The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXI, NO. 3 • JUNE 2016

Asymmetric Information about Collateral Values

JOHANNES STROEBEL*

ABSTRACT

I empirically analyze credit market outcomes when competing lenders are differentially informed about the expected return from making a loan. I study the residential mortgage market, where property developers often cooperate with vertically integrated mortgage lenders to offer financing to buyers of new homes. I show that these integrated lenders have superior information about the construction quality of individual homes and exploit this information to lend against higher quality collateral, decreasing foreclosures by up to 40%. To compensate for this adverse selection on collateral quality, nonintegrated lenders charge higher interest rates when competing against a better-informed integrated lender.

What is the impact on equilibrium credit market outcomes when competing lenders are differentially informed about the expected return from making a loan? I analyze this question by studying lender competition in the residential mortgage market. Empirical studies of the impact of asymmetric information in credit markets usually face the challenge of identifying the relative information of the different parties (Petersen and Rajan (1994), Ausubel (1999), Karlan and Zinman (2009)). I address this challenge by focusing on the competition between lenders to originate mortgages used to purchase newly developed properties. In this market, property developers regularly provide home buyers with financing offers through vertically integrated mortgage lenders (Agarwal et al. (2011), Gartenberg (2011)). These integrated lenders have potentially better information than nonintegrated lenders about two factors that affect the expected return from making a mortgage: the value of the house used to collateralize the mortgage, and characteristics of the borrower.[1] For example,

*Stroebel is at the New York University Stern School of Business. I am indebted to Caroline Hoxby, Monika Piazzesi, Martin Schneider, and John Taylor for their encouragement and guidance. I thank the Editor, Michael Roberts, and two anonymous referees for constructive comments and helpful feedback. Seminar participants at Stanford, Kellogg, Wharton, Princeton, MIT, Harvard Business School, UCLA Anderson, Chicago Booth, MIT Sloan, LSE, LBS, Michigan, Berkeley Haas, NYU Stern, Stanford GSB, Cornell, Bonn, Oxford, EUI, and the Chicago Fed as well as conference participants and discussants at the NBER Summer Institute, Wisconsin HULM, and the Philadelphia Fed Workshop on Consumer Credit and Payments provided insightful comments. I thank Trulia, Buildfax, Sumit Agarwal, and Chenxi Luo for providing data. Financial support through SIEPR and the Hoover Institution is gratefully acknowledged. I thank Miguel de Faria e Castro for excellent research assistance.

[1] The existing literature on information asymmetries in mortgage lending focuses on information about characteristics of borrowers, such as their income prospects (e.g., Keys et al. (2010), Elul

DOI: 10.1111/jofi.12288

an integrated lender might have access to the developer's property-level information about aspects of construction quality that are difficult for buyers and nonintegrated lenders to observe. Perhaps surprisingly, I find that such asymmetric information about collateral quality is a significant source of adverse selection in this market. In addition to testing for the presence of asymmetric information and uncovering its sources, I also quantify its impact on the cost of mortgages, which I find to be significant.

I first develop a set of empirical predictions in an environment where an integrated mortgage lender obtains an informative signal about housing collateral quality, and competes with nonintegrated lenders that only know average collateral quality. The integrated lender conditions its financing offer on its superior information, and thereby subjects nonintegrated lenders to adverse selection. In equilibrium, the integrated lender lends against above-average-quality collateral. As house quality is revealed over time, those homes financed by the integrated lender should thus outperform ex ante observationally similar homes financed by nonintegrated lenders. This effect will be bigger when the integrated lender's signal about collateral quality is more precise. Nonintegrated lenders need to charge higher interest rates to break even than if they were competing only against equally informed lenders. Interest rates rise by more for borrowers whose repayment is more sensitive to changes in collateral values, for example, because they make a smaller downpayment.

To test these predictions, I construct a data set of all housing transactions and associated mortgages in Arizona between 2000 and 2011. About 85% of new homes are in developments with an active integrated lender, and, when present, the average market share of these integrated lenders is about 73%. These data allow me to track the capital gain of properties following their initial sale, controlling for property and buyer characteristics. I find that, in developments with an integrated lender, those houses financed by the integrated lender do indeed outperform ex ante observationally similar houses in the same development financed by nonintegrated lenders by an average of 40 bps annually. Houses financed by nonintegrated lenders perform particularly badly when the borrower's mortgage application was rejected by the integrated lender. As a result, mortgages financed by integrated lenders are over 40% less likely to enter into foreclosure.

I address three main challenges to this identification of asymmetric information about collateral values. First, one might be concerned that the results can be explained by unobserved differences in borrower quality. If integrated lenders attract higher quality borrowers, who subsequently invest more in the property, this would lead to higher observed capital gains for the integrated lender's collateral portfolio. Second, some properties might be worse on some dimension, such as the view, that is fully observable to all market participants

(2011), Agarwal, Chang, and Yavas (2012)). The expected return from making a mortgage, however, depends both on the value of the housing collateral and the borrower's ability to make interest payments. Due to the illiquid and heterogeneous nature of housing assets, it is likely that there is also asymmetric information about collateral values.

but unobservable to the econometrician. If nonintegrated lenders were more willing to finance such properties that subsequently underperform, we would expect the integrated lender's collateral portfolio to experience higher capital gains, even in the absence of asymmetric information. Third, one might be worried that the results are explained by initial price bundling of the mortgage and the house, since any discounts on the house given to customers of the integrated lender would be observationally equivalent to true collateral outperformance when the house is subsequently sold for its true value.

I show that confounding unobserved borrower characteristics cannot explain my findings. First, I demonstrate that borrowers of integrated lenders are, if anything, worse on observable characteristics that affect the subsequent appreciation of the house, making it unlikely that they are better on confounding unobservable characteristics. Second, I compare the return and foreclosure probability for the ownership duration of the *second* owner of the house. The relative outperformance of houses *initially* financed by the integrated lender remains. This result suggests that the outperformance is explained by asymmetric information about the housing collateral, not the borrower, since the identity of a possible second owner of the house is not known to any lender at the time the mortgage is granted to the first owner. Third, I use data from building permits to show that there is no evidence of differential maintenance or investment between borrowers from integrated and nonintegrated lenders. Fourth, I show that the annual outperformance of the integrated lender's collateral portfolio is larger, at about 100 bps, among houses built on "expansive soil," a high clay content soil that makes housing returns more sensitive to unobservable aspects of construction quality such as the care with which the foundation was poured. Soil quality thus provides plausibly exogenous variation in the importance of information about collateral quality, but not about borrower characteristics, and therefore helps to identify the source of asymmetric information between competing lenders.

I also show that the results are unlikely to be driven by differences in the properties that are observable to all market participants but not the econometrician. First, after controlling for observable characteristics, there are no differences in the initial price paid for properties financed by integrated and nonintegrated lenders. This suggests that houses financed by different lenders do not systematically differ on characteristics observable to market participants but unobserved by the econometrician, since those characteristics would be reflected in the initial price. Second, if nonintegrated lenders were more willing to finance homes that everybody knew to be worse, we would expect them to charge higher interest rates for these homes. However, the results persist even after controlling for the interest rate charged on the mortgage. Third, I show that, while rejections by integrated lenders predict lower collateral returns, a rejection by nonintegrated lenders does not. This suggests that, while integrated lenders have information unobservable to the econometrician (e.g., about construction quality), nonintegrated lenders do not.

Finally, I show that price bundling, where borrowers of the integrated lender obtain a discount on the house, cannot explain my findings. First, 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. Second, I analyze the textual description of houses in property listings when these houses are resold by their initial owners. I find that houses initially financed by integrated lenders are significantly less likely to contain descriptions of damage to the property, suggesting that the integrated lender's outperformance can be best explained by differential depreciation rates of houses, rather than by differences in initial price. Third, I show that the outperformance of the integrated lender's collateral portfolio is larger for mortgages with a large downpayment, which is a prediction of a model with asymmetric information but not one of price bundling.

In an additional set of analyses, I also consider the cost to borrowers in terms of higher interest rates that result from this asymmetric information. I find that nonintegrated lenders charge an average interest rate premium of 10 bps annually for otherwise similar mortgages when competing against an integrated lender. This higher interest rate compensates nonintegrated lenders for the adverse selection resulting from the presence of an integrated lender. The magnitude of this increase is economically significant, and equivalent to the interest rate increase when reducing the downpayment from 25% to 20%. The interest rate increase is even larger, at 23 bps, 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. The interest rate increase is also larger for mortgages with a low downpayment, rising to almost 50 bps annually for mortgages with a downpayment of less than 3%. For those mortgages, the default probability is more sensitive to changes in collateral values. Nonintegrated lenders thus need to charge higher interest rates to compensate for the adverse selection when competing with an integrated lender.

Mortgage lending in new developments is a large and highly cyclical component of total household credit. Agarwal et al. (2011) show that in 2006 nearly a third of all families buying new homes obtained a mortgage from an integrated lender, and even more would have received financing offers from such lenders. Understanding the sources and magnitude of asymmetric information in this market is thus important, because such asymmetric information has the potential to disrupt lending markets, particularly during periods of falling house prices (see Fishman and Parker (2015) and Gorton and Ordonez (2014)).

This paper also provides insights into the behavior of mortgage lenders in the precrisis period between 2000 and 2007. It has sometimes been argued that due to a lack of "skin in the game" generated by securitization, many banks no longer had incentives to distinguish between borrowers and collateral of differential quality, reducing the average quality of originated mortgages (James (2010)). In contrast, the evidence presented in this paper is highly consistent with both integrated and nonintegrated lenders pricing cross-sectional differences in collateral quality in a sophisticated manner.

In addition to being interesting and important in its own right, the market for financing newly developed homes also provides an empirically rich environment that can help us understand credit markets with asymmetrically informed lenders more generally. For example, examining this market can contribute to our understanding of the market power that relationship lenders or physically proximate lenders might acquire by learning about hard-to-observe characteristics of particular borrowers or their collateral (Petersen and Rajan (2002), Berger and Udell (1995), Degryse and van Cayseele (2000), Degryse and Ongena (2005), Agarwal and Hauswald (2010)). Another setting with a similar information structure involves the practice of sell-side advisers offering "stapled financing" packages to buyers in M&A transactions (Povel and Singh (2010)). Since the bank providing the stapled financing advises on the sale of the asset, it could have superior information relative to other banks about the quality of the loan collateral. Similarly, firms providing trade credit to buyers of their product have superior information relative to other potential lenders about the quality of the underlying loan collateral (Smith (1987), Long, Malitz, and Ravid (1993)).

This paper also contributes to the literature on information asymmetries in real estate markets (Levitt and Syverson (2008), Ben-David (2011), Kurlat and Stroebel (2015)). From a policy perspective, the identification of collateral values as a key source of asymmetric information in mortgage lending can help develop policy proposals to improve the functioning of this market. For example, my findings suggest that better credit scoring technology and the more extensive sharing of borrower information through credit bureaus will not address all forms of asymmetric information. In addition, while the results of Keys et al. (2010) suggest that granting only full documentation mortgages might reduce some of the information asymmetry about borrower characteristics, policies that address information asymmetry 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.

The remainder of the paper is organized as follows. Section I outlines the possible sources of asymmetric information between different mortgage lenders. Section II discusses the empirical predictions of a model with asymmetric information about collateral values between competing lenders. Section III describes the data, while Section IV presents the baseline empirical results. Section V addresses a number of identification challenges; Section VI presents results on the effect of asymmetric information on interest rates. Section VII discusses the welfare implications of the paper's findings. Section VIII concludes.

I. Possible Sources of Asymmetric Information

Before describing the empirical analysis, I discuss evidence for possible sources of superior information of the integrated lender. In principle, this information could relate to either the quality of the housing collateral or the characteristics of the borrowers.

One component of the integrated lender's superior information about collateral quality concerns aspects of the construction quality of the house that are not observable to buyers and nonintegrated lenders at the time of purchase. *The Arizona Republic* (2001) describes shortcuts in the construction process regularly taken by builders in Arizona, which can generate differences in construction quality:

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. Builders sometimes add excess water to the cement mix used for the foundation, which makes it easier and faster to spread but more subject to cracking later.
3. Stucco is often applied too thinly, which can lead to subsequent cracking.

