The impact of asymmetric information about collateral values in mortgage lending*

Johannes Stroebel‡
Stanford University

JOB MARKET PAPER

Abstract

I empirically analyze the sources and magnitude of asymmetric information between competing lenders in residential mortgage lending. I exploit that property developers often cooperate with vertically integrated mortgage lenders to provide financing to buyers of their newly constructed homes. These integrated lenders might have superior information about both mortgage collateral quality and borrower characteristics. I construct a dataset of all housing transactions and associated mortgages in Arizona between 2000 and 2010. This allows me to test for asymmetric information by comparing the return of initially similar houses in the same development financed by different lenders. I find that houses financed by an integrated lender outperform similar houses financed by non-integrated competitors by 50 basis points annually. They are also less likely to enter into foreclosure. These differences persist during the ownership period of the second owner of the house. The outperformance of houses financed by an integrated lender is over twice as large amongst houses built on expansive soil, which makes housing return more sensitive to construction quality. Non-integrated lenders charge 10 basis points higher interest rates when competing against an integrated lender. This interest rate increase is larger for mortgages with a high loan-to-value ratio, for which repayment is more sensitive to subsequent changes in collateral value. These results are highly consistent with the presence of significant asymmetric information about collateral quality in mortgage lending.

*This version: November 8, 2011. For latest version, see www.stanford.edu/~stroebel/J_Stroebel_JMP.pdf. I am deeply indebted to my advisors Caroline Hoxby, Monika Piazzesi, Martin Schneider and John B. Taylor for their encouragement and guidance. I also benefited from conversations with Ran Abramitzky, Christoph Basten, Juliane Begenau, Doug Bernheim, Nick Bloom, Tim Bresnahan, Pascaline Dupas, Liran Einav, Bob Hall, Richard Hornbeck, Pete Klenow, Steven Kohlhagen, Siddharth Kothari, Theresa Kuchler, Pablo Kurlat, Tim Landvoigt, Juliana Salomao, Florian Scheuer, Paulo Somaini, Luke Stein, Arthur van Benthem, Gui Woolston and Ian Wright. Numerous seminar participants at Stanford and Trulia provided insightful comments. I thank Trulia for providing data. Financial support from the Kohlhagen Fellowship and the Bradley Fellowship through SIEPR, as well as the Working Group on Economic Policy at the Hoover Institution is gratefully acknowledged.

‡Stanford University, Department of Economics, 579 Serra Mall, Stanford, CA 94305, stroebel@stanford.edu.

1 Introduction

The impact of asymmetric information in financial markets has long been of interest to economists and the recent financial crisis has intensified research on the effects of asymmetric information in mortgage lending and securitization (e.g. Elul, 2011; Jiang et al., 2010; Keys et al., 2010). Most of this research has focused on analyzing asymmetric information about characteristics of the borrower such as their income prospects. The default risk of mortgages, however, depends both on the value of the housing collateral as well as on the borrower's ability to make interest payments (Deng et al., 2000). This importance of collateral values in determining mortgage default was particularly visible during the recent crisis, which precipitated many "strategic defaults" in which households that could afford to pay their mortgages chose to default once collateral values fell below the outstanding mortgage balance (Guiso et al., 2011).[1] Due to the highly illiquid and heterogeneous nature of housing as an asset it is also possible that there is asymmetric information about collateral values in mortgage lending, in particular given the significant resources that lenders spend on appraisals and property inspections to improve their valuation of the housing collateral prior to making a lending decision.

In this paper I empirically analyze the sources of asymmetric information between lenders that compete to originate mortgages used to purchase newly developed properties. This market is often characterized by the presence of a mortgage lender that is vertically integrated with the property developer and might have superior information about both borrower characteristics and collateral quality. Perhaps surprisingly, I find that asymmetric information about collateral quality is a significant source of adverse selection. In addition to testing for the presence of asymmetric information and uncovering its sources, my empirical approach allows me to quantify the impact of this asymmetric information on the cost of mortgages, which I find to be significant.

From a policy perspective, the identification of collateral values as a key source of asymmetric information in mortgage lending helps to develop and assess policy proposals to improve the functioning of this market. For example, it suggests that better credit scoring technology and the more extensive sharing of borrower information through credit bureaus (see, for example, Jappelli and Pagano, 2002) will not address all forms of asymmetric information. In addition, while the results of Keys et al. (2010) suggest that only granting full documentation mortgages might remove some asymmetric information about borrower characteristics, policies that address asymmetric information about collateral quality are also important. Indeed, a stronger focus on providing independent and reliable property assessments to all market participants might play an important role in mitigating the impact of asymmetric information in mortgage lending.

[1] The debate about the relative importance of illiquidity and negative equity in explaining recent trends in mortgage default is ongoing. Bajari et al. (2008) find evidence for both factors. Their results suggest that for a borrower with a 30-year fixed-rate mortgage and no downpayment, a 20% decline in house prices increases the probability of default by over 15% relative to a borrower whose housing collateral does not change in value. The survey data analyzed by Guiso et al. (2011) suggest that over one third of the mortgage defaults reported by the respondents in the September 2010 wave of the survey were strategic.

The competition between lenders to finance the purchase of newly developed residential properties is an ideal setting to examine asymmetric information in mortgage lending. Large property developers regularly provide home buyers with financing offers through a vertically integrated mortgage lender. Such integrated lenders are usually jointly owned with the developer and are likely to have more precise information than non-integrated lenders about the value of the house that is used to collateralize the mortgage. For example, an integrated lender might have access to the developer's information about aspects of construction quality that are difficult for non-integrated lenders to observe. In the process of guiding a borrower through the home purchase process an integrated lender might also acquire superior information about borrower characteristics such as their propensity to maintain the property.

I present a theoretical model to analyze the mortgage lending competition between an integrated lender and other non-integrated lenders, and use this model to generate a number of empirical predictions that I test in the main part of the paper. In the model, an integrated lender obtains an informative signal about the quality of the mortgage collateral, while competing lenders only know the average collateral quality. By conditioning its financing offer on its superior information, the integrated lender subjects non-integrated lenders to adverse selection. The portfolio of houses financed by an integrated lender should thus outperform a portfolio of ex-ante similar houses financed by non-integrated lenders. Non-integrated lenders need to charge a higher interest rate to break even than if they were to compete only against equally informed lenders. This raises the financing costs for buyers of newly developed properties. These effects are bigger in magnitude when the integrated lender's signal about collateral quality is more precise. The model also shows that the financing costs should rise by a larger amount for borrowers whose repayment is more sensitive to changes in collateral values, for example because they only make a small downpayment.

The key contribution of this paper is to show empirically that such asymmetric information between competing lenders about collateral values plays an important role in the financing of newly developed homes. I construct a dataset of all housing transactions and associated mortgages in Arizona between 2000 and 2010 to track the return of properties following their initial sale. About 85% of newly developed homes are in developments with an active integrated lender, and, when present, the average market share of these integrated lenders is about 70%. I find that in developments with an integrated lender, the portfolio of houses financed by the integrated lender outperforms a portfolio of ex-ante similar houses in the same development financed by non-integrated lenders by about 50 basis points annually. I also find that mortgages granted by the integrated lender are over 30% less likely to enter into foreclosure than ex-ante similar mortgages granted by non-integrated lenders. These estimates provide a measure of the extent of asymmetric information between competing lenders.

An important finding is that the outperformance of the integrated lender's collateral portfolio is especially large (over 100 basis points annually) amongst houses built on "expansive soil," a high clay content soil that makes the future price development more sensitive to unobservable

aspects of construction quality such as the care with which the foundation was poured. This result suggests that the construction quality of the housing collateral is a significant source of asymmetric information. The outperformance of houses financed by the integrated lender is also bigger when the borrower makes a large downpayment, which makes mortgage repayment less sensitive to changes in the house value. As a result, non-integrated lenders make fewer adjustments to their interest rate offers to avoid the winner's curse and in equilibrium lend against lower quality collateral.

To further analyze the sources of the asymmetric information that drives these differences in collateral return, I compare the collateral return and the foreclosure probability for the ownership duration of the *second* owner of these properties. The relative outperformance of those houses *initially* financed by the integrated lender remains. This result suggests that the outperformance of the integrated lender's collateral portfolio is to a large extent explained by asymmetric information about the housing collateral, not the borrower, since the identity of a possible second owner of the house was not known to *any* lender at the time of granting the mortgage to the initial owner.[2]

I also analyze the cost to borrowers in terms of higher interest rates that result from this asymmetric information. I find that non-integrated lenders charge an average interest rate premium of 10 basis points annually for otherwise similar mortgages when they compete against an integrated lender. This higher interest rate compensates non-integrated lenders for the adverse selection in the presence of an integrated lender which reduces the quality of their collateral portfolio. The interest rate increase is larger, at 25 basis points, for mortgages to purchase houses built on expansive soil. The return of those houses is particularly sensitive to aspects of construction quality about which the integrated lender could have superior information. As predicted by the model, the interest rate increase is also larger for mortgages with a low downpayment, rising to about 50 basis points annually for mortgages with a downpayment of less than 3%. For those mortgages the repayment probability is more sensitive to changes in collateral values. Non-integrated lenders thus need to charge higher interest rates to break even when facing adverse selection on collateral quality.

Understanding the sources of asymmetric information is important because such asymmetric information has the potential to disrupt lending markets significantly, in particular during periods of falling house prices. Fishman and Parker (2010) discuss a model in which asymmetric information about collateral values can introduce non-fundamental volatility in the supply of financing. Gorton and Ordonez (2011) show how in the presence of asymmetric information about collateral values, small shocks to the perceived average value of the collateral can lead to declines in output and consumption.[3] The current paper shows empirically that asymmetric information about collateral

[2]This specification also shows that my results are not driven by an initial price bundling of the mortgage and the house. Such bundling could be a concern, since any discounts on the house given to customers of the integrated lender would be observationally equivalent to a true collateral outperformance when the house gets subsequently sold. However, any such discounts would be capitalized in the transaction price between the first and second owners and should thus not contaminate the observed collateral return during the ownership of the second owner.

[3]This can have significant aggregate implications. In the United States, private residential fixed investment contributed about 6% of GDP in 2006, falling to about 3% in 2009. Over the same time period the annual construction of privately owned housing units fell from just under 2 million to about 800,000. The majority of these new houses

quality is a key feature of mortgage lending, and that thus the theoretical concerns of Fishman and Parker (2010) and Gorton and Ordonez (2011) might be relevant in this market. Mortgage lending in new developments also provides a fascinating environment that helps to understand lending competition under asymmetric information about collateral values in other financial market settings with a similar information structure. One example is the practice of sell-side advisors to offer "stapled financing" packages to buyers in many M&A transactions.[4] Since the bank providing the stapled financing offer advises on the sale of the asset, it presumably has superior information about the quality of the assets that underlie the loan. Consequently, one might expect the fear of adverse selection by other lenders to constrain the aggressiveness of their financing offers. This could lead to higher financing costs for buyers of assets that contain a stapled financing offer. Another example are repo markets that use non-standardized collateral such as mortgage-backed securities (MBS). The bank that originally assembled the MBS will have superior information about its quality. It can consequently subject other banks that might choose to lend against this MBS as repo collateral to adverse selection similar to that discussed in this paper.

1.1 Related Literature

This paper relates to an empirical literature that analyzes the impact on interest rates in corporate lending when one bank has superior information about a creditor firm. One set of papers considers the effect of relationship duration between borrower and lender and tests whether loan rates increase with relationship length as the information advantage of the incumbent bank increases. The empirical evidence is mixed. Petersen and Rajan (1994) find no effect of relationship length on loan rates in their analysis of the National Survey of Small Business Finance. Berger and Udell (1995) also analyze lending to small businesses and find that borrowers with longer relationships pay *lower* interest rates and are *less* likely to pledge collateral. Degryse and Van Cayseele (2000) consider loans to small businesses in Belgium and find loan rates to increase with relationship duration. Ioannidou and Ongena (2010) analyze a detailed dataset of loans in Bolivia and provide evidence that once a borrower is informationally captured, the bank will increase interest rates. A second set of papers considers physical proximity between banks and firms as a source of superior information for nearby lenders. Degryse and Ongena (2005) find that loan rates decrease with physical distance between the lending bank and the borrower. They argue that this is consistent with a model of information asymmetries that vary in the distance between lender and borrower. Petersen and Rajan (2002) and Agarwal and Hauswald (2010) also find a negative relationship between distance to the bank and loan rates. In my paper the information advantage of the integrated lender arises through its relationship with the developer. In addition to analyzing the price impact of the asymmetric

(about 70%) are built by developers such as the ones studied in this paper, which jointly sell the land and the house.

[4]Stapled financing consists of a loan commitment by a sell-side advisor which is available at pre-specified terms to any buyer of the asset. The loan's claims are only on the target's assets and cash flow. In 2004, stapled financing was provided in almost 40% of transactions in the U.S. that involved a private equity firm (Povel and Singh, 2010).

information, my data allows me to directly measure differences in collateral return.

Other papers consider the role of competition under asymmetric information in non-financial market settings. Hendricks and Porter (1988) analyze 144 auctions for oil field leases. They find evidence that a firm that has previously won a neighboring tract has superior information about the productivity of the currently auctioned tract and that this is represented in the bidding behavior of firms. Greenwald's (1986) analysis of adverse selection in the labor market considers a similar informational environment. Current employers have better information about the productivity of a worker and will try to keep high-quality workers from leaving. The set of job changers should thus be disproportionately composed of less able workers, which reduces wages for current employees.

Different aspects of the behavior of integrated lenders have been analyzed in a number of recent papers. Barron et al. (2008) discuss a model of an integrated lender that takes into account the return from making a loan as well as from selling the product, and would thus optimally set lower lending standards. They present evidence that car loans by integrated lenders default more often. Gartenberg (2011) analyzes whether integrated mortgage lenders lowered their lending standards during the housing boom in order to sell more houses. Consistent with my results, she finds that mortgages granted by integrated lenders were actually *less* likely to default. She concludes that this is might be explained by integrated lenders' organizational choices that reduced a loan officer's incentives to approve marginal applications, or by its desire to originate mortgages that could be easily securitized. Such incentive effects are a complementary explanation for lower default rates amongst integrated lender mortgages, but do not predict some of my key results.[5] Pierce (2011) shows that car leasing firms that are affiliated with a manufacturer have superior information about the timing of new model introductions, which allows them to profitably adjust the lease pricing of existing models, but demonstrates that their primary incentive is to subsidize the sale of low-quality vehicles. I argue that integrated mortgage lenders possess and exploit superior information about mortgage collateral quality, and consider the effect of this asymmetric information on loan pricing.

Asymmetric information about property values has been considered in previous empirical work, though most research focuses on its impact on the sales transaction rather than the financing process. Levitt and Syverson (2008) analyze the interaction between a home seller and her real estate agent who has better information about the value of the house. They show that real estate agents exploit this information asymmetry by advising homeowners to sell too quickly relative to when the agents sell their own home. Wong et al. (2011) argue that real estate markets in Hong Kong are more liquid when the more readily observable component of land makes up a larger fraction of total property values, reducing the scope for market shut-downs due to asymmetric information. Garmaise and Moskowitz (2004) focus on the commercial real estate market. They

[5]Incentive effects do not predict differential capital gains on housing collateral across lender types, in particular by soil quality and loan-to-value ratio. The latter are central predictions of a model with information asymmetries. In addition, I find that non-integrated lenders raise interest rates in response to the adverse selection on collateral quality, something one would not expect if the incentives of the non-integrated lenders' loan officers were to originate mortgages irrespective of collateral quality.

use regional variation in the quality of tax assessments to proxy for the importance of private information about property values and show that properties with less informative assessments attract more local buyers whose geographic proximity allows them to obtain a better valuation of the property. Ben-David (2011) shows that during the housing boom financially constrained home buyers exploited lenders' uncertainty about house values and paid inflated transaction prices while receiving non-collateralized side payments. This allowed them to to draw larger mortgages and obtain higher effective loan-to-value ratios than would have been possible if banks had had a more precise valuation of the property.

2 Theoretical Model

In this section I present a theoretical model of competition between differentially informed lenders to provide mortgage financing. This model adapts similar models by Sharpe (1990), Broecker (1990), von Thadden (2004) and Hauswald and Marquez (2006), as well as Engelbrecht-Wiggans et al.'s (1983) analysis of first-price sealed-bid common value auctions with differentially informed bidders. I first characterize the equilibrium interest rate offers of the integrated lender and competing non-integrated lenders. I then simulate the model to generate empirical predictions about the equilibrium quality of each lender's collateral portfolio, as well as the observed interest rates charged by non-integrated lenders when competing against an integrated lender. These predictions are tested in the empirical analysis, which constitutes the key contribution of this paper.

Houses: There is a unit mass of houses. Each house costs \$1 to purchase. Houses are either high quality ($\theta = h$) or low quality ($\theta = l$). High quality houses will be worth $H > 1$ with certainty next period. Low quality houses will be worth $L = 0$. Final house value is observable, but house type θ is unknown ex-ante. The fraction of houses that is high quality, q, is common knowledge.

Households: There is a unit mass of risk-neutral households. Households either live in a purchased house or in rented housing, the cost of which is normalized to zero. Households have no private resources in the first period and thus require a mortgage to purchase a house. Borrowers are indexed by γ, which represents the probability that they will repay the mortgage when the housing collateral declines in value (i.e. $\theta = l$). A borrower's γ is common knowledge.[6] The household's expected return from taking out a mortgage at rate R is equal to $q(H - R) - (1 - q)\gamma R$, which has to be bigger than zero, the cost of renting. This means that the maximum interest rate that a household is prepared to accept is $R(\gamma)_m = \frac{qH}{q+(1-q)\gamma}$.

Lenders: There are two types of risk-neutral lenders, an integrated lender that has some private

[6] In a more complex model, whether to repay the mortgage when collateral values fall would be an equilibrium decision by the household, not a characteristic. Stating γ as a characteristic is a reduced form way of capturing the probability of states of the world for which the household would optimally choose not to default when collateral values fall. In the empirical analysis, a borrower's initial downpayment proxies for γ: the larger the downpayment, the more the house can decline in value before the owner has an incentive to default. Another interpretation of γ is that it captures the agent's creditworthiness though unfortunately I do not observe a good proxy for this in the data.

information about the house and N non-integrated lenders that only know q. The private information of the integrated lender consists of a signal $\eta \in \{h, l\}$.[7] The precision of the signal is defined as $\phi = P(\eta = h|\theta = h) = P(\eta = l|\theta = l) > \frac{1}{2}$.[8] All lenders have access to funds at rate $R_f < qH$.[9]

Timing of the game: The households apply to the integrated lender and N non-integrated lenders for a mortgage. All lenders observe γ and q. The integrated lender also observes signal η. Lenders then compete by simultaneously offering loans at interest rate R. Lenders can also choose not to make an offer. Borrowers accept the lowest offer as long as it is below $R(\gamma)_m$.[10]

2.1 Equilibrium

I look for a Bayesian Nash equilibrium. Since the sensitivity of repayment with respect to collateral value, γ, is perfectly observable by all agents, I can solve the equilibrium separately for each value of γ. I then compare equilibrium outcomes across γ-types. A standard feature of these models is that the equilibrium bidding strategies of individual non-integrated lenders are indeterminate. What is determinate is the minimum of all non-integrated lenders' bids. Hence solving an equilibrium with many uninformed lenders is equivalent to solving the equilibrium of competition between the integrated lender and one representative non-integrated lender. The interest rate offer distribution of the representative non-integrated lender captures the minimum of the offers of the N non-integrated lenders (Engelbrecht-Wiggans et al., 1983).

Theorem 1 *There are no pure strategy equilibria.*

Proof Proof in Appendix A.

[7] I assume that integrated lenders cannot credibly convey this signal to borrowers, since borrowers know that the lender has an incentive to over-state the quality of the house in order to prompt them to accept $R > R(\gamma)_m$.

[8] In other words, ϕ captures the importance of the integrated lender's information in predicting movements in collateral values. In the empirical analysis I exploit geographic differences in soil type to proxy for ϕ. As explained in Section 3, low construction quality is especially problematic for houses built on expansive soil and can lead to particularly large declines in the value of those houses. I hence argue that the integrated lender's information about construction quality is particularly relevant in forecasting the price development of houses built on expansive soil.

[9] Barron et al. (2008) present some evidence that the marginal cost of funds between 1998 and 2002 were quite similar for banks and finance companies.

[10] This timing assumption makes the interest rate offer game resemble a first-price sealed-bid auction. If a non-integrated lender was able to observe the integrated lender's offer before making its own offer, it might be able to partially infer the integrated lender's signal (Hoerner and Jamison, 2008). However, while this game may not represent the optimal search strategy for the consumer, who might benefit from shopping around with the integrated lender's offer, it represents the actual mortgage shopping process of individuals reasonably well. Offered interest rates are "locked in" by lenders for a short period only. To have more time for comparing rates and attracting offers sequentially, borrowers have to pay to extend the "lock period." In a sample rate sheet for April 2000, studied by Woodward and Hall (2010), an extension of the lock period by 15 days costs 0.125% of the principal amount. In addition, people may not realize that they might be able to use the offer from the integrated lender to negotiate a better offer elsewhere. Woodward and Hall (2010) analyze the mortgage shopping behavior of individuals and find that most borrowers consider no more than two offers. The benefits from more search are so large that they conclude that it must be "confusion about how this market works that caused borrowers to shop too little." Another assumption of this set-up is that borrowers do not extract information from the integrated lender's offer about the quality of the house they purchase. Since interest rates vary with a large number of characteristics such signal extraction would be extremely complex, and, given the discussion above, probably beyond the skills of most borrowers.

The intuition for the absence of pure strategy equilibria proceeds in two steps: First, note that if a pure strategy equilibrium existed, both lenders would have to offer the mortgage at the same interest rate $\tilde{R}$. If one lender offered credit at an interest rate lower than the other lender, it could increase its payoff by raising its rate by ε. However, both lenders offering the same interest rate cannot constitute an equilibrium. If, conditional on observing η, it is profitable to lend at $\tilde{R}$, then the integrated lender would offer $\tilde{R} - \varepsilon$ and capture the entire market. If, conditional on η, it is unprofitable to lend at $\tilde{R}$, the integrated lender would increase its interest rate offer and subject the less informed lender to a winner's curse, leaving it with an expected loss. In other words, pure strategies are vulnerable to selective undercutting by the more informed party and cannot constitute an equilibrium for the non-integrated lender (von Thadden, 2004).

Theorem 2 *Let $W(R;\eta,\phi,\gamma)$ be the expected revenue to the integrated lender from lending at rate R to a type-γ borrower to buy a house with signal η. The interest rate offer game for a type-γ borrower when signal precision is ϕ has a unique mixed strategy equilibrium, such that:*

1. *The non-integrated lender breaks even, while the integrated lender earns positive expected profits.*

2. *There exists a value $\bar{\gamma}$ such that for borrowers with $\gamma < \bar{\gamma}$ the integrated lender rejects all mortgage applications to buy houses with $\eta = l$. When $\eta = h$, the integrated lender randomizes interest rate offers over $[R(\gamma)^b_a, R(\gamma)_m)$ using the following cumulative distribution function:*

$$F_i(R;h,\phi,\gamma) = 1 + \frac{P_i(l)[W(R;l,\phi,\gamma) - R_f]}{P_i(h)[W(R;h,\phi,\gamma) - R_f]}. \tag{1}$$

   $R(\gamma)^b_a = \frac{R_f}{q+\gamma(1-q)}$ is the break-even interest rate for lending to a type-γ agent to buy an average quality house. $P_i(\eta)$ is the probability of the integrated lender observing signal η. The integrated lender also makes interest rate offers with a point mass of $1 - F_i(R(\gamma)_m;h,\phi,\gamma)$ at $R(\gamma)_m$. The non-integrated lender randomizes interest rate offers over $[R(\gamma)^b_a, R(\gamma)_m)$ using the following cumulative distribution function:

$$F_u(R;\phi,\gamma) \quad = \quad 1 - \frac{W(R(\gamma)^b_a;h,\phi,\gamma) - R_f}{W(R;h,\phi,\gamma) - R_f}. \tag{2}$$

   With probability $1 - F_u(R(\gamma)_m;\phi,\gamma)$ the non-integrated lender does not make an offer.

3. *For borrowers with $\gamma > \bar{\gamma}$ both the integrated lender and the non-integrated lender always offer a mortgage. When $\eta = l$ the integrated lender offers the break-even interest rate $R(\gamma,\phi)^b_l$, defined implicitly by $R_f = W(R(\gamma,\phi)^b_l;l,\phi,\gamma)$. When $\eta = h$ the integrated lender randomizes its interest rate offers over $[R(\gamma)^b_a, R(\gamma)_m]$ using $F_i(R;h,\phi,\gamma)$. The non-integrated lender always randomizes over $[R(\gamma)^b_a, R(\gamma,\phi)^b_l)$ using $F_u(R;\phi,\gamma)$, with a point mass at $R(\gamma,\phi)^b_l$.*

Proof Proof in Appendix A.

There are a number of interesting observations about the unique mixed strategy equilibrium interest rate offers.[11] Consider the integrated lender's behavior when it receives a negative signal about house quality. At each level of signal precision, there exists a value $\bar{\gamma}$ such that when observing $\eta = l$ it is no longer profitable to lend to households with $\gamma < \bar{\gamma}$, even at $R(\gamma)_m$, the highest interest rate that borrowers are prepared to accept. This is because for low values of γ the probability of repayment is particularly sensitive to the value of the collateral. The cut-off value $\bar{\gamma}$ is increasing in the precision of the signal, ϕ, since for a more precise signal the probability of the collateral being of high quality when $\eta = l$ declines. For $\gamma \geq \bar{\gamma}$, the probability of repayment when $\eta = l$ remains high enough to make an expected profit when lending at $R(\gamma)_m$. This is because the probability of repayment is less sensitive to movements in collateral value. To these households the integrated lender always offers a mortgage at $R(\gamma, \phi)_l^b$, defined as the break-even interest rate when lending to a type-γ household purchasing a house with signal $\eta = l$. If it offered a higher rate R_i, the non-integrated lender could make an expected profit by always offering to lend at $R_i - \varepsilon > R(\gamma, \phi)_l^b$. This is the standard Bertrand undercutting argument. If it offered a lower interest rate, it would make an expected loss. When the integrated lender receives signal $\eta = h$ it makes mixed strategy interest rates offers using the cumulative distribution function F_i. The non-integrated lender does not receive a signal on which it could condition its interest rate offer and thus always makes mixed strategy interest rates offers using the cumulative distribution function F_u.

2.2 Empirical Predictions from Equilibrium Bank Behavior

To analyze equilibrium outcomes when lenders use the mixed strategies described in Theorem 2 I simulate the game for a range of parameter values.[12] This generates predictions about the expected quality of the equilibrium collateral portfolio of each lender and about equilibrium interest rates. I also show how collateral quality and interest rates vary with different values of γ, the probability of repayment when collateral values fall, and ϕ, the signal precision.

