Credit Supply Shocks in the US Housing Market

by Gianni La Cava

A thesis submitted in February 2013 in fulfilment of the

requirements for the degree of Doctor of Philosophy in Economics

at the London School of Economics and Political Science

Declaration

I certify that this thesis presented by me in February 2013 for examination for the PhD degree of the London School of Economics and Political Science is solely my own work.

Signature ....

Date 10/2/13 ....

Abstract

The aim of my thesis is to study the impact of three different types of credit supply shocks on the US economy during the recent boom-bust cycle. I apply rigourous identification strategies using a comprehensive mortgage loan-application level dataset that spans two decades to identify the causal effect of each shock on the US housing market – the market at the epicentre of the recent global financial crisis.

The first chapter examines how shocks to the geographic distance between borrowers and lenders affected the quality of loans originated by US banks. I show that US banks that originated loans at relatively long distances recorded relatively high shares of delinquent mortgages, which suggests that geographic distance hampers a bank's ability to screen borrowers.

The second chapter assesses how the supply of mortgage credit was affected by the closure of the private-label securitisation market in 2007. By impairing lender financing conditions, I demonstrate that the shock to the securitisation market caused US mortgage lenders to curtail new lending and that this adversely affected the aggregate level of new mortgage credit.

The third chapter explores whether shocks to bank lending standards affect household borrowing and spending. I show that the introduction of US state anti-predatory lending laws, by tightening lending standards, caused a decline in the level of household mortgage credit, although this had little subsequent impact on household spending.

Contents

List of Figures

List of Tables

Acknowledgements

Thank you to my PhD supervisor, Silvana Tenreyro, for her valuable feedback and constant encouragement. I am extremely grateful for her unwavering confidence in me.

I also thank my supervisor in the Centre for Economic Performance, Luis Garicano. I really valued his energy, enthusiasm and clarity of thought. I will also never forget the day that he cut our research meeting short in order to go and meet the Queen of England!

I would like to thank the following people for helpful discussions (in no particular order except alphabetical): Andrew Ainsworth, Johannes Boehm, Francisco Costa, Mardi Dungey, Luci Ellis, Angus Foulis, James Hansen, Ana McDowall, Daniel Osorio, Michele Piffer, Vanessa Rayner, Tom Rosewall, Kevin Sheedy, Grant Turner, Christoph Ungerer, and James Vickery. I also thank my family for their encouragement and support. It's nice to be back in Australia and spending time again with their non-Skype selves.

I am indebted to both the Reserve Bank of Australia and LSE Economics Department for their generous financial support. I take sole responsibility for any errors in this research.

Finally, my biggest thank you is reserved for Michelle, for all her editing, research advice, encouragement, support and, well, willingness to listen to me go on about how much I love difference-in-differences models.

This thesis is dedicated to the rest of my life with Michelle.

Chapter 1

Introduction

Economists have long recognised that the banking sector can generate, and propagate, economic shocks. Theoretical research has demonstrated that, when financial markets are imperfect, shocks to credit supply can affect the real economy. In practice, the recent financial crisis has also highlighted the importance of credit market shocks. Despite this, empirical research has generally lagged behind in adequately identifying the economic effect of these shocks.

Empirical studies face several challenges in tracing the channels through which credit supply shocks are transmitted to the real economy. Most importantly, there are significant endogeneity problems. For example, when economic conditions deteriorate, the balance sheets of both lenders and borrowers typically become impaired, which can reduce both the supply of and demand for credit. This makes it difficult to separate the effect of changes in credit supply from that of demand, and to determine the direction of causality between credit and economic activity.

The aim of my research is to use rigorous empirical methods to identify the causal effect of various credit supply shocks in the context of the US housing market during the recent boom-bust cycle. Despite a vast (and growing) literature, I provide new insights into the causes and effects of the recent US housing market crisis. The US housing credit market provides a natural testing ground to examine the impact of credit supply shocks, being the market at the epicentre of the global financial crisis. Most of the existing research on the crisis looks at the effect of changes in credit supply on the investment behaviour of large firms. But I believe household lending is more likely to explain the prolonged impact of the crisis on the US economy.

My three chapters address the following research questions:

1. Did a rise in geographic distance between borrowers and lenders, by hampering banks' ability to screen borrowers, lead to a decline in the quality of loans originated by US banks?

2. Did the closure of the private-label securitisation market in 2007, by impairing lender financing conditions, reduce the supply of new mortgage credit?

3. Did state anti-predatory lending laws, by tightening bank lending standards, reduce household borrowing and spending?

In addition to their common focus on credit supply shocks, the chapters share similar identification strategies to establish causality, based either on natural experiments or matched borrower-lender relationships. Each chapter makes use of a dataset that covers around half a billion mortgage loan applications spanning more than twenty years that I compiled from the Home Mortgage Disclosure Act (HMDA) loan application registry.

Chapter 2 documents some new stylised facts about the US housing market, most notably a dramatic increase in the geographic distance between lenders and borrowers (or "lending distance") over the past few decades, as well as a 'boom-bust' pattern in lending distance during the 2000s. To the extent that distance is related to the quality of information that banks have about their borrowers, I provide evidence that the increase in lending distance over the 2000s was associated with worse lending decisions and subsequently contributed to higher mortgage delinquencies.

Chapter 3 examines the impact of a shock to lender financing conditions, in particular, the loan securitisation market, on US new mortgage credit. Despite extensive evidence suggesting that a significant tightening in credit conditions occurred in the US in 2007-08, there has been surprisingly little formal testing for such a credit crunch in the context of the US housing market. I use the differential exposure of individual mortgage lenders to the collapse of the securitisation market in 2007 as a source of cross-lender variation in financing conditions and assess the impact on residential mortgage lending. Using disaggregated loan-level information to control for unobservable credit demand shocks, I show that mortgage lenders that were particularly reliant on loan securitisation reduced credit supply by more during the crisis.

Chapter 4 exploits a natural experiment on the US housing market to identify the causal effect of bank lending standards on the real economy. Prior to the subprime mortgage crisis, different US states enacted anti-predatory lending laws at different times. These lending laws were a government-enforced tightening of credit standards for subprime mortgages. I identify the causal effect of these shocks to residential mortgage credit standards by comparing the economic performance of contiguous counties across state borders where one state introduced the law ("the treatment group") and the other state did not ("the control group"). In a difference-in-differences modelling framework, I find that the introduction of these laws was associated with tighter credit conditions and less household borrowing. However, I find little evidence that the laws affected household spending.

Chapter 2

Lending Distance and Home Mortgage Delinquencies

The aim of this chapter is to identify a causal effect of geographic distance on credit quality in the US home mortgage market. I document a dramatic rise in the average distance between lenders and borrowers (or 'lending distance') during the US housing boom and a sharp decline in the subsequent bust. To the extent that distance is a proxy for the quality of information that banks have about their borrowers, I provide evidence that the increase in lending distance during the boom contributed to poorer lending decisions by banks, culminating in higher mortgage delinquencies in the bust.

My results suggest that the overall effect of lending distance on mortgage delinquencies reflects both the geographic distance between a bank's headquarters and its branches and the distance between a bank's branches and its borrowers. The former suggests that internal agency problems within banks contributed to the rise in mortgage delinquencies during the crisis. Despite advances in bank loan screening and monitoring technology, I also find that the effect of distance on credit quality has not fallen over time. If anything, the effect of distance was stronger during the global financial crisis than at any other time during the past two decades.

2.1 Introduction

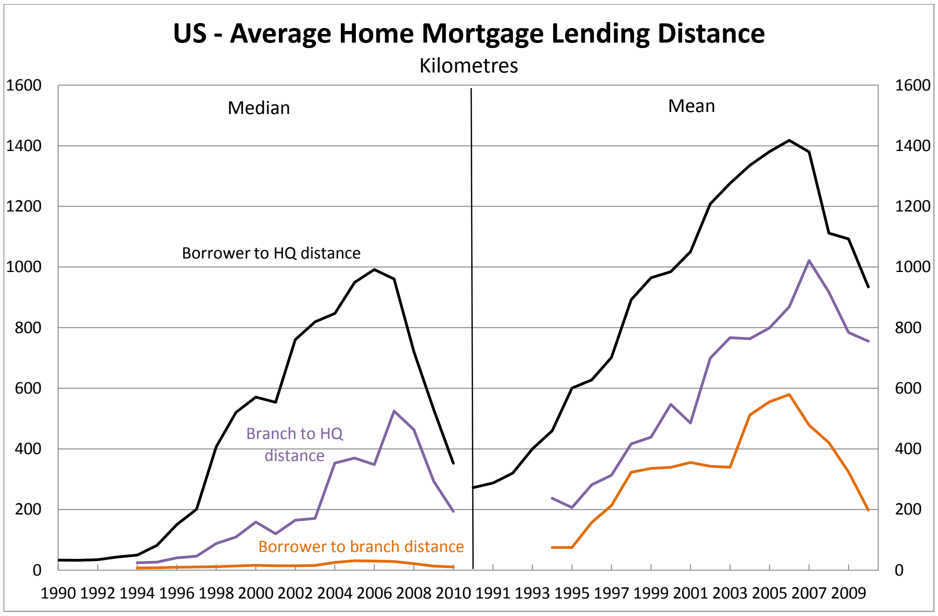
There is substantial evidence that excessive bank risk-taking contributed to the boom and bust in the US home mortgage market in the 2000s (e.g. Mian and Sufi [2009], Keys et al. [2010], Purnanandam [2011], Loutskina and Strahan [2011], Anderson et al. [2011], Dell'Ariccia et al. [2012]). One aspect of US bank behaviour during the boom that was potentially related to risk-taking, but has been largely overlooked in the literature, was the growing physical distance between the location of home mortgage borrowers and the headquarters (HQ) of their lenders (or 'lending distance'). The rise in lending distance in the US home mortgage market in the lead-up to the bust was staggering (Figure 2.1). The median lending distance rose from around 30 kilometres in 1990 to almost 1000 kilometres at the height of the housing credit boom in 2006 before falling back to around 350 kilometres in 2010 (the LHS panel of Figure 2.1). Decomposing the overall borrower-to-bank HQ distance into two components – the borrower-to-branch distance and the branch-to-bank HQ distance – the overall rise in geographic distance primarily reflected an increase in the distance between bank headquarters and their branches, with a smaller increase in the distance between branches and borrowers (Figure 2.1).

The remarkable increase in overall lending distance in the US over recent decades reflected the greater ability and willingness of financial institutions to expand geographically. This, in turn, was driven by several factors.[1] First, advances in loan production technology (e.g. credit bureaus, automated underwriting software) significantly lowered the costs associated with geographic distance, contributing to the dramatic increase in borrower-lender distance on the supply-side of the credit market. Second, the growing acceptance of internet banking, on the demand-side of the credit market, contributed to higher lending distance as borrowers were increasingly able to make online loan applications from anywhere in the country.[2]

[1] See Brevoort and Wolken [2008] for a more detailed discussion of the potential causes of rising lending distance in the US banking industry.

[2] See Amel et al. [2008] and DeYoung et al. [2007] for a more detailed discussion of the effect of internet banking on US bank location and performance.

Figure 2.1: Average Borrower-Lender Distance



Third, the consolidation of the banking industry through mergers and acquisitions led to a significant decline in the number of US banks, making it more difficult for a household to find a nearby bank, contributing to higher borrower-bank HQ distance. Fourth, US mortgage lenders increasingly relied on third-party originators – such as mortgage brokers – to originate new loans. The greater supply of mortgage brokers reduced the cost of bank geographic expansion and also led to higher borrower-lender distance. Finally, the removal of government restrictions on interstate bank branching in the mid-1990s allowed banks to expand geographically, which directly contributed to rising distance between borrowers and the headquarters of the lending institution. The geographic expansion of US banks brought many benefits, including increased efficiency and improved risk diversification, but may also have had an unintended cost in encouraging more risk-taking, leading to less financial stability (Allen and Gale [2003]).

A key feature of Figure 2.1 is the 'boom-bust' cycle in lending distance, with the peak in distance coinciding with the peak in activity in the subprime mortgage market in 2006. At least on the surface, the observed pro-cyclical pattern suggests that lending distance is positively related to bank risk-taking. This can be seen more clearly in Figure 2.2. The top left-hand panel is a scatter plot which shows the partial relationship between a bank's (log level) average lending distance and its (log level) average share of delinquent loans. Each dot on the scatterplot represents a specific bank in a particular year.[3] Casual observation suggests that there is a positive relationship between distance and the share of delinquent loans over the 1990-2010 period. The top right-hand panel considers the same data but in the form of a histogram, that is, it plots the average level of delinquent loans *within each percentile* of the distribution of lending distance. Similar to the scatterplot, the histogram points to a positive link between distance and bad loans – the average level of bad loans increases for banks that lend at relatively high distances. Interestingly, there is evidence that the average level of bad loans increases sharply at very high levels of lending distance.[4]

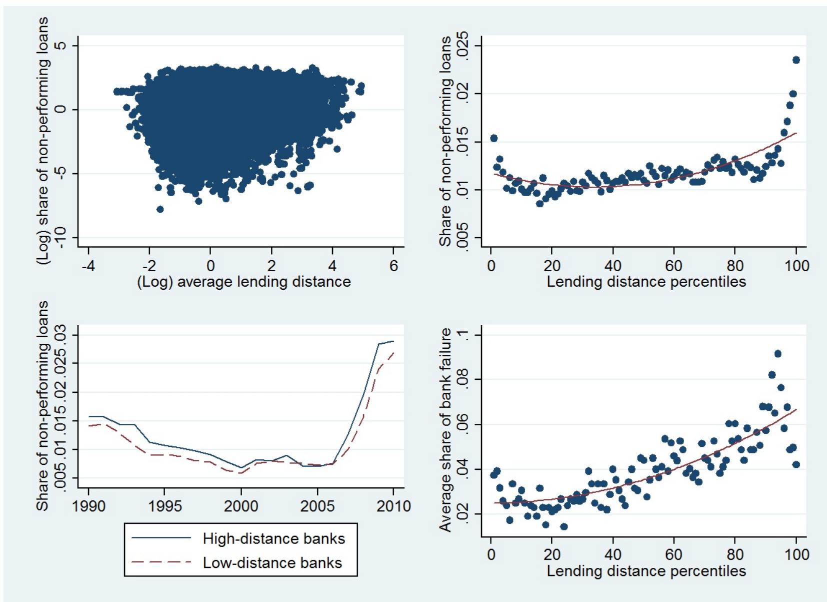
An alternative way to present the data is to separate banks into two groups depending on whether the bank's average lending distance is higher or lower than the median bank in a particular year. The associated time-series graph is shown in the bottom left-hand panel. Clearly, banks that lend to more distant borrowers, on average, consistently report higher shares of delinquent loans than those that lend to close borrowers.

Finally, the bottom right-hand panel graphs the average share of bank failures across different percentiles of the lending distance distribution. Clearly, banks that lent to distant borrowers experienced higher failure rates, on average, than banks that lent to close borrow-

[3]More specifically, the scatterplot shows on the y-axis the residuals from a regression of the (log) share of delinquent loans on the (log of) bank assets as well as bank and year fixed effects, and on the x-axis the residuals from a regression of the (log) average distance of each bank on the (log of) bank assets as well as bank and year fixed effects. In other words, the data in the scatter plot are purged of the effect of bank size and other unobservable bank-specific and time-specific effects.

[4]The mortgage lenders that typically lent at the longest distances, on average, included branches and subsidiaries of national banks such as CitiFinancial, Washington Mutual, IndyMac and Countrywide. These national banks were some of the most significant subprime mortgage lenders during the US housing boom period. This may suggest that the link between the share of bad loans and distance simply reflects each bank's exposure to subprime lending. It will therefore be important to control for the bank's business model in the later regression analysis.

Figure 2.2: Lending Distance and Bank Credit Quality



Notes: the top left panel is a scatterplot showing, for each bank and year, the (log of) the non-performing loan ratio (y-axis) against the (log of) the average lending distance (x-axis). The top right panel shows the average non-performing loan ratio (y-axis) within each percentile of the bank-level lending distance distribution (x-axis). The bottom left panel splits the sample of banks into 'high' and 'low' distance banks depending on whether each bank's average lending distance is higher or lower than the median bank each year. The time-series graph compares the non-performing loan ratios of the two groups across time. The bottom right panel shows the average proportion of banks that fail (y-axis) within each percentile of the bank-level lending distance distribution (x-axis).

ers. So, overall, these data are strongly suggestive of a link between lending distance and credit risk.

The purpose of this paper is to establish a causal relationship between lending distance and risk-taking. The theoretical and empirical literature typically assumes that lending distance is a proxy for the quality of lender information about borrower creditworthiness (Petersen and Rajan 1995, 2002). Specifically, lenders receive less precise signals of the credit risk of distant borrowers than of close borrowers, which makes distant loan applicants more difficult to screen and monitor (Hauswald and Marquez [2006]). This link between lending distance and screening ability reflects several factors. Firstly, proximity improves the ability of, and incentives for, loan officers to collect and analyse "soft" (or non-quantitative) information, which has been shown to reduce risk in home mortgage lending (Agarwal et al. [2011]).[5] Secondly, proximity is also associated with greater lender knowledge of local housing and labour market conditions (Ergungor and Moulton [2011]). Thirdly, households that live or work near a financial institution are more likely to have another type of account (e.g. deposit account) with that institution, providing additional "hard" (or quantitative) information to the lender (Ergungor and Moulton [2011], Puri et al. [2011]).

Generalising this logic, a potential explanation for the observed link between lending distance and credit quality is that greater lending distance causes lenders to produce lower-quality information about borrowers, resulting in worse lending decisions. This 'information frictions hypothesis' suggests that the link between credit quality and lending distance is causal and reflects information flows within banks that deteriorate with distance.

Another explanation is that the nationwide housing boom encouraged banks to expand geographically (leading to a rise in lending distance) *and* to lend to more marginal risky borrowers in order to grow market share (causing a rise in delinquent loans). This 'lending standards hypothesis' suggests that the link between credit quality and lending distance is

[5] Proximity reduces transaction costs and allows for more interpersonal communication between loan officers and borrowers (Petersen and Rajan [2002], Berger et al. [2005], Degryse and Ongena [2005], Agarwal and Hauswald [2010]).

not necessarily causal but reflects a common unobserved factor – a deliberate change in bank lending standards. While the 'information frictions hypothesis' and 'lending standards hypothesis' differ in detail, they both attribute the link between distance and risk-taking to the supply-side of the credit market. I provide evidence consistent with these supply-side explanations, and argue further that the weight of evidence favours the information frictions hypothesis.

Several empirical problems need to be addressed in order to establish a causal link between geographic distance and credit quality. Firstly, there may be selection bias. Lending distance is not necessarily exogenous because a loan origination is an equilibrium outcome of the optimal choices of borrowers and lenders. The borrower-lender 'match' could depend on both borrower and lender characteristics that are correlated with credit risk but unobservable to the econometrician (Tian [2011]). For instance, if borrowers know that their local banks are better informed about their creditworthiness than non-local banks, less creditworthy applicants may seek out more distant banks. In this case a positive link between distance and bad loans could reflect an adverse selection problem.[6] Part of the motivation for this paper is to rule out this "demand-side" explanation for the distance-credit quality relationship. I achieve this by using highly disaggregated loan-level information on lending distance to separate the effect of credit supply shocks from credit demand shocks. Specifically, I estimate a 2-stage model in which I isolate the credit supply component of lending distance in the first stage and then examine how this supply-side measure of distance affects the quality of loans originated by a bank in the second stage, after controlling for a range of observable and unobservable bank characteristics.

Secondly, there may be omitted variables bias. For example, distance may reflect the type of origination channel used by banks. Banks that choose to expand geographically

[6] A similar argument could be made that, because house prices tend to be highest in cities, the least-creditworthy, or most credit-constrained, borrowers are likely to locate furthest away from the cities in order to be able to afford to purchase a home. If banks are mainly located in cities then this spatial sorting may also yield a positive link between distance and the share of bad loans (Hanson and Turnbull [2012]). However, Hanson and Turnbull [2012] find mixed evidence for this spatial sorting hypothesis.

may be more likely to rely on mortgage brokers than local branch offices. In this case, the positive relationship between distance and bad loans may be due to principal-agent problems associated with lending through brokers and not to information problems within banks.[7] I provide evidence that the link between distance and credit quality holds, regardless of the type of origination channel used by banks.

Thirdly, if banks do optimally choose to lend to distant borrowers, they presumably have risk-management practices in place to mitigate the effect of distance on performance. For example, a bank may overcome the impact of distance by setting up a local branch in a distant market or giving more decision-making authority to officers located in branches that are located far away from the bank's headquarters. While I am unable to observe each bank's organisation or incentive structure, I control for these factors to the extent that they are constant over time.

Overall, I provide evidence that the banks that supplied credit to more distant borrowers experienced a relatively high share of delinquent loans, on average. Moreover, I find that the effect of distance on mortgage delinquencies is due to both the distance between a bank's headquarters and its branches and the distance between borrowers and the bank's branches. This suggests that increases in lending distance during the housing boom were associated with greater internal agency problems and that this contributed to poorer-quality information about borrowers being transmitted within banks, which led to bad lending decisions.

My paper makes the following contributions to the existing literature:

- **Information frictions, bank risk-taking and credit market outcomes**: my paper provides a new angle on the relationship between bank agency problems, risk-taking and the credit cycle. To the extent that geographic distance is associated with the quality of information produced by lenders, my paper sheds new light on how microeconomic factors (e.g. the organisational and incentive structures of banks) can have

[7]The documented rise in the distance between bank branches and borrowers (as opposed to the distance between bank headquarters and borrowers) could partly reflect the fact that banks have become increasingly reliant on brokers rather than local branches to originate loans.

implications for the credit cycle and financial stability. My paper adds to the growing list of studies that examine how agency problems (such as those in the securitisation process) reduce incentives for banks to screen (or monitor) risky unqualified borrowers (e.g. Keys et al. [2010], Purnanandam [2011], Loutskina and Strahan [2011]).

- **Lending distance in the housing market**: my paper is the first to document changes in lending distance for loans to households and to empirically identify the effect of distance on home mortgage credit quality. Developments in geographic distance between small business borrowers and bank lenders have been well documented (e.g. Petersen and Rajan [2002], Degryse and Ongena [2002], Brevoort and Hannan [2004]).[8] The links between distance and loan outcomes (e.g. Carling and Lundberg [2002], Qian and Yang [2010]) and distance and lender performance (e.g. DeYoung et al. [2008]) have also been previously studied in the context of small business lending.

The rest of this paper is organised as follows. Section 2 provides a more detailed discussion of how the structure and organisation of banks are related to lending distance. Section 3 discusses the data used to estimate the model. Section 4 describes the identification strategies behind the model and section 5 outlines the main results. Section 6 considers some extensions to the basic model and outlines some robustness tests. Section 7 concludes.

2.2 Institutional Background

In the introduction I discussed how lending distance was potentially related to the quality and type of information that lenders have about borrower creditworthiness. In the discussion there were some implicit assumptions about how banks are structured, how information flows between borrowers and lenders and how information flows between bank managers and loan

[8] Previous studies have examined how small business lending distance has changed over time. The evidence is mixed; some papers find that distance has increased in the US (e.g. Petersen and Rajan [2002]) while others do not (e.g. Brevoort and Hannan [2004]). The differences in results appear to reflect differences in the sample of firms and lenders – the change in technology that could give rise to greater distance appears to only have been adopted by a small group of banks.

officers within a bank. As such, it is helpful to provide a more detailed outline of the mortgage loan origination process, the types of information produced and used in this process, and the key role played by loan officers. The outline of the mortgage application process is based on the discussion by Agarwal et al. [2011].

Traditionally, the home mortgage loan origination process proceeds as follows: the borrower meets with a representative of the lending institution (the loan officer) who presents the borrower with a range of contracts with varying prices and credit terms. This meeting can take place either in person or over the phone or internet. Based on the range of contracts offered, the borrower chooses whether to submit an application or not. If the borrower submits an application, the loan officer conducts an initial screening of the loan application using "hard information" obtained from an automated underwriting system. Hard information is information that is easily observable and verifiable, such as income, debt levels and credit (FICO) scores. The system either accepts the application (in which case the loan is booked), rejects the application, or refers the application for secondary screening.

The marginal applications that are referred for secondary screening are returned to the loan officer who gathers "soft information" from discussions with the applicant. For example, a borrower's degree of commitment to make the mortgage payment or their understanding of their obligations are types of soft information. Unlike hard information, soft information is not readily observable or easily transmitted between agents. Loan officers assess the extent to which the information in a loan application meets the bank's lending criteria (or 'credit standards'). The criteria are determined not by the loan officer but by the central risk manager. Based on the hard information contained in the application and the soft information learned during negotiations, and after possibly consulting with the bank manager, the loan officer proposes a counter-offer contract (e.g. a higher interest rate or lower loan-to-valuation ratio) which the applicant either accepts or rejects. If the counter-offer is accepted, the loan is booked.

Based on this outline of the mortgage approval process, there are two broad channels through which lending distance can affect the bank's lending decision:

1. **Branch-borrower distance (or 'operational distance'):** the physical distance between the borrower and the branch office (and loan officer) affects the ability of loan officers to gather soft information. For example, greater operational distance is likely to be associated with higher transaction costs (e.g. transport and/or time costs). This is likely to adversely affect the incentives of loan officers to engage in interpersonal communication with the borrower, thereby reducing the extent of soft information produced and lowering the reliability of the signal of the borrower's creditworthiness, all other things being equal.

2. **Branch-headquarters distance (or 'hierarchical distance'):** the geographic distance between the bank's headquarters and the branch office affects the willingness of loan officers to use soft information and transmit it to the bank manager, assuming the bank manager is located in the bank's headquarters. For instance, an increase in hierarchical distance is likely to reduce the (ex ante) incentives for loan officers to collect soft information because the officers cannot credibly pass on the information to the bank manager (e.g. Aghion and Tirole [1997], Berger et al. [2005]) and may also increase the (ex post) costs of communicating that information (e.g. Radner [1993], Bolton and Dewatripont [1994], Garicano [2000]). In either case, soft information is less likely to be accurately transmitted from the branch office to the bank's headquarters, which again reduces the signal-to-noise ratio of the borrower's credit risk, all other things being equal.

While the process described above can be thought of as a general description of the origination process, in practice, the process is likely to vary across US banks, depending on factors such as the bank's organisational structure, how loan officers are paid and whether the bank relies on local branch offices or third-party originators, such as brokers, to originate loans. Theories based on incomplete contracting suggest that banks that are smaller and/or have a more centralised hierarchy have a comparative advantage in activities that make extensive use of soft information (e.g. Stein [2002], Berger et al. [2005]). As soft information is not verifiable and hard to communicate between loan officers and bank managers, soft information can be transmitted more easily within banks with smaller branch networks.[9]

However, small banks may not always obtain higher-quality signals of credit risk than large banks just because they are more likely to produce and use soft information; large banks can also implement strategies that allow them to 'harden' soft information (e.g. credit scoring). These strategies may overcome, or at least mitigate, the effect of distance on lender information (DeYoung et al. [2008]).

In other words, to correctly identify an effect of lending distance on performance, the econometric analysis needs to control for factors related to a bank's organisation and incentive structure. I will include controls for observable factors such as a bank's size and the extent of its branch network, as well as bank-level fixed effects to control for unobservable structural factors that influence credit quality. Essentially, I measure the effect of changes in lending distance on bank credit quality, *conditional* on the bank's organisational structure.

2.3 Data

The data in this paper are partly derived from the Home Mortgage Disclosure Act (HMDA) Loan Application Registry. Enacted by Congress in 1975, HMDA requires mortgage lenders located in metropolitan areas to collect data about their housing-related lending activity and to make these data publicly available. The HMDA dataset is generally considered to be the most comprehensive source of mortgage data in the US, and covers an estimated 80 per cent of all home loans nationwide and a higher share of loans originated in metropolitan statistical areas (Avery et al. [2007]). Whether a lender is covered depends on its size, the

[9]In small banks, the bank manager and loan officer are often the same agent, or at least operate at the same physical location. In large banks the bank manager and loan officer are typically separated in terms of physical location – the bank manager being located in the bank's headquarters while the loan officer is located in the bank's local branch (Agarwal and Ben-David [2012]). It will therefore be important to control for bank size in the regression analysis.

extent of its activity in a given metropolitan statistical area (or areas), and the weight of residential mortgage lending in its portfolio.

My sample of mortgage loan applications includes almost 500 million annual observations covering the period 1990-2010. For each application there is information on the loan application (e.g. the type of loan, the size of the loan, whether the loan is approved or not), the borrower (e.g. income, race, ethnicity), and the lending institution (e.g. the identity of the lender). Most importantly, the HMDA data identify both the Census tract of the borrower's property and the address of the lender's main office. However, the HMDA does not provide information on the separate branches of each bank.[10] However, the Summary of Deposits (SOD) dataset, provided by the Federal Deposit Insurance Corporation (FDIC), provides information on the address details of each US bank's headquarters and branches since 1994. I am therefore able to identify the location of all the bank branches of the HMDA lenders by matching the identities of lenders in both the HMDA and SOD.

I use geocoding software (provided by Stata and Google Maps) to separately calculate the geographic coordinates of the location of the borrower's Census tract, the location of the lender's main headquarters and the location of each of the lender's branch offices. Given the geographic location of the borrower's property and the location of the lender's headquarters and branches I can then calculate the geographic distance between each of them. Specifically, I calculate three separate measures of lending distance – the distance between the borrower's home and the nearest branch of the lender to which they applied (the borrower-branch distance), the distance between the nearest branch of the lender to which they applied and the headquarters of that lender (the lender-branch distance), as well as the overall distance between the borrower's home and the headquarters of the lender (the borrower-lender distance).[11]

[10] I do not know which branch office the borrower applied to, nor whether the application was made in person or over the phone or internet. I make the reasonable assumption that, if the bank has branches, the borrower applies to the bank branch that is nearest to their home.

[11] The measure of distance is based on a measure "as the crow flies" rather than a measure of road travel time.

The loan application data are not panel data because the HMDA only contains information on each individual loan at a given point in time. Specific borrowers cannot be tracked over time, although the lending behaviour of a particular bank can be observed across time. I create a pseudo-panel by aggregating the loan application data each year to the bank-county level. This means that I can track the behaviour of the *average* borrower in a given county across time. However, the HMDA does not provide information on the performance outcome of each individual loan. Instead, to obtain performance data I aggregate the loan-level information to the bank level and match this to bank financial statement information available in the Statistics on Depository Institutions (SDI) dataset, which is also provided by the FDIC. In doing so, I essentially compare the performance and distance of the *average loan* for each lender. The sample is restricted to depository institutions, as there are no financial data available for finance companies (or non-bank lenders).

From the FDIC financial statement data I obtain information on the fraction of home mortgage loans that are delinquent or, equivalently, non-performing. This measures the share of housing loans that are at least 90 days past due or that have been placed in non-accrual status. This acts as the key measure of (ex post) bank credit quality in all of the regressions. I also include a wide range of control variables for other observable lender characteristics. I include bank size (the log of total assets) and (the log of) the number of bank branches to control for a bank's organisational structure. I include a measure of bank mortgage loan growth (the annual change in each bank's total mortgage lending) to control for a bank's desire to expand its loan book. I also include indicators of the composition of the bank's loan portfolio – the share of housing loans in the loan portfolio and the share of home mortgage loans that are securitised – as controls for each bank's underlying business model. I also include the share of non-conforming (or jumbo) loans originated by each bank to proxy for a bank's ability to screen loans; banks need to exert more effort in screening non-conforming loans (Loutskina and Strahan [2011]) so banks with relatively high shares of jumbo loans would be expected to exert more screening effort on average. Finally, I include indicators of bank profitability (the net interest margin) and measures of funding structure and liquidity risk (the deposits to liabilities ratio and the risk-weighted capital to assets ratio). Details on each of these variables are provided in Appendix D. All variables in the models are expressed in natural logarithms, except variables that are measured as ratios (e.g. the ratio of home mortgages to total mortgages).[12] Some basic summary statistics for the variables used in the regression analysis are contained in Table 2.1.

Table 2.1: Bank-Level Summary Statistics

	Mean	Median	St Dev.	25th Pct.	75th Pct.
Non-performing loans ratio (%)	1.1	0.5	0.2	0.1	1.2
Borrower-HQ distance (kms)	70.1	21.8	222.6	13.9	41.1
Borrower-branch distance (kms)	34.0	12.3	141.2	8.7	19.2
Branch-HQ distance (kms)	43.4	11.1	180.8	3.2	25.8
Assets ($100,000)	16.7	1.8	244.5	0.9	4.2
No. of branches	15.5	4.0	107.0	2.0	9.0
Home mortgage share (%)	38.2	31.8	24.1	19.9	52.2
Home loan growth (%)	8.5	6.0	26.3	-2.5	16.2
Securitisation share (%)	19.2	0.0	29.7	0.0	32.9
Risk-weighted assets (%)	66.1	66.9	13.8	57.1	75.8
Deposit share (%)	91.4	94.5	9.7	8.8	98.7
Net interest margin (%)	2.5	2.4	2.5	2.1	2.8
Jumbo mortgage share (%)	6.0	1.6	12.6	0.0	5.7
Observations	71560				

Note: Based on full sample of US banks.

2.4 Identification

2.4.1 The OLS and fixed effects models

In Appendix A I outline a simple theoretical model that links the share of delinquent mortgages to lending distance. In the model, the share of delinquent loans is a function of the share of good borrowers in the population and the geographic distance between each bank

[12]This implies that I delete observations that have values of zero. This is likely to be the best approach to measuring the key variables – the share of delinquent loans and lending distance – as values of zero for these variables are likely to be errors. I have also estimated the models without taking natural logarithms and the results are qualitatively similar.

and borrower. We can loosely think of the share of good borrowers in the population as a measure of 'credit demand shocks' and the lending distance as an equilibrium outcome of both credit demand and supply shocks. Assuming that the county is the appropriate geographic measure of a mortgage lender's market, I would ideally estimate the following econometric specification at the bank-county level:

$$NPL_{bct} = D'_{bct}\beta + X'_{bct}\gamma + \epsilon_{bct} \text{ where } \epsilon_{bct} = \underbrace{\theta_b + \theta_c + \theta_t}_{\textit{Main effects}} + \underbrace{\theta_{bc} + \theta_{bt} + \theta_{ct}}_{\textit{Interaction effects}} + \nu_{bct} \tag{2.1}$$

Where b indexes banks, c indexes counties and t indexes years. The dependent variable (NPL_{bct}) is the share of delinquent (or non-performing) loans for bank b in county c at time t, the key independent variable (D_{bct}) is the average geographic distance between a borrower and lender in each county, bank and time period and X_{bct} is a vector of control variables that affect bank credit quality which also vary by bank, county and time period. The composite error term (ϵ_{bct}) includes unobservable bank-specific shocks (θ_b), county-specific shocks (θ_c) and time period shocks (θ_t), bank-county shocks (θ_{bc}), bank-time shocks (θ_{bt}), and county-time shocks (θ_{ct}), as well as an idiosyncratic error term (ν_{bct}) which is assumed to be *iid* and normally distributed. Based on my hypothesis I would expect the coefficient on distance to be significantly greater than zero ($\beta > 0$).

However, it is not possible to estimate the model at the bank-county-level as I only observe delinquent loans by bank rather than by bank *and* county (i.e. I do not observe *which* loans in a bank's portfolio are delinquent). So, instead, I aggregate the data to the bank-level by taking bank-time averages of all variables and estimate the following equation:

$$N\bar{P}L_{bt} = \bar{D}'_{bt}\beta + \bar{X}'_{bt}\gamma + \bar{\epsilon}_{bt} \text{ where } \bar{\epsilon}_{bt} = \underbrace{\bar{\theta}_b + \bar{\theta}_c + \bar{\theta}_t}_{\textit{Main effects}} + \underbrace{\bar{\theta}_{bc} + \theta_{bt} + \bar{\theta}_{ct}}_{\textit{Interaction effects}} + \bar{\epsilon}_{bt} \tag{2.2}$$

Here variables denoted with a 'bar' and the subscript "*bt*" are the bank-year averages of the bank-county data. So, for example, $N\bar{P}L_{bt}$ is the weighted average share of delinquent mortgage loans for bank b at time t, where the county-level weights are given by the share of mortgages originated in each county in each year.

To see the problems that arise when I estimate this bank-level equation rather than the bank-county-level equation, and to understand how I address these problems, consider what happens if I estimate the bank-level equivalent model by OLS. The OLS estimator of the lending distance coefficient would be:

$$\hat{\beta}_{OLS} = \beta + \frac{cov(\bar{D}_{bt}, \bar{\epsilon}_{bt})}{V(\bar{D}_{bt})}$$

The identifying assumption is that $cov(\bar{D}_{bt}, \bar{\epsilon}_{bt}) = 0$. This condition is violated if, for instance, a bank's attitude towards risk is associated with its choice of lending distance ($cov(\bar{D}_{bt}, \bar{\theta}_b) \neq 0$). In other words, pooled OLS does not control for unobservable bank-specific factors that may influence a bank's lending policies, such as its attitude towards credit risk. But, to the extent that these unobservable effects are bank-specific, they are dealt with by estimating a panel fixed effects regression at the bank-level with bank and time fixed effects in the estimating equation:

$$N\bar{P}L_{bt} = \bar{D}'_{bt}\beta + \bar{X}'_{bt}\gamma + \bar{\theta}_b + \bar{\theta}_t + \rho_{bt} \quad \text{where } \rho_{bt} = \bar{\theta}_c + \bar{\delta}_{bc} + \delta_{bt} + \bar{\delta}_{ct} + \bar{\nu}_{bt} \tag{2.3}$$

I will refer to this as the standard fixed effects specification. In this specification the effect of lending distance on delinquent loans is identified through variation *within* banks. But endogeneity could persist given the presence of credit demand shocks (e.g. $\bar{\theta}_c$) in the bank-level composite error term (ρ_{bt}). For instance, a region with a relatively high share of risky loan applicants could be more likely to seek out a less-informed distant lender than a region with a relatively low share of risky applicants ($cov(\bar{D}_{bt}, \bar{\delta}_c) \neq 0$). In this case, the positive correlation between distance and bad loans would be due to adverse selection in the pool of loan applicants and not due to information frictions.