Recent lawsuits provide mounting evidence that home builders were regularly aware of existing construction defects when selling a home. For example, a 2007 whistleblower lawsuit "alleges KB [Home] executives knew about, but concealed, 'life-threatening structural defects.' The suit, filed by a former human resources director, says he was fired when he refused to play along. The whistleblower, Ruben O'Neil, said an engineering report showed KB used 'substandard/inadequate materials in the construction of the rear-load-bearing wall of 50 to 60 townhomes.' O'Neil said sales agents were told to proceed with closing on homes anyway." (*Tampa Bay Tribune*, (2012)). *The Arizona Republic* (2001) discusses the lack of skilled workers to perform delicate construction tasks as a key factor in explaining low construction quality. Since a developer is aware of the skill of the crews working on different houses, this suggests it has the ability to identify low-construction-quality homes within a development.

A significant proportion of construction-related defects in Arizona involve insufficient care taken when building on expansive soil. Expansive soils have a high content of clays that 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 constructed can result in excessive foundation movements and the cracking of slabs and walls. *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." Since the developer knows whether the respective subcontractor is sufficiently skilled and experienced to conduct the more delicate procedures, one would expect that adverse selection is particularly important among houses built on expansive soil.

It is hard to empirically determine the precise channel through which such information is obtained by an integrated lender. However, given the significant resources spent by lenders on property appraisers and inspectors to acquire information about the quality of a house prior to making a lending decision, it seems natural to expect that they acquire additional relevant information from within their own organization. This is particularly likely for developers that co-locate regional construction teams and the integrated lender's loan officers, who often work on-site (and in adjacent offices) at each subdivision.

In addition to superior information about collateral quality, it is possible that, in the process of guiding borrowers through the house purchase process, an integrated lender also obtains superior information about borrower characteristics that might affect the return of the housing collateral. For example, the developer might learn about the buyer's propensity to maintain the property. However, in the empirical analysis I show that the outperformance of the collateral portfolio of the integrated lender can be best explained by its superior information about initial collateral quality, not borrower characteristics.

II. Empirical Predictions

In this section, I discuss the intuition for the empirical predictions tested in this paper. The Internet Appendix presents a theoretical model that formalizes this intuition.[2]

Consider a development with many newly constructed homes. Consistent with the evidence in the previous section, some of these homes have a construction defect such as a poorly laid foundation. Every period there is some probability that construction defects are revealed, and the affected homes will decline in value. Only the developer knows which houses have a construction defect. Homes look identical to all other market participants, and consequently sell at the same price.

Households are interested in buying and mortgage-financing a newly developed home, and simultaneously apply to two mortgage lenders: the integrated lender and a nonintegrated lender. Households accept the lowest-interest mortgage offer they receive. When homes decline in value due to a construction defect, some borrowers will default on the mortgage, leaving the lender with a loss due to foreclosure.[3] When lenders set the interest rate, they therefore trade off the revenue from interest charges with the loss in the event a construction defect is revealed. Since all homes look identical to the nonintegrated lender, their interest rate offers do not depend on the specific home chosen, but rather on the average rate of construction defects in a development. Integrated lenders, in contrast, condition their interest rate offer on their superior information, offering higher interest rates on mortgages to finance a house with

[2] The Internet Appendix may be found in the online version of this article.

[3] Even if mortgages are securitized, originating lenders retain exposure to their subsequent performance. First, in private label securitization, the issuer often retains tranches of varying seniority. In addition, securitizers are usually exposed through credit enhancements such as overcollateralization and excess spreads. Finally, the sale of mortgages to investors generally includes explicit recourse clauses that require the lender to take back loans if borrowers default.

construction defects.[4] Since integrated lenders generally require high interest rates for homes with construction defects, borrowers are more likely to accept the offer from a nonintegrated lender on those homes. Over time, as initial construction quality is revealed, houses financed by the integrated lender thus appreciate more and have fewer foreclosures than those financed by nonintegrated lenders.

PREDICTION 1: *The average ex post return of houses financed by integrated lenders is higher than the return of ex ante similar (conditional on a nonintegrated lender's information set) homes financed by nonintegrated lenders; homes financed by the integrated lender also experience fewer foreclosures.*

This setting contains two assumptions. First, borrowers do not infer the true quality of their home from the integrated lender's offer. This "bounded rationality" assumption is relatively weak given the complexity of the required signal extraction process. Second, nonintegrated lenders do not observe and learn from the integrated lender's offer, consistent with evidence in Woodward and Hall (2012) that borrowers do not generally engage in sequential mortgage shopping.

While nonintegrated lenders cannot identify homes with a construction defect, they are aware that adverse selection leads them to lend against below-average collateral quality in developments with an integrated lender. To compensate for this adverse selection, they charge higher interest rates when competing with integrated lenders.

PREDICTION 2: *Nonintegrated lenders charge higher interest rates when competing against an integrated lender relative to when competing only against equally informed lenders.*

Homes differ on the (fully observable) type of soil on which they are built. For homes built on expansive soil (see Section I), construction defects are more likely to lead to a loss in value; hence, for those homes, the integrated lender's signal about future value is more precise. Consequently, their equilibrium collateral portfolio is of particularly high quality relative to that of the nonintegrated lender. As a result of the increased adverse selection, the nonintegrated lender's interest rate compensation for competing with an integrated lender is higher for homes built on expansive soil.

PREDICTION 1(A): *Among homes built on expansive soil, the integrated lender's information about future returns is more precise, and the average ex post outperformance of the homes financed by the integrated lender is larger.*

PREDICTION 2(A): *Among homes built on expansive soil, the integrated lender's information about future returns is more precise, and the increase in the interest rate charged by nonintegrated lenders when competing against an integrated lender is larger.*

[4] The actual equilibrium interest rate offer strategies, which are discussed in the Internet Appendix, involve playing mixed strategies.

The loss to lenders from a decline in collateral value depends on the downpayment of the borrower. For mortgages with a large downpayment (low loan-to-value (LTV) ratio), even owners of homes with construction defects will not default when the defect is revealed. Hence, the return to lending is less affected by collateral quality, and nonintegrated lenders adjust their interest rates less to compensate for adverse selection. Since integrated lenders continue to fully exploit their information, they manage to assemble a particularly attractive collateral portfolio. On the other hand, when the downpayment is very small, small changes in the value of the collateral precipitate defaults and foreclosures. For those mortgages, nonintegrated lenders increase their interest rates a lot, improving their average collateral quality at the cost of making fewer mortgages.

PREDICTION 1(B): *For mortgages with a high LTV ratio, mortgage default is more sensitive to changes in collateral values, and the ex post outperformance of houses financed by the integrated lender is smaller.*

PREDICTION 2(B): *For mortgages with a high LTV ratio, mortgage default is more sensitive to changes in collateral values, and the increase in the interest rate charged by nonintegrated lenders when competing against an integrated lender is particularly large.*

III. Data Description

To conduct the empirical analysis, I combine a number of data sets. The first data set contains the universe of ownership-changing deeds in Arizona between 2000 and 2011. The property to which the deeds relate is uniquely identified via the Assessor Parcel Number (APN). The variables in this data set include property address, contract date, transaction price, type of deed (e.g., Intrafamily Transfer Deed, Warranty Deed, Foreclosure Deed), type of property (e.g., Apartment, Single-Family Residence), the name of the buyer and seller, and a classification of the buyer and seller (e.g., Husband and Wife, Company). This data set also reports the amount and duration of the mortgage, and the identity of the mortgage lender. For mortgages with a variable interest rate, I also observe the initial rate.

The second data set contains the universe of tax assessment records for 2010. Properties are again identified via their APN. This data set includes information on property characteristics such as construction year, owner-occupancy status, lot size, building size, and the number of bedrooms and bathrooms. The tax assessment records also include an estimate of the market value of each property for January 2009. To supplement this information on property characteristics, I use 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.[5] Soil expansiveness has significant geographic variation, often within developments, as seen in the Internet Appendix.

The third data set is the Home Mortgage Disclosure Act's (HMDA) Loan Application Registry, which provides details on every formal mortgage application in major Metropolitan Statistical Areas. This data set includes information on the census tract of the house, lender identity, application outcome, loan amount, property type, and the applicant's income, sex, and race. It also records whether the mortgage was sold or securitized within the same calendar year. I merge this data set with the deeds data, as described in the Internet Appendix. For borrowers in developments with an integrated lender, I obtain additional information from Acxiom Inc., a direct marketing firm. I observe the borrower's age, the number of children in the household, whether the borrower is a single parent, and whether the household owns a cat or a dog.

I focus on the state of Arizona, which was at the center of the recent boom-bust cycle, and which is an interesting region to study due to data quality and availability.[6,7] In the Internet Appendix, I describe the process of cleaning and merging the data, as well as the identification of integrated lenders. The resulting data set contains information on 102,818 single-family residences sold by developers in the 2000 to 2007 period that I can match to assessment records and HMDA data. Of these, 87,601 were in developments with an integrated lender, where a development is defined as all properties sold by a unique developer in a particular census tract.

Table I presents summary statistics for all control variables used in the regressions. The first three columns show the mean, standard deviation, and median for developments without an integrated lender, the next three columns for developments with an integrated lender. Properties in developments with an integrated lender are smaller, cheaper, and bought by poorer households. The last two columns focus on developments with an integrated lender, and show the average differences and associated standard errors between mort-

[5] 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 the 10% of houses built on soil in hydrologic group D (more than 40% clay, high shrink-swell potential) to the expansive soil category. See the Internet Appendix for details.

[6] There are a number of reasons to think that the results for Arizona are relevant for understanding U.S. mortgage lending more generally. First, most of the large property developers and integrated lenders operate nationally, so one would expect to observe similar behavior by the same actors in other states. Second, problems related to construction quality are not limited to Arizona. A survey by Criterium Engineers (2003) finds that, of all new homes in the United States, 21% have problems with roof installations, 15% have problems with the installation of sidings such as stucco, 23% have problems with the installation of windows and doors, and 14% have problems with the construction of the foundation.

[7] Since most of the information is originally recorded at the county level and field population varies widely, not all specifications could be tested on a larger geographic area. For example, a significant number of nondisclosure states do not report transaction prices. Other states, such as Georgia, do not allow me to identify sales by developers. Data from other states, such as Maryland, do 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.

Table I

Summary Statistics of Control Variables

This table shows mean, standard deviation, and median for the control variables in the analysis. The second-to-last column shows the coefficient on *IL* in the regression $Characteristic_i = \alpha + \beta IL_i + \psi_{Year \times Development}$ for developments with an integrated lender. The last column contains standard errors of the estimate, clustered at the development level and presented in parenthesis.

	No Integrated Lender			Has Integrated Lender				
	Mean	*SD*	Med.	Mean	*SD*	Med.	ΔIL	*SE*
Housing Characteristics								
Sales Price (*k* year-2000$)	227.2	120.8	188.8	185.9	82.1	162.6	−3.19	(0.7)
Lot Size (Sq ft)	8,824	4,903	7,716	7,528	3,584	6,777	−56.6	(31.5)
Building Area (Sq ft)	2,362	862.8	2,185	2,117	720.2	1,966	−53.3	(7.82)
Price/Sqft (*k* year-2000$)	29.4	13.3	26.2	27.5	9.7	25.5	−0.40	(0.17)
Garage Spaces	2.46	0.57	2	2.29	0.45	2	−0.01	(0.01)
Total Rooms	7.59	1.91	7	7.18	1.70	7	−0.14	(0.03)
Has Pool	0.39			0.29			0.01	(0.00)
Owner Occupied	0.80			0.80			0.06	(0.00)
Expansive Soil	0.06			0.10			−0.00	(0.00)
Financing Characteristics								
LTV Ratio	0.84	0.14	0.80	0.86	0.14	0.90	−0.00	(0.00)
LTI Ratio	2.49	1.01	2.41	2.58	1.05	2.50	0.19	(0.02)
Mortgage Duration (Years)	29.2	3.74	30	29.4	3.10	30	−0.22	(0.04)
FHA Insured	0.11			0.20			0.07	(0.01)
VA Insured	0.04			0.05			0.01	(0.04)
Jumbo Mortgages	0.01			0.00			−0.01	(0.00)
Borrower Characteristics								
Income (*k* year-2000$)	88.8	71.7	71.8	73.4	58.6	60.2	−11.4	(1.10)
Single Person	0.33			0.36			−0.06	(0.01)
Latino	0.10			0.12			−0.02	(0.00)
Asian	0.04			0.04			−0.01	(0.00)
Single Parent				0.09			−0.00	(0.00)
Number of Children				0.56	0.85	0	0.03	(0.01)
Age of Borrower				47.3	17.3	46	3.16	(0.26)
Dog Owner				0.17			0.03	(0.00)
Cat Owner				0.09			0.02	(0.00)
Census Tract Demographics								
Med. Tract Inc. (k $)	82.6	21.2	85.7	74.1	18.0	71.0		
High School Grad. Rate	0.91	0.10	0.95	0.89	0.09	0.92		

gages made by integrated and nonintegrated lenders in the same development and year. Overall, borrowers of integrated lenders appear *worse* in terms of observable characteristics: they have lower income, buy smaller and cheaper houses, and are more likely to qualify for FHA-insured mortgages. Borrowers of integrated lenders are also more likely to have pets, which can lead to faster depreciation of homes. There is no significant difference between integrated and nonintegrated lenders in the propensity to finance houses built on expansive soil.