The top row of Figure 1 plots the expected period-2 value of the equilibrium portfolios of houses financed by the two lenders as a function of ϕ and γ. The dashed line represents the integrated lender's portfolio, the solid line the non-integrated lender's portfolio. The dotted line shows the unconditional expected house price, qH. For all values of γ and ϕ the houses financed by the integrated lender are more likely to increase in value than those financed by the non-integrated lender (i.e. the dashed line is above the solid line). This is a direct result of the integrated lender conditioning its interest rate offers on the informative signal and the subsequent adverse selection.

[11] See Rubinstein (1991) for a discussion of possible interpretions of mixed strategy equilibria.

[12] To do this, I consider $100,000$ hypothetical mortgage applicants that apply for financing from the integrated lender and the non-integrated lender. A fraction q of agents apply to buy a house of high quality. When the agent applies to the integrated lender, the lender draws an informative signal η which has known precision ϕ. Both lenders draw an interest rate offer from their equilibrium distribution as defined in Theorem 2. The borrower accepts the lowest offer. The parameters of the economic environment are chosen such that $\bar{\gamma} < 0$, which means that $\gamma > \bar{\gamma}$ and all borrowers will receive an offer. The comparative statics are the same for $0 \leq \bar{\gamma} \leq 1$ and are available on request.

This implication is formalized in Prediction 1.

Prediction 1: *The average ex-post return of houses financed by an integrated lender is higher than the return of ex-ante similar (conditional on a non-integrated lender's information set) homes financed by a non-integrated lender.*

In the absence of an integrated lender, all lenders are equally informed about collateral quality. Each lender thus lends against collateral of average quality, with an expected period-2 value given by the dotted line. The non-integrated lender's expected collateral return in the presence of an integrated lender is thus lower than if it were only competing against equally informed lenders (i.e. the dashed line is below the dotted line). This is formalized in Prediction 2.

Prediction 2: *The average ex-post return of houses financed by a non-integrated lender is lower when it competes against an integrated lender relative to when it competes only against equally informed lenders.*

The middle row of Figure 1 plots the average interest rate spread over R_f for the non-integrated lender's mortgages (dashed line). It also shows the spread of $R(\gamma)_a^b$, the break-even interest rate for lending against average quality collateral (solid line). When all lenders are equally informed about collateral quality, Bertrand competition drives the interest rate down to $R(\gamma)_a^b$ where expected profits for all lenders are zero. When competing against a better informed integrated lender, a non-integrated lender lends against below average quality collateral. To ensure it continues to break even, the non-integrated lender must charge a higher average interest rate than when it competes only against equally informed lenders (i.e. the dashed line is above the solid line). This is formalized in Prediction 3.

Prediction 3: *A non-integrated lender charges a higher interest rate when it competes against an integrated lender relative to when it competes only against equally informed lenders.*

I next consider how equilibrium outcomes vary with ϕ, the precision of the integrated lender's signal of collateral quality. This is shown in the left column of Figure 1. The top left panel shows that the expected period-2 value of houses financed by the integrated lender is increasing in ϕ. The integrated lender only lends when it receives a good signal ($\eta = h$). As ϕ increases, this signal becomes more precise and allows the integrated lender to more accurately identify high and low quality collateral. The non-integrated lender correspondingly lends against ever lower quality collateral, both relative to the integrated lender and relative to the unconditional expected collateral quality. To continue to break even it needs to charge a higher equilibrium interest rate on the mortgages it makes. This is shown in the middle left panel of Figure 1. The resulting expected per-mortgage equilibrium profits are presented in the bottom left panel of Figure 1. The non-integrated lender always just breaks even while the integrated lender's profits are increasing in the precision of its signal. These insights are formalized in the following three predictions:

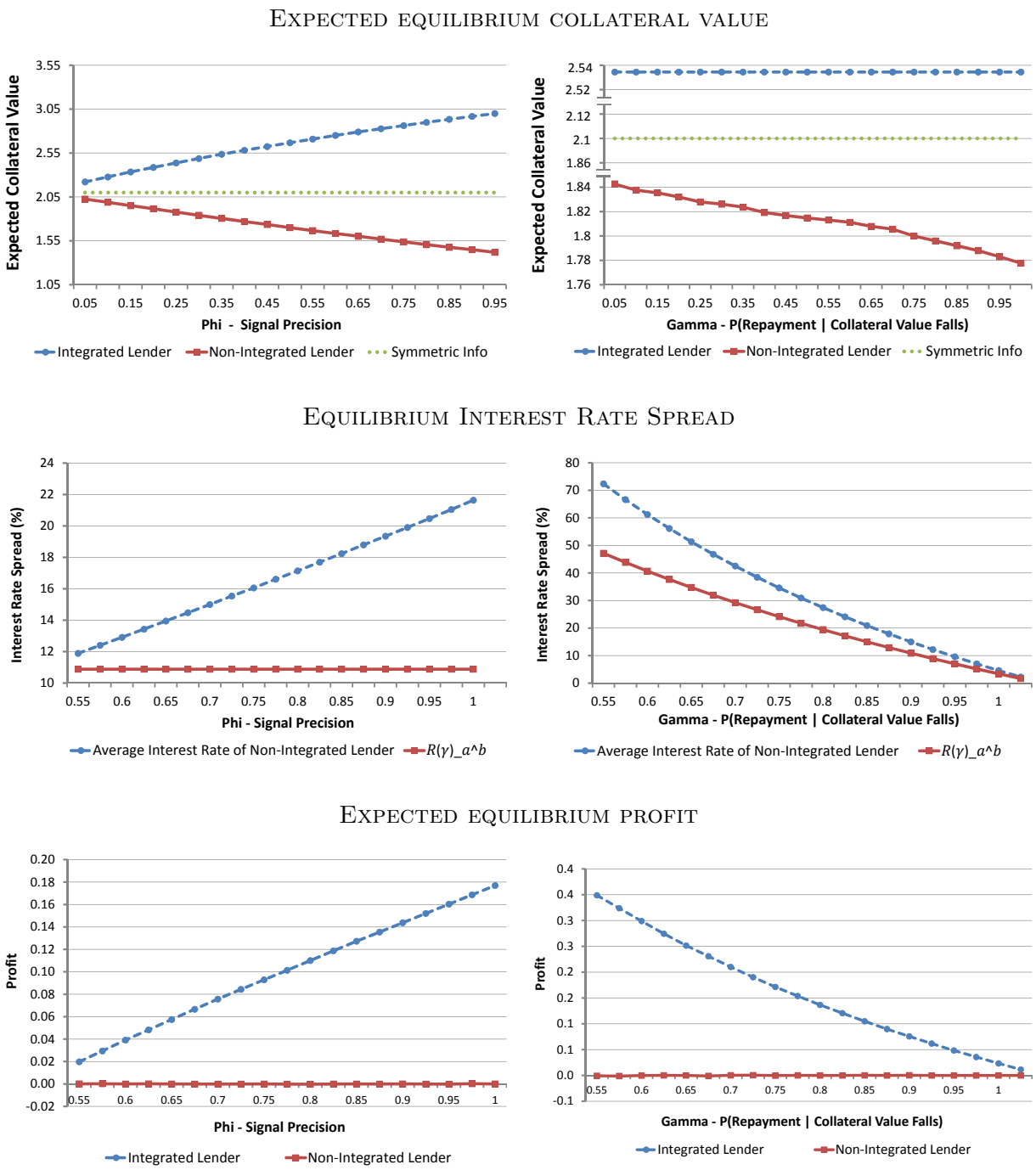


Figure 1: Equilibrium Model Outcomes

Note: The top row plots the expected period-2 price of a house in the integrated lender's equilibrium collateral portfolio (dashed line), the expected period-2 price of a house of average quality (dotted line) and the average expected period-2 price of a house in the non-integrated lender's equilibrium collateral portfolio (solid line). The middle row plots spreads of the average interest rate charged by the non-integrated lender over R_f (dashed line) and the break even rate when lending against average quality collateral, $R(\gamma)_a^b$ over R_f (solid line). The bottom row shows the expected profits of the integrated lender (dashed line) and non-integrated lender (solid line). In the left column ϕ varies along the horizontal axis. In the right column γ varies along the horizontal axis. If both lenders offer the same interest rate, I resolve the indifference in favor of the non-integrated lender. Other parameters are: $L = 0; H = 3; q = 0.7; R_f = 1.1$. I set $\phi = 0.7$ in the left panel and $\gamma = 0.7$ in the right panel.

Prediction 1(a): *When the integrated lender's information about future returns is more precise (high ϕ), the average ex-post outperformance of the homes financed by the integrated lender is larger.*

Prediction 2(a): *When the integrated lender's information about future returns is more precise (high ϕ), the return of houses financed by a non-integrated lender when competing against an integrated lender is particularly low relative to when it competes only against equally informed lenders.*

Prediction 3(a): *When the integrated lender's information about future returns is more precise (high ϕ), the increase in the interest rate charged by non-integrated lenders when competing against an integrated lender is particularly large.*

I also consider how equilibrium outcomes vary with γ, the sensitivity of the mortgage default probability with respect to changes in collateral values. The top right panel of Figure 1 shows that the return of the integrated lender's collateral is unaffected by γ. The integrated lender only lends when $\eta = h$ and the expected period-2 collateral value is $\frac{q\phi H}{q\phi+(1-q)(1-\phi)}$. The return of the non-integrated lender's equilibrium collateral portfolio declines as repayment becomes less sensitive to collateral values (higher γ). To follow the intuition for this result it is important to realize that mortgage lenders do not care about the value of the collateral per se, but only to the extent that it influences the repayment probability of the mortgage. When γ is low and the repayment probability is highly dependent on the value of the collateral, the non-integrated lender is particularly concerned about adverse selection on collateral quality. As a result it offers mortgages at higher interest rates to avoid the winner's curse, a behavior that is referred to as "bid shading" in the auction theory literature. This can be seen in the middle right panel of Figure 1, where the dashed line shows the spread over R_f of the average equilibrium interest rate charged by the non-integrated lender. Since the non-integrated lender lends against below-average quality collateral, this spread exceeds the spread of the break-even interest rate for lending against average quality collateral ($R(\gamma)_a^b - R_f$). As collateral values become less important in determining default probabilities, the spread over $R(\gamma)_a^b$ that the non-integrated lender needs to charge to break even declines. Since the integrated lender continues to exploit its superior information to the fullest degree, for larger values of γ, the non-integrated lender's equilibrium collateral is of lower quality. Put differently, the less the non-integrated lender shades its bid, the lower the quality of its equilibrium collateral portfolio. For very high values of γ collateral quality becomes a negligible driver of default and the non-integrated lender no longer shades its bid and charges R_f. The integrated lender's profit declines to zero as γ approaches one. These insights are formalized in the following three empirical predictions:

Prediction 1(b): *When mortgage repayment is more sensitive to changes in collateral values (low γ), the ex-post outperformance of houses financed by the integrated lender is smaller.*

Prediction 2(b): *When mortgage repayment is more sensitive to changes in collateral values (low γ), the return of houses financed by a non-integrated lender when competing against an integrated lender is particularly low relative to when it competes only against equally informed lenders.*

Prediction 3(b): *When mortgage repayment is more sensitive to changes in collateral values (low γ), the increase in the interest rate charged by non-integrated lenders when competing against an integrated lender is particularly large.*

Additional intuition for these predictions comes from thinking of collateralized mortgage lending in non-recourse states as involving the sale of a put option to the borrower.[13] The house represents the underlying security and, with some simplification (see Deng et al., 2000), the strike price equals the outstanding mortgage balance. The put option is "in the money" when the house value falls below the outstanding mortgage balance. The mortgage interest rate spread is the borrower's payment for the put option. Initially, the put options for mortgages with a high downpayment (low γ) are further out of the money than options for low downpayment mortgages, since prices have to fall by more before it might be valuable to exercise the option. Finance theory suggests that the option-δ (the sensitivity of option value to a change in house value) is smaller for options that are further out of the money. The presence of an integrated lender reduces non-integrated lender's expected value of the housing collateral, which makes the option more valuable. A fixed drop in the expected collateral value has a larger impact on the value of put options for low downpayment mortgages, which are closer to being in the money. Consequently, interest rates need to rise by more, as stated in Prediction 3(b). In the following sections I empirically test these predictions of the theoretical model. I argue that there is strong evidence for significant adverse selection on collateral quality in the financing of newly developed properties.

3 Possible Sources of Asymmetric Information

Before commencing the empirical analysis it is instructive to consider the possible sources of superior information of the integrated lender. This information could, in principle, relate to either the quality of the housing collateral or to characteristics of the borrowers.

3.1 Sources of Asymmetric Information about Collateral Quality

One possible source of superior information of the integrated lender about collateral quality is related to aspects of the construction quality of the house which are not (entirely) observable to buyers and non-integrated lenders at the time of purchase. The Arizona Republic (2001) describes a number of shortcuts in the construction process that are regularly taken by developers and subcontractors in Arizona. These shortcuts reduce construction costs and can generate differences in subsequent house price performance. They are most likely known to both the developer and the integrated lender, while being unobservable to buyers and non-integrated lenders at the time of purchase:

[13] Arizona is a non-recourse state, which means that a mortgage is only secured by the housing collateral and not by the borrower's income and other assets. Following a foreclosure, the lender cannot seek a deficiency judgment against the borrower for the difference between the outstanding mortgage balance and the value recovered in foreclosure.

1. The foundation is often poured without allowing the ground to settle, which saves time but can lead to subsequent shifting and cracking of the foundation.
2. Stucco is often applied too thinly, saving up to $\$1,500$ per house in construction costs. This can lead to subsequent cracking and is a common problem in Arizona.
3. Builders sometimes add excess water to the cement mix used for the foundation. This makes it easier and faster to spread, but more subject to cracking later.

In addition to cost considerations that may drive builders' decisions to take such shortcuts, The Arizona Republic (2001) also discusses the lack of skilled workers to perform delicate tasks such as "flashing" (the installation of a water-resistant protective layer around windows, doors and decks to prevent water intrusion and molding) as a factor in explaining initially unobservable differences in construction quality.[14] The integrated lender is likely to have superior information about the qualifications and skill of the work crews working simultaneously on different houses in a development. This generates another source of asymmetric information.[15]

A significant proportion of construction-related complaints in Arizona involve insufficient care taken when building on expansive soil (Phoenix New Times, 2006). Expansive soils have a high content of clays that attract and absorb large amounts of water into their surfaces. As the expansive soil absorbs water, it swells and exerts high pressure. Differential swelling and subsequent shrinkage of clay occurring under a property that is not properly designed to counteract those deformations can result in excessive foundation movements and the cracking of slabs and walls. The survey by Houston et al. (2011) suggests that most builders in Arizona were aware of the problematic soil conditions.[16] Rendon-Herrero (2011) describes countermeasures that can be taken to address expansive soil. However, there is significant evidence that these countermeasures were only selectively used in Arizona. The Phoenix New Times (2006), in its analysis of construction defects in Phoenix, concludes that: *"As bad as the results [from expansive soil] can be, experts agree that they're entirely avoidable. With proper engineering and careful attention, most soils in Maricopa County could be built on without too much trouble. The problem is that some builders aren't taking the trouble. [...] Builders frequently ignore their own [soil reports'] recommendations. The reports typically recommend stronger foundations, but some builders resist them, citing cost."*

[14] Problems caused by insufficiently qualified work crews were also discussed by Bloomberg (2011) in its nationwide analysis of construction boom flaws. Citing the founder of the International Association of Certified Home Inspectors, Bloomberg reported that "laborers became plumbers and plumbers became electricians."

[15] It is hard to empirically determine the precise channel through which the information is obtained by the integrated lender. Proponents of the knowledge-based view of the firm have argued that vertical integration facilitates knowledge flows within a firm (see, for example, Grant, 1996). Pierce (2011) finds evidence that car leasing subsidiaries have superior information about the timing of new model releases, which affects optimal lease pricing. Given the significant resources spent by lenders on property appraisers and inspectors to acquire information about the quality of a house prior to a lending decision, it seems natural to expect them to acquire additional relevant information from within their own organization. This is particularly likely for developers that co-locate regional sales teams and the integrated lender's loan officers, who sometimes work on-site (and in adjacent offices) at each subdivision (Gartenberg, 2011).

[16] In fact, problems related to construction on expansive soil are not limited to Arizona. Puppala and Cerato (2009) estimate the annual cost of damage due to expansive soils in the United States to be about $13 billion.

The Arizona Republic (2001) reports that only about 20% of all foundations poured in Phoenix in 2000 used post-tension slabs, one of the processes that can help to offset the effects of expansive soil. This often saves several thousand dollars per lot in construction costs. Insufficient care taken by builders when building on expansive soil is also the source of numerous of construction defect lawsuits in Arizona (e.g. Superior Court of Arizona, 2007). Since the integrated lender can know whether the appropriate countermeasures were taken, or whether the respective subcontractor was sufficiently skilled and experienced to conduct the more delicate procedures, one would expect that the effects of adverse selection are particularly prevalent amongst houses built on expansive soil. In section 5.5 I exploit differences in the return of houses built on expansive soil and those not built on expansive soil to provide evidence for Prediction 1(a), which suggested that amongst houses built on expansive soil, those financed by the integrated lender should outperform particularly.

Problems with construction quality are not isolated instances and are not limited to Arizona. A survey by Criterium Engineers (2003), which performs approximately 25,000 home inspections annually across the United States, found that of all new homes, 21% had problems with roof installations (which can lead to water intrusion), 15% had problems with the installation of sidings, such as stucco, 23% had problems with the installation of windows and doors, 14% had problems with the construction of the foundation and 19% had problems with soil preparation. In addition, such problems with construction quality are not a recent phenomenon. As early as 1979 the Federal Trade Commission and the builder Kaufman and Broad agreed on a consent order that alledged that: *"All housing sold by respondents [Kaufman and Broad] was not built in accordance with good construction practices in the housing industry. In some houses, foundation walls were not covered with membrane waterproofing to prevent water seepage into habitable spaces. In some houses, siding was not properly anchored, roof sheathing did not meet with roof edges, spaces between foundation walls and sill plates were not sealed to prevent the entry of air and moisture, or piping and bathroom fixtures were not properly installed. All land used by respondents for building sites was not free from severe limitations that may affect the use of such land for the construction of onsite residential housing sold by respondents. In some cases, such land was subject to frequent or continuous water saturation, slow run-off of surface water, ponding of water in various places or poor drainage that could result in frost-heave and shrink-swell* (Federal Trade Commission, 1979)."

A further information advantage of the integrated lender relates to the sales progress of other units within a development. If sales are slowing down, the resulting inventory will depress future home prices. The developer also has better information about its own plans for developing nearby parcels, which will affect the value of existing units. The empirical analysis provides evidence that the majority of the outperformance of the integrated lender's collateral portfolio is driven by information about collateral that varies at the property level rather than the development level. In Appendix C.1 I provide additional evidence for possible sources of asymmetric information about collateral quality, both in Arizona and in other parts of the United States. This evidence is sourced from construction defect lawsuits and consumer complaints websites.

3.2 Sources of Asymmetric Information about Borrower Characteristics

In addition to superior information about collateral quality, it is possible that the integrated lender also possesses superior information about borrower characteristics. For example, by guiding the borrowers through the house purchase process, the developer can obtain information about the buyer's propensity to maintain the property. This will affect the rate of depreciation and thus the return of the housing collateral. It is also conceivable that in the home sales process the integrated lender obtains superior information about aspects of the future default probability of the borrower that are unrelated to the price development of the collateral (though it is possible that the primary bank of the borrower possesses the best information about these characteristics). In the empirical analysis that follows I show that the outperformance of the collateral portfolio of the integrated lender as well as the lower incidence of foreclosures amongst the portfolio of mortgages financed by the integrated lender can be best explained by the integrated lender's superior information about initial collateral quality, not borrower characteristics.

4 Data Description

To conduct the empirical analysis, I combine three main datasets. The first dataset contains the universe of ownership-changing deeds in Arizona for the years 2000 to 2010. The property to which the deeds relate is uniquely identified via the Assessor Parcel Number (APN). The variables in this dataset include property address, contract date, transaction price, type of deed (e.g. Intra-Family Transfer Deed, Warranty Deed, Foreclosure Deed), the type of property (e.g. Apartment, Single-Family Residence, Multi-Family Dwelling), the name of the buyer and seller and a classification of the buyer and seller (e.g. Husband and Wife, Single Man, Company). It also reports the amount of the concurrent mortgage, the identity of the mortgage lender and the duration of the mortgage. For mortgages with a variable interest rate I also observe the initial interest rate.

The second dataset contains the universe of tax-assessment records for the year 2010. Properties are again identified via their APN. This dataset includes information on the owner of the property, as well as property characteristics such as construction year, lot size, building size, and the number of bedrooms, bathrooms and garage spaces. It also includes information on whether the property is a rental unit and whether it has a pool. The tax assessment records also include an estimate of the market value of the property for January 2009.

The third dataset contains information from the Home Mortgage Disclosure Act's (HMDA) Loan Application Registry, which provides details on every mortgage application in major Metropolitan Statistical Areas (MSAs). It includes information on the census tract of the house used as mortgage collateral, the lender identity, the loan amount, the property type, the loan purpose and the applicant's income, sex and race. It also records whether the mortgage was originated or denied as well as whether the mortgage was sold or securitized in the same calendar year as it was origi-

nated. I merge this dataset to the deeds data via the census tract of the house, the identity of the mortgage lender, the loan amount, loan type, borrower sex and race as well as occupancy status of the housing unit. This merging process is described in detail in Appendix B.2.

I focus on the state of Arizona, which was at the center of the recent boom-bust cycle, and which is an interesting focus of study due to data quality and availability.[17] Since most of the information is originally recorded at the county-level and field population varies widely, not all specifications can be tested on a larger geographic region.[18] In Appendix B I describe the process of cleaning and merging the data, as well as the identification of integrated lenders. The resulting dataset contains information on 98,706 single-family residences that were sold by developers between 2000 and 2007 and which I can match to assessment records and HMDA data. 84.8% of these properties are in developments with an integrated lender. For those houses that are in a developement with an integrated lender, the integrated lender has a market share of 72.8%. Summary statistics for the key variables are provided in Appendix B.4.

5 Outperformance of Integrated Lender's Collateral Portfolio

In this section I test for the presence of superior information about collateral quality by the integrated lender. In particular, I consider developments with an active integrated lender and compare the quality of the collateral against which the integrated lender lends to the quality of the collateral for observationally similar mortgages granted by non-integrated lenders. This tests empirical Predictions 1, 1(a) and 1(b) from Section 2. I analyze three different outcome variables that suggest that integrated lenders do indeed possess superior information about collateral quality.

The first approach compares the return of homes financed by an integrated lender to the return of ex-ante similar homes financed by non-integrated lenders.[19] I measure this return over

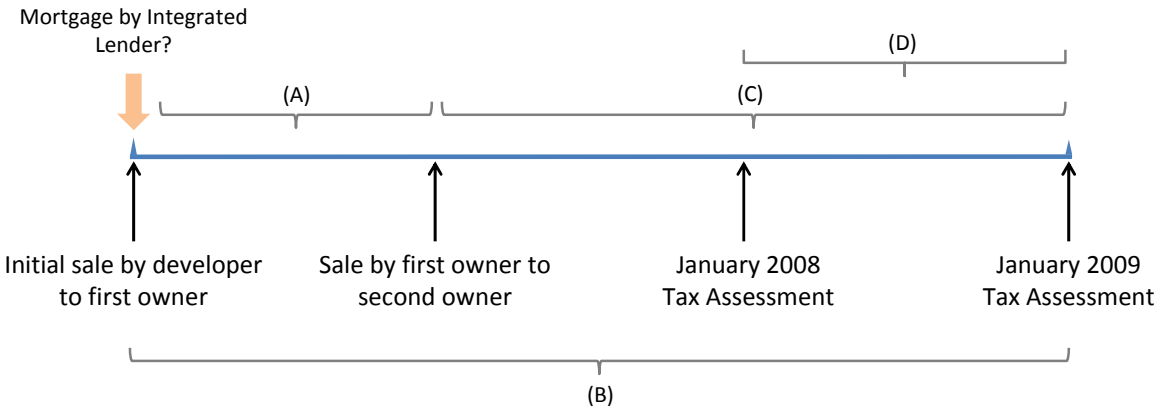
[17] There are a number of reasons to think that my results for Arizona are relevant for understanding mortgage lending in the rest of the United States. First, most of the large property developers and integrated lenders operate nationally, so to the extent that I find integrated lenders exploiting their superior information about collateral quality in Arizona, I would expect to observe similar behavior by the same actors in other states. Second, as discussed in Section 3, most of the unobservable aspects of construction quality that constitute the asymmetric information of the integrated lender (e.g. the quality of the foundations or of the roofing) are relevant in developments throughout the U.S. On the other hand, Arizona saw one of the larger boom-bust cycles in construction during the period under consideration. This means that my findings might be more representative for other states with a significant construction boom, such as California, Florida and Nevada, than for states with less of a boom, such as Texas. However, in many ways the states with the most significant boom-bust cycles were at the epicenter of the housing and mortgage crisis, and thus are of inherent interest even if it is not clear to what extent the results can be extrapolated to other states.

[18] For example, a number of non-disclosure states do not report transaction prices. These include Alaska, Idaho, Indiana, Kansas, Maine, Mississippi, Missouri, Montana, New Mexico, North Dakota, South Dakota, Texas, Utah and Wyoming. Other states, such as Georgia, do not allow me to identify sales by developers. The data from other states such as Maryland does not provide the identity of the mortgage lender. The changes following Proposition 13 in California mean that assessed property values cannot be interpreted to reflect true market values.

[19] Homes within the same development are often very similar to one another, due to developers' common practice of offering a choice from a number of model homes, the interior of which (e.g. the kitchen) is subsequently customized. In its 2004 10-K statement, KB Home, a large national homebuilder, describes a typical development: "The total number of lots in our domestic new home communities vary significantly, but typically range from 50 to 250 lots.

four different time horizons described below. Each of these time horizons allows me to address a different possible contaminating factor and jointly they provide strong evidence for the presence of asymmetric information about collateral quality by integrated lenders. For the first time horizon I focus on houses for which I observe a second armslength transaction in addition to the initial sale by the developer. For each such property I calculate the annualized return between the two sales. This corresponds to time period (A) in Figure 2. I then compare the return of properties that were financed by the integrated lender to the return of ex-ante similar properties in the same development financed by non-integrated lenders. I find that houses financed by an integrated lender outperform by about 50 basis points annually. One concern with this specification might be that the subset of houses for which a second sale is observed is not representative, perhaps because houses with lower construction quality are more likely to be sold again quickly by the initial owner. To address this concern I also consider the subset of homes for which the second sale is likely to have been prompted by an event that is plausibly exogenous to construction quality. In particular I focus on homes for which I observe a divorce or death of the initial owner prior to the second armslength transaction. The results from this specification confirm my initial findings.

