rules of the Game," *World Bank Policy Research Paper*.

MIAN, A., AND A. SUFI (2009): "The Consequences of Mortgage Credit Expansion: Evidence From the U. S. Mortgage Default Crisis," *The Quarterly Journal of Economics*, 124(November), 1449–1497.

MISRA, S., AND H. NAIR (2009): "Structural Model of Sales-Force Compensation Dynamics: Estimation and Field Implementation," *Quantitative Marketing and Economics*.

PAGANO, M. (1993): "Financial Markets and Growth: An Overview," *The European Economic Review*, 37, 613–622.

PETERSEN, M. A., AND R. G. RAJAN (1995): "The Effect of Credit Market Competition on Lending Relationships," *The Quarterly Journal of Economics*, 110(2), 407–443.

ROSSI, P. E., G. M. ALLENBY, AND R. E. MCCULLOCH (2006): *Bayesian Statistics and Marketing*. Wiley.

RUST, J. (1994): *Structural Estimation of Markov Decision Processes*vol. 4, pp. 3081–3143. North Holland.

SHAFFER, S. (1998): "The Winners Curse in Banking," *The Journal of Financial Intermediation*, 7(4), 359–392.

SILVERMAN, E., AND J. SKINNER (2004): "Medicare Upcoding and Hospital Ownership," *The Journal of Health Economics*, 23(2), 369–389.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

TANNER, M. A., AND W. H. WONG (1987): "The Calculation of Posterior Distributions by Data Augmentation," *The Journal of the American Statistical Association*, 82(398), 528–540.

TOWNSEND, R. M. (1979): "Optimal Contracts and Competitive Markets with Costly State Verification," *The Journal of Economic Theory*, 21(2), 265–293.

Appendix A

Implementing Bayesian Instrumental Variables Estimation

Two equations:

$$\begin{cases} y_i^* = AA_i\beta + x_i'\gamma + \varepsilon_{i,1} \\ AA_i = z_i'\delta + \varepsilon_{i,2} \end{cases}.$$

y^* is unobservable, but $y = 1\{y^* \geq 0\}$ is. Here $z_i = (dist_i, x_i)$. Assumptions:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

and priors:

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \sim N\left(\begin{pmatrix} \overline{\beta} \\ \overline{\gamma} \end{pmatrix}, A_{\beta\gamma}^{-1}\right),$$
$$\delta \sim N\left(\overline{\delta}, A_{\delta}^{-1}\right),$$
$$\Sigma \sim IW\left(\upsilon_0, V_0\right).$$

Gibbs sampling is used to obtain draws from the posterior, and there are four steps to it:

$$\begin{aligned} \{y_i^*\}_{i=1}^n &\mid \beta, \gamma, \delta, \Sigma, y, AA, z \\ \beta, \gamma &\mid \delta, \Sigma, \{y_i^*\}_{i=1}^n, AA, z \\ \delta &\mid \beta, \gamma, \Sigma, \{y_i^*\}_{i=1}^n, AA, z \\ \Sigma &\mid \beta, \gamma, \delta, \{y_i^*\}_{i=1}^n, AA, z, \end{aligned}$$

I detail each step below.

A.1 Step 1. Updating y^*.

For each i, conditional on $\varepsilon_{i,2}$, y_i^* is normal:

$$y_i^* \mid \varepsilon_{i,2} \sim N\left(AA_i\beta + x_i'\gamma + \frac{\sigma_{12}}{\sigma_2^2}\varepsilon_2, \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right).$$

Hence, I need to make draws from this distribution. The important part is to account for the data on y_i, so that I would not draw $y_i^* < 0$ is $y_i = 1$ and vice versa. This means draws must be made from the truncated normal distribution, which is done via the mixed rejection algorithm of Geweke (1991) covered in Appendix B.

A.2 Step 2. Updating β and γ.