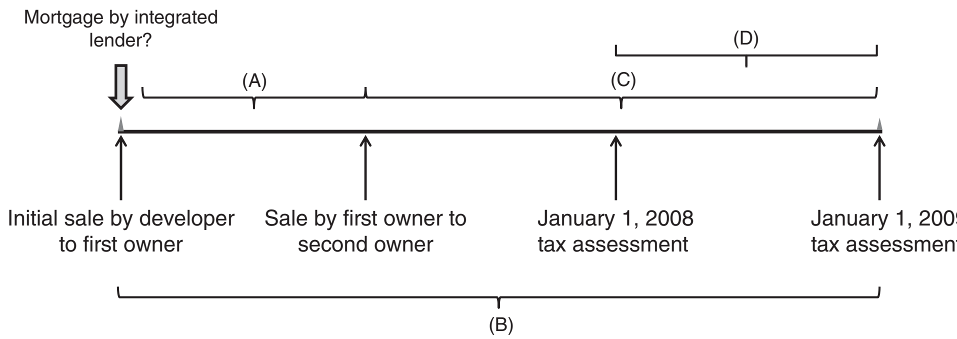

Figure 1. Measures of ex post price performance. This figure shows the four different time horizons over which ex post price performance is measured in the paper.

IV. Outperformance of Integrated Lender Collateral

This section presents evidence for the importance of superior information about collateral quality by the integrated lender. The empirical approach compares the ex post return of homes financed by an integrated lender to the return of ex ante observationally similar homes in the same development financed by nonintegrated lenders. I measure this return over different time horizons described below. Each of these horizons allows me to address a different possible challenge to my identification and interpretation, and together they provide strong evidence for the presence of asymmetric information about collateral quality by integrated lenders.

A. Measuring Collateral Return Using Repeat Sales

I first focus on the subset of homes in developments with an integrated lender for which I observe a second arms-length 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. This corresponds to period A in Figure 1.

I next regress this return on observable control variables and the dummy variable IL_i, which captures whether the mortgage was granted by the integrated lender:

$$Return_{i,q_1,q_2} = \alpha + \kappa IL_i + X_i\beta + \delta_{q_1,q_2} + \epsilon_i. \tag{1}$$

The unit of observation is house i, first sold in quarter q_1 and resold in quarter q_2. A set of fixed effects for each pair of sales quarters (e.g., first sale in Q1 2000, second sale in Q3 2008) is included as δ_{q_1,q_2}; this controls for general market movements in house prices over time. The vector X_i includes

Table II

Annualized Return (%): Between Repeat Sales, Period A

This table shows results from regression (1). The dependent variable is the annualized return of houses between two arms-length transactions. I include single-family residences first sold by a developer over the 2000 to 2007 period in developments with an integrated lender. In columns (6) and (7), I restrict the sample to those transaction pairs where the second transaction was preceded in the prior six months by a death or divorce of the initial owners. All specifications include sales quarter-pair fixed effects and county fixed effects. House characteristics (H) 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, whether the house is built on expansive soil, and whether the house is a rental unit. Buyer characteristics (B) include real income, number of children, age of the buyer, whether the property was purchased by an individual or a couple, whether the buyer is Asian or Latino, whether the buyer is a single parent, and whether the buyer is a dog or cat owner. Financing characteristics (F) include mortgage type, loan-to-value ratio, loan-to-income ratio, and mortgage duration. Census tract demographics (T) include median household income and the percentage of adults over 25 with a high school diploma. D1 includes developer fixed effects, D2 includes development fixed effects. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

						Forced Moves	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Integrated Lender	0.418***	0.455***	0.527***	0.420***	0.386***	0.409**	0.365*
	(0.155)	(0.148)	(0.133)	(0.115)	(0.109)	(0.170)	(0.198)
Controls		H	H, B, F	H, B, F T, D1	H, B, F, T, D2		H, B, F, T, D1
R^2	0.869	0.877	0.878	0.887	0.896	0.885	0.903
$\bar{y}$	7.450	7.450	7.450	7.450	7.450	5.436	5.436
N	30,398	30,398	30,398	30,398	30,398	3,289	3,289

observable characteristics of the house, the owner, and the mortgage. In Table II these are added sequentially to the regression. Standard errors are clustered at the developer level, yielding 109 clusters.[8]

Column (1) of Table II estimates equation (1) with only county fixed effects included as additional control variables in X_i. The magnitude of κ, the coefficient on IL_i, suggests that houses financed by the integrated lender outperform houses in the same county financed by nonintegrated lenders by over 40 bps (0.4 percentage points) annually. The effect is highly statistically significant.

[8] 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 were driven by unobserved developer characteristics that affect all houses of the developer in the same way, the developer fixed effects introduced in column (5) of Table II 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 among all houses of that type, but not necessarily across houses based on a different model. Clustering standard errors at the developer level takes into account this lack of independence across observations by allowing for an arbitrary correlation of residuals of houses built by the same developer.

The magnitude is also sizable relative to the average annual price increase during this period of 7.45%.

One concern is that the outperformance detected in column (1) could be the result of differences in borrower or house characteristics that make it more likely that both the mortgage was granted by the integrated lender and the house increases in value. For example, it could be the case that certain owners take better care of the house and are more likely to borrow from the integrated lender. To address such concerns, between columns (2) and (5) I add an increasing number of control variables to the vector X_i. The precise functional form of the controls is described in the Internet Appendix. Column (2) includes property characteristics such as initial sales price and building size.[9] This captures the fact that houses in different market segments experience a different return over the sample period. Column (3) adds controls for owner characteristics such as income, and mortgage financing characteristics such as the LTV ratio, which affect the borrower's ability and incentives to maintain the property. Column (4) adds census tract demographics, such as median income, as well as developer fixed effects. This prevents the results from being driven by a positive correlation between developer quality and the associated integrated lender's attractiveness. The coefficient κ is remarkably stable with respect to the addition of these observable controls and fixed effects, which reduces concerns that the results are driven by selection on unobservable buyer characteristics (Altonji, Elder, and Taber (2005)).

Column (5) adds development fixed effects. Homes in the same development are built by the same developer; they are also 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 κ is essentially unchanged by this addition.[10] 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 (see Section I for evidence that integrated lenders had knowledge of differential construction quality within developments). Superior information about characteristics that vary at the development level, such as overall development quality, which would quickly be revealed to nonintegrated lenders, does not appear to contribute to the outperformance of the integrated lender's collateral portfolio.

[9] Homes in the same development are often very similar to one another, due to developers' common practice of offering a choice from a small number of model homes. In its 2004 10-K statement, the homebuilder KB Home describes a development to "typically include two to four different model home designs."

[10] One might be concerned that the market development of house prices differs significantly by geography in a way that could bias κ. To show that this is not the case, rather than controlling for simple 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 (5) of Table II, I estimate κ to be 0.389 (clustered SE = 0.107). When controlling for $\delta_{q_1,q_2} \times city$ fixed effects, κ is equal to 0.371 (clustered SE = 0.137). When controlling for $\delta_{q_1,q_2} \times ZIP$ fixed effects, κ is 0.289 (clustered $SE = 0.176$).

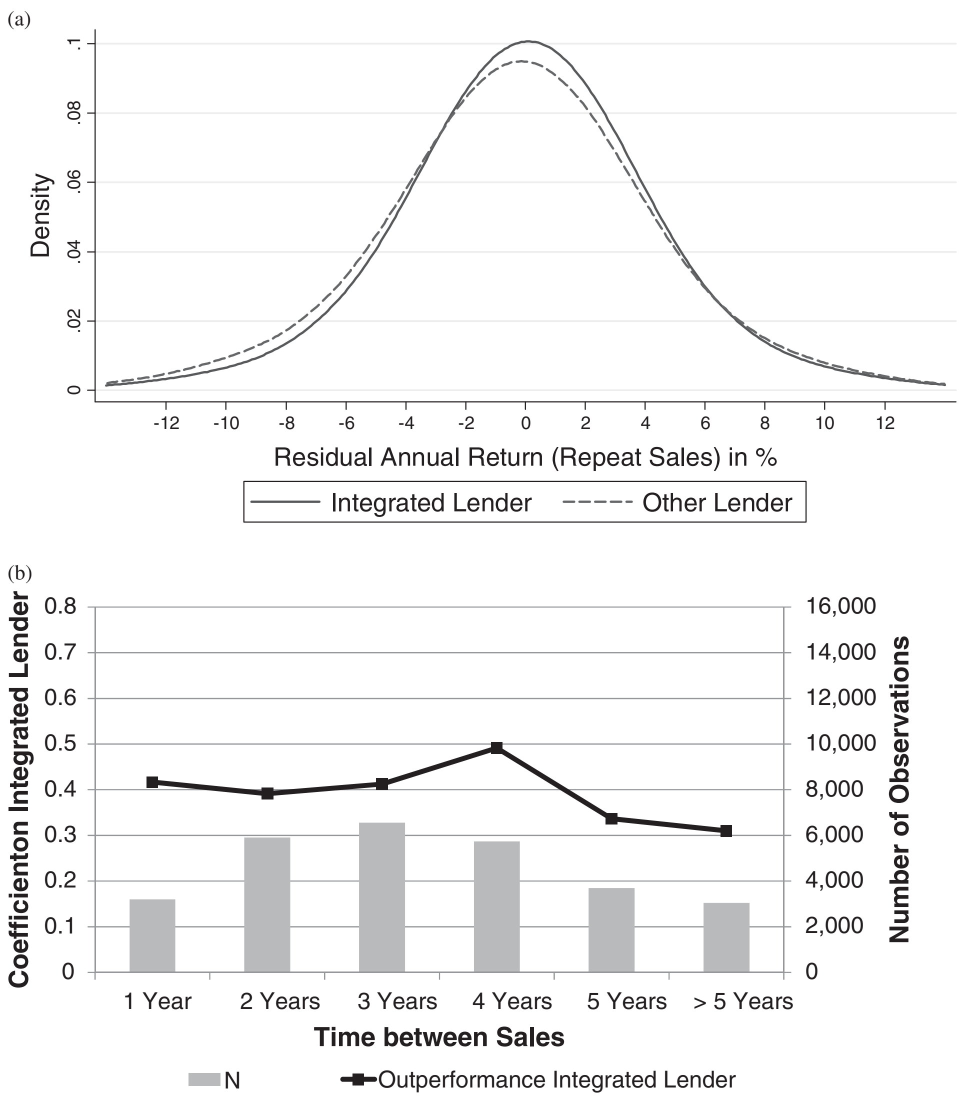

Figure 2. Density and timing of information release. The top panel plots the density of ϵ_i for regression (1) without the IL_i dummy by lender type. A Kolmogorov-Smirnov tests rejects equality of distributions with 99.9% probability. The bottom panel plots κ_j for regression (2); the line graph shows the coefficients, the bar chart the number of observations. Both specifications include the same controls as column (5) of Table II.

