Internet Appendix for "Asymmetric Information about Collateral Values"

JOHANNES STROEBEL*

I. Theoretical Model

In this section I present a theoretical model of the competition between differentially informed lenders to provide mortgage financing. This model formalizes the empirical predictions discussed in Section II of the published article. The model builds on the contributions in von Thadden (2004) and Hauswald and Marquez (2006), as well as Engelbrecht-Wiggans, Milgrom and Weber's (1983) analysis of first-price sealed-bid common value auctions with differentially informed bidders. I first characterize the equilibrium interest rate offers of the integrated and non-integrated lenders. I then simulate the model to generate empirical predictions about each lender's equilibrium collateral quality, and the observed interest rates charged by the non-integrated lenders.[1]

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

Households: Households are risk-neutral and live in either a purchased house or rented housing, the cost of which is normalized to zero. Households have no resources, and require a mortgage to purchase a house. They are indexed by γ, the probability that they will repay the mortgage when the value of their house falls (i.e., $\theta = l$). A household's γ is common knowledge. The household's expected return from borrowing at rate R is equal to $q(H - R) - (1 - q)\gamma R$, which has to be bigger than the cost of renting. The rate $R(\gamma)_m = \frac{qH}{q+(1-q)\gamma}$ is the maximum interest rate that a household would accept.

*Stroebel, Johannes, Internet Appendix for "Asymmetric Information about Collateral Values," Journal of Finance, DOI: 10.1111/jofi.12288. Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the author. Any query (other than missing material) should be directed to the author of the article.

Note that the parameter γ can be interpreted in two ways. First, in a model with only two house price realizations, it can be a reduced-form representation of the borrower's downpayment. In reality, house prices can take on a large number of values. Since negative equity is a necessary condition for default, how far house prices have to decline before borrowers default depends on their downpayment. When borrowers make a larger downpayment, they default in fewer states of the world, because in fewer states of the world do house prices fall by enough to push the borrower into negative equity. Higher values of γ would then correspond to higher downpayments, which, in a world of a continuous house price distribution, would generate a higher probability of repayment when prices fall. In fact, in Section V.C.2 of the published article, I take the borrower's LTV ratio as an empirical proxy for γ. A second interpretation of γ is that it captures observable characteristics of borrowers, such as their liquid assets or wealth, that make it less likely that they default when they have negative equity (see Elul et al. (2010) for a discussion of these factors).

Lenders: There are two types of risk-neutral lenders with access to funds at rate $R_f < qH$: an integrated lender that has some private information about the house, and N non-integrated lenders that only know q. The private information of the integrated lender consists of a nonconveyable signal $\eta \in \{h, l\}$. Signal precision is defined as $\phi = P(\eta = h|\theta = h) = P(\eta = l|\theta = l) > \frac{1}{2}$.

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

Note that this timing assumption makes the game resemble a first-price sealed-bid auction, in which non-integrated lenders are unable to observe the integrated lender's offer, which might have allowed them to infer its signal. While approaching lenders simultaneously may not represent the optimal search strategy for the consumer, who might benefit from shopping around with the integrated lender's offer, it is a reasonable representation of actual mortgage shopping behavior. Woodward and Hall (2012) find that most borrowers consider no more than two offers. The benefits from more search are so large that they conclude that it must be "confusion about how this market works that caused borrowers to shop too little." Another assumption is that borrowers themselves do not extract information from the integrated lender's offer about the quality of the house they purchase. If they did, they would be willing to pay less for a house on which they receive a high-interest offer from the integrated lender. However, since interest rates vary with a large number of characteristics, such signal extraction would be extremely complex and beyond the skills of most borrowers. In fact, the empirical analysis in Section V.B of the published article shows that such a "bounded rationality" assumption seems plausible, since, conditional on observable characteristics, house prices do not differ by the identity of the mortgage lender.

A. Equilibrium

I look for a Bayesian Nash equilibrium. Since the sensitivity of repayment with respect to collateral value, γ, is perfectly observable by all agents, I can solve the equilibrium separately for each value of γ and then compare equilibrium outcomes across γ-types.[2]

THEOREM IA.1: There are no pure strategy equilibria.

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

Note that if a pure strategy equilibrium existed, both lenders would have to offer the mortgage at the same interest rate $\tilde{R}$. If one lender offered credit at a rate lower than the other lender, it could increase its payoff by raising its rate by a small ε. However, both lenders offering the same $\tilde{R}$ cannot be an equilibrium. If, conditional on observing η, it is profitable to lend at $\tilde{R}$, then the integrated lender would offer $\tilde{R} - \varepsilon$ and capture the entire market. If, conditional on η, it is unprofitable to lend at $\tilde{R}$, the integrated lender would increase its interest rate offer and subject the less informed lender to a winner's curse, leaving it with an expected loss.

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

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

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

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

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

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

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

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

The following paragraphs present the proof for Theorem IA.2, as well as intuition for the individual steps. To find the unique mixed strategy equilibrium I follow a number of steps in similar proofs in von Thadden (2004) and Hauswald and Marquez (2006).

Define the probability of observing a positive signal, $\eta = h$, as $P_i(h) = q\phi + (1-q)(1-\phi)$ and the probability of observing a negative signal, $\eta = l$, as $P_i(l) = (1-q)\phi + q(1-\phi)$. The probability that a house is of high quality conditional on observing $\eta = h$ is $p(h, \phi) = Pr(\theta = H|\eta = h) = \frac{q\phi}{q\phi+(1-q)(1-\phi)}$. The probability that the house is of high quality conditional on observing $\eta = l$ is $p(l, \phi) = Pr(\theta = H|\eta = l) = \frac{q(1-\phi)}{(1-q)\phi+q(1-\phi)}$. Define the expected revenue from lending at interest rate R to a type-γ agent wanting to buy a house with signal η as:

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

When collateral values are high (house prices have increased), which happens with probability $p(\eta, \phi)$, all households repay. When collateral values are low, which happens with probability $(1 - p(\eta, \phi))$, the household will only repay with probability γ. The function $z(\eta, \phi, \gamma)$ is the repayment probability of the loan when observing η with precision ϕ.

Define $R(\gamma)_a^b = \frac{R_f}{q+\gamma(1-q)}$ as the break-even interest rate when lending to a type-γ agent to purchase an average house. It sets the cost of making the loan, R_f, which has to be paid in any case, equal to the expected return from the loan, which is the product of the interest rate and the repayment probability. Similarly, $R(\gamma, \phi)_l^b = \frac{R_f}{z(l,\phi,\gamma)}$ is the break-even interest rate for the integrated lender when lending to a type-γ agent who wants to purchase a house when $\eta = l$. This break-even interest rate is increasing in signal precision: when a negative signal becomes more precise, lenders require a higher break-even interest rate to finance the purchase. It is decreasing in γ: when the borrower makes a larger downpayment, lending against lower-quality collateral is profitable at lower interest rates.