Figure 2: Measures of Ex-Post Price Performance



Note: This figure shows the four different time horizons over which the relative return of the integrated lender's collateral portfolio is measured. Period (A) is analyzed in Section 5.1, period (B) in Section 5.2, period (C) in Section 5.3 and period (D) in Section 5.4.

The second time horizon over which I compare collateral return also addresses concerns about selection into observing repeat sales. It uses the estimated market value of each property for January 2009, which is provided in the tax-assessment records and is available for every property in the dataset. I calculate the annualized implied return of all newly-built single-family residences

These domestic developments typically include two to four different model home design."

from their initial sale to their January 2009 market value estimate.[20] This corresponds to period (B) in Figure 2. Using this measure I find that homes financed by the integrated lender outperform ex-ante similar homes financed by other lenders by about 40 basis points annually.

As discussed in Section 3, some of the outperformance of the the houses financed by the integrated lender could, in theory, be driven by superior information about relevant characteristics of the buyers. I next demonstrate that, in fact, the outperformance of the integrated lender's collateral portfolio is best explained by superior information about initial collateral quality. To show this, I consider the relative return of those houses *initially* financed by the integrated lender during the ownership period of their *second* owner. This corresponds to period (C) in Figure 2. Analyzing return over this period isolates the integrated lender's superior information about collateral values from any superior information that it might have about borrower characteristics. This is because there was no information, asymmetric or otherwise, about the identity of a possible second owner at the time of making the initial mortgage. The results suggest that the outperformance detected over periods (A) and (B) is primarily driven by asymmetric information about collateral values.

The specification that considers the return of the housing collateral over the ownership period of the second owner also addresses another important concern with comparing returns over periods (A) and (B). In particular, one might be worried that some of the observed outperformance of the integrated lender's collateral portfolio is driven by a bundling of the mortgage and the initial home sale by the developer. For example, if home buyers received a discount on the house when they finance through an integrated lender, the initial sales price would not reflect the true market value of the house. The return measured over period (C) does not suffer from such a possible contamination, since the first price is the sale by the first owner to the second owner. Any discounts received by the initial owner would already be capitalized in this transaction price and would thus do not contaminate this measure of return. Since the relative outperformance of those houses initially financed by the integrated lender remains over the ownership period of the second owner, this suggests that the observed outperformance over periods (A) and (B) is not driven by a bundling of the home sale and the mortgage. I also analyze a second specification that considers housing returns between two points subsequent to the initial sale by the developer. For a subset of counties I also observe a tax assessment value for January 2008. For those houses I consider the relative return of the integrated lender's collateral portfolio between the assessed values in January 2008 and January 2009, as given by period (D) in Figure 2. The result from this specification also suggests that the observed outperformance of the integrated lender's collateral portfolio is not driven by bundling the mortgage and the home sale.

Prediction 1(a) stated that the outperformance of the integrated lender's collateral portfolio should be larger amongst houses for which the integrated lender's information about construction quality is most relevant in determining future returns. As discussed in Section 3, complaints about

[20] The quality of the assessment data as a proxy for market value is discussed in Section 5.2 and Appendix B.5.

construction quality in Arizona often center around insufficient care taken by developers when building on expansive soil, which swells during rain and requires specially-strengthened foundations to prevent the cracking of foundations and walls. Whether or not the developer took measures to offset the impact of expansive soil is potentially part of the superior information of the integrated lender. I use precise geological data to show that amongst houses built on expansive soil the integrated lender's collateral portfolio outperforms by over 100 basis points annually. This provides support for Prediction 1(a). The adverse selection hypothesis also suggests that the outperformance of the integrated lender's collateral portfolio should be larger for mortgages that are more likely to be repaid if collateral values decline. I find that the outperformance over all four time horizons is particularly large amongst mortgages with a low loan-to-value ratio (high initial downpayment), for which the repayment is less sensitive to changes in collateral value. This provides evidence for Prediction 1(b).

I also compare the probability of observing a foreclosure between homes financed by the integrated lender and similar homes financed by non-integrated lenders, conditional on borrower and loan characteristics. I show that the probability of observing a foreclosure within three years of granting the initial mortgage is almost one percentage point lower for mortgages that were granted by the integrated lender. This is relative to an average sample probability of default within three years of about two percent. This difference could be due to either superior information about collateral quality or superior information about borrower characteristics. However, I also analyze the relative foreclosure probability of the second owner of houses initially financed by the integrated lender. I find that they are about two percentage points less likely to enter into foreclosure within three years, relative to a sample default probability of about 5.5 percent. This result suggests that at least some of the differences in foreclosure probability are driven by the integrated lender's superior information about collateral values.

As a third proxy for the initially unobservable quality of the housing collateral I consider the ownership duration of the initial occupants. I argue that owners that move again quickly are likely to have purchased a property that turned out to have undesirable characteristics, especially if these might not yet be fully observable to potential buyers of the property. I show that the probability of moving within two years is about 1.2 percentage points lower for owners living in a house financed by the integrated lender, relative to a sample probability of moving of about 7 percent.

5.1 Measuring Collateral Return Using Repeat Sales Transactions

In this section I focus on the subset of homes in developments with an integrated lender for which I observe a second armlength transaction subsequent to the initial sale by the developer to the first owner. For each such property I calculate the annualized return between the two sales.[21] This corresponds to period (A) in Figure 2. I then regress this return on a large number of observable

[21] I provide evidence below that considering *annualized* return is the appropriate econometric specification.

control variables and a dummy variable that captures whether the mortgage was granted by the integrated lender. The unit of observation is a house i, first sold in quarter q_1, resold in quarter q_2 and located in county c.

$$Return_i = \alpha + \kappa IntegratedLender_i + X_i\beta + \delta_{q_1,q_2} + \psi_c + \epsilon_i \tag{3}$$

In regression (3), δ_{q_1,q_2} represents a set of fixed effects for each pair of sales quarters (e.g. Q1 2000 to Q3 2008). This allows me to control for general market movements in house prices over time. I also include fixed effects for the county in which the house is located. The vector X_i controls for observable characteristics that help to explain the return of the house. In Table 1 these are added sequentially to the regression. Standard errors are clustered at the developer level.[22]

Column (1) of Table 1 estimates equation (3) without any additional controls included in X_i. The magnitude of the coefficient on $IntegratedLender_i$ suggests that houses financed by the integrated lender outperform ex-ante similar houses financed by a non-integrated lender by about 0.5 percentage points (50 basis points) annually.[23] This is significant compared to the mean of the dependent variable during the sample period, which implies an average annual house price growth of around 8.6%. It is also large compared to the long-term average growth rate of house prices between 1950 and 2000 of 1.7% reported by Gyourko et al. (2006).

One concern is that the outperformance detected in column (1) could be the result of some spurious correlation based on characteristics of the house or the owner that make it more likely that the mortgage was granted by the integrated lender and that the house increases in value. For example, it could be that certain owners take better care of the house and are more likely to borrow from the integrated lender. To address such a concern, between column (1) and column (6) I add an increasing number of control variables to the vector X_i. Controls are included by adding dummy variables for groups of values. The precise functional forms are discussed in Appendix B.4. The outperformance of the integrated lender's collateral portfolio is remarkably stable to the addition of these control variables.[24] Column (2) includes property characteristics such as

[22]Clustering standard errors at the developer level addresses possible concerns about the correlation of regression residuals across different houses built by the same developer. If this correlation was driven by unobserved developer characteristics that affected all houses of the developer in the same way, the developer fixed effects introduced in column (5) of Table 1 would pick this up and robust OLS standard errors would be consistent. However, if the residuals were only imperfectly correlated across houses built by the same developer, OLS standard errors would be biased even after controlling for developer fixed effects. For example, if there was a design defect in only one of the model houses of a developer, we would expect the residuals to be correlated amongst all houses of that type, but not necessarily across houses based off a different model. Clustering standard errors at the developer level (a total of 106 clusters) takes account of this lack of independence of observations by allowing for an arbitrary correlation of residuals of houses built by the same developer.

[23]The coefficient κ on $IntegratedLender_i$ remains very similar in all specifications in Table 1 when I control for time effects with pairs of sales month fixed effects. However, given the significant increase in fixed effects using this specification, the variance-covariance matrix with clustered standard errors is nearly singular. The coefficients remain highly significant with non-clustered standard errors.

[24]In the main body of the paper I do not discuss the coefficients on the control variables, since they do not constitute the primary focus of my analysis. In appendix C.2 I show the values for the control variables for the specification in

Table 1: Annualized Collateral Return (%) - Between Repeat Sales - Period (A)

							Forced Moves	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Integrated Lender	0.498***	0.473***	0.547***	0.513***	0.485***	0.439***	0.610***	0.447**
	(0.166)	(0.158)	(0.141)	(0.134)	(0.124)	(0.108)	(0.203)	(0.211)
Quarter-Pair Fixed Effect	✓	✓	✓	✓	✓	✓	✓	✓
County Fixed Effect	✓	✓	✓	✓	✓	✓	✓	✓
House Characteristics	·	✓	✓	✓	✓	✓	·	✓
Owner Characteristics	·	·	✓	✓	✓	✓	·	✓
Financing Characteristics	·	·	✓	✓	✓	✓	·	✓
Census Tract Demographics	·	·	·	✓	✓	✓	·	✓
Developer Fixed Effects	·	·	·	·	✓	·	·	✓
Development Fixed Effects	·	·	·	·	·	✓	·	·
R-squared	0.870	0.878	0.880	0.886	0.889	0.898	0.883	0.901
Mean Dependent Var.	8.609	8.609	8.609	8.609	8.609	8.609	6.821	6.821
N	26,174	26,174	26,174	26,174	26,174	26,174	2,793	2,793

Note: This table shows results from regression (3). The dependent variable is the annualized return of houses between two armslength transactions. I include single-family residences in Arizona first sold by the initial developer between 2000 and 2007 in developments with an integrated lender. House characteristics include real initial sales price, building size, lot size, price per square foot, number of garage spaces, average size of bedrooms and bathrooms, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

initial sales price and building size. This captures that houses in different price segments had a different return over the sample period. Column (3) controls for owner characteristics such as income, and mortgage financing characteristics such as the loan-to-value ratio, which might affect the borrower's ability and incentives to maintain the property. Column (4) includes census tract demographics, capturing the median income and high school graduation rate in the 6-digit census tract for each house. Column (5) also controls for developer fixed effects. This prevents the results being driven by a positive correlation between developer quality and the associated integrated lender's aggressiveness. The coefficient on $IntegratedLender_i$ is unaffected by the addition of these controls and fixed effects.

column (6) of Table 1. It is reassuring to see that, where applicable, these coefficients are consistent with the existing literature. For example, I find that houses with an initially low sales price had a higher return over my sample period. This is consistent with evidence presented by Landvoigt et al. (2011) for San Diego.

Similar to Altonji et al. (2005) and others I argue that the stability of κ with respect to the addition of these observable characteristics reduces the likelihood that the results are driven by selection on unobservable characteristics. In addition, to the degree that one might expect selection on unobservable borrower characteristics, many of the more plausible stories suggest that my empirical approach underestimates the true effect of asymmetric information. For example, it might be that less sophisticated agents are more likely to engage in suboptimal mortgage shopping and just accept the integrated lender's offer. If less sophisticated agents were also more likely to overpay for a house (or less likely to maintain it properly), this selection would imply that the coefficient on $IntegratedLender_i$ does not capture the full extent of asymmetric information.

Column (6) adds development fixed effects. Homes in the same development are very similar in terms of school quality, crime and local amenities. Including development fixed effects thus removes further possible biases due to unobservables that might affect housing returns and the propensity of borrowers to select an integrated lender. The coefficient capturing the outperformance of the integrated lender's collateral portfolio is essentially unchanged by this addition.[25] This result suggests that the majority of the outperformance of the integrated lender can be attributed to superior information about characteristics that vary at the property level, such as construction quality. Superior information of the integrated lender about characteristics that vary at the development level, such as current demand conditions and plans for future nearby developments, does not appear to contribute significantly to the outperformance of the integrated lender's collateral portfolio.

One might be worried that the sample of houses for which I observe a resale is not a representative subsample of all newly developed homes. In fact, I argue below that a fast resale by the initial owner is evidence for the revelation of initially unobservable and undesirable characteristics of the property.[26] To address this concern about non-random selection into observing a resale of the house, columns (7) and (8) of Table 1 restrict the sample to transaction pairs where the second sale is precipitated by a plausibly exogenous event, such as the death of the owner or a divorce of the owners (Appendix B describes how I use the deeds data to identify such events). About 10% of all transaction pairs are classified as such "forced moves," most of them through a divorce of the initial owners. When measured in the subsample of forced moves, the integrated lender's collateral portfolio outperforms by about the same amount as it does in the full sample.

[25] One might be concerned that the general time development of house prices differed significantly by geography in a way that could bias the coefficient on $IntegratedLender_i$. To show that this is not the case, rather than controlling for one set of δ_{q_1,q_2} fixed effects, I interact these fixed effects with an ever tighter set of geographic identifiers. When controlling for $\delta_{q_1,q_2} \times county$ fixed effects and including the same covariates as column (6) of Table 1, I find that $\kappa = 0.447$ (t-statistic with clustered SE = 4.22). When controlling for $\delta_{q_1,q_2} \times city$ fixed effects, $\kappa = 0.454$ (t-statistic with clustered SE = 3.34). When controlling for $\delta_{q_1,q_2} \times ZIP$ fixed effects, $\kappa = 0.404$ (t-statistic with clustered SE = 2.28). When controlling for $\delta_{q_1,q_2} \times Development$ fixed effects, $\kappa = 0.453$ (t-statistic with clustered SE = 1.26).

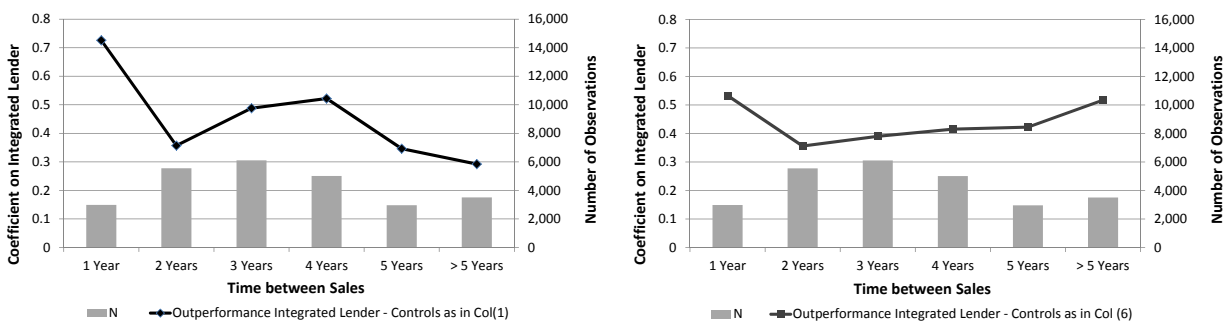
[26] This suggests that the coefficient on $IntegratedLender_i$ is biased downwards from its true value. To see this, consider three sets of houses a few years after their initial sale: those that are obviously good, those that are obviously bad and those that the owners know to be bad, but that might still appear good to potential buyers. I expect houses in the last category to be significantly overrepresented amongst the set of houses that are resold quickly and hence that transaction prices understate the true quality differences between houses.

I also analyze the time horizon over which the asymmetric information is revealed. The central specification presented in regression (3) implicitly assumes that the superior information of the integrated lender is revealed at a constant rate over time. However, it is also possible that a larger part of the asymmetric information is revealed in the first few years after the property is built. To test this, I re-run regression (3), but instead of including a simple dummy variable for $IntegratedLender_i$ I interact this dummy with the time difference between the two sales, measured in years, as given by regression (4).

$$\begin{aligned} Return_i &= \alpha + \sum_{j=1}^{6} \kappa_j \times IntegratedLender_i \times TimeBetweenSales_{i,j} \\ &\quad + X_i\beta + \delta_{q_1,q_2} + \psi_c + \epsilon_i \end{aligned} \tag{4}$$

Figure 3 presents the coefficients for two specifications that include the same controls as columns (1) and (6) of Table 1. There is no clear pattern in the development of the annualized outperformance with respect to the time difference between the two sales. The F-statistics for a Wald test of the equality of all coefficients test are 0.46 (p-value of 0.81) and 0.05 (p-value of 0.99) for the specifications corresponding to columns (1) and (6) of Table 1 respectively. Hence I cannot reject the null hypothesis that the asymmetric information that generates the outperformance of the integrated lender is revealed at a constant rate over time.[27]

Figure 3: Timing of Information Release



Note: These figures plot the coefficient on $IntegratedLender_i$ for regression (4). The line graph plots the coefficients on the left axis. The left panel corresponds to coefficients from a regression with the same control variables as column (1) of Table 1. The right panel corresponds to the coefficients from a regression with the same controls as column (6) of Table 1. The bar chart plots the number of observations in each group on the right axis.

[27]The longest time period between two sales in this analysis is 10 years. I would expect that the rate of revelation of the asymmetric information declines over time for longer time horizons. However, for the time period under consideration the data suggest that regression (3) is correctly specified with respect to the timing of the outperformance.

5.2 Measuring Collateral Return Using Assessed Home Values

The second analysis of the ex-post performance of the housing collateral considers the implied annualized return between the initial sale of a house and its estimated market value in January 2009 as recorded in the tax assessment records. In Arizona, the "full cash value" determined by the tax assessor is defined in A.R.S 42-11001(6) to be "synonymous with market value, which means the estimate of value that is derived annually by using standard appraisal methods and techniques."[28] I run regression (5) for all houses i located in county c and initially sold in month m in a development with an integrated lender.

$$Return_i = \alpha + \kappa IntegratedLender_i + X_i\beta + \delta_m + \psi_c + \epsilon_i \tag{5}$$

Table 2 presents the results from regression (5). As before, the coefficient on $IntegratedLender_i$ is not significantly affected by the addition of a large set of control variables between columns (1) and (5). Reassuringly, the magnitude of κ using this measure of return is similar to the magnitude obtained using the return between repeat sales: the houses financed by the integrated lender outperform ex-ante similar houses financed by non-integrated lenders by over 40 basis points annually. While the inclusion of developer fixed effects does not significantly affect the coefficient on $IntegratedLender_i$, the inclusion of development fixed effects reduces the observed outperformance of the houses financed by the integrated lender. This is explained by the use of the "comparables" methodology employed by assessors in arriving at the assessed market values. In this method the valuation of each house uses transaction prices of recently-sold similar homes, often from within the same development. The identity of the mortgage lenders is not one of the dimensions along which similarity is assessed. This means that the sale of a high quality home financed by the integrated lender will also impact the assessed values of lower quality homes in the same development financed by non-integrated lenders. Without development fixed effects, additional identification comes from the fact that developments with a higher percentage of mortgages financed by the integrated lender should generally outperform other developments. The fact that κ remains positive after controlling for development fixed effects does provide evidence that the assessor has some success at detecting differential collateral quality within a development.

5.3 Source of Asymmetric Information: Borrower or Collateral?

In Section 3 I discussed that the integrated lender might have superior information about the initial quality of the collateral as well as about characteristics of the borrower. Both of these sources of superior information could, in principle, explain why the return of houses financed by the integrated lender is higher than the return of houses financed by non-integrated lenders. In

[28] Appendix B.5 describes the assessment process in Arizona in more detail and analyzes the accuracy of assessments. Also see Garmaise and Moskowitz (2004) for a discussion of the usefulness and accuracy of tax assessments.

Table 2: Annualized Collateral Return (%) - Between Initial Sale and Assessment - Period (B)

	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	0.401***	0.447***	0.473***	0.407***	0.332***	0.196***
	(0.111)	(0.105)	(0.098)	(0.106)	(0.091)	(0.048)
Month of Sale Fixed Effect	✓	✓	✓	✓	✓	✓
County Fixed Effect	✓	✓	✓	✓	✓	✓
House Characteristics	·	✓	✓	✓	✓	✓
Owner Characteristics	·	·	✓	✓	✓	✓
Financing Characteristics	·	·	✓	✓	✓	✓
Census Tract Demographics	·	·	·	✓	✓	✓
Developer Fixed Effects	·	·	·	·	✓	·
Development Fixed Effects	·	·	·	·	·	✓
R-squared	0.810	0. 827	0.831	0.880	0.891	0.935
Mean Dependent Var.	-6.602	-6.602	-6.602	-6.602	-6.602	-6.602
N	83,669	83,668	83,668	83,668	83,668	83,668

Note: This table shows results from regression (5). The dependent variable is the annualized return of houses between their initial sale and their January 2009 assessed market value. I include single-family residences in Arizona first sold by the initial developer between 2000 and 2007 in developments with an integrated lender. House characteristics include real initial sales price, building size, lot size, price per square foot, number of garage spaces, average size of bedrooms and bathrooms, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

this section I provide evidence that, in fact, asymmetric information about initial collateral quality is likely to be the key driver of the outperformance of the integrated lender's collateral portfolio.

I focus on the subset of houses for which I observe at least two armslength transactions and calculate the annualized return of these properties between the second transaction price and the assessed market value in January 2009, given by period (C) in Figure 2. In other words, I compare the collateral return over the ownership period of the *second* owner of these properties. Since the identity of this owner was unknown to all lenders at the time of granting the initial mortgage, any outperformance of the integrated lender's collateral portfolio over this period can only be explained by superior information about initial collateral quality. Table 3 shows the results of a regression of this return on the identity of the initial mortgage lender. As before, I control for house characteristics, owner characteristics and financing characteristics, as well as the identity of

the developer.[29] In columns (4) and (5) I restrict the sample to those houses for which the second sale was prompted by a divorce or death of the initial owners. As in Section 5.1, this addresses possible concerns about non-random selection into observing repeat sales. The magnitude of the κ coefficients is similar to those reported in Tables 1 and 2. These results suggest that the results found in Sections 5.1 and 5.2 are driven by asymmetric information about the value of the collateral and not the borrower.

Table 3: Annualized Collateral Return (%) - Between Second Sale and Assessment - Period (C)

				Forced Moves	
	(1)	(2)	(3)	(4)	(5)
Integrated Lender	0.355***	0.287***	0.144	0.539**	0.432**
	(0.129)	(0.104)	(0.091)	(0.229)	(0.204)
Control Variables (See Note)	✓	✓	✓	✓	✓
Developer Fixed Effects	·	✓	·	·	✓
Development Fixed Effects	·	·	✓	·	·
R-squared	0.890	0.898	0.946	0.875	0.885
Mean Dependent Var.	-11.00	-11.00	-11.00	-12.54	-12.54
N	16,766	16,766	16,766	1,561	1,561

Note: This table shows results from a regression of the annualized return of houses from their second sale to the assessed value in January 2009 on control variables and a dummy variable of whether the initial mortgage was made by an integrated lender. I include single-family residences in Arizona first sold by the initial developer between 2000 and 2007 in developments with an integrated lender, and resold prior to 2008. Each specification includes fixed effects for construction quarter, resale month and county, as well as controls for characteristics of the house, owner, financing and census tract . House characteristics include real initial and resale price, building size, lot size, average size of bedrooms and bathrooms, price per square foot, number of garage spaces, whether the house has a pool and whether it is a rental unit. Owner characteristics include whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio and mortgage duration. I also control for characteristics of the initial owner and financing as in Table 1. Census tract demographics include median household income and the percentage of adults over 25 with a high-school diploma. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

5.4 Bundling of Home and Mortgage

One concern with specifications that consider the return of the housing collateral starting from the initial sale by the developer (e.g. periods (A) and (B) in Figure 2) is that price bundling of the house and the mortgage by the developer could contaminate these measures of return. For

[29] The set of owner characteristics and financing characteristics is smaller than in Sections 5.1 and 5.2. This is because a significant number of second purchasers did not use a mortgage. This makes it impossible to retrieve income information from the Home Mortgage Disclosure Act (HDMA) data. Since the integrated lender only interacted with the initial owner I am not concerned about a spurious correlation along characteristics of the second owner. In addition, the results in Tables 1 and 2 show that selection on observable owner and financing characteristics is limited. Nevertheless, in Table 3 I also control for characteristics of the initial owner and the initial financing.

example, developers might offer a discount to home buyers if they finance the purchase through the integrated lender. If the house was thus sold to customers of the integrated lender at below market value, this would also show up in the data as outperformance of the integrated lender's collateral portfolio when the house subsequently gets sold for its true market value. The results presented in Figure 3 partially address this concern: if the outperformance detected in Section 5.1 was driven by an initial price discount that gets subsequently capitalized, one would expect the *annualized* outperformance to decline with the time between the two sales. Since the annualized outperformance does not vary with the time between sales, it cannot be explained primarily by price bundling. In addition, the results in Section 5.3 show that the outperformance persists over the ownership period of the second owner. Since any initial price discounts for customers of the integrated lender would be capitalized in the sales price between the first and the second owner, it should not contaminate this measure of return.[30]

In this section I present additional results that address possible concerns about bundling. I again compare the collateral return between two points subsequent to the initial sale. For a subset of homes I observe an assessment value for January 1, 2008.[31] For house i, built in quarter q in county c, I calculate the return between the assessed market value in 2008 and the assessed market value in 2009, given by period (D) in Figure 2. I regress this return on an integrated lender dummy and a set of control variables, as shown in equation (6). Any discount on the initial sales price given to customers of the integrated lender would be included in the market value in January 2008. Thus, any outperformance of the integrated lender's collateral portfolio in the year 2008 is driven by differences in collateral quality rather than an initial bundling of mortgage and home sale.

$$Return_2008_i = \alpha + \kappa IntegratedLender_i + X_i\beta + \delta_q + \psi_c + \epsilon_i \tag{6}$$

Table 4 shows the results from this regression. The year 2008 was the year of the most dramatic decline in house prices in Arizona. The mean return of all houses in 2008 was -25%. Over this period those houses financed by the integrated lender outperformed ex-ante similar homes financed by non-integrated lenders by over 70 basis points. This is the same order of magnitude as the outperformance detected in Tables 1 and 2. As in regression (5), which also relies on assessed values, the inclusion of development fixed effects reduces the measured outperformance significantly.

[30] Another bundling strategy would be to subsidize the mortgage to be able to sell more houses. Such a behavior would suggest that integrated lenders make worse mortgages, and would bias against detecting an outperformance of the integrated lender's collateral portfolio. Barron et al. (2008) presents a model of an integrated lender that takes into account the return from making the loan as well as from selling the product, and would thus optimally set lower lending standards. Pierce (2011) presents evidence that car manufacturers use their leasing subsidiaries in such a way. During the period I consider (properties sold between 2000 and 2007), credit was available to a wide range of buyers, and thus subsidizing the financing may not have been necessary. While there is anecdotal evidence that integrated lenders did lower their lending standards at the onset of the crisis, Gartenberg (2011) argues that integrated lenders had put in place effective incentive mechanisms to limit such behavior.