In another test, I exclude the dummy IL_i from regression (1), and analyze the residuals ϵ_i by lender type, as plotted in the top panel of Figure 2. A Kolmogorov-Smirnov test rejects equality of the two distributions with 99.9% probability. This graph shows that the 40 bps mean return difference is driven by a thicker left return tail for houses financed by nonintegrated lenders. This is confirmed by a quantile regression analysis, presented in the Internet Appendix, which shows that the integrated lender's collateral portfolio outperforms most for lower percentiles of the conditional return distribution, with the 1st percentile having 3.79 percentage points higher annualized return. Over the average holding period in my regression sample of 3.8 years, this suggests that there are more nonintegrated lender homes that experience adverse price movements of 20% or more. This result is consistent with a story of asymmetric information about collateral values, where houses financed by nonintegrated lenders experience significant structural problems at an above-average frequency. This is important, because it is those significant declines in the value of the collateral that can induce borrowers to default on their mortgage. In addition, alternative explanations of the capital gains differences based on differential maintenance are hard to reconcile with such large declines in house values.[11]

I also consider the time horizon over which the asymmetric information is revealed. Regression (1) implicitly assumes that there is a constant probability of revelation of the integrated lender's initially private property-level information, for example, because the foundation cracks. This translates into a constant annualized outperformance at the collateral portfolio level. 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 conjecture, I re-run regression (1), but instead of including the simple dummy variable IL_i, I interact it with the number of years between the two sales:

$$Return_{i,q_1,q_2} = \alpha + \sum_{j=1}^{6} \kappa_j \times IL_i \times Time\ Between\ Sales_{i,j} + X_i\beta + \delta_{q_1,q_2} + \epsilon_i. \quad (2)$$

The bottom panel of Figure 2 plots the coefficients for a specification that includes the same controls as column (5) of Table II. There is no clear pattern in the annualized outperformance with respect to the time difference between the two sales. The F-statistic for a Wald test of the equality of all coefficients is 0.21 (p-value of 0.96). 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.[12]

[11] The evidence in Wilhelmsson (2008) suggests that, in the first year following construction, the estimated depreciation rate is 0.77% per year for a well-maintained property and 1.10% for a property that is not maintained indoors or outdoors. This is consistent with Campbell, Giglio, and Pathak (2011), who suggest that poor maintenance will, on average, reduce the value of a house by at most 5%.

[12] This is consistent with insights from the civil engineering literature (American Concrete Institute (2001)). The longest time between sales in this sample is 11 years. The rate of revelation of the asymmetric information might decline over time for longer time horizons. However, for the current sample the data suggest that regression (1) is correctly specified with respect to the timing of the outperformance.

B. Selection into Observing Repeat Sales

One might be worried that the subsample of houses for which I observe a resale is not representative of all newly developed homes, and that such a selection might be correlated with the ϵ in regression (1) in a way that might bias the estimate of κ, the outperformance of the integrated lender's collateral portfolio. To address this concern, columns (6) and (7) of Table II restrict the sample to sales pairs in which the second sale is precipitated by a plausibly exogenous event. Specifically, I identify "forced moves" as resales preceded in the prior six months by a death or divorce of the initial owners (see the Internet Appendix). Such sales comprise approximately 10% of all sales pairs, most due to a divorce of the initial owners. When limiting attention to the subsample of forced moves, the integrated lender's collateral portfolio outperforms by about the same amount as it does in the full sample.

In addition, I consider the implied annualized return between the initial sale of a house and its estimated market value in January 2009 (period B in Figure 1), as recorded in the tax assessment records. Such estimated market values are available for *all* houses in the data set.[13] In particular, I run the following regression for all houses i initially sold in month m (e.g., May 2004) in a development with an integrated lender:

$$Return_{i,m} = \alpha + \kappa IL_i + X_i\beta + \delta_m + \epsilon_i. \tag{3}$$

Table III presents the results from regression (3). As before, the outperformance of the integrated lender's collateral portfolio, as measured by the coefficient κ on the integrated lender dummy IL_i, is not significantly affected by the addition of control variables between columns (1) and (5). The magnitude of κ using this measure of return is very similar to that obtained using the return between repeat sales: houses financed by integrated lenders outperform ex ante similar houses financed by nonintegrated lenders by about 40 bps annually.[14]

[13] The Internet Appendix describes the assessment process in Arizona, and analyzes assessment accuracy, which I find to be high. One key mechanism through which assessors learn about differential property conditions is through an elaborate complaints process, which allows homeowners to appeal their tax assessments if they feel their house is worth less than the assessed value.

[14] Unlike before, however, the inclusion of development fixed effects reduces the measured outperformance of the integrated lender's collateral portfolio. This is explained by the "comparables" methodology used by assessors to arrive at the assessed values. This method uses transaction prices of recently sold similar homes, often from within the same development, to predict the market value of homes that have not recently transacted. 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 nonintegrated 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 outperform other developments. The fact that κ remains positive after controlling for development fixed effects suggests that the assessor has some success at detecting differential house quality even within a development, most likely driven by the use of the appeals process described in the Internet Appendix.

Table III

Annualized Return (%): Initial Sale to Assessment, Period B

This table shows results from regression (3). The dependent variable is the annualized return of houses between the initial sale and the January 2009 assessed market value. I include single-family residences first sold by a developer in the 2000 to 2007 period in developments with an integrated lender. All specifications include month-of-sale fixed effects and county fixed effects. Controls for house characteristics (H), buyer characteristics (B), financing characteristics (F), and census tract demographics (T) are defined as in Table II. D1 includes developer fixed effects, D2 includes development fixed effects. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	0.395***	0.445***	0.457***	0.367***	0.285***	0.173***
	(0.111)	(0.101)	(0.091)	(0.088)	(0.077)	(0.046)
Controls		H	H, B, F	H, B, F, T	H, B, F, T, D1	H, B, F, T, D2
R^2	0.811	0.828	0.832	0.881	0.892	0.936
$\bar{y}$	−6.346	−6.347	−6.347	−6.347	−6.347	−6.347
N	87,601	87,601	87,601	87,601	87,601	87,601

C. Foreclosures

A second test for adverse selection on collateral values in mortgage lending is to analyze the performance of the mortgage directly. Since negative equity is a necessary condition for mortgage default, mortgages backed by low-quality collateral should default more frequently. While the deeds data do not track when a mortgage becomes delinquent, I do observe when there is a foreclosure. For every mortgage made over the 2000 to 2007 period in a development with an integrated lender, I determine whether I observe a foreclosure within three years of the initial sale. I then run a probit regression of *Foreclosure3Years* on the dummy variable for whether the loan was made by the integrated lender, IL_i, month-of-sale fixed effects, δ_m, and control variables X_i:

$$Foreclosure\ 3Years_{i,m} = \alpha + \kappa IL_i + X_i\beta + \delta_m + \epsilon_i. \quad (4)$$

Table IV shows average marginal probit coefficients from regression (4). As can be seen, κ is consistently negative and highly significant. Conditional on observables, a mortgage made by an integrated lender is about 1 percentage point, or 40%, less likely to default than a mortgage made by a nonintegrated lender.[15] As before, the addition of control variables and fixed effects between columns (1) and (4) does not affect the magnitude of κ.

[15] The magnitude of the effect on foreclosure rates is large relative to the expected *average* 120 bps price difference after three years. However, Figure 2 shows that the 40 bps annual average return difference is largely driven by a significantly thicker left tail in the return for nonintegrated lender properties. This is confirmed by the quantile regression evidence in the Internet Appendix. It is those homes that experienced particularly severe drops in value that are most likely to end up in foreclosure.

Table IV

Relative Foreclosure Probability of Integrated Lender Mortgages

This table shows average marginal effects from probit regression (4). The dependent variable is whether a foreclosure was observed within three years of purchase. I include single-family residences sold by a developer over the 2000 to 2007 period in developments with an integrated lender. Each specification controls for month-of-sale fixed effects and county fixed effects. House characteristics (H), buyer characteristics (B), financing characteristics (F), and census tract demographics (T) are defined as in Table II. D1 includes developer fixed effects, D2 includes development fixed effects. Columns (5) and (6) restrict the sample to mortgages securitized in the year in which they were originated. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

					Securitized	
	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	−0.011***	−0.009***	−0.009***	−0.009***	−0.010***	−0.018***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)
Controls		H,B,F, T	H,B,F, T,D1	H,B,F, T,D2	H,B,F, T,D1	H,B,F, T,D2
$\bar{y}$	0.017	0.018	0.018	0.021	0.019	0.028
N	71,714	70,397	68,354	59,319	12,322	8,638

One concern with this analysis is that a foreclosure requires a strategic decision by the lender about whether to foreclose on a delinquent mortgage. Since integrated lenders usually hold many mortgages in the same development, they might be reluctant to initiate a foreclosure if doing so would depress prices for neighboring homes. One might thus observe fewer foreclosures for integrated lenders, without their mortgages performing any better. To address this concern, in columns (5) and (6) I restrict the sample to mortgages securitized within the same calendar year in which they were originated. For these mortgages, the decision of whether to foreclose is usually outside the discretion of the originator. The magnitude of the effect in this sample is, if anything, somewhat larger than in the full data.

D. Information on Rejections

I have so far compared the average return and foreclosure probability across all houses that end up being financed by either an integrated or a nonintegrated lender. In practice, however, not all buyers of homes financed by nonintegrated lenders experience the same degree of adverse selection: some are formally rejected by the integrated lender, some receive a high-interest offer from the integrated lender, and some never formally apply to the integrated lender. In this section, I use additional information on all mortgage applications in the HMDA data, including those that do not lead to a mortgage origination, to exploit this heterogeneity. This analysis provides further evidence for the importance of asymmetric information about collateral quality.

Table V
Effect of Rejection by Integrated Lender

Columns (1) to (3) show results from regression (1), where the dependent variable is the annualized return of houses between two arms-length transactions. The controls are as in Table II. Columns (4) to (6) show average marginal effects from probit regression (4), where the dependent variable is whether a foreclosure was observed within three years of purchase. The controls are as in Table IV. I include single-family residences sold by a developer over the 2000 to 2007 period in developments with an integrated lender. The excluded category consists of houses financed by nonintegrated lenders, where the owner applied for and received a mortgage offer from the integrated lender but where this offer was dominated by an offer from a nonintegrated lender. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Return: Period A			Foreclosure		
	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	0.387***	0.409***	0.389***	−0.008***	−0.009***	−0.009***
	(0.118)	(0.120)	(0.114)	(0.001)	(0.001)	(0.002)
Nonintegrated Lender +	−0.450**	−0.429**	−0.368*	0.007***	0.007***	0.007***
Rejection by IL	(0.214)	(0.211)	(0.212)	(0.002)	(0.002)	(0.002)
Nonintegrated Lender +		0.083	0.120		−0.002	−0.002
No Application to IL		(0.154)	(0.142)		(0.002)	(0.002)
Controls	H,B,F, T,D1	H,B,F, T,D1	H,B,F, T,D2	H,B,F, T,D1	H,B,F, T,D1	H,B,F, T,D2
R^2	0.887	0.887	0.896			
$\bar{y}$	7.450	7.450	7.450	0.018	0.018	0.021
N	30,398	30,398	30,398	68,354	68,354	59,319

I treat all mortgage applications in the same year and census tract that have the same loan amount, applicant income, and applicant ethnicity as coming from the same individual.[16] Of those people borrowing from nonintegrated lenders, 7% were rejected for a mortgage by the integrated lender, while 30% never formally applied for a mortgage with the integrated lender. Table V, presents results when including dummies for these two categories in regressions (1) and (4). The excluded category comprises individuals who applied to integrated lenders and received an offer that was dominated by an offer from a nonintegrated lender. Among properties financed by a nonintegrated lender, the return of those properties bought by households that were rejected by the integrated lender is a highly significant 40 bps lower than the return of properties bought by households that were not rejected. Properties that the integrated lender refused to finance are also 0.7 percentage points more likely to end up in foreclosure. The annual outperformance of the houses financed by the integrated lender relative to the houses that the integrated lender refused to finance is almost 1 percentage point. Houses bought by individuals that never applied to the integrated lender, and therefore never faced the adverse

[16] Data in HMDA are only recorded when a borrower makes a formal mortgage application to a lender. This means I am likely to miss more informal applications of borrowers who were discouraged from applying at a relatively early stage.

Table VI

Informativeness of Rejections

This table shows the effect of being rejected by a lender on the return over period A in Figure 1 in columns (1) and (2), and the probability of a foreclosure within three years in columns (3) and (4). I include single-family residences sold by a developer over the 2000 to 2007 period in developments with an integrated lender. In columns (1) and (3), I restrict the sample to houses where the buyers applied for a mortgage with the integrated lender; in columns (2) and (4), I restrict to houses where the buyers applied for a mortgage with a nonintegrated lender. The controls are as in column (4) of Table II for columns (1) and (2), and as in column (3) of Table IV for columns (3) and (4). Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Return: Period A		Foreclosure	
	(1)	(2)	(3)	(4)
Rejected by Integrated Lender	−0.376**		0.007***	
	(0.165)		(0.001)	
Rejected by Nonintegrated Lender		0.018		0.003
		(0.173)		(0.003)
Controls	√	√	√	√
Sample Applied to:	Integrated Lender	Nonintegrated Lender	Integrated Lender	Nonintegrated Lender
R^2	0.893	0.885		
$\bar{y}$	7.450	6.444	0.018	0.029
N	22,844	12,066	48,477	21,876

selection, perform somewhat better than houses bought by people that received an offer from the integrated lender that was dominated by an offer from a non-integrated lender. However, the return difference of about 10 bps annually is not statistically significant.