Let $F_i(R; \eta, \phi, \gamma)$ represent the cumulative distribution function (cdf) of the integrated lender's distribution of interest rate offers R for a type-γ borrower wanting to buy a house with signal η when the integrated lender's signal precision is ϕ. Let $F_n(R; \phi, \gamma)$ be the cdf of the non-integrated lender's distribution over interest rate offers R for a type-γ borrower wanting to buy a house when the integrated lender's signal precision is ϕ. Both $F_i(R; \eta, x, \gamma)$ and $F_n(R; \phi, \gamma)$ are continuous, strictly increasing, and atomless on a common support $[\underline{R}, \bar{R}]$. For each signal precision ϕ, there is a marginal household with $\gamma = \bar{\gamma}$ to whom it is no longer profitable to lend at the highest possible rate if the collateral signal is negative. This cutoff is defined as the solution to $R(\bar{\gamma}, \phi)_l^b = R(\bar{\gamma})_m$. The parameter $\bar{\gamma}$ is increasing in ϕ. Intuitively, for higher signal precision, the probability that house is truly of low quality $(\theta = l)$ when the integrated lender observes a bad signal $(\eta = l)$ is higher. When the lender is more certain that they are lending against bad collateral, they will only lend to borrowers that make larger and larger downpayments (larger γ). $\bar{\gamma}$ is also decreasing in q and H.

Common Support of the Bidding Distribution: Since a less-informed bidder cannot profit from a sealed-bid auction against a better-informed competitor,[3] the non-integrated lender must break even in equilibrium. This allows us to calculate the lower bound of the support of the common bidding distribution. When offering $\underline{R}$, the non-integrated lender wins almost surely. Since it needs to make a profit of zero, we have $\underline{R} = R(\gamma)_a^b$, the break-even interest rate for lending to a type-γ borrower buying a house of average quality.

The upper bound of the lenders' bidding distribution, $\bar{R}$, depends on γ. When $\gamma \geq \bar{\gamma}$, a repeated undercutting argument similar to Bertrand competition shows that for $\eta = l$ the integrated lender offers $R(\bar{\gamma}, \phi)_l^b$, and makes zero profit. If it offered a higher interest rate $R^* > R(\bar{\gamma}, \phi)_l^b$ with positive probability, the non-integrated lender would be able to profitably undercut it in some states of the world, by always offering R^{**} such that $R^* > R^{**} > R(\bar{\gamma}, \phi)_l^b$, and generate a positive profit in expectation. When $\eta = h$, the integrated lender mixes offers on the support of $[R(\gamma)_a^b, R(\gamma, \phi)_l^b]$. When $\gamma < \bar{\gamma}$, the integrated lender never makes an offer if $\eta = l$ (it would make a loss in expectation). For any ϕ, the common support is thus given by $[R(\gamma)_a^b, \min\{R(\gamma, \phi)_l^b, R(\gamma)_m\})$. The min-operator ensures that no lender will ever offer an interest rate larger than what borrowers would be willing to accept.

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

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

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

$$\begin{aligned} \pi_u(R; \phi, \gamma) &= [(\text{Prob. } i \text{ has } \eta = l) \times (\text{Expected profit when } i \text{ has } \eta = l)] + \\ &\quad [(\text{Prob. } i \text{ has } \eta = h) \times (\text{Prob. of winning}) \times \\ &\quad (\text{Expected Profit when } i \text{ has } \eta = h)] \\ &= P_i(l)[W(R; l, \phi, \gamma) - R_f] + P_i(h)[1 - F_i(R; h, \phi, \gamma)][W(R; h, \phi, \gamma) - R_f]. \end{aligned} \tag{IA.2}$$

Since the non-integrated lender cannot make a profit, we have that $\forall (R, \gamma) : \pi_u(R; \phi, \gamma) = 0$. In addition, since the mixing distributions are strictly increasing, equilibrium profit for each lender must be the same for every interest rate offered on the support of the offer distribution: $\pi_i(R; h, \phi, \gamma) = \bar{\pi}(\phi, \gamma)$.

If we now evaluate $\pi_i(R; h, \phi, \gamma)$ at the lower bound of the support, since $F_n(R(\gamma)_a^b; \phi, \gamma) = 0$, we have that $\bar{\pi}(\phi, \gamma) = W(R(\gamma)_a^b; h, \phi, \gamma) - R_f$ from equation (IA.1). Plugging this back into equation (IA.1) and solving for $F_n(R; \phi, \gamma)$ gives:

$$F_n(R; \phi, \gamma) = 1 - \frac{W(R(\gamma)_a^b; h, \phi, \gamma) - R_f}{W(R; h, \phi, \gamma) - R_f}. \tag{IA.3}$$

Similarly, solving equation (IA.2), by setting $\pi_u(R; \phi, \gamma) = 0$, gives:

$$F_i(R; h, \phi, \gamma) = 1 + \frac{P_i(l)[W(R; l, \phi, \gamma) - R_f]}{P_i(h)[W(R; h, \phi, \gamma) - R_f]}. \tag{IA.4}$$

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

Probability of offer when $\gamma < \bar{\gamma}$: The integrated lender never bids when it receives a negative signal ($\eta = l$). When it receives a positive signal ($\eta = h$), $F_i(\bar{R}; h, \phi, \gamma) = F_i(R(\gamma)_m; h, \phi, \gamma) < 1$. For a lender to not make an interest rate offer in some instances, it must be indifferent between bidding and not bidding. Since the integrated lender can make a profit in expectation when $\eta = h$, it is never indifferent between bidding and not bidding, which generates expected profits of zero. Thus, unlike the non-integrated lender, which is indifferent between bidding and not bidding, the integrated lender will never not bid when $\eta = h$. Hence, the integrated lender randomizes over $[R(\gamma)_a^b, R(\gamma)_m)$ for $\eta = h$ houses, without any atoms, but with point mass at $R(\gamma)_m$, where the mass is equal to $1 - F_i(R(\gamma)_m; h, \phi, \gamma)$. The non-integrated lender bids with probability $F_n(R(\gamma)_m; \phi, \gamma) < F_i(R(\gamma)_m; h, \phi, \gamma) < 1$ for all agents. With probability $1 - F_n(R(\gamma)_m; \phi, \gamma)$ the non-integrated lender does not make an interest rate offer and the household is rationed.