[31] Properties from Maricopa, Pima and Yuma counties in Arizona have assessed values for January 2008. Properties from Pinal county, which constitute about 20% of my baseline data, do no include this additional information.

Table 4: Annualized Collateral Return (%) - Between 2008 and 2009 Assessment - Period (D)

	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	0.949***	0.786***	0.929***	0.816***	0.686***	0.350***
	(0.288)	(0.225)	(0.218)	(0.190)	(0.190)	(0.119)
Quarter of Construction Fixed Effect	✓	✓	✓	✓	✓	✓
County Fixed Effect	✓	✓	✓	✓	✓	✓
House Characteristics	·	✓	✓	✓	✓	✓
Owner Characteristics	·	·	✓	✓	✓	✓
Financing Characteristics	·	·	✓	✓	✓	✓
Census Tract Demographics	·	·	·	✓	✓	✓
Developer Fixed Effects	·	·	·	·	✓	·
Development Fixed Effects	·	·	·	·	·	✓
R-squared	0.211	0.304	0.331	0.573	0.605	0.795
Mean Dependent Variable	-27.23	-27.23	-27.23	-27.23	-27.23	-27.23
N	66,497	66,497	66,497	66,497	66,497	66,497

Note: This table shows results from regression (6). The dependent variable is the return between the January 2008 and January 2009 assessed market values. I include single-family residences in Arizona first sold by the initial developer between 2000 and 2007 in developments with an integrated lender. House characteristics include the estimated house value in January 2008, building size, lot size, price per square foot, number of garage spaces, average size of bedrooms and bathrooms, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

5.5 Outperformance Varies with Importance of Asymmetric Information

Prediction 1(a) stated that the outperformance of the integrated lender's collateral portfolio should be higher when its information about collateral quality is more important in determining housing returns. I test this prediction by exploiting exogenous differences in the type of soil on which houses are built. Section 3 explained that the return of houses built on expansive soil is particularly sensitive to whether or not appropriate measures to strengthen the foundation were taken during the construction process. I use highly detailed data on the geographic distribution of soil types from the USDA's Soil Survey database to determine which houses are built on expansive soil.[32]

[32]The data identify four hydrologic soil groups, which are characterized by their intake of water under conditions of maximum yearly wetness and the maximum swelling of expansive clays. I assign houses constructed on soil in hydrologic group D to be built on expansive soil. Group D soil typically has more than 40% clay, and a high shrink-swell potential (United States Soil Conservation Service, 1985). About 10% of observations have hydrologic soil group

Figure 4: Map of Representative Development and Soil Type

Note: These figures show a map of a representative housing development in Arizona. Each blue circle (•) and red cross (✚) represents a sale by one of two developers building in this development that appears in my dataset. The right panel also presents the soil type for each house. Houses built on the light gray, striped land are built on expansive soil while houses built on the dark green land are not built on expansive soil.

Soil expansiveness has significant geographic variation, often within developments. This can be seen in Figure 4, which shows a representative development in Arizona. In the right panel I include information on soil expansiveness. In regression (7) I include a dummy for "expansive soil" as well as its interaction with the integrated lender dummy.

$$\begin{aligned} Return_i &= \alpha + \kappa_1 \times IntegratedLender_i + \kappa_2 \times ExpansiveSoil_i + \\ &\quad \kappa_3 \times IntegratedLender_i \times ExpansiveSoil_i + X_i\beta + \delta_{q_1,q_2} + \psi_c + \epsilon_i \end{aligned} \tag{7}$$

The results from regression (7) are shown in Table 5. Columns (1) - (3) measure housing return over period (A) and correspond to Table 1. Columns (4) - (6) measure housing return over period (B) and correspond to Table 2.

D. Expanding the "expansive soil" category to include hydrologic group C (that includes between 20% and 40% clay), adds another 10% of the observations to the "expansive soil" category, and does not change the empirical results.

Table 5: Annualized Collateral Return (%) by Soil Type - Periods (A) and (B)

	Repeat Sales - Period (A)			Assessor Data - Period (B)		
	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	0.458***	0.425***	0.399***	0.341***	0.276***	0.172***
	(0.144)	(0.129)	(0.111)	(0.090)	(0.090)	(0.048)
Expansive Soil	-0.173	-0.166	-0.792**	-1.359*	-1.072*	-0.877***
	(0.528)	(0.507)	(0.345)	(0.752)	(0.643)	(0.195)
Integrated Lender × Expansive Soil	0.623**	0.668**	0.438	0.609**	0.557**	0.265**
	(0.293)	(0.277)	(0.272)	(0.268)	(0.233)	(0.101)
Control Variables (See Note)	✓	✓	✓	✓	✓	✓
Developer Fixed Effects	·	✓	·	·	✓	·
Development Fixed Effects	·	·	✓	·	·	✓
R-squared	0.886	0.889	0.898	0.881	0.891	0.935
Mean Dependent Var.	8.609	8.609	8.609	-6.602	-6.602	-6.602
N	26,174	26,174	26,174	83,668	83,668	83,668

Note: This table shows results from regression (7). $ExpansiveSoil_i$ is equal to one for houses built on hydrologic soil group D. Columns (1) - (3) show results corresponding to Table 1 and include quarter-pair fixed effects. Columns (4) - (6) show results corresponding to Table 2 and include month of sale fixed effects. Each specification controls for county fixed effects, house characteristics, owner characteristics, financing characteristics and census tract demographics. House characteristics include real initial sales price, building size, lot size, price per square foot, number of garage spaces, average size of bedrooms and bathrooms, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

The coefficient on $ExpansiveSoil_i$ is negative and sometimes statistically significant. More importantly, the positive and significant coefficient on the interaction between $IntegratedLender_i$ and $ExpansiveSoil_i$ shows that the outperformance of the integrated lender is particularly large for those properties where the integrated lender's information about construction quality is most informative about subsequent housing returns. For houses built on expansive soil, the integrated lender's collateral portfolio outperforms an ex-ante similar portfolio of houses financed by a non-integrated lender by almost one percentage point annually ($\kappa_1 + \kappa_3$). This is further evidence that a significant part of the asymmetric information is likely to relate to the initial construction quality of the housing collateral.[33] The coefficient on $IntegratedLender_i$ also remains positive. This result suggests that some of the outperformance of the integrated lender relates to information about characteristics of the collateral that also affect the return of houses not built on expansive soil (for example, information about the quality of the electric wiring or the flashing).

[33]Table 16 in Appendix C.3 shows that the increased outperformance of the integrated lender's collateral portfolio on expansive soil remains when I measure performance over periods (C) and (D) in Figure 2.

In columns (3) and (6) I also include development fixed effects. The positive coefficient on the interaction remains in these specifications, which exploit within-development variation in soil type. The lower statistical significance of the interaction in this specification can be explained by possible measurement and classification errors that arise from the fact that while soil type is inherently continuous, my measure is discrete.[34] The gradual change in soil type implies that soil in different hydrologic groups near classification boundaries will be rather similar. When including development fixed effects, the identification relies more strongly on differences in return for houses closer to soil boundaries and so the actual differences in soil expansiveness are smaller.

5.6 Outperformance Varies with Relevance of Collateral Quality

The previous sections showed that the portfolio of mortgages granted by the integrated lender is backed by higher quality collateral than the portfolio of mortgages granted by non-integrated lenders. Prediction 1(b) suggests that this outperformance should be particularly large for houses backing mortgages for which repayment is *less* dependent on collateral quality (high-γ mortgages). To test this prediction I use the initial downpayment on the mortgage to proxy for γ. When the downpayment is high (and the loan-to-value is low) the repayment probability is less sensitive to collateral quality, since house prices have to fall by more to create incentives for default. Put differently, the option-δ for the put option to default is lower, since the option is further out of the money. Hence non-integrated lenders are less concerned about the collateral quality for mortgages with a low loan-to-value (LTV) ratio and offer more aggressive financing (they shade their bids less). This allows the integrated lender to assemble a particularly attractive collateral portfolio for low LTV ratio mortgages. On the other hand, for high LTV ratio mortgages a small decline in house prices can already generate incentives for default. Non-integrated lenders thus offer less aggressive financing to avoid the winner's curse (they shade their bid more), which, in equilibrium, improves their average collateral quality. I divide borrowers into four LTV ratio groups: Less than or equal to 80%, between 80% and 90%, between 90% and 97% and above 97%.[35] I then run regresssion (8), which allows me to separately determine the outperformance of the integrated lender's collateral portfolio for each of these LTV ratio groups.

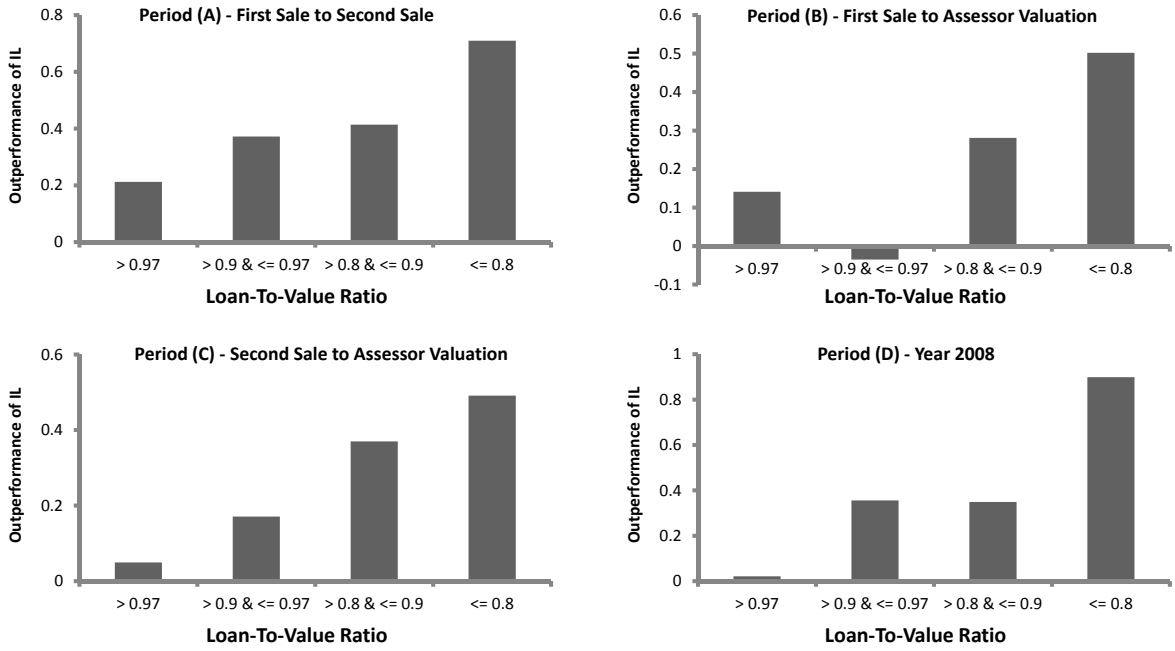
$$Return_i = \alpha + \sum_{j=1}^{4} \kappa_j \times IntegratedLender_i \times LTVDummy_{i,j} + \sum_{j=2}^{4} \omega_j \times LTVDummy_{i,j} \quad (8)$$
$$+ X_i\beta + \delta_{q_1,q_2} + \psi_c + \epsilon_i$$

[34] The data documentation states that "the locational accuracy of soil delineations on the ground varies [...]. For example, on long gently sloping landscapes the transition occurs gradually over many feet. Where landscapes change abruptly, the transition will be narrow."

[35] This cutoff was chosen since the minimum downpayment on an FHA-insured loan is 3.5%. The conclusions are not sensitive to the choice of the cutoff.

The top left panel of Figure 5 shows the resulting κ_j coefficients when I include the same control variables in X_i as in column (5) of Table 1. The houses financed by the integrated lender outperform similar houses financed by non-integrated lenders for all LTV ratios. The outperformance is particularly large for houses backing those low LTV ratio mortgages for which a decline in the collateral value does not lead to a significant increase in the default probability. In Sections 5.2 to 5.4 I also measure the outperformance of the integrated lender's collateral portfolio over three other time horizons, which allows me to address concerns about selection into resale and concerns related to a bundling of mortgage and home sale. The other three panels of Figure 5 show how the outperformance of the integrated lender's collateral portfolio varies with the LTV ratio over these time horizons. For all measures of return, the outperformance of houses financed by the integrated lender is larger for low LTV ratio mortgages. A Wald test rejects the null-hypothesis of equality of κ_1 and κ_4. The F-statistics for the test when the return is measured over periods (A) - (D) in Figure 2 are 2.62 (p-value of 0.109), 8.43 (p-value of 0.005), 11.08 (p-value of 0.001) and 16.71 (p-value of 0.000) respectively.

Figure 5: Relative Collateral Return of Integrated Lender by LTV Ratio



Note: These graphs plot the κ_j coefficients for regression (8), measuring the outperformance of houses financed by the integrated lender for different values of the LTV ratio. The top left panel uses return measured over period (A) in Figure 2 and includes the same covariates as column (5) of Table 1. The top right panel uses return measured over period (B) and includes the same covariates as column (5) of Table 2. The bottom left panel uses return measured over period (C) and includes the same covariates as column (2) of Table 3. The bottom right panel uses return measured over period (D) and includes the same covariates as column (5) of Table 4.

5.7 Collateral Quality - Foreclosure Event

A further test for adverse selection in mortgage lending is to analyze the performance of the mortgage directly. One would expect mortgages backed by low quality collateral to show higher delinquency rates than mortgages backed by high quality collateral. This effect of adverse selection on delinquency rates is of primary interest to the non-integrated lenders. Unfortunately the deeds data does not track when a mortgage becomes delinquent. However, I do observe when there is a foreclosure, since foreclosures involve a transfer of ownership to the mortgage lender. For every mortgage made between 2000 and 2007 in a development with an integrated lender I determine whether I observe a foreclosure within 3 years of the initial sale. I then run a probit regression of $Foreclosure3Years_i$ on a number of explanatory variables as well as a dummy variable for whether or not the loan was made by the integrated lender, as given in regression (9).

$$Foreclosure3Years_i = \alpha + \kappa IntegratedLender_i + X_i\beta + \delta_m + \psi_c + \epsilon_i \tag{9}$$

Table 6 shows average marginal probit coefficients from regression (9). The vector of control variables X_i includes the same characteristics of property, mortgage, owner and neighborhood as in Section 5.1. The coefficient on $IntegratedLender_i$ is consistently negative and highly significant. Conditional on observables, a mortgage made by an integrated lender is about one percentage point less likely to default than a mortgage made by a non-integrated lender. These results are highly consistent with those presented by Gartenberg (2011). This analysis provides additional evidence that integrated lenders obtain information that allows them to identify mortgages that are less likely to default. As before, the sequential addition of control variables and fixed effects between specifications (1) and (4) does not affect the magnitude of κ, the coefficient on $IntegratedLender_i$.

One concern with this analysis is that a foreclosure event also requires a strategic decision by the lender that has to choose whether to foreclose on a delinquent mortgage. Integrated lenders might be particularly reluctant to initiate a foreclosure if this depresses prices for neighboring homes (Campbell et al., 2011). Since integrated lenders usually hold many mortgages in the same development, this decline in house prices might precipitate further defaults in their portfolio. In other words, it is possible that integrated lenders internalize the externality of foreclosures. One might thus observe fewer foreclosures for integrated lenders, without their mortgages performing any better. To test whether this is the case I exploit that over my sample period a significant fraction of mortgages were sold or securitized into mortgage-backed securities subsequent to their origination. For these mortgages, the decision of whether to foreclose is usually made by an independent loan servicing company. This means that the decision of whether or not to initiate a foreclosure is outside the discretion of the lender that originated the mortgage, even if that lender remains exposed to the default (see the discussion in Section 8.4). In column (5) I restrict the sample to mortgages that were securitized by the originator within the same calendar year as they were originated, as

Table 6: Relative Foreclosure Probability of Integrated Lender Mortgages

	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	-0.011***	-0.009***	-0.009***	-0.010***	-0.008*	-0.008***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.005)	(0.001)
Month of Sale Fixed Effects	✓	✓	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓	✓	✓
House Characteristics	·	✓	✓	✓	✓	✓
Owner Characteristics	·	·	✓	✓	✓	✓
Financing Characteristics	·	·	✓	✓	✓	✓
Census Tract Demographics	·	·	✓	✓	✓	✓
Developer Fixed Effects	·	·	✓	·	✓	✓
Development Fixed Effects	·	·	·	✓	·	·
Model and Sample	Probit	Probit	Probit	Probit	Probit Securitized	LPM
Mean Dependent Var.	0.019	0.019	0.020	0.022	0.021	0.015
N	66,633	65,467	63,917	56,092	11,548	83,702

Note: This table shows average marginal effects from probit regression (9). The dependent variable is the probability of observing a foreclosure within three years of purchase. I include single-family residences in Arizona sold by the initial developer between 2000 and 2007 in developments with an integrated lender. House characteristics include real initial sales price, building size, lot size, price per square foot, number of garage spaces, average size of bedrooms and bathrooms, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Column (5) is restricted to mortgages that were securitized in the year they were originated. Column (6) estimates regression (9) as a linear probability model (LPM). Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

reported in the HMDA data.[36] The magnitude of the effect in this subsample is very similar to that in the full sample of mortgages. This result alleviates concerns that the lower default probabilities of mortgages originated by the integrated lender that I observe in columns (1) to (4) are driven by a strategic choice of the integrated lender to delay foreclosure. In column (6) I estimate regression (9) as a linear probability model (LPM) rather than as a probit model. The estimated difference in the default probability of the integrated lender's mortgage portfolio is very similar to the difference estimated using the probit model.

[36]The originator only reports what it does to the mortgage, but does not track a mortgage over time. In this sample I thus only consider mortgages that were directly securitized by the originator. Some of the mortgages that are first sold (either to affiliate companies or to other financial institutions) will also end up getting securitized.

Table 7: Relative Foreclosure Probability of Integrated Lender Mortgages - 2nd Owner

	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	-0.011**	-0.014***	-0.016***	-0.019***	-0.019***	-0.017***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Month of Sale Fixed Effects	✓	✓	✓	✓	✓	✓
Quarter of Construction Fixed Effects	✓	✓	✓	✓	✓	✓
County Fixed Effects	✓	✓	✓	✓	✓	✓
House Characteristics	·	✓	✓	✓	✓	✓
Owner Characteristics	·	·	✓	✓	✓	✓
Financing Characteristics	·	·	✓	✓	✓	✓
Census Tract Demographics	·	·	✓	✓	✓	✓
Developer Fixed Effects	·	·	·	✓	·	✓
Development Fixed Effects	·	·	·	·	✓	·
Model	Probit	Probit	Probit	Probit	Probit	LPM
Mean Dependent Var.	0.045	0.047	0.048	0.050	0.056	0.044
N	12,526	11,616	11,616	11,050	9,762	12,594

Note: This table shows average marginal effects from probit regression (9) for the second owners of the house. I include single-family residences in Arizona sold by the initial developer between 2000 and 2007 in developments with an integrated lender and where the second sale occured before 2007 and was mortgage financed. House characteristics include real initial and resale price, building size, lot size, price per square foot, number of garage spaces, average size of bedrooms and bathrooms, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. I also control for the same characteristics of the initial owner and financing as in Table 6. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Column (6) estimates regression (9) as a linear probability model (LPM). Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

I also explore to what degree the lower default probability of mortgages granted by the integrated lender is explained by the integrated lender's superior information about collateral quality or about borrower characteristics. To do this, I analyze the probability of the *second* owner of the house entering into foreclosure within 3 years of purchasing the house from the initial owner. I only include those observations with a mortgage-financed second sale, since only those might end up in default. Table 7 shows the results from the probit regression. In this sample the average probability of default within three years is around 5%, which is more than twice as large as for the previous sample that considered the default probability for the first owner. This increase in the baseline

probability of default is because more of the resales happened within three years of the large decline in house prices that began in 2007. This specification compares the default probability of two similar mortgages, neither of which was granted by the integrated lender. The mortgages differ in whether or not they are backed by housing collateral that was *initially* financed by the integrated lender. Mortgages that are backed by such collateral are almost two percentage points less likely to enter into foreclosure than mortgages that are backed by collateral that was initially financed by a non-integrated lender. This difference in default probabilities must be driven by the integrated lenders's superior information about collateral quality not borrower characteristics, since no lender could have had any information about the identity of a possible second owner. In addition, this specification provides further evidence that the lower foreclosure rate of the mortgages financed by the integrated lender is not driven by a strategic choice of the integrated lender to not foreclose. This is because second owners will not have financed through the integrated lender.

5.8 Collateral Quality - Owner Moving Decision

The third measure of relative collateral quality comes from analyzing the initial owner's moving decision. I argue that the discovery of negative aspects of housing quality subsequent to the initial purchase increases the probability of a quick voluntary move by the first owner, especially if that aspect might not yet be fully observable to possible buyers. If some of these negative aspects were known to the integrated lender but unobservable to non-integrated lenders at the time of the initial sale, one would expect that owners who took their mortgage from the integrated lender would be less likely to move again quickly. To test this prediction I consider all houses that were sold between 2000 and 2007 in a development with an integrated lender and construct a dummy variable that is equal to one when the house is resold voluntarily within two years of the initial sale.[37] I then run a probit regression of this variable on a large number of controls and an indicator variable to capture whether the house was initially financed by the integrated lender, as given by equation (10).

$$MoveTwoYears_i = \alpha + \kappa IntegratedLender_i + X_i\beta + \delta_m + \psi_c + \epsilon_i \tag{10}$$

The vector of control variables X_i includes the same characteristics of property, mortgage, owner and neighborhood as in Section 5.1. The results are shown in Table 8. Conditional on observables, agents that purchased a house financed by an integrated lender have a probability of moving within 2 years that is over one percentage point lower than that of agents whose mortgage was financed by a non-integrated lender. This effect is sizable when compared to the average probability of moving within two years of 6.9%. As before, the magnitude of the coefficient is largely unaffected by the sequential addition of control variables between column (1) and column (6). This specification provides additional evidence for the presence of superior information of the integrated lender.

[37] The empirical results are very similar when I use a period of one year or three years instead of two years.

Table 8: Relative Moving Probability for Customers of Integrated Lender

	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	-0.015***	-0.013***	-0.012***	-0.012***	-0.012***	-0.012***
	(0.002)	(0.002)	(0.000)	(0.002)	(0.002)	(0.002)
Month of Sale Fixed Effect	✓	✓	✓	✓	✓	✓
County Fixed Effect	✓	✓	✓	✓	✓	✓
House Characteristics	·	✓	✓	✓	✓	✓
Owner Characteristics	·	·	✓	✓	✓	✓
Financing Characteristics	·	·	✓	✓	✓	✓
Census Tract Demographics	·	·	·	✓	✓	✓
Developer Fixed Effects	·	·	·	·	✓	·
Development Fixed Effects	·	·	·	·	·	✓
Mean Dependent Var.	0.069	0.069	0.069	0.069	0.069	0.069
N	83,650	83,439	83,439	83,439	83,183	83,439

Note: This table shows average marginal effects from probit regression (10). I include single-family residences sold by the initial developer in Arizona between 2000 and 2007 in developments with an integrated lender. House characteristics include real initial sales price, building size, lot size, price per square foot, number of garage spaces, average size of bedrooms and bathrooms, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

5.9 Robustness Check - Control for Interest Rates

In the previous sections I showed that houses financed by the integrated lender outperformed observationally similar houses in the same development financed by non-integrated lenders. However, I have not so far conditioned on the pricing of the mortgage. This might be a concern if there were fundamental differences in the risk preferences or strategies of integrated and non-integrated lenders. Such differences might explain some of the observed outperformance even if information about collateral quality was completely symmetric. For example, it could be that integrated lenders are more risk-averse and choose not to lend against low-quality collateral. If non-integrated lenders were more willing to lend against low-quality collateral, but at higher interest rates, we might observe an outperformance of the integrated lender's collateral portfolio, even if collateral quality was perfectly observed by all competing lenders. Ideally, I would thus like to control for the pricing of mortgages when analyzing the difference in the return of the two collateral portfolios.

Unfortunately, interest rates are not observed for all mortgages. In Arizona, mortgage information is reported in a "Deed of Trust," which is recorded alongside the ownership-changing deed. This Deed of Trust includes information on the borrower, beneficiary, loan amount and maturity of the loan. It can also include mortgage riders, which are added when (i) the mortgage is an adjustable rate mortgage and (ii) when the mortgage is a hybrid-ARM.[38] For all mortgages with a rider, I observe the initial interest rate charged on the mortgage.

In this section I present robustness checks for the subset of mortgages for which I observe the initial interest rate.[39] I show that including the pricing of the mortgage as an additional control variable reinforces my conclusions about the relative outperformance of the integrated lender's collateral portfolio. I first construct a "mortgage spread" variable that is equal to the spread of the initial mortgage rate over the average relevant (i.e. adjustable or hybrid-adjustable) interest rate in that month. I then include this spread as an additional covariate in the regressions that produce the key results in Tables 1, 2, 3, 4 and 6. The results are shown in Table 9.

Table 9: Robustness Check - Control For Interest Rate

	Period (A)	Period (B)	Period (C)	Period (D)	Foreclosure
	(1)	(2)	(3)	(4)	(5)
Integrated Lender	0.399*	0.356***	0.521***	0.617***	-0.006**
	(0.233)	(0.116)	(0.164)	(0.187)	(0.003)
Mortgage Spread	-0.520***	-0.175***	-0.024	-0.240**	0.011***
	(0.119)	(0.030)	(0.063)	(0.079)	(0.002)
Controls	Col (5)	Col (5)	Col (2)	Col(5)	Col (3)
	Table 1	Table 2	Table 3	Table 4	Table 6
R-squared	0.893	0.880	0.861	0.580	
Mean Dependent Var.	6.700	-8.426	-14.49	-27.43	0.037
N	6,999	23,355	3,296	18,174	16,249

Note: This table shows robustness checks, controlling for the initial interest rate for the subset of mortgages for which this information is available. The respective tables and the appropriate control variables are indicated. The dependent variables in columns (1) to (4) are the annualized collateral return over periods (A) - (D) respectively. The dependent variable in column (5) is the probability of observing a foreclosure within three years of the initial sale. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Houses financed with a higher interest rate mortgage have a lower return. This result suggests that there are observable aspects of collateral quality that lenders take into account when pricing mortgages. Columns (1) - (4) show that after the inclusion of the interest rate spread as an additional control variable, those houses financed by the integrated lender continue to outperform

[38] Hybrid-ARM mortgages (e.g. a 5/1 ARM is a mortgage with a fixed interest rates for the first five years and then an annually adjusted rate after that) were popular during the recent housing boom. For an example of an interest rate rider, see https://www.efanniemae.com/sf/formsdocs/documents/ridersaddenda/pdf/3187.pdf.