Given δ, I can compute ε_2 as $\varepsilon_{i,2} = AA_i - z_i'\delta$. Given the output of the data augmentation step, I treat y^* as observables now, and condition the equation for y^* on ε_2:

$$y_i^* = AA_i\beta + x_i'\gamma + \frac{\sigma_{12}}{\sigma_2^2}\varepsilon_{i,2} + \xi_{i,1|2},$$

where $Var\left[\xi_{1|2}\right] = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}$. Denote $\tau^2 \equiv \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}$, then rewrite the above equation as

$$\frac{y_i^* - \frac{\sigma_{12}}{\sigma_2^2}\varepsilon_{2i}}{\tau} = \frac{AA_i}{\tau}\beta + \frac{x_i{}'}{\tau}\gamma + \zeta_i,$$

and $\zeta_i \sim N(0,1)$. Thus I can now use the standard Bayesian linear regression algebra. Given the assumption of normal prior, I have

$$\begin{pmatrix} \beta & \gamma \end{pmatrix}' \sim N\left(\begin{pmatrix} \tilde{\beta} & \tilde{\gamma} \end{pmatrix}', \left(\tilde{X}'\tilde{X} + A_{\beta,\gamma}\right)^{-1}\right)$$

where

$$
\begin{aligned}
\begin{pmatrix} \tilde{\beta} & \tilde{\gamma} \end{pmatrix}' &= \left(\tilde{X}'\tilde{X} + A_{\beta,\gamma}\right)^{-1} \left[\tilde{X}'\tilde{y}^* + A_{\beta,\gamma} \begin{pmatrix} \bar{\beta} & \bar{\gamma} \end{pmatrix}'\right] \\
\tilde{X} &= \begin{bmatrix} \frac{AA}{\tau} & \frac{x}{\tau} \end{bmatrix} \\
\tilde{y}^* &= \frac{y^* - \frac{\sigma_{12}}{\sigma_2^2}\varepsilon_2}{\tau}.
\end{aligned}
$$

A.3 Step 3. Updating δ.

I rewrite the system of two equations in such a way that they would have the same parameters on the RHS, namely, δ. Nothing has to be done with the equation for AA, but the equation for y^* is a bit more tricky. I start by substituting the instrument equation into the main:

$$
\begin{aligned}
y_i^* &= \left[z_i'\delta + \varepsilon_{i,2}\right]\beta + x_i'\gamma + \varepsilon_{i,1} \\
&= \beta z_i'\delta + x_i'\gamma + \beta\varepsilon_{i,2} + \varepsilon_{i,1}.
\end{aligned}
$$

Now I transform the last equation as follows:

$$
\begin{aligned}
y_i^* - x_i'\gamma &= \beta z_i'\delta + \beta\varepsilon_{i,2} + \varepsilon_{i,1} \\
\frac{y_i^* - x_i'\gamma}{\beta} &= z_i'\delta + \left(\frac{1}{\beta}\varepsilon_{i,1} + \varepsilon_{i,2}\right),
\end{aligned}
$$

and denote $\widehat{y}^* \equiv (y_i^* - x_i'\gamma)/\beta$. The system is now of the form

$$\begin{cases} \widehat{y}^* = z_i'\delta + u_{1i} \\ x_i = z_i'\delta + u_{2i} \end{cases},$$

where

$$Var\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \equiv \mathbf{\Omega} = \begin{pmatrix} \frac{1}{\beta} & 1 \\ 0 & 1 \end{pmatrix} Var\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \begin{pmatrix} \frac{1}{\beta} & 1 \\ 0 & 1 \end{pmatrix}'.$$

Given that $\mathbf{\Omega}$ is symmetric and positive definite, its Cholesky root $\mathbf{F}$ exists and is unique, so:

$$\mathbf{\Omega} = \mathbf{F}\mathbf{F}'.$$

Therefore, I take the transformed system, premultiply it by $\mathbf{F}$, stack observations and obtain the standard normal Bayesian linear regression with unit variance:

$$\left(\mathbf{F}^{-1}\right)'\begin{pmatrix} \hat{y}^* \\ x \end{pmatrix} = \left(\mathbf{F}^{-1}\right)'\begin{pmatrix} z' \\ z' \end{pmatrix}\delta + \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix},$$

where $\psi = (\psi_1, \psi_2) \sim N(0, I)$. Given the assumption of normal prior:

$$\delta \sim N\left[\tilde{\delta}, \left(\tilde{z}'\tilde{z} + A_\delta\right)^{-1}\right],$$

where

$$
\begin{aligned}
\tilde{\delta} &= \left(\tilde{z}'\tilde{z} + A_\delta\right)^{-1}\left[\tilde{z}'\overleftrightarrow{y} + A_\delta\bar{\delta}\right], \\
\tilde{z} &= \begin{bmatrix} z' & z' \end{bmatrix}' \\
\overleftrightarrow{y} &= \begin{pmatrix} \frac{y_i^* - x_i'\gamma}{\beta} & x \end{pmatrix}'.
\end{aligned}
$$

A.4 Step 4. Updating Σ.