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

B. *Empirical Predictions from Equilibrium Bank Behavior*

To analyze equilibrium outcomes when lenders use the mixed strategies of Theorem IA.2, I simulate the game for a range of parameter values. This generates predictions about the expected quality of the equilibrium collateral portfolio of each lender, about the equilibrium interest rates, and about how these outcomes vary with different values of γ, the probability of repayment when collateral values fall, and ϕ, the signal precision.[4]

The top row of Figure IA.1 plots the expected period-2 value of the equilibrium portfolios of houses financed by the two lenders as a function of ϕ and γ. The dashed line represents the integrated lender's portfolio, the solid line the non-integrated lender's portfolio. The dotted line shows the unconditional expected house value, qH. For all values of γ and ϕ, the houses financed by the integrated lender are more likely to increase in value than those financed by the non-integrated lender. This is a direct result of the integrated lender

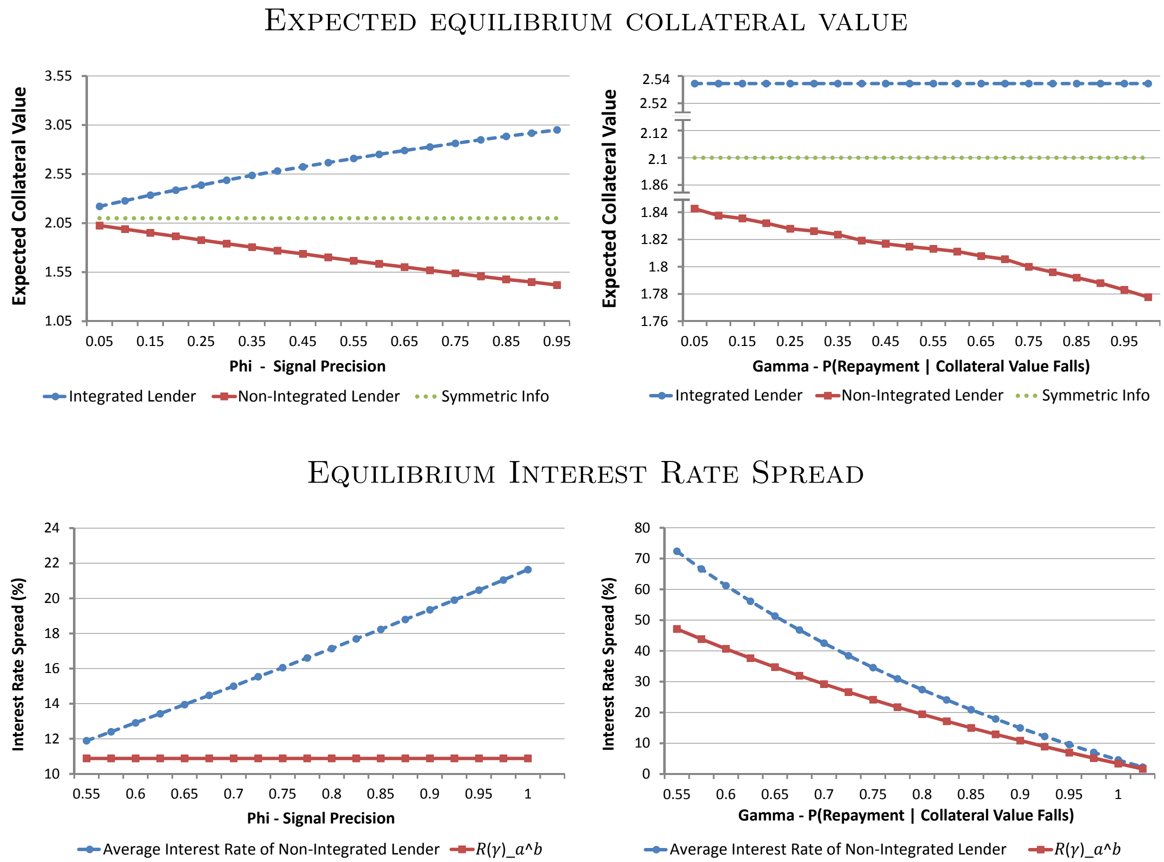

Figure IA.1. Equilibrium model outcomes. The top row plots the expected period-2 price of a house in the integrated lender's equilibrium collateral portfolio (dashed line), the expected period-2 price of a house of average quality (dotted line), and the expected period-2 price of a house in the non-integrated lender's equilibrium collateral portfolio (solid line). The bottom row plots spreads of the average interest rate charged by the non-integrated lender over R_f (dashed line), and the break-even rate when lending against average quality collateral, $R(\gamma)_a^b$, over R_f (solid line). In the left column ϕ varies along the horizontal axis. In the right column γ varies along the horizontal axis. If both lenders offer the same interest rate, I resolve the indifference in favor of the non-integrated lender. The model parameters are: $H = 3; q = 0.7; R_f = 1.1$. I set $\gamma = 0.7$ in the left panel and $\phi = 0.7$ in the right panel.

conditioning its interest rate offers on the informative signal and the subsequent adverse selection. This implication is formalized in Prediction 1, reproduced here.

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 non-integrated lender's information set) homes financed by non-integrated lenders; homes financed by the integrated lender* *also experience fewer foreclosures.*

The bottom row of Figure IA.1 plots the average interest rate spread over R_f for the non-integrated lender's mortgages (dashed line). It also shows the spread of $R(\gamma)_a^b$, the break-even interest rate for lending against average quality collateral (solid line). When lenders are equally informed about collateral quality, Bertrand competition drives interest rates to $R(\gamma)_a^b$. When competing against a better-informed integrated lender, a non-integrated lender lends against below-average-quality collateral and must charge a higher interest rate to continue to break even. This is formalized in Prediction 2.