[39] The interest rate captures the most salient aspect of mortgage pricing. I do not observe other aspects of mortgage pricing such as the closing costs, "lock-in periods" or prepayment penalties.

by about 40 or 50 basis points annually. Column (5) shows that the probability of foreclosure is also larger for higher interest rate mortgages. The mortgages in the integrated lender's portfolio continue to have a significantly lower foreclosure probability than those originated by non-integrated lenders. These results provide evidence that the findings from the previous sections are not driven by differential risk-preferences of integrated lender and non-integrated lenders.[40]

5.10 Conclusion on Outperformance of Integrated Lender Portfolio

In this section I showed that houses financed by integrated lenders outperform ex-ante similar houses financed by competing non-integrated lenders by about 50 basis points annually. This provides evidence for Prediction 1. By also considering the housing return during the ownership period of the second owner, I ensure that the observed outperformance is not driven by an initial bundling of the mortgage and home sale. This specification also suggests that the outperformance can be best explained by the integrated lender's superior information about collateral quality, not borrower characteristics. I also analyze how the outperformance of the integrated lender's collateral portfolio varies with exogenous soil type and the initial downpayment. The results provide strong support for Predictions 1(a) and 1(b). Mortgages granted by the integrated lender are also significantly less likely to end up in foreclosure than ex-ante similar mortgages granted by non-integrated lenders. This is true both during the ownership period of the first owner and of the second owner. In addition, borrowers of the integrated lender are less likely to move again than otherwise similar households borrowing from non-integrated lenders. I interpret these results as further evidence that the integrated lender possesses and exploits superior information about collateral quality.

6 Adverse Selection in the Presence of Integrated Lender

In the previous section I considered developments with an active integrated lender and showed that this integrated lender lends against higher quality collateral than non-integrated competitors. In this section I present further evidence of adverse selection on collateral quality in mortgage lending. I compare the return of houses financed by non-integrated lenders when these lenders compete against an integrated lender to the return of ex-ante similar houses financed by non-integrated lenders in developments without an integrated lender. In the latter case, each lender should lend against average quality collateral. On the other hand, non-integrated lenders that compete against an integrated lender will lend against below average quality collateral. Hence Prediction 2 states that houses financed by non-integrated lenders should underperform in developments with an active integrated lender. To test this, I run regression (11) for all newly developed houses financed by a non-integrated lender between 2000 and 2007.

[40] I also run these regressions by (i) including the actual interest rate charged, rather than a spread and (ii) including the spread over a different base rate (the Federal Funds rate for variable rate mortgages and the average national fixed-rate mortgage rate provided by Freddie Mac's Primary Mortgage Market Survey for hybrid-ARMs). The conclusions using either of these two other measures of loan pricing are very similar to the ones presented above.

$$Return_i = \alpha + \kappa HasIntegratedLender_i + X_i\beta + \delta_{q_1,q_2} + \psi_c + \epsilon_i \tag{11}$$

The key explanatory variable, $HasIntegratedLender_i$, captures whether the non-integrated lender competes against an integrated lender. It is set to one when an integrated lender makes loans in the same development and year. The results from this regression are shown in columns (1) and (2) of Table 10. As before, standard errors are clustered at the developer level. Column (1) shows that amongst ex-ante similar homes financed by non-integrated lenders, those financed in the presence of an integrated lender underperform those financed in the absence of an integrated lender by over 60 basis points annually.[41] Column (2) adds developer fixed effects. In this specification the identification comes from analyzing the return of houses built by developers that sometimes, but not always, cooperate with an integrated lender. Adding these fixed effects strengthens the conclusion about the presence of adverse selection on collateral quality.[42] Columns (4) and (5) consider the outperformance over period (C) in Figure 2 and analyze the return over the ownership period of the second owner. Again, houses financed by a non-integrated lender underperform significantly when the non-integrated lender competes against an integrated lender. Columns (3) and (6) include an interaction between $HasIntegratedLender_i$ and $ExpansiveSoil_i$. Prediction 2(a) stated that the underperformance of the collateral of mortgages granted when competing against an integrated lender should be bigger for houses built on expansive soil, which makes return particularly sensitive to initial construction quality. The results provide some evidence for this hypothesis, though the coefficient on the interaction in column (3) is not statistically significant.

Prediction 2(b) stated that non-integrated lenders should end up with particularly low-quality collateral when competing against integrated lenders for mortgages for which repayment is less sensitive to movements in collateral quality (e.g. mortgages with a high downpayment). For these mortgages the price of the collateral has to fall by more before there are incentives for default. Consequently, non-integrated lenders do not increase their interest rate offers as much in order to avoid the winner's curse. To test this prediction I run regression (12).

$$\begin{aligned} Return_i &= \alpha + \sum_{j=1}^{4} \kappa_j \times HasIntegratedLender_i \times LTVDummy_{i,j} \\ &\quad + \sum_{j=2}^{4} \omega_j \times LTVDummy_{i,j} + X_i\beta + \delta_{q_1,q_2} + \psi_c + \epsilon_i \end{aligned} \tag{12}$$

[41] The coefficient on $HasIntegratedLender_i$ does not vary significantly when I add the control variables sequentially, as, for example, in Table 1 (results available from author). This result suggests that none of the observable differences between developments with and without an integrated lender are correlated with $HasIntegratedLender_i$ in a way that might make the observed results spurious.

[42] I do not include development fixed effects, since these are nearly collinear with $HasIntegratedLender_i$, which varies at the development-year level. In other words, most developments either do or do not have an integrated lender in all years of operation.

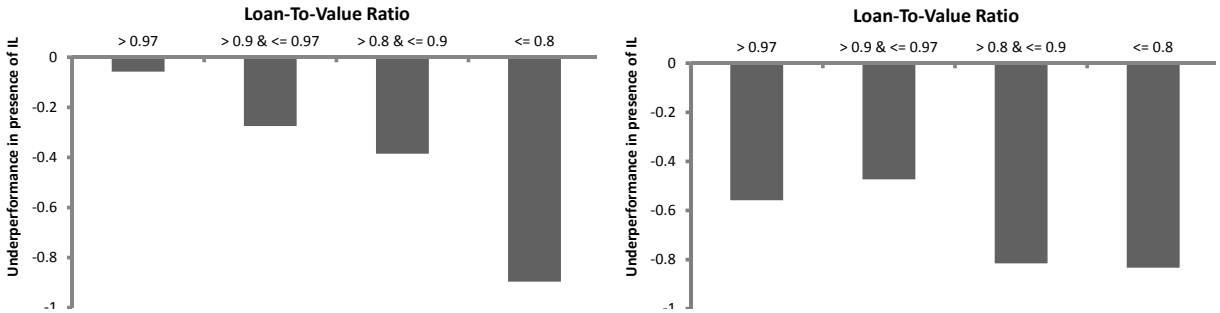
Table 10: Annualized Return (%) of Collateral of Non-Integrated Lenders

	Return Period (A) - First Owner			Return Period (C) - Second Owner		
	(1)	(2)	(3)	(4)	(5)	(6)
Has Integrated Lender	-0.616**	-0.744*	-0.589*	-0.639*	-1.065**	-0.300
	(0.300)	(0.419)	(0.301)	(0.330)	(0.463)	(0.256)
Has Integrated Lender × Expansive Soil			-0.214			-2.546***
			(0.773)			(0.932)
Expansive Soil			0.009			-0.115
			(0.604)			(0.575)
Control Variables (See Note)	✓	✓	✓	✓	✓	✓
Developer Fixed Effects	·	✓	·	·	✓	·
R-squared	0.869	0.877	0.869	0.890	0.904	0.894
Mean Dependent Var.	9.233	9.233	9.233	-10.02	-10.02	-10.02
N	12,483	12,483	12,483	7,957	7,957	7,957

Note: This table shows results from regression (11). The dependent variable in columns (1) to (3) is the annualized return over period (A) in Figure 2 and the regressions include sales quarter-pair fixed effects. The dependent variable in columns (4) to (6) is the annualized return over period (C) and the regressions include quarter of construction and month of resale fixed effects. I include single-family residences sold by a developer in Arizona between 2000 and 2007 that were financed by non-integrated lenders. Each specification controls for county fixed effects, house characteristics, owner characteristics, financing characteristics and census tract demographics. House characteristics include real initial sales price, building size, lot size, average size of bedrooms and bathrooms, price per square foot, number of garage spaces, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Figure 6 plots the κ_j coefficients from regression (12). The left panel shows the coefficients from a regression similar to that of column (1) of Table 10. The results suggest that the relative underperformance of non-integrated lenders when competing against an integrated lender is indeed larger for mortgages with a lower LTV ratio. For mortgages with a high LTV ratio for which even a small decline in prices could precipitate a default, the non-integrated lender adjusts its behavior significantly and avoids originating too many mortgages backed by low-quality collateral. For those mortgages the relative underperformance of the housing collateral financed when competing against an integrated lender is smaller. The right panel shows the coefficients from a regression similar to that of column (4) of Table 10, measuring housing return over the ownership period of the second owner. On this measure the differences in the relative performance by loan-to-value ratio are not statistically significant. The F-statistics for a Wald test of the equality of κ_1 and κ_4 are equal to 3.57 (p-value of 0.058) for the left panel and 0.46 (p-value of 0.498) for the right panel.

Figure 6: Relative Collateral Return in Presence of Integrated Lender by LTV Ratio



Note: These graphs plot the point estimates of the κ_j coefficients for regression (12), measuring the relative performance of houses financed by a non-integrated lender when it is competing against an integrated lender relative to when there is no integrated lender. The left panel uses return measured over period (A) in Figure 2 and includes the same covariates as column (1) of Table 10. The right panel uses return measured over period (C) in Figure 2 and includes the same covariates as column (4) of Table 10.

7 Interest Rate Response to Integrated Lender

Sections 5 and 6 analyzed the effects of adverse selection on the collateral quality of integrated and non-integrated lenders. I showed that in developments with an integrated lender the integrated lender lends against collateral that outperforms ex-ante similar collateral of non-integrated lenders by about 50 basis points annually. I also found that the collateral return of mortgages granted by non-integrated lenders is lower when they compete against an integrated lender. In this section I test the model's predictions for the equilibrium interest rates charged by non-integrated lenders. Prediction 3 stated that this interest rate should be higher when the non-integrated lender competes against an integrated lender and thus lends against below average quality collateral. To test this prediction I analyze the interest rates charged by non-integrated lenders by running regression (13).

$$InterestRate_i = \alpha + \kappa HasIntegratedLender_i + X_i\beta + \delta_{m,f} + \tau_l + \psi_c + \epsilon_i \tag{13}$$

The dependent variable is the mortgage interest rate. I include month by rate-type (adjustable or hybrid-adjustable) fixed effects, $\delta_{m,f}$, to capture the general interest rate environment at the time of making the mortgage. I also include a set of lender fixed effects τ_l. These are important because lenders with different funding sources and strategies might be more or less aggressive in their interest rate offers. The regression thus compares the lending behavior of the same lender making similar mortgages to purchase properties in two developments: one in which the developer cooperates with an integrated lender and one in which it does not. In a first specification, standard errors are clustered at the lender level. This allows for an arbitrary correlation between the residuals of mortgages granted by the same lender.[43] There might also be a concern that mortgages granted

[43]The lender fixed effects τ_l control for unobservable lender characteristics that impact the interest rate of all mortgages by the lender in exactly the same way. If such factors were the only source of correlation between residuals

around the same time should not be considered as independent observations since they might be affected by correlated unobserved shocks in a way that is not picked up by the $\delta_{f,m}$ fixed effects. To address this concern, I also report a second set of standard errors that are clustered at both the month level and the lender level. This uses a method proposed by Cameron et al. (2011) and described and implemented by Petersen (2009). This two-level clustering assumes that observations by different lenders in different months are independent, but allows an arbitrary correlation between all mortgages made in the same month as well as between all mortgages by the same lender.[44]

The results, which are presented in Table 11, suggest that lenders charge about 10 basis points higher interest rates when they compete against a better-informed integrated lender. These results provide a measure of the cost of the asymmetric information to borrowers in terms of higher interest rates.[45] Between columns (1) and (6) I sequentially add control variables. The magnitude of the coefficient on $HasIntegratedLender_i$ changes little, which suggests that the results are not driven by a different composition of houses or borrowers in developments with an integrated lender relative to developments without an integrated lender. In column (6), when I also include developer fixed effects, the coefficient on $HasIntegratedLender_i$ is identified by considering houses built by developers that sometimes cooperate with an integrated lender, but do not always do so. In column (7) I also add the interaction between $HasIntegratedLender_i$ and $ExpansiveSoil_i$. This tests Prediction 2(a) from Section 2 which stated that the interest rate premium for competing against an integrated lender should be particularly large for those houses where the integrated lender's information about construction quality has a significant impact on future price development. The results in column (7) provide evidence for this prediction. The interest rate increase in response to the presence of the integrated lender is more than twice as large for houses built on expansive soil.

Prediction 2(b) stated that the interest rate increase in response to competing against an integrated lender should be larger for mortgages for which the adverse selection on collateral quality is more costly. In particular, non-integrated lenders should raise interested rates more for high LTV ratio mortgages, for which a small decline in collateral value precipitates a significant increase in

of the same lender, robust OLS standard errors would be consistent. However, some elements of lender strategy could affect different mortgages differentially. For example, consider a lender that decides to focus on low-income households by offering them lower interest rates. This strategy will influence the residuals of all mortgages to low-income households by that lender, but not the residuals of other mortgages by that lender. In such a situation robust OLS standard errors will be biased. Standard errors that are clustered at the lender level allow for an arbitrary correlation in residuals of mortgages by the same lender and will be consistent if the number of clusters is large enough (in this case there are 338 clusters).

[44] It is not obvious whether it is preferable to cluster at the month level, the quarter level or the year level to allow for a possible correlation of residuals of mortgages made around the same time. As a robustness check to the results in Table 11 I also chose the quarter level and the year level as the second cluster dimension in addition to the lender level. This had little effect on the statistical significance of the coefficient on $HasIntegratedLender_i$ (results available from the author). However, in these specifications one should be concerned about the small number of clusters in the time dimension (7 clusters when choosing the year level, 28 clusters when choosing the quarter level), since the asymptotic consistency of clustered standard errors relies on the number of clusters going to infinity (Petersen, 2009).

[45] While the higher interest rates generate informational rents for the integrated lender, it is possible that the profits of the integrated lender are competed away in the construction sector. Section 8.3 discusses the possible welfare effects of the presence of the integrated lender.

Table 11: Impact of Integrated Lender on Interest Rates of Non-Integrated Lender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Has Integrated Lender	0.120**	0.116**	0.100**	0.099**	0.091**	0.093**	0.081*
	(0.054)	(0.053)	(0.047)	(0.046)	(0.045)	(0.044)	(0.043)
	[0.054]	[0.053]	[0.046]	[0.046]	[0.044]	[0.042]	[0.043]
Has Integrated Lender × Expansive Soil							0.149**
							(0.066)
							[0.066]
Expansive Soil							-0.093
							(0.066)
							[0.067]
Fixed Effects (See Note)	✓	✓	✓	✓	✓	✓	✓
Financing Characteristics	·	✓	✓	✓	✓	✓	✓
House Characteristics	·	·	✓	✓	✓	✓	✓
Owner Characteristics	·	·	·	✓	✓	✓	✓
Census Tract Demographics	·	·	·	·	✓	✓	✓
Developer Fixed Effects	·	·	·	·	·	✓	·
R-squared	0.551	0.580	0.586	0.587	0.588	0.593	0.588
Mean Dependent Var.	6.629	6.629	6.629	6.629	6.629	6.629	6.629
N	15,342	15,342	15,339	15,339	15,339	15,339	15,339

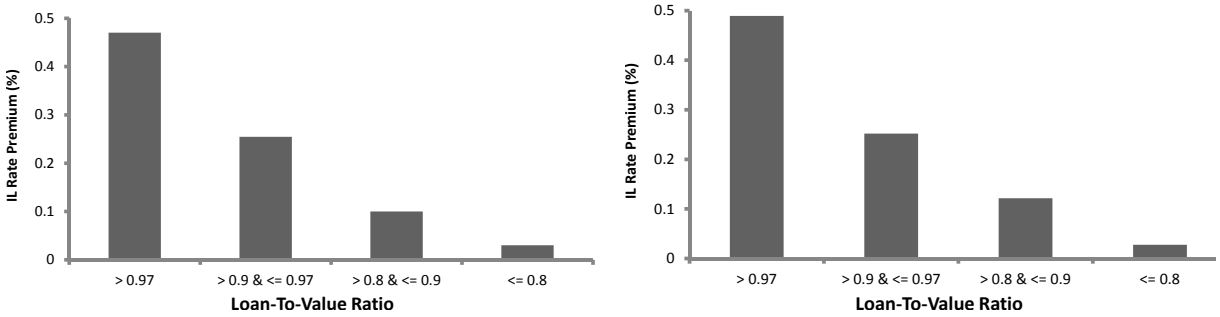
Note: This table shows results from regression (13). The dependent variable is the mortgage interest rate. I include single-family residences sold by a developer in Arizona between 2000 and 2007 that were financed by non-integrated lenders. Each specification includes month × rate-type (adjustable or hybrid-adjustable), county and lender fixed effects. House characteristics include real initial sales price, building size, lot size, average size of bedrooms and bathrooms, price per square foot, number of garage spaces, whether the house has a pool and whether it is a rental unit. Owner characteristics include real income, whether the property was purchased by an individual or a couple and whether the owners are Asian or Latino. Financing characteristics include mortgage type, loan-to-value ratio, loan-to-income ratio and mortgage duration. Census tract demographics include median household income and the percentage of adults over 25 with a high school diploma. Standard errors in parentheses are clustered at the lender level. Standard errors in square brackets are clustered at both the lender level and the month level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

default risk. To test whether this is the case, I run regression (14).

$$\begin{aligned} InterestRate_i &= \alpha + \sum_{j=1}^{4} \kappa_j \times HasIntegratedLender_i \times LTVDummy_{i,j} \\ &\quad + \sum_{j=2}^{4} \omega_j \times LTVDummy_{i,j} + X_i\beta + \delta_{m,f} + \tau_l + \psi_c + \epsilon_i \end{aligned} \tag{14}$$

Figure 7 plots the κ_j coefficients from these regressions. The regressions for the left and right panels includes the same control variables as columns (5) and (6) of Table 11 respectively. The results support the model predictions. For high LTV ratio mortgages the interest rate premium charged by the non-integrated lender when competing against an integrated lender is the largest, at almost half a percentage point annually. For these mortgages the option-δ is bigger since they are closer to being in the money. The F-statistics for a Wald test of the equality of κ_1 and κ_4 are equal to 19.6 (p-value of 0.00) for the left panel and 21.0 (p-value of 0.00) for the right panel.

Figure 7: Impact of Integrated Lender on Interest Rates of Non-Integrated Lender by LTV Ratio



Note: These graphs plot the point estimates of the κ_j coefficients for regression (14), measuring the interest rate increase of a non-integrated lender when competing against an integrated lender for different values of the LTV ratio. The left panel includes the same covariates as column (5) of Table 11. The right panel includes the same covariates as column (6) of Table 11.

8 Discussion of Results

8.1 Interpretation of Non-Integrated Lender Behavior

One question is whether the results in Sections 6 and 7 imply that non-integrated lenders are explicitly aware of the adverse selection they face when competing against an integrated lender. Under the Nash equilibrium concept employed in Section 2, the non-integrated lender accounts for the adverse selection by conditioning its belief about the true collateral quality on whether its interest rate offer is lower than that of the integrated lender.[46] However, many studies show that economic agents often do not account for selection in this way. For example, Greene (1998) argues that many credit scoring models do not adjust for the fact that the sample of individuals who received credit in the past is a selected sample of all applicants. This then raises the question of whether we might observe the pricing behavior of lenders and the collateral performance that I detect when lenders do not explicitly understand the adverse selection they face. A lender that fails to take account of the winners' curse might be naively modeled as using the unconditional

[46]This implies that the non-integrated lender understands that by adjusting its own interest rate offer it affects the quality of the collateral of the set of mortgages that it makes in equilibrium. In other words, lenders know the precise schedule of interest rate offers and associated expected collateral qualities.

expectation of collateral quality to make lending decisions. However, such behavior would generate a loss in expectation and is unlikely to persist in equilibrium. If lenders learn from previous lending they adjust their behavior based on their updated valuation of collateral quality. In equilibrium, a lender that updates its beliefs in such a way ends up lending against collateral that is exactly the quality that it expects. This does not necessarily mean, however, that the lender is aware that adjusting its interest rate offers would attract a different set of equilibrium collateral quality.[47]

The actual pricing of mortgages usually relies on statistical models that use past data to compute expected future default probabilities and adjust interest rates accordingly.[48] For a review of such pricing approaches, see Deng et al. (2000). This statistical pricing is very similar to the process of updating beliefs about true collateral values that is described above. When mortgages on houses built by a specific developer default more regularly, this would lead the non-integrated lender to charge higher interest rates for mortgages to purchase homes built by this developer. From the perspective of the lender, this could be because the quality of homes built by the developer is generally low or because of the adverse selection described in this paper. To price these mortgages correctly (i.e. so that default realizations do not contradict default expectations), the non-integrated lender would not need to differentiate between these two explanations. Hence we might observe the above results on interest rate adjustments and differences in the collateral performance without non-integrated lenders being necessarily aware of the channel through which the adverse selection generates the decline in collateral quality that they price in.

8.2 Magnitude of Effects

Sections 5 and 6 discuss the effect of adverse selection on the average quality of houses financed by integrated and non-integrated lenders. I now consider whether the effect of this difference in collateral quality on the value of the mortgage is consistent with the changes in the non-integrated lenders' pricing behavior discussed in Section 7. To address this question, it is again instructive to

[47] Esponda (2008) introduces the relevant equilibrium concept. His "behavioral equilibrium" involves lenders that adjust their behavior when experience contradicts their beliefs. Expected collateral quality is consistent with the actual equilibrium quality of collateral against which the non-integrated lender lends at the interest rates it offers in equilibrium. This is different from the unconditional expected quality, q. A lender that is not aware of the underlying selection, though, does not realize that making a lower interest rate offer would increase average collateral quality. Esponda (2008) shows that if lenders behave in such a "naive" way and fail to take account of the selection explicitly, this exacerbates the negative effects of asymmetric information on market thickness and market functioning.

[48] Since these pricing models are usually propriatory, it is difficult to directly ascertain how lenders adjust their pricing to the presence of an integrated lender. However, some information can be gathered by looking at the "rate sheets" published by some lenders, which provide basic pricing guidance, though the final interest rate will depend on more borrower and property-specific characteristics. A rate sheet from March 8, 2011 for mortgages by *ING Mortgage* shows that maximum allowable LTV ratios are adjusted downwards by 15 percentage points for mortgages financing the purchase of newly constructed (never occupied) homes. A rate sheet from March 9, 2011 by *Luxury Mortgage* shows that mortgages used to purchase homes in new developments have a 12.5 basis points higher interest rate compared to similar mortgages used to purchase existing homes. This provides some evidence that lenders do adjust their mortgages, at least when lending in new developments compared to lending to purchase existing homes. Loan officers usually have some additional discretion in adjusting mortgages rates from rate sheets, by charging overages or underages. Black et al. (2003) describe this process in detail.

think of a mortgage from the perspective of the lender as the combination of a making a risk-free loan and selling a put option on the house (abstracting from complications due to the borrower's prepayment option). Differences in collateral quality impact the value of this put option. When the collateral appreciates by 50 basis points less annually the put option is in the money in more states of the world and thus worth more. Consequently, the lender that sells the put option charges a higher interest rate. Calculating this change in the option value requires inputs such as expected house price appreciation and volatility. Banks regularly conduct similar calculations when deciding how to adjust interest rates for borrowers with different downpayments (a lower initial downpayment also moves the put option closer to being in the money). The average life time of a 30-year mortgage is about 7-8 years, as borrowers move or prepay their mortgage. Section 6 showed that relative to lending in a development without an integrated lender, the presence of an integrated lender causes a non-integrated lender to lend against collateral that underperforms by about 60 basis points annually. After the average life time of a mortgage, a non-integrated lender's collateral in developments with an integrated lender is thus worth about $4-5\%$ less than if the non-integrated lender did not face the adverse selection. I would thus expect the pricing adjustment to the presence of an integrated lender to be roughly similar to the interest rate increase when raising the initial loan-to-value ratio by 5 percentage points. This indeed appears to be the case, at least anecdotally. A rate sheet by *US Bank* from June 2008 shows pricing adjustments to changes in the LTV ratio for 5/1 ARMs. For a borrower with a credit score between 700 and 719, increasing the LTV ratio from 70% to 75% and from 75% to 80% involves an increase in the annual interest rate of 10 basis points each. An increase from 80% to 85% involves rate increase of 20 basis points, and an increase in the LTV ratio from 85% to 90% involves a rate increase of 25 basis points.. These magnitudes are consistent with the interest rate increases detected in Section 7.

8.3 Welfare Effects and Policy Responses

The welfare consequences of an integrated lender's presence and the resulting adverse selection are unclear. In this section I discuss channels through which the integrated lender impacts the functioning of mortgage and housing markets. Before policy conclusions about regulating and possibly restricting the behavior of integrated lenders can be reached, these additional factors would need to be quantified. This is beyond the scope of the present paper.

The first channel through which the integrated lender impacts welfare is through its direct effect on borrowers. I found that borrowers on average pay about 10 basis points higher interest rates in the presence of an integrated lender. It is possible that the resulting rents to the integrated lender are competed away in the construction sector. If this were the case borrowers might pay higher interest rates on their mortgages, but pay less for homes than in the absence of the integrated lender. Since I showed that the interest rate increase is particularly large for borrowers with small downpayments, this might still involve a welfare transfer from households that make low

downpayments to households that make high downpayments or do not require a mortgage at all.