Another prediction of a model with asymmetric information about collateral values is that a rejection by an integrated lender should be more predictive of future declines in collateral value than a rejection by a nonintegrated lender. To test this conjecture, I run separate regressions on two (possibly overlapping) samples of houses: those for which I observe a mortgage application to the integrated lender, and those for which I observe a mortgage application to a nonintegrated lender. I include homes independent of the outcome of the mortgage application. Within each of these samples, I then estimate the relative return and foreclosure probability of those houses that were rejected by the respective lender.

Table VI shows the results. Among buyers applying for mortgages from the integrated lender, those that were rejected by the integrated lender experience a 40 bps lower collateral return than those that were not rejected; they are also 0.7 percentage points more likely to enter into foreclosure within three years. In contrast, conditional on observable characteristics, a rejection by a nonintegrated lender is not predictive of future declines in collateral value; there is a small and statistically insignificant increase in foreclosure probability associated with a rejection by a nonintegrated lender. This evidence suggests that integrated lenders possess superior information about collateral quality, while nonintegrated lenders do not.

Overall, this section provides substantial evidence that the channel through which the integrated lender is able to assemble a high-quality collateral portfolio is by rejecting mortgage applications for homes that are of below-average quality on unobservable characteristics.

V. Addressing Identification Challenges

To test for asymmetric information about collateral quality, ideally one would randomly assign identical borrowers to observationally identical houses within a development, which differ only on unobservable construction quality. When these home buyers then apply for mortgages with both the integrated and nonintegrated lenders, any variation in information across lenders would be orthogonal to all borrower characteristics and observable characteristics of the home.

In this section, I address three important challenges to my identification that arise from deviations of my empirical setting from this ideal experiment. First, one might be concerned that, despite the large number of controls for borrower characteristics, the results are explained by unobserved differences in borrower quality. For example, if integrated lenders attract higher-quality borrowers who subsequently invest more in the property, this would lead to higher subsequent returns for the integrated lender's collateral portfolio. Second, some properties might be worse on some dimension, such as the view, that is fully observable to all market participants but unobservable to the econometrician. If nonintegrated lenders were more willing to finance such properties (while charging higher interest rates), and they underperform ex post, one would expect the integrated lender's portfolio to experience higher capital gains even in the absence of asymmetric information. Third, one might be concerned that the results are explained by initial price bundling of the mortgage and the house, since any discounts on the house given to customers of the integrated lender would be observationally equivalent to true collateral outperformance when the house is subsequently resold. In Sections V.A to V.C, I present evidence that rules out these alternative explanations.

A. Identification Concern 1: Unobserved Owner Characteristics

A property's capital gain depends on both its initial construction quality and the subsequent investment and maintenance by the owners. Thus, houses financed by an integrated lender could outperform both because they are of superior construction quality, and because their owners have higher willingness and ability to maintain the property. In this section, I further investigate the sources of the outperformance estimated in Section IV. I show that the results cannot be explained by the integrated lender financing borrowers that are superior on characteristics not observed by the econometrician.

There are a number of reasons to expect integrated lenders to lend to *worse* borrowers, which, if anything, would lead me to underestimate the importance of asymmetric information about collateral quality. First, it is unclear how integrated lenders would identify a borrower pool that is superior to that of nonintegrated lenders. While integrated lenders are likely to have better information about the construction quality of the house through their relationship with the developer, one might expect nonintegrated "relationship lenders" to have better information about the borrower relative to integrated lenders that have not previously interacted with the borrower (Elyasiani and Goldberg (2004)).

Second, to the extent that selection is based on buyer characteristics that are observable to all lenders but unobservable to the econometrician, many plausible stories suggest that integrated lenders should be *more* willing to lend to low-quality borrowers, if doing so allows the developer to sell another home. This prior is strengthened by evidence in Table I, which shows that borrowers of integrated lenders are poorer and have higher loan-to-income ratios.[17] In addition, there does not appear to be any pattern in lender identity by stage of the development, which might have been a concern if integrated lenders primarily financed early buyers who were viewed as good initial stewards for the development. Rather, within a development, those homes financed by integrated lenders are sold on average 20 days later than homes financed by nonintegrated lenders.[18]

In the following, I provide direct evidence that differences in borrower characteristics do not explain my findings, and that the integrated lenders' superior information about collateral quality is the key driver explaining the better performance of their mortgages.

A.1. Collateral Return and Foreclosure during Second Ownership Period

First, I consider the relative return of houses *initially* financed by an integrated lender over the ownership period of the *second* owner of the house. Since the identity of this second owner is unknown to all lenders at the time the initial mortgage is granted, any outperformance of the integrated lender's collateral portfolio over this period can be attributed to superior information about collateral quality.

I focus on the subset of houses for which I observe at least two arms-length sales, and calculate the annualized return of these properties between the second sale and the assessed market value in January 2009, which corresponds to period C in Figure 1. Since integrated lenders only lend to buyers of new properties, all of the second owners finance through nonintegrated lenders. Columns (1) and (2) in Table VII show results of a regression of this return on

[17] Consistent with this finding, Agarwal et al. (2011) use a propriatory data set to show that borrowers of integrated lenders had lower FICO scores than borrowers of nonintegrated lenders.

[18] This is determined by a regression of a continuous month-of-sale variable on development fixed effects and an integrated lender dummy. The estimate is highly statistically significant.

Table VII

Annualized Return (%): Second Ownership Period

This table shows results from a regression of the annualized return and foreclosure probability of houses during the ownership period of the second owner. I include all single-family residences initially sold by a developer over the 2000 to 2007 period in developments with an integrated lender. "Integrated Lender" is equal to one if the mortgage to the first owner was made by an integrated lender. Columns (1) and (2) measure the return from the second sale to the assessed value in January 2009, and restrict attention to houses resold prior to 2008. Columns (3) and (4) measure the return between the second and third sales. In columns (5) and (6), the dependent variable is an indicator for whether a foreclosure was observed within three years of purchase by the second owner. Control variables include fixed effects for month of resale in columns (1), (2), (5), and (6), and sales-quarter pairs in columns (3) and (4). All specifications include fixed effects for construction quarter and county as well as characteristics of the house (H) and census tract (T) as in Table II, and characteristics of the buyer (B2) and mortgage (F2). Buyer characteristics include whether the second buyer was an individual or a couple and whether the second buyer was Asian or Latino, in addition to characteristics of the first buyer (B) as in Table II. Financing characteristics include the loan-to-value ratio of the second mortgage in addition to details of the first mortgage (F) as in Table II. D1 includes developer fixed effects, D2 includes development fixed effects. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Return: Period C		2^{nd} to 3^{rd} Sale		Foreclosures	
	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	0.296***	0.178**	0.463*	0.320	−0.025***	−0.024***
	(0.084)	(0.086)	(0.252)	(0.231)	(0.007)	(0.009)
Controls	✓	✓	✓	✓	✓	✓
Other Fixed Effects	D1	D2	D1	D2	D1	D2
R^2	0.902	0.947	0.881	0.891		
$\bar{y}$	−10.87	−10.87	3.323	3.323	0.061	0.069
N	18,339	18,339	5,390	5,390	9,537	8,358

the identity of the *initial* mortgage lender. As before, I control for characteristics of the house, the first and second owner, the first and second mortgage, and the month of the resale.[19] The magnitude of κ over this time horizon is similar to that reported in Tables II and III, with the familiar decline upon the addition of development fixed effects.

In columns (3) and (4), I focus on those homes for which I observe at least three arms-length transactions, and consider the return of the collateral between the second and third sales. While this specification has the advantage that it only relies on market transactions rather than assessed values, the decline in market activity in Arizona since the crash of 2007/8 means that the number of houses for which I observe three sales is limited, and the standard

[19] The set of characteristics of the second owner and mortgage is smaller than that of the first owner and mortgage in Sections IV.A and IV.B, because a significant number of second purchasers did not use a mortgage. This makes it impossible to retrieve income information from the HDMA data. However, since the integrated lender only interacted with the initial owner, a spurious correlation along characteristics of the second owner seems implausible.

errors are consequently larger. The annual outperformance of houses initially financed by an integrated lender remains at around 40 bps during the ownership of the second owner. These estimates suggest that the observed outperformance of the integrated lender's collateral portfolio is driven by asymmetric information about the value of the collateral and not differences in borrower characteristics.

I also analyze the probability of the *second* owner of the house entering into foreclosure within three years of purchasing the property. I only include those observations with a mortgage-financed second sale, since only those might end up in default. Columns (5) and (6) of Table VII show the results from this probit regression. This specification compares the default probability of two similar mortgages, neither of which was granted by the integrated lender. The mortgages differ in whether they are backed by housing collateral that was *initially* financed by the integrated lender. Mortgages that are backed by such collateral are more than 2 percentage points less likely to enter into foreclosure. 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.

A.2. Differential Investment in House

I next provide a direct test of whether differential investment by buyers can explain the observed outperformance of the integrated lender's collateral. For homes in Maricopa, Pima, and Yuma Counties, I observe assessment records for 2008 in addition to 2009. For each house in these counties, I calculate the implied return between the assessed market values in 2008 and 2009, given by period D in Figure 1. I then regress this return on an integrated lender dummy and a set of control variables, as follows:

$$Return\_2008_i = \alpha + \kappa IL_i + X_i\beta + \epsilon_i. \tag{5}$$

Columns (1) and (2) of Table VIII show the results from this regression. As before, houses financed by an integrated lender outperform ex ante similar homes financed by nonintegrated lenders by between 40 and 60 bps (average market price movements over this period were significant). As in regression (3), which also relies on assessed values, the inclusion of development fixed effects reduces the measured outperformance.

I next test whether this differential return can be explained by differential investment in the house over this period. My first measure of investment is an indicator of whether the reported building area of the house changes between the 2008 and 2009 assessor reports, which would provide evidence for construction activity. About 0.4% of properties experience a change in building area. Second, using data obtained from Buildfax on the universe of all building permits filed in Phoenix, for each property I check whether a permit is filed in

Table VIII

Annualized Return (%): 2008 to 2009 Assessment, Period D

The dependent variable in columns (1) to (3) is the return between the assessed market values in January 2008 and 2009, as given by regression (5), that in column (4) is a dummy variable for whether building size changed in 2008, and that in column (5) is a dummy variable for whether a building permit was filed in 2008. I include single-family residences first sold by a developer over the 2000 to 2007 period in developments with an integrated lender. In columns (1) to (4) I restrict the sample to properties in Maricopa, Pima, and Yuma Counties, for which I observe the 2008 assessed values. In column (5) I restrict the sample to properties in the city of Phoenix, for which I can observe building permits. All specifications control for the assessed house value in January 2008, in addition to other characteristics of the house (H), the buyer (B), the financing (F), and the census tract (T) as in Table II. D1 includes developer fixed effects, D2 development fixed effects. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Return: Period D			Δ Building	Has Permit
	(1)	(2)	(3)	Area (4)	(5)
Integrated Lender	0.633***	0.334***	0.326***	−0.0006	0.0002
	(0.196)	(0.127)	(0.122)	(0.0010)	(0.0027)
Δ Building Area			0.961		
			(0.818)		
Controls	D1	D2	D2	D2	D2
R^2	0.613	0.797	0.801	0.103	0.331
$\bar{y}$	−27.45	−27.45	−27.47	0.0042	0.0141
N	69,949	69,949	69,876	69,876	7,654

2008, which is the case for about 1% of properties. The results in column (3) suggest that houses that experience a change in building area do have larger returns, though the results are not statistically significant. Columns (4) and (5) test whether houses initially financed by an integrated lender are more likely to experience an investment event in 2008. I regress a dummy variable capturing the change in building area or the filing of a permit on whether the house was financed by an integrated lender, as well as control variables. There is no significant difference in the probability of investment for houses financed by different lenders. This suggests the outperformance of the integrated lender's collateral portfolio is not driven by differential investment by the owners.