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

The left column of Figure IA.1 shows how equilibrium outcomes vary with ϕ, the precision of the integrated lender's signal. The top left panel shows that the expected period-2 value of houses financed by the integrated lender is increasing in ϕ: as the signal becomes more precise, the integrated lender is better at identifying high-quality collateral. The non-integrated lender correspondingly lends against lower-quality collateral. To continue to break even, it needs to charge a higher interest rate on the mortgages it makes, as shown in the bottom left panel. As discussed in Section I of the published article, the prices of homes built on expansive soil are particularly sensitive to initial construction quality. Therefore, for those homes, the integrated lender's signal about future home values is particularly precise. These insights are formalized in the following predictions:

PREDICTION 1(a): *Among homes built on expansive soil, the integrated lender's information about future returns is more precise (high ϕ), 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 (high ϕ), and the increase in the interest rate charged by non-integrated lenders when competing against an integrated lender is larger.*

The right column of Figure IA.1 shows how equilibrium outcomes vary with γ, the sensitivity of the mortgage default probability with respect to changes in collateral value. In the empirical implementation I use the mortgages' loan-to-value ratio to proxy for γ: for mortgages with a high loan-to-value ratio (and a low downpayment), small movements in collateral value are sufficient to generate incentives for default. Borrowers with a large downpayment will only default in the event of very large drops in collateral values. The top right panel shows that the return of the integrated lender's collateral is unaffected by γ, since the integrated lender only lends when $\eta = h$. The return of the non-integrated lender's collateral declines as repayment becomes less sensitive to collateral value. To follow the intuition for this result it is important to realize that mortgage lenders only care about collateral values to the extent that they influence the repayment probability of the mortgage. When γ is low, and the repayment probability is highly dependent on the value of the collateral, the non-integrated lender is particularly concerned about adverse selection on collateral quality. As a result it offers mortgages at higher interest rates to avoid the winner's curse ("bid shading"), as shown in the bottom right panel. As default probabilities become less sensitive to collateral value, the break-even spread charged by the non-integrated lender declines. Since the integrated lender continues to exploit its superior information to the fullest degree, for larger values of γ the non-integrated lender's equilibrium collateral is of lower quality. Put differently, the less the non-integrated lender shades its bid, the lower the quality of its equilibrium collateral portfolio. These insights are formalized in the following empirical predictions:

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

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

II. Data Appendix

I begin with a data set that contains 3.34 million ownership-changing deeds recorded in Arizona between 2000 and 2011. The data include both arms-length market transactions and transfers in divorce, estate settlements, and foreclosures. For each deed with sufficient information to uniquely identify the property, the address is geocoded to determine the property's precise location. For 91.7% of the deeds the address information is sufficiently detailed to determine the exact latitude and longitude. For another 2.1% of the deeds the street number is missing and a latitude and longitude is assigned that locates the property at the geographic midpoint of the street. The 6.2% of the deeds with insufficient address information to assign a location are dropped (many of them refer to the sale of vacant land). I then merge each deed via its assessor parcel number (APN) and county to the underlying property's tax assessment record for 2010.

A. Soil Data

In the next step, I use detailed data on the geographic distribution of soil types from the U.S. Department of Agriculture's (USDA) Soil Survey database, combined with the latitude and longitude of each property, to determine which houses are built on expansive soil. The underlying ArcGIS shape files for soil distribution come from `http://soildatamart.nrcs.usda.gov/County.aspx?State=AZ`. 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.[5] Soil expansiveness has significant geographic variation. Figure IA.2 is produced by the USDA's Natural Resources Conservation Service, and shows the distribution of soil types in the Phoenix region, which makes up most of my sample.

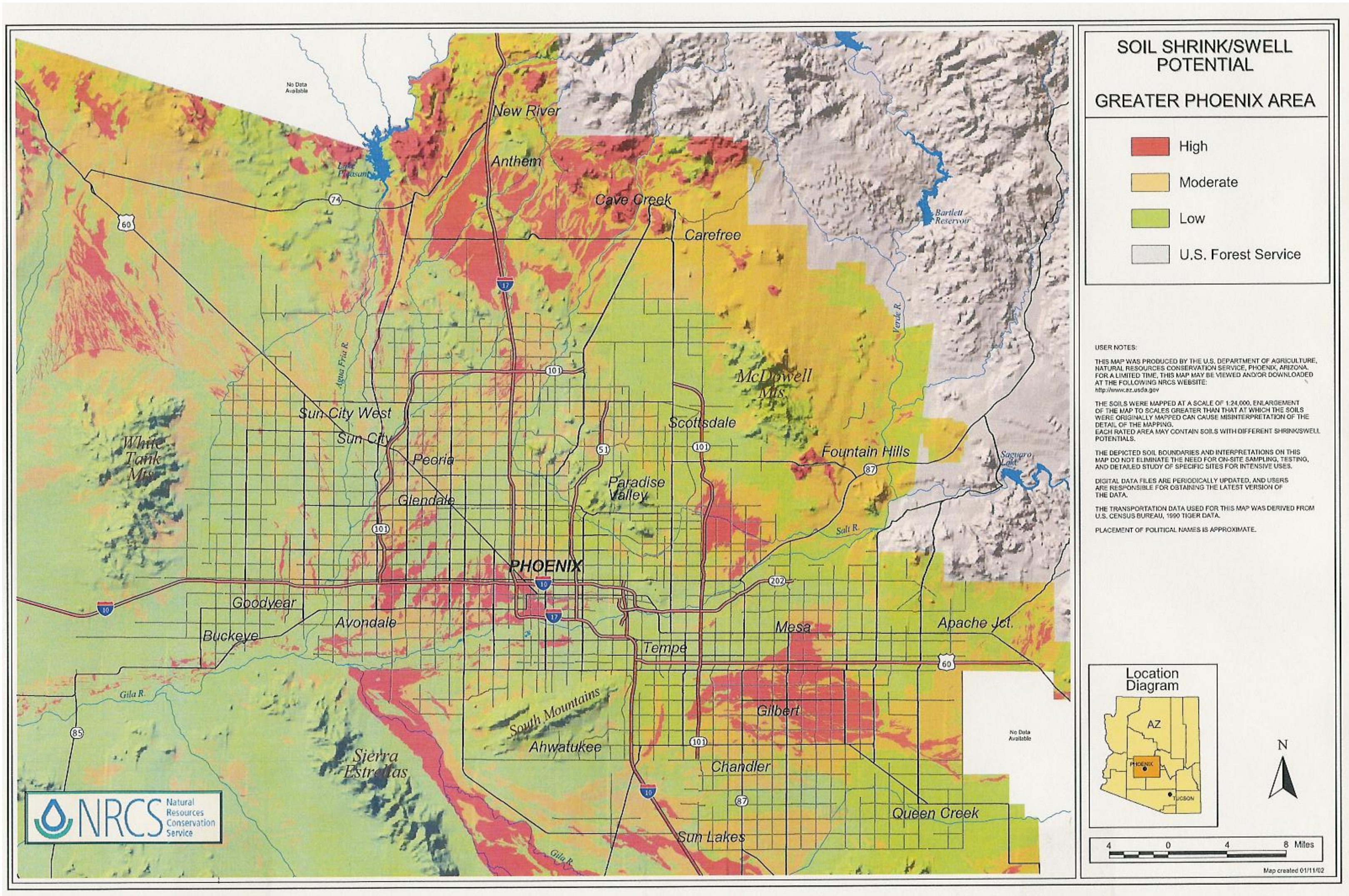

Figure IA.2. Map showing soil distribution in Phoenix.