I first obtain the residuals for each equation:

$$
\begin{cases} e_{i,1} = y_i^* - AA_i\beta - w_i'\gamma \\ e_{i,2} = AA_i - z_i'\delta \end{cases},
$$

and then compute

$$
S = \sum_{i=1}^{n} \begin{pmatrix} e_{i,1} \\ e_{i,2} \end{pmatrix} \begin{pmatrix} e_{i,1} & e_{i,2} \end{pmatrix}.
$$

Then the properties of the inverse Wishart distribution guarantee that the posterior for Σ will be

$$
\Sigma \sim IW\left(\upsilon_0 + n, V_0 + S\right).
$$

Appendix B

Sampling from the Truncated Normal Distribution

Let $\phi(\cdot)$ be a standard normal pdf, and $\Phi(\cdot)$ be the standard normal cdf, then $\Phi^{-1}(\cdot)$ is the inverse of Φ. Let $U_{[a,b]}$ denote the uniform distribution on $[a,b]$: $a < b$. Denote $TN_{[a,b]}(\mu, \sigma^2)$ to be the normal distribution with parameters μ and σ^2, truncated to the interval $[a,b]$. Its density at x is

$$\left[\Phi^{-1}(b) - \Phi^{-1}(a)\right]^{-1} \phi(x) \, 1\{x \in [a,b]\}.$$

It is immediately apparent that when $x \sim TN_{[a,b]}(\mu, \sigma^2)$ then

$$z \equiv \frac{x - \mu}{\sigma} \sim TN_{\left[\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}\right]}(0, 1),$$

and hence I need to be able to draw from a truncated standard normal density only (which will be denoted $TN_{[a,b]}$ for brevity). There are several ways of doing that, but my interest is in sampling in an efficient manner.

B.1 Inverse CDF Sampling

Let $x \sim TN_{[a,b]}$. Then $x = \Phi^{-1}(u)$ where $u \sim U_{[\Phi(a),\Phi(b)]}$. This method works fine when a is far enough from $-\infty$ and b is far from ∞. Otherwise numerical issues arise: if $|w| > 8$, then $w = \Phi^{-1}(p)$ usually cannot be solved numerically precisely enough.

B.2 Mixed Rejection Sampling

This algorithm, which was developed by Geweke (1991), uses a variant of importance sampling. Two key concepts are the *target* density (the one I want to draw from), and the *proposal* (also sometimes called instrumental) density (the one I use to simplify the drawing process).

Denote the target density by $f(x)$ and the proposal density by $g(x)$. Assume that $\exists M : \forall x \quad f(x) \leq Mg(x)$. Then the importance sampling scheme works as follows:

- draw $x \sim g$ and $u \sim U_{[0,1]}$,
- accept $y = x$ if $u \leq \frac{f(x)}{Mg(x)}$,

- else redraw x and u, repeat.

The mixed rejection algorithm alternates between two different sampling schemes:

1. Normal Rejection Sampling — draw x from $N(0,1)$ and accept it if $x \in [a,b]$, redraw otherwise. In other words, target density is the truncated normal density, proposal density is the untruncated standard normal density. This works great if only a small portion of probability mass at the tails actually gets truncated.

2. Exponential Rejection Sampling — covered in Section B.2.1 below.

The choice between sampling methods depends on the values of a and b. As long as I consider only the cases of left truncation and right truncation (i.e. either $a = -\infty$ or $b = \infty$), the cut-off values are as follows:

- if draws are needed from $TN_{[-\infty,b]}$, sample via normal rejection if $b \geq -0.45$, and via exponential rejection otherwise;

- if draws are needed from $TN_{[a,\infty]}$, sample via normal rejection if $a \leq 0.45$, and via exponential rejection otherwise.

It remains to discuss how the exponential rejection sampling must be administered.