A.3. Importance of Collateral Quality: Expansive Soil

The integrated lender's information about construction quality is particularly relevant for the return of homes built on expansive soil, while the importance of any differences in borrower characteristics should not vary by soil type (see Prediction 1(a) and Section I). Therefore, if integrated lenders have superior information about construction quality, the outperformance of their collateral portfolio should be higher for homes built on expansive soil; if their primary information relates to borrower characteristics, one would not expect this to be the case. To test this prediction, in regression (6) I include a dummy for expansive soil, ES_i, as well as its interaction with the integrated lender dummy:

$$Return_{i,q_1,q_2} = \alpha + \kappa_1 IL_i + \kappa_2 ES_i + \kappa_3 IL_i \cdot ES_i + X_i\beta + \delta_{q_1,q_2} + \epsilon_i. \quad (6)$$

The results from regression (6), measuring return over periods A to D in Figure 1, are reported in Table IX.[20] The coefficient on ES_i is negative and usually statistically significant, indicating a lower average return for houses built on expansive soil. More importantly, the positive and significant coefficient on the interaction between IL_i and ES_i shows that, for houses built on expansive soil, the integrated lender's collateral portfolio outperforms an ex ante observationally similar portfolio of houses financed by a nonintegrated lender by almost 1 percentage point annually ($\kappa_1 + \kappa_3$). This evidence further suggests that the outperformance of the integrated lender's collateral is attributable to its superior information about the initial construction quality of the housing collateral. The coefficient κ_1 also remains positive, which suggests that some of the outperformance of the integrated lender relates to information about collateral characteristics that also affect the return of houses not built on expansive soil, for example, information about the quality of the electric wiring.[21]

For each period I also include a specification with development fixed effects that exploits within-development variation in soil type. The positive coefficient on the interaction term remains. The lower statistical significance of the estimate results from possible measurement and classification errors that arise from the soil type being inherently continuous, while my measure of soil expansiveness is discrete.[22] 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 returns for houses closer to soil boundaries, and so actual differences in soil expansiveness are smaller.

[20] In the Internet Appendix, I show that the information contained in a rejection by an integrated lender is also greater in the sample of houses built on expansive soil (see Section IV.D).

[21] It is important to recall that soil quality is, in fact, observable to all market participants. What is not observable is the construction quality of the property. Some aspects of construction quality are particularly important for houses built on expansive soil (e.g., the quality of the foundation). If present, these types of construction defects generate a particularly large outperformance of the integrated lender's collateral for houses located on expansive soil. Other aspects of construction quality are equally important across all soil types (e.g., the quality of the stucco, roofing, plumbing, and electric wiring). These construction defects generate an outperformance of the integrated lender's collateral even among homes not built on expansive soil. The evidence in Table IX is consistent with the presence of both types of construction defects.

[22] 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."

Table IX

Annualized Return (%) by Soil Type

This table shows results from regression (6). ES_i is equal to one for houses built on hydrologic soil group D. Columns (1) and (2) correspond to Table II, columns (3) and (4) correspond to Table III, columns (5) and (6) correspond to Table VII, and columns (7) and (8) correspond to Table VIII. Control variables are included as indicated. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Period A		Period B		Period C		Period D	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expansive Soil	−0.223	−0.768**	−1.037*	−0.886***	−1.843***	−0.812***	−3.352***	−0.827
	(0.509)	(0.338)	(0.598)	(0.187)	(0.406)	(0.294)	(1.272)	(0.643)
Integrated Lender	0.363***	0.353***	0.224***	0.145***	0.213**	0.166*	0.493**	0.116
	(0.119)	(0.107)	(0.080)	(0.046)	(0.089)	(0.084)	(0.147)	(0.105)
Integrated Lender ×	0.564**	0.324	0.595***	0.293**	0.776***	0.0551	1.274***	0.663**
Expansive Soil	(0.266)	(0.227)	(0.207)	(0.125)	(0.255)	(0.151)	(0.448)	(0.307)
Controls	Table II	Table II	Table III	Table III	Table VII	Table VII	Table VIII	Table VIII
	Col (4)	Col (5)	Col (5)	Col (6)	Col (3)	Col (4)	Col (1)	Col (2)
R^2	0.887	0.896	0.891	0.936	0.902	0.947	0.654	0.831
$\bar{y}$	7.450	7.450	−6.347	−6.347	−10.87	−10.87	−28.09	−28.09
N	30,398	30,398	87,601	87,601	18,339	18,339	69,949	26,809

Table X

Test for Differential Riskiness: Control for Interest Rate

This table shows results when controlling for the initial interest rate, for the subset of ARMs and hybrid ARMs for which this information is available. The dependent variables in columns (1) to (4) are the annualized collateral return over periods A to D, respectively. The dependent variable in column (5) is the probability of observing a foreclosure within three years of the initial sale. The control variables are as indicated. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Period A (1)	Period B (2)	Period C (3)	Period D (4)	Foreclosure (5)
Integrated Lender	0.462*	0.340***	0.471***	0.395**	−0.007**
	(0.238)	(0.098)	(0.159)	(0.170)	(0.003)
Mortgage Spread	−0.450***	−0.167***	−0.086	−0.134*	0.011***
	(0.099)	(0.027)	(0.060)	(0.078)	(0.001)
Controls	Col (4) Table II	Col (5) Table III	Col (2) Table VII	Col (2) Table VIII	Col (2) Table IV
R^2	0.888	0.881	0.869	0.593	
$\bar{y}$	5.370	−8.340	−14.43	−27.59	0.038
N	8,002	23,864	3,439	6,885	16,012

B. Identification Concern 2: Unobserved House Characteristics

A second concern with my identification of asymmetric information about collateral values is that, despite the extensive set of control variables, there could be systematic differences in property characteristics that are unobserved by the econometrician but observed by all market participants. For example, it could be the case that integrated lenders are more risk-averse and choose not to lend against low-quality collateral, while nonintegrated lenders are more willing to lend against low-quality collateral but at higher interest rates. If my property controls do not pick up this difference, one might observe outperformance of the integrated lender's collateral portfolio even if collateral quality were perfectly observed by all lenders.

Ideally, I would thus control for the pricing of mortgages when analyzing the difference in collateral returns. Unfortunately, interest rates in Arizona are only recorded when the mortgage is an adjustable-rate mortgage (ARM) or a hybrid ARM.[23] I therefore present results only for the subset of mortgages for which I observe the initial interest rate. I construct a "mortgage spread" variable that equals 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 II, III, IV, VII, and VIII. The results are shown in Table X.

Houses financed with a higher interest rate mortgage have a lower return. This result suggests that there are aspects of collateral quality that are

[23] In addition, while interest rates capture the most salient aspect of mortgage pricing, I do not observe other components such as the closing costs.

observable to lenders and unobservable to the econometrician that lenders take into account when pricing mortgages. Columns (1) to (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 by about 40 to 50 bps annually. Column (5) shows that, the probability of foreclosure is also larger for mortgages with a higher interest rate. Mortgages originated by integrated lenders continue to have a significantly lower foreclosure probability than those originated by nonintegrated lenders.

In interpreting these results, one might be concerned about the ability of nonintegrated lenders to correctly price the risk they are bearing when financing lower-quality homes. If interest rates were not increased sufficiently to compensate for the additional risk, some outperformance of the integrated lender's collateral could persist even after controlling for the interest rate. However, even if lenders did not completely price the risk inherent in the lower quality mortgages, the fact that the point estimates of the outperformance remain virtually unchanged makes it unlikely that the outperformance is driven by collateral characteristics of which nonintegrated lenders were aware.

To further show that integrated lender properties do not differ systematically on characteristics observed by all market participants but unobserved by the econometrician, I argue that such characteristics, if present, should affect the initial sales price. In that case, houses financed by integrated lenders would sell at a different price, conditional on the control variables. This is not the case. Similar to the hedonic pricing analysis of Rosen (1974), I regress the log of the initial price of a house on whether the house was financed by an integrated lender as well as the same house-specific control variables as in Table II. I also include development by sales-quarter fixed effects:

$$Log\ Price_{i,q_1} = \alpha + \kappa IL_i + X_i\beta + \delta_{q_1,Development} + \epsilon_i. \tag{7}$$

The Internet Appendix shows the coefficients on the hedonic variables. The coefficient κ is equal to 0.001, with a standard error of 0.002. Price increases monotonically in building size and lot size. All else equal, properties built on expansive soil sell at a 2.5% discount. The regression has a relatively high R^2 of 94%, suggesting that the control variables capture most of the observable variation across properties. It is thus unlikely that there are significant differences between houses financed by integrated and nonintegrated lenders that are not captured by my control variables but are observable to market participants at the time of the initial sale. Any differences in subsequent returns, therefore, are due to initially unobservable differences in house characteristics.[24]

In addition to the two tests presented in this section, previously discussed results also provide evidence against an alternative story that does not rely on

[24] This result also provides further support for the assumption that borrowers do not extract the integrated lender's information about collateral quality from its mortgage offer. If they did, one would expect them to pay less for houses when the integrated lender made an unattractive mortgage offer.

asymmetric information. In particular, if all lenders had the same information about collateral quality but differential risk preferences, one would expect a rejection by either lender to be equally predictive of future collateral returns.[25] In contrast, the evidence presented in Section IV.D shows that, while a rejection by an integrated lender is predictive of lower future collateral gains, whether a mortgage application gets rejected by a nonintegrated lender is not correlated with future house appreciation.

C. Identification Concern 3: Bundling of Home and Mortgage

One concern when measuring collateral return starting from the initial sale by the developer is that price bundling of the house and the mortgage by the developer could contaminate these measures.[26] If such bundling involved discounts on the house price for customers of the integrated lender, it would be observationally equivalent to true collateral outperformance when the house is subsequently sold for its actual value.

A number of results discussed above already address this concern. If the outperformance was indeed driven by an initial price discount that is subsequently capitalized, the *annualized* outperformance should decline with the time between sales; it would also lead to an equal increase in the return of all houses financed by the integrated lender. The results presented in Figure 2, conversely, show that the annualized outperformance does not differ by the time between sales, and that it is driven by a thicker left tail in the return distribution of nonintegrated lenders. In addition, the results in Section V.A.1 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 second owners, it should not contaminate these returns.[27] Furthermore, unless initial price discounts

[25] If anything, if the results in Section IV.A were explained by a willingness of nonintegrated lenders to finance lower-quality collateral, one would expect a rejection by those lenders to be *even more* predictive of lower future capital gains. This is because houses that even the nonintegrated lender was unwilling to finance would be of particularly bad quality.

[26] Section 8 of the Real Estate Settlement and Proceedings Act (RESPA) regulates bundling in the mortgage lending process. Section 8(c)4 deals with affiliated businesses. A customer referred to an affiliated business must not be required to use the affiliate, and its use may not be required as a condition for the availability of any other settlement service for which the consumer will pay. Offering a package of services to a consumer at a discount, or offering another consumer incentive, however, would not constitute required use and thus would be acceptable under RESPA. Any such incentive must be a true discount and not result in higher prices elsewhere in the settlement process. In general, a price discount of a few percent would be sufficient to translate into a 40 basis points annual outperformance over the horizons considered in this paper. While there is some evidence that home builders did offer incentives such as reduced closing costs to borrowers of the integrated lender, I found less evidence that those borrowers paid less for their house. In any case, the quantitative impact of such discounts on my results is an inherently empirical question.

[27] This is subject to the caveat that the second house price might be affected by the initial discount through an anchoring effect (Kahneman (1992), Lambson, McQueen, and Slade (2004), Beggs and Graddy (2009)), where people take the initial price as a reference point for subsequent home value. Evidence on the magnitude of such anchoring effects in real estate markets is inconclusive.

for choosing the integrated lender were particularly large for homes built on expansive soil, the evidence in Section V.A.3 can only be reconciled with a story of asymmetric information about collateral values. Finally, if the results were explained by price bundling, all houses financed by the nonintegrated lender should have the same return, since none of them received the price discount. However, in Section IV.D, I show that, even among houses financed by nonintegrated lenders, those whose owners were rejected by the integrated lender significantly underperform, providing evidence against a bundling-based story and for an adverse selection story.

C.1. Direct Evidence for Property Damage

Further evidence that my findings cannot be explained with initial price bundling comes from looking at direct evidence of differential depreciation, rather than relying on prices to capture this depreciation. To do so, I scan the textual descriptions of all property listings on the online real estate listings platform Trulia.com between October 2005 and August 2010 for evidence of damage to the property. I identify three categories of evidence for property damage. The first category includes listings that propose an "as-is" sale, whereby the buyer accepts the house "with all faults," whether or not immediately apparent. The second category includes listings with a description that includes at least one of the phrases "repair," "damage," "broken," "leak," "peeling," "crack," "needs work," "fix-up," and "TLC." The third category includes listings that include terms suggesting the home is particularly suited for a special buyer such as "handyman," "right buyer," and "investor."