Figure IA.3 presents a map of two representative housing developments in Arizona, and shows that there is significant within-development variation in soil type. Each blue circle (•) and red cross (+) represents a sale by one of the two developers that appear in my data set. The right panel also presents the soil type for each house. Houses built on the light gray, striped land are built on expansive soil while houses built on the dark green land are not built on expansive soil.

Figure IA.3. Map showing two representative developments and soil type

B. Deeds Data to HMDA Merge

I next merge the deeds to data from the Home Mortgage Disclosure Act's (HMDA) Loan Application Registry (LAR). This allows me to obtain additional characteristics of the home owners, as well as information on the subsequent securitization of mortgages. The LAR is a mortgage-level data set and identifies a mortgage by year, census tract, mortgage amount, and mortgage lender.[6] Bayer et al. (2011) use these characteristics to merge a data set similar to my deeds data to the LAR. This procedure allows them to uniquely match about 70% of all sales. I use additional characteristics to improve match rates and quality. First, both the deeds and HMDA data report whether mortgages are FHA-insured or VA-guaranteed. Second, HMDA data identify whether a house is purchased as a rental property, while the assessor data provide information about whether the property is owneroccupied in 2009. Third, HMDA data contain information about whether the mortgage was applied for by a male, female, or two applicants. The deeds data also identify purchasers as male, female, or a married couple. Fourth, the HMDA data contain information about the race and ethnicity of applicants. In the deeds data I do not have this information, but I do observe the names of buyers. I match the surnames of buyers to the 1,000 most common Asian and Latino surnames from the 2000 U.S. Census. Using these four additional characteristics allows me to confirm $64,947$ unique matches. Despite the use of additional match variables, my unique match rate is lower than that reported by Bayer at al. (2011). There are a number of reasons for this. First, since integrated lenders make a significant number of mortgages in new developments, the power of using lender identity to merge deeds to HMDA data declines. Second, lenders in new developments might be more likely to fall below the asset reporting threshold. For my main data set, for those mortgages where more than one match is possible, I match each deed randomly to one of the possible records in the HMDA data. I can merge a total of $102,818$ deeds to HMDA data. In a previous version of the paper I show that the key empirical results are robust to considering (i) only the sample of houses with a unique HMDA merge, and (ii) the full sample of houses in my data, without requiring an HMDA merge and without conditioning on owner characteristics.

C. Data Cleaning + Identifying Transaction Types

Arms-length Transactions: I identify all deeds that contain information about arms-length transactions in which both buyer and seller act in their best economic interest. This ensures that transaction prices reflect the market value of the property. I include all deeds that are one of the following: "Grant Deed," "Condominium Deed," "Individual Deed," "Warranty Deed," "Joint Tenancy Deed," "Special Warranty Deed," "Limited Warranty Deed," and "Corporation Deed." This excludes intra-family transfers and foreclosures. I drop all observations that are not a Main Deed or only transfer partial interest in a property. This leaves 1.73 million arms-length transactions.

Newly Developed Single-Family Residences: Among the arms-length transactions I identify mortgage-financed purchases of newly developed properties. This includes all deeds in which the seller is identified as a company or partnership, but that are not real-estate-owned (REO) resales (i.e., sales by a bank following a foreclosure). I exclude sales in which the construction date of the house (as reported in the assessor data) precedes the sales date by more than two years. These transactions usually involve a developer that renovates and resells existing properties. I also exclude transactions in which the buyer is identified as a company. I also only consider single-family residences, which make up about 85% of newly developed properties in Arizona. This leaves 240,803 observations. For each newly developed property I collect subsequent arms-length sales to track their future return.[7]

Divorce and Death: I identify those repeat sales pairs for which I observe a divorce or death of the owners up to six months before the second sale. I identify divorces through the presence of an “Intra-Family Transfer & Dissolution” deed that transfers property rights from initially joint ownership to one of the initial owners. A death of an owner is identified if either (i) the seller on a deed is classified as an “Estate,” “Executor,” “Deceased,” or “Surviving joint owner,” or (ii) if I observe “Affidavit of Death of Joint Tenant,” or “Executor’s Deed.”

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

Data Cleaning: I identify houses in the same development by combinations of seller identity and census tract. I only consider houses that were first sold before 2008 and are located in developments with more than 30 units. I drop a few observations that are likely to have misreported loan or sales price details (i.e., when the sales price is less than $25,000 or more than $10 million and when the LTV ratio is more than 1.3 or less than 0.3). In addition, I only keep observations with a full set of control variables in the assessor data.[8] This leaves 158,785 observations.

D. *Identifying Integrated Lenders*

To identify integrated lenders, I follow a number of steps. First, developers usually own their integrated lenders (e.g., the developer “Shea Homes” owns ”Shea Mortgage”). For each developer, I determine whether there is joint ownership with its largest lender, using OneSource North American Business Browser and SEC filings. If I can confirm joint ownership, I assign the lender to be the integrated lender of this developer. This procedure allows me to identify 45,266 mortgages granted by integrated lenders. I also analyze instances in which the market share of a single lender in a development exceeds 50%, but the developer does not own this lender. In these cases, I also assign the lender to be integrated, which assigns another 18,550 transactions as having mortgages granted by an integrated lender. Using this process to identify integrated lenders, 85.1% of newly built houses are in a development with an integrated lender. For houses in developments with an integrated lender, the integrated lender has a market share of 72.9%. I believe that this process of identifying integrated lenders is appropriate: when analyzing the distribution of the market share of the largest lender for lending to purchase *existing* homes, I find that there are essentially no census tracts in which the largest lender has a market share in excess of 35%. Consequently, in any development in which a lender attains more than 50% of all mortgages, it is very likely that this lender is only able to obtain such a market share through an integrated lender arrangement.[9]

E. Sample Overview

Table IA.I shows how the observations in my data are distributed over time and across counties. It includes all observations with a successful HMDA merge and the full set of covariates. Summary statistics are split up for developments with and without an integrated lender. For developments with an integrated lender, the results are given separately for the integrated lender and for non-integrated lenders. The top panel shows that the majority of observations are from Maricopa and Pinal counties, which constitute the Phoenix MSA. Pima county (including Tuscon) only contributes a few observations. This is because for Pima I only observe building and lot size for a small number of observations in the assessment data. These variables are important controls in my main specifications. In order to estimate all models with a common sample, observations with missing data on home characteristics were dropped.[10] The bottom panel shows the distribution of observations by year of sale. The number of newly developed properties sold increased up to 2005, the peak of Arizona's housing boom, and then declined markedly during the financial crisis.