2.4.2 The "purged" fixed effects models

To address these remaining endogeneity problems, I utilise the loan-level information on lending distance. While I only observe the dependent variable – the share of bad loans – at the bank level I observe the key explanatory variable – lending distance – at the more disaggregated bank-county level. This additional information allows me to adopt a 2-stage procedure where, in the first stage, I regress lending distance at the bank-county level (D_{bct}) on county fixed effects as well as bank-time and county-time fixed effects. I recover the estimated bank-time fixed effects from this equation and, in the second stage, I use these bank-time fixed effects as the key explanatory variable in identifying the effect of distance on delinquent loans. In doing so, I essentially estimate a bank-level equation that is "purged" of unobservable demand-side (county) effects. To be clear I estimate the following first-stage regression:

$$D_{bct} = Z'_{bct}\alpha + \delta_c + \delta_{ct} + \delta_{bt} + \mu_{bct} \tag{2.4}$$

This equation includes observable determinants of lending distance that vary by bank-county-year (Z_{bct}) and, more importantly, county-year dummies (δ_{ct}) that control for unobservable (time-varying) credit demand shocks, such as changes in local housing prices that may affect the risk profile of local borrowers. From this first-stage equation I recover the estimated bank-time fixed effects ($\hat{\delta}_{bt}$). This bank-level measure of distance will effectively be purified of demand shocks as the first-stage regression implicitly controls for borrower effects.[13] I then use the estimated bank-year fixed effects in a second-stage equation:

$$N\bar{P}L_{bt} = \hat{\delta}'_{bt}\beta + \bar{X}'_{bt}\gamma + \bar{\theta}_b + \bar{\theta}_t + \bar{\epsilon}_{bt} \tag{2.5}$$

My main interest is in the estimated coefficient β, which captures the effect of distance on the share of bad loans after eliminating the effect of credit demand shocks. The

[13] I obtain very similar results when the first-stage regression is estimated on bank-state level data rather than bank-county level data.

identifying assumption is that $cov(\hat{\delta}_{bt}, \bar{\epsilon}_{bt}) = 0$. The standard errors are estimated by block-bootstrapping at the bank level to account for the fact that the bank-year fixed effects are estimated regressors (Pagan [1984]).[14]

2.5 Results

Table 2.2 summarises the results of estimating a range of model specifications, including pooled OLS, standard fixed effects and 'purged' fixed effects. As the table indicates, lending distance, as measured as the distance between a bank's headquarters and the borrower, has a significant positive effect on the share of delinquent loans. This result is consistent with my main hypothesis that greater geographic distance is associated with more credit risk and hence leads to more bad loans, on average. The sign and magnitude of the estimated effect is robust across each specification. The effect of distance on bad loans is economically significant; based on the estimates from the 2SLS regression (with 'purged' fixed effects) (column 3), a one standard deviation rise in borrower-lender distance, which is about 275 kilometres (from a mean of 80 kilometres) is associated with about a 0.1 percentage point increase in the share of bad loans on average (so the average delinquent loan ratio would rise from around 1.2 per cent to almost 1.3 per cent). For the average US bank in my sample in 2010, the stock of home mortgage loans was around $480 million, so a one standard deviation increase in lending distance would correspond to an expected fall in bank profits for a typical US bank of around $0.48 million (= $480*0.1/100) on average.

Comparing columns 1 and 2, the effect of distance is robust to the inclusion of bank fixed effects. However, the inclusion of bank fixed effects causes the sensitivity of the mortgage delinquencies ratio to lending distance to fall by more than 50 per cent. This may indicate

[14] It is more difficult to control for potential endogeneity caused by correlation between bank-level lending distance and the bank-time specific effects (i.e. $cov(\hat{\delta}_{bt}, \theta_{bt}) \neq 0$). For instance, such endogeneity would arise if a bank chose to lend to more distant regions and expand lending to more risky borrowers. In other words, the strategy does not control for changes in bank lending standards that may be correlated with distance. However, I provide evidence in the robustness tests to suggest that lending standards are not driving the relationship between distance and credit quality.

Table 2.2: Non-Performing Loans and Lending Distance

	(1) OLS	(2) FE	(3) Purged FE
HQ-borrower distance	0.104***	0.0455**	0.0474**
	(8.81)	(3.03)	(2.61)
Size	-0.0860***	0.0588	0.0596*
	(-4.80)	(1.41)	(1.99)
Home mortgage share	-0.639***	-0.925***	-0.922***
	(-10.79)	(-6.83)	(-6.58)
Home loan growth	-0.608***	-0.483***	-0.481***
	(-18.25)	(-15.21)	(-15.66)
Securitisation share	0.0700*	0.154***	0.159***
	(2.38)	(3.97)	(4.66)
Risk-weighted assets	-0.00598	-0.224	-0.214
	(-0.06)	(-1.51)	(-1.43)
Deposit share	-0.182	0.777***	0.779***
	(-1.54)	(5.44)	(5.57)
Net interest margin	7.796***	7.117**	6.971**
	(3.59)	(2.77)	(3.24)
No. of branches	-0.0638***	0.0680	0.0702*
	(-3.37)	(1.95)	(2.55)
Jumbo mortgage share	0.151	-0.481***	-0.471***
	(1.62)	(-4.49)	(-5.53)
R^2	0.127	0.515	0.145
Observations	53292	53292	53225

t statistics in parentheses

Notes: Standard errors are clustered at the bank level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

that banks have risk-management strategies in place to partly mitigate the effect of distance on their ability to screen borrowers. More importantly, comparing columns 2 and 3, it appears that the estimated effect of lending distance on the share of delinquent loans is barely affected by the elimination of credit demand shocks. If anything, the effect of distance becomes slightly stronger in the "purged" bank fixed effects specification (column 3).[15]

The estimated coefficients on the control variables are generally statistically significant. Banks that have a high share of home mortgages in their loan portfolios typically experience fewer loan problems. This may be because these banks are specialists in home lending and are therefore better able to screen borrowers. Banks that experience faster home lending growth tend to have fewer credit problems, which appears to be inconsistent with the notion that excessive lending by commercial banks is an important determinant of bad loans. However, the negative correlation between loan growth and the share of NPLs may also reflect a short-term compositional effect; borrowers tend not to default on mortgage loans within the first year so the credit problems associated with a bank growing its loan book quickly may only be revealed with a delay. Banks that rely more on securitisation tend to have more bad loans, which is consistent with the findings of other recent studies on the US housing market that suggest that business models that are highly reliant on securitisation (the "originate-to-distribute" model) are associated with more lax lending standards and hence lower (ex-post) credit quality (e.g. Keys et al. [2010], Purnanandam [2011]).

I find that banks with a relatively high share of non-conforming (or jumbo) loans in their loan books are likely to have fewer credit problems on average. This supports the idea

[15] Based on unreported results, I find that the effect of lending distance is unchanged when I include a broader set of control variables based on the bank's location, such as the number of banks per capita or the level of local competition (measured by the Hirschman-Herfindahl index). The result is also upheld when I identify the relationship using variation within each bank's location through the inclusion of bank location fixed effects (e.g. zipcode dummies) rather than bank fixed effects. The result also stands when I control for the 'opaqueness' of each bank's borrowers. We might expect the effect of distance would be stronger for banks that lend to relatively 'opaque' borrowers (e.g. subprime borrowers). To address this, I include an interaction term for whether the bank is a subprime lender scaled by distance. I find that the interaction term is positive, implying a stronger effect for subprime lenders, but the differential effect is not significant. This might be because subprime lenders specialise in lending to less-creditworthy borrowers and adopt relatively more stringent credit tests.

that these banks exert relatively more screening and monitoring effort and hence make fewer bad lending decisions. A bank's net interest margin is positively correlated with bad loans, which may indicate that this measure of bank returns is a proxy for bank credit risk, and that higher returns compensate for additional risk-taking. In contrast, the risk-weighted capital ratio, which is usually taken to be a proxy for a bank's willingness to take risk, is generally an insignificant determinant of credit quality.[16]

The estimated effect of bank size (assets) and the number of bank branches reverses when bank fixed effects are included in the specification. Both variables are negatively correlated with the share of bad loans in the pooled OLS regression (column 1) but positively signed when the bank dummies are included (columns 2 and 3). The combination of these results may indicate that large banks typically have better risk management strategies than small banks, on average. But if these risk strategies are largely time-invariant, then the fixed effects regressions, by controlling for such unobservables, induce a positive effect of bank size on the share of delinquent loans. In other words, the regression results are consistent with the argument that, conditional on the bank's risk-management practices, banks that have relatively large branch networks also have more internal agency problems, which leads to more bad lending decisions.

2.6 Robustness Tests and Model Extensions

In this section I test the robustness of the main results and explore some extensions of the basic model. Firstly, I consider the evidence as to *why* there is a link between geographic distance and credit quality. The estimated model in the paper explores the link between distance and credit quality but is silent on the underlying mechanism through which distance

[16] I also find that the link between distance and credit quality holds when I instead measure credit quality through each bank's net charge-off ratio (i.e. net charge-offs are equal to loan write-offs less recoveries). The net charge-off ratio is designed to measure both the probability of default, as with the non-performing loan ratio, and the loss given default. Note also that the non-performing loan ratio excludes off-balance sheet, or securitised, loans that have turned bad. I have also estimated the model using each bank's lending distance for on-balance sheet loans and I again find very similar results to those presented here.

and credit quality may be related. I consider the following explanations:

- Internal problems in producing and transmitting soft information between bank employees (the information frictions hypothesis)
- Deliberate changes in bank lending policies (the credit standards hypothesis)
- Adverse selection in the pool of loan applicants (the winner's curse hypothesis)
- Principal-agent problems associated with the origination channel (the mortgage broker channel hypothesis)

I address these explanations in reverse order, beginning with the mortgage broker channel hypothesis. I do this because I believe the evidence against the mortgage broker and winner's curse hypotheses is the most compelling. In this section I also examine how the relationship between lending distance and credit quality has evolved over time, as this provides further insight into the mechanisms that underpin the relationship. Finally, in a robustness test, I also demonstrate that the link between lending distance and credit quality persists even when I utilise an alternative source of information on mortgage delinquencies.

2.6.1 Mortgage broker principal-agent problems

In this subsection I consider whether the link between measured lending distance and credit quality is due to banks' reliance on mortgage brokers. If banks that lend at relatively long distances are more likely to originate loans through brokers, then the positive correlation between distance and bad loans may ultimately stem from principal-agent problems associated with mortgage brokers rather than distance-related information problems.

Ideally, we would directly observe the origination channels used by each bank and address this issue by simply including the share of loans that each bank originates through brokers ('broker loans') as a control variable. But there is limited bank-level information available on broker loans. I therefore adopt two approaches to proxy for bank-level broker loans.

My first measure estimates each bank's reliance on broker loans based on whether its securitised loans include loans originated by brokers.[17] This is a direct measure of a bank's reliance on brokers, although the measure is based on *securitised* mortgage loans, which may not be representative of all mortgage loans originated by the bank.[18] I construct a dummy variable which is equal to one if the bank securitised broker loans and is equal to zero otherwise. I interact this dummy variable with the (log of) distance to see whether the share of bad loans is more sensitive to distance for the banks that securitise broker loans. I then re-estimate the panel regression given by equation (2.5) including the new interaction term ('Distance x broker') as an additional RHS variable.

My second proxy for a bank's reliance on the mortgage broker channel is a more indirect measure, which combines three sources of information: the cross-county distribution of each bank's mortgage lending, the cross-county distribution of each bank's branch network, and employment data disaggregated by county and industry. Using information from the HMDA, I firstly measure the volume of loans originated by each bank in each county (i.e. the bank's cross-county loan portfolio). Using the SOD, I then identify all the counties in which a bank has local branches (i.e. the bank's cross-county branch network). Finally, I use the Quarterly Census of Employment and Wages (QCEW) to calculate the share of all banking sector employees that work as mortgage brokers in each county.[19] These data allow me to construct an 'employment-based' measure of each bank's reliance on broker loans. I firstly proxy the volume of mortgage broker loans originated by each bank by multiplying each bank's *total* lending in each county by the share of banking sector employees that are brokers

[17] I use financial statement data from the US Call Reports to calculate the share of securitised broker loans. Specifically, I measure the share of securitised broker loans as the total volume of wholesale originations and purchases of closed-end 1-4 family residential mortgage loans for sale (RCONF068 + RCONF069) divided by the total volume of originations and purchases of closed-end 1-4 family residential mortgage loans held for sale (RCONF072+RCONF073). This measure includes both first lien and junior lien home mortgages. I match the data from the US Call Reports to the SDI data using the bank's FDIC identification number.

[18] Moreover, the variable can only be constructed for a small sample of banks because most banks do not report these data items. More importantly, this measure of securitised lending is only available from 2006 onwards. To obtain a longer time series I assume that banks that securitised broker loans in 2006 also relied on brokers in the period prior to 2006.

[19] I loosely define the banking sector to include mortgage broker firms (NAICS code 522310), depository institutions (NAICS code 5221) and non-depository institutions (NAICS codes 5222).

in that county.[20] I then adjust this employment-based estimate depending on whether the bank has local branches in the county – if the bank has no branches in a given county, I assume that all loans in that county are originated through brokers.[21] I then re-estimate the panel regression given by equation (2.5) but include my constructed estimate of the share of loans originated through mortgage brokers ('Broker share') as another RHS variable. I also interact this employment-based measure of broker loans with the (log of) distance ('Distance x broker share') to determine if the effect of distance is solely coming through broker loans. The results of re-estimating equation (2.5) with the inclusion of these proxies for mortgage broker lending are given in Table 2.3. I include all the same RHS variables as before (though these are not shown for parsimony).

Table 2.3: Non-Performing Loans and Mortgage Brokers

	(1) Direct	(2) Indirect
HQ-borrower distance	0.0388**	0.0597***
	(2.79)	(3.43)
Dist x Broker	0.0687	
	(1.51)	
Broker share		-0.172
		(-1.17)
Dist x Broker share		0.00795
		(0.21)
R^2	0.498	0.502
Observations	63142	53066

t statistics in parentheses

Notes: Standard errors are clustered at the bank level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results imply that the effect of lending distance on the share of bad loans is stronger for mortgage broker-originated loans. The coefficient on the interaction term in both specifications is positively signed.[22] However, the effect is not statistically significant. Banks that

[20]This assumes that each bank that competes in a particular county is equally reliant on brokers in that county.

[21]This implicitly assumes that borrowers do not apply for loans at branches outside their home county.

[22]Note that the broker share variable is not included in the 'direct' model because, being a dummy variable, it is perfectly collinear with the bank fixed effects.

originate more broker loans also appear to have fewer credit problems, on average, as implied by the negatively-signed coefficient estimate for the broker share variable. Again though, the effect is not statistically significant.[23] More importantly, the main effect of lending distance on the share of bad loans (shown in the first row of each column) is not materially affected by the inclusion of the mortgage broker variables. Overall, it appears that the link between distance and credit quality is not due to principal-agent problems associated with banks and their mortgage brokers.

2.6.2 The winners' curse hypothesis

I now test whether the link between distance and bad loans is due to adverse selection in the pool of borrowers. This 'selection bias' is due to risky borrowers being more likely to approach distant banks than close banks because the borrowers know the distant banks are less informed about their creditworthiness.[24] I adopt two different approaches to gauge the extent to which the distance-performance link is affected by selection bias. The first approach relies on variation in the competitive conditions of county-level housing markets. The second approach relies on variation in the potential 'screening ability' of banks.

The first approach is based on the observation that, if the winner's curse hypothesis is true, banks that lend to outside markets dominated by local lenders should report higher shares of bad loans than banks lending to outside markets in which there are very few local lenders.

To test this idea I construct a measure of each bank's 'outside lending' in 'locally-dominated markets'. I need two pieces of information to construct this estimate. Firstly, I measure the volume of lending by each bank in each county in which it is *not* headquartered

[23]The lack of significant results for the mortgage broker variables may reflect the fact that reliance on mortgage brokers is captured by a bank's business model, and with these models being largely time-invariant, any effect is absorbed through the bank dummies. This is consistent with the fact that pooled OLS estimates show the broker variables to be statistically significant. Regardless, the main effect of distance remains unchanged.

[24]It may not be that the borrowers themselves know which banks are more lax in their credit screening, but that mortgage brokers 'steer' borrowers towards banks that the brokers know are more lax. In this case, the adverse selection problem would be an example of the mortgage broker principal-agent problem.

(i.e. each bank's lending outside its home county). Secondly, I measure the share of total lending in each county that is due to lenders that are headquartered there (i.e. each county's local lending). I can calculate both measures because I know the number of loans originated by each bank in each county and I know the county location of each bank's headquarters. For each bank, I multiply the number of loans originated in each outside county by the share of loans made by lenders that are local to those counties. For each bank, I then aggregate across all counties to obtain the bank-level share of outside lending in locally-dominated markets. This measure will be relatively high for banks that either originate a lot of loans outside their home county or originate loans in counties that are dominated by local lenders.

I re-estimate equation (2.5) again, but now include the share of lending of each bank outside its home county ('outside share'), the share of lending by each bank outside its home county in locally-dominated markets ('L.D. outside share'), and the interaction between (the log of) lending distance and the share of lending by each bank outside its home county in locally-dominated markets ('distance x L.D. outside share'). As before, I include bank fixed effects, time fixed effects and all the usual control variables. If the winner's curse hypothesis is true, we should observe that the positive effect of distance on the share of bad loans is relatively high for banks lending in outside, but locally-dominated, markets, all other things being equal. More specifically, we should observe that the interaction variable for distance and the share of outside lending by banks in locally dominated markets is *positively* correlated with the share of bad loans.

The second approach is based on a generalisation of the theoretical model outlined in the Appendix. One implication of the model is that each bank's equilibrium market share is solely a function of exogenous factors, such as the state of the credit screening technology and the share of good borrowers in the population. A simplifying assumption is that banks are equally capable of screening borrowers; the model abstracts from potential variation in bank characteristics, such as size or age or market presence, that may influence a bank's ability to screen borrowers. I now make the more realistic assumption that banks that are larger and/or older and/or have a larger market presence are more likely to have accumulated knowledge about borrowers, are more capable of screening borrowers and therefore less likely to suffer the winner's curse (Shaffer [1998]).

I re-estimate equation (2.5) again but now include variables that are correlated with a bank's screening ability, such as size, age and average market share, as well as the interaction between these variables and (the log of) lending distance. I measure a bank's size as the (log of) assets, which is identical to the benchmark model. I measure a bank's age as the (log of) the number of years since the bank was established. I measure a bank's market presence as the average share of total lending in each county in which it originates loans (i.e. the bank's *average* market share).[25] If the winner's curse hypothesis is true, distance should have less of an impact on the share of bad loans for the banks with more 'screening ability', ceteris paribus. More specifically, we should observe that the interaction variable for distance and 'screening ability' is *negatively* correlated with the share of bad loans.

Overall, the results provide little support for the winner's curse hypothesis under either approach (Table 2.4). The results of the first approach are summarised in the first column. The positive coefficient on the 'outside share' variable implies that banks that lend outside their home county are *more* likely to report bad loans, but the negative coefficient on the 'LD outside share' variable suggests that banks that lend outside their home county and in locally-dominant markets are *less* likely to report bad loans. Moreover, although the interaction coefficient is positive, implying that the effect of distance is stronger for banks that lend to locally-dominated outside markets, the effect is not statistically significant. Note also that the main effect of distance remains positive and significant.

The results of the second approach also provide little support for the adverse selection effect. Regardless of the measure of screening ability, in columns 2 to 4, the interaction coefficient is positive (and significant in the case of size). This implies that the effect of

[25] I calculate each bank's share of total home mortgage lending in each US county using the HMDA and then average these county-level estimates across all the counties in which the bank lends to get a bank-level estimate.

Table 2.4: Non-Performing Loans and the Winner's Curse

	(1)	(2)	(3)	(4)
	Outside Lending	Size	Age	Market Share
HQ-borrower distance	0.0573**	-0.332**	0.0210	0.0354*
	(3.10)	(-3.27)	(0.91)	(2.43)
Dist. x LD outside .	0.0442			
	(0.27)			
Outside share	0.0472			
	(0.77)			
LD outside share	-1.052			
	(-1.62)			
Size		-0.0896		
		(-1.74)		
Distance x size		0.0297***		
		(3.75)		
Age			-0.00330	
			(-1.42)	
Distance x age			0.000429	
			(1.42)	
Market share				-3.969***
				(-3.98)
Distance x mkt. sh				0.541
				(1.92)
R^2	0.516	0.516	0.516	0.517
Observations	55154	55154	55154	55154

t statistics in parentheses

Notes: Standard errors are clustered at the bank level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

distance in increasing the share of bad loans is stronger for 'more capable' banks, which directly contradicts the winner's curse hypothesis. This suggests that these banks are not actually more capable of screening borrowers. Rather, it seems that particularly large and well-established banks that have national branch networks are more likely to experience internal agency problems.

2.6.3 The information frictions and credit standards hypotheses

To distinguish between the information frictions and credit standards hypotheses, I re-estimate the benchmark panel regression model replacing the lending distance variable with its decomposition – the branch to borrower distance (or operational distance) and the branch to HQ distance (or hierarchical distance). The fixed-effects specification is re-written to be:

$$NPL_{bt} = BRDIST'_{bt}\beta_1 + HQDIST'_{bt}\beta_2 + X'_{bt}\beta_3 + \theta_b + \lambda_t + \epsilon_{bt} \tag{2.6}$$

where $BRDIST_{bt}$ is the branch-to-borrower distance and $HQDIST_{bt}$ is the branch-to-headquarters distance and all other variables are as defined earlier.[26]

This decomposition of lending distance is designed to help me to identify the cause of the link between distance and credit quality. The specification allows me to disentangle the winner's curse and information frictions hypotheses. From a borrower's perspective, the distance to the nearest bank branch is likely to be more important in their loan application decision than is the distance to the nearest bank headquarters. So if the winner's curse is important and risky borrowers seek out distant banks then this effect is more likely to be correlated with the branch-to-borrower distance than the branch-to-HQ distance. In contrast, if internal agency problems are important to loan outcomes then this effect is

[26]This method introduces an implicit sampling bias as the decomposition requires banks to have at least one branch. Banks that rely solely on brokers to originate loans will be excluded from the analysis. This sampling bias may also explain why my measure of overall borrower-HQ distance displays larger cyclical fluctuations than either the borrower-branch or branch-HQ distance during the subprime boom-bust cycle; it is likely that some part of the observed overall fluctuations was due to changes in the share of banks that were totally reliant on brokers to originate loans.

probably more likely to be correlated with the branch-to-HQ distance. This is an imperfect test as bank internal problems could equally affect the incentives of loan officers to gather the necessary information and to transmit the information. But, in general, if the *acquisition* of borrower information matters to the performance of a bank, the coefficient on the branch-borrower distance should be positive and significant ($\beta_1 > 0$). If the *transmission* of borrower information is important, then the coefficient on the branch-HQ distance should be positive and significant ($\beta_2 > 0$).

The decomposition of lending distance also helps to distinguish between the lending standards and information frictions hypotheses. Under both hypotheses the branch-HQ distance should be positively correlated with the share of bad loans ($\beta_2 > 0$) as the bank expands into new markets. However, under the lending standards hypothesis, the branch-borrower distance should be *negatively* correlated with the share of bad loans ($\beta_1 < 0$). This is because a bank that implements lax credit standards to chase market share would be expected to *reduce* the distance between its branches and potential borrowers, on average.[27] So the sign of the coefficient on the branch-borrower distance may provide some further insight into the underlying drivers of the distance-performance relationship.

The results of estimating the model after decomposing lending distance into the branch-borrower distance and the branch-HQ distance are reported in Table 2.5. It appears that the aggregate (positive) effect of lending distance reflects *both* the branch-borrower distance and the branch-HQ distance. More importantly, the positive correlation between the branch-borrower distance and the share of bad loans argues against the lending standards hypothesis. The results are only suggestive, but the positive link between the share of bad loans and the branch-HQ distance supports the idea that bank internal agency problems contributed to the decline in average credit quality.

[27] This argument assumes, to some extent, that banks view local branches and brokers as complementary ways in which to expand geographically. Of course, it may be the case that banks expand aggressively through brokers, while leaving the branch network unchanged, in which case we might expect a positive link between branch-borrower distance and the share of bad loans. However, the results of the mortgage broker robustness tests suggests that the broker channel is not driving the key results.

Table 2.5: The Decomposition of Lending Distance

	(1) OLS	(2) FE
Branch-borrower distance	0.100***	0.0393**
	(7.31)	(2.77)
Branch-HQ distance	0.0279*	0.0488*
	(2.27)	(1.99)
R^2	0.133	0.526
Observations	43920	43920

t statistics in parentheses

Notes: Standard errors are clustered at the bank level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

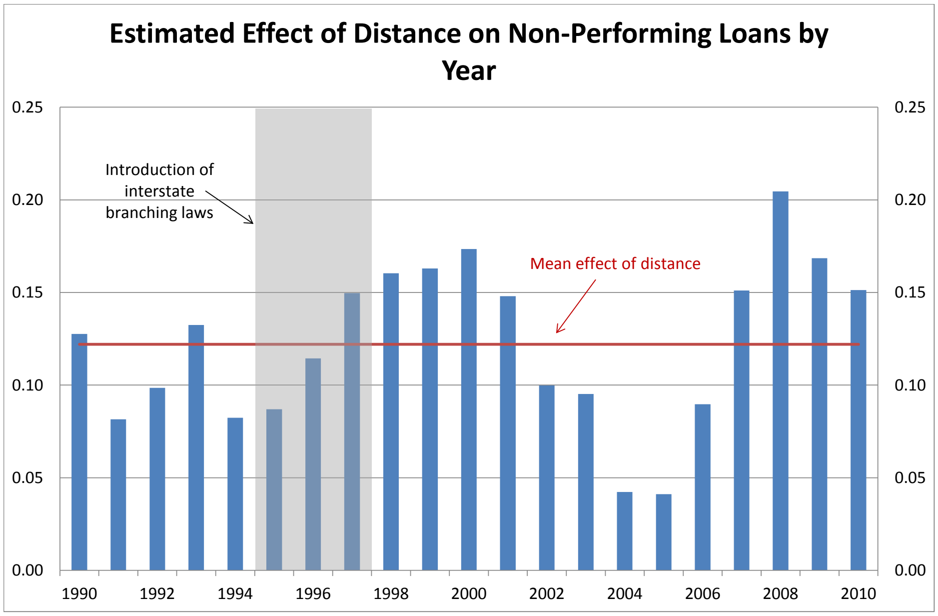
2.6.4 The varying effect of lending distance on credit quality

I also examine how the sensitivity of bad loans to distance varies over time. The time variation in this sensitivity may provide further insights into the link between credit quality and geographic distance. To do so, I re-estimate equation 2.3 by OLS, but with the distance variable interacted with each of the time fixed effects. This is designed to capture how the link between distance and credit quality varies by year. More specifically, I estimate the equation:

$$NPL_{bt} = \sum_{t=1}^{T} (D_{bt} * \theta_t)'\beta_t + X'_{bt}\gamma + \theta_b + \epsilon_{bt} \tag{2.7}$$

The resulting OLS coefficient estimates for each year (i.e. the $\beta_t's$) are displayed in Figure 2.3. Notably, the positive effect of distance on bad loans appears to be strongest in two important periods – the period immediately following the enactment of the interstate branching laws in 1997-2000, as well as the period since the subprime crisis in 2007-2010. The deregulation of branching laws acted as a positive exogenous shock to lending distance, encouraging banks to expand geographically into new regions. This may have contributed to more bad lending decisions. The figure also suggests that the variation in the sensitivity of loan quality to lending distance is partly countercyclical, with the sensitivity being above-average during both the 2001 and 2008-09 recessions.

Figure 2.3: The Varying Sensitivity of Bad Loans to Lending Distance



This countercyclical pattern to the credit quality-distance sensitivity may point to the adverse selection effect as being a contributing factor to the positive correlation between lending distance and the share of non-performing loans. It could be argued that there will be a relatively high proportion of "bad" applicants during a recession, so adverse selection may be strongest at this time. On the other hand, risky borrowers may be discouraged from borrowing during a recession given the higher level of uncertainty. Alternatively, the pattern could support the lending standards hypothesis as banks that expanded the most aggressively during the boom could have experienced the largest increases in both lending distance and risky lending. Under this scenario, the results of these riskier decisions were only revealed when economic conditions deteriorated during the downturn.

One of the most interesting features of Figure 2.3 is the lack of evidence that the sensitivity of bad loans to distance has fallen over time. If anything, the effect has been strongest during the financial crisis. This suggests that advances in loan production technology, such as the ‘hardening’ of soft information through credit scoring and automated underwriting, has done little on average to mitigate the effect of lending distance on bank credit quality.

2.6.5 Mortgage delinquencies by lender and region

To identify the mechanism underlying the relationship between credit quality and lending distance it would be ideal to have information on mortgage delinquencies across lenders *and* borrowers. This variation would allow for the separate identification of credit demand and supply shocks (Khwaja and Mian [2008]). Such disaggregated borrower-lender data are generally unavailable, but some limited information can be gleaned from the US Department of Housing and Urban Development (HUD)’s Neighbourhood Watch programme. As part of this programme the HUD collects information on the stock of federally-insured home mortgages that are delinquent by lender and geographic region. Importantly, this information can be matched with estimates of lending distance by lender and region from the HMDA dataset. This allows me to examine the relationship between credit quality and distance by region and lender.[28]

There are a couple of caveats with the Neighbourhood Watch data. Firstly, the decomposition of mortgage delinquencies by region and lender is only available for one particular month (October 2012). I therefore cannot utilise time-series variation to identify the relationship between distance and credit quality. Secondly, the HUD data cover only a subset of residential mortgages known as Federal Housing Administration (FHA) mortgages. These are home mortgages that are insured by the government under the FHA mortgage insurance programme. The FHA is a federal government agency that was created to make affordable housing available to low and average income borrowers, so FHA loans have historically been for borrowers with relatively low credit scores. However, the FHA insures these private

[28] I use the most recent HMDA data for 2011 to construct the lending distance measure, whereas the HUD data on mortgage delinquencies apply to 2012. This lag between measured lending distance and credit quality is unlikely to be a significant problem as the delinquent loans were originated in the two years prior to being reported. Regardless, the lag should reduce the chance of finding a significant relationship between credit quality and distance.

mortgages, so mortgage lenders are guaranteed repayment if the borrower defaults. In other words, despite the fact that FHA borrowers are relatively less creditworthy, the loans carry limited credit risk to the lender.

Traditionally, FHA loans have accounted for a relatively small share of the residential mortgage market (of less than 10 per cent). However, following the subprime crisis, the share of FHA loans rose significantly to around 25 per cent of all residential mortgages in 2011 (Avery et al. [2012]). Moreover, there has been a shift in the characteristics of FHA applicants towards more creditworthy borrowers (Avery et al. [2012]). Both of these factors suggest that the results of the analysis based on the FHA mortgage market may carry over to the overall housing market. More importantly, the limited credit risk for the bank (due to the explicit government guarantee) suggests that FHA lending is not related to bank credit standards. A finding that there is a link between lending distance and credit quality in this particular market may therefore provide evidence against the lending standards hypothesis.[29]

The following cross-sectional regression is estimated to examine the link between mortgage delinquencies and lending distance at the state-lender level:

$$NPL_{bs} = D'_{bs}\beta + X'_{bs}\gamma + \theta_s + \epsilon_{bs} \tag{2.8}$$

Where NPL_{bs} is the share of mortgage delinquencies recorded by lender b in state s, D_{bs} is the lending distance of loans originated by each lender to the average borrower in each state, and X_{bs} is a set of lender-state control variables. The key measure of HQ-borrower lending distance is constructed from the HMDA dataset as before, though the measure now applies specifically to FHA loans and varies by state *and* lender. I also use the HMDA data to construct a measure of the average income of loan applicants that varies across lenders within each state. This is designed to control for any remaining borrower risk at the state-lender

[29] Another useful feature of the Neighbourhood Watch data is that the information covers loans originated by non-bank lenders (i.e. mortgage companies). Mortgage companies do not have branch networks and typically lend at longer distances than banks, so their inclusion could be important in identifying why there is a link between credit quality and distance.

level. I include state fixed effects (θ_s) to control for unobservable state-level factors, such as average credit scores, that may influence the average level of delinquencies. This test relies on variation across lenders *within* states to identify the effect of distance on delinquencies. Essentially, the estimation procedure compares the mortgage delinquency outcomes of two lenders lending to borrowers residing in the same state, but where one lender is located further away than the other.

Table 2.6: Non-Performing Loans and Lending Distance by Lender and State

	(1) OLS	(2) State FE
Lending distance	0.0142*	0.0204**
	(2.28)	(3.20)
Household income	0.0469	0.0536
	(1.88)	(1.88)
Constant	-0.135	-0.206
	(-1.28)	(-1.64)
R^2	0.039	0.177
Observations	827	827

t statistics in parentheses

Notes: Standard errors are clustered at the state and lender levels.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results indicate that, controlling for unobservable credit demand shocks at the state level, lending distance still has a positive effect on the share of mortgage delinquencies. In fact, the effect is stronger when I control for state fixed effects (column 2) compared to the case of OLS (column 1). A one standard deviation increase in lending distance is associated with a 2.2 percentage point increase in the share of bad loans (from a mean of 15.7 per cent). This suggests that distance has an economically meaningful impact on mortgage performance. The fact that mortgage lenders retain little credit risk in originating FHA loans suggests that the supply of FHA loans is not directly related to risk-taking. Given the state fixed effects largely absorb unobservable variation in credit demand, the results suggest that the link between mortgage delinquencies and geographic distance is due to other credit supply factors, such as internal agency problems.

2.7 Conclusion

The aim of this paper was to identify the causal effect of lending distance on bank credit quality for US mortgage lenders. I document some new stylised facts about the US home mortgage market. I show that the geographic distance between the headquarters of US mortgage lenders and their borrowers rose rapidly during the housing boom of the 2000s, but has fallen sharply since the subprime crisis.

I also show how changes in bank credit quality are related to these changes in lending distance. Assuming that geographic proximity is correlated with the quality of signals that banks receive about borrowers, the rapid increase in lending distance over the boom period appears to have been associated with banks receiving lower-quality signals of borrower creditworthiness when approving mortgage applications and hence contributing to greater bank risk-taking.

I find a clear and robust positive link between a bank's average lending distance and its share of delinquent loans, controlling for a range of observable and unobservable bank characteristics. The relationship between geographic distance and credit quality is also robust to various model specifications.

I do not find clear-cut evidence to explain why there is a link between distance and credit quality. Based on a battery of tests, the weight of evidence suggests that bank internal agency problems associated with geographic expansion were a contributing factor, but I cannot rule out the possibility that some omitted variable may be driving the relationship. For instance, banks that lend at long distances may also be more likely to expand aggressively by relaxing their lending standards and this intention to take on more risk may have also contributed to the rise in bad loans. Nevertheless, the identification strategy appears to rule out demand-side explanations for the observed relationship, such as risky borrowers being relatively more likely to approach distant banks.

When I decompose the overall geographic distance between the borrower and bank headquarters into the borrower-branch and branch-HQ distance I find evidence that both components are positively correlated with the share of bad loans. The positive link between the branch-borrower distance and the share of bad loans argues against the lending standards hypothesis as banks that were aggressively chasing marginal borrowers through branch expansion should have experienced a rise in bad loans but a *decline* in the distance between branches and borrowers.

I also find evidence that the sensitivity of credit quality to distance varies with the cycle, and that the strength of the relationship between mortgage delinquencies and distance has not declined over time, despite advances in bank screening and monitoring technology.

Appendices

A Theoretical Model

A.1 Basic framework

To derive testable hypotheses of how geographic distance affects bank loan quality I outline a simple model. The model is based on the original Salop (1979) spatial model of product differentiation and the setup of my model closely follows that of Dell'Ariccia [2001] and Hyytinen [2003].[30] There are n banks in the economy, with each bank indexed by i. These n banks are located on a circle of unit circumference and are assumed to be equidistant so that the distance between any two banks is $1/n$.[31] There are N potential borrowers (households) that are uniformly distributed around the circle. Both households and banks are assumed to be risk-neutral.[32]

There are two types of borrowers in the population (where θ denotes the type) – both

[30] We can interpret the spatial dimension literally as representing banks with different locations or we can interpret it metaphorically as representing some other qualitative difference in the services provided.

[31] Note that maximal differentiation is exogenously imposed and I do not allow the bank to optimally choose its location.

[32] The more realistic assumption of risk-averse households would reduce the model's tractability.

good ($\theta = G$) and bad ($\theta = B$) borrowers. The difference between the two types of borrowers is that good borrowers purchase a housing asset that generates a cash flow ($R > 0$) with probability equal to one, while bad borrowers purchase a housing asset that generates no cash flow with probability equal to one. Instead, the bad borrower extracts a private benefit ($B > 0$) from having a loan. As such, it is never profitable for a bank to lend to a bad borrower, which provides an incentive to the bank to screen borrowers. In the total population of borrowers (n) the proportion of good borrowers is λ and the proportion of bad borrowers is $1-\lambda$, with these population shares being common knowledge to all lenders. I will sometimes refer to the good (bad) borrowers as low (high)-risk borrowers.

Banks are monopolistically competitive and compete with each other in terms of lending rates. The banks post a loan interest rate of r (the lending rate) and can obtain limitless funding at the risk-free interest rate of ρ. Banks compete locally in that, when applying for a loan, the marginal borrower decides between the two closest banks based on the posted lending rates and the distance to each bank.

The location of a borrower is denoted d, but this variable can also be thought of as the lending distance between the bank and the borrower. When approaching a bank for a loan, borrowers incur a transportation cost (τ) per unit of distance travelled. I assume that borrowers have no initial wealth and must therefore borrow the full amount of the value of the home, which is normalised to one.

A.2 Credit screening technology

I assume the bank cannot perfectly observe the type of a mortgage loan applicant. However, the bank has access to a screening technology that allows it to infer the type of the household at no cost. I assume that all banks have access to the same screening technology that generates a binary signal (s) of the borrower's creditworthiness. The screening technology generates a signal $s = G$ when the borrower is tested to be good and $s = B$ when the borrower is tested to be bad, where s corresponds to the true type θ with probability q.

$$q = P(s = G|\theta = G) = P(s = B|\theta = B)$$
$$(1 - q) = P(s = B|\theta = G) = P(s = G|\theta = B)$$

I assume that $1/2 < q < 1$ so that signals are noisy but informative. The higher is q the more accurate is the credit signal; if $q = 1/2$ then the signal is uninformative as the borrower is equally likely to be good or bad, while if $q = 1$ then the signal is a perfect indicator of a borrower's type. The signal is also assumed to be symmetric in that banks are equally likely to grant a loan to a bad borrower as to reject a good borrower (i.e. Type I and Type II errors are equally likely). Banks observe the noisy signal and only lend if the test claims the borrower is good.

Crucially, I assume that the accuracy of the signal is a function of lending distance, d:

$$q = 1 - d/2 \quad \text{where d} \in \text{(0,1/n)}$$

This specification captures the idea that banks enjoy an informational advantage in market segments in which they operate, but this advantage diminishes as banks lend to customers located further away.

The expected utility of a good borrower from being granted a loan by bank i at distance d is given by:

$$U_G = q\pi(r_i) - \tau d = q(R - r_i) - \tau d$$

Similarly, the expected utility of a bad borrower from being granted a loan by bank i at distance d is given by:

$$U_B = (1 - q)B - \tau d$$

Note that the expected utility of borrowers is only dependent on the loan interest rate for good borrowers, but not for bad borrowers. The assumption that only good borrowers are sensitive to changes in lending rates is strong, but could be relaxed.[33]

A.3 Nature of competition and timing of events

The time structure of the model is as follows:

1. **Stage 1:** Banks simultaneously post loan interest rates to compete for customers
2. **Stage 2:** Borrowers observe the menu of lending rates and choose a bank to travel to based on the observed interest rates and the distance to each bank
3. **Stage 3**: i) *(First round screening)* Banks screen loan applicants. If an applicant is approved for a loan, the game ends, but if an applicant is rejected the applicant decides whether to approach another bank or exit.
4. **Stage 3**: ii) *(Second round screening)* Previously rejected applicants are either accepted by the next nearest bank or rejected again. Either way the game ends.[34]

There are a few things to note about the setup of the model. Firstly, the loan interest rate may be set depending on the perceived characteristics of a borrower. This is consistent with lending practices in the subprime lending market, where risk-based loan pricing is extensive. Secondly, geographic distance plays two roles in the model – greater geographic distance reduces the likelihood that a borrower will approach a given bank, given the higher transportation costs involved, but greater geographic distance also reduces the likelihood that a borrower will be correctly classified as either a good or bad type. In other words, banks are more likely to make both Type I and Type II errors about distant borrowers than

[33] For example, the bad borrowers could be defined as being "over-confident" about future (positive) house price growth and therefore less sensitive to variations in the cost of borrowing than good borrowers (Hyytinen [2003]).