In addition, the adverse selection can introduce costly frictions in the home financing process. When economic fundamentals are strong (in my model this could be represented by high γ), asymmetric information about the collateral might not significantly impact the value of a particular mortgage, as repayment is less sensitive to collateral values. This is similar to the notion of information insensitiveness of Gorton and Pennacchi (1990). When a negative shock hits the economy the repayment probability becomes more sensitive to information about collateral quality. Interest rates and haircuts (required downpayments) will rise and leverage will decline more rapidly than in the absence of asymmetric information. Through a similar mechanism, asymmetric information about asset quality impacts the trading of these assets (Fishman and Parker, 2010; Gorton and Ordonez, 2011). Hence, asymmetric information about collateral quality can be an important propagation mechanism for shocks and can introduce significant volatility in the demand for newly constructed homes.[49]

To the extent that developers determine construction quality, the presence of the integrated lender might also provide a valuable quality guarantee function. The put option inherent in mortgages by the integrated lender provides buyers the option to return low quality homes to the developer. This generates incentives for developers to produce higher quality homes than they might otherwise do.[50] This mechanism might be more effective than competing mechanisms such as warranties.[51] If full payment (rather than just the downpayment) is received by the developer at the time of sale, the buyer must convince the developer that quality is not as promised. This question regularly requires length and expensive lawsuits to be settled. In addition, the developer must still be in business to fulfill its guarantee. When the integrated lender makes the mortgage the decision to return the collateral is transferred to the buyer who can choose when to default.[52]

A developer with some market power might also use its integrated lender to solve the durable goods monopoly problem outlined by Coase (1972). A developer that sells multiple houses over time charges as high a price as it can get in every period, until the last unit is sold at marginal cost. When a developer cannot commit that it will not sell similar houses for less in the future, households may be unprepared to pay more than marginal cost for early units. Consequently, the

[49] Einav et al. (2011) show that consumer demand for used cars is very responsive to changes in borrowing constraints. Geanakoplos (2009) shows that variations in leverage can have a significant impact on assets prices.

[50] Since the price of a house is affected by the value of neighboring homes, this "guarantee channel" generates incentives to produce high-quality homes even when the mortgages are financed by non-integrated lenders.

[51] Warranties for new homes are usually limited in scope and time. The 2008 10-K statement of Beazer, a large U.S. home builder, states that: "We currently provide a limited warranty (ranging from one to two years) covering workmanship and materials per our defined performance quality standards. In addition, we provide a limited warranty (generally ranging from five years up to the period covered by the applicable statute of response) covering only certain defined construction effects."

[52] This mechanism relates to the literature on the extension of trade credit which allows buyers to delay payment. Smith (1987) argues that such delayed payment can facilitate exchange by allowing the buyer to verify product quality before paying. Long et al. (1993) provide empirical evidence for such a theory, by showing that firms producing products where quality requires more time to assess extend more trade credit.

developer can lose all its market power. Bulow (1982) shows that if rather than selling the houses the monopolist leases it to consumers, it can retain its market power. This is because if the leasor reduces the price for later units, it suffers a capital loss on existing units and thus internalizes the cost of future price reductions. By providing mortgage financing through an integrated lender the developer remains exposed to the return on houses in a way similar to a leasing firm. This means that the presence of the integrated lender might provide the developer with a mechanism to retain market power, which could lead to higher prices for newly constructed homes.

8.4 Effect of Securitization

During the period under consideration, the years 2000 - 2007, a significant proportion of mortgages were securitized and sold as mortgage-backed securities (MBS). It is often argued that this ability to securitize mortgages led to moral hazard in mortgage origination (Keys et al., 2010), since originators would not face the costs of default. This would also suggest that the incentives of non-integrated lenders to avoid the winner's curse may be reduced. In this section I argue that while securitization can *reduce* the exposure to an eventual default of the mortgage, it usually does not *eliminate* it.[53] This explains why I do observe lenders responding to concerns about adverse selection on collateral quality. There are a number of reasons for this.

First, in private label securitization, which made up 56% of all MBS issued in 2006, the issuer oftentimes retains tranches of varying seniority, generating direct exposure to the performance of the underlying mortgages. In addition, securitizers usually retain exposure to credit risk through credit enhancements such as overcollateralization and excess spread, which are used to cover default losses.[54] Gorton and Souleles (2007) propose that a sponsor's support of a securitized product might be provided by a relational contract, or implicit recourse, since secondary market buyers will not buy from sponsors whose securities frequently default. They find that the spreads demanded by investors for securitized products are decreasing in the rating of the sponsor, even though the investors have no legal recourse against the sponsor.[55]

[53] In future work it would be interesting to use data from the HMDA to test whether mortgage originators are more likely to securitize mortgages backed by low-quality collateral (and whether such an effect is bigger for integrated lenders, who are likely to have the best information about collateral quality).

[54] Average excess spread for subprime MBS in 2006 was 2.5%, and 61% of all MBS issued by private firms in 2006 were overcollateralized (Rosen, 2007). There is a debate about the effectiveness of credit enhancements in generating incentives for optimal screening of borrowers if they can subsequently be sold. Section 941 of the Dodd-Frank Act introduces a formal risk-retention requirement. For most mortgages, it requires the sponsor of a securitization to retain credit risk corresponding to 5% of the securitized assets. The rule prohibits a sponsor from transferring or selling any assets held to fulfil risk retention requirements. It does, however, permit the sponsor to allocate some its risk retention obligation to any originator that contributed at least 20% of assets in the pool (Alvarez, 2011).

[55] Gorton and Souleles (2007) compare the impact of the senior unsecured credit rating of a sponsor on the initial spread paid on observationally similar AAA issuances of securitized credit card loans. They find that relative to sponsors in the top quartile of the quality distribution, the riskist 25 percent of sponsors must pay an additional 46 basis points on average. This is about the same size as the average spread for these securities. They cite the rating agency FitchIBCA: "Although not legally required, issuers may feel compelled to support a securitization and absorb credit risk beyond the residual exposure. In effect, there is moral recourse, since failure to support the securitization may impair future access to capital markets." The recourse will often not be made explicit to be

Second, the sale of mortgages to investors often includes recourse clauses that require the lender to take back loans if specific events such as borrower default occur (Michaud, 1996). Similarly, collateral substitution clauses require lenders to substitute performing loans for loans that go into default (Engel and McCoy, 2006). Standard purchase contracts with the government-sponsored enterprises Fannie Mae and Freddie Mac include "reps and warranties" in which the originator guarantees certain standards of origination. A violation of the reps and warranties requires the seller to buy back the affected mortgages from the trust. This is sometimes referred to as a mortgage put-back. Such put-backs have occurred frequently since the unfolding of the crisis.[56] Third, since a significant part of a mortgage lender's income derives from mortgage servicing, lenders remain exposed to the default of mortgages, even after they are securitized. Gorton (2008) shows that on March 31, 2008, about 9% of Countrywide's assets were comprised of mortgage servicing rights and that Countrywide took write-downs worth hundreds of millions of dollars against such mortgage servicing rights when foreclosures started to increase. Fourth, D'Silva and Gordon (2008) discuss the exposure of originators to credit risk of mortgages over the holding period of several months prior to securitization. This means that even for mortgages that are eventually securitized, the originator is exposed to a possible default for several months after the initial loan is made.

9 Conclusion

In this paper I analyze the sources and magnitude of asymmetric information in mortgage lending. I exploit that property developers often cooperate with an integrated mortgage lender that might have superior information about the quality of the housing collateral and about borrower characteristics.

able to treat the transfer of assets to the SPV as a sale rather than a loan, which allows a reduction of regulatory capital and the maintenance of bankruptcy remoteness. In addition to losing access to financial markets, lenders with excessive default rates may also lose their ability to originate FHA-insured mortgages. 12 U.S.C. 1735f-11 directs HUD to review the rates of "early serious defaults and claims" on FHA-insured loans. HUD's Credit Watch Terminiation Initiative was set up to identify lender branch offices with unacceptably high default and claim rates for loans originated within the preceeding 24 months. Under Credit Watch, HUD can terminate a lender's authority to originate FHA-insured loans in a geographic area if the default rate exceeded both the national average and the average local HUD field office default rate by 200%.

[56]For example, the 2010 10-K filings of Ryland Group, a large U.S. home builder with an integrated mortgage lender, shows that it incurred over $30 million in such indemnification expenses since the beginning of the crisis. It states that "the Company [Ryland] is required to indemnify its investors to which mortgage loans are sold if it is shown that there has been undiscovered fraud on the part of the borrower; if there are losses due to origination deficiencies attributable to RMC [Ryland Mortgage Company]; or if the borrower does not make a first payment." MDC Holdings, the parent company of Richmond American, another large U.S. home builder that cooperates with the integrated lender HomeAmerican, states in its 2010 10-K filing regarding major business risks: "Further uncertainty in the mortgage lending industry, including repurchase requirements associated with HomeAmericans sale of mortgage loans, could negatively impact our results of operations. These risks may include, among other things, compliance with mortgage loan underwriting criteria and the associated homebuyers performance, which could require HomeAmerican to repurchase certain of those mortgage loans or provide indemnification. Repurchased mortgage loans could have a substantial impact on HomeAmericans results of operations, liquidity and cash flow as the existence of a defect that necessitated repurchase may require additional effort and expense incurred by HomeAmerican to cure the defect, the passage of time in order to cure or reduce the impact of an identified defect, a discounted sale of the repurchased loan due to the existence of a defect or, in the event that the loan has a defect and is non-performing, foreclosure and re-sale of the subject property."

By conditioning their interest rate offers on this superior information, integrated lenders subject competing non-integrated lenders to adverse selection. I show that in developments with an active integrated lender the portfolio of houses financed by the integrated lender is of above average quality along a number of key dimensions.[57] In particular, its annual return is about 50 basis points higher than that of houses financed by non-integrated lenders. This outperformance is particularly large for houses built on expansive soil, which makes housing return more sensitive to construction quality. The outperformance is also larger for mortgages with a low loan-to-value ratio for which repayment is less sensitive to changes in collateral values. The adverse selection also translates into higher foreclosure rates for mortgages granted by non-integrated lenders. By analyzing the differences in return and foreclosure probabilities over the ownership period of the second owner, I show that a significant amount of the asymmetric information is about collateral quality, not borrower characteristics. I also compare the performance of ex-ante similar housing collateral of non-integrated lenders when all lenders are equally informed to the performance when the non-integrated lender competes against an integrated lender. I find that in the latter case the collateral underperforms by about 60 basis points annually. To compensate for lending against below-average quality collateral, non-integrated lenders charge about 10 basis points higher interest rates when competing against an integrated lender. This interest rate increase is larger for houses built on expansive soil and for houses financed with high loan-to-value ratio mortgages, the repayment of which is more sensitive to changes in collateral values.

My results highlight the pervasive nature of asymmetric information in mortgage markets. While the debate surrounding such asymmetric information usually focuses on information about the borrower, I show that asymmetric information about the mortgage collateral is common and of significant magnitude. This suggests that policy proposals aimed at reducing the amount of asymmetric information in mortgage markets should also focus on providing better information about collateral quality, potentially by making more detailed property inspection records available publicly. Some of the relevant information is already being collected by insurance claims information systems such as the Automated Property Loss Underwriting System (A-PLUS) or the Comprehensive Loss Underwriting Exchange (CLUE),[58] which track past insurance claims against a property. This information is currently only available to insurance companies for the performance of underwriting functions. Expanding access to such information to all housing market participants (including secondary market purchasers of MBS) might be a first step at reducing the impact of asymmetric information about collateral values in mortgage lending.

[57] While the activity of integrated lenders in new developments provides a clearly identifiable measure of relative information and thus facilitates an empirical assessment of the sources and magnitude of asymmetric information in mortgage lending, it is likely that there is similar asymmetric information about housing collateral values in lending to purchase existing properties. For example, lenders often acquire superior information about local demand factors that will impact future house prices in a specific geographic or price segment.

[58] For a sample report, see http://www.iso.com/Products/A-PLUS/Sample-A-PLUS-Property-Report.html.

A Theoretical Appendix

I begin by introducing some notation. Recall that q is the unconditional probability of a house being of high quality ($\theta = h$) and that ϕ is the pecision of the integrated lender's signal. The probability of observing signal $\eta = h$ is defined as $P_i(h) = q\phi + (1-q)(1-\phi)$. The probability of observing signal $\eta = l$ as $P_i(l) = (1-q)\phi + q(1-\phi)$. The probability that a house is high quality conditional on having received a positive signal ($\eta = h$) is given by $p(h,\phi) = Pr(\theta = H|\eta = h) = \frac{q\phi}{q\phi+(1-q)(1-\phi)}$. The probability that the house is house is high quality conditional on having received a negative signal ($\eta = l$) is given by $p(l,\phi) = Pr(\theta = H|\eta = l) = \frac{q(1-\phi)}{(1-q)\phi+q(1-\phi)}$.

The expected revenue from lending at interest rate R to a type-γ agent wanting to buy a house with signal η is defined as:

$$\begin{aligned} W(R;\eta,\phi,\gamma) &= p(\eta,\phi)R + [1-p(\eta,\phi)]\gamma R \\ &= [p(\eta,\phi)(1-\gamma)+\gamma]R \\ &= z(\eta,\phi,\gamma)R \end{aligned}$$

The term $z(\eta,\phi,\gamma)$ represents the probability of repayment of the loan conditional on having observed signal η with signal precision ϕ. I define $R(\gamma)_a^b = \frac{R_f}{q+\gamma(1-q)}$ to be the break-even interest rate when lending to a type-γ agent to purchase the average house. In addition, I define $R(\gamma,\phi)_l^b = \frac{R_f}{z(l,\phi,\gamma)}$ to be the break-even interest rate for the integrated lender when lending to a type-γ agent who wants to purchase a house with signal $\eta = l$.

Theorem 1 *There are no pure strategy equilibria.*

Proof The proof follows by contradiction. Let pure strategies be $R_i(\eta)$ for the integrated lender and R_u for the non-integrated lender. The only possible pure strategy equilibrium is $R_a = R_u = R_i(h) = R_i(l)$. Assume otherwise. If $R_u < R_i(h), R_i(l)$, the non-integrated lender can increase its expected return by offering $R_u' = R_u + \varepsilon$. On the other hand, if $R_i(\eta) < R_u$, the integrated lender can increase its expected return by offering $R_i(\eta)' = R_i(\eta) + \varepsilon$. However, each lender offering R_a is also not an equilibrium. If $R_a < R(\gamma)_a^b$, each lender would be better off not offering a mortgage at all. If $R_a > R(\gamma,\phi)_l^b$, then the integrated lender would be better off by offering interest rates $R_i(l)' = R_a - \varepsilon$ and $R_i(h)' = R_a - \varepsilon$. If $R(\gamma)_a^b < R_a < R(\gamma,\phi)_l^b$ the integrated lender would be better off offering $R_i(l)' = R_a + \varepsilon$ and $R_i(h)' = R_a - \varepsilon$, subjecting the non-integrated lender to a winner's curse. The non-integrated lender would make a loss in expectation.

Theorem 2 *Let $W(R;\eta,\phi,\gamma)$ be the expected revenue to the integrated lender from lending at rate R to a type-γ borrower to buy a house with signal η. The interest rate offer game for a type-γ borrower when signal precision is ϕ has a unique mixed strategy equilibrium, such that:*

1. *The non-integrated lender breaks even, while the integrated lender earns positive expected profits.*

2. *There exists a value $\bar{\gamma}$ such that for borrowers with $\gamma < \bar{\gamma}$ the integrated lender rejects all mortgage applications to buy houses with $\eta = l$. When $\eta = h$, the integrated lender randomizes interest rate offers over $[R(\gamma)_a^b, R(\gamma)_m)$ using the following cumulative distribution function:*

$$F_i(R; h, \phi, \gamma) = 1 + \frac{P_i(l)[W(R; l, \phi, \gamma) - R_f]}{P_i(h)[W(R; h, \phi, \gamma) - R_f]}. \tag{15}$$

$R(\gamma)_a^b = \frac{R_f}{q+\gamma(1-q)}$ is the break-even interest rate for lending to a type-γ agent to buy an average quality house. $P_i(\eta)$ is the probability of the integrated lender observing signal η. The integrated lender also makes interest rate offers with a point mass of $1 - F_i(R(\gamma)_m; h, \phi, \gamma)$ at $R(\gamma)_m$. The non-integrated lender randomizes interest rate offers over $[R(\gamma)_a^b, R(\gamma)_m)$ using the following cumulative distribution function:

$$F_u(R; \phi, \gamma) \quad = \quad 1 - \frac{W(R(\gamma)_a^b; h, \phi, \gamma) - R_f}{W(R; h, \phi, \gamma) - R_f}. \tag{16}$$

With probability $1 - F_u(R(\gamma)_m; \phi, \gamma)$ the non-integrated lender does not make an offer.

3. *For borrowers with $\gamma > \bar{\gamma}$ both the integrated lender and the non-integrated lender always offer a mortgage. When $\eta = l$ the integrated lender offers the break-even interest rate $R(\gamma, \phi)_l^b$, defined implicitly by $R_f = W(R(\gamma, \phi)_l^b; l, \phi, \gamma)$. When $\eta = h$ the integrated lender randomizes its interest rate offers over $[R(\gamma)_a^b, R(\gamma)_m]$ using $F_i(R; h, \phi, \gamma)$. The non-integrated lender always randomizes over $[R(\gamma)_a^b, R(\gamma, \phi)_l^b)$ using $F_u(R; \phi, \gamma)$, with a point mass at $R(\gamma, \phi)_l^b$.*

The competitive game amongst lenders is similar to that of a first-price sealed bid common value auction, analyzed by Milgrom and Weber (1982) and others. There the authors show that in a similar setup, when the information set of the less informed competitors is less finely partitioned, the less informed lenders will make zero profit in equilibrium. In the following, I find the unique mixed strategy equilibrium. The proof of this theorem follows a number of steps in similar proofs in Hauswald and Marquez (2006), von Thadden (2004) and others.

Proof Let $F_i(R; \eta, \phi, \gamma)$ represent the cumulative distribution function (cdf) of the integrated lender's distribution of the interest rate offers R for a type-γ borrower wanting to buy a house with signal η when the integrated lender's signal precision is ϕ. Let $F_u(R; \phi, \gamma)$ be the cdf of the non-integrated lender's distribution over interest rate offers R for a type-γ borrower wanting to buy a house when the integrated lender's signal precision is ϕ. Both $F_i(R; \eta, x, \gamma)$ and $F_u(R; \phi, \gamma)$ are continuous, strictly increasing and atomless on a common support $[\underline{R}, \bar{R}]$ (see von Thadden, 2004).

The intuition for the existence of $\bar{\gamma}$ relies on the idea that for each level of signal precision ϕ there is a household to whom it is no longer valuable to lend at the highest possible rate if the signal about collateral quality is negative. The cut-off level of γ below which the integrated lender does not want to lend to agents wanting to borrow to purchase a house with $\eta = l$ is defined as the solution

to $R(\bar{\gamma},\phi)_l^b = R(\bar{\gamma})_m$. The cut-off level $\bar{\gamma}$ is increasing in the precision of the integrated lender's signal, ϕ. For higher signal precision, the probability that house is truly $\theta = l$ when the integrated lender observes $\eta = l$ is higher. In addition, $\bar{\gamma}$ is decreasing in q, the proportion of houses that are of high quality. It is also decreasing in H.

Since a less informed bidder cannot profit from a sealed-bid auction against a better-informed competitor, the non-integrated lender must break even in equilibrium. This allows us to calculate the lower bound of the support. Note that when offering $\underline{R}$ the non-integrated lender wins almost surely and since it needs to make a profit of 0, we have that $\underline{R} = R(\gamma)_a^b$. The upper bound on distribution, $\bar{R}$, depends on the value of γ. When $\gamma \geq \bar{\gamma}$ a repeated undercutting argument, which is described in the main text and is similar to standard Bertrand competition, shows that for $\eta = l$ the integrated lender offers $R(\bar{\gamma},\phi)_l^b$ and makes zero profit. When $\eta = h$, the integrated lender mixes offers on the support of $[R(\gamma)_a^b, R(\gamma,\phi)_l^b)$. When $\gamma < \bar{\gamma}$, on the other hand, the integrated lender never makes an interest rate offer if $\eta = l$ and mixes over $[R(\gamma)_a^b, R(\gamma)_m)$ when $\eta = h$. For any ϕ, the common support is thus given by $[R(\gamma)_a^b, \min\{R(\gamma,\phi)_l^b, R(\gamma)_m\})$.

The expected profit for the integrated lender from offering an interest rate R when $\eta = h$ can be stated as (recalling that the integrated lender will make zero profits if $\eta = l$):

$$\begin{aligned} \pi_i(R;h,\phi,\gamma) &= \text{Probability of winning} \times \text{Expected Profit when Winning} \\ &= [1 - F_u(R;\phi,\gamma)] \times [W(R;h,\phi,\gamma) - R_f] \end{aligned} \tag{17}$$

Equivalently, the expected profit for the non-integrated lender from offering interest rate R is:

$$\begin{aligned} \pi_u(R;\phi,\gamma) &= [(\text{Prob. } i \text{ has } \eta = l) \times (\text{Expected Profit when } i \text{ has } \eta = l)] + \\ &\quad [(\text{Prob. } i \text{ has } \eta = h) \times (\text{Prob. of winning}) \times \\ &\quad (\text{Expected Profit when } i \text{ has } \eta = h)] \\ &= P_i(l)[W(R;l,\phi,\gamma) - R_f] + \\ &\quad P_i(h)[1 - F_i(R;h,\phi,\gamma)][W(R;h,\phi,\gamma) - R_f] \end{aligned} \tag{18}$$

Since the non-integrated lender must break even, we have that $\forall(R,\gamma) : \pi_u(R;\phi,\gamma) = 0$. In addition, since the mixing distributions are strictly increasing, equilibrium profit for each lender must be the same for every interest rate offered in the support: $\pi_i(R;h,\phi,\gamma) = \bar{\pi}(\phi,\gamma)$. If we now evaluate $\pi_i(R;h,\phi,\gamma)$ at the lower bound of the support, since $F_u(R(\gamma)_a^b;\phi,\gamma) = 0$, we have that $\bar{\pi}(\phi,\gamma) = W(R(\gamma)_a^b;h,\phi,\gamma) - R_f$.

Plugging this into equation (17) and solving for $F_u(R;\phi,\gamma)$ gives:

$$F_u(R;\phi,\gamma) = 1 - \frac{W(R(\gamma)_a^b;h,\phi,\gamma) - R_f}{W(R;h,\phi,\gamma) - R_f} \tag{19}$$

Similarly, solving equation (18), by setting $\pi_u(R;\phi,\gamma) = 0$ gives:

$$F_i(R;h,\phi,\gamma) = 1 + \frac{P_i(l)[W(R;l,\phi,\gamma) - R_f]}{P_i(h)[W(R;h,\phi,\gamma) - R_f]} \tag{20}$$

Since both lenders randomize over the full support of the distribution functions, they cannot profitably deviate from their mixed strategies. Hence, the preceding distributions represent the unique equilibrium for a borrower of type γ.

Probability of making an offer: $\gamma < \bar{\gamma}$

For $\gamma < \bar{\gamma}$, $F_i(\bar{R};h,\phi,\gamma) = F_i(R(\gamma)_m;h,\phi,\gamma) < 1$. Hence the integrated lender randomizes over $[R(\gamma)_a^b, R(\gamma)_m)$ for $\eta = h$ houses, without any atoms, but with point mass at $R(\gamma)_m$, where the mass is equal to $1 - F_i(R(\gamma)_m;h,\phi,\gamma)$.[59]

The integrated lender never bids for $\eta = l$ agents and bids with probability 1 for $\eta = h$ agents. The non-integrated lender bids with probability $F_u(R(\gamma)_m;\phi,\gamma) < F_i(R(\gamma)_m;h,\phi,\gamma) < 1$ for all agents. With probability $1 - F_u(R(\gamma)_m;\phi,\gamma)$ the non-integrated lender does not make an interest rate offer and the household gets rationed.

Probability of making an offer: $\gamma \geq \bar{\gamma}$

For $\gamma \geq \bar{\gamma}$, both lenders always make an offer to the borrower. I argued above that for $\eta = l$ the integrated lender always offers credit at $R(\gamma(\phi),\phi)_l^b$, making zero profit. For $\eta = h$ we have $F_i(R(\gamma,\phi)_l^b;h,\phi,\gamma) = 1$, since $R_f = W(R(\gamma,\phi)_l^b;l,\phi,\gamma)$ and $\bar{R} = R(\gamma,\phi)_l^b$ for $\gamma \geq \bar{\gamma}$. Hence the informed lender will make an offer by randomizing over the full support without atoms. Similarly, $F_u(R(\gamma,\phi)_l^b;\phi,\gamma) < 1$, so the uninformed lender will also randomize over the full support, with a mass point of $1 - F_u(R(\gamma,\phi)_l^b;\phi,\gamma)$ at $R(\gamma,\phi)_l^b$.

[59] In order for a lender to not make an interest rate offer in some instances, it must be indifferent between bidding and not bidding. Since the integrated lender makes a profit in expectation when making an offer on the $\eta = h$, it is never indifferent between bidding and not bidding, which generates expected profits of zero. Thus, unlike the non-integrated lender, it will never not bid.

B Data Appendix

I begin with a dataset that contains all 3.19 million ownership-changing deeds recorded in Arizona between 2000 and 2010. The data include both armslength market transactions, as well as transfers in divorce, estate settlements and foreclosures. For each deed with sufficient information to uniquely identify the property, the address is geocoded to determine the property's precise location. For 92.5% of deeds the address information is sufficiently detailed to determine the exact latitude and longitude. For another 1.9% of deeds the street-number is missing and a latitude and longitude is assigned that locates the property at the geographic midpoint of the street. The 5.6% of deeds with insufficient address information to assign a location are dropped (many of them refer to to the sale of parcels of vacant land). I then merge each deed via its asssessor parcel number (APN) and county to the underlying property's tax assessment record for the year 2010.

B.1 Identifying Transaction Types

Armslength Transactions: I identify all deeds that contain information about armslength transactions in which both buyer and seller act in their best economic interest when agreeing upon a price. This ensures that observed transaction prices reflect the market value of the property. The process of identifying armslength transactions follows a procedure similar to Caplin et al. (2008). I include all deeds that are one of the following: "Grant Deed," "Condominium Deed," "Individual Deed," "Warranty Deed," "Joint Tenancy Deed," "Special Warranty Deed," "Limited Warranty Deed" and "Corporation Deed." This excludes intra-family transfers and foreclosures. I drop all observations that are not a Main Deed, that only transfer partial interest in a property, or that transfer ownership in more than one parcel. This leaves 1.67 million armslength transactions.

Newly Developed Single-Family Residences: Amongst the armslength transactions I identify sales of newly developed properties. This includes all deeds in which the seller is identified as a company or a partnership, but that are not REO resales (i.e. sales by a bank following a foreclosure). I exclude sales in which the construction date of the house (as reported in the assessor data) precedes the sales date by more than two years. These transactions usually involve a developer that renovates and resells existing properties. I also exclude transactions where the buyer is identified as a company. In addition, I only consider single-family residences, which make up about 85% of newly developed properties in Arizona. This leaves me with 228,952 observations. For each newly developed property I collect subsequent armslength sales to track the their future return.[60]

Divorce and Death: I identify those repeat sales pairs for which I observe a divorce or death of the owners up to six months before the second sale. I identify divorces through the presence of an "Intra-Family Transfer & Dissolution" deed that transfers property rights from initially joint

[60] I exclude repeat sales pairs for which the time difference between the two sales is less than 270 days. Such sales often precede or follow the redevelopment of a property. For similar reasons, the calculations of the Case-Shiller house price index exclude transaction pairs with less than six months time difference (Standard & Poor's, 2009).

ownership to one of the initial owners. The death of an owner is identified if either (i) the seller on a deed is classified as an "estate", "executor", "deceased" or "surviving joint owner" or (ii) if I observe one of the following: "Affidavit of Death of Joint Tenant" or "Executor's Deed."