B.2.1 Exponential Rejection Sampling

The motivating example is $TN_{[a,\infty]}$, where $\Phi(a)$ is close to 1. As $a \to \infty$, $TN_{[a,\infty]}$ converges to the exponential distribution on $[a,\infty)$ with kernel $\exp(-\lambda z)$ for $z \geq a$. Thus, the target and proposal densities are

$$
\begin{aligned}
f(x) &= c \exp\left(-\frac{1}{2}x^2\right) 1(x > a), \\
g(x) &= \lambda \exp\left[-\lambda(x-a)\right] 1(x > a),
\end{aligned}
$$

where c is the normalizing constant so that $f(x)$ would integrate to 1. The ratio is

$$
\begin{aligned}
\frac{f(x)}{g(x)} &= \frac{c}{\lambda} 1(x > a) \exp\left[-\frac{1}{2}x^2 + \lambda(x-a)\right] \\
&\leq \frac{c}{\lambda} 1[x > a] \exp\left\{\max_{x \geq a}\left[-\frac{1}{2}x^2 + \lambda(x-a)\right]\right\} \\
&= \frac{c}{\lambda} \exp\left(\frac{1}{2}\lambda^2 - \lambda a\right) 1(\lambda > a) + \frac{c}{\lambda} \exp\left(-\frac{1}{2}a^2\right) 1[\lambda \leq a] \\
&= M_1(\lambda) 1(\lambda > a) + M_2(\lambda) 1(\lambda \leq a).
\end{aligned}
$$

Minimizing this in λ yields the smallest probability of rejection. Geweke (1991) notes, however, that from the computational standpoint it is better to minimize only the second term, which yields $\lambda = a$.

The exponential rejection then proceeds in drawing $x \sim g$ and $u \sim U_{[0,1]}$, and accepting x as long as

$$u \leq \frac{f(x)}{g(x \mid \lambda = a) M_2(a)} = \exp\left\{-\frac{1}{2}\left(x^2 + a^2\right) + ax\right\},$$

and rejecting the x otherwise.

B.3 Algorithm

This algorithm is presented for the case of drawing a scalar random variable. It is in principle vectorizable. Consider first the case of left-truncation, i.e. $TN_{[a,\infty]}$:

1. Compare a with 0.45. Suppose first that $a \leq 0.45$:
   (a) Draw $z \sim N(0,1)$
   (b) If $z > a$, done, else return to step 1a.
2. Suppose now that $a > 0.45$:
   (a) Draw z from $g(z) = a \exp\{-a(z-a)\}$, say, via inverse c.d.f. transformation:
      i. draw $w \sim U_{[0,1]}$;
      ii. use the fact that the cdf that corresponds to $g(z)$ is $G(z) = 1 -$

$\exp\left[-a\left(z-a\right)\right]$ to do the inversion:

$$\begin{aligned} w &= 1-\exp\left[-a\left(z-a\right)\right] \\ z &= a-\frac{1}{a}\log\left(1-w\right) \end{aligned}$$

One can replace w with $(1-w)$ in the last line, since if $w \sim U_{[0,1]}$, then so is $(1-w)$.

(b) Draw $u \sim U_{[0,1]}$;

(c) If $u < \exp\left[-\frac{1}{2}\left(z^2-a^2\right)+az\right]$, done, else return to Step 2a.

Now consider the case of right-truncation, i.e. $TN_{[-\infty,b]}$. Use the fact that if $z_1 \sim TN_{[-b,\infty]}$, then $z_2 = -z_1 \sim TN_{[-\infty,b]}$. Thus, proceed as follows:

1. Sample z_1 from $TN_{[-b,\infty]}$ as detailed above

2. Set $z_2 = -z_1$.

Appendix C

Supplementary Evidence

I present the summary of the first stage regressions for the two-stage least squares procedure discussed in Section 1.6.1. Table C.1 below documents the results. The “distance to nearest branch” variable has an expected negative sign, indicating that when the loan collateral is far away from the bank branch, there is a lower chance this loan will fall into the CRA assessment area. The high values of the first stage F-statistics suggest that there is no weak instruments issue in my application. Staiger and Stock (1997) claim that F-statistics below 10 are usually associated with weak instruments; the smallest one that I obtain is 713.869.