[34] Under the assumption of independent screening, the probability that a previously rejected applicant will pass the test in the second round is given by $q(1-q)$.

close borrowers. Thirdly, applicants are only allowed to re-apply for loans once, which is fairly realistic in the context of the US home mortgage market, where both the monetary and reputational cost of being rejected for a loan is typically high.[35]

A crucial assumption is that banks cannot distinguish between initial and rejected applicants. This assumption is realistic in most credit markets, as banks have an interest in not sharing private information about individual borrowers. However, it may be less realistic in the US home mortgage market as credit bureaus often collect information on the rejection rate of borrowers, and this information can be made available to lenders.

A.4 Borrower loan demand

I solve the model backwards and so begin by deriving the loan demand functions for rejected applicants during the second round of credit screening. Consider the problem of the good applicant that has been rejected in the first round of screening and is deciding between reapplying for a loan at another bank or exiting the market. To prevent rejected borrowers from exiting I assume the following condition holds:

$$q\pi(r_i) - \tau(1/n) \geq 0 \tag{2.9}$$

This condition means that the expected return from reapplying for a loan exceeds the return from exiting *regardless* of the borrower's location. In other words, even if the borrower is forced to travel the maximum possible distance to a bank ($1/n$), they will still want to reapply for a loan rather than exit. By assuming this 'participation constraint' holds, I ensure that all good, but rejected, borrowers reapply for loans. A similar necessary condition holds for rejected, and bad, loan applicants:

$$(1-q)B - \tau(1/n) \geq 0 \tag{2.10}$$

[35] I also assume that borrowers may reapply for loans at a rival bank, but not at the same bank that originally rejected them. For a model in which borrowers may reapply to all banks in the market, see Shaffer [1998].

Again, when this condition holds, a bad borrower will want to reapply for a loan rather than exit. Now consider the problem of the first-time borrower. When deciding whether to initially apply for a loan, first-time borrowers will be forward-looking and will anticipate the possibility of rejection, in which case they may want to reapply to another bank. An initial good borrower will want to apply for a loan from bank i rather than its closest rival if:

$$(q\pi(r_i) - \tau d) + (1-q)[q\pi(r) - \tau(1/n - d)] \geq (q\pi(r) - \tau(1/n - d)) + (1-q)[q\pi(r_i) - \tau d] \quad (2.11)$$

This 'incentive compatability constraint' requires some explanation. The first term on the left-hand side is the expected return for a borrower if accepted by bank i during the first round of screening, with probability q at interest rate r_i and located at a distance d. The second term on the left-hand side is the 'forward-looking component' which consists of the expected return to the borrower if rejected by bank i with probability $1-q$ but accepted by the rival bank j in the second round of screening with probability q at interest rate r and located at a distance $1/n - d$. A similar interpretation applies to the right-hand side of the equation, but in reverse: the first term is the expected return for a first-time applicant that approaches bank j and is accepted and the second term is the expected return if the applicant is rejected by bank j but is successful in obtaining a loan from bank i.

We can also derive the equivalent incentive compatability constraint for a bad borrower:

$$((1-q)B - \tau d) + q[(1-q)B - \tau(1/n - d)] \geq (1-q)B - \tau(1/n - d) + q[(1-q)B - \tau d] \quad (2.12)$$

This equation has essentially the same interpretation as that of the good borrower, except that the probabilities are reversed and the term for a good borrower's profit, $\pi(r_i)$, is replaced by that of a bad borrower, B. By assuming that the incentive compatability constraints hold for both types of borrowers, I can then determine the equilibrium location of the marginal good and bad borrowers (i.e. the good and bad borrowers that are indifferent between approaching bank i and the nearest rival, bank j).

$$d_G^* \leq 1/2n + q(1/2\tau)(\pi(r_i) - \pi(r)) \tag{2.13}$$

$$d_B^* \leq 1/2n \tag{2.14}$$

These conditions ensure that both good and bad borrowers are, at worst, indifferent between bank i and bank j. Given the equilibrium location of both types of borrowers, I can then determine each bank's market share of initial loan applicants.

Bank i's market share of initial good loan applicants is given by:[36]

$$s_{iG} = 2d_G^* = 1/n + q(1/\tau)(\pi(r_i) - \pi(r)) \tag{2.15}$$

Bank i's market share of initial bad loan applicants is given by:

$$s_{iB} = 2d_B^* = 1/n \tag{2.16}$$

Aggregate mortgage loan demand faced by each bank will also depend on demand for mortgages from the backlog of rejected applicants. The backlog of rejected applicants that will approach bank i during the second round of screening comprises borrowers that have been rejected by the two nearest rival banks (on either side of bank i) *and* are sufficiently close to bank i. Recall that the maximum lending distance to a given borrower is $1/n$, so there are $2/n$ good borrowers in total that are potentially within bank i's geographic reach. But we have already established that a proportion s_{iG} of good borrowers will approach bank i for first-round screening, so the remaining $2/n - s_{iG}$ good borrowers must be screened by the two rival banks. Given the imperfect screening technology, the rival banks will incorrectly

[36]Borrowers are located at a distance d on both sides of bank i so to obtain bank i's market share I need to multiply by 2 each of the above conditions.

classify a share $(1-q)$ of these good applicants as bad borrowers. Under the maintained assumption that rejected applicants seek to reapply rather than exit the market, the backlog of good borrowers that will approach bank i for second round screening is therefore given by the share $(1-q)(2/n - s_{iG})$.

In contrast, the backlog of rejected bad loan applicants will consist of those correctly rejected by the two closest rival banks. Following the same logic as before, there are $2/n$ bad borrowers potentially within reach of bank i but bank i will screen $s_{iB} = 1/n$ bad borrowers during the first round. This leaves $1/n$ bad borrowers to approach neighbouring banks and a share q of these bad borrowers will be correctly rejected by these banks. Bank i will therefore face a total of q/n rejected bad borrowers in the second round of screening under the assumption that rejected applicants reapply rather than exit.

Now I can determine the post-screening level of loan demand faced by each bank. The loan demand of applicants consists of the demand from first-time applicants, as well as the backlog of rejected applicants. Recall that there are N borrowers in the total population and a share λ of these borrowers are known to be good borrowers. The bank will correctly screen a proportion q of these good borrowers. Bank i's post-screening loan demand from good applicants is therefore given by:

$$L_{iG} = qN\lambda[s_{iG} + (1-q)(2/n - s_{iG})] \tag{2.17}$$

The first term in the squared brackets is bank i's market share of initial borrowers and the second term is bank i's market share of rejected borrowers. Using the same logic, bank i's post-screening loan demand from bad applicants is:

$$L_{iB} = (1-q)N(1-\lambda)[s_{iB} + q(2/n - s_{iB})] = N(1-\lambda)[(1+q)/n] \tag{2.18}$$

A.5 Symmetric Nash equilibrium

In the symmetric Nash equilibrium, bank i chooses its lending interest rate, r_i to maximise expected profits ($\pi(r_i)$) taking as given the interest rate choices of all its competitors and the post-screening loan demands of both good and bad borrowers (which includes both initial and rejected applicants):

$$\pi_i = r_i L_{iG} - \rho(L_{iG} + L_{iB}) \tag{2.19}$$

Note that a bank only earns revenue by lending to good borrowers (as bad borrowers do not repay their loans) but funds loans to both good and bad borrowers. The first-order condition from this optimisation problem is:

$$\frac{d\pi_i}{dr_i} = (r_i - \rho)\frac{dL_{iG}}{dr_i} + L_{iG} - \rho\frac{dL_{iB}}{dr_i} = 0$$

Using this FOC and imposing symmetry across banks ($r_i = r$), the equilibrium interest rate spread, $r^* - \rho$, is given by:

$$r^* - \rho = \frac{\tau(2-q)}{nq^2} \tag{2.20}$$

Consistent with standard product differentiation models, a rise in transporation costs (τ) leads to higher equilibrium lending rate spreads. This is because higher transportation costs imply more product differentiation, and so more market power for banks which, in turn, implies higher loan margins. Note that if there are no transportation costs ($\tau = 0$) then the loan rate is equal to the risk-free rate and the model collapses to the case of perfect competition. A rise in the precision of the credit signal (q) also implies lower equilibrium spreads. A more precise signal implies that the bank is better informed about the borrower so there is less risk in lending. In other words, more distant borrowers will face higher lending rate spreads, all other things being equal. This gives me the first testable implication of the model:

Proposition 1: *The loan interest rate spread $(r - \rho)$ is increasing in lending distance (d).*

My loan-level dataset has limited information on interest rates so I only consider the tests of this proposition in the Appendix. But, based on the model, I can also derive the equilibrium levels of lending to both good (L^*_{iG}) and bad (L^*_{iB}) borrowers, and therefore the equilibrium aggregate level of lending, $L^*_i = (L^*_{iG} + L^*_{iB})$:

$$L^*_{iG} = q(2-q)\lambda\frac{N}{n}$$

$$L^*_{iB} = (1-q^2)(1-\lambda)\frac{N}{n}$$

$$L^*_i = [q(2\lambda - q) + (1-\lambda)]\frac{N}{n}$$

Finally, I can derive bank i's equilibrium share of delinquent loans (NPL^*_i) conditional on the quality of the credit signal. By assumption, all bad borrowers default on their debt while all good borrowers repay, so the share of bad loans is equivalent to the share of loans made to bad borrowers. The equilibrium share of bad loans is therefore given by:

$$NPL^*_i = \frac{L^*_{iB}}{(L^*_{iG} + L^*_{iB})} = \frac{(1-q^2)(1-\lambda)}{(1-\lambda) + q(2\lambda - q)} \tag{2.21}$$

Using equation 2.21 we can ascertain that the share of delinquent loans is inversely related to the precision in the signal of credit quality:

$$\frac{dNPL^*_i}{dq} = \frac{-2\lambda(1-\lambda)[(q-1)^2 + q]}{[(1-\lambda) + q(2\lambda - q)]^2} \leq 0$$

Given the inverse relationship between distance (d) and the reliability of the credit signal

(q), we have the main hypothesis of this paper:

Proposition 2: *A bank's share of delinquent loans ($NPLs$) is increasing in lending distance (d)*

There are two reasons for the positive link between borrower-lender distance and the share of delinquent mortgages. Firstly, there is a direct (supply-side) effect – as the borrower gets further away from the bank the signal of credit quality deteriorates, resulting in a greater chance of the bank approving bad credit risks. Secondly, there is an indirect (demand-side) effect – because the market size is geographically fixed, the borrower effectively moves closer to other rival banks as the distance from any given bank rises, so that the rival banks become more likely to correctly identify good and bad applicants. As a result, the bank will become more likely to be approached by bad applicants that were rejected by the rival banks, which is the "winner's curse effect". Therefore, based on the model, mortgage loans that are made to more distant borrowers are more likely to default than loans made to nearby borrowers. The model therefore predicts that the geographic expansion of US banks over the past couple of decades may have contributed to a higher share of bad loans, all other things being equal.[37]

This simple model predicts that credit quality will be low for mortgage lenders that lend at long distances. However, the model ignores the possibility that local lenders may have a comparative advantage in lending to high-risk borrowers. There is empirical evidence for this in the international finance literature (e.g. Mian [2006]). In other words, theory could be ambiguous with regards to the link between credit quality and geographic distance. For this reason, in the main part of the paper, I focus on the empirical evidence as it relates to the US residential mortgage market.

[37]Note that I do not allow banks to optimally choose their screening level. Nevertheless, I have considered a model in which banks can choose how well they screen – the implications of this model are qualitatively the same, but it is more difficult to obtain a closed-form solution for the share of bad loans.

B Lending Distance, Bank Risk-Taking and Interest Rates

In this Appendix I explore how lending distance is related to bank credit pricing using the loan-level data. Specifically, I test proposition 1 derived from the theoretical model that the loan interest rate spread is increasing in lending distance. The highly disaggregated nature of the loan-level data allows for formal testing of the distance-credit pricing relationship as *both* the interest rate spread and lending distance are observed at the loan level. I can therefore isolate unobservable credit demand shocks from supply shocks through the use of separate region and bank fixed effects. However, there are two caveats to this analysis as I only have information on loan pricing for a subset of my sample period (2004 to 2010) and for a subset of (high-risk) loans. I estimate the following equation:

$$R_{bct} = D'_{bct}\beta + X'_{bct}\gamma + \theta_{bt} + \theta_{ct} + \epsilon_{bct} \tag{2.22}$$

Where R_{bct} is the interest rate spread for bank b to the average borrower in county c at time t, D_{bct} is the borrower-lender distance and X_{bct} is a vector of observable market characteristics that vary by bank, county and time, such as the income of loan applicants. I also control for unobservable (time-varying) lender and borrower characteristics by including a full set of bank-time (θ_{bt}) and county-time dummies (θ_{ct}). The inclusion of county-time fixed effects is important as they control for time-varying credit demand shocks, such as changes in local housing prices.

To illustrate the effect of controlling for the various unobservable shocks, I present the results of estimating the model with simple OLS (column 1), with county-specific time trends (column 2), with bank-specific time trends (column 3) and with both county-specific and bank-specific time trends (column 4). I cluster the standard errors at *both* the bank and county levels. The results are shown below.

Consistent with the theoretical model, the loan interest rate spread is positively correlated

Table 2.7: The Effect of Distance on the Interest Rate Spread

	(1) OLS	(2) County FE	(3) Bank FE	(4) County & Bank FE
Borrower-HQ distance	0.166***	0.143***	-0.00304	0.0303***
	(7.66)	(6.26)	(-0.35)	(3.74)
Household income	-0.395***	-0.447***	-0.0117	-0.0569***
	(-6.08)	(-4.95)	(-0.46)	(-4.60)
Constant	5.448***	5.806***	4.871***	-0.0377***
	(18.82)	(15.21)	(37.06)	(-19.52)
R^2	0.057	0.150	0.571	0.525
Observations	430079	430079	430079	430079

t statistics in parentheses

Notes: Standard errors are clustered at the bank and county levels.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

with lending distance in most specifications. Based on the model that includes both bank and county time trends (column 4) I find that a one per cent increase in lending distance is associated with an increase in the interest rate spread of around 3 basis points. This effect is small but statistically significant. Notably, the explanatory power of the model greatly improves when I include the bank time trends which suggests that most of the variation in interest rates is bank-specific rather than county-specific. I also find that higher-income loan applicants are offered lower interest rate spreads, presumably as household income is positively correlated with creditworthiness.

The positive link between the interest rate spread and distance indicates that long-distance loans are relatively more risky from the lender's perspective. The finding of a positive link is consistent with other studies on US syndicated business lending (e.g. Knyazeva and Knyazeva [2012]) but contradicts the observed negative relationship found in some US small business lending studies (e.g. Petersen and Rajan [2002], Agarwal and Hauswald [2010]). The studies that observe an inverse relationship point to this as evidence that banks engage in spatial price discrimination; banks can effectively 'lock in' local borrowers and charge them relatively high interest rates if there are significant search costs for borrowers in applying for loans. However, spatial price discrimination is probably less likely in US home mortgage lending – the large number of mortgage lenders and the widespread availability of online loan application technology means that there are relatively low search costs. As such, we are probably less likely to observe a negative correlation between the interest rate and distance in home mortgage lending.

C Regression Model Variables

The table below provides a detailed outline of the variables used to estimate the regression models.

Table 2.8: Description of Model Variables

Variable	Source	Description
Mortgage delinquency (or non-performing loan) ratio	Statistics on Depository Institutions (SDI)	Total loans secured by 1-4 family residential properties past due 90+ days (p9reres) + total loans secured by 1-4 family residential properties no longer accruing interest (narreres) / Total loans secured by 1-4 family residential properties (lnreres)
HQ-borrower distance	Home Mortgage Disclosure Act (HMDA)	(Log of) average distance between location of lender's headquarters and borrower's Census tract
Branch-borrower distance	Home Mortgage Disclosure Act (HMDA)	(Log of) average distance between location of each lender's domestic branch and borrower's Census tract
HQ-branch distance	Home Mortgage Disclosure Act (HMDA)	(Log of) average distance between location of lender's headquarters and location of each lender's domestic branch
Size	SDI	(Log of) total assets (asset)
No. of branches	SDI	(Log of) total domestic branches (including headquarters) of the lender (offdom)
Home mortgage share	SDI	Total loans secured by 1-4 family residential properties (lnreres) / Total loans (less unearned income and loan loss allowances) (lnlsnet)
Home loan growth	SDI	Annual change in the (log of) total loans secured by 1-4 family residential properties (lnreres)
Risk-weighted asset ratio	SDI	Total risk-weighted assets (rwaj) / Total assets (asset)
Deposit share	SDI	Total deposits (deposit) / Total liabilities (liab)
Net interest margin	SDI	(Total interest income (intinc) - total interest expense (eintexp))/ Average earning assets (ernast)
Jumbo mortgage share	HMDA	Total non-conforming conventional 1-4 family residential property loans / Total conventional 1-4 family residential property loans
Securitisation share	HMDA	Sales of conventional 1-4 family residential property loans / Total conventional 1-4 family residential property loans

Chapter 3

Liquidity Shocks and Mortgage Credit Supply

There is extensive anecdotal evidence to suggest that a significant tightening in credit conditions, or a "credit crunch", occurred in the US following the collapse of the loan securitisation market in 2007. However, there has been surprisingly little formal testing for the existence of a credit crunch in the context of the US housing market.

In this chapter I examine the evidence for the existence of a credit crunch; namely, that the fall in mortgage credit over 2007-08 was caused by a reduction in credit supply which, in turn, can be traced to a fall in the amount of financing available to mortgage lenders. I use the differential exposures of individual mortgage lenders to the collapse of the securitisation market in 2007 as a source of cross-lender variation in bank financing conditions and assess the impact on residential mortgage lending.

Using loan-level information to control for unobservable credit demand shocks, I show that mortgage lenders that were particularly reliant on loan securitisation disproportionately reduced the supply of mortgage credit. The negative liquidity shock caused by the shutdown of the securitisation market explains a relatively large share of the total decline in new mortgage credit during the crisis. The result is robust to a battery of tests, and holds at both the aggregate and disaggregate level.

3.1 Introduction

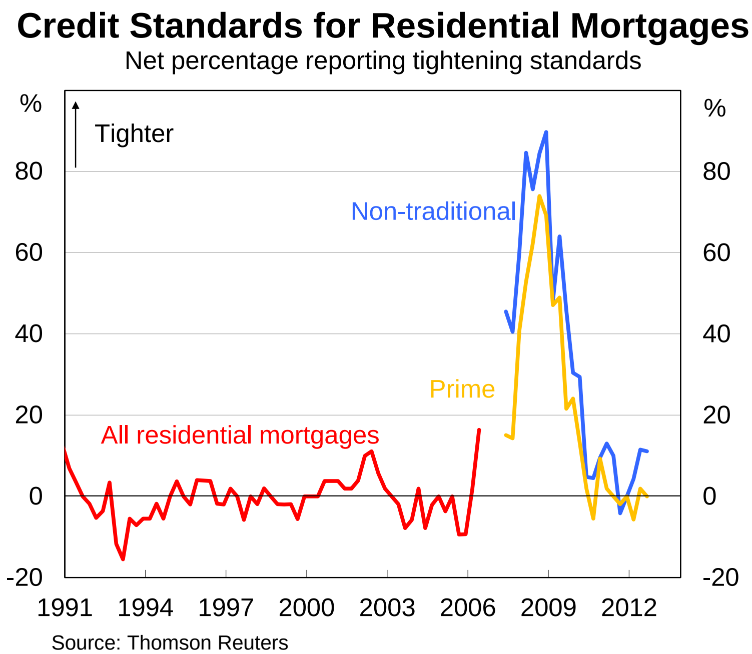
There is extensive anecdotal evidence to suggest that a significant tightening in credit conditions, or a "credit crunch", occurred in the US housing market following the collapse of the loan securitisation market in late 2007. For example, the Federal Reserve's Senior Loan Officer Survey indicates that the number of US banks that tightened lending standards rose sharply for both prime and subprime mortgages in the December quarter 2007 (Figure 3.1). The term "credit crunch" has become so commonplace that the Economist magazine has created a Credit Crunch board game and the term is now officially part of the English language, having been recently included in the Concise Oxford English Dictionary.

Despite its popularity as a concept, there has been surprisingly little formal testing of whether a credit crunch, in fact, occurred in the US housing market in 2007-08. This is probably because the necessary conditions to test the hypothesis are quite strict. The most generally accepted definition of a credit crunch is attributable to Bernanke and Lown [1991] who define it as a "general reduction in the supply of credit, holding constant both the risk-free interest rate and the quality of potential borrowers". This definition requires two key conditions to be satisfied to establish a credit crunch: first, there should be a fall in credit which is caused by a decline in credit supply rather than demand, and, second, the fall in credit supply must be exogenous in the sense that it is not caused by an increase in the credit risk of potential borrowers. In other words, the fall in credit supply will generally be caused by factors on the liability side of financial institutions' balance sheets, such as a tightening in financing conditions.[38]

The other reason why there has been little formal testing for a credit crunch is that it is difficult to conduct adequate econometric tests. There are two common econometric problems in identifying a credit crunch. Firstly, a crunch typically coincides with a general decline in economic activity, which also causes the demand for credit to fall (simultaneity

[38]It is effectively a leftward shift of the credit supply curve where the quantity of credit is measured on the x-axis and the loan interest spread is measured on the y-axis.

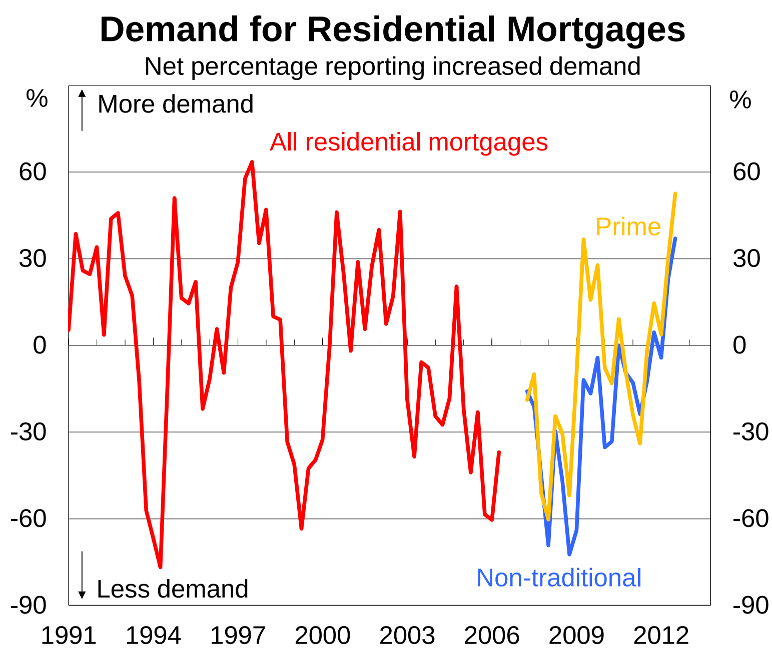
Figure 3.1: Credit Standards for US Residential Mortgages



bias). Secondly, even if the decline in credit can be traced to a fall in supply, this may be an endogenous response by lenders to a decline in the quality of potential borrowers associated with the economic downturn (selection bias). The difficulty in separately identifying the effects of changes in credit supply and demand is highlighted by the Federal Reserve's Senior Loan Officer Survey, which shows that the demand for mortgage credit also fell sharply around 2007-08 (Figure 3.2). It is therefore possible that this decline in demand drove the overall fall in lending rather than a decrease in credit supply.

This paper tests whether the fall in mortgage credit over 2007-08 was caused by a reduction in credit supply which, in turn, can be traced to a fall in the level of financing available to US mortgage lenders, which caused them to become liquidity constrained. I will refer to this as the 'liquidity constraints hypothesis'. I use application-level information on new mortgage loans to assess how US banks' lending behaviour changed as a result of the tightening in financing conditions. The analysis is divided into two parts. In the first part, I estimate a

Figure 3.2: US Residential Mortgage Demand



model that identifies the liquidity shock based on each bank's reliance on securitised lending in the period prior to 2007. In this framework the closure of the securitisation market acts as the "treatment" with the "treatment group" being the mortgage lenders that were reliant on securitisation before 2007 and the "control group" being the mortgage lenders that were not dependent on securitisation. For expositional purposes, I will refer to the treated lenders (those reliant on securitised lending) as the originate-to-distribute (or OTD) lenders, and the control group of lenders, that were not dependent on securitised lending, as the non-OTD lenders.

The novel aspect of this study is that the causal effect of the liquidity shock on mortgage lending is identified through variation in the lending policies of OTD and non-OTD lenders that grant credit to the *same* borrower (where a particular region is broadly defined as a "borrower"). Specifically, I assume that mortgage lenders that originate loans in the same Census tract (a tract is similar to a postcode) face the same demand conditions and the same risk profile of loan applicants. Under this assumption, a reduction in credit by OTD lenders relative to non-OTD lenders *within a tract* implies that a negative bank balance sheet shock, and hence a decline in credit supply, caused the overall fall in mortgage credit.

I find strong evidence to indicate that the OTD lenders disproportionately reduced mortgage credit supply following the liquidity shock. The negative liquidity shock caused by the shutdown of the securitisation market explains about 14 per cent of the average decline in mortgage credit during the crisis. Moreover, I find that the link between bank liquidity and mortgage lending holds even after controlling for unobservable lender characteristics, such as changes in bank risk assessment.

I also examine which borrowers were most affected by the reduction in credit supply due to the tightening in lender financing conditions. Theory suggests that lenders re-balance their portfolios towards less risky loans when economic conditions deteriorate (Bernanke et al. [1996]). During recessions, the share of credit flowing to borrowers with more severe asymmetric information and agency problems, such as small firms, decreases. This "flight to quality" has been identified in a range of empirical studies (Lang and Nakamura [1995], Popov and Udell [2010]). However, more recent research indicates that there may also be a "flight to home" effect when economic conditions deteriorate (Giannetti and Laeven [2012], Haas and Horen [2012]). Specifically, lenders re-balance their asset portfolios towards local borrowers when the economy weakens. The flight home effect co-exists with, but is distinct from, the flight to quality effect. To the best of my knowledge, this paper is the first to examine whether the flight to quality and flight to home effects are relevant to residential mortgage lending.

I find that all mortgage lenders reduced credit to risky borrowers, though the effect was not disproportionately larger for the OTD lenders. This points to a general "flight to quality" by US mortgage lenders during the crisis. I find limited evidence for a "flight to home" effect; while the OTD lenders increased the share of credit to local borrowers relative to the non-OTD lenders, the differential effect is not significant.

The rest of this paper is organised as follows. Section 2 provides some institutional background regarding the US securitisation market and discusses the business models of US mortgage lenders. Section 3 discusses the data to be used in estimating the difference-in-differences model. Section 4 provides an outline of the identification strategy. Sections 5 and 6 discuss the results, section 7 outlines some robustness tests and section 8 concludes.

3.2 Institutional Background

Mortgage securitisation refers to the process of pooling mortgages into securities that are sold to investors on the secondary market. The securities are backed by the cash flow generated by the borrowers' mortgage payments. Securitisation is essentially a process that allows a loan originator to transform cash flows from a pool of non-tradable assets into tradable debt instruments. In doing so, securitisation provides financial institutions with an additional method of financing mortgages – in this case through the issuance of mortgage-backed bonds rather than unsecured bonds or deposits.

Securitised bonds backed by home mortgages are known as 'residential mortgage-backed securities' (RMBS). In the US, the RMBS market can be divided into two sectors: agency and non-agency (or private-label) RMBS. The agency market includes mortgages securitised by the government-sponsored enterprises, such as the Federal National Mortgage Association (Fannie Mae) and the Federal Home Loan Mortgage Corporation (Freddie Mac). The government-sponsored enterprises (or GSEs) have traditionally been private corporations with a public charter, operating with the implicit backing of the US government. Historically, the government agencies have purchased residential mortgage loans on the secondary market from loan originators (e.g. banks) and then packaged these loans into securities which they either sell to other investors or hold in their own portfolios. In this paper I will sometimes refer to the loans securitised by the government agencies as 'public securitisations'. In contrast, the non-agency market comprises mortgages securitised by private financial institutions, such as commercial and investment banks. I will refer to these securitised loans as 'private securitisations'.

Loans that are securitised by the GSEs must meet certain eligibility criteria, based on factors such as loan size and other underwriting guidelines. Residential mortgages that are eligible to be purchased by the GSEs are known as 'conforming mortgages'. Mortgages that are non-conforming because their size exceeds the purchasing limit are known as 'jumbo' mortgages. Mortgages that are non-conforming because they do not meet other underwriting guidelines, such as credit quality or loan-to-value ratio, are often called 'subprime' mortgages.

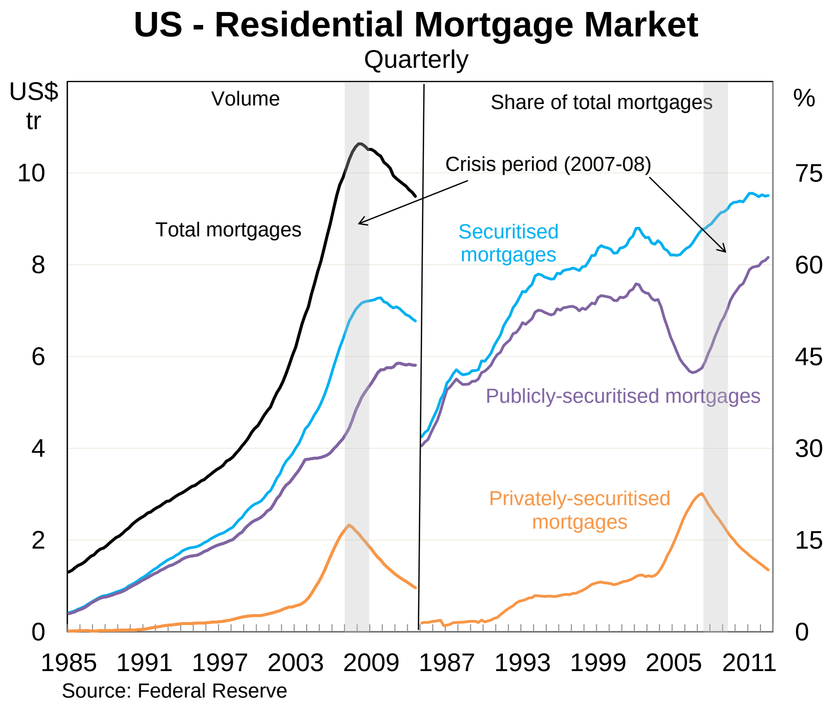
The private securitisation market developed to facilitate the sale of mortgages that did not meet the government agencies' eligibility criteria. This included jumbo mortgages that the government agencies were not allowed to purchase and non-traditional mortgages that may have satisfied the conforming loan limit but did not meet other GSE underwriting criteria, or for which private investors were willing to offer a higher price. [39]

According to US flow of funds data, around 60 per cent of residential mortgages have been securitised on average over the past two decades (Figure 3.3). The bulk of these loans have been securitised by the government agencies. However, the most recent housing cycle in the US caused significant changes in the composition of US securitised home mortgage debt. The share of mortgages that were privately securitised rose rapidly around 2004-2006, coinciding with the 'hottest' period in the US housing boom (Figure 3.3). The increase in the share of private securitisations is likely to reflect several factors, such as an increase in demand for non-conforming mortgages by borrowing households, an increase in demand for non-conforming mortgage securities by private investors, and a relaxation of lending standards by mortgage originators (Nadauld and Sherlund [2009]).

The share of private securitisations fell dramatically when the subprime mortgage market collapsed in the first half of 2007. The private securitisation market was forced to essentially

[39] Historically, the conforming mortgage loan limit has been adjusted periodically in line with changes in average US home prices. Higher limits apply for mortgages secured by homes that are: (i) located in high-cost housing areas, (ii) multifamily dwellings, and (iii) located in Alaska, Hawaii, Guam and the US Virgin Islands.

Figure 3.3: US Residential Mortgage Market



shut down by late 2007. In contrast, the public securitisation market continued to function due to implicit government backing and, eventually, interventions by the Federal Reserve and US Treasury. This substitution away from private securitised lending to public securitised lending during the crisis suggests that the government agencies were partly able to step into the breach caused by the evaporation of the private-label market. This substitution could be important in identifying the effect of a liquidity shock on lending and will be discussed in a robustness test later.

Home mortgage lenders that are reliant on loan securitisation to fund the origination of new loans are often referred to as 'originate-to-distribute' (or OTD) lenders (Purnanandam [2010]). The mortgage lenders that, instead, rely on other forms of funding, such as retail deposits, to originate loans are known as 'originate-to-hold' (or non-OTD) lenders. The non-OTD lenders keep, rather than sell, most of the loans on their balance sheets. This distinction between the two groups of lenders – the OTD and non-OTD lenders – is important in this study. I assume the OTD lenders were more affected by the disruption to the securitisation market in 2007 than the non-OTD lenders. The distinction, therefore, provides a way in which to identify the effect of the liquidity shock stemming from the securitisation market. As the OTD lenders were highly dependent on securitisation to finance new lending, these lenders would have become relatively more liquidity constrained when investors withdrew funding from the secondary market.

3.3 Literature Review

In examining the relationship between bank liquidity and lending, this paper relates to several branches of the macroeconomic literature. The theoretical literature provides a framework in which banks' financing conditions can affect overall lending due to credit market imperfections (Bernanke and Blinder [1988], Holmstrom and Tirole [1997], Stein [1998]). But empirical studies face a challenge in tracing the channels through which credit supply shocks are transmitted. Traditionally, empirical research has relied on either time-series or cross-sectional variation in the balance sheet positions of banks to identify the effect of bank financing conditions on lending.[40] For example, Peek and Rosengren [1995] use US bank-level data to document a positive relationship between bank capital and credit growth during the 1990-91 recession. However, the evidence is not compelling, as banks that face more creditworthy borrowers will likely experience fewer loan losses. The lower losses could translate into higher levels of capital and may also encourage more lending. In other words, endogeneity could be a problem as differences in bank-level credit growth may reflect differences in the risk profile of borrowers or other demand conditions. Other studies that use instrumental variables (Paravisini [2008]) or natural experiments (Peek and Rosengren

[40] There is a close analog to the literature on the credit channel of monetary policy. The credit channel of monetary policy can be divided into two channels – the "bank lending channel" and the "borrower balance sheet channel". The bank lending channel measures the effect of monetary policy shocks on the real economy through their effect on the balance sheets of lenders. In contrast, the borrower balance sheet channel measures the effect of monetary policy shocks on the real economy through their effect on borrower balance sheets. The effect of liquidity shocks on lending is sometimes loosely referred to as "a bank lending channel" in the literature, despite the fact that neither monetary policy nor the real economy are considered.

[2000]) generally provide more compelling evidence that exogenous (to demand) liquidity supply shocks affect lending. For instance, Peek and Rosengren [2000] demonstrate that US subsidiaries of Japanese banks were more likely than domestic US banks to cut credit to the US commercial real estate sector following a negative balance sheet shock to their Japanese parent. As the shock stems from overseas, it is likely to be exogenous to demand conditions in the US. However, it is still possible that the Japanese subsidiaries and the domestic US banks lend to different pools of borrowers within the US commercial property sector, so that differences in demand conditions across banks could still drive the results.

My paper belongs to a growing literature that uses loan-level information to identify the causal effect of credit supply shocks. The increasing availability of loan-level data has allowed researchers to implement more sophisticated identification strategies than empirical studies that rely on either aggregate or bank-level data. The seminal paper in this branch of the literature is Khwaja and Mian [2008]. They examine the impact of liquidity shocks on bank lending by exploiting cross-bank liquidity variation induced by unanticipated nuclear tests in Pakistan in 1998. The nuclear tests caused the Pakistani government (in anticipation of balance of payment problems) to restrict withdrawals of dollar-denominated deposit accounts to local currency only, and at an unfavourable exchange rate. The collapse of the dollar-denominated deposit market disproportionately affected banks that relied more on dollar-denominated deposits for liquidity. They show that, for the same firm borrowing from two different banks, the bank exposed to the larger decline in liquidity was more likely to reduce lending. To the extent that the within-firm comparison fully absorbs firm-specific changes in credit demand, the estimated difference in loan growth between banks can be attributed to differences in bank liquidity shocks. This "within borrower" identification scheme has now been adopted in a range of empirical studies that have access to loan-level information (Albertazzi and Marchetti [2010], Iyer et al. [2010], Jimenez et al. [2011], Schnabl [2012], Cetorelli and Goldberg [2012]). To the best of my knowledge, this is the first study to apply this 'within-borrower' identification strategy to household lending.

The US housing credit market provides a natural testing ground to examine the impact of credit supply shocks cause it was the market at the epicentre of the global financial crisis. Most of the existing research on the current crisis has looked at the effect of changes in credit supply on the investment behaviour of large corporate borrowers (e.g. Ivashina and Scharfstein [2010], Duchin et al. [2010], Campello et al. [2012]). But this is unlikely to be the primary channel through which the financial crisis affected the real economy. Instead, the prolonged period of weak economic conditions in the US economy was more likely due to developments in residential mortgage finance.

There are a few other recent papers that similarly treat the shutdown of the securitisation market in 2007 as a negative liquidity shock and examine how this affected bank lending (e.g. Goetz and Gozzi [2010], Calem et al. [2011], Dagher and Kazimov [2012]). However, my paper covers a wider cross-section of lenders, a longer time series, and utilises loan-level information which allows me to better control for variation in the distribution of borrowers across banks.

Goetz and Gozzi [2010] focus on small banks that lend within their own local markets while I examine the behaviour of all lenders, regardless of size, location or geographic reach. Restricting the sample to small local banks is likely to bias the causal effect of bank liquidity shocks for two reasons. Firstly, if affected borrowers are able to switch to large banks when small banks cut off their funding then restricting the sample to only small banks may overstate the true aggregate effect of the shock. Secondly, if liquidity-constrained banks are relatively more likely to cut lending to non-local borrowers, then the focus on local lending could understate the true effect of a credit supply shock.

Calem et al. [2011] rely on bank-level variation in liquidity and lending and hence only control for unobservable variation in the characteristics of each bank's average borrower. In contrast, I control for changes in the distribution of each bank's (unobservable) borrower characteristics. This will be important if the effect of the supply shock varies across different borrowers within a bank's loan portfolio.

Dagher and Kazimov [2012] similarly treat the shutdown of the securitisation market as an exogenous negative liquidity shock, but use each bank's share of non-deposit funding, rather than the share of securitisation funding, to identify the treatment group of mortgage lenders. Moreover, their identification strategy assumes that loan applicants that reside within the same metropolitan statistical area share similar characteristics, whereas my strategy assumes that applicants residing in the same Census tract are similar, which is more likely to be true given the tracts are defined based on residents sharing similar characteristics. And, unlike my study, Dagher and Kazimov [2012] do not consider which borrowers were most affected by the credit supply shock.