Foreclosures: I mark those properties that experience a foreclosure within three years of the initial sale by the developer. A foreclosure event is identified (i) if the deed is either a "REO Repossession", "REO Resale", "Foreclosure Deed", "Deed in Lieu of Foreclosure", "Trustee's Deed" or (ii) when the buyer is identified as a "beneficiary."

Data Cleaning: I identify houses in the same development by combinations of seller identity and census tract. I only consider houses that were first sold before 2008 and are located in developments with more than 30 units. I drop a few observations that are likely to have misreported loan or sales price details (i.e. when the sales price is less than $\$25,000$ or more than $\$10$ million and when the LTV ratio is more than 1.3 or less than 0.3).[61] In addition, I only keep observations with a full set of control variables in the assessor data.[62] This leaves me with $152,215$ observations.

B.2 Deeds Data to HMDA Merge

I next merge the deeds to data from the Home Mortgage Disclosure Act (HMDA). This allows me to obtain additional characteristics of the home owners, as well as information on the subsequent securitization of mortgages. The HMDA is a mortgage-level dataset and identifies a mortgage by year, census tract, mortgage amount and mortgage lender.[63] Bayer et al. (2011) use these characteristics to merge a dataset similar to my deeds data to the HMDA. This allows them to uniquely match about 70% of all sales. I use additional characteristics to improve match rates and quality. First, both the deeds and HMDA data report whether mortgages are FHA-insured or VA-guaranteed. Second, HMDA data identifies whether a house is purchased as a rental property, while the assessor data has information about whether it was owner-occupied in 2009. Third, the HMDA contains information about whether the mortgage was applied for by a male, female, or two applicants. The deeds data also identifies purchasers as male, female or a married couple. Fourth, the HMDA data has information about the race and ethnicity of applicants. In the deeds data I do not have this information, but I do observe the names of buyers. I match the surnames of buyers to the 1000 most common Asian and Latino surnames from the 2000 U.S. Census.[64] Using these four additional characteristics allows me to confirm $63,054$ unique matches.

[61] This procedure is similar to that in Levitt and Syverson (2008) and Ben-David (2011).

[62] This primarily drops observations from Pima county (city of Tuscon), which does not usually provide lot size and building size in the assessment records. See discussion in Appendix B.4.

[63] The Federal Reserve's Regulation C, which governs the HMDA, applies to most depository institutions with a branch office in a metropolitan area. Banks below $39 million in assets are exempt from reporting requirements, as are nondepository institutions (mainly mortgage companies) with assets below $10 million. Avery et al. (2005) estimates that about 20% of all mortgages granted in 2004 were thus not covered by the HMDA.

[64] The underlying dataset, which includes the number of occurances and self-reported ethnic breakdown for every surname occuring more than 100 times in the 2000 census is provided by Word et al. (2008).

Despite the use of additional match variables, my unique match rate is lower than the one reported by Bayer et al. (2011). There are a number of possible reasons for this: First, since the integrated lender makes a significant number of mortgages in new developments, the power of using lender identity to merge deeds to HMDA data declines. Second, lenders in new developments might be more likely to fall below the asset reporting threshold. These explanations suggest that those deeds that I can match uniquely may not represent a random sample of all mortgages, since they are more likely to have been granted by the integrated lender. For my main data set, for those mortgages where more than one match is possible, I match each deed randomly to one of the possible records in the HMDA data. I can merge a total of 98,706 deeds to HMDA data. In Appendix D.1 I show that the key empirical results that use this data set are robust to considering (i) only the sample of houses with a unique HMDA merge, and (ii) the full sample of houses in my data, without requiring an HMDA merge and without conditioning on owner characteristics.[65]

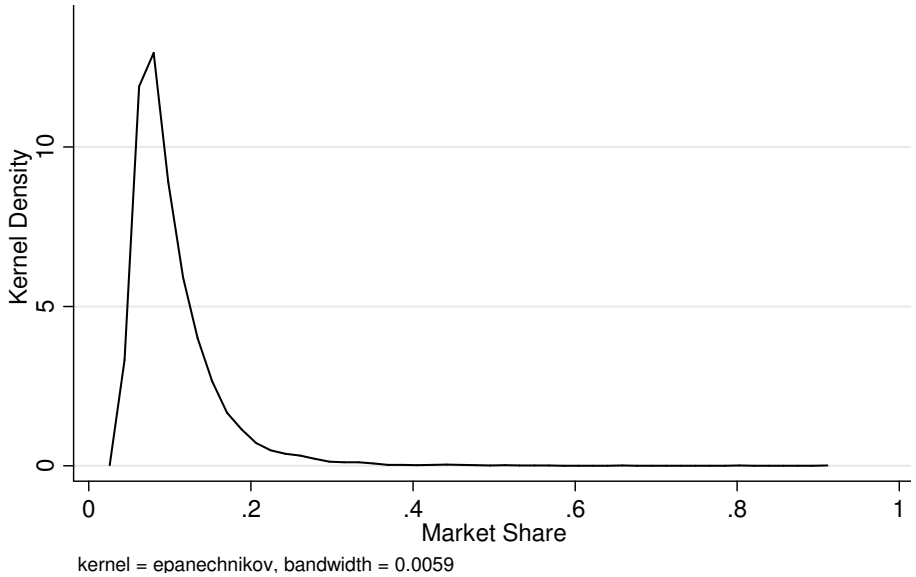
B.3 Identifying Integrated Lenders

To identify integrated lenders, I follow a number of steps: First, developers usually own their integrated lenders (e.g. the developer "Shea Homes" owns "Shea Mortgage"). For each developer, I determine whether there is joint ownership with its largest lender, using OneSource North American Business Browser and information from SEC filings. If I can confirm joint ownership, I assign the lender to be the integrated lender of this developer. This allows me to assign integrated lenders to 32 developers and identify 51,707 mortgages that were granted by integrated lenders. I also analyze instances in which the market share of a single lender in a development exceeds 50%, but in which the developer does not own this lender. In these cases, I also assign the lender to be integrated. This assigns another 9,186 transactions to have mortgages granted by an integrated lender. Using this process of identifying integrated lenders, 84.8% of newly developed single-family residences are in a development with an integrated lender. For those houses that are in developements with an integrated lender, the integrated lender has a market share of 72.8%. I believe that this process of identifying integrated lenders is appropriate: Figure 8 shows the distribution of the market share of the largest lender for lending to purchase *existing* homes. The unit of observation is a census-tract/year combination. There are essentially no census tracts in which the largest lender has a market share in excess of 35% for mortgages to purchase existing homes. Consequently, in any development in which a lender attains more than 50% of all mortgages, it is very likely that this lender is only able to obtain such a market share through an integrated lender arrangement.[66]

[65] Not conditioning on owner income is unlikely to contaminate my results. The results in Section 5.1 and the robustness check using only unique merges suggest that there is no selection into the integrated lender portfolio along observable owner characteristics that also affect the subsequent performance of the housing collateral.

[66] The additional lenders identified through this channel are usually independent companies that specialize in providing financing for developers in a integrated lender role. One example is *IMortgage*, which states on its website: "We partner with homebuilders across the country to establish and manage their mortgage operations. We originate, underwrite, process and close mortgages on newly constructed homes."

Figure 8: Market Share of Biggest Lender in Census Tracts with Existing Homes



Note: This figure shows the kernel density of the market share of the biggest lender for lending to purchase existing homes. The unit of observation is a census-tract/year combination with at least 50 mortgages granted. I include all such combinations in Arizona between 2000 and 2007. N = 6121.

In Appendix D.2 I present a robustness check of my results that only considers integrated lenders identified through joint ownership with the developer.

B.4 Summary Statistics

Table 12 shows how the observations in my data are distributed over time and across counties. It includes all observations for which I have an HMDA-merge and which contain the full set of covariates. Summary statistics are split up by developments with and without an integrated lender. For developments with an integrated lender, the results are given separately for the integrated lender and for other non-integrated lenders. The top panel shows that the majority of observations are from Maricopa and Pinal county, which constitute the Phoenix MSA. Pima county (including Tuscon) only contributes a few observations. This is because for Pima I only observe building and lot size for a small number of observations in the assessment data. These variables are important controls in my main specifications. In order to estimate all models with a common sample, observations with missing data on home characteristics were dropped.[67] The bottom panel shows the distribution of observations by year of sale. The number of newly developed properties sold increased up to 2005, the peak of Arizona's housing boom, and then declined markedly during the financial crisis.

[67] A robustness check shows that the main results are unaffected when including observations from Pima county and dropping the relevant control variables with incomplete field population from the empirical model. This adds about 6,000 observations in developments with an integrated lender, 1,800 of them with a repeat sale.

Table 12: Number of Observations by County and Year

	No Integrated Lender		Has Integrated Lender				*Total*
			Integrated Lender		*Other Lender*		
	No.	%	No.	%	No.	%	No.
County							
Cochise	95	64.2	35	23.6	18	12.2	148
Coconino	151	100.0	0	0.0	0	0.0	151
Maricopa	12,130	15.5	47,725	60.8	18,578	23.7	78,433
Mohave	26	27.7	45	47.9	23	24.5	94
Pima	9	20.0	20	44.4	16	35.6	45
Pinal	1,127	6.3	12,829	71.4	4,000	22.3	17,956
Yavapai	541	71.1	148	19.4	72	9.5	761
Yuma	958	85.7	91	8.1	69	6.2	1,118
Total	15,037	15.2	60,893	61.7	22,776	23.1	98,706
Year Sold							
2000	1,759	21.6	4,593	56.4	1,786	21.9	8,138
2001	1,853	17.7	6,704	64.0	1,913	18.3	10,470
2002	1,750	16.8	6,732	64.7	1,919	18.5	10,401
2003	2,035	16.7	7,659	63.0	2,462	20.3	12,156
2004	2,749	16.8	9,077	55.6	4,496	27.5	16,322
2005	2,822	17.7	9,048	56.8	4,072	25.5	15,942
2006	1,286	9.3	8,722	63.0	3,842	27.7	13,850
2007	783	6.9	8,358	73.1	2,286	20.0	11,427
Total	15,037	15.2	60,893	61.7	22,776	23.1	98,706

Note: This table shows the number of observations in the primary dataset used in this paper. It includes observations with a successful HDMA merge and a full set of covariates.

Table 13 shows summary statistics for the key control variables used in the regressions. Most of these variables are not included linearly in the regression, but by splitting them into groups of values that are represented by dummy variables. This allows a more flexible functional form than could be achieved by including these variables linearly. The results are not sensitive to the exact definition of groups. All dollar amounts are in year-2000 dollars.

House Characteristics: Controls for initial sales price are included by adding dummy variables for $\$10,000$ buckets. Lot size and building size are controlled for by adding dummy variables for 20 equally sized groups. To control for garage spaces, I add a dummy variable for each possible value.

Borrower and Financing Characteristics: Income is controlled for by adding dummy variables for 50 equally sized groups. The loan-to-income (LTI) ratio is included by adding dummy variables for mortgages with LTI ratio ≤ 1.5, between 1.5 and 2, between 2 and 2.5, between 2.5 and 3, between 3 and 3.5 and > 3.5. The loan-to-value (LTV) ratio is included by dummy variables for mortgages with an LTV $\leq 80\%$, between 80% and 90%, between 90% and 97% and $> 97\%$, (where 3.5% is the minimum downpayment for FHA mortgages).

Census Tract Demographics: I control for the median income as well as the proportion of

adults over 25 with at least a high school diploma.[68] I control for census tract demographics by including dummy variables for the following (roughly equally sized) median income groups: $\leq$ \$35k, \$35k - \$50k, \$50k - \$65k, \$65k - \$75k, \$75k - \$100k and $\geq$ \$100k. Dummy variables for high-school graduation rates are: $\leq 75\%, 75\% - 80\%, 80\% - 90\%, 90\% - 95\%, \geq 95\%$.

Table 13: Summary Statistics of Control Variables

	No Integrated Lender			Has Integrated Lender					
				Integrated Lender			*Other Lender*		
	Mean	SD	Med.	Mean	SD	Med.	Mean	SD	Med.
Housing Characteristics									
Sales Price (k year-2000\$)	232.3	132.9	192.6	202.6	94.23	178.8	219.0	114.9	188.5
Lot Size (Sqft)	8,554	5,115	7,620	7,252	3,318	6,578	7,495	3,871	6,600
Building Area (Sqft)	2,280	843	2,112	2,149	719	2,011	2,248	773	2,113
Price / Sqft (k year-2000\$)	32.9	19.0	28.0	33.1	15.5	19.3	34.5	16.0	31.0
Garage Spaces	1.97	1.10	2	1.79	1.07	2	1.89	1.04	2
Total Rooms	7.58	1.84	7	7.15	1.70	7	7.36	1.71	7
Has Pool	0.31			0.21			0.21		
Owner Occupied	0.78			0.81			0.75		
Financing Characteristics									
LTV Ratio	0.83	0.14	0.80	0.85	0.15	0.86	0.84	0.12	0.80
LTI Ratio	2.53	1.06	2.45	2.75	1.09	2.66	2.56	1.09	2.50
Mortgage Duration (Years)	29.4	3.45	30	29.6	2.80	30	29.8	2.86	30
FHA Insured	0.09			0.18			0.07		
VA Insured	0.04			0.04			0.03		
Jumbo Mortgages	0.008			0.002			0.01		
Borrower Characteristics									
Income (k year-2000\$)	89.9	75.3	70.5	72.3	54.8	59.9	87.5	72.9	68.0
Single Person	0.34			0.36			0.41		
Latino	0.13			0.12			0.14		
Asian	0.04			0.03			0.05		
Census Tract Demographics									
Med. Census Tract Inc. (k \$)	76.4	21.8	76.9	71.4	17.7	72.2	72.4	17.3	68.6
Highschool Grad. Rate	0.89	0.09	0.93	0.89	0.09	0.92	0.89	0.08	0.91

Note: This table shows the mean, standard deviation and median for the control variables. Note that some of the differences between developments with and without integrated lenders might be driven by differences in the variable over time and a declining share of developments without integrated lenders.

[68] These are from the 2005 - 2009 estimates of the American Community Survey, a nationwide continuous survey that samples nearly 3 million addresses annually. The 2005 - 2009 estimates are based on 15 million sampled addresses.

B.5 Tax Assessment Process in Arizona

Arizona Revised Statues (A.R.S) 42-11054 (C) tasks tax assessors to annually compute the so-called "full cash value" of each residential property: "In applying prescribed standard appraisal methods and techniques, current usage shall be included in the formula for reaching a determination of full cash value." A.R.S 42-11001(6) specifies the full cash value to be "synonymous with market value, which means the estimate of value that is derived annually by using standard appraisal methods and techniques." The term "market value" has been defined further by courts in Department of Revenue v. Transamerica Title Insurance Company, Court of Appeals of Arizona, 50 p2d 797, 177 Ariz 26 (1977) as "the highest price estimated in terms of money which a property will bring if exposed for sale in an open market, allowing a reasonable time to find a purchaser who buys with knowledge of all uses to which it is adapted and for which it is capable of being used."[69]

The full cash value provided in the tax assessment records is set at 82% of the assessed market value for residential properties, as mandated by A.R.S 42 - 11054(C). The procedure for arriving at these valuations is described by the assessor of Mohave County, Arizona, as follows: "Between January and March of each year, the Assessors Office is required (by Arizona State Statute) to notify property owners of their assessed values for the following tax year. For residential and land parcels, this is accomplished by first collecting sales data in the area in which a property is located. Elements of comparability such as location, view, size, quality and condition are taken into consideration, and a mass appraisal mathematical model is used to arrive at each parcel's value. The market is driven by actual sales that have occurred in a time window established by Department of Revenue guidelines. Increases or decreases in sale prices impact the final assessed valuation." Coconino County's website outlines that there are three appraisal methods used: sales comparison, replacement cost (less depreciation) and income (primarily used for commercial property).

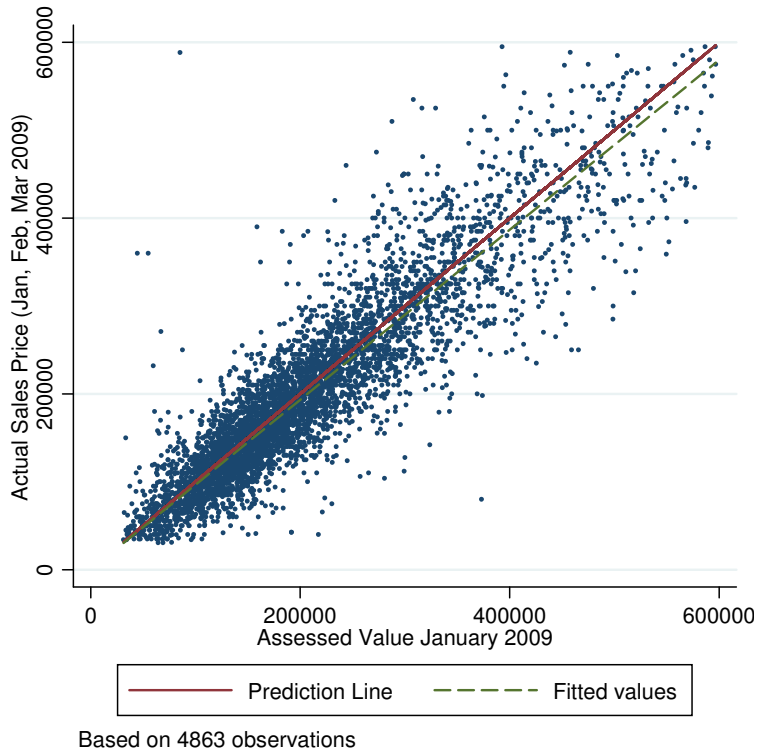
There exist a number of procedures through which a homeowner can challenge a tax assessment if she feels that the house was valued too highly. The appeals process provides a mechanism through which the assessor obtains information about differential depreciation of housing units. According to The Arizona Republic (2009), in 2009 there were 19,801 assessment appeals in Maricopa County, up from 17,213 in 2008 and 13,251 in 2007. This means that about 1.3% of valuations get appealed annually. In 2008, Maricopa County assessors reduced property valuations by a total of $1.9 billion, while the second stage of the appeals process, the Arizona State Board of Equalization, reduced valuations by an additional $2 billion (Arizona Capitol Times, 2009).

In the following I test how well these assessed values in Arizona capture true market values. To

[69] To prevent property taxes from rising too quickly, a significant amount of the tax burden is based on the "Limited Property Value (LPV)." This includes taxes raised for the maintenance and operations budgets of state and local governments and schools. The Full Cash Value is the tax-base for funding voter-approved bond issues, voter-approved budget overrides and special district levies (e.g. for flood, library, jail or fire districts). The LPV is limited in the amount it can increase each year. Growth is limited to the greater of (i) 10% more than the prior year's LPV or (ii) 25% of the difference between the prior year's LPV and the current year's Full Cash Value. The Limited Property Value cannot exceed the Full Cash Value (A.R.S. 42-13301).

do this I consider those properties that were sold in an armslength transaction between January and March 2009 (in this section I use all sales in Arizona, not just those pertaining to newly developed properties). I compare the transaction price with the assessed value in January 2009. In Figure 9, each dot represents such a transaction. The solid line represents the 45° line - if assessments were 100% correct, all observations would lie on this line. It is not suprising that there is a signfiicant spread around the 45° line. Unlike homogenous goods such as stocks and bonds, houses are heterogeneous assets that are sold in a search market. By adjusting the time that a seller is prepared to wait, she can influence the final transaction price. The dashed line represents the prediction from an ordinary least squares regression. The fact that it is very close to the 45° line suggests that on average assessed values capture current market values reasonably well.[70]

Figure 9: Quality of Assessment Values



Note: This figure test for the accuracy of the estimated market value in the assessment data. Each dot represents an observation of a house that was sold in the first three months of 2009 and for which I observe an assessed value in January 2009. On the horizontal axis is the assessed value and on the vertical axis the corresponding transaction price. The solid line represents the 45° line. The dashed line represents the linear prediction of a regression of sales price on transaction price.

[70] In addition, as long as any bias of assessed values from true market values does not differ by the identity of the initial mortgage lender, this would only pose a problem for the interpretation of the finding as "annual return," but not for the detection of asymmetric information.

C Further Results and Discussion

C.1 Sources of Asymmetric Information: Additional Evidence

In Section 3 I presented evidence that there are aspects of construction quality of the housing collateral about which developer and integrated lender could have superior information relative to buyers and non-integrated lenders. Table 14 introduces more examples from construction defect lawsuits, some of which specifically address complaints about construction quality when building on expansive soil. Stauber Associates (2011) provides a more comprehensive list of recent cases and settlements in construction lawsuits in California, Nevada, New York and Florida.

Additional evidence for problems with the construction quality of newly developed homes can be obtained by looking at the complaints of buyers of new homes on websites such as `www.poorlybuiltbypulte.info`, though clearly those have to be considered with caution. A home owner from Pulte's "Stetson Valley" development in Phoenix, AZ, complained about "Walls bowed, broken roof tiles, damaged baseboards, sloppy paint on exterior and interior, block wall has many chips and unpainted areas, yard was too low, some rooms are constantly much warmer than others when using AC, no exhaust fan works, windows are very cheap and made poorly, front door sags, cabinets and counter have defects." A home owner from Del Webb's "Sun City Anthem" development in Henderson, NV, commented that "Foundation has a crack running from one side of the house to the other. Shingles have blown off the roof in high winds. Many cracks in the stucco of the house." An owner from Pulte's "Sheely Farms" development in Phoenix, AZ, complained that "Pulte [filled in] the last half of the subdivision with junk 2x4 homes, destroying our home values."

C.2 Coefficients on Control Variables

In Section 5 I focus on presenting and discussing the impact of $IntegratedLender_i$ on housing returns, controlling for a large number of characteristics of the house, borrower and mortgage. While the impact of these characteristics on return is not a primary focus of my analysis, it is interesting to consider whether their coefficients in my key specification are sensible.

Table 15 shows the coefficients on a number of key control variables. All else equal, homes that are financed with a higher loan-to-income ratio have a lower return. This is consistent with those households contributing a larger fraction of their income towards interest payments, leaving fewer resources for maintenance. Homes owned by single people and by Latinos underperform. Homes in census tracts with higher median income appreciate at a faster rate. A number of characteristics (e.g. initial sales price, income, building size) are not controlled for linearly, but by including dummy variables capturing different value buckets. Figure 10 plots the coefficient on different values of the initial sales price and the building size. All else equal, cheaper houses appreciated at a faster rate over this period, which is consistent with findings in Landvoigt et al. (2011).

Table 14: Construction Defect Lawsuits

Lawsuit	Plaintiffs complaint
Edwin and Vera Bloch *et al.* vs. Del Webb California Corp *et al.*; Calif. Super., Riverside Co.; Case No: 1101735	"The original design, materials, and/or workmanship is typical in that homes have defects in the concrete slabs, stucco, water intrusion membranes, roofs, floors/floor covering, walls, ceilings, drywalls, cabinets, doors and windows, sliding glass doors, shear walls, concrete framework, sheet metal, insulation, electrical systems, heating, ventilation and air conditioning systems, pavement system, plumbing and plumbing fixtures, irrigation systems, soils, grading, framing, stairs, foundations, garage doors, shower doors, mirrors, drainage, paint, fences, fireplaces/chimneys, decks, and structural systems, among other areas." "The resulting damage are typical in that the homes have had water damage inside from, including, but not limited to, cracked concrete, cracked stucco, leaking roofs, leaking condensate lines, leaking windows and sliding glass doors, leaking plumbing fixtures and pipes, and leaking shower assemblies." "Plaintiffs are informed and believe, and based theron allege, that developer defendants, and each of them, did inspect and market said individual residential units with full knowledge of the causes and effects of defects in their construction of the subject properties, the deficiencies in design, installation and supervision thereof and, in willfull and reckless disregard of defective conditions, causes and results."
Mike Crump and Natalie Crump vs. Lennar Homes of Texas Land and Construction Ltd, *et al.*, Judicial District Court, Williamson Co., TX; Case No: 08-1023-C26	"Prior to the above real estate transaction with Plaintiffs, Defendants had actual knowledge that the soil in the Hutto Parke Subdivision was highly expansive and/or collapsible. Moreover, prior to the above real estate transaction with Plaintiffs, Defendants had actual knowledge that arsenic was present within the soil in the Hutto Parke Subdivision in sufficient amounts to warrant remediation. At no time prior to the execution of the real estate transaction with Plaintiffs did the Defendants disclose to the Plaintiffs that the soil in the Hutto Parke Subdivision was highly expansive or collapsible. To the contrary, the Defendants affirmatively represented to the Plaintiffs that the soil did not have unstable, expansive or collapsible soil."
Johnston et al. vs. Del Webb's Coventry Homes of Nevada *et al.*; Distr. Court, Clark Co., NV; Case No: A581561	"As a direct and proximate result of the breach of warranties (written and oral) by Defendants, and each of them, as herein and above alleged, Plaintiffs suffered damages stemming from, among other items, defectively built roofs, leaking windows, dirt coming through windows, drywall cracking, stucco cracking, stucco staining, water and insect intrusion through foundation slabs, and other poor workmanship."
Shuette vs. Beazer Homes Holdings Corp; Nevada Supreme Court, Case Nos: 4161, 41768	"In April 2000, three homeowners, individually and as proposed class representatives, filed a complaint against Beazer Homes alleging constructional defects to their homes. The homeowners claimed that their houses' foundations and concrete slabs were damaged by expansive soils, a condition in which the soils beneath a house expand when exposed to water and contract when the soil dries. This condition can cause a house's foundation and concrete slab to crack and separate. The homeowners also alleged over 30 additional constructional defects unrelated to the soils condition."
Joan Kirsch *et al.* vs. Del Webb Coventry Homes Inc, *et al.*; Arizona Super., Maricopa Co.; Case No: CV2007-0323536	"To date, 88 homes in the Anthem Arizona community have experienced failures in their underslab unsleeved copper water pipes. The exact same problem occurred in another Del Webb community in Sun City Summerlin, Las Vegas, Nevada." "The Summerlin homeowners began reporting underslab pipe leaks to Del Webb in February 1996." "There are 9,364 single family homes in the Anthem Community. Approximately 3,623 homes have unsleeved copper water pipe." "Homes with unsleeved copper pipe closed escrow between July 9, 1999 and December 2003."