I next demonstrate the evidence of convergence of my MCMC procedure discussed in Section 1.6.2. For brevity I only present the results for the CRA effect β that is of primary interest. The top panel plots the first 2000 simulated draws from the posterior distribution. It is clear that the chain takes relatively few iterations to

Table C.1: Two Stage Least Squares: First Stage Results

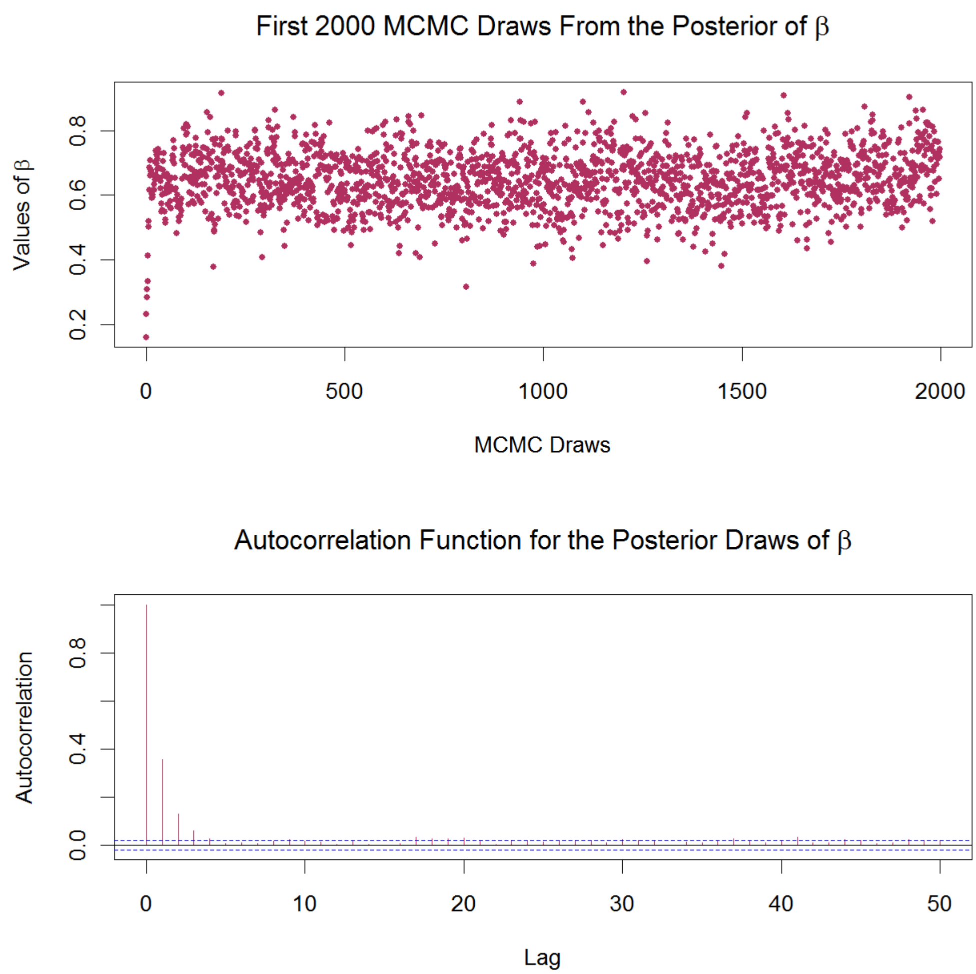
	Column 1	Column 2	Column 3	Column 4	Column 5
	All Observations	All	Lower-Income Tracts Closest 5	Lower-Income Tracts Closest 2	Closest 2
Distance to Nearest Branch	-0.0003*	-0.002*	-0.002*	-0.002*	-0.002*
	(0.000004)	(0.00003)	(0.00004)	(0.00009)	(0.00004)
Loan Size, \$100k	0.00004*	-0.061*	-0.040*	-0.018*	-0.005
	(0.000008)	(0.002)	(0.002)	(0.003)	(0.003)
Annual Income, \$100k	-0.008†	0.023*	0.003	0.008*	0.0002
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Applicant Female	-0.076*	-0.002	-0.004	-0.003	-0.008
	(0.002)	(0.005)	(0.006)	(0.010)	(0.007)
Applicant Not White	0.044*	-0.083*	-0.034*	-0.042*	0.018†
	(0.002)	(0.005)	(0.006)	(0.010)	(0.007)
Applicant Hispanic	0.023*	0.009	0.042*	-0.005	-0.016 ⋄
	(0.002)	(0.004)	(0.005)	(0.009)	(0.007)
Has a Co-Applicant	-0.002*	0.053*	0.033*	0.040*	0.021*
	(0.00001)	(0.005)	(0.006)	(0.009)	(0.006)
Constant	0.75*	0.820*	0.924*	0.909*	0.861*
	(0.003)	(0.007)	(0.007)	(0.011)	(0.010)
Tract Fixed Effects	No	No	No	No	Yes
First stage F-statistics	21089.1	3522.03	2332.13	713.87	2383.70
Number of observations	169,859	44,546	20,867	6,533	6,533