The housing market is also a potentially interesting area in which to identify any home bias in lending because the location of the asset (the home) is a fundamental determinant of its price (and hence its collateral risk). Geographic location is therefore potentially a significant determinant of credit risk in home mortgage lending. There is an extensive literature identifying a home bias in the global allocation of capital (Coeurdacier and Rey [2011]) but, to the best of my knowledge, there is little research on home bias in residential mortgage lending. Moreover, only recently has evidence emerged that this home bias increases when economic conditions worsen (i.e. that there is a "flight to home" effect). For example, in response to an adverse shock to financing conditions, international banks in the syndicated lending market shifted their lending activity towards their home country, regardless of the perceived risk of the borrowers (Giannetti and Laeven [2012], Haas and Horen [2012]). Broadly speaking, there are two possible explanations for a home bias in credit markets – information asymmetries and behavioural biases. If lenders cannot perfectly observe borrower risk and it is costly to collect information on the creditworthiness of borrowers then lenders may be better informed about local borrowers than non-local borrowers. Under this explanation, geographic distance is essentially a proxy for credit risk, and, similar to the flight to quality, lenders will re-balance their portfolios towards local borrowers when economic conditions deteriorate. Alternatively, certain lenders may specialise in lending to distant borrowers and have more sophisticated loan screening and monitoring technologies than local lenders. In this case, local lenders would not have an informational advantage, so any home bias may be better explained by a behavioural bias towards familiar assets rather than by information asymmetries (see Coeurdacier and Rey [2011] for a detailed overview of the competing theories behind the home bias).

3.4 Data

3.4.1 The Home Mortgage Disclosure Act

The data underpinning the regression analysis are derived from the Home Mortgage Disclosure Act (HMDA) Loan Application Registry. Enacted by Congress in 1975, HMDA requires mortgage lenders located in metropolitan areas to collect data about their housing-related lending activity and make these data publicly available. The HMDA dataset is generally considered to be the most comprehensive source of mortgage data in the US, and covers about 80 per cent of all home loans nationwide and a higher share of loans originated in metropolitan statistical areas. Whether a lender is covered depends on its size, the extent of its activity in a metropolitan statistical area, and the weight of residential mortgage lending in its portfolio.

The underlying sample of mortgage loan applications includes almost 300 million annual observations covering the period 2000-2010. For each application there is information on the loan application (e.g. the type of loan, the size of the loan, whether the loan is approved or not), the borrower (e.g. income, race, ethnicity), and the lending institution (e.g. the identity of the lender).[41] Most importantly for the purposes of this paper, I can identify whether a loan is sold to another financial institution or not. I assume that loan sales and loan securitisations are equivalent, so that I can directly observe the extent of securitisation

[41] I restrict the sample to conventional owner-occupier one-to-four family residential mortgages, which is consistent with numerous other studies.

activity by each mortgage lender.[42] I can also identify the type of institution that purchased the loan, which allows me to split loan securitisations into private and public securitisations. In particular, I identify 'public' securitisations as any loans that were sold to the four government agencies – Fannie Mae, Ginnie Mae, Freddie Mac and Farmer Mac. I classify the remaining loan sales as 'private' securitisations.

The raw loan application data are not panel data as the behaviour of specific borrowers cannot be tracked over time, although a given lender can be observed each year. To create a pseudo-panel I aggregate the annual loan application data so that the data vary by lender and Census tract. This means that I track the lending of a given loan originator to the *average* borrower in a given Census tract across time. A Census tract is a very narrowly-defined geographic region. The tracts are designed, for the purpose of taking the Census, to be relatively homogeneous units in terms of population characteristics, economic status, and living conditions. In the US, there are about $73,000$ Census tracts, with each tract having between $2,500$ and $8,000$ residents. Several tracts commonly exist within a county, with the boundaries of a tract usually coinciding with the limits of cities and towns. The very narrow geographic focus of Census tracts supports my identification strategy, as different borrowers in the same tract are likely to share similar characteristics. This ensures that two different lenders that originate loans in the same tract are likely to face the same demand conditions and borrower risk profiles.

The HMDA dataset covers bank and non-bank lenders (i.e. mortgage companies). The non-bank lenders are an important segment of the US mortgage market. Over the sample period, they originated more than half of all new residential mortgage loans. Moreover, with no access to depository funding, non-bank lenders typically operate under the originate-to-distribute (OTD) business model and hence are much more reliant on loan securitisation for funding. The market share of these lenders typically varies with the credit cycle, so

[42]Loan sales and securitisations are closely related concepts. A loan sale involves the lender selling the loan in its entirety to another institution. If that institution wants to sell it again, they have to find a buyer and negotiate a price. In contrast, a loan securitisation involves the lender selling a loan (or portfolio of loans) to investors where the loan (or portfolio) is converted into rated securities, which are publicly traded.

including these non-bank lenders in the sample reduces the probability of sample selection bias in identifying the effect of financing conditions on credit supply.

3.4.2 Measuring mortgage lending and bank liquidity

As will be discussed in the next section, in the main regression model, the dependent variable is a measure of the change in lending activity by each mortgage lender in each tract during the crisis. My preferred measure of lending activity is the number of new loans.[43] I proxy the liquidity shock through the average propensity of each mortgage lender to securitise loans in the pre-crisis period. For each lender and each year, I calculate the ratio of the number of new loans that are sold to the total number of new loans and then, for each bank, average across all the years of the pre-crisis period. This averaging process is partly aimed at transforming the flow of loan sales into an approximate measure of each lender's stock of loan sales in the pre-crisis period, as the stock determines each lender's exposure to the liquidity shock. I define the pre-crisis period to be 2000 to 2006. However, the results are not sensitive to the length of the pre-crisis period. For example, similar results are obtained when the pre-crisis period is defined as 2004 to 2006.

My set of control variables includes lender-tract controls, such as the average growth in income of the borrowing household and the share of minority household applicants[44] faced by each lender in each tract, as well as lender-level controls, such as the average (log) number of loan applications, which acts as a proxy for lender size. I exclude other lender-level variables, such as measures of profitability, as these data are unavailable for non-bank lenders. The non-bank lenders are an important segment of the US mortgage market and could be critical to the relationship between loan securitisation and credit supply given they typically operate under the OTD business model.

The second part of the analysis relies on a measure of borrower risk to test the flight to

[43] I have also estimated the regressions using alternative measures of lending activity, such as the total value of new loans and the share of applications that are approved. The results are qualitatively very similar.

[44] I define minority households as all African-American or hispanic households.

quality hypothesis. I measure the borrower risk faced by each lender through the share of high-priced loans originated by each lender in each tract. In 2004, information was added to the HMDA dataset on the interest rate spread.[45] Mortgages with a reported spread are "higher-priced" loans. As the interest rate spread on a loan largely reflects the credit risk of a borrower, the share of high-priced mortgages is often viewed as an indicator of risky or subprime lending (Mayer and Pence [2008]). This measure of risky lending is only available since 2004, so this necessarily restricts the time series available to establish the pre-treatment period in the second part of the analysis.[46]

To test the flight to home hypothesis I construct a measure of each bank's average lending distance based on the detailed information provided by the HMDA. The HMDA includes information on the Census tract of the residence of each loan applicant, as well as the full address details of the headquarters of each mortgage lender. This allows me to determine the geographic coordinates of each borrower and lender, and therefore estimate the geographic distance between each borrower and lender using geocoding software provided by STATA and Google Maps. I then calculate, for each lender in each year, the average distance across all its borrowers within a given Census tract, which provides a measure of "lending distance" at the lender-tract level.

The setup of the regression model implies that the sample is restricted to tracts in which there is at least one OTD lender and one non-OTD lender originating new loans. In other words, I implicitly exclude tracts in which there is only one type of lender. The final sample comprises about $5,000$ mortgage lenders that lend to more than $60,000$ tracts in the US. The table below provides summary statistics for the key variables used in the panel regression

[45] Higher-priced loans are those with an interest rate spread to the comparable-maturity Treasury for first-lien mortgages with an annual percentage rate (APR) three percentage points over the Treasury benchmark and for junior liens with an APR five percentage points over the benchmark. A lien is the legal claim of the lender upon the property for the purpose of securing debt repayments. The lien given the highest priority for repayment is the first lien; any other liens are junior liens. Because junior liens are less likely to be repaid, they are a higher risk to the lender than the first lien. Junior liens include home equity loans and home equity lines of credit.

[46] There are a couple of problems with using the share of high-priced loans as an indicator of risky or subprime lending. I talk about these issues in more detail in Appendix D.

Table 3.1: Variable Summary Statistics

Variable	Mean	Median	25th Pct.	75th Pct.	St Dev.
Pre-crisis, 2000-06					
Sale share (%)	36.1	14.3	0.0	80.2	40.5
Private sale share (%)	29.6	1.8	0.0	63.5	39.1
Public sale share (%)	6.5	0.0	0.0	0.4	17.6
Minority ratio (%)	12.0	0.0	0.0	13.3	22.7
No. of applications (log level)	4.7	4.7	3.4	6.0	2.2
Post-crisis, 2007-08					
No. of loans (% change)	-18.7	-2.4	-67.2	14.4	66.7
Household income (% change)	7.1	6.8	-17.6	31.1	48.6

Source: Home Mortgage Disclosure Act

modelling.

The summary statistics show that, on average, 36.1 per cent of all approved loans were sold in the pre-crisis period (2000-06) and that 12.0 per cent of household loan applicants were from a minority group.[47] Moreover, the number of new loans fell by 18.7 per cent over the crisis period (2007-08) while average household income rose by 7.1 per cent, on average.

3.5 Testing the liquidity constraints hypothesis

3.5.1 Identification

I estimate a difference-in-differences panel regression model to examine the causal effect of bank liquidity shocks on mortgage lending. The setup of the model is based on an experimental design in which there are two groups of lenders – a "treatment group" (the OTD lenders) and a "control group" (the non-OTD lenders) – as well as a "treatment" (the closure of the securitisation market in 2007). I first write the regression model in terms of

[47] The estimated share of loan sales in the HMDA (36 per cent) is significantly lower than the share of loan securitisations suggested by the US flow of funds (63 per cent) over the corresponding period. This is mainly because the flow of funds estimate is based on aggregate mortgage data while the HMDA estimate is based on bank-level mortgage data. The different estimates reflect the distribution of loan sales across lenders of different sizes. There is a large number of small banks in the US that sell few loans, which implies that the bank-level mean estimate is lower than the aggregate estimate. Aggregating the HMDA data to the national level, the share of loans sold is about 60 per cent, which is similar to the flow of funds estimate.

levels:

$$
\begin{aligned}
L_{ijt} = \alpha + SALESHARE_i'\beta + CRISIS_t'\gamma + SALESHARE_i * CRISIS_t'\rho \\
+X_{ijt}'\phi + \underbrace{\theta_j + \eta_{jt} + \epsilon_{ijt}}_{\nu_{ijt}}
\end{aligned}
\tag{3.1}
$$

Where the dependent variable is the (log) number of new loans of lender i to tract j in period t (L_{ijt}). The explanatory variables include the share of loans that were sold by lender i in the pre-crisis period ($SALESHARE_i$), a dummy variable for whether the period is pre or post-crisis ($CRISIS_t$) and an interaction variable ($SALESHARE_i * CRISIS_t$). I also include a set of control variables (X_{ijt}) that vary by lender, tract and time, such as the average income of loan applicants and the share of loan applicants that are from a minority group, as well as controls that vary by lender and time only, such as lender size. The composite error term (ν_{ijt}) consists of a tract-specific effect (θ_j), a tract-specific time trend (η_{jt}) and an idiosyncratic term (ϵ_{ijt}). The tract-specific effect captures unobservable factors in each tract that do not vary with time (e.g. cultural factors) while the tract-specific time trend captures unobservable factors in each tract that do vary with time (e.g. tract-level housing prices or employment prospects).

I collapse the time-series information into two periods – the pre-shock ('before') and post-shock ('after') periods – by taking the average of all observations before and after the crisis. The pre-shock period covers the years 2000 to 2006 while the post-shock period covers the years 2007 to 2008. Difference-in-differences estimation that uses many periods of data and focuses on serially-correlated outcomes can produce inconsistent standard errors (Bertrand et al. [2004]). Collapsing the data in this way smoothes out variation and generates conservative standard errors.

To aid computation, I take the first difference over time between the pre and post-crisis periods to obtain the equation in growth rates:

$$\triangle L_{ij} = \gamma + SALESHARE_i'\rho + \triangle X_{ij}'\phi + \underbrace{\triangle\eta_j + \triangle\epsilon_{ij}}_{\triangle\nu_{ij}} \tag{3.2}$$

Where the dependent variable is the percentage change in the number of new mortgage loans by lender i to tract j over 2007-08 ($\triangle L_{ij}$). The key explanatory variable is the share of loans sold by lender i on average in the pre-crisis period ($SALESHARE_i$). This variable is the proxy for the liquidity shock. The main coefficient of interest is the difference-in-differences estimator (ρ) which measures the causal effect of the liquidity shock on mortgage lending during the crisis. The test of the liquidity constraints hypothesis is a test of whether banks that were reliant on securitisation pre-crisis, and hence more exposed to the negative liquidity shock, reduced lending by more than banks that were not reliant on securitisation (i.e $\rho < 0$). The equation is reduced form in nature but can be derived as an equilibrium condition by explicitly modeling the credit supply and demand schedules (see Appendix A for the derivation).

The OLS estimator of ρ will be biased if unobservable credit demand shocks are correlated with a bank's reliance on loan sales (i.e. $corr(SALESHARE_i, \triangle\eta_j) \neq 0$). It is difficult to determine the direction of this bias. For example, a bank's reliance on loan sales and the credit demand shocks could be positively correlated if it is relatively more expensive to fund a loan through securitisation than through retail deposits (e.g. due to deposit insurance) so that only banks that experience particularly rapid growth in loan demand turn to securitisation. In this case, the OLS bias will be positive and the effect of the liquidity shock will be underestimated unless the demand shocks are controlled for. Alternatively, a bank's reliance on loan sales and the credit demand shocks could be negatively correlated. For instance, the OTD lenders may have been more likely to lend to risky borrowers that became particularly discouraged from borrowing when economic conditions deteriorated. In this case, there would be a negative correlation between the reliance on loan sales and the credit demand shock so the OLS bias will be negative, resulting in the effect of the liquidity shock being overestimated. More generally, variation in borrower composition across OTD and non-OTD banks that directly affects credit demand biases the estimated coefficient on the loan sale share variable.

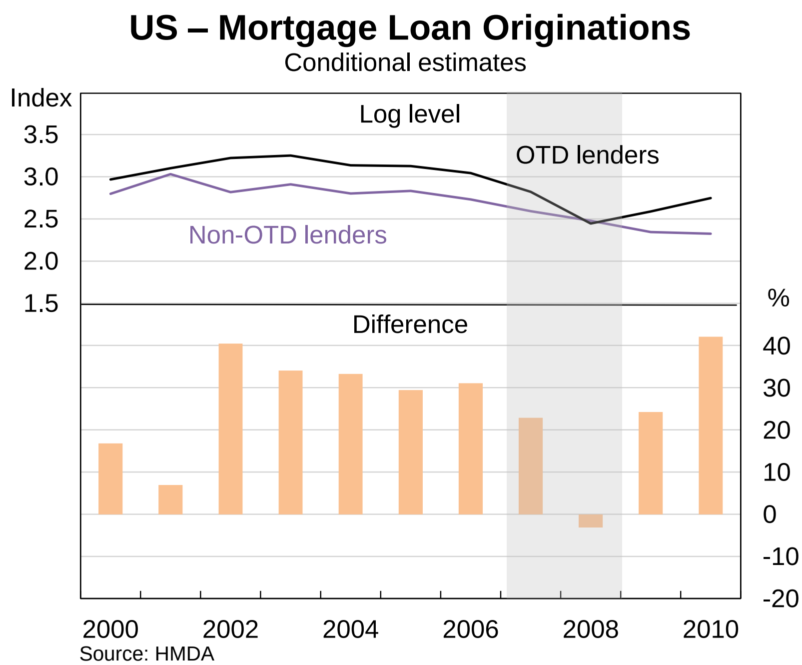
To address this issue, I include tract dummies ($\triangle\eta_j$) in the estimating equation that fully absorb all regional demand shocks, such as shocks to growth in local housing prices or local unemployment rates. The identification strategy assumes that changes in credit demand are felt proportionately across different banks that lend to borrowers *in the same tract*. The model then identifies the causal effect of the liquidity shock through variation in the lending behaviour of OTD and non-OTD lenders within the same tract. The remaining identifying assumption is that the financial crisis was not anticipated, so that a lender's reliance on the secondary market and lender-tract shocks are uncorrelated (i.e. $corr(SALESHARE_i, \triangle\epsilon_{ij}) = 0$). Put differently, US mortgage lenders did not adjust their financing structures in anticipation of the shock.

3.5.2 Graphical analysis

Before turning to the econometric analysis it is instructive to visually inspect the trends in the disaggregated loan-level data which underpin the regression. The graphical analysis is designed to see whether the difference-in-differences regression is driven by appropriate identification assumptions. The key identifying assumption in my empirical strategy is that the trends in mortgage lending are the same for the OTD lenders and non-OTD lenders in the absence of the shock to the securitisation market. This is known as the common (or parallel) trends assumption. Specifically, I compare the trends in the mortgage lending of OTD and non-OTD lenders, both before and after the credit crisis.

To aid comparisons with the regression analysis, I construct *conditional* estimates of the lending of both types of lenders. Specifically, I split lenders into OTD and non-OTD lenders based on whether the share of loan sales is above or below the lender-mean each year. I then separately calculate the average level of lending for both OTD and non-OTD lenders each

Figure 3.4: New Mortgage Lending by Type of Lender



year, and plot the logarithm of this mean estimate over time.

The aggregate trends in Figure 3.4 illustrate the impact of the liquidity shock on lending and generally support the identification strategy. On average, the OTD lenders supply more credit than the non-OTD lenders. But, more importantly, prior to the crisis, the trends in average lending for the two types of lenders were very similar, with the gap in lending between the two groups being relatively constant over time. This constant gap in lending activity supports the common trends assumption. As the crisis hit, the OTD lenders reduced new lending by significantly more than the non-OTD lenders, particularly in 2008. As the US economy emerged from recession in 2009, the lending of the two groups then reverted back to their pre-crisis levels. This overall time-series pattern of lending is consistent with the hypothesis that the OTD lenders became liquidity-constrained when a major source of funding – loan securitisation – evaporated, and this caused them to reduce credit supply relative to the non-OTD lenders that were not liquidity constrained.

Table 3.2: New Mortgage Lending

	(1) OLS	(2) FE
Sale share	-0.0937*	-0.0774*
	(-2.74)	(-2.17)
Income	0.0719***	0.0679***
	(8.75)	(8.63)
Minority share	-0.0398*	-0.0840***
	(-2.15)	(-3.66)
Lender size	-0.0208**	-0.0217**
	(-3.22)	(-3.14)
Constant	0.0824	0.0875
	(1.36)	(1.31)
Tract fixed effects	N	Y
R^2	0.015	0.062
Observations	1848528	1848528

t statistics in parentheses

Notes: Standard errors are clustered at the lender and tract levels.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

3.5.3 Econometric analysis

I next take my empirical methodology to the data using regression analysis. Table 3.2 below summarises the results of estimating the benchmark difference-in-differences equation for the causal effect of the liquidity shock. The first column provide the OLS estimates as a benchmark and the second column provides the preferred tract fixed effects estimates.

Overall, the results strongly support the hypothesis that the negative shock to lender financing conditions caused a reduction in mortgage credit supply. Each model specification suggests that mortgage lenders that were particularly reliant on loan sales cut lending by relatively more during the crisis. The estimates of the effect of the liquidity shock on the number of new loans are shown in the first row of the table. The fixed effect estimate, which controls for unobservable trends in credit demand, is -0.077 (column 2). The OLS estimate of the causal effect is -0.094 (column 1). The OLS estimate is further from zero than the fixed effects estimate, which suggests that the OLS bias is negative. In other words, there is a negative correlation between the share of loans sold and the tract-specific demand shocks.

This implies that the borrowers of OTD lenders reduced their credit demands by more than the borrowers of non-OTD lenders.

The coefficient estimate from the fixed effects specification implies that a one percentage point increase in the share of loans sold is associated with a decline in the level of new mortgage lending of about 7.7 per cent during the crisis. Alternatively, a one standard deviation increase in the share of loans that are sold (the standard deviation is 34 per cent) is associated with a fall in the level of new mortgage lending of about 2.7 per cent (=-0.077*0.34*100) during the crisis. At the lender-tract level, the total number of new loans fell by around 18.7 per cent, on average, over the crisis period. In other words, the estimates imply that a one standard deviation liquidity shock explains about 14 per cent (=2.7/18.7*100) of the total decline in new mortgage credit. This effect appears economically meaningful.

The coefficient estimates on the control variables are statistically significant. The income growth of loan applicants is positively correlated with growth in the number of new loans over the crisis period. The share of loan applicants that are from a minority has a negative effect on lending activity during the crisis. Larger mortgage lenders, as measured by the average number of loan applications, cut lending by more than smaller lenders during the crisis.

3.6 Testing the flight to quality and flight to home hypotheses

3.6.1 Identification

I now re-estimate the benchmark equation but augment it with indicators of borrower risk to determine how the effects of the liquidity shock vary across different classes of borrowers. I estimate the following panel regression model:

$$
\begin{aligned}
\triangle L_{ij} = \alpha_0 + RISKY'_{ij}\alpha_1 + DISTANCE'_{ij}\alpha_2 + SALESHARE'_{i}\beta_0 \\
+RISKY_{ij} * SALESHARE'_{i}\beta_1 + DISTANCE_{ij} * SALESHARE'_{i}\beta_2 \\
+\triangle X'_{ij}\phi + \triangle\eta_j + \triangle\epsilon_{ij}
\end{aligned}
\tag{3.3}
$$

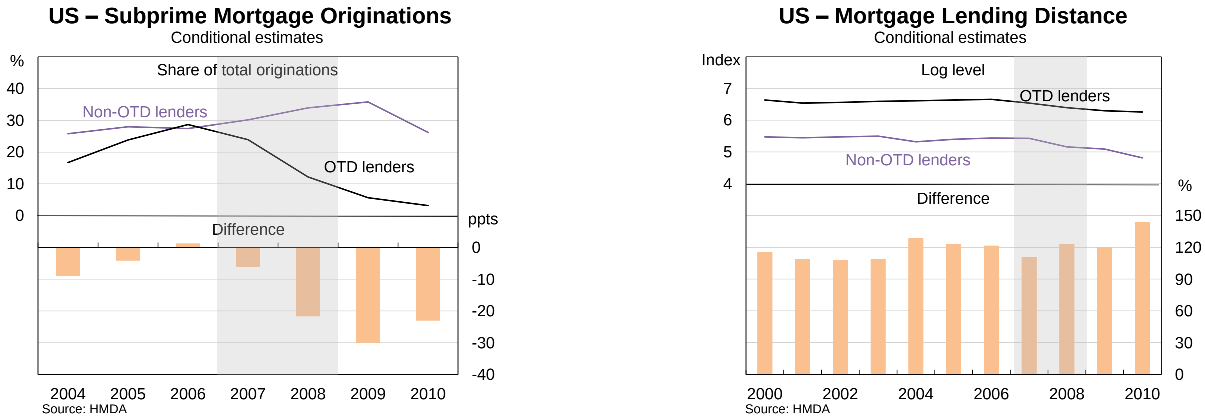
The dependent variable ($\triangle L_{ij}$) is again a measure of the change in lending activity during the crisis for lender i in tract j. Amongst the independent variables, I again include the liquidity shock variable ($SALESHARE_i$) but I now also interact this variable with indicators for whether the lending is risky or not and the distance between the borrower's home tract and the location of the lender's headquarters. More specifically, as explanatory variables I include a variable that measures the ratio of high-cost (or risky) lending to total lending ($RISKY_{ij}$) of lender i in tract j and a measure of the (log) number of kilometres between the headquarters of lender i and tract j ($DISTANCE_{ij}$). I also include the same set of pre-crisis control variables (X_{ij}) as in the benchmark equation.

If the liquidity shock caused a "flight to quality", the coefficient on the interaction variable $RISKY * SALESHARE$ should be significantly less than zero ($\beta_1 < 0$) as OTD lenders should shift lending away from risky borrowers by more than non-OTD lenders. If the liquidity shock caused a "flight to home", the coefficient on the interaction term $DISTANCE * SALESHARE$ should be significantly less than zero ($\beta_2 < 0$) as OTD lenders should be more likely to shift lending away from distant borrowers.

3.6.2 Graphical analysis

Next, I consider graphically how the liquidity shock affected lending to different groups of mortgage borrowers during the crisis. I follow the same procedure in splitting lenders into the two groups – OTD and non-OTD lenders – within each tract and year, but now examine how different types of lending evolved. In the LHS panel of Figure 3.5, I plot the evolution of the share of subprime lending for both OTD and non-OTD lenders. In the RHS panel, I plot

Figure 3.5: New Mortgage Lending by Type of Borrower



the evolution of the average lending distance for both groups. There is some evidence that the OTD lenders reduced their exposure to subprime lending by more than the non-OTD lenders around the time of the crisis (LHS panel). The overall share of subprime lending was broadly similar for the two groups in the pre-crisis period but a wedge emerged in 2007 as the subprime lending of the OTD lenders shrunk while the subprime lending of the non-OTD lenders remained elevated. The difference in the share of subprime lending persisted through the post-crisis period. This shift away from risky lending to less-risky lending is consistent with a "flight to quality" by the mortgage lenders affected by the liquidity shock.

On the other hand, there is little aggregate evidence that the OTD lenders shifted their mortgage lending towards closer borrowers by more than the non-OTD lenders (RHS panel). The conditional estimates indicate that, on average, the OTD lenders originate loans at more than twice the distance of non-OTD lenders. The average lending distance of the OTD lenders fell slightly relative to the non-OTD lenders in 2007, but quickly recovered the next year. In other words, based on this simple graphical analysis, there is little evidence in favour of the "flight to home" hypothesis.

Table 3.3: New Mortgage Lending by Type of Borrower

	(1) OLS	(2) FE
Risky share	-0.364*	-0.429*
	(-2.31)	(-2.70)
Distance	0.0211	0.0423*
	(1.52)	(2.86)
Sale share	-0.145+	-0.0309
	(-1.75)	(-0.32)
Sale X risky share	-0.279	-0.227
	(-1.25)	(-0.99)
Sale X distance	0.0102	-0.00345
	(0.55)	(-0.16)
Income	0.0775***	0.0725***
	(7.62)	(7.48)
Minority share	-0.0271	-0.141*
	(-1.04)	(-3.09)
Lender size	-0.0449***	-0.0513***
	(-4.11)	(-4.59)
Constant	0.217*	0.154*
	(2.92)	(2.05)
R^2	0.041	0.108
Observations	1270287	1270287

t statistics in parentheses

Notes: Standard errors are clustered at the lender and tract levels.

$^{+}$ $p < 0.10$, * $p < 0.05005$, *** $p < 0.001$

3.6.3 Econometric analysis

I have so far documented a sizeable negative impact of the liquidity shock stemming from the securitisation market on mortgage lending. Next, I consider which borrowers were most affected by the shock. The results of estimating equation 3.3 are shown in Table 3.3 below.

The negative coefficient on the $RISKY$ variable indicates that there was a tendency for all lenders to reduce credit to risky borrowers during the crisis. Moreover, the negative coefficient on the $SALE{*}RISKY$ interaction variable suggests that OTD lenders cut lending to risky borrowers by more than the non-OTD lenders. However, the differential effect is not statistically significant. In other words, there is evidence that *all* lenders reduced credit to risky borrowers, rather than just the OTD lenders, which is inconsistent with the specific flight to quality hypothesis considered here. The positive coefficient on the $DISTANCE$ variable suggests, surprisingly, that there was a tendency to increase lending to distant borrowers during the crisis, although the negative coefficient on the $SALE * DISTANCE$ interaction variable implies that the OTD lenders were relatively more likely to cut lending to distant borrowers. However, the differential effect is again not statistically significant. In other words, there is limited evidence to support the flight to home hypothesis.

3.7 Robustness Tests

My main hypothesis is that the OTD lenders became liquidity constrained when the securitisation market effectively shut down in 2007, and this increase in liquidity constraints, in turn, caused the OTD lenders to disproportionately reduce new mortgage lending. A key assumption underpinning this hypothesis, and the identification strategy, is that the OTD and non-OTD lenders are similar in all respects except the extent to which they became liquidity constrained during the crisis. There remains some possibility that the characteristics of OTD and non-OTD lenders differ along other unobservable dimensions, and that these differences could be driving the variation in lending behaviour across the two lender groups during the crisis. For example, OTD lenders may have been more willing to take risk or they may have been more reliant on specific loan origination channels that may have been associated with greater risk-taking (e.g. mortgage brokers). It could be these systematic differences, rather than differences in financing constraints, causing the observable variation in lending behaviour during the crisis. The purpose of the following series of tests is to rule out such alternative explanations for the observed link between securitisation and mortgage lending during the crisis.

3.7.1 Changes in mortgage lending standards

Some empirical studies suggests that securitisation contributed to bad lending by reducing the incentives of lenders to carefully screen borrowers (Mian and Sufi [2009], Keys et al. [2010], Purnanandam [2010], Rosen [2010]). These studies argue that, by creating distance between loan originators and the investors who bear the default risk, securitisation weakened lenders incentives to screen borrowers, leading to asymmetric information between loan originators and final investors and, subsequently, moral hazard problems. This suggests that OTD lenders may have had weaker incentives to screen borrowers than non-OTD lenders. It could be this pre-crisis difference in lending standards between the two groups, and not differences in financing constraints, that caused difference in lending behaviour during the crisis.

The purpose of this robustness test is to rule out this alternative 'lending standards' explanation for the link between securitisation and mortgage lending. To do so I need to make some subtle changes to the estimating equation. I firstly re-write the equation in levels:

$$\begin{aligned} L_{ijt} = \alpha + SALESHARE'_{ij}\beta + CRISIS'_t\gamma + SALESHARE_{ij} * CRISIS'_t\rho \\ + X'_{ijt}\phi + \underbrace{\theta_i + \theta_j + \eta_{it} + \eta_{jt} + \epsilon_{ijt}}_{\nu_{ijt}} \end{aligned} \tag{3.4}$$

There are two small, but important, differences between equation 3.1 and equation 3.4. First, the subscript on the loan sale variable ($SALESHARE_{ij}$) indicates that the reliance on loan sales varies by lender *and* tract, rather than just lender. The highly disaggregated nature of the loan-level data means that the share of loans that are sold can be constructed for each lender in each tract. Second, the additional variation in the loan sales share variable due to this re-specification means that lender-specific time trends (η_{it}) can be added to the error term. The lender-specific time trends control for unobservable bank-specific factors that vary over time, such as changes in bank lending standards. The original specification could not include lender-specific dummies as these would be perfectly collinear with the sales share variable which only varied by lender. This equation can again be written in growth rates by taking first differences over time:

$$\triangle L_{ij} = \gamma + SALESHARE'_{ij}\rho + \triangle X'_{ij}\phi + \underbrace{\triangle\eta_i + \triangle\eta_j + \triangle\epsilon_{ij}}_{\triangle\nu_{ij}} \tag{3.5}$$

As discussed earlier, the OLS estimator of ρ is biased if unobservable credit demand shocks are correlated with a lender-tract's reliance on loan sales (i.e. $corr(SALESHARE_{ij}, \triangle\eta_j) \neq 0$). But, as before, this is easily handled by including tract dummies in equation 3.5. However, the OLS estimator of ρ can now also be biased if unobservable credit supply shocks are correlated with a bank-tract's reliance on loan sales (i.e. $corr(SALESHARE_{ij}, \triangle\eta_i) \neq 0$). For example, if banks that were reliant on the originate-to-distribute model were also more likely to reassess their risk exposures and tighten lending standards during the crisis there could be a negative correlation between a bank-tract's reliance on loan sales and bank-specific lending growth (i.e. $corr(SALESHARE_{ij}, \triangle\eta_i) < 0$). By including a dummy variable for each lender in the growth rate equation, I control for unobservable changes in bank lending policies, including changes in lending standards. In other words, I now identify the relationship between the pre-crisis reliance on securitisation and mortgage lending during the crisis after controlling for *both* unobservable region-specific and lender-specific shocks. This eliminates any potential source of endogeneity caused by differences in the national lending policies of OTD and non-OTD lenders. The results of estimating equation 3.5 are shown below.

The results of estimating the model using the lender-tract share of loan sales are very similar to the benchmark model that uses the lender share of loan sales. Based on the specification with just tract fixed effects (column 2) the coefficient estimate is -0.108 which is similar to the estimate of -0.077 obtained from the benchmark fixed effects model. More importantly, the negative relationship between pre-crisis securitisation and lending activity

Table 3.4: Loan Sales by Tract and Lender

	(1) OLS	(2) Tract FE	(3) Lender FE	(4) Tract & Lender FE
Sale share	-0.111***	-0.108***	-0.0620	-0.102**
	(-4.32)	(-4.20)	(-1.63)	(-2.90)
Income	0.0678***	0.0631***	0.0747***	0.0700***
	(7.43)	(6.97)	(24.31)	(26.54)
Minority share	-0.0288	-0.0660*	0.00338	-0.0213*
	(-1.50)	(-2.27)	(0.31)	(-2.23)
Lender size	-0.220***	-0.222***	-0.278***	-0.300***
	(-12.51)	(-11.82)	(-19.26)	(-19.75)
Constant	0.172***	0.178***	0.215***	-0.00185
	(8.23)	(8.41)	(6.55)	(.)
Tract fixed effects	N	Y	N	Y
Lender fixed effects	N	N	Y	Y
R^2	0.138	0.176	0.227	0.226
Observations	1848528	1848528	1848528	1848528

t statistics in parentheses

Notes: Standard errors are clustered at the lender and tract levels.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

during the crisis still holds even when I include lender fixed effects (column 3) as well as both lender and tract fixed effects (column 4). In other words, the OTD lenders do not appear to have disproportionately reduced lending because of a relatively large (unobservable) tightening of bank lending standards. Rather, the link between loan sales and lending activity remains even after controlling for changes in bank lending policies, so the evidence favours the liquidity constraints hypothesis. The estimates from the specification with both lender and tract fixed effects (column 4) suggests that a one standard deviation shock to the share of loans sold is associated with a 21 per cent decline in new mortgage credit.

3.7.2 Private versus public securitisation

A key assumption underpinning the benchmark model is that US mortgage lenders cannot easily substitute between different sources of funding, so the lenders that were dependent on securitisation would have become more liquidity constrained when the private-label market shut down in 2007.[48] But this assumption may not hold if the mortgage lenders were able to substitute towards other less-affected sources of finance. For example, lenders that were particularly reliant on their ability to privately securitise mortgages may have been able to instead sell their loans to the government-sponsored enterprises. The flow of funds data presented earlier suggested that, in aggregate, there was a substitution away from private securitised lending to public securitised lending during the crisis as the government agencies stepped into the breach caused by the evaporation of the private-label market. In other words, the flow of funds evidence suggests that at least some lenders were able to substitute away from the worst-affected sources of finance.

The cross-sectional variation in the dependence of mortgage lenders on private versus public securitisation is useful in further testing the liquidity constraints hypothesis. The HMDA can be used to identify each lender's reliance on both public and private loan sales. If the liquidity constraints hypothesis is true, the lenders most reliant on *private* securitisation in the pre-crisis period should have become more liquidity constrained than lenders dependent on public securitisation and hence should have scaled back credit by relatively more during the crisis.

I re-estimate the benchmark equation but, for each bank, I split the share of loan sales into two components – the share of loans that are sold to the government-sponsored agencies ($PUBSHARE$) and the share of loans that are sold to private financial institutions ($PRIVSHARE$):

$$\triangle L_{ij} = PRIVSHARE_i'\beta_1 + PUBSHARE_i'\beta_2 + \triangle X_{ij}'\phi + \triangle\eta_j + \triangle\epsilon_{ij} \tag{3.6}$$

If the liquidity constraints hypothesis is true, then we should observe a significant negative

[48]There is an additional assumption that borrowers are unable to perfectly offset funding shocks by substituting other sources of external finance. This assumption is more likely to hold in housing finance than in corporate finance as corporations typically have much greater access than households to other funding sources (e.g. public debt and equity markets). Moreover, it is generally costly to re-apply for credit if the borrower's initial application is rejected. In other words, if there is a supply-side effect of the liquidity shocks, it is likely to be particularly important for household lending.

Table 3.5: Private and public securitisation

	(1) OLS	(2) FE
Private sale share	-0.102**	-0.0816*
	(-2.85)	(-2.19)
Public sale share	-0.0527	-0.0583
	(-0.67)	(-0.72)
Income	0.0719***	0.0679***
	(8.92)	(8.71)
Minority share	-0.0360*	-0.0824***
	(-2.15)	(-3.83)
Lender size	-0.0226**	-0.0226**
	(-3.29)	(-3.16)
Constant	0.0960	0.0945
	(1.53)	(1.40)
Tract fixed effects	N	Y
R^2	0.016	0.062
Observations	1848528	1848528

t statistics in parentheses

Notes: Standard errors are clustered at the lender and tract levels.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

effect of the private sale share variable ($PRIVSHARE$) on lending (i.e. $\beta_1 < 0$). If the effect of the public loan sale variable is also negative and significant, then this would be evidence that there is some confounding factor, related to the OTD business model, that causes all such lenders to cut credit. The results are shown below.

Comparing the coefficient estimates on the private and public sale share variables (rows 1 and 2) suggests that the effect of the liquidity shock on lending activity is significantly larger for the private sale share than for the public sale share. Moreover, the estimated effect of the private sale share on lending activity is economically larger than the effect for the total sale share shown in Table 3.2. A one standard deviation shock to the private sale share variable is associated with a 16 per cent decline in new mortgage credit. This supports the hypothesis that OTD lenders reduced lending because they became financially constrained in 2007 and not because of some other unobservable confounding factor affecting all lenders that are reliant on the originate-to-distribute model.

3.7.3 Afilliated non-bank mortgage lenders

To test the relative merits of the liquidity constraints hypothesis I construct another test in which I focus specifically on OTD lenders. Many OTD lenders are mortgage companies that specialise in home mortgage lending, whereas most non-OTD lenders are depository institutions. These mortgage companies typically rely solely on securitisation to finance their lending, and generally cannot fund themselves through alternative sources of finance, such as deposits. However, there are important differences *within* the pool of OTD lenders. For instance, some mortgage companies are affiliated with depository institutions, such as Citigroup, whereas others are not. The mortgage companies that are affiliated with a bank potentially have access to a more diversified funding base (through internal capital markets) than mortgage companies that are not affiliated. We might therefore expect the affiliated mortgage companies to be relatively less vulnerable to a funding shock that is specific to the securitisation market than the non-affiliated companies.