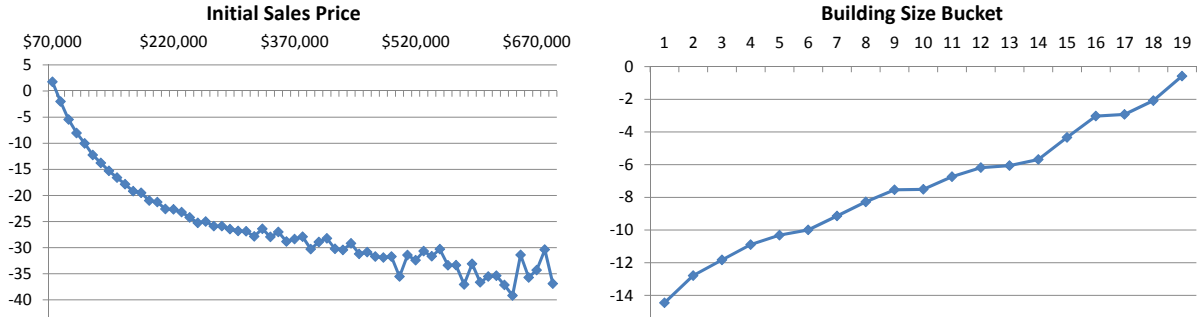
Note: This is a list of construction defect lawsuits compiled by the author, including verbatim statements from plaintiffs' petitions or other court filings related to the case. Note that some of these lawsuits are ongoing, or have been settled without admission of guilt by the defendant.

Table 15: Coefficients on Control Variables

LTV $\in (80\%, 90\%]$	0.275 (0.169)	Single	-0.328*** (0.099)
LTV $\in (90\%, 97\%]$	0.130 (0.151)	Latino	-0.392** (0.149)
LTV $> 97\%$	0.0689 (0.225)	Asian	-0.0483 (0.287)
LTI $\in (1.5, 2]$	-0.409** (0.206)	Median Income $\in (35\text{k}, 50\text{k}]$	3.431*** (1.006)
LTI $\in (2, 2.5]$	-0.675*** (0.220)	Median Income $\in (50\text{k}, 65\text{k}]$	4.729*** (0.916)
LTI $\in (2.5, 3]$	-0.768*** (0.273)	Median Income $\in (65\text{k}, 75\text{k}]$	4.307*** (0.839)
LTI $\in (3, 3.5]$	-0.873*** (0.324)	Median Income $\in (75\text{k}, 100\text{k}]$	5.638*** (0.834)
LTI > 3.5	-1.165*** (0.319)	Median Income $> 100\text{k}$	6.362*** (0.860)
VA	-0.168 (0.205)	Highschool Grad Rate $\in (75, 80]$	3.233** (1.513)
FHA	-0.586*** (0.154)	Highschool Grad Rate $\in (80, 90]$	0.722 (0.971)
Jumbo	-1.472 (1.044)	Highschool Grad Rate $\in (90, 95]$	-0.196 (1.100)
Duration $\in (10, 30)$	-0.560 (0.706)	Highschool Grad Rate > 95	1.323 (1.334)
Duration $== 30$	-0.722 (0.589)	Owner-Occupied	-0.251* (0.144)
Duration > 30	-1.834** (0.725)	Pool	1.785*** (0.126)

Note: This table reports a subset of the coefficients on control variables for the regression corresponding to Column (6) of Table 1.

Figure 10: Coefficients on Control Variables



Note: These graphs plot the coefficients on the respective variables for the regression corresponding to Column (6) of Table 1. For the left panel, the excluded category is houses with a sales price below $\$70,000$, in the right panel the excluded category is the highest 5% of houses in terms of building size.

C.3 Soil Conditions

Table 5 shows that the outperformance of houses financed by the integrated lender was particularly large amongst houses built on expansive soil. In that table the return of the housing collateral is measured over periods (A) and (B) in Figure 2, which the capture return between two armslength sales and between the initial sale and the January 2009 assessment value, respectively. In this section I confirm that these results persist when the return of the housing collateral is measured over period (C), between the second sale and the January 2009 assessment value, and over period (D), between the assessment values in January 2008 and January 2009. The results are presented in Table 16 and are highly consistent with those discussed in Section 5.5. During the ownership of the second owner, amongst houses built on expansive soil, those financed by the integrated lender outperform ex-ante equivalent homes financed by a non-integrated lender by a full percentage point annually. During the year 2008, when prices of recently developed properties fell by an average of 25%, the integrated lender's collateral portfolio outperformed by over one percentage point amongst houses built on expansive soil. Amongst houses not built on expansive soil, those financed by the integrated lender outperformed by less, though the coefficients are still large and significant.

Table 16: Annualized Collateral Return (%) by Soil Type - Periods (C) and (D)

	Second Owner - Period (C)			Year 2008 - Period (D)		
	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	0.232**	0.203**	0.128	0.691***	0.560***	0.270**
	(0.116)	(0.102)	(0.0894)	(0.182)	(0.178)	(0.108)
Expansive Soil	-2.398***	-1.977***	-0.997***	-3.549***	-3.640***	-1.714**
	(0.575)	(0.485)	(0.323)	(1.115)	(1.245)	(0.817)
Integrated Lender × Expansive Soil	0.830***	0.719***	0.147	0.754*	0.936**	0.766***
	(0.257)	(0.231)	(0.153)	(0.425)	(0.411)	(0.289)
Control Variables (See Note)	✓	✓	✓	✓	✓	✓
Developer Fixed Effects	·	✓	·	·	✓	·
Development Fixed Effects	·	·	✓	·	·	✓
R-squared	0.893	0.899	0.946	0.581	0.612	0.796
Mean Dependent Var.	-11.00	-11.00	-11.00	-27.23	-27.23	-27.23
N	16,764	16,764	16,764	66,497	66,497	66,497

Note: The dummy variable "expansive soil" is equal to one for houses built on soil belonging to hydrologic soil group D. Columns (1) - (3) show results corresponding to Table 3. Columns (4) - (6) show results corresponding to Table 4. Standard errors are clustered at the developer level. Each specification controls for county fixed effects, quarter of construction fixed effects, house characteristics, owner characteristics, financing characteristics and census tract demographics. Columns (1) to (3) also include month of resale fixed effects. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

D Robustness Checks

In this appendix I present robustness checks to the empirical analysis. In Appendix D.1, I focus on issues related to the HMDA-deeds merge. In particular, I show that the key results are robust to considering (i) the full set of observations, without requiring a successful HMDA-deeds merge and (ii) only the set of observations for which I observe a unique HMDA-deeds merge. In Appendix D.2 I show that the results are also robust to a different processes for identifying the integrated lender.

D.1 Robustness Check - HMDA-deeds merge

This section presents two robustness checks that focus on issues related to the HMDA-deeds merge described in Appendix B.2. The first robustness check addresses possible concerns with the restriction of the main dataset to those mortgages that can be matched to at least one observation in the HMDA data, which allowed me to control for borrower characteristics in the empirical analysis. This was important because one might have expected selection on observable borrower characteristics which could introduce a spurious correlation between lender identity and collateral return. I next show that restricting the sample to the subset of houses that can be merged to HMDA data does not affect the results. I repeat the key regressions from the analysis above for the full sample of houses, but only control for those borrower characteristics that I can observe in the deeds data (i.e. whether they are married or single and whether the last name is amongst the 1000 most common Asian and Latino names). The regressions have a substantially larger sample size compared to those presented in the main body of the paper, due to the addition of mortgages that cannot be matched to HMDA data.

Table 17 presents robustness checks for the key results from Tables 1, 2, 3, 4, 6 and 7. Columns (1) and (2) compare the annualized return of the housing collateral between two armslength transactions. The magnitude of the coefficient on $IntegratedLender_i$ is very similar with this larger sample and somewhat smaller set of controls: the portfolio of houses financed by the integrated lender outperforms an ex-ante similar portfolio of houses financed by a non-integrated lender by about 50 basis points annually.[71] The outperformance persists when measured over periods (B), (C) and (D), as show in columns (3) - (6). Column (7) shows that for the full sample, the probability of observing a foreclosure within 3 years of granting the mortgage is about one percentage point lower for mortgages financed by the integrated lender. This is similar in magnitude to the effect found for the restricted sample.

Table 18 presents robustness checks for Tables 10 and 11. As before, a non-integrated lender's collateral portfolio underperforms when it competes against an integrated lender. In response the non-integrated lender charges about 10 basis points higher interest rates.

[71] As before, the sequential addition of the control variables does not influence the magnitude of the measured outperformance, and restricting the sample to those repeat sales that were prompted by an exogenous divorce or death of the owner confirms the findings from the full sample. These results are available on request.

Table 17: Robustness Check - No HMDA-Merge

	Return (A)		Return (B)		Return (C)+(D)		Foreclosure	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Integrated Lender	0.443***	0.424***	0.241***	0.143***	0.197*	0.471***	-0.009***	-0.019***
	(0.105)	(0.085)	(0.078)	(0.036)	(0.110)	(0.146)	(0.001)	(0.005)
Specification	Col (5) Table 1	Col (6) Table 1	Col (5) Table 2	Col (6) Table 2	Col (1) Table 3	Col (5) Table 4	Col (7) Table 6	Col (8) Table 7
R-squared	0.888	0.898	0.882	0.932	0.881	0.583		
Mean Dependent Var.	7.890	7.890	-6.903	-6.903	-11.09	-27.22	0.019	0.052
N	39,574	39,574	127,387	127,387	24,785	101,673	110,338	16,831

Note: This table shows robustness checks for key results in the main body of the paper (respective specifications indicated) using all observations without requiring a successful HMDA merge. Control variables are added as indicated, though income and loan-to-income ratio cannot be controlled for in this specification. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table 18: Robustness Check - No HMDA-Merge - "Has Integrated Lender"

	Return Period (A)		Return Period (C)	Mortgage Interest Rate	
	(1)	(2)	(3)	(4)	(5)
Has Integrated Lender	-0.488*	-0.820*	-0.345	0.131***	0.109***
	(0.268)	(0.419)	(0.326)	(0.045)	(0.040)
Specification	Col (1) Table 10	Col (2) Table 10	Col (4) Table 10	Col(1) Table 11	Col (5) Table 11
R-squared	0.868	0.876	0.880	0.545	0.576
Mean Dependent Var.	8.740	8.740	-10.11	6.690	6.690
N	21,369	21,369	13,500	25,645	25,642

Note: This table shows robustness checks for key results in the main body of the paper (respective specifications indicated) using all observations without requiring a successful HMDA merge. Control variables are added as indicated, though income and loan-to-income ratio cannot be controlled for in this specification. In columns (1) - (3), standard errors are clustered at the developer level, in columns (4) - (5) at the lender level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

For a number of observations, a unique merge between deeds data and HMDA data was not possible, since there is more than one HMDA entry with the same value of the merge variables. As described in Appendix B.2, I match each such deed randomly to one of the possible HDMA records. In a second robustness check, I repeat the key regressions from the paper using just the sample of houses that could be matched *uniquely* to the HMDA data. Unlike the first robustness check, this reduces the sample relative to the main body of the paper. Table 20 shows results corresponding to Tables 1, 2, 3, 4, 6 and 7. As before, the integrated lender's collateral portfolio outperforms by about 50 basis points annually and foreclosure probabilities are about one percentage point lower. This result suggests that the random assignment of ambiguous merges did not introduce a bias into the results. Table 20 shows that the regression in Tables 10 and 11 are similarly robust to only considering the set of observations that can be uniquely merged to the HMDA data.

Table 19: Robustness Check - Unique HMDA-Merge

	Return (A)		Return (B)		Return (C)+(D)		Foreclosure	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Integrated Lender	0.537***	0.495***	0.239***	0.158***	0.437***	0.518***	-0.011***	-0.027***
	(0.143)	(0.129)	(0.0826)	(0.0527)	(0.137)	(0.166)	(0.002)	(0.007)
Specification	Col (5) Table 1	Col (6) Table 1	Col (5) Table 2	Col (6) Table 2	Col (1) Table 3	Col (5) Table 4	Col (7) Table 6	Col (8) Table 7
R-squared	0.887	0.896	0.890	0.934	0.885	0.595		
Mean Dependent Var.	8.253	8.253	-7.117	-7.117	-11.22	-27.53	0.024	0.055
N	15,631	15,631	51,434	51,434	9,704	41,116	35,751	5,685

Note: This table shows robustness checks for a number of results in the main body of the paper (respective specifications indicated), using only those observations with a unique HMDA merge. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table 20: Robustness Check - Unique HMDA-Merge - "Has Integrated Lender"

	Return Period (A)		Return Period (C)	Mortgage Interest Rate	
	(1)	(2)	(3)	(4)	(5)
Has Integrated Lender	-0.723**	-0.803*	-0.735**	0.119**	0.093**
	(0.300)	(0.472)	(0.348)	(0.048)	(0.043)
Specification	Col (1) Table 10	Col (2) Table 10	Col (4) Table 10	Col(1) Table 11	Col (5) Table 11
R-squared	0.870	0.877	0.884	0.556	0.588
Mean Dependent Var.	9.010	9.010	-10.36	6.664	6.665
N	10,069	10,069	6,316	13,002	12,999

Note: This table shows robustness checks for key results in the main body of the paper (respective specifications indicated), using only those observations with a unique HMDA merge. In columns (1) - (3), standard errors are clustered at the developer level, in columns (4) - (5) at the lender level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

D.2 Discussion and Robustness Check - Definition of "Integrated Lender"

Appendix B.3 describes the process of identifying integrated lenders in the primary dataset. For most developers, the integrated lender is identified through verifying joint ownership between the developer and the integrated lender using OneSource North American Business Browser and SEC filings. In addition, any lender that grants more than 50% of the mortgages in a specific development is also identified as an integrated lender. In this appendix I show that my results are robust to only identifying integrated lenders through joint ownereship with the developer.

Table 21 presents robustness checks for the key results from Tables 1, 2, 3, 4, 6 and 7. The sample size falls by about 20%, dropping those developments for which I previously identified the integrated lender via its market share. Columns (1) and (2) show that the outperformance of the integrated lender's collateral portfolio remains at about 50 basis points. The other results are similarly stable with respect to the alternative identification of integrated lenders. Table 22

presents robustness checks for Tables 10 and 11, which test differences in the return and behavior of non-integrated lenders when competing against an integrated lender. The coefficients are likely biased downwards, since a number of developments with a lender that is likely to be integrated (since it has over 50% market share) are assigned to not have an integrated lender. As before, a non-integrated lender's collateral portfolio underperforms when it competes against an integrated lender. In response, the non-integrated lender charges an average of about 8 basis points higher interest rates.

Table 21: Robustness Check - Alternative "Integrated Lender" Definition

	Return (A)		Return (B)		Return (C)+(D)		Foreclosure	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Integrated Lender	0.547***	0.508***	0.363***	0.169***	0.270**	0.677***	-0.011***	-0.011*
	(0.175)	(0.160)	(0.109)	(0.0511)	(0.119)	(0.222)	(0.002)	(0.006)
Specification	Col (5) Table 1	Col (6) Table 1	Col (5) Table 2	Col (6) Table 2	Col (1) Table 3	Col (5) Table 4	Col (7) Table 6	Col (8) Table 7
R-squared	0.889	0.898	0.890	0.939	0.891	0.603		
Mean Dependent Var.	8.802	8.802	-6.307	-6.307	-11.30	-27.96	0.023	0.050
N	20,639	20,639	64,417	64,417	13,203	50,511	25,953	4,343

Note: This table shows robustness checks key results in the main body of the paper (respective specifications indicated), using only integrated lenders identified via joint ownership with the developer. Standard errors are clustered at the developer level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

Table 22: Robustness Check - Alternative "Integrated Lender" Definition - "Has Integrated Lender"

	Return Period (A)		Return Period (C)	Mortgage Interest Rate	
	(1)	(2)	(3)	(4)	(5)
Has Integrated Lender	-0.568*	-0.552	-0.761**	0.078*	0.077*
	(0.312)	(0.440)	(0.305)	(0.047)	(0.043)
Specification	Col (1) Table 10	Col (2) Table 10	Col (4) Table 10	Col(1) Table 11	Col (5) Table 11
R-squared	0.869	0.877	0.890	0.542	0.583
Mean Dependent Var.	9.233	9.233	-10.02	6.688	6.629
N	12,483	12,483	7,957	15,342	15,339

Note: This table shows robustness checks for key results in the main body of the paper (respective specifications indicated), using only integrated lenders identified via joint ownership with the developer. In columns (1) - (3), standard errors are clustered at the developer level, in columns (4) - (5) at the lender level. Significance Levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

References

Agarwal, S. and R. Hauswald (2010): "Distance and private information in lending," *Review of Financial Studies*, 23, 2757–2788.

Altonji, J., T. Elder, and C. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 113, 151–184.

Alvarez, S. (2011): "Testimony before the Subcommittee on Capital Markets," *U.S. House of Representatives*.

Arizona Capitol Times (2009): "Maricopa County rejects most tax appeals," *Published on October 2, 2009*.

Avery, R., G. Canner, and R. Cook (2005): "New information reported under HMDA and its application in fair lending enforcement," *Federal Reserve Bulletin*, 91, 344–394.

Bajari, P., C. Chu, and M. Park (2008): "An empirical model of subprime mortgage default from 2000 to 2007," *NBER Working Paper*, 14625.

Barron, J., B. Chong, and M. Staten (2008): "Emergence of captive finance companies and risk segmentation in loan markets: theory and evidence," *Journal of Money, Credit and Banking*, 40, 173–192.

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2011): "A dynamic model of demand for houses and neighborhoods," *NBER Working Paper*, 17250.

Ben-David, I. (2011): "Financial constraints and inflated home prices during the real estate boom," *American Economic Journal: Applied Economics*, 3, 55–87.

Berger, A. and G. Udell (1995): "Relationship lending and lines of credit in small firm finance," *Journal of Business*, 68, 351–381.

Black, H., T. Boehm, and R. DeGennaro (2003): "Is there discrimination in mortgage pricing? The case of overages," *Journal of Banking & Finance*, 27, 1139–1165.

Bloomberg (2011): "Homebuilder shares undermined by creeping cost of construction boom flaws," *Published on February 10, 2011*.

Broecker, T. (1990): "Credit-worthiness tests and interbank competition," *Econometrica*, 58, 429–452.

Bulow, J. (1982): "Durable-goods monopolists," *Journal of Political Economy*, 90, 314–332.

Cameron, A., J. Gelbach, and D. Miller (2011): "Robust inference with multiway clustering," *Journal of Business and Economic Statistics*, 29, 238–249.

Campbell, J., S. Giglio, and P. Pathak (2011): "Forced sales and housing prices," *American Economic Review*, 101, 2108–2131.

Caplin, A., S. Chopra, J. Leahy, Y. LeCun, and T. Thampy (2008): "Machine learning and the spatial structure of house prices and housing returns," *Working Paper*.

Coase, R. (1972): "Durability and monopoly," *Journal of Law & Economics*, 15, 143–49.

Criterium Engineers (2003): "If it's new, is it good?" http://criteriumhomeinspection.com/new-construction-is-it-good,.

Degryse, H. and S. Ongena (2005): "Distance, lending relationships, and competition," *Journal of Finance*, 60, 231–266.

Degryse, H. and P. Van Cayseele (2000): "Relationship lending within a bank-based system: Evidence from European small business data," *Journal of Financial Intermediation*, 9, 90–109.

Deng, Y., J. Quigley, and R. Order (2000): "Mortgage terminations, heterogeneity and the exercise of mortgage options," *Econometrica*, 68, 275–307.

D'Silva, A. and B. Gordon (2008): "Hedges in the warehouse: the banks get trimmed," *Chicago Fed Letter*.

Einav, L., M. Jenkins, and J. Levin (2011): "Contract pricing in consumer credit markets," *Working Paper*.

Elul, R. (2011): "Securitization and mortgage default," *Federal Reserve Bank of Philadelphia Working Paper*, 09–21.

Engel, K. and P. McCoy (2006): "Turning a blind eye: Wall Street finance of predatory lending," *Fordham Law Review*, 75, 2039.

Engelbrecht-Wiggans, R., P. Milgrom, and R. Weber (1983): "Competitive bidding and proprietary information," *Journal of Mathematical Economics*, 11, 161–169.

Esponda, I. (2008): "Behavioral equilibrium in economies with adverse selection," *American Economic Review*, 98, 1269–1291.

Federal Trade Commission (1979): "Federal Trade Commission Consent Order," 93 F.T.C 235 - Docket No. C-2954.

Fishman, M. and J. Parker (2010): "Valuation, adverse selection, and market collapses," *Working Paper*.

Garmaise, M. and T. Moskowitz (2004): "Confronting information asymmetries: Evidence from real estate markets," *Review of Financial Studies*, 17, 405–437.

Gartenberg, C. (2011): "Tempted by scope? Homebuilder mortgage affiliates, lending quality and the housing crisis," *Working Paper*.

Geanakoplos, J. (2009): "The leverage cycle," *NBER Macroeconomics Annual*.

Gorton, G. (2008): "The panic of 2007," *NBER Working Paper*, 14358.

Gorton, G. and G. Ordonez (2011): "Collateral crises," *Working Paper*.

Gorton, G. and G. Pennacchi (1990): "Financial intermediaries and liquidity creation," *Journal of Finance*, 45, 49–71.

Gorton, G. and N. Souleles (2007): "Special purpose vehicles and securitization," in *The Risks of Financial Institutions*, ed. by M. Carey and R. Stulz, University of Chicago Press.

Grant, R. (1996): "Toward a knowledge-based theory of the firm," *Strategic management journal*, 17, 109–122.

Greene, W. (1998): "Sample selection in credit-scoring models," *Japan and the World Economy*, 10, 299–316.

Greenwald, B. (1986): "Adverse selection in the labour market," *Review of Economic Studies*, 53, 325.

Guiso, L., P. Sapienza, and L. Zingales (2011): "The determinants of attitudes towards strategic default on mortgages," *Journal of Finance*, forthcoming.

Gyourko, J., C. Mayer, and T. Sinai (2006): "Superstar cities," *NBER Working Paper*, 12355.

Hauswald, R. and R. Marquez (2006): "Competition and strategic information acquisition in credit markets," *Review of Financial Studies*, 19, 967–1000.

Hendricks, K. and R. Porter (1988): "An empirical study of an auction with asymmetric information," *American Economic Review*, 78, 865–883.

Hoerner, J. and J. Jamison (2008): "Sequential common-value auctions with asymmetrically informed bidders," *Review of Economic Studies*, 75, 475–498.

Houston, S., H. Dye, C. Zapata, K. Walsh, W. Houston, et al. (2011): "Study of expansive soils and residential foundations on expansive soils in Arizona," *Journal of Performance of Constructed Facilities*, 25, 31–45.

Ioannidou, V. and S. Ongena (2010): "Time for a change: Loan conditions and bank behavior when firms switch banks," *Journal of Finance*, 65, 1847–1877.

Jappelli, T. and M. Pagano (2002): "Information sharing, lending and defaults: Cross-country evidence," *Journal of Banking & Finance*, 26, 2017–2045.

Jiang, W., A. Nelson, and E. Vytlacil (2010): "Securitization and loan performance: A contrast of ex ante and ex post relations in the mortgage market," *Working Paper*.

Keys, B., T. Mukherjee, A. Seru, and V. Vig (2010): "Did securitization lead to lax screening? Evidence from subprime loans," *Quarterly Journal of Economics*, 125, 307–362.

Landvoigt, T., M. Piazzesi, and M. Schneider (2011): "The housing market(s) of San Diego," *Working Paper*.

Levitt, S. and C. Syverson (2008): "Market distortions when agents are better informed: The value of information in real estate transactions," *Review of Economics and Statistics*, 90, 599–611.

Long, M., I. Malitz, and S. Ravid (1993): "Trade credit, quality guarantees, and product marketability," *Financial Management*, 22, 117–127.

Michaud, S. (1996): "Rating agency perspective on asset-backed securitization," in *Asset-backed securities*, ed. by A. Bhattacharya and F. Fabozzi, Wiley, 269–276.

Milgrom, P. and R. Weber (1982): "The value of information in a sealed-bid auction," *Journal of Mathematical Economics*, 10, 105–114.

Petersen, M. (2009): "Estimating standard errors in finance panel data sets: Comparing approaches," *Review of Financial Studies*, 22, 435.

Petersen, M. and R. Rajan (1994): "The benefits of lending relationships: Evidence from small business data," *Journal of Finance*, 49, 3–37.

——— (2002): "Does distance still matter? The information revolution in small business lending," *Journal of Finance*, 57, 2533–2570.

Phoenix New Times (2006): "Cracked Houses - Homes all over Arizona are falling apart," *Published on March 16, 2006*.

Pierce, L. (2011): "Organizational structure and the limits of knowledge exploitation: Evidence from consumer automobile leasing," *Management Science*, forthcoming.

Povel, P. and R. Singh (2010): "Stapled finance," *Journal of Finance*, 65, 927–953.

Puppala, A. and A. Cerato (2009): "Heave distress problems in chemically-treated sulfate-laden materials," *Geo-Strata*, 10, 28–30.

Rendon-Herrero, O. (2011): "Special issue on construction on expansive soils," *Journal of Performance of Constructed Facilities*, 25, 2.

Rosen, R. (2007): "The role of securitization in mortgage lending," *Chicago Fed Letter*.

Rubinstein, A. (1991): "Comments on the interpretation of game theory," *Econometrica*, 909–924.

Sharpe, S. (1990): "Asymmetric information, bank lending and implicit contracts: A stylized model of customer relationships," *Journal of Finance*, 45, 1069–1087.

Smith, J. (1987): "Trade credit and informational asymmetry," *Journal of Finance*, 42, 863–872.

Standard & Poor's (2009): "S&P/CASE-SHILLER Home Price Indices - Index Methodoloy," .

Stauber Associates (2011): "Recent trends in construction defect litigation," http://www.stauberassociates.com/products.asp?cat=6&hierarchy=0,.

Superior Court of Arizona (2007): "Del Webb Home Construction, Inc v. Arizona State Registrar of Contractors/Doe," *LC2006-000577-001 DT*.

The Arizona Republic (2001): "Construction Quality Varies Widely," *Published on November 20, 2001*.

——— (2009): "Appeals of property valuations soar in Maricopa County," *Published on November 10, 2009*.

United States Soil Conservation Service (1985): *National Engineering Handbook: Section 4: Hydrology*, Nat. Engineering Publ. from the U.S. Soil Conservation Service.

von Thadden, E. (2004): "Asymmetric information, bank lending and implicit contracts: The winner's curse," *Finance Research Letters*, 1, 11–23.

Wong, S., C. Yiu, and K. Chau (2011): "Liquidity and information asymmetry in the real estate market," *Journal of Real Estate Finance and Economics*, 1–14.

Woodward, S. and R. Hall (2010): "Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence," *NBER Working Paper*, 16007.

Word, D., C. Coleman, R. Nunziata, and R. Kominski (2008): "Demographic aspects of surnames from census 2000," http://www.census.gov/genealogy/www/surnames.pdf,.