F. Functional Form of Control Variables

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

Borrower and Financing Characteristics: I control for income by adding dummy variables for 50 equally-sized groups. The loan-to-income (LTI) ratio is included by adding dummy variables for mortgages with LTI ratio ≤ 1.5, between 1.5 and 2, between 2 and 2.5, between 2.5 and 3, between 3 and 3.5, and > 3.5. The loan-to-value (LTV) ratio is included by adding dummy variables for mortgages with an LTV ratio $\leq 80\%$, between 80% and 90%, between 90% and 97%, and $> 97\%$. Age is controlled for by adding dummies in buckets of two years.

Table IA.I
Number of Observations by County and Year

This table shows the number of observations in the primary dataset used in the published article. It includes observations with a successful HDMA merge and a full set of covariates.

	No Integrated Lender		Has Integrated Lender				*Total*
			Integrated Lender		*Other Lender*		
	No.	%	No.	%	No.	%	No.
County							
Cochise	94	63.9	35	23.8	18	12.2	147
Coconino	154	100.0	0	0.0	0	0.0	154
Maricopa	12,367	15.1	50,263	61.3	19,377	23.6	82,007
Mohave	27	26.0	52	50.0	25	24.0	104
Pima	9	20.0	20	44.4	16	35.6	45
Pinal	1,113	6.1	13,184	71.7	4,081	22.2	18,378
Yavapai	609	70.8	171	19.9	80	9.3	860
Yuma	963	85.8	91	8.1	69	6.1	1,123
Total	15,336	14.9	63,816	62.1	23,666	23.0	102,818
Year Sold							
2000	1,896	20.4	5,327	57.3	2,075	22.3	9,298
2001	2,088	17.7	7,596	64.4	2,111	17.9	11,795
2002	1,759	16.0	7,231	65.7	2,009	18.3	10,999
2003	2,060	16.2	8,083	63.5	2,582	20.3	12,725
2004	2,700	16.4	9,238	56.0	4,567	27.7	16,505
2005	2,729	17.0	9,155	57.2	4,134	25.8	16,018
2006	1,320	9.4	8,792	62.8	3,877	27.7	13,989
2007	784	6.8	8,394	73.1	2,311	20.1	11,489
Total	15,336	14.9	63,816	62.1	23,666	23.0	102,818

Census Tract Demographics: I control for the median income as well as the proportion of adults over 25 with at least a high school diploma. This information comes from the 2005 to 2009 estimates of the American Community Survey. I control for census tract demographics by including dummy variables for the following (roughly equally sized) median income groups: $\leq$ \$35k, \$35k to \$50k, \$50k to \$65k, \$65k to \$75k, \$75k to \$100k, and $\geq$ \$100k. Dummy variables for high-school graduation rates are: $\leq 75\%, 75\%$ to $80\%, 80\%$ to $90\%, 90\%$ to 95%, and $\geq 95\%$.

G. *Tax Assessment Process in Arizona*

Arizona Revised Statutes (A.R.S.) 42-11054 (C) requires that tax assessors annually compute the so-called “full cash value” of each residential property. A.R.S. 42-11001(6) specifies the full cash value to be “synonymous with market value, which means the estimate of value that is derived annually by using standard appraisal methods and techniques.” The full cash value provided in the tax assessment records is set at 82% of the assessed market value for residential properties. The procedure for arriving at these valuations is described by the assessor of Mohave County, Arizona, as follows: "Between January and March of each year, the Assessor's Office is required (by Arizona State Statute) to notify property owners of their assessed values for the following tax year. For residential and land parcels, this is accomplished by first collecting sales data in the area in which a property is located. Elements of comparability such as location, view, size, quality and condition are taken into consideration, and a mass appraisal mathematical model is used to arrive at each parcel's value. The market is driven by actual sales that have occurred in a time window established by Department of Revenue guidelines. Increases or decreases in sale prices impact the final assessed valuation."

A number of procedures allow homeowners to challenge a tax assessment if they feel the house was valued too highly. The appeals process provides a mechanism through which the assessor obtains information about differential depreciation of housing units. In 2009 there were 19,801 assessment appeals in Maricopa County, up from 17,213 in 2008 (The Arizona Republic (2009)). Overall, about 1.3% of valuations are appealed annually. In 2008, Maricopa County assessments were reduced by a total of $3.9 billion. In the following, I test how well assessed values in Arizona capture true market values. I analyze all properties that were sold in an arms-length transaction between January and March 2009. I compare the transaction price with the assessed value in January 2009. In Figure IA.4, each dot represents such a transaction, with assessed values on the horizontal axis, and transacted values on the vertical axis. The solid line is the 45° line – if assessments were 100% correct, all observations would lie on this line. It is not surprising that there is a significant spread around the 45° line. Unlike homogeneous goods such as stocks and bonds, houses are heterogeneous assets that are sold in a search market. By adjusting the time that a seller is prepared to wait, she can influence the final transaction price. The dashed line represents the prediction from an OLS regression. The fact that it is very close to the 45° line suggests that, on average, assessed values capture current market values reasonably well.

H. Quantile Regression Analysis

Figure 2 in the published article presents evidence that the mean return difference between mortgages financed by the integrated and non-integrated lenders comes from a longer left tail of the return distribution for integrated lender mortgages. In this section, I conduct a quantile regression analysis to formalize these conclusions. A quantile regression allows me to consider how the quantiles of the conditional return distribution differ across

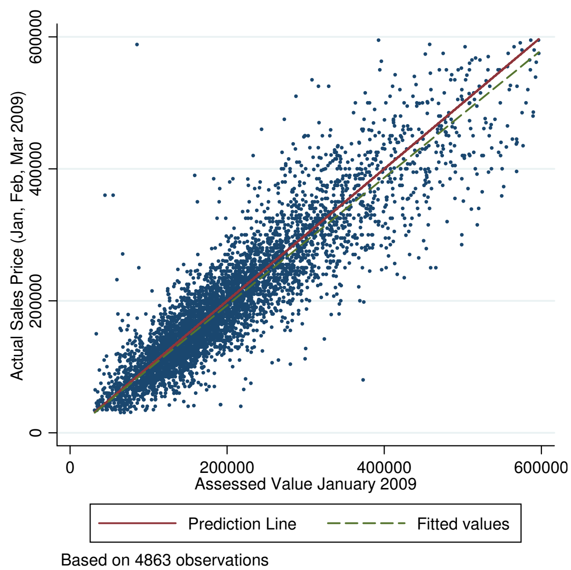

Figure IA.4. Quality of assessment values. This figure tests for the accuracy of the estimated market value in the assessment data. Each dot represents an observation of a house that was sold in the first three months of 2009 and for which I observe an assessed value in January 2009. On the horizontal axis is the assessed value and on the vertical axis the corresponding transaction price. The solid line represents the 45° line. The dashed line represents the linear prediction of a regression of sales price on assessed value.