This variation amongst OTD lenders in the ability to diversify funding risk allows me to test the liquidity constraints hypothesis against alternative explanations for the link between pre-crisis reliance on securitisation and post-crisis mortgage lending. If the liquidity constraints hypothesis is true, we would expect the lending of non-affiliated mortgage companies (that are solely reliant on securitisation) to be relatively more responsive to the liquidity shock caused by the closure of securitisation market than the lending of affiliated mortgage companies. So I re-estimate the benchmark equation to again test the liquidity constraints hypothesis, but now focus on the subset of mortgage companies (that are predominantly OTD lenders):

$$\begin{aligned} \triangle L_{ij} &= NONAFF_i'\beta_0 \\ &+NONAFF_i * SALESHARE_i'\beta_1 + AFF_i * SALESHARE_i'\beta_2 \\ &+\triangle X_{ij}'\phi + \triangle\eta_j + \triangle\epsilon_{ij} \end{aligned} \tag{3.7}$$

Table 3.6: New Mortgage Lending by Non-Bank Lenders

	(1) OLS	(2) FE
Non-affiliated	0.742	0.730
	(1.50)	(1.36)
Non-aff. x sale share	-0.114	-0.0843
	(-1.31)	(-0.92)
Aff. x sale share	0.627	0.630
	(1.20)	(1.12)
Income growth	0.0613**	0.0604***
	(3.22)	(3.34)
Lender size	-0.0361*	-0.0410*
	(-2.16)	(-2.34)
Minority share	-0.0720**	-0.0515^{+}
	(-2.90)	(-1.87)
Constant	-0.516	-0.483
	(-1.00)	(-0.87)
R^2	0.022	0.133
Observations	549118	549118

t statistics in parentheses

Notes: Standard errors are clustered at the lender and tract levels.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

where I include two dummy variables for whether the lender is affiliated with a commercial bank or not. The dummy variable $NONAFF$ takes a value of one if the lender is not affiliated with a bank and is zero otherwise, while the dummy variable AFF takes a value of one if the lender is affiliated with a bank and is zero otherwise. I also interact these dummy variables with the share of loans sold by each lender. All the other variables are as before. Under the liquidity constraints hypothesis, the negative effect of the liquidity shock should be larger for the non-affiliated lenders than for the affiliated lenders (i.e. $\beta_1 - \beta_2 < 0$). The results are shown in the table below.

In both columns 1 and 2, the coefficient estimate on the interaction term $NONAFF * SALESHARE$ is negatively signed, indicating that the non-affiliated mortgage companies reduced their lending during the crisis. In contrast, the affiliated mortgage companies expanded credit over the period, as demonstrated by the positive coefficient estimate on the interaction term $AFF * SALESHARE$. In other words, the non-affiliated lenders that lacked alternative funding sources had a greater tendency than the affiliated lenders to reduce lending in response to the liquidity shock. However, a one-sided joint test suggests that the difference between the coefficient estimates is not statistically significant. So, overall, the results provide some tentative evidence to support the liquidity constraints hypothesis.

3.7.4 The aggregate effect of the liquidity shock

The results of estimating the benchmark equation indicate that the liquidity shock, caused by the closure of the securitisation market, had a negative effect on new mortgage lending. The analysis is based on a comparison of the behaviour of OTD and non-OTD lenders that grant credit within each tract. However, this "within-tract" identification strategy does not provide the complete picture of the aggregate effect of the liquidity shock on mortgage lending. This is because the strategy implicitly does not allow borrowers to substitute between different lenders (where Census tracts are thought of as "borrowers"). But borrowers might compensate for any reduction in credit from OTD lenders by obtaining alternative finance from non-OTD lenders. This substitution towards unaffected lenders could limit the effect of the liquidity shock on aggregate mortgage lending.

One approach to identify the aggregate effect of the liquidity shock on mortgage lending would be to estimate the relationship based on a simple tract-level regression. A regression of tract-level lending growth on the tract-level share of loans that are sold would implicitly allow borrowers to substitute between lenders. However, as discussed earlier, the estimates from such a regression will be biased if changes in total mortgage credit at the tract level reflect both changes in credit demand and supply.

Jimenez et al. [2011] have recently proposed a method to estimate the unbiased aggregate effect of the liquidity shock on new lending growth. The model is estimated at the tract-level, but the implied coefficient estimates are adjusted for bias using the (unbiased) coefficient estimates obtained at the bank-tract level. This approach effectively separates the impact of supply from demand while taking into account general equilibrium adjustments by borrowers. The approach is described in more detail in the Appendix. Basically, I estimate the tract-level version of equation 3.2:

$$\bar{\triangle L_j} = \gamma + SALE\bar{S}HARE'_j\rho + \bar{X}'_j\phi + \eta_j + \bar{\epsilon}_j \tag{3.8}$$

Where $\bar{\triangle L_j}$ denotes the log change in credit for tract j across all mortgage lenders. It is essentially a weighted average of the growth rate of credit at the bank-tract level, where the weights are given by each lender's share of loans within each tract. Similarly, $SALE\bar{S}HARE_j$ denotes the (weighted) average pre-crisis reliance on loan sales of lenders that grant credit to tract j. The specification also includes a set of tract-level control variables, such as the average growth in the income of loan applicants ($\bar{X}_j$). The same credit demand shock (η_j) appears in equations 3.2 and 3.8 under the assumption that the shock equally affects a tract's borrowing from each lender. I then adjust the implied estimates using the adjustment formula outlined in the Appendix:

$$\bar{\beta} = \hat{\bar{\beta}}_{OLS} - (\hat{\beta}_{OLS} - \hat{\beta}_{FE}) * \frac{V(SALESHARE_i)}{V(SALE\bar{S}HARE_j)}$$

The results of estimating equation 3.8 are shown in Table 3.7. As the first row of the table indicates, the OLS estimate of the aggregate effect is -0.364, but this estimate needs to be adjusted to account for any bias caused by demand shocks. Recall that the bank-tract level OLS estimate is -0.094 and the fixed effects estimate is -0.077 so, using the above adjustment formula, the unbiased estimate of the aggregate effect of the liquidity shock is -0.207. This is obtained from the calculation: $-0.207 = -0.364 - (-0.094 + 0.077) * (0.012/0.116)$.

The fact that the unbiased estimate of the aggregate effect of the liquidity shock (-0.207) is larger than the unbiased estimate of the partial equilibrium effect (-0.077) implies that borrowers, somewhat surprisingly, substituted towards affected OTD lenders and away from non-OTD lenders.

Table 3.7: Aggregate New Mortgage Lending

	(1) OLS
Sale share	-0.364***
	(-48.70)
Income	0.139***
	(24.75)
Minority share	-0.0803***
	(-23.62)
Lender size	-0.0304***
	(-28.90)
Constant	0.401***
	(45.89)
Tract fixed effects	N
R^2	0.194
Observations	63269

t statistics in parentheses

Notes: Standard errors are clustered at the tract level.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$

Overall, the results imply that a one standard deviation shock to the share of loans sold in aggregate is associated with a decline in new mortgage lending of around 2.3 per cent, on average. In aggregate, new mortgage lending fell by around 16.7 per cent so a one standard deviation shock to the share of loans sold can explain about 14 per cent of the aggregate fall in new mortgage credit. This estimated (general equilibrium) effect is very similar to the (partial equilibrium) effect identified at the more disaggregated lender-tract level.

3.8 Conclusion

The flow of new residential mortgage lending in the US contracted sharply over 2007-08, which is often attributed to a significant tightening in credit conditions. But, to the best of my knowledge, there has been little formal testing of the extent to which this decline in new credit can be attributed to a reduction in credit supply. I fill this gap in the empirical literature by examining the extent to which the fall in US residential mortgage credit over

2007-08 was caused by a reduction in credit supply which, in turn, can be traced to a fall in liquidity. I use the differential exposure of US mortgage lenders to the collapse of the securitisation market in 2007 as a source of cross-lender variation in lender financing conditions and assess the impact on residential mortgage lending. The assumption is that lenders that were reliant on funding from the secondary market would have become more liquidity constrained when the market collapsed in late 2007. If alternative sources of bank finance were relatively expensive, this would have caused lenders to reduce the supply of credit.

Controlling for unobservable credit demand shocks that may be correlated with borrower risk, I show that mortgage lenders that were particularly reliant on loan securitisation reduced credit disproportionately during the crisis. This main result is robust to a series of tests that are designed to rule out alternative explanations for the link between pre-crisis reliance on securitisation and mortgage lending during the crisis. For example, the results are upheld when I control for unobservable changes in bank risk-taking through lender-specific time trends. In other words, systematic differences in lending standards between OTD and non-OTD lenders do not appear to explain the variation in lending behaviour. Moreover, focussing just on lenders reliant on the originate-to-distribute business model, I find that mortgage companies that were affiliated with a commercial bank were less likely to cut credit than non-affiliated companies. Assuming the affiliated mortgage companies were more able to access alternative sources of funding when the crisis hit, this provides further support for the liquidity constraints explanation.

I also find that the effect of the liquidity shock on lending activity was particularly strong for lenders that were reliant on selling loans to private, rather than public, financial institutions. This is again consistent with the liquidity constraints hypothesis as it was the private-label securitisation market that was most adversely affected when capital market investors withdrew funding in late 2007. In contrast, the public securitisation market largely remained liquid due to active government support. Furthermore, I find little evidence that borrowers substituted away from the affected OTD lenders, which implies that the adverse liquidity shock had a significant impact on the aggregate level of new mortgage credit as well.

However, I do not find that the most affected lenders disproportionately reduced credit to risky borrowers, which is inconsistent with a "flight to quality" caused by the liquidity shock. Furthermore, I do not find evidence to indicate that the affected lenders disproportionately reduced credit to non-local borrowers, which appears to contradict the "flight to home" hypothesis.

Appendices

D Identifying the effect of credit supply shocks

In this Appendix I outline a model that describes how credit demand and supply shocks affect mortgage lending. The model is closely related to that of Khwaja and Mian [2008]. The purpose of the model is to highlight the identification problem and explain the construction of the estimator that controls for credit demand shocks. In this simple model I assume that each bank i lends to region j at time t the amount L_{ijt}. There are two time periods. I assume that each bank can lend to only one region, but a region can borrow from multiple banks. On the credit supply side, I assume that each bank loan to a particular region is financed through a combination of securitisation S_{it} and other forms of external financing W_{it}. These assumptions generate the following flow of funds constraint:

$$L_{ijt} = S_{it} + W_{it}$$

I assume that the bank can securitise loans at no cost, but, importantly, there is a quantitative limit on how much the bank can securitise (i.e. $S_{it} < \bar{S}_j$). Beyond this limit, the bank must turn to costly external finance to support higher levels of lending. These assumptions ensure that the level of securitisation matters to the lending decision of each bank. If there was no limit on securitisation and/or no cost of accessing external finance, then the funding structure of the bank would not affect lending.[49] Under this set of assumptions, the marginal cost of lending for the bank is solely a function of the volume of wholesale debt (i.e. external finance):

$$MC_{ijt} = \gamma W_{it}$$

The cost parameter ($\gamma > 0$) denotes the slope of the marginal cost curve. On the credit demand side, I assume the marginal loan return is given by the following equation:

$$MR_{ijt} = \bar{r}_{jt} - \alpha L_{ijt}$$

The borrower quality parameter ($\bar{r}_{jt}$) allows for variation in loan returns across regions. Given the slope parameter (α) is a positive constant, the formulation assumes that there are diminishing marginal returns to borrowing. I solve for the first-period equilibrium by equating marginal revenue with marginal cost and substituting the flow of funds identity:

$$L^*_{ijt} = \frac{1}{\alpha + \gamma}(\bar{r}_{jt} + \gamma S_{it})$$

Where the superscript '*' denotes equilibrium. At the end of the first period, the credit market experiences two types of shocks:

1. Credit demand shock: $\bar{r}_{jt+1} = \bar{r}_{jt} + \bar{\eta} + \eta_j$

2. Credit supply shock: $S_{it+1} = S_{it} + \bar{\delta} + \delta_i$

[49] I further assume that the bank cannot fund loans through internal finance (e.g. deposits). While this appears to be a strong assumption, it is only made to simplify the algebra – the results still hold if I instead assume that banks finance lending through deposits (D_{it}), where deposits are the cheapest form of funding. In that case, the key assumptions are that there is also a quantitative limit on internal finance (i.e. $D_{it} < \bar{D}$) and that securitisation is a cheaper form of funding than other forms of external finance.

The credit demand shock consists of two terms – an aggregate shock that is common to all regions ($\bar{\eta}$) and an idiosyncratic shock that is specific to each region (η_j). In terms of the econometric framework, the aggregate credit demand shock might be an unexpected change in US monetary policy while the region-specific demand shock might be a shock to regional house prices. The credit supply shock also consists of two terms – an aggregate shock that is common to all banks ($\bar{\delta}$) and a bank-specific shock (δ_i). The aggregate credit supply shock might reflect some change in financial regulation that affects the ability of banks to securitise loans while the bank-specific credit supply shock could reflect each bank's ability to securitise assets.

Following the same approach as before, I solve for the second-period equilibrium:

$$L^*_{ijt+1} = \frac{1}{\alpha+\gamma}(\bar{r}_{jt+1} + \gamma S_{it+1})$$

As the two period solutions are linear I can then take the difference in (equilibrium) lending over time ($\triangle L^*_{ij} = L^*_{ijt+1} - L^*_{ijt}$) to obtain:

$$\triangle L^*_{ij} = \frac{(\bar{\eta}+\eta_j)}{(\alpha+\gamma)} + \frac{\gamma(\bar{\delta}+\delta_i)}{(\alpha+\gamma)}$$

The change in loan amount consists of two terms. The first term on the right-hand side denotes the impact of the region-specific credit demand shocks on lending. The second term denotes the impact of the bank-specific supply shocks. Note that if there is no cost of external finance ($\gamma = 0$) the credit supply shocks will not affect the equilibrium growth rate of lending; lending growth will only be a function of demand shocks. In other words, credit supply shocks only matter if there are financing frictions on the lender side.

Now suppose I re-arrange the equation to combine all the aggregate shocks in a single term and have two separate terms for the bank-specific and region-specific shocks:

$$\triangle L^*_{ij} = \frac{(\bar{\eta}+\gamma\bar{\delta})}{(\alpha+\gamma)} + \frac{\gamma\delta_i}{(\alpha+\gamma)} + \frac{\eta_j}{(\alpha+\gamma)}$$

If I assume that the share of loans that are securitised ($SALESHARE_i$) by each bank is a suitable proxy for the bank-specific credit supply shock (δ_i) then I could run the OLS regression:

$$\triangle L^*_{ij} = \beta_0 + SALESHARE'_i\beta_1 + \underbrace{\eta_j + \epsilon_{ij}}_{\nu_{ij}}$$

Where there is an intercept that captures all the aggregate effects ($\beta_0 = \frac{(\bar{\eta}+\gamma\bar{\delta})}{(\alpha+\gamma)}$), a slope coefficient ($\beta_1 = \frac{\gamma}{(\alpha+\gamma)}$) that captures the relationship of interest and a composite error term (ν_{ij}) which consists of a region-specific component (η_j) and a bank-region specific component (ϵ_{ij}). If the share of loans that are securitised ($SALESHARE_i$) is correlated with the unobservable credit demand shocks (η_j) then the OLS estimate of β_1 will be biased. But suppose the region borrows from both an OTD lender and a non-OTD lender. Denote the OTD lender with subscript O and the non-OTD lender with subscript N. Consider a region j that has a loan from each type of bank and compute the within-region difference in lending growth:

$$\triangle L_{Oj} - \triangle L_{Nj} = (SALESHARE_O - SALESHARE_N)'\beta_1 + (\eta_j - \eta_j) + (\epsilon_{Oj} - \epsilon_{Nj})$$

This equation eliminates both the effect of the aggregate shocks and the unobservable region-specific credit demand shocks. Equivalently, I can run the OLS regression with the inclusion of borrower-specific fixed effects to control for the unobservable credit demand shocks (η_j). I can obtain an unbiased estimate of the causal effect of the credit supply shock under the identifying assumption that each bank's loan securitisation share is uncorrelated with the bank-region specific errors ($corr(SALESHARE_i, \epsilon_{ij}) = 0)$).

E Identifying the separate effects of liquidity and lending standards shocks

I now modify the model to allow for changes in bank lending standards. I maintain all the assumptions on the demand side of the credit market, but make two slight changes to the supply side. Firstly, identification now requires that each bank loan to a particular region is funded by external finance that is obtained from that specific region. This implies that the flow of funds constraint becomes:

$$L_{ijt} = S_{ijt} + W_{ijt}$$

Secondly, I assume that the bank must exert some costly "screening effort" (E_{it}) to originate each loan. The effort exerted in screening borrowers can be loosely thought of as the bank's lending standards; a bank that exerts more effort has stricter lending standards. I assume that the cost of screening is given by a convex function ($\frac{\phi E_{it}^2}{2}$). Under this set of assumptions, the marginal cost of lending for the bank is a function of the volume of external finance *and* the level of screening effort:

$$MC_{ijt} = \gamma W_{ijt} + \phi E_{it}$$

As before, the marginal loan return is given by:

$$MR_{ijt} = \bar{r}_{jt} - \alpha L_{ijt}$$

Solving for the first-period equilibrium by equating marginal revenue with marginal cost:

$$L^*_{ijt} = \frac{1}{\alpha + \gamma}(\bar{r}_{jt} + \gamma S_{ijt} - \phi E_{it})$$

At the end of the first period, the credit market now experiences three shocks. There is the same demand shock as before, but now there are two types of credit supply shocks - a liquidity shock and a 'new' shock to screening effort (or 'lending standards shock'):

1. Credit demand shock: $\bar{r}_{jt+1} = \bar{r}_{jt} + \bar{\eta} + \eta_j$
2. Credit supply (liquidity) shock: $S_{ijt+1} = S_{ijt} + \bar{\delta} + \delta_{ij}$
3. Credit supply (lending standards) shock: $E_{it+1} = E_{it} + \bar{\psi} + \psi_i$

The liquidity shock is similar to before, except that it is now specific to each bank and region (δ_{ij}) rather than specific to each bank. The new lending standards shock comprises two components: an aggregate shock ($\bar{\psi}$), such as a change in loan screening technology (e.g. the availability of credit scoring), and a bank-specific shock (ψ_i), such as a change in bank risk preferences.

Following the same approach as before, I solve for the second-period equilibrium:

$$L^*_{ijt+1} = \frac{1}{\alpha + \gamma}(\bar{r}_{jt+1} + \gamma S_{ijt+1} - \phi E_{it+1})$$

Taking the difference in (equilibrium) lending over time I obtain:

$$\triangle L^*_{ij} = \frac{(\bar{\eta} + \eta_j)}{(\alpha + \gamma)} + \frac{\gamma(\bar{\delta} + \delta_{ij})}{(\alpha + \gamma)} - \frac{\phi(\bar{\psi} + \psi_i)}{(\alpha + \gamma)}$$

Compared to the basic model, this equation now has an additional third term on the right-hand side, which denotes the impact of the lending standards shock. Note that if there is no cost of screening loans ($\phi = 0$), then the lending standards shock has no effect on the equilibrium growth rate of lending.

Next, I re-arrange the equation to combine all the aggregate shocks in a single term and have three separate terms for the bank-specific, region-specific and bank-region specific shocks:

$$\triangle L^*_{ij} = \frac{(\bar{\eta} + \gamma\bar{\delta} - \phi\bar{\psi})}{(\alpha + \gamma)} + \frac{\gamma\delta_{ij}}{(\alpha + \gamma)} + \frac{\eta_j}{(\alpha + \gamma)} - \frac{\phi\psi_i}{(\alpha + \gamma)}$$

If I assume that the share of loans that are sold ($SALESHARE_{ij}$) by each bank in each region is a suitable proxy for the liquidity shock (δ_{ij}) then I obtain the estimating equation:

$$\triangle L^*_{ij} = \beta_0 + SALESHARE'_{ij}\beta_1 + \underbrace{\eta_j + \psi_i + \epsilon_{ij}}_{\nu_{ij}}$$

Where there is an intercept that captures all the aggregate effects ($\beta_0 = \frac{(\bar{\eta}+\gamma\bar{\delta}-\phi\bar{\psi})}{(\alpha+\gamma)}$), a slope coefficient ($\beta_1 = \frac{\gamma}{(\alpha+\gamma)}$) that captures the relationship of interest, and a composite error term (ν_{ij}) which includes an unobservable region-specific component (η_j), an unobservable bank-specific component (ψ_i) and a bank-region specific component (ϵ_{ij}).

I can estimate this equation by OLS and include borrower-specific fixed effects to control for the unobservable credit demand shocks (η_j) *and* bank-specific fixed effects to control for the unobservable lending standards shocks (ψ_i). I can obtain an unbiased estimate of the causal effect of the liquidity shock assuming that the share of loans sold by each bank in each region is uncorrelated with the bank-region specific errors (i.e. $corr(SALESHARE_{ij}, \epsilon_{ij}) = 0$).

F Estimating the unbiased aggregate effect of the liquidity shock

In this Appendix, I outline the methodology used by Jimenez et al. [2011] to estimate the unbiased aggregate effect of the liquidity shock on new lending growth. The model is estimated at the tract-level, but the implied coefficient estimates are adjusted for bias using the (unbiased) coefficient estimates obtained at the bank-tract level. The approach separates the impact of supply from demand, while taking into account general equilibrium adjustments by borrowers.

It is helpful to outline the methodology in a few steps. For simplicity, suppose there are no control variables. Recall equation 3.2 (without controls) estimated at the bank-tract level:

$$\triangle L_{ij} = SALESHARE_i'\beta + \eta_j + \epsilon_{ij} \tag{3.9}$$

But I want to estimate the tract-level version of this equation:

$$\bar{\triangle L}_j = SALES\bar{H}ARE_j{}'\beta + \eta_j + \bar{\epsilon}_{ij} \tag{3.10}$$

Where $\bar{\triangle L}_j$ denotes the log change in credit for tract j across all mortgage lenders. It is a weighted average of the growth rate of credit at the bank-tract level, where the weights are given by each lender's share of loans within each tract. It is not a simple unweighted average of $\triangle L_j$ because tracts can start borrowing from new mortgage lenders. The tract-level measure of new lending is constructed by adding up the total number of new loans originated by each mortgage lender within a given tract each year. Similarly, $SALES\bar{H}ARE_j$ denotes the (weighted) average pre-crisis reliance on loan sales of lenders that grant credit to tract j. This variable is slightly more complicated to construct as it requires converting a measure of the share of loans that are sold by each lender ($SALESHARE_i$) into a measure of the share of loans that are sold within each tract ($SALES\bar{H}ARE_j$). The tract-specific measure of loans sold is constructed using the following formula:

$$SALES\bar{H}ARE_j = \sum_{i=1}^{N_j} w_{ij} * SALESHARE_i$$

where $w_{ij} = L_{ij}/L_j$ is the share of new loans originated by lender i within each tract j and N_j is the set of lenders that originate loans in tract j. Note also that the same credit demand shock (η_j) appears in both equations 3.9 and 3.10 under the assumption that the shock equally affects a tract's borrowing from each lender.

Recall that the OLS estimate of the (partial equilibrium) effect of the liquidity shock at the bank-tract level is given by:

$$\hat{\beta}_{OLS} = \beta + \frac{cov(SALESHARE_i, \eta_j)}{V(SALESHARE_i)} \tag{3.11}$$

Also, recall that the fixed-effects (FE) estimate (that controls for credit demand shocks) provides an unbiased estimate of the effect of the liquidity shock:

$$\hat{\beta}_{FE} = \beta \tag{3.12}$$

Combining these two conditions we obtain:

$$\hat{\beta}_{OLS} - \hat{\beta}_{FE} = \frac{cov(SALESHARE_i, \eta_j)}{V(SALESHARE_i)} \tag{3.13}$$

Now the OLS estimate of the aggregate (general equilibrium) effect of the liquidity shock at the tract level is given by:

$$\hat{\bar{\beta}}_{OLS} = \bar{\beta} + \frac{cov(SALE\bar{S}HARE_j, \eta_j)}{V(SALE\bar{S}HARE_j)} \tag{3.14}$$

But this will be biased if there is any correlation between the share of loans sold in a particular tract and unobservable tract-specific trends, such as shocks to local housing prices. We cannot follow the same procedure as before and estimate a fixed-effects version of equation 3.10 because the unobservable tract-specific fixed effect (η_j) is collinear with the key explanatory variable ($SALE\bar{S}HARE_j$). Assuming that the covariance between the share of loans sold by each lender and the tract-specific demand shocks is the same across all lenders, then the following condition holds:

$$\begin{aligned} cov(SALE\bar{S}HARE_j, \eta_j) &= cov(\sum_{i=1}^{N_j} w_{ij} * SALESHARE_i, \eta_j) \\ &= cov(SALESHARE_i, \eta_j) \end{aligned} \tag{3.15}$$

Combining equations 3.13, 3.14, 3.15 we obtain the aggregate bias-adjustment formula:

$$\bar{\beta} = \hat{\bar{\beta}}_{OLS} - (\hat{\beta}_{OLS} - \hat{\beta}_{FE}) * \frac{V(SALESHARE_i)}{V(SALE\bar{S}HARE_j)} \tag{3.16}$$

This is the formula used to obtain the unbiased estimate of the aggregate effect of the liquidity shock presented in the paper. Importantly, both the variance of the bank-specific liquidity shocks ($V(SALESHARE_i)$) and the variance of the (weighted) tract-specific liquidity shocks ($V(SALE\bar{S}HARE_j)$) can be obtained from the data. This means that all the terms on the right-hand side of the equation can be estimated, providing an unbiased estimate of the aggregate effect of the liquidity shock ($\bar{\beta}$).

G The measurement of subprime mortgage lending

As briefly discussed in the data section of the paper, there are a couple of problems with using the share of high-priced loans as an indicator of risky (or subprime) lending. Firstly, it may be biased as the share of high-priced mortgages can change over time due to changes in the yield curve rather than changes in bank lending policies (Mayer and Pence [2008]). The share of high-priced loans is affected by the way in which the HMDA measures the interest rate spread on each loan. The HMDA does not collect information directly on the interest rate spread, but rather estimates the spread from information it collects on the interest rate of each loan. For example, to calculate the interest rate spread on an adjustable-rate mortgage (ARM) with a contract maturity of 30 years, the HMDA uses the interest rate on a 30-year Treasury bond even though the interest rate on the loan may actually be priced off a shorter-term security. In other words, the comparable maturity of the loan corresponds to the maturity in the loan contract, and not the expected maturity of the loan (which is more likely to be used by the lender). As short-term rates are generally lower than long-term rates, subprime ARMs are likely to be underreported in the data (because there will be fewer loans reported with a sufficiently large interest rate spread). More problematically, the extent of this bias shifts over time as the slope of the yield curve changes. For instance, if longer-term rates fall relative to short-term rates (i.e. the yield curve becomes flatter), the measured share of subprime ARMs will rise as a result of this bias. It is estimated that at least 13 per cent of the increase in the number of higher-priced loans in the HMDA data between 2004 and 2005 is attributable to a flattening of the yield curve (Avery et al. [2007]).

Chapter 4

State Anti-Predatory Lending Laws, Bank Lending Standards and Household Spending

I exploit a natural experiment on the US housing market to identify the causal effect of bank lending standards on the real economy. Prior to the subprime mortgage crisis, different US states enacted anti-predatory lending laws – equivalent to a government-enforced tightening of credit standards for subprime mortgages – at different times. By comparing the economic performance of contiguous counties across state borders where one state introduces the law ("the treatment group") and the other state does not ("the control group") I identify the economic effect of shocks to residential mortgage credit standards. Using a difference-in-difference modelling framework, I find that the introduction of the state lending laws is associated with tighter credit conditions but little evidence that the laws affected household spending.

4.1 Introduction

"We at the Federal Reserve were aware as early as 2000 of incidents of some highly irregular subprime mortgage underwriting practices. But regrettably we viewed it as a localized problem subject to standard prudential oversight, not the precursor of the securitized subprime mortgage bubble that was to arise several years later."

(Greenspan 2010, "The Crisis", p.7)

"Stronger regulation and supervision aimed at problems with underwriting practices and lenders' risk management would have been a more effective and surgical approach to constraining the housing bubble than a general increase in interest rates."

(Bernanke 2010, speech at the Annual Meeting of the American Economic Association, Atlanta, 3 January)

The objective of this paper is to identify a causal effect of bank lending standards shocks on the real economy through a natural experiment on the US housing market. Previous academic research that has attempted to identify a causal effect of lending standards shocks on economic activity has faced two significant empirical challenges. Firstly, there may be reverse causality from economic conditions to credit standards, as the credit decisions of both lenders and borrowers are likely to be influenced by the state of the economy. Secondly, there may be omitted variables bias as it can be difficult to separate the effects of bank lending standards from that of other unobservable factors that might affect both credit supply and the real economy (e.g. changes in banking competition).

I exploit a natural experiment that addresses these endogeneity problems. I examine the effect of the introduction of state anti-predatory lending laws (APLs) on the US housing market over the past decade. The lending laws are equivalent to an exogenous (government-enforced) tightening of residential mortgage credit standards as they restrict the number of high-risk loans available to borrowers. These state-level restrictions on the riskiness of credit may, in turn, affect consumer spending if high-risk households are generally financially constrained and therefore reliant on banks to finance their expenditures. For example, tighter

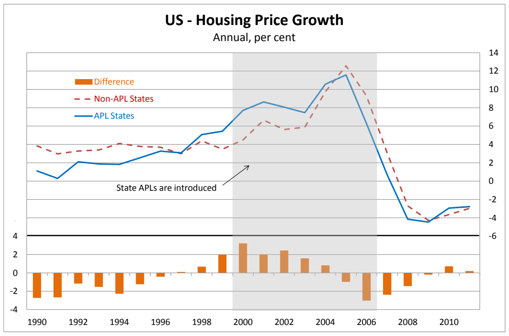
(c) Real GDP

(d) Real Retail Trade

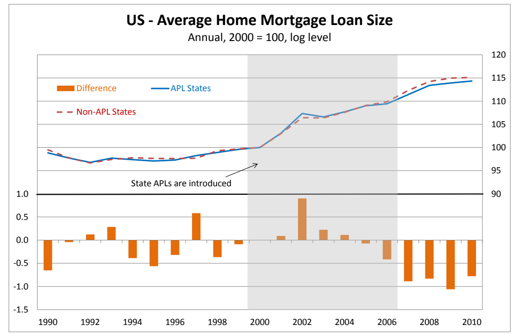
Figure 4.1: Real GDP and Consumer Spending Growth by APL and Non-APL States

mortgage credit standards may make it more difficult to refinance a mortgage, reducing the potential cash flow of the borrower relative to a situation in which credit standards were not tightened. Over 30 US states introduced an APL some time between 2000 and 2006 and this variation in the presence of APLs across states and time helps me to identify a causal link from lending standards to economic activity.

As a prelude to the experiment, I present some preliminary evidence that states with APLs experienced weaker consumer spending, and softer economic conditions more generally, than other states immediately after the lending laws were introduced (Figure 4.1). Over the 1990s real GDP grew at a similar rate for APL states (those that introduced the lending laws) and non-APL states (those that did not). But real GDP growth clearly slowed for the APL states relative to the non-APL states between 2000 and 2006, which coincides with the period in which each of the state APLs were introduced. A similar pattern emerges for the growth in real consumer spending (as measured by retail trade); APL states experience faster spending growth prior to 2000 but then experience slower growth through the 2000-06 period before rebounding again after 2007. The cross-sectional differences in consumer spending and economic activity do not appear to be directly attributable to wealth effects from changes in housing prices. In fact, housing prices and credit in APL and non-APL states display a very different cross-sectional pattern to that of GDP and retail trade, with



(a) Housing Prices



(b) Average Home Mortgage Loan Size

Figure 4.2: Housing Prices and Lending by APL and Non-APL States

the APL states experiencing *faster* housing price and credit growth relative to the non-APL states between 2000 and 2006 (Figure 4.2).

Another notable feature of the GDP and consumer spending graphs is that the relatively weak period of growth in the APL states disappears after 2007 as the performance gap between APL and non-APL states narrows. This seems to imply that the subprime mortgage crisis had less of an impact on the economic performance of the APL states than the non-APL states. Taken together, the observed fluctuations in GDP and consumer spending suggest that the tightening of credit standards reduced short-term growth but may have led to better medium-term growth in the APL states.

It is tempting to conclude that mortgage lenders in the more highly-regulated APL states originated fewer risky new loans than their counterparts in the non-APL states after the lending laws were introduced, and this reduction in risk-taking subsequently led to fewer mortgage defaults and improved economic performance in APL states. But we cannot infer a causal link between bank lending standards and economic growth by simply staring at these graphs. The observed patterns of GDP and consumer spending growth across APL and non-APL states may not be attributable to differences in residential mortgage credit standards, but due to some other unobservable factor(s). For example, APL states may have more 'vulnerable' populations, such as higher shares of subprime mortgage borrowers or minority households. This may make these states more sensitive to adverse economic shocks *and* more likely to have state policymakers that enact laws to mitigate high-risk lending. The relative slowing in real GDP growth in the APL states roughly corresponds to the nationwide recession in 2001. So it may be the combination of an adverse demand shock (the US recession) and the demographic composition of APL states that explains the observed relationship between APLs and economic growth. The observed patterns in housing prices and credit may lend support to this argument – the policymakers in APL states may have been more inclined to introduce the APLs as they knew their households were more exposed to rising house prices and debt in these states. In other words, this analysis is open to the criticism that there is some omitted variables bias.

The natural experiment is designed to deal with such endogeneity problems. In my research design, the counties in a state that introduces an anti-predatory lending law are effectively 'treated' to stricter lending standards vis-a-vis counties in states that do not introduce the law. More specifically, I compare the economic outcomes of *contiguous* counties separated by state borders where one state introduces the lending law (the "treatment group") and the other state does not (the "control group"). This "border county matching" methodology has at least two important benefits. Firstly, state border counties are immediately adjacent neighbours and so should be similar in both observable and unobservable characteristics. They should therefore follow similar economic growth paths in the absence of any change in lending laws. If this assumption holds, any differences in economic outcomes between the contiguous counties can be attributed to the introduction of the legislation and hence to differences in bank lending standards. Secondly, the decision of a US state policymaker to introduce a lending law may be influenced by the conditions that prevail in the *average* county of that state, but the decision is less likely to be influenced by conditions in a particular *border* county. In other words, by focussing on individual counties my methodology reduces the potential endogeneity of the lending laws.

By examining the effect of bank lending standard shocks I hope to shed new light on the long-standing research question of whether the supply of credit can affect economic activity independently of the price of credit. This line of research dates back to at least the "credit availability doctrine" (Bernanke and Gertler [1995]) of the 1950s. The underlying assumption of this research is that credit markets do not operate like other product markets; asymmetric (or imperfect) information between lenders and borrowers mean that lenders may choose to tighten lending standards and ration credit to marginal borrowers rather than raise interest rates when faced with greater credit risk (Stiglitz and Weiss [1981]). This is because higher loan rates can drive out the most creditworthy applicants (adverse selection) or elicit riskier behaviour by borrowers (moral hazard). If these marginal borrowers are financially constrained and reliant on banks for finance, the unexpected tightening of lending standards will restrict the availability of credit and reduce spending by borrowers. In other words, changes in lending standards can have a procyclical effect on the real economy.[50]

Macroeconomists' interest in 'non-price' credit shocks tended to wane after the 1980s. However, the subprime mortgage crisis has led to a revival of macroeconomic research on credit supply shocks, reflecting both the nature of the crisis and the increased availability of microeconomic banking data. The general consensus from this revitalised body of research is that the boom-and-bust cycle in the US housing market was directly related to the expansion of the riskier subprime mortgage market as well as a decline in lending standards within the subprime market (e.g. Demyanyk and Hemert [2008], Mian and Sufi [2009], Rajan et al. [2010], Keys et al. [2010], Mian and Sufi [2011], Dell'Ariccia et al. [2012]).[51]

[50] I will often use the terms "credit supply" and "credit standards" interchangeably, even though there is a subtle distinction; *credit supply* generally relates to the *quantity* of credit, while *credit standards* relate to the *quality* (or riskiness) of credit. Despite this distinction, there is a close association between credit standards and credit supply – to the extent that high (low) risk borrowers tend to borrow more (less) on average, a tightening (loosening) of credit standards will be equivalent to a decrease (increase) in credit supply. I will further discuss this in the context of my model.

[51] Until recently there had been little empirical research into the specific link between lending standards and the economy. Notable exceptions include studies showing that changes in aggregate commercial lending standards affect the level of credit and economic activity (Lown and Morgan [2006]), that business loan collateralization increases during recessions and decreases during booms (Asea and Blomberg [1998]), and that banks lower collateral requirements during expansions so that riskier borrowers are more likely to obtain credit (Jimnez et al. [2006]). Moreover, prior to the subprime mortgage crisis, there had been no research

As the opening quotes by the two most recent Federal Reserve chairmen indicate, the financial crisis has also stimulated policymakers' interest in bank credit standards. If changes in loan standards have the potential to destabilise the economy and contribute to boom-bust credit cycles then bank supervisors and/or central bankers may wish to offset the effect of the changes in standards either directly (e.g. by setting maximum loan-to-valuation ratios) or indirectly (e.g. through offsetting changes in policy rates or bank capital standards). As an example of this, the Hong Kong Monetary Authority (HKMA) has recently introduced direct controls on residential mortgage loan-to-valuation ratios, which are an important component of lending standards. In fact, lending standards are the main tool through which the HKMA manages capital flows and monetary conditions. A part of the motivation for this paper is to consider whether changes in US state government policy that were designed to limit risk-taking by home mortgage lenders actually had such a stabilising effect on the US economy.

To summarise, my paper makes the following contributions:

- **Motivation:** I contribute to a growing academic literature that identifies whether credit supply shocks affect the business cycle (e.g. Mian and Sufi [2009]). To the best of my knowledge this is the first paper to test whether shocks to mortgage lending standards affect housing market activity and, subsequently, regional economic growth.[52] Previous studies have focused almost exclusively on commercial lending standards (e.g. Lown and Morgan [2006]).

- **Methodology:** My "geographic matching" of US border counties is a methodological

examining the specific role of *residential mortgage* lending standards in affecting economic conditions, other than studies looking at whether banks discriminate against minority groups. The body of research has also been largely limited to studying the US economy. A recent exception to this rule is a study of the varying effect of lending standards across the Euro Area (Cappiello et al. [2010]).

[52]Choi [2011], who wrote his paper concurrently with mine, exploits a similar natural experiment. He finds that states with stricter lending laws had lower mortgage refinancing volume but exhibited no difference in home purchase mortgage volume. He also tests whether, by restricting mortgage refinancing, the laws affected household spending. He finds some limited evidence that the laws reduce household spending. However, his analysis uses a different methodology and is undertaken at the MSA-level. He also uses a different measure of household spending.

improvement on studies that estimate the effect of banking shocks on the local economy (e.g. Jayaratne and Strahan [1996]). These studies typically use the US state as their geographic unit of analysis, with very diverse US states forming the treatment and control groups. By instead focussing on contiguous border counties, my treatment and control groups are likely to be 'better matched', increasing the likelihood that any change in the relative difference between the treatment and control groups following a policy treatment is due to that specific policy. To the best of my knowledge there are only two other published papers that use the same border county matching methodology. Huang [2008] uses such an identification strategy to examine the effect of interstate branching laws on economic growth in the US. Dube et al. [2010] looks at whether state-enforced changes in minimum wages affect employment and earnings.[53]

- **Policy implications:** While its not the primary focus of this paper, I also consider the role of bank regulation in the financial crisis by examining the relative contribution of state anti-predatory lending legislation. In particular, I consider how the states that had tougher mortgage credit regulation fared post-crisis relative to states that had weaker regulation. If lax credit standards contributed to excessive household spending in the housing boom period, then states that imposed tougher restrictions on bank lending during this time may have fared better in the post-crisis period than states that did not impose such policies (or imposed relatively weak lending policies).