The results in Table XI show that, among all houses listed on Trulia.com, those initially financed by the integrated lender are 2.7 percentage points less likely to propose an as-is sale, relative to a baseline probability of around 14%. They are also about 1 percentage point less likely to include words that indicate damage to the property (baseline of 6%), and about 1 percentage point less likely to suggest the property is particularly attractive for a special buyer (baseline of 5%). This provides direct evidence that the higher returns of houses financed by the integrated lender are indeed driven by superior collateral quality and not by bundling of the home and mortgage.

C.2. Outperformance and Relevance of Collateral Quality

Prediction 1(b) posits that, in the presence of asymmetric information about collateral quality, the outperformance of the integrated lender will vary with the down payment. When the down payment is large, the repayment probability is less sensitive to collateral quality, since house prices have to decrease by more to create incentives for default. Hence, nonintegrated lenders are less

In particular, it is unclear whether one would expect a strong bias in an environment where the developer clearly signals that the initial price was below the true value of the house, and where the price discount is directly tied to the use of the integrated lender.

Table XI
Evidence for Property Damage

This table shows the average marginal effects from a probit regression explaining damage indicators in property listings. I analyze three indicators of property damage: whether the house is sold "as-is" (columns (1) and (2)), whether there were details of damage in the description (columns (3) and (4)), and whether the property was said to be attractive for special buyers (columns (5) and (6)). I include single-family residences first sold by a developer in the 2000 to 2007 period in developments with an integrated lender and that were listed for resale on Trulia.com between October 2005 and August 2010 (but at least one year after the first sale), and that include a textual description. All specifications include month of sale fixed effects, county fixed effects, and a control for the time between the initial sale and the listing. They also include house characteristics (H), buyer characteristics (B), financing characteristics (F), and census tract demographics (T) as in Table II. D1 includes developer fixed effects, D2 includes development fixed effects. Standard errors are clustered at the developer level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	"As-Is"		Damage Indicator		Special Buyer	
	(1)	(2)	(3)	(4)	(5)	(6)
Integrated Lender	−0.027***	−0.025***	−0.009**	−0.010*	−0.012**	−0.016***
	(0.007)	(0.008)	(0.005)	(0.006)	(0.005)	(0.006)
Controls	D1	D2	D1	D2	D1	D2
$\bar{y}$	0.138	0.143	0.055	0.063	0.046	0.055
N	11,299	10,742	10,908	9,379	10,759	8,808

concerned about collateral quality, and offer more aggressive financing. This allows the integrated lender to assemble a particularly attractive collateral portfolio. In contrast, for high-LTV ratio mortgages a small decline in house prices can generate incentives for default. Nonintegrated lenders thus offer less aggressive financing to avoid the winner's curse, which, in equilibrium, improves their average collateral quality. If the integrated lender's outperformance were instead driven by a bundling of home and mortgage, one would not expect differential outperformance by initial downpayment.

To test which of these predictions is borne out in the data, 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%, the upper limit for FHA-insured mortgages. Regression (8) estimates the outperformance of the integrated lender's collateral portfolio for each LTV ratio group:

$$Return_i = \alpha + \sum_{j=1}^{4} \kappa_j IL_i \cdot LTV_{i,j} + \sum_{j=2}^{4} \omega_j LTV_{i,j} + X_i\beta + \delta_{q_1,q_2} + \epsilon_i. \quad (8)$$

Figure 3 depicts the κ_j coefficients, which estimate the outperformance of the integrated lender's collateral portfolio for each LTV ratio group, when measuring the return over different time horizons. The gray bars show the coefficients when the return is measured over period A in Figure 1, using the same control variables as in column (4) of Table II. The houses financed by the integrated

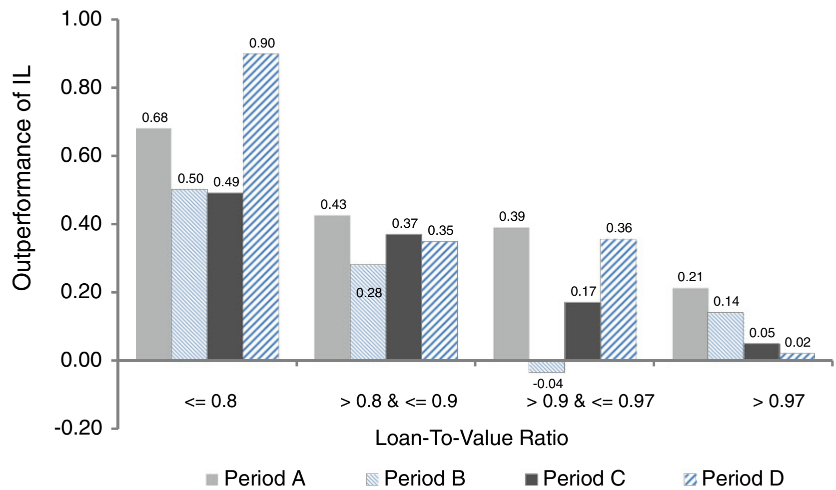

Figure 3. Relative collateral return of integrated lender by LTV ratio. This graph plots the κ_j coefficients from regression (8). I measure return over periods A, B, C, and D in Figure 1 and include the same covariates as column (4) of Table II, column (5) of Table III, column (2) of Table VII, and column (1) of Table VIII, respectively.

lender outperform for all LTV ratio groups. The outperformance is particularly large for houses backing those low-LTV ratio mortgages, for which a decline in collateral value does not lead to a significant increase in the default probability. Similar results obtain when measuring the return over periods B, C, and D. 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 to D are 4.20 (p-value of 0.04), 10.92 (p-value of 0.00), 7.74 (p-value of 0.01) and 16.24 (p-value of 0.00) respectively.

VI. Interest Rate Response to Integrated Lender

In Section IV, I analyzed the effects of adverse selection on the collateral quality of integrated and nonintegrated lenders. Section V rules out a number of alternative explanations and addresses important identification challenges. In this section, I test the predictions for the equilibrium interest rates charged by nonintegrated lenders. Prediction 2 posits that this interest rate should be higher when the nonintegrated lender competes against an integrated lender, and thus lends against below-average-quality collateral. I analyze the interest rates charged by nonintegrated lenders by running regression (9), which compares developments with and without integrated lenders:

$$MortRate_{i,m,f} = \alpha + \kappa HasIL_i + X_i\beta + \delta_{m,f} + \tau_l + \epsilon_i. \tag{9}$$

Table XII

Impact of Integrated Lender on Interest Rates of Other Lenders

This table shows results from regression (9). The dependent variable is the mortgage interest rate. I include single-family residences sold by a developer in Arizona over the 2000 to 2007 period that were financed by nonintegrated lenders. Each specification includes month × rate-type (adjustable or hybrid adjustable) fixed effects, county fixed effects, and lender fixed effects. Finance characteristics (F), house characteristics (H), buyer characteristics (B), and census tract demographics (T) are as in Table II. D1 includes developer fixed effects. Standard errors are clustered at the lender level and presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	(1)	(2)	(3)	(4)	(5)	(6)
Has Integrated Lender	0.116**	0.114**	0.099**	0.088**	0.092**	0.077*
	(0.055)	(0.054)	(0.045)	(0.043)	(0.042)	(0.042)
Expansive Soil			−0.000	0.023	0.001	−0.089*
			(0.036)	(0.038)	(0.033)	(0.048)
Has Integrated Lender ×						0.150***
Expansive Soil						(0.054)
Controls	·	F	F, H, B	F, H, B, T	F, H, B, T D1	F, H, B, T
R^2	0.555	0.583	0.591	0.592	0.597	0.592
$\bar{y}$	6.638	6.638	6.638	6.638	6.638	6.638
N	15,620	15,620	15,620	15,620	15,620	15,620

The dependent variable is the initial mortgage interest rate for ARMs or hybrid ARMs (see Section V.B). The key explanatory variable, $HasIL_i$, captures whether the nonintegrated lender competes against an integrated lender to make a particular mortgage. This variable is set to one when an integrated lender makes loans in the same development and year. I include month by rate-type (adjustable or hybrid adjustable) fixed effects, $\delta_{m,f}$, to control for the contemporaneous interest rate environment. I also include lender fixed effects, τ_l. The regression thus compares the lending behavior of the same nonintegrated 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. Standard errors are clustered at the lender level.

The results, which are presented in Table XII, suggest that, nonintegrated lenders charge about 10 bps higher interest rates when competing against a better-informed integrated lender. This is economically significant, and, based on a rate sheet by U.S. Bank from June 2008 for 5/1 ARMs, equivalent in magnitude to the increase in interest rates experienced by borrowers when they reduce their downpayment from 25% to 20%. Between columns (1) and (6), I sequentially add control variables. The magnitude of the estimated interest rate increase does not change. This finding is important, because it suggests that the results are not driven by a different composition of houses or borrowers in developments with or without an integrated lender. In column (5), when I include developer fixed effects, κ is identified by considering houses built by developers that sometimes, but not always, cooperate with an integrated lender.[28] The evidence is highly consistent with a story of asymmetric information about collateral quality.

These results also provide further evidence alleviating the identification challenges discussed in Section V. First, they address concerns about a spurious correlation on characteristics of the house that are observable to lenders but unobservable to the econometrician (see Section V.B). If different risk preferences lead nonintegrated lenders to finance lower-quality collateral, one would not expect them to adjust their interest rates depending on whether they are competing with an integrated lender. The results presented above are thus consistent with a model in which nonintegrated lenders face adverse selection in the presence of integrated lenders, but are not consistent with a model in which all lenders are fully aware of differences in collateral quality. Second, the results in Table XII are inconsistent with a story of price bundling discussed in Section V.C, which would not involve nonintegrated lenders pricing differentially depending on the presence of integrated lenders.

In column (6), I add the interaction between $HasIL_i$ and ES_i, the dummy variable capturing whether the house was built on expansive soil. This specification tests Prediction 2(a), which posits that the interest rate premium for competing with an integrated lender will be particularly large for those houses for which the integrated lender's information about construction quality has the most impact on future house prices. Consistent with 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.

The interest rate increase for competing against an integrated lender should also be larger when the adverse selection on collateral quality is more important—see Prediction 2(b). In particular, nonintegrated lenders should raise interest rates more for high-LTV ratio mortgages, for which a small decline in collateral value precipitates a larger increase in default risk. To test whether this is the case, I run the following regression:

$$Mort\ Rate_{i,m,f} = \alpha + \sum_{j=1}^{4} \kappa_j HasIL_i \cdot LTV_{i,j} + \sum_{j=2}^{4} \omega_j LTV_{i,j} + X_i\beta + \delta_{m,f} + \tau_l + \epsilon_i. \quad (10)$$

Figure 4 plots the κ_j coefficients from this regression. The regressions for the top and bottom panels include the same control variables as columns (4) and (5) of Table XII, respectively. The results support the model prediction. For high-LTV ratio mortgages, the interest rate premium charged by the nonintegrated lender when competing against an integrated lender is the largest, at almost half a percentage point annually. The F-statistics for a Wald test of the equality of κ_1 and κ_4 are equal to 16.2 (p-value of 0.00) for the left panel and 17.4 (p-value of 0.00) for the right panel.

[28] I do not include development fixed effects, since these are nearly collinear with $HasIL_i$, 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.