mortgages made by the integrated and non-integrated lenders:

$$Return_i^{\text{QUANTILE}} = \alpha + \kappa IL_i + X_i\beta + \delta_{q_1,q_2} + \epsilon_i \tag{IA.1}$$

The results of regression (IA.1) are presented in Table IA.II. Column (1) shows that the 75^{th} percentile of the conditional return distribution is the same for mortgages made by integrated and non-integrated lenders. As the quantiles get smaller (i.e., as we consider homes with lower and lower conditional returns), the difference between homes financed by integrated and non-integrated lenders increases. At the 5^{th} percentile of the conditional return distribution, houses financed by the integrated lender outperform by an annualized 1.5 percentage points. At the 0.1^{th} percentile, they outperform by an annualized 10.3 percentage points.

Table IA.II
Quantile Regression Analysis

This table show results from quantile regression (IA.2) for different quantiles. Control variables and fixed effects are as in Column 5 of Table II in the published article. Standard errors are bootstrapped with 2,000 replications. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

	(1)	(2)	(3)	(4)	(5)	(6)	(6)
Percentile	0.75	0.50	0.25	0.10	0.05	0.01	0.001
Has Integrated Lender	0.038	0.218***	0.491***	1.065***	1.553***	3.791***	10.30**
	(0.069)	(0.061)	(0.079)	(0.125)	(0.215)	(0.976)	(4.664)
Controls	✓	✓	✓	✓	✓	✓	✓
$\bar{y}$	7.450	7.450	7.450	7.450	7.450	7.450	7.450
N	30,398	30,398	30,398	30,398	30,398	30,398	30,398

The results confirm that the average annualized outperformance of the integrated lender does indeed come from the integrated lender's homes experiencing large price declines at a lower frequency than homes financed by non-integrated lenders. As discussed in Section IV.C of the published article, this skewness also explains why the relatively modest average return differences lead to significantly lower foreclosure rates among homes financed by the integrated lender.

I. Importance of Construction Quality

Predictions 1(b) and 2(b) in Section II of the published article posit that the impact of adverse selection on equilibrium interest rates and relative collateral returns depends on the importance of construction quality in determining mortgage default. The parameter capturing this relationship is modeled as γ in Section I of this Internet Appendix. In the main empirical implementation, I use the loan-to-value ratio to proxy for γ (see Sections V.C.2 and VI of the published article). From the perspective of a lender, the construction quality of the housing collateral is more important when the borrower makes a smaller downpayment. Conversely, when the downpayment is large, non-integrated lenders increase interest rates less when competing with integrated lenders; they consequently face more adverse selection, but the resulting difference in collateral quality has a smaller impact on their profit from making the mortgage.

In this section, I discuss results using a different proxy for the importance of construction quality in determining default. In particular, properties differ by the share of total value comprised of the structure and the land. Adverse selection about construction quality only affects the value of the structure. Consequently, construction quality has a greater percentage impact on property values when the structure is a larger component of total value. Similar to Prediction 2(b), one would thus expect banks to increase interest rates more when competing with an integrated lender to finance a house with a significant structure component. In contrast, when land comprises most of the value of the property, overall capital gains are less affected by movements in the value of the structure, and non-integrated lenders respond less to the adverse selection.

It is theoretically unclear whether the relative outperformance of properties financed by the integrated lender is larger or smaller for high-structure-value properties. On the one hand, as the value of the structure makes up a larger part of total value, a lender that can better predict construction defects should generate higher outperformance. On the other hand, as discussed above, for those properties the non-integrated lender also responds by increasing his interest rate more, reducing the impact of the asymmetric information on equilibrium collateral quality. Which effect dominates is an empirical question.

To test this, I need to construct a measure of the share of structure and land in total property value. I do so by assuming that constructing a house is equally expensive in all parts of Phoenix (i.e., that labor and materials are equally expensive). As a result, variation in the price per square foot is driven by variation in the land value. When the price per square foot is lower, the structure component is a more important contributor to overall value.[11]

I test how the integrated lender's outperformance, and the non-integrated lenders' interest response, differ by the (inverse of the) price per square foot of the property. This variable, which I call "structure share," captures the number of square feet one can purchase for \$1, and has a mean of 0.035, a standard deviation of 0.015, and a p5-p95 range of 0.016 to 0.060. The results are presented in Table IA.III. I control for the structure share by including it linearly, as well as by including dummy variables for tertiles of the distribution; this shows the robustness of the results to a variety of parametric assumptions. Columns (1) to (3) present adjustments to regression (9) in the published article. These results show that non-integrated lenders do indeed increase interest rates more to compete with integrated lenders for properties when the structure share of total value is larger. Columns (4) to (6) present adjustments to regression (1) in the published article. Among homes with a large structure share, non-integrated lenders use their interest rate adjustment to assemble a relatively higher-quality collateral portfolio. The equilibrium collateral return is thus closer to that of the integrated lender than when the structure share component is large. These findings provide further confidence in the importance of asymmetric information about collateral values for explaining my findings.

Table IA.III
Impact of Structure Share

Columns (1) to (3) show results from regression (9) in the published article. The dependent variable is the mortgage interest rate. I include single-family residences sold by a developer in Arizona in the 2000 to 2007 period that were financed by non-integrated lenders. Columns (4) to (6) show results from regression (1) in the published article. The dependent variable is the annualized return of houses between two arms-length transactions. I include single-family residences first sold by a developer in the 2000 to 2007 period in developments with an integrated lender. Control variables and fixed effects are as indicated. Standard errors in Columns (1) to (3) are clustered at the lender level, and in columns (4) to (6) at the developer level. Significance levels: * (p<0.10), ** (p<0.05), *** (p<0.01).