Dependent variable: assessment area indicator. Data from the 2005 HMDA, CRA and Summary of Deposits.
First stage F-statistics test for weak instruments, as suggested by Staiger and Stock (1997).
Standard errors in parentheses. ⋄ $p < 0.05$, † $p < 0.01$, * $p < 0.001$.

get to its stationary distribution and then just stays there. The second panel presents the autocorrelation function for the "thinned" set of β draws. It demonstrates that there is virtually no leftover serial dependence in the chain.

Taken together, these two plots suggest that the MCMC procedure had been successful in arriving at the stationary distribution for the underlying parameters. The corresponding plots for the other parameters of the model looked extremely similar to the ones presented and, hence, are not shown.

Figure C.1: MCMC Convergence for β



The top panel presents the first 2000 draws from the posterior for β. The bottom panel contains the ACF for the "thinned" draws.

Next, I provide the tables with all the information on differences in loan-level observables across lenders and samples. Table 1.3 on page 47 is a condensed version of the four tables that follow. And finally, Table C.6 from CRA Reference

(2005) provides the other criteria that are relevant for the CRA enforcement.

Table C.2: Decomposition of Average Loan Characterstics By Lender, All Observations

	Loan Amount, $1000	Annual Income, $1000	Applicant Female	Applicant Not White	Applicant Hispanic	Has A Co-Applicant
			Inside the Assessment Areas			
Bank Of America	336.04	124.48	0.311	0.306	0.269	0.549
Downey Savings & Loan	338.02	109.24	0.517	0.509	0.554	0.371
First Bank	302.64	97.337	0.332	0.252	0.544	0.322
First Federal Bank of CA	450.57	150.05	0.348	0.306	0.415	0.373
Fremont Investment & Loan	287.07	96.614	0.368	0.335	0.556	0.340
Guaranty Bank	329.49	111.13	0.288	0.259	0.260	0.456
Provident Savings Bank	267.11	93.535	0.286	0.181	0.414	0.390
US Bank	335.72	111.57	0.391	0.414	0.357	0.609
Union Bank of CA	524.84	203.35	0.346	0.309	0.257	0.601
Wachovia	319.61	105.63	0.372	0.368	0.460	0.440
Washington Mutual Bank	428.26	139.20	0.385	0.351	0.345	0.500
Wells Fargo Bank	341.28	118.10	0.291	0.310	0.268	0.552
			Outside the Assessment Areas			
Bank Of America	480.08	168.33	0.336	0.427	0.224	0.535
Downey Savings & Loan	395.55	125.71	0.462	0.509	0.504	0.365
First Bank	285.52	90.415	0.384	0.293	0.608	0.307
First Federal Bank of CA	399.79	128.38	0.375	0.472	0.368	0.385
Fremont Investment & Loan	345.32	111.74	0.379	0.390	0.498	0.321
Guaranty Bank	402.39	131.03	0.332	0.331	0.272	0.354
Provident Savings Bank	336.84	109.64	0.317	0.271	0.460	0.327
US Bank	422.86	167.84	0.436	0.436	0.419	0.620
Union Bank of CA	736.44	301.96	0.355	0.369	0.282	0.531
Wachovia	364.14	118.79	0.397	0.432	0.432	0.426
Washington Mutual Bank	535.39	167.01	0.397	0.436	0.311	0.470
Wells Fargo Bank	473.07	151.55	0.293	0.414	0.183	0.555

Data from the 2005 HMDA and CRA.
Every cell contains the average value of the corresponding loan-level observable for a given bank and treatment status.