The paper is organised as follows. Section 2 discusses the institutional background and, in particular, state anti-predatory lending legislation in the US. Section 3 outlines the data used in the experiment. Section 4 considers the identification strategy and empirical specification of the econometric model. Section 5 discusses the key results. Section 6 considers some

[53]Similar border identification strategies have been used to explore the impact of state sales taxes on retail sales (Fox [1986]), state minimum wages on employment in the fast-food industry (Card and Krueger [1994]), state right-to-work laws on the location of manufacturing (Holmes [1998]), school quality on house prices (Black [1999]) and state foreclosure laws on average mortgage loan size (Pence [2006]). However, these papers compare economic outcomes across two strips of land on opposite sides of a state border and do not necessarily match contiguous counties. See Huang [2008] for a discussion of the methodological issues.

extensions and robustness tests. Section 7 concludes.

4.2 Institutional Background

4.2.1 State anti-predatory lending laws

The introduction of high-cost lending laws by different states at different points in time acts as the key 'shock' to lending standards in my model. As such, it is useful to consider the nature of these laws, and the type of lending that they are designed to rule out. Predatory loans are typically grouped into four categories (Pennington-Cross and Ho [2008]):

- **Loan flipping**: loans that are repeatedly refinanced within a short period of time. With each refinance, high fees are wrapped into the new loan amount, thereby reducing the equity left in the home.
- **Excessive fees and 'packing'**: fees that are added to the financed amount (wrapped) instead of being paid upfront.
- **Lending without regard for the ability to repay**: loans that are originated under terms that the borrower would never be able to meet.
- **Fraud**: appraisers and brokers conspire to inflate property values above the market price.

In response to reported abuses by mortgage lenders in the late 1980s, the US Congress enacted the first comprehensive statute – the Homeownership Equity Protection Act (HOEPA) – in 1994 to replace state mortgage regulations with federal safeguards against high-cost lending. This law, known as an "anti-predatory lending law" (APL), specifically regulates loan terms and conditions other than interest rates. It is intended to curb predatory practices in the subprime mortgage market while allowing the development of non-abusive mortgage lending (Li and Ernst [2007]).

However, the Federal legislation was generally perceived to not provide adequate protection to vulnerable borrowers because it applied to a very small segment of the subprime mortgage market (Pennington-Cross and Ho [2008]). The federal HOEPA covered only mortgages with very high annual percentage rates (i.e. 8-10 percentage points above rates on comparable Treasury securities) or very high prepaid points and fees (i.e. 8 per cent of the loan size).[54] Furthermore, the Federal HOEPA only applied to refinance and home improvement loans and not to mortgages for home purchase. Notably, the price thresholds were sufficiently high that the federal HOEPA covered less than one per cent of subprime home loans (McLean [2009]).

As a result, many states adopted their own individual "mini-HOEPA" laws to redress predatory mortgage lending (Ding et al 2011). The state APLs followed the HOEPA approach of restricting mortgage terms for loans above a price threshold, but the price thresholds were generally set *below* those in HOEPA. North Carolina was the first state to adopt a "mini-HOEPA" law in 1999. By 2007, more than 30 states had some type of APL in place (Bostic, Engel, McCoy, Pennington-Cross and Wachter 2008). In total, only six states – Arizona, Delaware, Montana, North Dakota, Oregon and South Dakota – had no mini-HOEPA statutes or other laws or regulations in place (Bostic et al 2008).

There is significant variation in state APLs in terms of their coverage, restriction and enforcement provisions (Bostic et al. [2008]). Although some states adopted the same level of coverage as the original HOEPA legislation, APLs in most states covered a wider range of high-cost mortgages. For example, North Carolina reduced the threshold from 8 percentage points above comparable Treasury security yields (the Federal level) to 5 percentage points. The North Carolina APL therefore covered a larger segment of the mortgage market than the Federal legislation. But not all states extended regulations in this way. For example,

[54] More specifically, high-cost loans are loans that exceed one of the following two "triggers": where the annual percentage rate (APR) at consummation exceeds the yield on comparable Treasury securities plus 8 (10) per cent for first-lien (subordinate-lien) loans or where the total points and fees exceeds the greater of 8 per cent of the total loan amount or $400 (which is subject to annual indexing). HOEPA also gave the Federal Reserve broad power to regulate mortgages below these price thresholds, but the Federal Reserve did not exercise that authority until 2008.

the laws in Indiana and Kentucky maintained the same price level as the Federal HOEPA (8 percentage points) but covered home purchase loans in addition to refinance and home improvement mortgage loans. Prohibited practices also varied widely across states in terms of the level of restrictions involved. For example, some states banned prepayment penalties altogether while other states only banned prepayment penalties after five years of the loan being originated. Enforcement provisions also varied considerably. For instance, some state APLs only provided for government enforcement while others gave injured borrowers the right to sue.

Early research on state anti-predatory lending laws examined how APLs affected the mortgage market, including the flow of credit, the cost of credit, and the choice of mortgage products. Some of these studies focused on one jurisdiction‘s law while others analysed national outcomes. Studies that focused on the first state anti-predatory lending law, in North Carolina, generally found that the subprime market in the state diminished in size as a result of the passage of the law (Keith Ernst and Stein [2002], Elliehausen and Staten [2004], Harvey and Nigro [2004], Quercia et al. [2005]). However, taken as a whole, the national data show no clear relationship between state laws and the cost or availability of subprime credit (Ho and Pennington-Cross [2006], Li and Ernst [2007], Pennington-Cross and Ho [2008], Elliehausen et al. [2008], Bostic et al. [2008]).

The lack of consensus on the effects of APLs may be due to two important factors. Firstly, most studies do not control for the difference between states that extended restrictions on subprime mortgages *beyond* federal requirements and states that simply copied existing federal HOEPA restrictions; there should be no discernible effect of the state lending laws on the mortgage market in those states that just copied existing federal laws (Ding et al 2011). Secondly, in addition to examining overall credit flows, it is important to consider the impact of the lending laws *within* each state's subprime market. Since APLs were intended to reduce the number of predatory or abusive subprime loans *but* also stimulate non-abusive subprime lending, the overall size of the subprime market need not change; though loan quality should improve (Keith Ernst and Stein [2002], Quercia et al. [2005], Bostic et al. [2008]).

In the aftermath of the subprime crisis more recent research on APLs has focussed on how the presence of state laws affects mortgage foreclosure rates. A number of loan product features, such as prepayment penalties and very high interest rates and fees, are known to be associated with relatively high default risk (Calhoun and Deng [2002], Quercia et al. [2005], Ambrose et al. [2005], Immergluck [2009], Pennington-Cross and Ho [2010]). To the extent that APLs rule out these loan product features, there is some evidence that tougher APLs are associated with lower default risk and therefore lower state foreclosure rates (Goodman and Smith [2010]). In a later section of this paper, I take this idea a step further by examining whether the states with APLs have fared better post-crisis in terms of economic growth.[55]

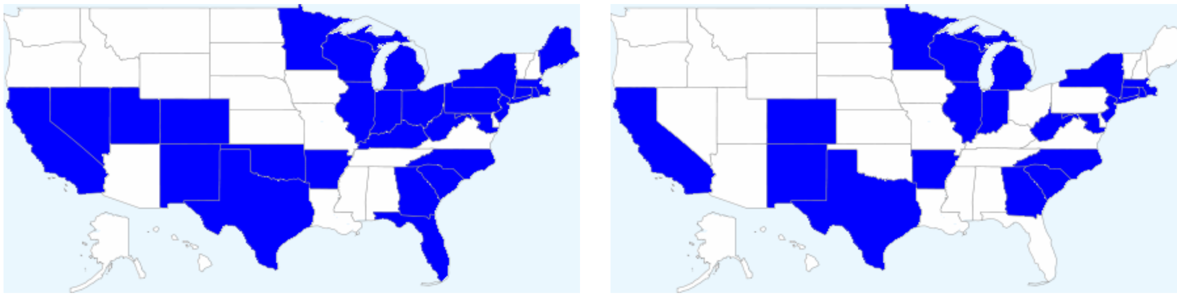
Table 4.1 includes a list of all states that currently have an 'effective' APL in place and the year in which the law was introduced; a state APL is considered to be effective when the state law is sufficiently different to the original Federal HOEPA legislation (Ding et al. [2011]).

Table 4.1: Year that State Anti-Predatory Lending Law is Introduced

Year of Introduction				
2000-01	**2002**	**2003**	**2004**	**2005-06**
North Carolina	California	Arkansas	Illinois	Indiana
West Virginia	Connecticut	Colorado	Massachusetts	Wisconsin
Texas	DC	Georgia	New Mexico	Rhode Island
	Maryland	Minnesota	South Carolina	
	Michigan	New Jersey	Oklahoma	
	Florida	New York	Utah	
	Ohio	Kentucky		
	Pennsylvania	Nevada		

Figure 4.3 illustrates the geographic clustering of state anti-predatory lending laws and compares states that have an anti-predatory lending law to those that have an effective anti-

[55]It should be noted that, as part of the Dodd-Frank Wall Street Reform and Consumer Protection Act, the US enacted the Mortgage Reform and Anti-Predatory Lending Act in 2010. This federal legislation established anti-predatory lending laws that were national in scope and were designed to replace the various state-level lending laws. As such, *every* US state should have been effectively treated to stricter mortgage laws by 2010. However, I find that my results are barely affected by the inclusion of the 2010 data in the sample.



(a) All State Anti-Predatory Lending Laws (b) Effective State Anti-Predatory Lending Laws

Figure 4.3: Anti-Predatory Lending Laws by State

predatory lending law (i.e. a state law that is sufficiently different to the Federal law). The states with laws are indicated by the darker shades.

4.3 Data

4.3.1 County-level measures of economic activity and consumer spending

To be able to conduct the natural experiment I need a measure of county-level economic activity. This choice is limited by data availability. For instance, GDP and consumer spending data are not readily available at the US county level. I rely on two specific measures of economic activity – real personal income growth across all industries and real personal income growth in the retail trade industry. Real personal income growth is an income-based measure of GDP that uses the total earnings of all employees to approximate economic activity. Real retail trade personal income growth is a similar measure of total earnings of employees in the retail trade sector.[56] The advantages of using real personal income

[56]The Bureau of Economic Analysis (BEA) officially defines *personal income* as the income received by all persons from all sources. Personal income is the sum of net earnings by place of residence; dividends interest, and rental income (property income) of persons; and personal current transfer receipts. Net earnings is earnings by place of work (the sum of wage and salary disbursements (payrolls), supplements to wages and salaries, and proprietors" income) less contributions for government social insurance, plus an adjustment to convert earnings by place of work to a place-of- residence basis. Personal income is measured before the deduction of personal income taxes and other personal taxes and is reported in current dollars (no adjustment is made for price changes).

growth is that it is the broadest indicator of county-level economic performance available, is highly correlated with real GDP growth at the national, state and metropolitan statistical area (MSA) levels, and is most commonly used in other studies that examine the effect of banking shocks on local economic growth (e.g. Huang [2008]). The disadvantage of using real personal income growth is that it is not closely aligned with the theory of how mortgage credit standards shocks affect the economy. As my model suggests, the effect of credit standards shocks should specifically come through a consumer borrowing and spending channel. This is the reason why I also look at retail trade income growth – fluctuations in retail trade income are very highly correlated with fluctuations in consumer spending at the national, state and MSA levels, which suggests that it is likely to be a good proxy for consumer spending.[57]

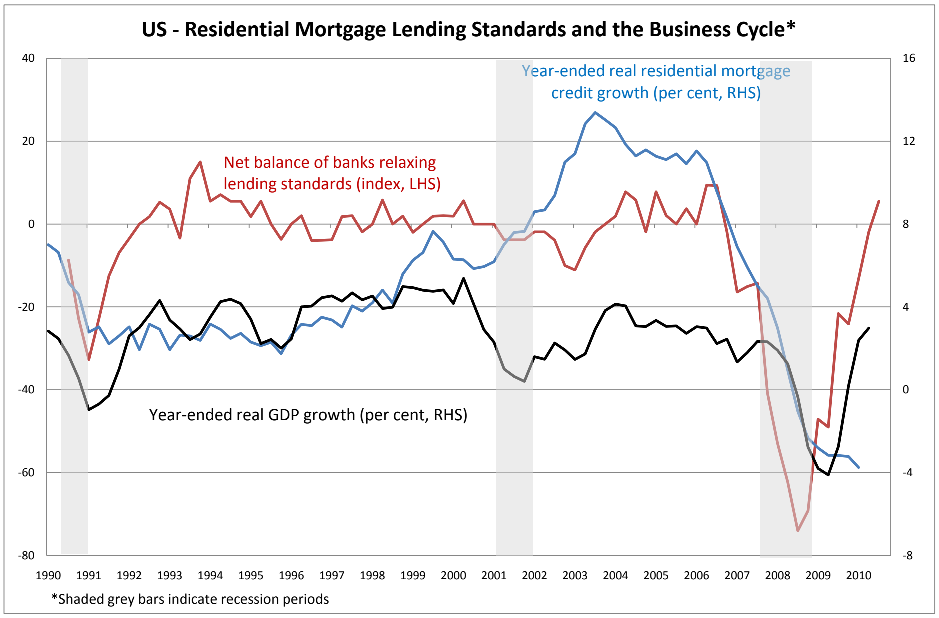
4.3.2 Survey-based measures of bank lending standards

In this section, I discuss the various available measures of US bank lending standards. One of the reasons for the limited academic research into lending standards is because lending standards are difficult to define. I broadly define lending standards to be a "vector of non-price loan criteria" that are set in a contract between a lender and a borrower (Lown and Morgan [2006]). Lending standards are separate from the interest rate and encompass a wide range of loan terms, such as the loan to valuation ratio, collateral, the term to maturity and other loan limits. Another reason for the lack of research into lending standards has been the limited data availability. The most comprehensive (and certainly best-known) measure of aggregate credit standards in the US economy is the Federal Reserve's Loan Officer Survey. Since 1990 economists at the Federal Reserve (FRS) have asked a sample of senior loan officers at large banks supervised by the Federal Reserve whether their bank

[57] The correlation between real personal income growth and GDP is 0.52 at the MSA level over the period 2001-2010. For the retail trade sector, the correlation between real personal income growth and GDP is even higher at 0.87 over the corresponding period. This suggests that personal income growth is a relatively good proxy for business cycle conditions. As a cross-check on these measures of local economic growth, I also estimated the models using two labour force indicators – total employment growth and employment growth in the retail trade industry – as proxies for spending. I found very similar results to those presented in the paper.

has tightened mortgage lending standards or not.[58] The net percentage of lenders reporting weaker standards derived from this loan officer survey is clearly countercyclical, with banks tightening standards in recessions and relaxing them in booms (Figure 4.4).[59]

Figure 4.4: Federal Reserve Senior Loan Officer Survey



Despite a general consensus that lending standards were being progressively relaxed during the housing credit boom of the early 2000s, there is no compelling evidence of a decline in standards in the Federal Reserve survey. This is likely to reflect the fact that this survey only covers large national commercial banks that are supervised by the Federal Reserve, which typically served only a small part of the subprime mortgage market during the pre-crisis period. In other words, the survey does not cover non-depository institutions, such as independent mortgage companies, that are generally cited as the 'main culprits' for the

[58]Federal Reserve economists ask loan officers the following question: "over the past three months, how have your bank's credit standards for approving loan applications for mortgage loans changed?" 1) Tightened considerably 2) tightened somewhat 3) remained unchanged 4) eased somewhat 5) eased considerably.

[59]Another bank supervisor, the Office of the Comptroller of the Currency (OCC), conducts a similar (annual) survey of lending standards which covers national mortgage lenders under its own supervision. There is a very similar counter-cyclical pattern to credit standards in this survey too.

decline in credit standards during the housing boom (e.g. Immergluck [2009]).

4.3.3 Loan-level based measures of bank lending standards

I do not rely on the Federal Reserve survey to construct my measure(s) of lending standards. This is partly because the Federal Reserve survey sample is restricted to a particular segment of the mortgage market, but also because my identification scheme relies on regional variation in credit standards, and the Federal Reserve does not make the disaggregated survey data available to external researchers.

I instead rely on *loan-level* indicators of lending standards. Specifically, I derive my measures of mortgage lending standards from the annual Home Mortgage Disclosure Act (HMDA) Loan Application Registry. Enacted by Congress in 1975, HMDA requires mortgage lenders located in metropolitan areas to collect data about their housing-related lending activity and make these data publicly available. These data are generally considered to be the most comprehensive source of mortgage data, and cover an estimated 80 per cent of all home loans nationwide and a higher share of loans originated in metropolitan statistical areas (Avery et al. [2007]).[60] The data that are available in the HMDA dataset are listed in Table 4.2.

Table 4.2: Home Mortgage Disclosure Act Data

HMDA Information			
Lender	**Property**	**Loan**	**Applicant**
Bank ID	Census tract	Action (e.g. loan approved)	Race/ethnicity
Regulator	County	Amount	Gender
	MSA/MD	Type (e.g. owner-occupier)	Income
	State	Purpose (e.g. home purchase)	Family structure
		Property type	

For each application there is information on the borrower (e.g. income, ethnicity), the property (e.g. zipcode location) and the lending institution (e.g. the name of the lender).

[60] Whether a lender is covered depends on its size, the extent of its activity in a metropolitan statistical area (MSA), and the weight of residential mortgage lending in its portfolio.

Most importantly for my purposes, the HMDA data separately identify mortgage loan applications and originations.[61] This difference allows me to identify whether a mortgage application is approved or not and, if the application is approved, the size of the loan relative to the income of the borrower. This gives me my measures of credit standards:

- **Average loan size**: the average size of loan originated by each lender in each US county in each year.
- **Average loan-to-income ratio**: the average size of loan originated (as a share of the applicant's annual income) by each lender in each US county in each year.
- **Loan approval rate**: the proportion of loan applications approved by each mortgage lender in each US county in each year.

While these measures do not fully encapsulate the broad definition of credit standards as *any* non-price lending terms, they are very important components of credit standards. We can loosely think of the loan approval rate as being the 'intensive margin' and the average loan size as the 'extensive margin' of the lending decision. I also measure credit standards through the average loan-to-income (LTI) ratio, as the LTI is a more common measure of lending standards than simply the average loan size.[62]

For the purposes of my experiment, it would be better to use information on subprime mortgage lending rather than total mortgage lending because the APLs are designed to rule out the most risky loans within the subprime market. But, for most of the sample period, HMDA does not require lenders to report information (such as loan interest rates or points) that might be used to identify subprime loans. For these reasons, I estimate the models based on the overall mortgage market rather than just the subprime market.

[61]There is also information on the reasons why each loan was denied (e.g. the borrower's credit history, insufficient down-payment, employment history etc).

[62]Ideally, I would like a measure of the loan-to-valuation ratio (LTV) rather than the loan-to-income ratio but the HMDA data do not include the value of the home. Notice also that the loan approval rate can be an imperfect measure of credit standards. This is because lenders often undetake aggressive marketing campaigns when they relax their lending policies, which can cause the number of loan applications to rise faster than the number of originations, implying a *fall* rather than a rise in the loan approval rate when banks relax credit standards.

4.3.4 Control variables

I also include control variables to provide additional explanatory power and to increase the number of observable factors that are likely to be associated with economic activity which should minimise the likelihood of endogeneity in the panel regression model. The control variables include annual growth in the county-level population (population growth) and annual growth in state-level housing prices (house price growth) as measured by the Federal Housing Association.[63]

4.3.5 Sample

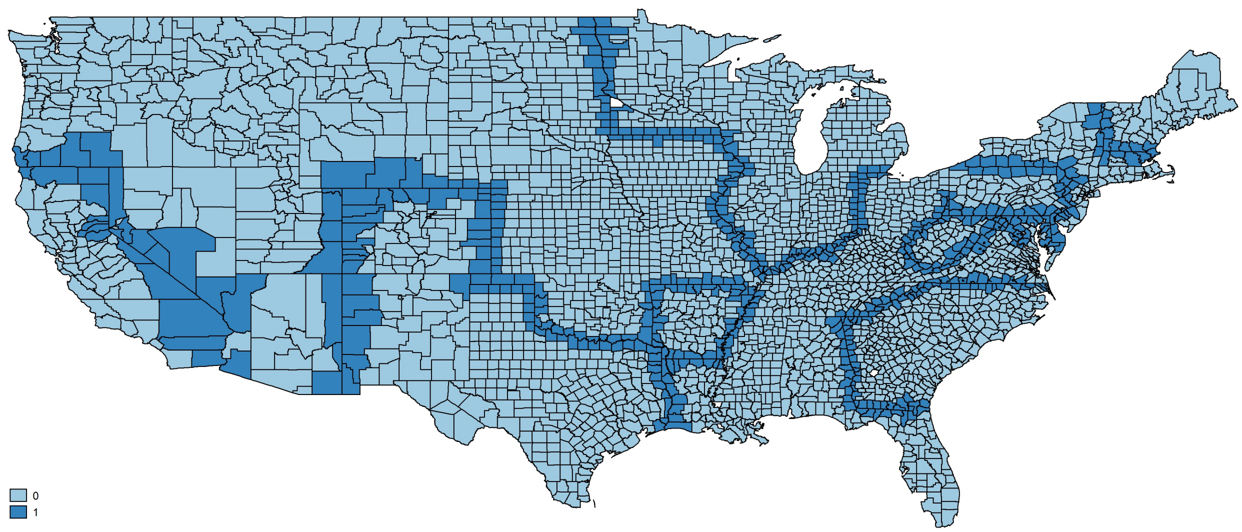
The HMDA sample is very large, covering almost half a billion loan applications over the period 1990-2010. The two-decade sample period covers several regional housing cycles as well as the national boom-bust cycle of the 2000s. I focus only on 'pairs' of matched border counties - counties that share a common state border but are on opposite sides of that state border. There are more than 1100 such border county pairs in the US, but my sample consists of the 542 *treated* border county pairs. These are the border pairs in which one of the two counties is in an APL state and the other is in a non-APL state. In other words, around half of all US border pairs are "never treated" because it is either the case that a) neither of the two states introduced an APL or b) both states introduced an APL at some time. The geographic distribution of the sample of treated border counties is shown in Figure 4.5 (the treated county pairs are shown in the dark blue colour).

The benefit in using border counties as the geographic unit of analysis is that the decision by state legislators to introduce an APL could be affected by characteristics of the average county in that state but is much less likely to be affected by conditions in specific border counties, especially border counties that are very different to the average county.[64] Impor-

[63] I have also experimented with including other explanatory variables, such as county-level measures of the share of the household population that is of African American or Hispanic origin, the share of the population that is aged below 16 or above 65 years and the manufacturing share of total employment. The results of estimating these models are both qualitatively and quantitatively similar.

[64] Admittedly, this may not be true for states with few and/or very large counties. To test whether this is

Figure 4.5: Sample of Treatment and Control Border Counties



tantly, each border county is more similar to its matched 'twin' than to the average county in its own US state on almost all dimensions of my dataset. For instance, the (absolute) personal income growth differential between each border county and the average county in its state exceeds the (absolute) personal income growth differential between each border county and its matched twin by around 1.7 percentage points, on average. This gives me confidence that the border counties are relatively homogeneous.

4.4 Identification

In this section I outline the difference-in-difference (DD) regression models to be estimated on the matched border counties. I am interested in assessing the impact of the lending standards shocks (as represented by the anti-predatory lending laws) on the real economy. However, before doing so, I want to provide evidence that the laws actually are associated with changes in credit conditions. I therefore separately outline the effect of APL shocks

a problem, in unreported results, I re-estimated the model excluding states with only a few large counties. I found qualitatively similar results. Another problem with the border county methodology is that large cities that are not near state borders are not included in the estimation. This will bias the results if the reaction to a state anti-predatory lending law is systematically different in the excluded cities.

on credit standards (the credit standards model) and the effect of APL shocks on economic activity (the economic activity model) as there are subtle differences between the two models. The main difference is that I estimate the credit standards model at the lender-county level while I estimate the economy activity model at the county level. So the credit standards model uses more disaggregated data than the economic activity model. I can more precisely gauge the effect of APL shocks on credit standards because credit standards vary by county, time *and* lender while economic activity only varies by county and time.

I begin with the most general specification of each regression model and then outline the necessary steps in each case to get to the estimating equations. To begin with, notice that my dataset spans the period 1990-2010 so I have many years of observations both before and after the policy treatment because all the APLs were introduced between 2000 and 2005. But difference-in-differences estimation that uses many periods of data and focuses on serially correlated outcomes has been shown to produce inconsistent standard errors (Bertrand et al. [2004]). Before estimation I therefore follow the prescription of Bertrand et al. [2004] and collapse the time-series information into two periods – the "pre-treatment" and "post-treatment" periods – by taking the average of all observations before and after the APL is introduced. Collapsing the data smoothes out variation and generates conservative standard errors.

4.4.1 The credit standards model

I begin with the general specification of the credit standards model:

$$L_{ipst} = \underbrace{APL'_{it}\alpha + X'_{ipst}\beta}_{\textit{Observable factors}} + \underbrace{\theta_{ip} + \theta_s + \lambda_{pt} + \epsilon_{ipst}}_{\textit{Unobservable factors}}$$

Where i indexes either the control or treatment county ($i = C, T$), p indexes the border county pair, s indexes the mortgage lender and t indexes the pre or post-treatment period ($t = PRE, POST$). The LHS variable (L_{ipst}) is a measure of credit standards (i.e. the loan approval rate or loan size or loan-to-income ratio) for each bank (s) in each county (i) in each matched border pair (p) in either the pre or post-treatment period (t). The RHS variables include both observable and unobservable factors. The observable factors include the following:

- **Policy variable (APL_{it})**: This is a binary indicator that captures the effect of the anti-predatory lending law. The indicator is essentially an interaction dummy which is equal to unity if the lender and county are in the treatment group (i.e. an APL state) in the post-treatment period (i.e. after the introduction of the APL) and is equal to zero otherwise. The policy variable varies between the treatment and control counties *within* each matched border pair and across time but *does not* vary by lender or border pair.

- **Lender-county-specific covariates (X_{ipst})**: these are observable variables that vary along each dimension of the data. In other words, factors that vary by county and border pair and mortgage lender and time, such as the average household income of loan applicants that approach a particular lender.

The unobservable factors include:[65]

- **County fixed effects (θ_{ip})**: variables that capture county-specific factors that do not vary with time such as culture and geography.

- **Mortgage lender fixed effects (θ_s)**: capture bank-specific factors that are constant over time and across regions, such as bank risk preferences.

- **Border pair-period fixed effects (λ_{pt})** capture border-specific time trends. For example, some state border pairs may be more sensitive to macroeconomic shocks than others due to their demographic or industrial structure.

[65] Note that there are a number of higher-order interaction effects I could include in this regression, but I choose to ignore them as they are likely to be of lesser importance and will simply soak up degrees of freedom. In any case, I effectively eliminate any higher-order interactions that are time invariant as my estimation method requires first differencing the data.

- **Lender-county-period-specific errors (ϵ_{ipst})**: these are factors that vary along each dimension of the data and are assumed to be white noise.

I eliminate all unobservable time-invariant factors by taking the difference in each variable over time between the pre and post-treatment periods:

$$\triangle L_{ips} = \triangle APL_i'\alpha + \triangle X_{ips}'\beta + \triangle\lambda_p + \triangle\epsilon_{ips} \tag{4.1}$$

The key policy variable ($\triangle APL_i$) becomes an indicator variable equal to one if the border county is in the treatment group and equal to zero if the border county is in the control group. The average treatment effect (ATE) is captured by the estimate of α which is the main coefficient of interest. This is essentially a difference-in-differences model where I also include pair-specific dummies ($\triangle\lambda_p$) to control for border pair-specific trends in lending conditions and condition on observable variables including trends in house prices and average household income given by $\triangle X_{ips}$. The key identifying assumption to obtain a consistent estimate of the average treatment effect (α) is that the change in treatment status within each county pair is uncorrelated with the change in the error term (i.e. $E(\triangle APL_i'\triangle\epsilon_{ips}) = 0$). This condition could be violated if there are certain characteristics of the treatment counties that make them more likely to receive treatment than the corresponding control counties on the other side of the state border.[66] In other words, there may be reasons to believe that the counties are not randomly assigned to treatment because, for instance, the treated counties are in states that have a history of consumer protection activism. One of the benefits of using the matched border county methodology is that the policy decision to 'treat' a state may be influenced by conditions in the state's *average* county, but is less likely to be influenced by conditions in a *border* county. In other words, I argue that the policy treatment is 'as good as randomly assigned' once I condition on other observable factors and eliminate unobservable

[66] To see this more clearly we could represent the error term as $\epsilon_{ipst} = \theta_{it} + \mu_{ipst}$ where θ_{it} is an unobservable shock that varies by treatment status and year, such as state policymakers' willingness to introduce lending laws, and μ_{ipst} is a white noise error term. The identification condition then implies that $E(\triangle APL_i'\triangle\theta_i) = 0$ which will not hold if treatment assignment depends on the willingness of state policymakers to change policy.

factors that could affect the policy decision.

4.4.2 The economic activity model

To understand the regression model for the effect of APL shocks on economic activity I again begin with the most general specification of the model:

$$Y_{ipt} = \underbrace{APL_{it}'\alpha + X_{ipt}'\beta}_{\text{Observable factors}} + \underbrace{\theta_{ip} + \lambda_{pt} + \epsilon_{ipt}}_{\text{Unobservable factors}}$$

Here the LHS variable (Y_{ipt}) is a measure of economic conditions (e.g. real personal income growth) in each county i in each matched border pair p in either the pre or post-treatment period t. In other words, the data now vary by county and year but not by lender because it is not possible to obtain lender-specific measures of economic activity. The RHS variables are very similar to those in the credit standards model. For instance, the policy variable (APL_{it}) is again a binary indicator equal to unity if the county is in the treatment group in the post-treatment period and is equal to zero otherwise.

I proceed by eliminating many of the unobservable variables by taking the (first) difference between the treated and control county *within* each border pair at a given point in time (i.e. I take the 'within pair' difference) which eliminates the unobservable border pair-time effects (λ_{pt}). This procedure also implicitly eliminates any unobserved period effects that are common to all counties, such as macroeconomic shocks.

$$\underbrace{(Y_{Tpt} - Y_{Cpt})}_{\hat{Y}_{pt}} = \underbrace{(APL_{Tt} - APL_{Ct})'}_{A\hat{P}L_t}\alpha + \underbrace{(X_{Tpt} - X_{Cpt})'}_{\hat{X}_{pt}}\beta + \underbrace{(\theta_{Tp} - \theta_{Cp})}_{\hat{\theta}_p} + \underbrace{(\epsilon_{Tpt} - \epsilon_{Cpt})}_{\hat{\epsilon}_{pt}}$$

Where a variable with a 'hat' denotes the within pair-year difference between the treated and control county (i.e. $\hat{x}_{pt} = (x_{Tpt} - x_{Cpt})$). I then take the (second) difference over time between the pre and post treatment periods which eliminates the pair-specific effects ($\hat{\theta}_p$):

$$\underbrace{\hat{Y}_{p,post} - \hat{Y}_{p,pre}}_{\triangle \hat{Y}_p} = \underbrace{(A\hat{P}L_{post} - A\hat{P}L_{pre})}_{\triangle A\hat{P}L}{}^{'} \alpha + \underbrace{(\hat{X}_{p,post} - \hat{X}_{p,pre})}_{\triangle \hat{X}_p}{}^{'} \beta + \underbrace{(\hat{\epsilon}_{p,post} - \hat{\epsilon}_{p,pre})}_{\triangle \hat{\epsilon}_p}$$

The resulting equation to be estimated is again a difference-in-difference (DD) panel regression:

$$\triangle \hat{Y}_p = \triangle A\hat{P}L^{'} \alpha + \triangle \hat{X}_p^{'} \beta + \triangle \hat{\epsilon}_p \tag{4.2}$$

After these manipulations, the policy variable ($\triangle A\hat{P}L$) becomes a constant. The coefficient for the average treatment effect (α) captures any change in the trend growth rate of the treatment counties relative to the control counties, conditional on other observables such as population growth and house price growth ($\triangle \hat{X}_p$). The identifying assumption to obtain a consistent estimate of the average treatment effect is that the change in treatment status is uncorrelated with the change in the error term (i.e. $E(\triangle A\hat{P}L^{'} \triangle \hat{\epsilon}_p) = 0$). In other words, the difference in the law within each county pair is uncorrelated with the difference in residual economic activity between the two counties. This condition is essentially the same as in the credit standards model. The identifying assumption would be violated if, for example, there are certain characteristics of the counties in the treatment states that made them more likely to receive treatment than the corresponding (matched) counties in the control states. To reiterate, the major advantage of using variation between contiguous counties within a matched border pair is that the policy decision should not be endogenous to economic conditions in the border counties.

For the OLS estimate of the APL effect to also be efficient I require there to be no serial correlation or heteroscedasticity in the error ($\triangle \hat{\epsilon}_p$). But OLS standard errors are potentially subject to three distinct sources of bias in this setting (Dube et al. [2010]). Firstly, there is positive serial correlation in economic growth at the county level. Secondly, there is positive serial correlation in the treatment effect (APL) within each state. Thirdly, the presence of a single county in multiple pairs along a border segment means that different border county pairs can be correlated. In other words, the residuals are not independent if the border counties are within the same state (within-state correlation) or if two different county pairs are within the same border segment (within-border correlation). I am able to partly deal with these correlations by collapsing the data into the pre and post-treatment periods. To account for any remaining correlation within the residuals, I cluster the standard errors for the estimates on the state and border segment separately.

4.5 Results

4.5.1 Graphical analysis

Before turning to the results, consider the underlying data to be used in the estimated regressions. Figure 4.6 displays the relative trends in the various measures of credit standards over time for the treatment and control border counties. Figure 4.7 shows the equivalent patterns for the measures of local economic growth and consumer spending (total personal income growth and growth in retail trade personal income), as well as two labour market indicators (total employment growth and retail trade employment growth) and two control variables (state-level house price growth and county-level population growth). Notice that the data have been normalised so that the APL shock occurs in year 'zero' which allows us to most clearly see how credit standards and economic growth evolve both before and after the policy treatment. Also, recall that the APLs were introduced in different states at different times – the data presented here are effectively aggregated over all border counties and time periods before and after the APL shock, so they 'mix together' different states and years. For example, the year 'zero' corresponds to the year 2002 for counties on the border of California but the year 2005 for counties on the border of Indiana. This minimises the impact of structural factors (e.g. geography) and macroeconomic factors (e.g. monetary policy) that could explain the relative trends in credit standards and economic conditions.

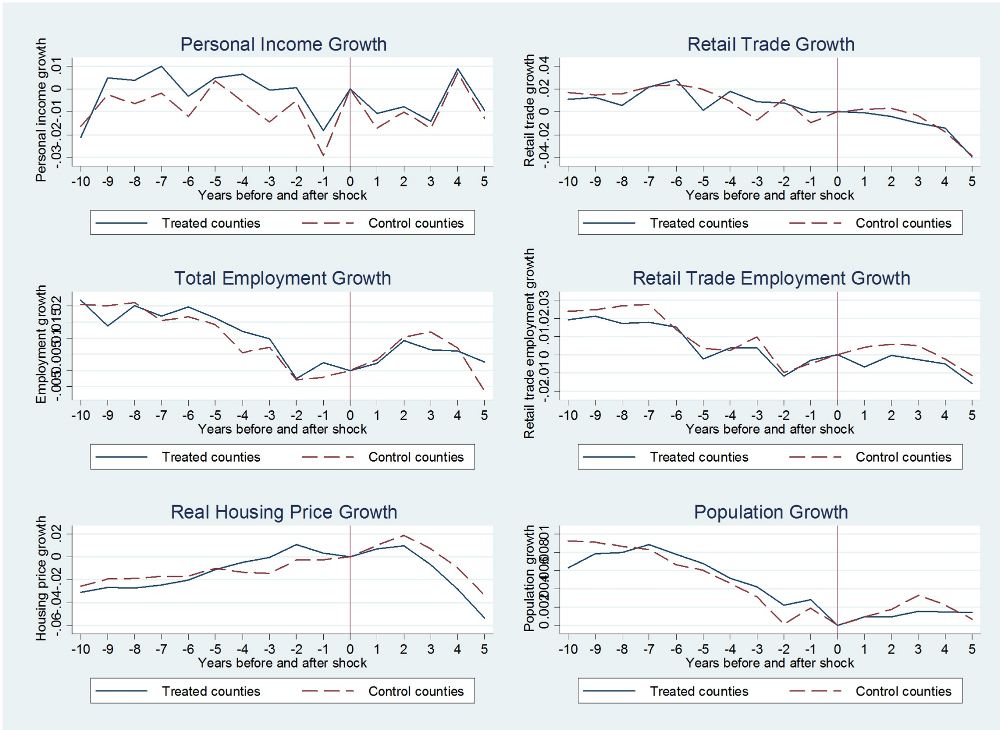
Figure 4.6: Time Series of Credit Standards in Border Counties

Clearly, both the average loan size and the loan-to-income ratio display a similar pattern across the treatment and control counties. Both measures are rising during the period prior to the APL shock, reflecting the nationwide boom in house prices over the late 1990s and early 2000s. It also appears that the average loan is rising more rapidly in the APL states than in the non-APL states during this period. After the APL shock, both the average loan size and the loan-to-income ratio grow at a slower pace across all counties than before the shock. More importantly, the lending measures grow more slowly in the treatment counties relative to the control counties. This suggests the APL effectively tightened lending standards and reduced the extent of lending in the treatment counties as compared to the control counties.

Prior to the introduction of the APL the loan approval rate was generally declining across both the treatment and control counties which, on the surface, appears inconsistent with the general consensus that there was a nationwide relaxation of lending standards through the late 1990s and early 2000s.[67] But, the most important aspect of the figure is that, after the

[67]However, the fall in the aggregate loan approval rate during the boom period could be explained by two

Figure 4.7: Time Series of Consumer Spending and Economic Activity in Border Counties



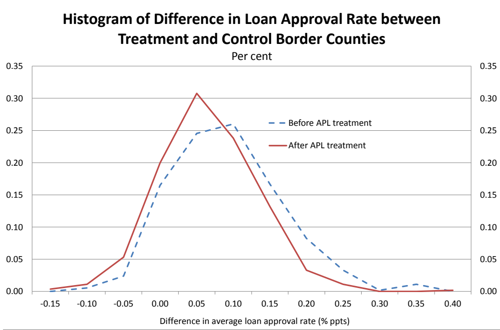
APL shock, the loan approval rate rises by less in the treatment counties than in the control counties which is consistent with the hypothesis that the APLs tightened lending standards.

For completeness, I also show the evolution of total mortgage lending volume (i.e. the average loan size multiplied by the number of loans). The overall pattern is fairly similar to the other measures of lending conditions, though it is more difficult to detect different trends in mortgage volume for APL states relative to non-APL states.