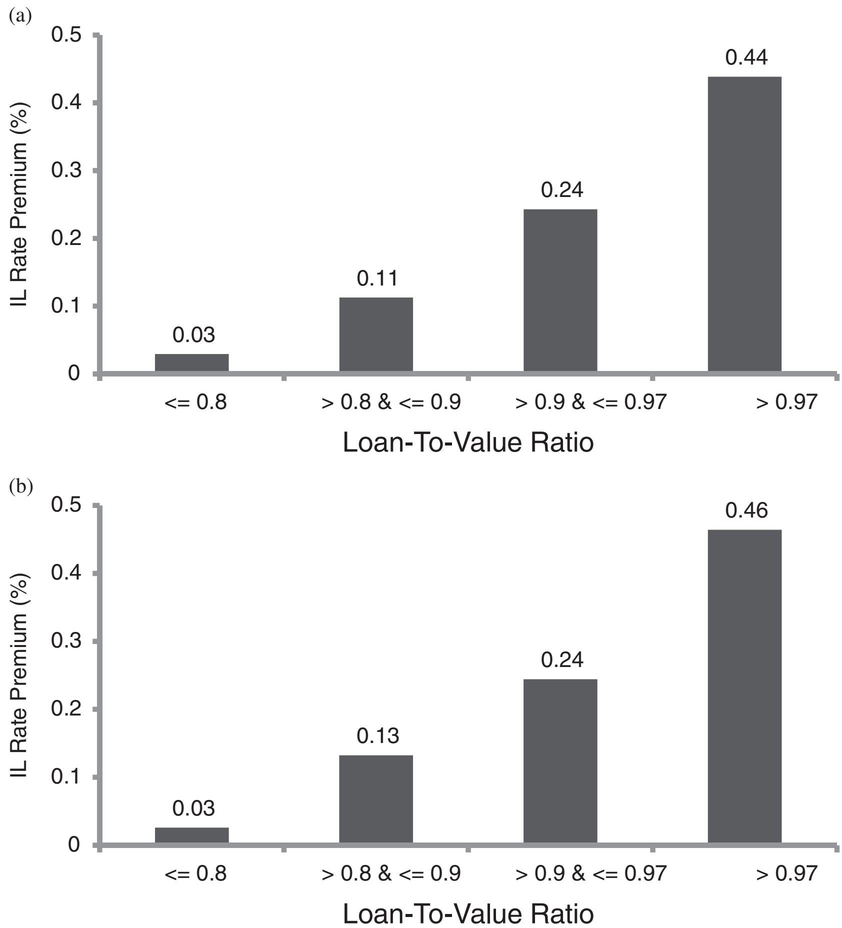

Figure 4. Interest rate increase by nonintegrated lenders by LTV ratio. These graphs plot the estimates of the κ_j coefficients for regression (10), measuring the interest rate increase for mortgages by nonintegrated lenders when they compete with an integrated lender, for different groups of the LTV ratio. The top panel includes the same covariates as column (4) of Table XII. The bottom panel includes the same covariates as column (5) of Table XII.

While I find that nonintegrated lenders adjust interest rates when competing with an integrated lender, these results do not necessarily imply that nonintegrated lenders are explicitly aware of the underlying adverse selection. The pricing of mortgages usually relies on statistical models that use past data to compute expected future default probabilities, and adjust interest rates accordingly. When mortgages on houses built by a certain developer default more often, the nonintegrated lender would charge higher interest rates for mortgages to purchase those homes. From the perspective of the nonintegrated lender, this could be because the quality of homes built by the developer is generally low, or because of adverse selection on collateral quality. To price these mortgages correctly (i.e., so that default realizations do not contradict default expectations), nonintegrated lenders would not need to differentiate between these two explanations.

VII. Discussion and Welfare Effects

The welfare consequences of the existence of integrated lenders, and the resulting adverse selection, are ambiguous. While borrowers pay at least 10 bps 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, pushing down house prices.[29] However, since the interest rate increase is particularly large for borrowers with high LTV ratios, this might involve a welfare transfer from households that make small downpayments to households that make large downpayments.

In addition, the adverse selection can introduce costly frictions in the home financing process. When economic fundamentals are strong, asymmetric information about the collateral may not significantly impact the value of a particular mortgage, as repayment is less sensitive to collateral values. When a negative shock hits the economy, the repayment probability becomes more sensitive to information about collateral quality. Interest rates and required down payments will rise, and leverage will decline more rapidly than in the absence of asymmetric information (Fishman and Parker (2010), Gorton and Ordonez (2011)). In this environment, lending markets can break down completely, and some borrowers will not receive a mortgage offer.[30] Hence, asymmetric information about collateral quality can be an important propagation mechanism for shocks, can introduce volatility in the demand for new homes, and can have important negative welfare consequences for households unable to obtain a mortgage.

To the extent that developers determine the construction quality, the presence of the integrated lender may also provide a valuable quality guarantee function, since it provides buyers the option to return low-quality homes to the developer through defaulting. This generates incentives for developers to produce higher quality homes than they might otherwise do. This mechanism is more effective than competing mechanisms such as warranties, for which the

[29] On the other hand, a developer with market power might also use its integrated lender to solve Coase's (1972) durable goods monopoly problem, creating upward pressure on house prices. When a developer cannot commit to not sell houses for less in the future, households may be unprepared to pay more than marginal cost for early units. Consequently, the developer can lose all its market power. Bulow (1982) shows that if, rather than selling the product, the monopolist leases it to consumers, it can retain its market power. This is because if the lessor 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 similar way.

[30] In the model in the Internet Appendix, this is captured by the fact that, when γ is low (proxying a bad economic state when repayment is very sensitive to collateral values), some borrowers will not receive any mortgage.

buyer must convince the developer that quality is not as promised. In addition, the developer must still be in business to fulfill its guarantee.[31]

VIII. Conclusion

In many credit markets, lenders are differentially informed about the expected returns from making a loan. From a theoretical perspective, this has important implications for the equilibrium outcomes in these markets. Empirical tests of the impact of such information asymmetries are complicated by key data challenges, including the importance of identifying the relative information of different lenders, and the need to evaluate the quality of different loan portfolios along dimensions that are initially unobservable to some lenders.

I address these challenges by focusing on the important market to finance the purchase of newly developed homes. In this market, property developers usually cooperate with vertically integrated mortgage lenders that have access to information about the construction quality of the housing collateral. By conditioning their interest rate offers on this superior information, integrated lenders subject competing nonintegrated lenders to adverse selection. I show that such adverse selection on collateral quality is a key feature of this market. In developments with an active integrated lender, houses financed by the integrated lender have a 40 bps higher capital gain than ex ante observationally similar houses in the same development financed by nonintegrated lenders. Their owners are also 40% less likely to enter into foreclosure. These results are not explained by confounding unobserved borrower characteristics, characteristics of the house that are observable to market participants but not the econometrician, or price bundling of the house and the mortgage. To compensate for lending against below-average-quality collateral, nonintegrated lenders charge 10 bps higher interest rates when competing against an integrated lender.

These results provide a novel test of lender behavior in credit markets when they are differentially informed about the profitability of making a loan. I show that, consistent with theoretical predictions, better-informed lenders subject competing lenders to adverse selection on loan quality. This is associated with significantly higher equilibrium borrowing costs. My results also highlight the pervasive nature of asymmetric information in mortgage markets, and identify collateral values as a novel source of such asymmetric information. More generally, 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 area or price segment

[31] 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, Malitz, and Ravid (1993) provide empirical evidence for such a theory by showing that firms selling products where quality requires more time to assess extend more trade credit.

(Loutskina and Strahan (2011)). The resulting increase in market power for better-informed lenders might lead to a segmentation of the mortgage market, with different lenders specializing to lend against certain collateral. It may also limit the ability of distant online-lenders to compete in this market. This would increase the equilibrium cost of credit for borrowers.

Initial submission: October 16, 2013; Accepted: October 1, 2014
Editors: Bruno Biais, Michael R. Roberts, and Kenneth J. Singleton

REFERENCES

Agarwal, Sumit, Gene Amromin, Claudine Gartenberg, Anna Paulson, and Sriram Villupuram, 2011, Homebuilders, affiliated financing arms and the mortgage crisis, Working paper, Federal Reserve Bank of Chicago.

Agarwal, Sumit, Yan Chang, and Abdullah Yavas, 2012, Adverse selection in mortgage securitization, *Journal of Financial Economics* 105, 640–660.

Agarwal, Sumit, and Robert Hauswald, 2010, Distance and private information in lending, *Review of Financial Studies* 23, 2757–2788.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, 2005, Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools, *Journal of Political Economy* 113, 151–184.

American Concrete Institute, 2001, Control of cracking in concrete structures, ACI 224R-01. Available at http://www.bpesol.com/bachphuong/media/images/book/224r_01.pdf.

The Arizona Republic, Construction quality varies widely, November 20, 2001.

Ausubel, Lawrence M., 1999, Adverse selection in the credit card market, Working paper, University of Maryland.

Beggs, Alan, and Kathryn Graddy, 2009, Anchoring effects: Evidence from art auctions, *American Economic Review* 99, 1027–1039.

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

Berger, Allen N., and Gregory F. Udell, 1995, Relationship lending and lines of credit in small firm finance, *Journal of Business* 68, 351–381.

Bulow, Jeremy, 1982, Durable-goods monopolists, *Journal of Political Economy* 90, 314–332.

Campbell, John Y., Stefano Giglio, and Parag Pathak, 2011, Forced sales and house prices, *American Economic Review* 101, 2108–2131.

Coase, Ronald H., 1972, Durability and monopoly, *Journal of Law & Economics* 15, 143–149.

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

Degryse, Hans, and Steven Ongena, 2005, Distance, lending relationships, and competition, *Journal of Finance* 60, 231–266.

Degryse, Hans, and Steven Ongena, 2008, Competition and regulation in the banking sector: A review of the empirical evidence on the sources of bank rents, in Anjan V. Thakor and Arnoud W.A. Boot, eds.: *Handbook of Financial Intermediation and Banking* (Elsevier Science, Arnsterdom.

Degryse, Hans, and Patrick van Cayseele, 2000, Relationship lending within a bank-based system: Evidence from European small business data, *Journal of Financial Intermediation* 9, 90–109.

Elul, Ronel, 2011, Securitization and mortgage default, Federal Reserve Bank of Philadelphia Working Paper 09–21.

Elyasiani, Elyas, and Lawrence Goldberg, 2004, Relationship lending: A survey of the literature, *Journal of Economics and Business* 56, 315–330.

Fishman, Michael J., and Jonathan A. Parker, 2015, Valuation, adverse selection, and market collapses, Review of Financial Studies 28, 2595–2607.

Garmaise, Mark J., and Tobias J. Moskowitz, 2004, Confronting information asymmetries: Evidence from real estate markets, *Review of Financial Studies* 17, 405–437.

Gartenberg, Claudine, 2011, Tempted by scope? Homebuilder mortgage affiliates, lending quality and the housing crisis, Working paper, New York University.

Gorton, Gary B., and Guillermo Ordonez, 2014, Collateral crises, American Economic Review 04, 343–398.

James, Christopher M., 2010, Mortgage-backed securities: How important is skin in the game? FRBSF Economic Letter 37. Available at http://www.frbsf.org/economic_research/files/e/2010_39.pdf.

Kahneman, Daniel, 1992, Reference points, anchors, norms, and mixed feelings, *Organizational Behavior and Human Decision Processes* 51, 296–312.

Karlan, Dean, and Jonathan Zinman, 2009, Observing unobservables: Identifying information asymmetries with a consumer credit field experiment, *Econometrica* 77, 1993–2008.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2010, Did securitization lead to lax screening? Evidence from subprime loans, *Quarterly Journal of Economics* 125, 307–362.

Kurlat, Pablo, and Johannes Stroebel, 2015, Testing for information asymmetries in real estate markets, *Review of Financial Studies*, 28, 2429–2461.

Lambson, Val E., Grant R. McQueen, and Barrett A. Slade, 2004, Do out-of-state buyers pay more for real estate? An examination of anchoring-induced bias and search costs, *Real Estate Economics* 32, 85–126.

Levitt, Steven D., and Chad 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, Michael S., Ileen B. Malitz, and S. Abraham Ravid, 1993, Trade credit, quality guarantees, and product marketability, *Financial Management* 22, 117–127.

Loutskina, Elena, and Philip E. Strahan, 2011, Informed and uninformed investment in housing: The downside of diversification, *Review of Financial Studies* 24, 1447–1480.

Petersen, Mitchell A., and Raghuram G. Rajan, 2002, Does distance still matter? The information revolution in small business lending, *Journal of Finance* 57, 2533–2570.

The *Phoenix New Times*, Cracked houses: Homes all over Arizona are falling apart, March 16, 2006.

Povel, Paul, and Rajdeep Singh, 2010, Stapled finance, *Journal of Finance* 65, 927–953.

Rosen, Sherwin, 1974, Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy* 82, 34–55.

Smith, Janet K., 1987, Trade credit and informational asymmetry, *Journal of Finance* 42, 863–872.

Tampa Bay Tribune, 2007 lawsuit says builder knew of Willowbrook home defects, September 14, 2012.

Wilhelmsson, Mats, 2008, House price depreciation rates and level of maintenance, *Journal of Housing Economics* 17, 88–101.

Woodward, Susan E., and Robert E. Hall, 2012, Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence, *American Economic Review* 102, 3249–3276.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.