Table C.3: Decomposition of Average Loan Characterstics By Lender, Lower-Income Tracts

	Loan Amount, $1000	Annual Income, $1000	Applicant Female	Applicant Not White	Applicant Hispanic	Has A Co-Applicant
			Inside the Assessment Areas			
Bank Of America	248.57	88.118	0.323	0.316	0.372	0.482
Downey Savings & Loan	291.89	94.911	0.540	0.494	0.596	0.317
First Bank	271.59	85.779	0.357	0.251	0.635	0.293
First Federal Bank of CA	359.43	120.01	0.419	0.337	0.570	0.302
Fremont Investment & Loan	249.48	85.475	0.362	0.297	0.637	0.311
Guaranty Bank	284.81	97.887	0.294	0.244	0.344	0.369
Provident Savings Bank	225.53	81.52	0.302	0.141	0.470	0.347
US Bank	239.8	91.81	0.337	0.368	0.337	0.537
Union Bank of CA	375.31	135.79	0.417	0.281	0.251	0.506
Wachovia	268.59	92.405	0.384	0.353	0.522	0.396
Washington Mutual Bank	322.83	103.84	0.421	0.340	0.409	0.426
Wells Fargo Bank	273.04	91.314	0.312	0.341	0.336	0.492
			Outside the Assessment Areas			
Bank Of America	315.54	103.91	0.343	0.478	0.339	0.467
Downey Savings & Loan	347.1	110.32	0.455	0.526	0.557	0.305
First Bank	242.39	77.19	0.372	0.256	0.687	0.238
First Federal Bank of CA	325.59	102.86	0.412	0.456	0.458	0.287
Fremont Investment & Loan	308.62	97.612	0.382	0.391	0.561	0.271
Guaranty Bank	316.97	100.25	0.358	0.358	0.352	0.264
Provident Savings Bank	284.83	90.558	0.305	0.244	0.599	0.265
US Bank	394.12	108.74	0.346	0.385	0.423	0.538
Union Bank of CA	500.21	169.19	0.303	0.349	0.330	0.495
Wachovia	305.62	100.97	0.418	0.464	0.513	0.382
Washington Mutual Bank	355.01	103.32	0.426	0.424	0.429	0.374
Wells Fargo Bank	345.02	105.08	0.363	0.466	0.236	0.433

Data from the 2005 HMDA and CRA.
Every cell contains the average value of the corresponding loan-level observable for a given bank and treatment status.

Table C.4: Decomposition of Average Loan Characterstics By Lender, "Closest-Five Tracts" Sample

	Loan Amount, $1000	Annual Income, $1000	Applicant Female	Applicant Not White	Applicant Hispanic	Has A Co-Applicant
			Inside the Assessment Areas			
Bank Of America	242.35	84.443	0.318	0.324	0.382	0.480
Downey Savings & Loan	285.6	94.79	0.524	0.496	0.609	0.336
First Bank	269.52	82.564	0.35	0.242	0.646	0.260
First Federal Bank of CA	345.96	116.65	0.367	0.388	0.531	0.327
Fremont Investment & Loan	240.04	81.898	0.365	0.296	0.650	0.306
Guaranty Bank	250.89	82.99	0.306	0.255	0.367	0.327
Provident Savings Bank	205.1	73.034	0.328	0.147	0.539	0.333
US Bank	230.21	67.235	0.293	0.328	0.276	0.534
Union Bank of CA	319.76	102.35	0.461	0.343	0.343	0.480
Wachovia	261.56	92.983	0.406	0.354	0.537	0.384
Washington Mutual Bank	313.2	98.817	0.419	0.363	0.430	0.431
Wells Fargo Bank	264.55	88.165	0.302	0.333	0.350	0.491
			Outside the Assessment Areas			
Bank Of America	333.06	113.80	0.361	0.465	0.229	0.444
Downey Savings & Loan	325.22	105.54	0.451	0.488	0.591	0.319
First Bank	213.92	69.906	0.381	0.220	0.732	0.261
First Federal Bank of CA	309.89	100.36	0.389	0.418	0.449	0.291
Fremont Investment & Loan	280.43	94.863	0.330	0.345	0.483	0.330
Guaranty Bank	283.57	89.63	0.311	0.338	0.392	0.311
Provident Savings Bank	282.28	91.506	0.272	0.213	0.494	0.322
US Bank	338.54	86.077	0.154	0.231	0.462	0.615
Union Bank of CA	496.54	163.79	0.375	0.333	0.458	0.417
Wachovia	284.35	102.53	0.404	0.418	0.503	0.450
Washington Mutual Bank	364.33	110.57	0.449	0.436	0.329	0.394
Wells Fargo Bank	347.25	109.55	0.350	0.398	0.188	0.416

Data from the 2005 HMDA and CRA.
Every cell contains the average value of the corresponding loan-level observable for a given bank and treatment status.