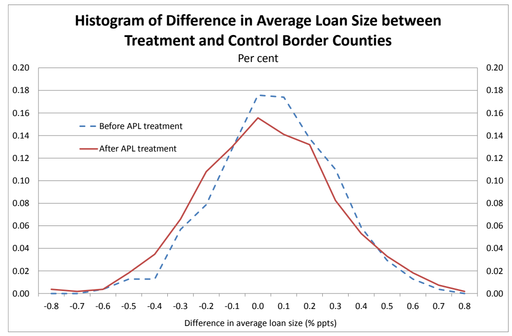
Next, consider the trends in total personal income growth and retail trade income growth across counties (Figure 4.7). Before the APL shock, personal income is growing at a faster pace in the treatment counties than in the control counties. However, soon after the APL shock, the gap in economic performance between the treatment and control counties narrows

factors. Firstly, mortgage lenders engaged in aggressive marketing of new loans during the housing boom which may have caused applications to rise more quickly than originations. As the number of applications is in the denominator of the loan approval rate, this would mechanically lead to a lower approval rate, all other things being equal. Secondly, if the marginal loan applicant is risky, then the aggressive marketing push may have contributed to an increase in the riskiness of potential new lending and further contributed to fewer applications being approved.

Figure 4.8: Differences in Credit Standards Between Treatment and Control Counties



(a) Loan Approval Rate



(b) Average Loan Size

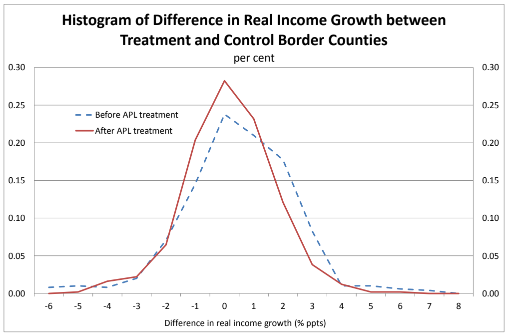
as personal income grows more slowly in the treatment counties relative to the control counties. This is consistent with the hypothesis that the APL causes tighter lending standards which, in turn, leads to weaker growth in economic activity. A similar story emerges for retail trade income growth – retail trade income is growing more quickly in the treatment counties than in the control counties before the treatment, but this growth pattern reverses soon after the APL shock with the control counties growing more quickly than the treatment counties.

However, a similar pattern also emerges for both growth in state house prices and in population growth in the bottom row of the panel which may suggest that differences between the treatment and control counties in these variables, or some other omitted variable, are driving the differences in economic activity rather than the APL shock itself. It will therefore be important to control for these factors in the regressions.

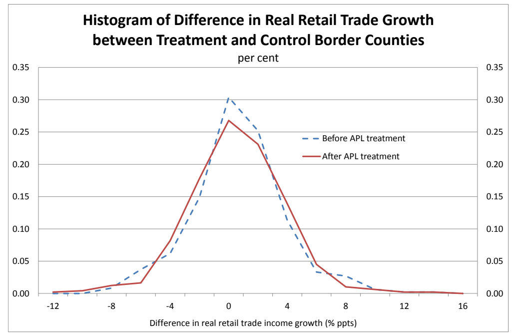
An alternative way to measure the effect of the APL treatment shock is to collapse the time-series information on credit standards and economic activity into the pre- and post-treatment periods for both the treatment and control groups and then look at the cross-sectional distribution of average differences between the treatment and control counties before and after the APL shock (Figures 4.8 and 4.9). To be clear, consider the example of the loan approval rate. I firstly calculate the average loan approval rate for each of the treatment and control counties in both the pre and post-treatment periods (so I have four estimates in total for every matched county pair). I then calculate the difference in the average loan approval rate between the treatment and control counties within each matched pair in the pre and post-treatment periods (which gives me two estimates in total for every county pair). The histogram is then a graphical representation of how the differences between the treatment and control counties are distributed across the matched county pairs in the two periods. An estimate of zero implies that the two counties within each pair are growing at the same rate, on average. So we can see how the difference in outcomes between treatment and control counties changes after the policy treatment by comparing the histograms before and after treatment. Note that I measure the average difference as the outcome in the treatment county less the outcome in the control county, so a positive differential implies that the treatment county has a higher outcome (e.g. higher economic growth) than the control county.

Firstly, consider the histograms of credit standards. We can see a clear leftward shift in the histograms for both the loan approval rate and the average loan size after the APL shock. This is consistent with the hypothesis that both the average mortgage loan size and the loan approval rate fall in the treated counties relative to the control counties after the APL shock. Secondly, consider the histograms of economic activity. Here there is a clear leftward shift in the real personal income growth distribution after the APL shock, implying that growth in the treatment counties falls relative to the control counties. This, again, is consistent with tighter lending standards being associated with weaker economic conditions. In contrast, there is no discernible shift in the histogram of growth in retail trade personal income. If anything, the tails of the distribution simply become 'thinner' after the policy treatment, so this potentially argues against any effect of the credit standards shock on household spending.

Figure 4.9: Differences in Economic Activity Between Treatment and Control Counties



(a) Total Personal Income Growth



(b) Retail Trade Personal Income Growth

4.5.2 Econometric analysis

Given the limitations of this graphical analysis, we need to turn to the regression models to uncover the relationships between credit standards and economic conditions. The other benefit of the regression models is that we can test whether the effect of the state APLs on economic growth is statistically significant. I firstly consider the effect of the APLs on my measures of credit standards – the average loan size, loan-to-income ratio and loan approval rate. I then show the effect of the APLs on my measures of economic activity – personal income across all industries, as well as in the retail trade industry.

The average treatment effects (and statistical significance levels) from the border county panel regression for the credit standards variables are shown in Table 4.3. The average treatment effects are shown in the first row of the table. Clearly, the introduction of an APL results in both lower loan sizes and lower loan-to-income ratios, on average, which is consistent with the graphical analysis presented above. Moreover, the effect is statistically significant in both cases. The average treatment effects also appear to be economically significant, as the introduction of an APL causes the average loan size to fall by 1.7 per cent (column 1) and by 7.5 per cent relative to income (column 2) in treated counties relative to control counties. The APL shock also appears to lower the average loan approval rate, but the effect is not statistically significant.

Table 4.3: The Effect of an APL Shock on Credit Conditions

	(1) Average Loan Size	(2) Loan to Income Ratio	(3) Loan Approval Rate
APL	-0.0173***	-0.0752***	-0.00129
	(-9.28)	(-3.48)	(-0.91)
House Prices	0.651***	1.173***	-0.0534***
	(9.34)	(5.53)	(-4.02)
Household Income	0.407***	-1.193***	0.0561***
	(18.27)	(-7.83)	(10.26)
Minority Ratio	0.423**	1.520*	0.0126
	(2.48)	(1.81)	(0.19)
Constant	0.0669***	0.115**	-0.0216***
	(3.39)	(2.66)	(-5.97)
R^2	0.152	0.030	0.019
Observations	69030	67750	77758

t statistics in parentheses

Notes: Standard errors are clustered at both the US state and mortgage lender levels.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, I briefly consider the effect of the control variables. Firstly, higher house prices are associated with larger mortgages and higher mortgage loan-to-income ratios, on average. More surprisingly, higher house prices are associated with a lower probability of being approved for a loan. Secondly, higher average household incomes are associated with higher loan approval rates and loan sizes, but associated with lower loan-to-income ratios. Taken together, these results imply that high-income households take out larger loans than low-income households, on average, but that high-income households take out smaller loans *relative to their income* than low-income households.

Now we can assess the average treatment effect on the various measures of economic activity. Firstly, we see that the introduction of a state APL is associated with slower real personal income growth, on average (Table 4.4). The OLS estimate of the average treatment effect for real personal income growth (in column 1) implies that the introduction of an APL

Table 4.4: The Effect of an APL Shock on Economic Activity

	(1) Personal Income	(2) Personal Income	(3) Retail Trade	(4) Retail Trade
APL	-0.00259	-0.000647	-0.00357	-0.00282
	(-0.70)	(-0.24)	(-0.78)	(-0.65)
House Prices		0.212		0.118
		(1.63)		(0.90)
Population		0.833***		0.905***
		(4.26)		(3.18)
Constant	-0.00644***	-0.00170	-0.0255***	-0.0212***
	(-4.10)	(-0.91)	(-12.66)	(-9.57)
R^2	0.700	0.766	0.580	0.604
Observations	1012	1012	978	978

t statistics in parentheses

Notes: Standard errors are clustered at both the US state and border segment levels.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

lowers annual average growth by 0.26 percentage points in the treated counties relative to the control counties. However, the treatment effect falls to 0.06 percentage points when I include house price growth and population growth as control variables (column 2). In contrast, the OLS estimate of the ATE for real retail trade personal income implies that growth in consumer spending declines by around 0.36 percentage points. The implied treatment effect is stronger for retail trade, supporting the idea that the credit standards shock affects the economy through a household borrowing/spending channel. However, in each case the treatment effect is not statistically significant. Based on the regression results, we can see that both house price growth and population growth have a positive impact on short-run local economic growth. However, only the effect of population growth is statistically significant.

4.6 Robustness Tests and Extensions

I now run a series of robustness tests. The first test is designed to assess the robustness of the credit standards model and the result that the APL shock causes a tightening in credit conditions. The remaining tests are designed to determine why I find statistically insignificant results for the economic activity model. As discussed earlier, identifying the effect of anti-predatory lending laws using variation within matched pairs of border counties limits the potential for endogeneity, both because state policymakers are less likely to consider conditions in border counties when setting state laws (selection bias) and because border counties are likely to be very similar in both observable and unobservable characteristics (omitted variables bias). But there are also costs associated with this identification strategy. The robustness tests are designed to address each of these problems.

First, my research design implicitly assumes that borrowing households in treated counties respond to the credit standards shock by reducing spending in their own county but do not reduce their spending in any nearby control counties. But the APL shock could affect treated households' spending in both the treated county and the neighbouring control county. These cross-border spending spillovers are more likely in border counties given their geographic proximity, which will bias the average treatment effect towards zero. Second, the treatment effect of the APL may be spurious if the treatment was not randomly assigned to counties. In other words, there may be unobservable characteristics that determined that one border county would be treated while the neighbouring county across the state border was not. To address this, I conduct placebo tests that exploit the fact some APls were 'ineffective' in the sense that the state lending laws were basically the same as the pre-existing federal law. If I find that these placebo APLs also have significant treatment effects then this would be evidence that the relationship is spurious. Finally, county-level proxies for household expenditure are very limited. Therefore, I also estimate the economic activity model at the state-level, both to provide a point of comparison to the county-level estimates, and to determine the extent to which the insignificant county-level estimates reflect measurement error, as personal income may be a poor proxy for household expenditure.

4.6.1 A 'side experiment': federal preemption of state lending laws

To provide further evidence that state anti-predatory lending laws affect lending standards I construct a 'side experiment' that is similar in design to the main experiment but essentially 'digs deeper' by comparing the lending behaviour of mortgage lenders *within* each county housing market. I can do this because the federal pre-emption of state law – the invalidation of US state law when it conflicts with federal law – means that subprime mortgages originated by national mortgage lenders are not subject to the anti-predatory lending laws even if the lenders operate in states that have the laws. This is useful because it provides additional control groups of mortgage lenders and this extra variation in group assignment allows me to better gauge the impact of state APLs on bank lending standards.

Federal preemption first occurred in 1996 when the Office of Thrift Supervision (OTS) issued a regulation that broadly exempted federally-chartered savings and loan institutions and their operating subsidiaries from state laws regulating credit. OTS-regulated lenders were therefore free to disregard state laws. In August 2003, the Office of Comptroller of the Currency (OCC) issued a Preemption Determination and Order stating that the Georgia mini-HOEPA statute would not apply to National City Bank or to its subsidiaries, including non-bank subprime mortgage lender, First Franklin Financial Company. The OCC then issued a broad preemption regulation in February 2004 that exempted all national banks and their subsidiaries from most state laws regarding mortgage credit. As a result, prior to August 2003, national banks were likely subject to state mortgage laws, while after February 2004 they were not.

In other words, there are effectively two groups of federally-chartered mortgage lenders that do not need to comply with state APLs even though they, or their subsidiaries, operate in states where the laws are in place. The two groups are national banks that are supervised by the Office of the Comptroller of the Currency (OCC) and Federal thrifts that are supervised by the Office of Thrift Supervision (OTS). Table 4.5 provides a brief summary of the national and local regulatory agencies that regulate residential mortgage lenders.

Table 4.5: US Bank Supervisory Agencies

Supervisor	**Responsibility**	**Subject to State Laws**
Office of the Comptroller of the Currency (OCC)	Federal banks (and the federal branches and agencies of foreign banks)	Not since 2004
Office of Thrift Supervision (OTS)	Federal and many state-charted thrift institutions (includes savings banks and savings and loan associations)	Not since 1996
Federal Reserve Board (FRB)	State banks that are members of the Federal Reserve	Yes
Federal Deposit Insurance Corporation (FDIC)	State banks that are not members of the Federal Reserve	Yes
National Credit Union Administration (NCUA)	All federal and many state-charted credit unions	Yes
Department of Housing and Urban Development (HUD)	Finance companies	Yes

In this side experiment I refer to the lenders under the supervision of the OCC and OTS as "federal" lenders and all other lenders as "state" lenders. I effectively have four different groups of mortgage lenders:

1. **Treatment group**: state mortgage lenders in APL counties
2. **Control group 1**: federal mortgage lenders in APL counties
3. **Control group 2**: state mortgage lenders in non-APL counties
4. **Control group 3**: federal mortgage lenders in non-APL counties

I test the hypothesis that state APLs affect lending standards by comparing the lending behaviour of state versus federal banks (first difference) in border counties with APLs versus border counties without APLs (second difference) and over time (third difference). This is essentially a difference-in-difference-in-difference (DDD) model. To illustrate the setup of the model, consider Table 4.6.

Table 4.6: Difference-in-Difference-in-Difference (DDD) Model Example

State	Group	Before	After	Difference	Diff.-in-Diff.	Diff.-in-Diff.-in-Diff.
APL	**State**	B_1	A_1	$D_1 = A_1 - B_1$	$DD_1 = D_1 - D_3$	$DDD_1 = DD_1 - DD_2$
	Federal	B_2	A_2	$D_2 = A_2 - B_2$	$DD_2 = D_2 - D_4$	
Non-APL	**State**	B_3	A_3	$D_3 = A_3 - B_3$		
	Federal	B_4	A_4	$D_4 = A_4 - B_4$		

In the table, the first difference compares the lending outcomes before and after treatment (which gives me the estimates D_1, D_2, D_3 and D_4). The second difference compares the first difference between APL and non-APL counties (which gives the estimates DD_1 and DD_2). Finally, the third difference compares the second difference between state and federal lenders (which gives the estimate DDD). The Average Treatment Effect (ATE) is the final DDD estimate. Intuitively, this is a test of whether there is a common time trend in lending standards for state and federal mortgage lenders in APL counties relative to non-APL counties. To assess whether the introduction of a state APL results in stricter mortgage lending standards, I conduct a (one-sided) test of the null hypothesis that the DDD estimate is equal to zero:

$$H0 : DDD = [(D1 - D3) - (D2 - D4)] = 0.$$

The introduction of a state APL is equivalent to a tightening of lending standards if I reject the null hypothesis in favour of the alternative hypothesis. This would imply that state lenders in APL counties tighten lending standards relative to both federal lenders in APL counties *and* state lenders in non-APL counties. In other words, the introduction of an APL should lead to state lenders approving fewer loan applications and offering smaller-sized loans in that county relative to federal lenders in the same county and relative to state lenders in non-APL counties. In order to test this hypothesis I augment the credit standards model with a dummy for whether the lender is a state lender or not and also interact this dummy with the key policy variable:

$$\triangle L_{ips} = \triangle APL_i'\alpha + \triangle(APL_i * STATEBANK_s)'\gamma + \triangle X_{ips}'\beta + \triangle\lambda_p + \triangle\epsilon_{ips} \quad (4.3)$$

Where the variable $STATEBANK_s$ indexes the type of lender and is equal to one if the lender is a state lender and equal to zero if the lender is a federal lender. Note that the main effect of the dummy variable $STATEBANK_s$ has already been removed through time differencing the data, so all that remains is the interaction effect with the APL dummy variable, which is given by $\triangle APL_i * STATEBANK_s$.

The coefficient γ is the DDD estimate which measures the average change in lending standards for state lenders in APL counties (the treatment group) relative to all other lenders (the control groups) after the APL shock. Under the hypothesis that only state lenders should tighten lending standards in response to the change in state government policy I would expect that $\gamma < 0$. The coefficient α now has a different interpretation as it measures the effect of introducing an APL on federal lenders in APL counties. I would now expect that $\alpha = 0$ as federal lenders are a control group and should not be affected by the policy change. As before, I separately estimate the equation for the three different measures of lending standards – the loan approval rate, the average loan size and the average loan-to-income ratio.

Overall, there is evidence that state-chartered mortgage lenders increase their lending standards in response to the introduction of the state anti-predatory lending laws (Table 4.7). The coefficient on the interaction term $APL * STATEBANK$ is negatively signed in every case, although only significant in the case of the average loan size. This is consistent with the hypothesis that only state lenders tighten credit standards in response to the change in state legislation. More surprisingly, the coefficient on the term APL is positively signed, and significant for the loan size, which implies that federally-chartered banks approve larger loans in APL states after the policy change. In other words, the federal lenders are affected

Table 4.7: The Effect of an APL Shock on Credit Conditions

	(1) Average Loan Size	(2) Loan to Income Ratio	(3) Loan Approval Rate
APL	0.0247*	-0.0382	0.0134
	(1.69)	(-0.70)	(1.24)
APL x STATEBANK	-0.0718**	-0.0633	-0.0243
	(-2.37)	(-0.75)	(-1.24)
House Prices	0.647***	1.169***	-0.0549
	(9.40)	(4.10)	(-1.09)
Household Income	0.408***	-1.192***	0.0563***
	(18.22)	(-4.53)	(5.92)
Minority Ratio	0.403**	1.502*	0.00700
	(2.38)	(1.79)	(0.10)
Constant	0.0683***	0.116**	-0.0212**
	(3.45)	(2.08)	(-2.08)
R^2	0.154	0.030	0.020
Observations	69030	67750	77758

t statistics in parentheses

Notes: Standard errors are clustered at both the US state and mortgage lender levels.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

by the policy change, which is inconsistent with the null hypothesis. The combination of these results suggests that high-risk applicants switch their borrowing from state lenders to federal lenders as the federal lenders were offering more relaxed credit terms after the policy change. In any case, there is evidence that the state laws affected bank credit standards.

4.6.2 Cross-border spillovers in household spending

While there is a reasonable body of evidence that the lending laws affected credit standards, the main regression results suggest there is little evidence that this, in turn, affected outcomes in the real economy. The lack of statistically significant results for the average treatment effect in the economic activity model may be due to spillovers in "cross-border spending". My research design implicitly assumes that borrowing households in treated counties respond to the credit standards shock by reducing their spending in their own county and has no effect on their spending in control counties across the border. If, instead, households in the treatment counties respond to the shock by curtailing spending in both their own treated county *and* the control counties then the average treatment effect will be biased towards zero. By focussing on matched border counties, my methodology is particularly susceptible to these cross-border spillovers.

To determine the extent of any cross-border spillovers I re-estimate the economic activity model and replace each border control county with its closest neighbouring *hinterland* county. Each hinterland county shares a border with a control county and is in the same state but is further in-land and so does not share a border with the treated county. As the treated border county and the control hinterland county are geographically close they should still be similar in both observable and unobservable characteristics *but* there should be less scope for cross-border spillovers as the control hinterland counties are more distant than the control border counties.

The specification of the regression equation is identical to the main part of the paper, except that each control border county observation is replaced with its closest hinterland county observation. I again remove unobservable county fixed effects through first differencing and unobservable pair-specific trends through within-pair differencing. The resulting equation to be estimated is again a difference-in-difference (DD) panel regression. The hypothesis is that, if there are cross-border spending spillovers, the average treatment effect of the APL shock will be negative (and larger in magnitude) in the hinterland county regression than in the border county regression. The results of estimating the equation are shown in Table 4.8. The first two columns contain the estimates based on the sample of control hinterland counties. The last two columns contain the original estimates based on the sample of border counties as a point of comparison.

Table 4.8: The Effect of an APL Shock on Economic Activity

	(1) Pers. Income (Hinterland)	(2) Ret. Trade (Hinterland)	(3) Pers. Income	(4) Ret. Trade
APL	-0.00201	0.00157	-0.000647	-0.00282
	(-0.61)	(0.21)	(-0.24)	(-0.65)
House Prices	0.347*	0.273	0.212	0.118
	(1.95)	(1.10)	(1.63)	(0.90)
Population	0.667***	0.675*	0.833***	0.905***
	(3.04)	(1.87)	(4.26)	(3.18)
Constant	0.00180	-0.0236***	-0.00170	-0.0212***
	(0.63)	(-6.68)	(-0.91)	(-9.57)
R^2	0.726	0.577	0.766	0.604
Observations	973	927	1012	978

t statistics in parentheses

Notes: Standard errors are clustered at both the US state and border segment levels.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Comparing columns 1 and 3, the estimated treatment effect on total personal income growth suggest that the APL shocks have negative spillover effects on control counties. As expected, the treatment effect is larger based on the hinterland county sample because the matched counties are more geographically distant, reducing the probability of cross-border spillovers. However, there is less evidence for spillovers based on the retail trade estimates.

Comparing columns 2 and 4, we see that the APL shock has a stronger (negative) effect on retail trade income growth when based on the sample of border counties compared to the sample based on the hinterland counties. In fact, the estimates in column 2 for the hinterland counties suggest that the policy change has a positive effect on growth in retail trade.

Compared to the case of cross-border spending, the inability to find a significant effect of APL shocks on economic activity in the main part of the paper is unlikely to be due to households in APL states engaging in "cross-border borrowing" from mortgage lenders in non-APL states. At first blush, it may seem that borrowers in treated border counties could respond to any policy change by accessing credit from a lender in the untreated neighbouring county, which would again bias the effect of the policy change towards zero. However, the APL legislation applies to the location of the property and not the location of the borrower or lender.[68]

4.6.3 Placebo test of "ineffective" state lending laws

As another robustness exercise I conduct a placebo test based on the economic activity model. The test is designed to rule out a spurious treatment effect of the anti-predatory lending laws. For instance, the anti-predatory lending laws may affect households' expectations of future housing demand (and future house prices) and it could be this change in households' expectations that affect their current spending rather than a tightening in credit standards. The placebo test exploits the fact that some states introduced APLs that were essentially the same as the pre-existing federal legislation. If the state and federal laws were effectively the same, then we should observe no change in credit standards and hence no change in economic outcomes in the counties with these 'ineffective' laws relative to the control counties. If, instead, there is a significant effect of the placebo laws, then this is

[68]The only way in which this type of cross-border borrowing could reduce the chance of finding a significant treatment effect is if households respond to the APL by literally moving to a new property in a non-APL state. But it would be surprising if households made their housing investment decisions purely on the basis of whether an APL exists or not, especially as most APL laws relate to refinance loans rather than house purchase loans.

evidence that the treatment effect is due to some confounding factor.

Ding et al (2011) identify eight states with state laws that, they judge, were not sufficiently different to the existing federal law.[69] I re-estimate equation 4.2 based on the subset of matched counties that lie on the borders of these placebo states. The results of conducting the placebo tests are shown in table 4.9.

Table 4.9: The Effect of a Placebo APL Shock on Economic Activity

	(1) Personal Income	(2) Personal Income	(3) Retail Trade	(4) Retail Trade
Placebo APL	0.00309	0.00289	-0.00155	-0.00149
	(1.18)	(1.54)	(-0.24)	(-0.24)
House Prices		0.210**		0.172
		(2.48)		(0.53)
Population		0.782***		1.097**
		(3.56)		(2.60)
Constant	-0.0109***	-0.00635***	-0.0270***	-0.0224***
	(-6.98)	(-4.05)	(-8.06)	(-7.19)
R^2	0.649	0.728	0.553	0.589
Observations	706	706	679	679

t statistics in parentheses

Notes: Standard errors are clustered at both the US state and border segment levels.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, I find that I cannot recreate the results of the natural experiment in the placebo tests. In fact, when economic activity is measured through total personal income growth, I find a *positive* average treatment effect. In other words, the placebo laws appear to lead to higher growth outcomes. For instance, the placebo treatment effect is about 0.3 percentage points for average annual personal income growth (columns 1 and 2). This positive placebo effect is surprising. A potential explanation is that the presence of an anti-predatory lending law reassures borrowers that they are protected from predatory lending and therefore encourages more borrowing and spending. More importantly, these results suggest that the

[69] These identified 'placebo' states are (with the year of each APL in parentheses): Florida (2002), Kentucky (2003), Maine (1995), Nevada (2003), Ohio (2002), Oklahoma (2004), Pennsylvania (2002) and Utah (2004).

inability to find a significant treatment effect for the *effective* APLs is because of such a confounding positive confidence effect on borrowing and spending.[70]

4.6.4 State expenditure estimates

I next present the results of estimating the regression models using state-level data. These tests are designed as a point of comparison to the county-level results and to determine the extent to which the insignificant county-level estimates reflect measurement error in my measure of economic activity. The regression equation is precisely the same as in equation 4.2 but the unit of observation is the US state rather than US county. Essentially, the treatment group is the state that introduces the law and the control group is the neighbouring state that does not. Identification relies on variation in economic activity within each neighbouring state pair before and after the law. I measure economic activity through state-level growth in total GDP and retail trade GDP. To aid comparisons with the county-level estimates, I also provide the estimates based on state-level total personal income and retail trade income growth. The results are presented in Table 4.10.

As column 1 indicates, GDP growth falls by 0.25 percentage points in the APL states relative to the non-APL states, on average, following a tightening of lending standards. The implied treatment effect is substantially larger than the estimate of 0.08 percentage points based on state-level personal income growth (column 2). This suggests that personal income is an imprecise measure of county-level spending and may partly explain the lack of statistical significance. However, the GDP estimates are also statistically insignificant. Moreover, the average treatment effects identified on the basis of state-level retail trade GDP is wrongly signed. Overall, it appears that the lack of direct measures of household spending at the county level is not driving the results.

[70] In unreported results I have also conducted two alternative placebo tests where I estimate the main equation under the assumption that the treatment occurs one year before the actual APL introduction and one year after the actual introduction. In both cases I find the estimated treatment effects to be weaker than those reported in the paper. This provides further evidence that I am capturing the 'true' impact of the shock to lending standards on economic conditions, even if the true treatment effect is small.

Table 4.10: Effect of APLs on State GDP Growth

	(1) GDP	(2) Personal Income	(3) Retail Trade GDP	(4) Retail Trade Income
APL	-0.00248	-0.000821	0.00219	0.00447
	(-0.88)	(-0.41)	(0.61)	(0.89)
House Prices	0.106	0.235**	0.278**	0.238
	(0.62)	(2.30)	(2.31)	(1.19)
Population	1.393**	1.479***	1.487***	1.671**
	(2.48)	(4.96)	(4.10)	(2.22)
Observations	90	90	90	90
R2	0.767	0.910	0.798	0.749

t statistics in parentheses

Notes: Standard errors are clustered on both states in each matched pair.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6.5 Mortgage credit regulation and the financial crisis

Next, I assess how the states that introduced the APLs fared during the 2008-09 financial crisis relative to the states that did not. The introduction of an APL may be an example of "short-term pain, long-term gain" if the credit standards shock causes weaker spending in the short-term but leaves the local economy, in the longer-term, less susceptible to other types of shocks. This may also explain why the effect of the credit standards shock on local economic growth is insignificant – the model collapses the post-treatment period into a single observation for each matched border pair so it makes no distinction between the short-term and longer-term impact of the policy change.

To address this question I modify the difference-in-differences model of economic activity. The equation is now specified as:

$$Y_{ipt} = APL_{it}'\alpha + CRISIS_t'\beta + APL_{it} * CRISIS_t'\gamma + X_{ipt}'\rho + \theta_{ip} + \lambda_{pt} + \epsilon_{ipt}$$

Similar to before, the dependent variable (Y_{ipt}) is a measure of economic conditions in each county i in each matched border pair p in year t. The policy variable (APL_{it}) is a binary indicator equal to unity if county i is in the treatment group in year t and is equal to zero otherwise. This specification also includes a dummy for the period since the onset of the crisis in 2007 ($CRISIS_t$) and an interaction dummy for the period since the crisis and if the county is in an APL state ($APL_{it} * CRISIS_t$) which measures whether the effect of the APL treatment on economic conditions is different before and after the crisis.[71] The specification also includes unobservable county effects (θ_{ip}) and pair-time effects (λ_{pt}). I eliminate the unobservable pair-time effects by taking the difference between the treated and control county within each border pair at a given point in time. The estimating equation becomes:

$$\hat{Y}_{pt} = A\hat{P}L_t{}'\alpha + A\hat{P}L_t{}'CRISIS_t\gamma + \hat{X}'_{pt}\rho + \hat{\theta}_p + \hat{\epsilon}_{pt}$$

As before, a variable with a 'hat' denotes the within pair-year difference between the treated and control county (i.e. $\hat{x}_{pt} = (x_{Tpt} - x_{Cpt})$). After these manipulations, the variable APL_t becomes a dummy variable that is equal to one in any year after the APL is introduced, while the variable $APL'_tCRISIS_t$ becomes a dummy variable that is equal to one in any year after 2007. In this model there are three periods for each border county pair – the period before the APL treatment, the period after the APL treatment, but before the crisis, and the period after both the APL treatment and crisis. The estimate of α now measures the effect of the credit standards shock on economic activity *prior* to 2007. The estimate of γ measures the effect of the APL shock on economic conditions *after* 2007. Based on the hypothesis that the credit standards shock leads to weaker economic growth in the short-term, but stronger growth in the medium-term, we would expect $\alpha < 0$ and $\gamma > 0$. The results of estimating the difference-in-differences model with the financial crisis interaction term are shown in Table 4.11.

In the period after the APL shock but before the crisis (i.e. 2000 to 2006), growth in total personal income falls in the treated counties relative to the control counties. Personal

[71] Note that the financial crisis occurred in 2007 after all the state anti-predatory lending laws were enacted, so this rules out a number of potential interaction effects.

Table 4.11: The Effect of APL Shocks Before and After the Financial Crisis

	(1) Personal Income	(2) Personal Income	(3) Retail Trade	(4) Retail Trade
APL	-0.00422	-0.00371	-0.000128	0.00133
	(-1.25)	(-1.12)	(-0.02)	(0.33)
APL x CRISIS	0.00485	0.00466	-0.00422	-0.00455
	(1.10)	(1.05)	(-0.63)	(-0.75)
CRISIS	-0.0264***	-0.0220***	-0.0154	0.0101
	(-4.95)	(-3.35)	(-1.56)	(0.98)
House Prices		0.0504		0.253***
		(1.42)		(4.64)
Population		0.512		0.832***
		(1.62)		(3.08)
Constant	0.00269	0.00299	-0.0132***	-0.0165***
	(1.02)	(1.13)	(-2.83)	(-4.16)
R^2	0.405	0.443	0.218	0.316
Observations	2070	2070	1986	1986

t statistics in parentheses

Notes: Standard errors are clustered at both the US state and border segment levels.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

income growth shrinks by about 0.4 percentage points in the APL counties relative to the non-APL counties (columns 1 and 2). This suggests that the APLs cause slower growth in the period immediately after their introduction. However, the positive coefficient on the interaction term for the APL and financial crisis period indicates that growth in total personal income is higher in the treatment counties than the control counties after 2007. Relative to the control counties, personal income growth rebounds by around 0.5 percentage points in the treated counties in the post-crisis period. In other words, the financial crisis had a disproportionately larger negative impact on the control counties than on the treatment counties. This is consistent with the state anti-predatory lending laws helping to mitigate the impact of the financial crisis. However, none of the results are statistically significant.

These results for total personal income are partly contradicted by the estimates based on the retail trade industry. Here we see that retail trade growth is generally stronger in the treatment counties than in the control counties in both the period after the APL shock but before the crisis (column 4). This casts doubt on whether the effect of the APL shocks is being transmitted through a household borrowing/spending channel. So, overall, there is only tentative evidence that the introduction of an APL resulted in weaker spending growth, on average, and that the states with the tougher credit regulations fared better in the post-financial crisis period.[72]

4.7 Conclusion

I conduct a natural experiment on the US housing market to identify the causal effect of shocks to credit standards on economic activity. Prior to the subprime mortgage crisis, different US states enacted anti-predatory lending laws at different times and the introduction of these lending laws acted like a government-enforced tightening of credit standards for home mortgages. By comparing the economic performance of contiguous counties along

[72]In unreported results, I also find little evidence that the share of mortgage delinquencies in the APL states fell relative to the non-APL states after the passage of the law. So this also argues against the laws having a positive impact on the stability of local housing markets.

state borders where one state introduced the law ("the treatment group") and the other state did not ("the control group") I identify the economic effect of shocks to residential mortgage credit standards.

Using a difference-in-difference modelling framework, I find evidence that the introduction of the state lending laws is associated with tighter credit standards and less household borrowing. In particular, I find that the laws reduce both the average size of loans offered to borrowers and the proportion of loans offered to applicants, on average.

While I find that the laws matter for household borrowing I find only limited evidence that this subsequently translates into weaker growth in household spending or, more generally, economic activity. A placebo test using lending laws that should have no impact on spending suggests that there may be a confounding effect of the laws on borrower confidence, which works in the opposite direction to boost spending. Any spurious confidence effect will bias the estimated average treatment effect towards zero.

The insignificant results may also reflect the fact that households affected by the credit standards shock can scale back their spending in both their own state and in neighbouring states that do not introduce the laws. This 'cross-border spending' biases the effect of the APL shock towards zero.

The insignificant effect of the lending laws on economic performance may also reflect the fact that the laws restrain household spending in the short-term but boost spending in the medium-term. States that adopt tougher mortgage regulation may do better in the medium-term than states without such mortgage regulation because they are better able to withstand other types of banking sector shocks. To examine this issue I assess how the treated and control counties fared after the financial crisis. I find some tentative evidence that the APL states performed better than the non-APL states after the crisis, which is consistent with the notion that the laws help to stabilise local economies in the face of other financial shocks.

Appendices

H Derivation of Theoretical Model

To put the main ideas in a conceptual framework, I outline a simple theoretical model to demonstrate how shocks to credit standards affect household borrowing and consumption. The model is based on the consumer loan models of Jaffee and Russell [1976], Webb [1984], Milde and Riley [1988]. It is also similar to the mortgage market model of Brueckner [2000]. The framework is intended to convey the intuition behind the empirical strategy in the paper and not to explain optimal household borrowing decisions or the effect of asymmetric information on credit markets. I use the framework to examine how the introduction of laws that constrain lending to high-risk borrowers affects the distribution of credit and hence the aggregate level of household spending.

H.1 General Setup

Households live for two periods, are risk-averse, and earn income and consume in both periods.[73] In the first period, the lender simply offers a loan at an exogenously set interest rate (R).[74] The household takes the interest rate, their current income and their expected income as given and decides to borrow L to purchase a home in period 1 which costs P_1.[75] The loan balance (RL) is due in period 2. The risk in the model relates to the realised outcome for the home price (P) in period 2. I assume that the home price is stochastic and has finite support ($P \in \underline{P}, \overline{P}$) and a cumulative distribution, $F(P)$ which is differentiable

[73] This is a partial equilibrium model in which I ignore the behaviour of households that choose to save rather than borrow. To do so, I assume that at least some households are sufficiently impatient that they optimally choose to borrow at any of the relevant interest rates.

[74] I have also solved the model in the case where the lender sets the interest rate based on the observable characteristics of the borrower. The model's predictions are similar, though the comparative statics are considerably more difficult to derive. This is because the borrower, when choosing their desired loan size, must account for the fact that the loan decision will impact on the probability of default and therefore the interest rate charged by the lender.

[75] Households make a necessary downpayment given by $P_1 - L$.

and strictly increasing wherever $0 < F < 1$. The household may choose to default on the loan once they learn the outcome for the home price in period 2.

I assume there is a continuum of households with the type of each household denoted by θ which has finite support ($\theta \in 0, 1$). Households differ solely in terms of the disutility (or cost) of defaulting on a loan. The borrower-specific cost function ($C(\theta)$) is assumed to be increasing and convex in the borrower's type ($C'(\theta), C''(\theta) > 0$). This implies that 'high-type' borrowers, that find it more costly to default, are 'low-risk' borrowers. Let the utility of the household in period one be $u(Y_1 - P_1 + L)$ and utility in period two be $u(Y_2 + P - RL)$ if the household repays the loan and $u(Y_2) - C(\theta)$ if the household defaults.

The loan default rule in this model is relatively straightforward – in period 2 the type-θ borrower ex post defaults on a loan of value $RL(\theta)$ if the utility of the realised wealth from default ($u(Y_2) - C(\theta)$) exceeds the utility of the realised wealth from repaying the loan ($u(Y_2 + P - RL(\theta))$). Therefore, for each borrower, there is a function $G(\hat{P(\theta)}; \theta)$ that implicitly defines the home price level ($\hat{P}(\theta)$) at which the borrower is indifferent between repaying and defaulting on the loan:

$$G(\hat{P(\theta)}; L(\theta)) \equiv u(Y_2 + \hat{P}(\theta) - RL(\theta)) - u(Y_2) + C(\theta) = 0$$

The household repays the loan when the realised value of the home price exceeds the cut-off level ($P \geq \hat{P}$) and defaults when the realised home price is below the cut-off level ($P < \hat{P}$).[76] This decision rule is highly simplified; empirical evidence suggests that negative home equity is a neccesary but not sufficient condition for borrowers to default. Instead, default generally occurs when there is a "double trigger" – negative equity *and* some unexpected household cash flow problem (Foote et al. [2008]). Nevertheless, the rule is chosen to maintain

[76] I assume that in the event of default the lender can monitor the consumption of the household to ensure that the borrower did not default in the most favourable state (at a cost which is passed fully on to consumers via the default cost term, C). If households could 'strategically default', then they would borrow as much as possible in period 1 and default in every state in period 2. Consumption in period 1 would then be unbounded. To rule this out I assume there is a very high penalty if the borrower defaults in the most favourable state. This puts an upper limit on desired borrowing and hence consumption in period 1.

tractability in the model and is the standard approach taken in the consumption/housing literature (e.g. Brueckner [2000]). Applying the implicit function theorem we can infer from this default decision rule that the cut-off house price ($\hat{P}$) is increasing in the loan size ($\partial \hat{P}/\partial L > 0$) because a larger loan implies a greater chance of default, but decreasing in the borrower's type ($\partial \hat{P}/\partial \theta < 0$) because higher borrower types find it more costly to default.

Finally, I assume that household preferences are given by a concave additively separable expected utility function (V) (with discount factor $\beta \in (0,1)$):[77]

$$V(\theta) = u(Y_1 - P_1 + L(\theta)) + \beta \int_{\hat{P}(\theta)}^{\overline{P}} u(Y_2 + P - RL(\theta))\, dF(P) + \beta \int_{\underline{P}}^{\hat{P}(\theta)} u(Y_2) - C(\theta)\, dF(P)$$

I solve the model backwards. In period 2, the household takes the loan size and interest rate as given and chooses whether to repay the loan given the realised shock to the house price. In period 1 the household chooses their desired loan size given the interest rate and expected future income and house prices.