Table C.5: Decomposition of Average Loan Characterstics By Lender, "Closest-Two Tracts" Sample

	Loan Amount, $1000	Annual Income, $1000	Applicant Female	Applicant Not White	Applicant Hispanic	Has A Co-Applicant
			Inside the Assessment Areas			
Bank Of America	240.95	86.402	0.308	0.329	0.406	0.481
Downey Savings & Loan	282.87	92.588	0.529	0.484	0.588	0.363
First Bank	249.77	79.106	0.409	0.167	0.591	0.288
First Federal Bank of CA	334.26	111.20	0.316	0.368	0.632	0.368
Fremont Investment & Loan	236.39	87.929	0.359	0.284	0.659	0.317
Guaranty Bank	264.37	87.50	0.200	0.233	0.300	0.533
Provident Savings Bank	185.43	70.833	0.315	0.167	0.407	0.352
US Bank	226.88	91.905	0.417	0.333	0.250	0.417
Union Bank of CA	409.10	152.78	0.425	0.250	0.250	0.500
Wachovia	264.95	89.25	0.393	0.320	0.556	0.406
Washington Mutual Bank	327.86	101.22	0.464	0.372	0.414	0.432
Wells Fargo Bank	277.96	93.97	0.298	0.305	0.314	0.471
			Outside the Assessment Areas			
Bank Of America	327.22	112.43	0.311	0.500	0.233	0.411
Downey Savings & Loan	305.69	100.43	0.480	0.547	0.581	0.338
First Bank	213.57	68.53	0.317	0.149	0.772	0.208
First Federal Bank of CA	276.32	96.37	0.375	0.409	0.511	0.250
Fremont Investment & Loan	254.32	85.144	0.309	0.325	0.491	0.359
Guaranty Bank	273.31	86.531	0.406	0.375	0.500	0.219
Provident Savings Bank	294.98	97.018	0.264	0.209	0.464	0.318
US Bank	277.50	91.75	0.250	0.500	1.000	0.500
Union Bank of CA	425.33	134.33	0.333	0.667	1.000	0.000
Wachovia	258.06	93.917	0.400	0.385	0.561	0.410
Washington Mutual Bank	319.10	95.681	0.530	0.512	0.339	0.333
Wells Fargo Bank	310.23	98.686	0.369	0.449	0.222	0.438

Data from the 2005 HMDA and CRA.
Every cell contains the average value of the corresponding loan-level observable for a given bank and treatment status.

Table C.6: Lending Performance Ratings

Characteristic	Performance Ratings				
	Outstanding	High Satisfactory	Low Satisfactory	Needs to Improve	Substantial Noncompliance
Lending Activity	Levels reflect EXCELLENT responsiveness to AA credit needs	GOOD	ADEQUATE	POOR	VERY POOR
Assessment Area(s) Concentration	SUBSTANTIAL MAJORITY of loans are made in the bank's AA	HIGH percentage	ADEQUATE percentage	SMALL percentage	VERY SMALL percentage
Geographic Distribution of Loans	Reflects EXCELLENT penetration throughout the AA	GOOD	ADEQUATE	POOR	VERY POOR
Borrower's Profile	Distribution of borrowers reflects EXCELLENT penetration among customers of different income levels	GOOD	ADEQUATE	POOR	VERY POOR
Responsiveness to Credit Needs of Low-Income Individuals and Areas	Exhibits an EXCELLENT record	GOOD	ADEQUATE	POOR	VERY POOR
Community Development (CD) Lending Activities	A LEADER in making CD loans	Makes A RELATIVELY HIGH level	Makes AN ADEQUATE level	Makes A LOW level	Makes FEW, if ANY
Product Innovation	Makes EXTENSIVE USE of innovative and/or flexible lending practices	USE	LIMITED USE	LITTLE USE	NO USE

Adapted from the CRA Reference (2005)