H.2 Competitive Equilibrium

The equilibrium of this two-period consumption model is for each borrower to choose the loan size (L^*) that maximises their expected utility, taking as given the interest rate, income, house prices, and the default decision rule.

$$\max_{L(\theta)} V(\theta) = u(Y_1 - P_1 + L(\theta)) + \beta \int_{\hat{P}(\theta)}^{\overline{P}} u(Y_2 + P - RL(\theta))\, dF(P) + \beta \int_{\underline{P}}^{\hat{P}(\theta)} (u(Y_2 - C(\theta))\, dF(P)$$

I apply Leibniz's rule to obtain the household's first order condition:

[77]The assumption of a concave, additively separable utility function ensures that consumption in periods 1 and 2 are normal goods.

$$
\begin{aligned}
\frac{\partial V}{\partial L} =& u'(Y_1 - P_1 + L^*) \\
& - \beta R \int_{\hat{P}}^{\overline{P}} u'(Y_2 + P - RL) dF(P) \\
& - \beta \frac{\partial \hat{P}}{\partial L} u(Y_2 + \hat{P} - RL^*) f(\hat{P}) + \beta \frac{\partial \hat{P}}{\partial L} u(Y_2 - C(\theta)) f(\hat{P}) = 0
\end{aligned}
$$

Making use of the default decision rule this simplifies to:

$$
J(L^*(\theta); \theta) \equiv \frac{\partial V}{\partial L} = u'(Y_1 - P_1 + L^*) - \beta R \int_{\hat{P}^*}^{\overline{P}} u'(Y_2 + P - RL^*) dF(P) = 0 \tag{4.4}
$$

The optimal loan size in the competitive equilibrium (L^*) is implicitly defined by this Euler equation.

H.3 Comparative Statics

Before considering the effect of the anti-predatory lending laws in the model it is useful to derive some comparative statics. In particular, I consider how the optimal loan size (L^*) varies with the observed riskiness of the borrower (θ). Totally differentiating the household's first order condition:

$$
\begin{aligned}
dJ =& u''(Y_1 - P_1 + L^*) dL^* + \beta R^2 dL^* \int_{\hat{P}^*}^{\overline{P}} u''(Y_2 + P - RL^*) dF(P) \\
& + \beta R dL^* \frac{\partial \hat{P}^*}{\partial L^*} u'(Y_2 + \hat{P} - RL^*) dF(\hat{P}) + \beta R d\theta \frac{\partial \hat{P}}{\partial \theta} u'(Y_2 + \hat{P} - RL^*) dF(\hat{P}) = 0
\end{aligned}
$$

Re-arranging I obtain the marginal effect of a change in borrower riskiness (or creditworthiness) on the optimal loan size:

$$\frac{dL^*}{d\theta} = -[\frac{u''(Y_1 - P_1 + L^*) + \beta R^2 \int_{\hat{P}^*}^{\overline{P}} u''(Y_2 + P - RL^*)dF(P) + \beta R \frac{\partial \hat{P}^*}{\partial L^*} u'(Y_2 + \hat{P} - RL^*)dF(\hat{P})}{\beta R \frac{\partial \hat{P}}{\partial \theta} u'(Y_2 + \hat{P} - RL^*)dF(\hat{P})}]$$

It is not possible to sign this derivative *a priori* without making an additional assumption. Note that the first two terms in the numerator capture the effect of a marginal increase in household borrowing (L^*) on the marginal utility of consumption in the two periods. These effects are totally standard and appear in the consumption model with no default option. In equilibrium the marginal increase in utility from the higher consumption in period 1 is offset by a marginal decrease in utility due to lower (discounted) consumption in period 2. These terms are both negative given the assumption of diminishing marginal utility and because they appear on the same side of the equation. The complication in signing the derivative is due to the third term in the numerator. This term captures the extensive margin of an increase in borrowing when there is default risk. A higher level of debt increases the likelihood of default ($\partial \hat{P}/\partial L > 0$) and *increases* future consumption if the borrower chooses not to repay the loan. This last term is therefore positively signed. It essentially captures the 'moral hazard' or 'incentive' effect of higher household debt. In contrast, the denominator captures the 'direct' effect of a marginal change in the risk profile of the borrower (θ) on the marginal utility of consumption. As the riskiness of the borrower falls (i.e. θ rises), they become more likely to repay the loan in the second period, which lowers consumption and increases expected marginal utility in the second period. The denominator is therefore negatively signed.

In order to sign the partial derivative, I assume the moral hazard effect of an increase in borrowing is relatively small (i.e. $\partial \hat{P}/\partial L$ is small). In other words, I assume that most borrowers do not take out a loan with the explicit aim of defaulting in the future. If this assumption holds, then we can determine that the overall derivative is negatively signed:

$$\frac{dL^*}{d\theta} \leq 0$$

This implies that riskier borrowers (i.e. low θ types) take out larger loans than safe borrowers (i.e. high θ types) on average. The riskiness of the borrower affects the intertemporal trade-off in consumption when there is default risk. Essentially, a less risky borrower is more likely to repay the loan which reduces their expected future consumption by more than in the case where there is no default option. The risk-averse borrower smooths the additional burden on future consumption by choosing a smaller loan in the first period. Intuitively, a borrower who finds default more costly reduces the probability of default by choosing to borrow less.

H.4 The Effect of Introducing High-Risk Lending Laws

In this setting I argue that, by ruling out high-risk loans, an anti-predatory lending law is equivalent to an exogenous (government-imposed) interest rate ceiling. Under the reasonable assumption that riskier borrowers are charged higher interest rates, the setting of an interest rate ceiling (R_{max}) is equivalent to imposing a lower bound on the riskiness of the borrower distribution (θ_{min}). To the extent that some borrowers are deemed too risky under the law, the imposition of a ceiling will lead to a fall in the aggregate level of credit and in the share of borrowers that are approved for loans, as lenders are forced to ration credit. Less obviously, it will also lead to a lower average loan size for approved loans and a lower average level of consumer spending. To see this, denote the following:

- Optimal loan size of each borrower: $L^*(\theta)$
- No. of originations: $O_{PRE} = \int_0^1 O(\theta)d\theta > O_{POST} = \int_{\theta_{min}}^1 O(\theta)d\theta$
- No. of applications: $A_{PRE} = A_{POST} = \int_0^1 A(\theta)d\theta$
- Lending volume: $L_{PRE} = \int_0^1 O(\theta) * L^*(\theta)d\theta > L_{POST} = \int_{\theta_{min}}^1 O(\theta) * L^*(\theta)d\theta$

In my simple model, the anti-predatory lending law results in fewer loans being originated after the law (O_{POST}) than before the law (O_{PRE}) due to the tighter minimum lending standards required to comply with the law. By the same logic, the loan approval rate (i.e. the number of loan originations divided by the number of loan applications) will also fall after the law is introduced.[78] In my model the imposition of an anti-predatory lending law also leads to a fall in the *average loan size* ($\bar{L}^*_{POST} = \frac{L^*_{POST}}{O_{POST}} < \bar{L}^*_{PRE} = \frac{L^*_{PRE}}{O_{PRE}}$). Unlike the effect of the law on the volume of credit and the approval rate, this model prediction is not mechanical but relies crucially on the fact that high-risk borrowers take out larger loans on average than low-risk borrowers. As the riskier borrowers that demand larger loans are rationed, the average loan size falls.[79] Given the direct mapping from the average loan size to the average consumption of each household, the introduction of the law leads to a lower average level of consumer spending. The main predictions of the model are therefore:

Proposition: *Compared to counties in states with no such laws, counties in states that introduce anti-predatory lending laws will observe after the law is introduced:*

- Smaller average loan sizes
- Lower average household consumption
- Lower average loan approval rates

[78]These results hinge on at least two key assumptions that are implicit in the model. Firstly, riskier borrowers are assumed to not endogenously respond to the introduction of the law by applying for smaller loans than initially desired. This would work to limit the impact of the law on both the total volume of credit and the loan approval rate. But presumably at least some high-risk applicants would be rationed, so that there would be an overall negative impact of the law on loan volume and the approval rate. Secondly, the model assumes there is no 'encouraged applicant effect' of the anti-predatory lending law. For example, some previously-discouraged applicants may feel more 'protected' from predatory lending if a law is in place, which might lead to an increase in the number of loan applications after the law is introduced. Refer to Ho and Pennington-Cross [2006] for a theoretical model that allows for such an effect. There is also some empirical evidence for the encouraged applicant effect (e.g. Bostic et al. [2008]). But while this would also work to limit the effect of the law on the volume of credit, the loan approval rate would presumably still fall. This is because the encouraged applicants are more likely to be high-risk borrowers, and so are relatively more likely to be rationed by lenders even if they are encouraged to apply. As such, we should still observe a fall in the loan approval rate.

[79]If, instead, the less risky borrowers demanded larger loans, then the introduction of the law would result in a rise in average loan size.

Chapter 5

Conclusion

In this thesis I provide new insights into the causes and effects of the recent US housing market crisis. Nevertheless, I believe there is still scope for future research in this large (and rapidly expanding) branch of the economic literature.

For instance, my first chapter documents new stylised facts about the US housing market, such as a dramatic 'boom-bust' pattern in the geographic distance between borrowers and lenders during the 2000s. Assuming distance affects the quality of information that banks have about borrowers, I provide evidence that the increase in lending distance over the 2000s was associated with relatively worse lending decisions and subsequently contributed to more mortgage delinquencies. This chapter could be strengthened if I could find an appropriate natural experiment to test whether geographic distance affects credit quality. For instance, I have considered an experiment in which the deregulation of US interstate bank branching laws in the mid-1990s acts as a positive "lending distance shock". The branching deregulation by certain US states allowed out-of-state banks to establish branches and originate loans in those particular markets. In the experiment, each bank's geographic distance to the deregulating states in the pre-deregulation period acts as the instrument for the change in each bank's lending distance during the deregulation. The logic is that banks located further away from the deregulating states would need to travel relatively far to take advantage of the new lending opportunities. But this experiment currently suffers from a couple of problems. First, the deregulation allowed banks to expand geographically, which contributed to a rise in lending distance, but it also caused other factors to change, such as the level of competition amongst mortgage lenders. Second, the deregulation allowed banks to expand geographically in the mid-1990s, which is more than a decade before the rapid rise in mortgage delinquencies that occurred during the subprime mortgage crisis. The effect of distance on information production and risk-taking may not persist for such a long time.

In the second chapter I showed that lenders that were reliant on securitisation to finance new lending (the OTD lenders) disproportionately reduced mortgage credit supply during the credit crisis. I attributed the relatively large decline in credit to greater financing constraints for OTD lenders when the securitisation market closed. However, the paper is unable to completely rule out the alternative hypothesis that the OTD lenders took more risk in lending during the boom and that the relatively large decline in credit supply during the crisis was caused by a subsequently greater need for the OTD lenders to reduce their risk exposure. The paper would therefore benefit from a cleaner test that distinguishes between these competing hypotheses.

In the third chapter I found that the introduction of the state lending laws was associated with tighter credit standards and less household borrowing, but little evidence that the shocks significantly affected household spending. I identified several possible reasons for the lack of significant results, including cross-border spending spillovers, borrower confidence effects and laws that were not sufficiently binding to affect local economic conditions. It is possible that using more direct measures of household spending at the county-level would resolve these problems. For example, it is possible to collect information on county-level retail sales from various state government departments. But this exercise is very time consuming and may not reap significant rewards as my state-level regressions (which use more direct measures of spending) also generate insignificant treatment effects. This suggests that measurement error in the dependent variable may not be the source of statistical insignificance.

Overall, I believe my research helps to bridge the gap between microeconomics and macroeconomics by examining interesting macroeconomic research questions using identification strategies based on micro-econometric evidence. In doing so, I have compiled a comprehensive mortgage application dataset covering two decades. I intend to continue using this dataset to uncover other interesting facts about the US housing market. For instance, I am currently using the dataset to examine how the human capital of bank employees affected the quality of loans originated by US banks duing the subprime mortgage boom-bust cycle. This is joint research with Professor Luis Garicano of the LSE's Department of Management. My preliminary results indicate that the US banks that were relatively dependent on young and inexperienced workers generally performed worse during the crisis. In future research, I hope to continue to demonstrate that, when financial markets are imperfect, credit supply shocks can significantly affect the real economy.

Chapter 6

References

References

Sumit Agarwal and Itzhak Ben-David. Do loan officers incentives lead to lax lending standards? Technical report, 2012. URL http://www.econ.yale.edu/~shiller/behfin/2012-04-11/Agarwal_Ben-David.pdf.

Sumit Agarwal and Robert Hauswald. Distance and private information in lending. *Review of Financial Studies*, 23(7):2757–2788, 2010. URL http://rfs.oxfordjournals.org/content/23/7/2757.abstract.

Sumit Agarwal, Brent W. Ambrose, Souphala Chomsisengphet, and Chunlin Liu. The role of soft information in a dynamic contract setting: Evidence from the home equity credit market. *Journal of Money, Credit and Banking*, 43(4):633–655, 2011. ISSN 1538-4616. URL http://dx.doi.org/10.1111/j.1538-4616.2011.00390.x.

Philippe Aghion and Jean Tirole. Formal and real authority in organizations. *Journal of Political Economy*, 105(1):1–29, February 1997. URL http://ideas.repec.org/a/ucp/jpolec/v105y1997i1p1-29.html.

Ugo Albertazzi and Domenico J. Marchetti. Credit supply, flight to quality and evergreening: an analysis of bank-firm relationships after lehman. Temi di discussione (Economic working papers) 756, Bank of Italy, Economic Research and International Relations Area, Apr 2010. URL http://ideas.repec.org/p/bdi/wptemi/td_756_10.html.

Franklin Allen and Douglas M. Gale. Competition and Financial Stability. *SSRN eLibrary*, 2003.

Brent W. Ambrose, Michael LaCour-Little, and Zsuzsa R. Huszar. A note on hybrid mortgages. *Real Estate Economics*, 33(4):765–782, December 2005. URL http://ideas.repec.org/a/bla/reesec/v33y2005i4p765-782.html.

Dean F. Amel, Arthur B. Kennickell, and Kevin B. Moore. Banking market definition: evidence from the survey of consumer finances. Finance and Economics Discussion Series 2008-35, Board of Governors of the Federal Reserve System (U.S.), 2008. URL http://ideas.repec.org/p/fip/fedgfe/2008-35.html.

Charles D. Anderson, Dennis R. Capozza, and Robert Van Order. Deconstructing a mortgage meltdown: A methodology for decomposing underwriting quality. *Journal of Money, Credit and Banking*, 43(4):609–631, 2011. ISSN 1538-4616. doi: 10.1111/j.1538-4616.2011.00389.x. URL http://dx.doi.org/10.1111/j.1538-4616.2011.00389.x.

Patrick K. Asea and Brock Blomberg. Lending cycles. *Journal of Econometrics*, 83(1-2):89–128, 1998. URL http://ideas.repec.org/a/eee/econom/v83y1998i1-2p89-128.html.

Robert B. Avery, Kenneth P. Brevoort, and Glenn B. Canner. Opportunities and issues in using hmda data. *Journal of Real Estate Research*, 29(4):351–380, 2007. URL http://ideas.repec.org/a/jre/issued/v29n42007p351-380.html.

Robert B. Avery, Neil Bhutta, Kenneth P. Brevoort, and Glenn B. Canner. The mortgage market in 2011: highlights from the data reported under the home mortgage disclosure act. *Federal Reserve Bulletin*, (Sept), 2012. URL http://ideas.repec.org/a/fip/fedgrb/y2012iseptnv.98no.3.html.

Allen N. Berger, Nathan H. Miller, Mitchell A. Petersen, Raghuram G. Rajan, and Jeremy C. Stein. Does function follow organizational form? evidence from the lending practices of large and small banks. *Journal of Financial Economics*, 76(2):237–269, May 2005. URL http://ideas.repec.org/a/eee/jfinec/v76y2005i2p237-269.html.

Ben Bernanke, Mark Gertler, and Simon Gilchrist. The financial accelerator and the flight to quality. *The Review of Economics and Statistics*, 78(1):1–15, February 1996. URL http://ideas.repec.org/a/tpr/restat/v78y1996i1p1-15.html.

Ben S Bernanke and Alan S Blinder. Credit, money, and aggregate demand. *American Economic Review*, 78(2):435–39, May 1988. URL http://ideas.repec.org/a/aea/aecrev/v78y1988i2p435-39.html.

Ben S Bernanke and Mark Gertler. Inside the black box: The credit channel of monetary policy transmission. *Journal of Economic Perspectives*, 9(4):27–48, Fall 1995. URL http://ideas.repec.org/a/aea/jecper/v9y1995i4p27-48.html.

Ben S. Bernanke and Cara S. Lown. The credit crunch. *Brookings Papers on Economic Activity*, 22(2):205–248, 1991. URL http://ideas.repec.org/a/bin/bpeajo/v22y1991i1991-2p205-248.html.

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275, 2004. doi: 10.1162/003355304772839588. URL http://qje.oxfordjournals.org/content/119/1/249.abstract.

Sandra E. Black. Do better schools matter? parental valuation of elementary education. *The Quarterly Journal of Economics*, 114(2):577–599, May 1999. URL http://ideas.repec.org/a/tpr/qjecon/v114y1999i2p577-599.html.

Patrick Bolton and Mathias Dewatripont. The firm as a communication network. *The Quarterly Journal of Economics*, 109(4):809–39, November 1994. URL http://ideas.repec.org/a/tpr/qjecon/v109y1994i4p809-39.html.

Raphael W. Bostic, Kathleen C. Engel, Patricia A. McCoy, Anthony N. Pennington-Cross, and Susan M. Wachter. State and local anti-predatory lending laws: The effect of legal enforcement mechanisms. *Journal of Economics and Business*, 60(1-2):47–66, 2008. URL http://EconPapers.repec.org/RePEc:eee:jebusi:v:60:y:2008:i:1-2:p:47-66.

Kenneth P. Brevoort and Timothy H. Hannan. Commercial lending and distance: evidence from community reinvestment act data. Finance and Economics Discussion Series 2004-24, Board of Governors of the Federal Reserve System (U.S.), 2004. URL http://ideas.repec.org/p/fip/fedgfe/2004-24.html.

Kenneth P. Brevoort and John D. Wolken. Does distance matter in banking? Finance and Economics Discussion Series 2008-34, Board of Governors of the Federal Reserve System (U.S.), 2008. URL http://ideas.repec.org/p/fip/fedgfe/2008-34.html.

Jan K Brueckner. Mortgage default with asymmetric information. *The Journal of Real Estate Finance and Economics*, 20(3):251–74, May 2000. URL http://ideas.repec.org/a/kap/jrefec/v20y2000i3p251-74.html.

Paul Calem, Francisco Covas, and Jason Wu. The impact of a liquidity shock on bank lending: The case of the 2007 collapse of the private-label rmbs market. Working papers, 2011. URL http://www.federalreserve.gov/events/conferences/2011/rsr/papers/CalemCovasWu.pdf.

Charles A Calhoun and Yongheng Deng. A dynamic analysis of fixed- and adjustable-rate mortgage terminations. *The Journal of Real Estate Finance and Economics*, 24(1-2):9–33, 2002. URL http://econpapers.repec.org/RePEc:kap:jrefec:v:24:y:2002:i:1-2:p:9-33.

Murillo Campello, Erasmo Giambona, John R. Graham, and Campbell R. Harvey. Access to liquidity and corporate investment in europe during the financial crisis. *Review of Finance*, 16(2):323–346, 2012. doi: 10.1093/rof/rfr030. URL http://rof.oxfordjournals.org/content/16/2/323.abstract.

Lorenzo Cappiello, Arjan Kadareja, Christoffer Kok Srensen, and Marco Protopapa. Do bank loans and credit standards have an effect on output? a panel approach for the euro area. Working Paper Series 1150, European Central Bank, Jan 2010. URL http://ideas.repec.org/p/ecb/ecbwps/20101150.html.

David Card and Alan B Krueger. Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *American Economic Review*, 84(4):772–93, September 1994. URL http://ideas.repec.org/a/aea/aecrev/v84y1994i4p772-93.html.

Kenneth Carling and Sofia Lundberg. Bank lending, geographical distance, and credit risk: An empirical assessment of the church tower principle. Working Paper Series 144, Sveriges Riksbank (Central Bank of Sweden), Dec 2002. URL http://ideas.repec.org/p/hhs/rbnkwp/0144.html.

Nicola Cetorelli and Linda S. Goldberg. Follow the money: Quantifying domestic effects of foreign bank shocks in the great recession. *American Economic Review*, 102(3):213–18, May 2012. URL http://ideas.repec.org/a/aea/aecrev/v102y2012i3p213-18.html.

Hyun-Soo Choi. The impact of anti-predatory lending laws on mortgage volume. Working paper, Princeton University, 2011. URL http://www.princeton.edu/bcf/newsevents/seminar/Choi%20HyunSoo.pdf.

Nicolas Coeurdacier and Hlne Rey. Home bias in open economy financial macroeconomics. NBER Working Papers 17691, National Bureau of Economic Research, Inc, Dec 2011. URL http://ideas.repec.org/p/nbr/nberwo/17691.html.

Jihad Dagher and Kazim Kazimov. Banks' liability structure and mortgage lending during the financial crisis. IMF Working Papers 12/155, International Monetary Fund, Jun 2012. URL http://ideas.repec.org/p/imf/imfwpa/12-155.html.

Hans Degryse and Steven Ongena. Distance and competition. *Proceedings*, (May):405–422, 2002. URL http://ideas.repec.org/a/fip/fedhpr/y2002imayp405-422.html.

Hans Degryse and Steven Ongena. Distance, lending relationships, and competition. *Journal of Finance*, 60(1):231–266, 02 2005. URL http://ideas.repec.org/a/bla/jfinan/v60y2005i1p231-266.html.

Giovanni Dell'Ariccia. Asymmetric information and the structure of the banking industry. *European Economic Review*, 45(10):1957–1980, December 2001. URL http://ideas.repec.org/a/eee/eecrev/v45y2001i10p1957-1980.html.

Giovanni Dell'Ariccia, Deniz Igan, and Luc Laeven. Credit booms and lending standards: Evidence from the subprime mortgage market. *Journal of Money, Credit and Banking*, 44(2-3):367–384, 2012. ISSN 1538-4616. doi: 10.1111/j.1538-4616.2011.00491.x. URL http://dx.doi.org/10.1111/j.1538-4616.2011.00491.x.

Yuliya Demyanyk and Otto Van Hemert. Understanding the subprime mortgage crisis. *Proceedings*, (May):171–192, 2008. URL http://ideas.repec.org/a/fip/fedhpr/y2008imayp171-192.html.

Robert DeYoung, William W. Lang, and Daniel L. Nolle. How the internet affects output and performance at community banks. *Journal of Banking & Finance*, 31(4):1033–1060, April 2007. URL http://ideas.repec.org/a/eee/jbfina/v31y2007i4p1033-1060.html.

Robert DeYoung, Dennis Glennon, and Peter Nigro. Borrower-lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. *Journal of Financial Intermediation*, 17(1):113–143, January 2008. URL http://ideas.repec.org/a/eee/jfinin/v17y2008i1p113-143.html.

Lei Ding, Roberto Quercia, Carolina Reid, and Alan White. State antipredatory lending laws and neighborhood foreclosure rates. *Journal of Urban Affairs*, 33(4):451–467, 2011. ISSN 1467-9906. doi: 10.1111/j.1467-9906.2011.00556.x. URL http://dx.doi.org/10.1111/j.1467-9906.2011.00556.x.

Arindrajit Dube, T. William Lester, and Michael Reich. Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics*, 92(4):945–964, November 2010. URL http://ideas.repec.org/a/tpr/restat/v92y2010i4p945-964.html.

Ran Duchin, Oguzhan Ozbas, and Berk A. Sensoy. Costly external finance, corporate investment, and the subprime mortgage credit crisis. *Journal of Financial Economics*, 97(3):418–435, September 2010. URL http://ideas.repec.org/a/eee/jfinec/v97y2010i3p418-435.html.

Gregory Elliehausen and Michael E. Staten. Regulation of subprime mortgage products: An analysis of north carolina's predatory lending law. *The Journal of Real Estate Finance and Economics*, 29(4):411–433, December 2004. URL http://ideas.repec.org/a/kap/jrefec/v29y2004i4p411-433.html.

Gregory Elliehausen, Michael E. Staten, and Jevgenijs Steinbuks. The effect of prepayment penalties on the pricing of subprime mortgages. *Journal of Economics and Business*, 60(1-2):33–46, 2008. URL http://ideas.repec.org/a/eee/jebusi/v60y2008i1-2p33-46.html.

Ozgur Emre Ergungor and Stephanie Moulton. Do bank branches matter anymore? *Economic Commentary*, (Aug 4):2011–13, 2011. URL http://EconPapers.repec.org/RePEc:fip:fedcec:y:2011:i:aug4:n:2011-13.

Christopher L. Foote, Kristopher Gerardi, and Paul S. Willen. Negative equity and foreclosure: theory and evidence. Public Policy Discussion Paper 08-3, Federal Reserve Bank of Boston, 2008. URL http://ideas.repec.org/p/fip/fedbpp/08-3.html.

William F Fox. Tax structure and the location of economic activity along state borders. Technical report, 1986.

Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108(5):874–904, October 2000. URL http://ideas.repec.org/a/ucp/jpolec/v108y2000i5p874-904.html.

Mariassunta Giannetti and Luc Laeven. The flight home effect: Evidence from the syndicated loan market during financial crises. *Journal of Financial Economics*, 104(1):23–43, 2012. URL http://ideas.repec.org/a/eee/jfinec/v104y2012i1p23-43.html.

Martin Goetz and Juan Carlos Gozzi. Liquidity shocks, local banks, and economic activity: Evidence from the 2007-2009 crisis. Working papers, 2010. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1709677.

Allen C. Goodman and Brent C. Smith. Housing default: theory works and so does policy. Working Paper 10-10, Federal Reserve Bank of Richmond, 2010. URL http://ideas.repec.org/p/fip/fedrwp/10-10.html.

Ralph De Haas and Neeltje Van Horen. International shock transmission after the lehman brothers collapse: Evidence from syndicated lending. *American Economic Review*, 102(3): 231–37, May 2012. URL http://ideas.repec.org/a/aea/aecrev/v102y2012i3p231-37.html.

Kurt Schniera Hanson, Andrew and Geoffrey Turnbull. Drive until you qualify: Credit quality and household location. *Regional Science and Urban Economics*, 42(1-2):63–77, 2012. URL http://www.sciencedirect.com/science/article/pii/S0166046211000718.

Keith D. Harvey and Peter J. Nigro. Do predatory lending laws influence mortgage lending? an analysis of the north carolina predatory lending law. *The Journal of Real Estate Finance and Economics*, 29(4):435–456, December 2004. URL http://ideas.repec.org/a/kap/jrefec/v29y2004i4p435-456.html.

Robert Hauswald and Robert Marquez. Competition and strategic information acquisition in credit markets. *Review of Financial Studies*, 19(3):967–1000, 2006. URL http://rfs.oxfordjournals.org/content/19/3/967.abstract.

Giang Ho and Anthony Pennington-Cross. The impact of local predatory lending laws on the flow of subprime credit. *Journal of Urban Economics*, 60(2):210–228, September 2006. URL http://ideas.repec.org/a/eee/juecon/v60y2006i2p210-228.html.

Thomas J. Holmes. The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy*, 106(4):667–705, August 1998. URL http://ideas.repec.org/a/ucp/jpolec/v106y1998i4p667-705.html.

Bengt Holmstrom and Jean Tirole. Financial intermediation, loanable funds, and the real sector. *The Quarterly Journal of Economics*, 112(3):663–91, August 1997. URL http://ideas.repec.org/a/tpr/qjecon/v112y1997i3p663-91.html.

Rocco R. Huang. Evaluating the real effect of bank branching deregulation: Comparing contiguous counties across us state borders. *Journal of Financial Economics*, 87(3):678–705, March 2008. URL http://ideas.repec.org/a/eee/jfinec/v87y2008i3p678-705.html.

Ari Hyytinen. Information production and lending market competition. *Journal of Economics and Business*, 55(3):233–253, 2003. URL http://ideas.repec.org/a/eee/jebusi/v55y2003i3p233-253.html.

D. Immergluck. *Foreclosed: high-risk lending, deregulation, and the undermining of America's mortgage market*. G - Reference, Information and Interdisciplinary Subjects Series. Cornell University Press, 2009. ISBN 9780801447723. URL http://books.google.com.au/books?id=aBMBQm5mevQC.

Victoria Ivashina and David Scharfstein. Bank lending during the financial crisis of 2008. *Journal of Financial Economics*, 97(3):319–338, September 2010. URL http://ideas.repec.org/a/eee/jfinec/v97y2010i3p319-338.html.

Rajkamal Iyer, Samuel Lopes, Jose-Luis Peydro, and Antoinette Schoar. Interbank liquidity crunch and the firm credit crunch: evidence from the 2007-2009 crisis. Working papers, 2010. URL http://sites.kauffman.org/efic/conference/PT_CreditCrunch_20100612_Paper.pdf.

Dwight M Jaffee and Thomas Russell. Imperfect information, uncertainty, and credit rationing. *The Quarterly Journal of Economics*, 90(4):651–66, November 1976. URL http://ideas.repec.org/a/tpr/qjecon/v90y1976i4p651-66.html.

Jith Jayaratne and Philip E Strahan. The finance-growth nexus: Evidence from bank branch deregulation. *The Quarterly Journal of Economics*, 111(3):639–70, August 1996. URL http://ideas.repec.org/a/tpr/qjecon/v111y1996i3p639-70.html.

Gabriel Jimenez, Atif Mian, Jose-Luis Peydro, and Jesus Saurina. Local versus aggregate lending channels: the effects of securitization on corporate credit supply. Banco de Espaa Working Papers 1124, Banco de Espaa, Oct 2011. URL http://ideas.repec.org/p/bde/wpaper/1124.html.

Gabriel Jimnez, Vicente Salas-Fums, and Jess Saurina. Credit market competition, collateral and firms' finance. Banco de Espaa Working Papers 0612, Banco de Espaa, June 2006. URL http://ideas.repec.org/p/bde/wpaper/0612.html.

John Farris Keith Ernst and Eric Stein. North carolinas subprime home loan market after predatory lending reform. Discussion paper, Center for Responsible Lending, 2002. URL http://www.responsiblelending.org/mortgage-lending/research-analysis/HMDA_Study_on_NC_Market.pdf.

Benjamin J. Keys, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. Did securitization lead to lax screening? evidence from subprime loans. *The Quarterly Journal of Economics*, 125(1):307–362, February 2010. URL http://ideas.repec.org/a/tpr/qjecon/v125y2010i1p307-362.html.

Asim Ijaz Khwaja and Atif Mian. Tracing the impact of bank liquidity shocks: Evidence from an emerging market. *American Economic Review*, 98(4):1413–42, September 2008. URL http://ideas.repec.org/a/aea/aecrev/v98y2008i4p1413-42.html.

Anzhela Knyazeva and Diana Knyazeva. Does being your banks neighbor matter? *Journal* *of Banking and Finance*, 36(4):1194–1209, April 2012. URL http://www.sciencedirect.com/science/article/pii/S0378426611003220.

William W. Lang and Leonard I. Nakamura. 'flight to quality' in banking and economic activity. *Journal of Monetary Economics*, 36(1):145–164, August 1995. URL http://ideas.repec.org/a/eee/moneco/v36y1995i1p145-164.html.

Wei Li and Keith S Ernst. Do state predatory lending laws work? a panel analysis of market reforms. *Housing Policy Debate*, 18(2):347–392, 2007.

Elena Loutskina and Philip E. Strahan. Informed and uninformed investment in housing: The downside of diversification. *Review of Financial Studies*, 24(5):1447–1480, 2011. URL http://rfs.oxfordjournals.org/content/24/5/1447.abstract.

Cara Lown and Donald P. Morgan. The credit cycle and the business cycle: New findings using the loan officer opinion survey. *Journal of Money, Credit and Banking*, 38(6):1575–1597, September 2006. URL http://ideas.repec.org/a/mcb/jmoncb/v38y2006i6p1575-1597.html.

Christopher J. Mayer and Karen Pence. Subprime mortgages: What, where, and to whom? NBER Working Papers 14083, National Bureau of Economic Research, Inc, Jun 2008. URL http://ideas.repec.org/p/nbr/nberwo/14083.html.

Robert A McLean. Subprime mortgages: America's latest boom and bust. *Business Economics*, 44(1):57–58, January 2009. URL http://ideas.repec.org/a/pal/buseco/v44y2009i1p57-58.html.

Atif Mian. Distance constraints: The limits of foreign lending in poor economies. *Journal of Finance*, 61(3):1465–1505, 06 2006. URL http://ideas.repec.org/a/bla/jfinan/v61y2006i3p1465-1505.html.

Atif Mian and Amir Sufi. The consequences of mortgage credit expansion: Evidence from the u.s. mortgage default crisis. *The Quarterly Journal of Economics*, 124(4):1449–1496, 2009. URL http://qje.oxfordjournals.org/content/124/4/1449.abstract.

Atif Mian and Amir Sufi. House prices, home equity-based borrowing, and the us household leverage crisis. *American Economic Review*, 101(5):2132–56, August 2011. URL http://ideas.repec.org/a/aea/aecrev/v101y2011i5p2132-56.html.

Hellmuth Milde and John G Riley. Signaling in credit markets. *The Quarterly Journal of Economics*, 103(1):101–29, February 1988. URL http://ideas.repec.org/a/tpr/qjecon/v103y1988i1p101-29.html.

Taylor D. Nadauld and Shane M. Sherlund. The role of the securitization process in the expansion of subprime credit. Finance and Economics Discussion Series 2009-28, Board of Governors of the Federal Reserve System (U.S.), 2009. URL http://ideas.repec.org/p/fip/fedgfe/2009-28.html.

Adrian Pagan. Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, 25(1):221–47, February 1984. URL http://ideas.repec.org/a/ier/iecrev/v25y1984i1p221-47.html.

Daniel Paravisini. Local bank financial constraints and firm access to external finance. *Journal of Finance*, 63(5):2161–2193, October 2008. URL http://ideas.repec.org/a/bla/jfinan/v63y2008i5p2161-2193.html.

Joe Peek and Eric Rosengren. Bank regulation and the credit crunch. *Journal of Banking and Finance*, 19(3-4):679–692, June 1995. URL http://ideas.repec.org/a/eee/jbfina/v19y1995i3-4p679-692.html.

Joe Peek and Eric S. Rosengren. Collateral damage: Effects of the japanese bank crisis on real activity in the united states. *American Economic Review*, 90(1):30–45, March 2000. URL http://ideas.repec.org/a/aea/aecrev/v90y2000i1p30-45.html.

Karen M. Pence. Foreclosing on opportunity: State laws and mortgage credit. *The Review of Economics and Statistics*, 88(1):177–182, April 2006. URL http://ideas.repec.org/a/tpr/restat/v88y2006i1p177-182.html.

Anthony Pennington-Cross and Giang Ho. Predatory lending laws and the cost of credit. *Real Estate Economics*, 36(2):175–211, 06 2008. URL http://ideas.repec.org/a/bla/reesec/v36y2008i2p175-211.html.

Anthony Pennington-Cross and Giang Ho. The termination of subprime hybrid and fixed-rate mortgages. *Real Estate Economics*, 38(3):399–426, 2010. URL http://ideas.repec.org/a/bla/reesec/v38y2010i3p399-426.html.

Mitchell A Petersen and Raghuram G Rajan. The effect of credit market competition on lending relationships. *The Quarterly Journal of Economics*, 110(2):407–43, May 1995. URL http://ideas.repec.org/a/tpr/qjecon/v110y1995i2p407-43.html.

Mitchell A. Petersen and Raghuram G. Rajan. Does distance still matter? the information revolution in small business lending. *Journal of Finance*, 57(6):2533–2570, December 2002. URL http://ideas.repec.org/a/bla/jfinan/v57y2002i6p2533-2570.html.

Alexander Popov and Gregory F. Udell. Cross-border banking and the international transmission of financial distress during the crisis of 2007-2008. Working Paper Series 1203, European Central Bank, Jun 2010. URL http://ideas.repec.org/p/ecb/ecbwps/20101203.html.

Manju Puri, Jrg Rocholl, and Sascha Steffen. On the importance of prior relationships in bank loans to retail customers. Working Paper Series 1395, European Central Bank, Nov 2011. URL http://ideas.repec.org/p/ecb/ecbwps/20111395.html.

Amiyatosh Purnanandam. Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies*, 2010. doi: 10.1093/rfs/hhq106. URL http://rfs.oxfordjournals.org/content/early/2010/10/13/rfs.hhq106.abstract.

Amiyatosh Purnanandam. Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies*, 24(6):1881–1915, 2011. URL http://rfs.oxfordjournals.org/content/24/6/1881.abstract.

Philip Strahan Qian, Jun and Zhishu Yang. The impact of organizational and incentive structures on soft information: Evidence from bank lending. Technical report, 2010. URL https://www2.bc.edu/~qianju/China-Bank-QSY-12Nov10.pdf.

Roberto G. Quercia, Michael A. Stegman, and Walter R. Davis. The impact of predatory loan terms on subprime foreclosures: The special case of prepayment penalties and balloon payments, 2005.

Roy Radner. The organization of decentralized information processing. *Econometrica*, 61(5):1109–46, September 1993. URL http://ideas.repec.org/a/ecm/emetrp/v61y1993i5p1109-46.html.

Uday Rajan, Amit Seru, and Vikrant Vig. Statistical default models and incentives. *American Economic Review*, 100(2):506–10, May 2010. URL http://ideas.repec.org/a/aea/aecrev/v100y2010i2p506-10.html.

Richard J. Rosen. The impact of the originate-to-distribute model on banks before and during the financial crisis. Working Paper Series WP-2010-20, Federal Reserve Bank of Chicago, 2010. URL http://ideas.repec.org/p/fip/fedhwp/wp-2010-20.html.

Philipp Schnabl. The international transmission of bank liquidity shocks: Evidence from an emerging market. *The Journal of Finance*, 67(3):897–932, 2012. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2012.01737.x. URL http://dx.doi.org/10.1111/j.1540-6261.2012.01737.x.

Sherrill Shaffer. The winner's curse in banking. *Journal of Financial Intermediation*, 7 (4):359–392, October 1998. URL http://ideas.repec.org/a/eee/jfinin/v7y1998i4p359-392.html.

Jeremy C. Stein. An adverse-selection model of bank asset and liability management with implications for the transmission of monetary policy. *RAND Journal of Economics*, 29(3):466–486, Autumn 1998. URL http://ideas.repec.org/a/rje/randje/v29y1998iautumnp466-486.html.

Jeremy C. Stein. Information production and capital allocation: Decentralized versus hierarchical firms. *Journal of Finance*, 57(5):1891–1921, October 2002. URL http://ideas.repec.org/a/bla/jfinan/v57y2002i5p1891-1921.html.

Joseph E Stiglitz and Andrew Weiss. Credit rationing in markets with imperfect information. *American Economic Review*, 71(3):393–410, June 1981. URL http://ideas.repec.org/a/aea/aecrev/v71y1981i3p393-410.html.

Xuan Tian. The causes and consequences of venture capital stage financing. *Journal of Financial Economics*, 101(1):132–159, July 2011. URL http://ideas.repec.org/a/eee/jfinec/v101y2011i1p132-159.html.

David C. Webb. Imperfect information and credit market equilibrium. *European Economic Review*, 26(1-2):247–258, 1984. URL http://ideas.repec.org/a/eee/eecrev/v26y1984i1-2p247-258.html.