	Interest Rate			Return - Period A		
	(1)	(2)	(3)	(4)	(5)	(6)
Has Integrated Lender	-0.005 (0.059)	-0.029 (0.077)				
Structure Share	0.656 (1.433)	0.128 (1.992)		56.33*** (16.38)	46.11*** (15.24)	
Has Integrated Lender × Structure Share	2.854* (1.498)	3.850** (1.556)				
Structure Share 1st Tertile			0.054 (0.043)			-1.427*** (0.360)
Structure Share 2nd Tertile			0.026 (0.039)			-0.353* (0.190)
Has Integrated Lender × Structure Share 1st Tertile			0.045 (0.049)			
Has Integrated Lender × Structure Share 2nd Tertile			0.076 (0.049)			
Has Integrated Lender × Structure Share 3rd Tertile			0.181*** (0.057)			
Integrated Lender				1.106*** (0.305)	0.980*** (0.280)	
Integrated Lender × Structure Share				-18.09*** (6.555)	-15.69** (6.420)	
Integrated Lender × Structure Share 1st Tertile						0.688*** (0.252)
Integrated Lender × Structure Share 2nd Tertile						0.355* (0.189)
Integrated Lender × Structure Share 3rd Tertile						0.281** (0.123)
Control Variables	Col (4) Table XII	Col (5) Table XII	Col (4) Table XII	Col (4) Table II	Col (5) Table II	Col (4) Table II
R^2	0.591	0.597	0.592	0.887	0.896	0.887
$\bar{y}$	6.638	6.638	6.638	7.450	7.450	7.450
N	15,620	15,620	15,620	30,398	30,398	30,398

J. Hedonic Regression Coefficients

In this section, I present the full set of coefficients on the control variables in the hedonic regression conducted in Section V.B in the published article. The coefficients are shown in Table IA.IV. Transaction price is monotonically increasing in building size and lot size. Houses with a pool trade at a 3.5% premium. Conditional on home size, the number of bedrooms and bathrooms is not systematically related to the value of the house. Houses built on expansive soil trade at a 2.5% discount to otherwise identical houses not built on expansive soil. There is no initial price difference between houses financed by integrated and non-integrated lenders.

K. Rejections and Expansive Soil

Section IV.D of the published article shows that those houses that were rejected for a mortgage by the integrated lender subsequently underperformed; rejections by the non-integrated lender, on the other hand, did not contain information about the quality of the collateral. Section V.A.3 of the published article shows that houses financed by an integrated lender outperformed more when they were built on expansive soil, where asymmetric information about collateral quality was more important.

In this section, I consider whether the information on which mortgage applications were rejected by the integrated lender are more informative for subsequent returns for those houses built on expansive soil. To do so, I repeat the key analyses presented in Tables V and VI of the published article, restricting the sample to houses built on expansive soil. The results are presented in Table IA.V. The sample size drops significantly, since only about 10% of all properties are built on expansive soil. The magnitude of the correlation between being rejected by an integrated lender and subsequent mortgage default, estimated in column (1), is significantly larger among the sample of houses built on expansive soil. The relationship between rejection by an integrated lender and subsequent price appreciation is about twice as large, but no longer statistically significant in this smaller sample. Column (3) focuses on the sample of houses whose buyer applied for a mortgage with the integrated lender (see Table VI). The correlation between being rejected by an integrated lender and subsequent price appreciation is again an order of magnitude larger in the sample of houses built on expansive soil than it is in the full sample. As before, and as shown in column (4), whether a mortgage applicant was rejected by a non-integrated lender is not predictive of the subsequent capital gains of the housing collateral.

Table IA.IV
Hedonic Regression Table

This table shows coefficients and standard errors from a hedonic regression analyzed in Section V.B of the published article: $LogPrice_{i,q_1} = \alpha + \kappa IL_i + X_i\beta + \delta_{q_1,Development} + \epsilon_i$. There are 133,614 observations. The regression has an R^2 of 94%. Standard errors are clustered at the developer level.

Variable	Coefficient	SE	Variable	Coefficient	SE
Building Size			**Lot Size**		
5 pctl - 10pctl	0.060	0.004	5 pctl - 10pctl	0.034	0.009
10 pctl -1 5pctl	0.088	0.004	10 pctl -1 5pctl	0.050	0.013
15 pctl - 20ctl	0.118	0.004	15 pctl - 20ctl	0.046	0.014
20 pctl - 25pctl	0.147	0.006	20 pctl - 25pctl	0.063	0.013
25 pctl - 30pctl	0.161	0.004	25 pctl - 30pctl	0.067	0.014
30 pctl - 35pctl	0.187	0.005	30 pctl - 35pctl	0.076	0.013
35 pctl - 40pctl	0.210	0.006	35 pctl - 40pctl	0.080	0.014
40 pctl - 45pctl	0.241	0.006	40 pctl - 45pctl	0.093	0.015
45 pctl - 50pctl	0.250	0.007	45 pctl - 50pctl	0.090	0.015
50 pctl - 55pctl	0.287	0.007	50 pctl - 55pctl	0.104	0.016
55 pctl - 60pctl	0.297	0.006	55 pctl - 60pctl	0.113	0.016
60 pctl - 65pctl	0.311	0.010	60 pctl - 65pctl	0.120	0.016
65 pctl - 70pctl	0.344	0.010	65 pctl - 70pctl	0.139	0.017
70 pctl - 75pctl	0.379	0.009	70 pctl - 75pctl	0.149	0.017
75 pctl - 80pctl	0.420	0.012	75 pctl - 80pctl	0.168	0.017
80 pctl - 85pctl	0.467	0.011	80 pctl - 85pctl	0.192	0.017
85 pctl - 90pctl	0.500	0.012	85 pctl - 90pctl	0.217	0.019
90 pctl - 95pctl	0.553	0.016	90 pctl - 95pctl	0.262	0.021
95 pctl - 100pctl	0.642	0.019	95 pctl - 100pctl	0.357	0.022
Number of Rooms			**Has Pool**	0.035	0.002
2	0.005	0.039			
3	0.023	0.028	**Garage number of Cars**		
4	-0.040	0.029	1	-0.023	0.012
5	0.019	0.020	2	-0.016	0.010
6	0.028	0.019	3	0.032	0.009
7	0.014	0.018	4	0.064	0.014
8	0.001	0.018			
9	-0.004	0.017	**Expansive Soil**	-0.025	0.015
10	0.005	0.020			
			Integrated Lender	0.001	0.002
Number of Bathrooms					
2	0.053	0.039			
3	0.114	0.019			
4	0.122	0.019			
5	0.135	0.022			
6	0.188	0.025			

Table IA.V
Effect of Rejections - Expansive Soil Properties

Column (1) shows average marginal effects from probit regression (4) in the published paper, where the dependent variable is whether a foreclosure was observed within three years of purchase. Controls as in Table IV of the published article. Columns (2) to (4) show results from regression (1) in the published paper, where the dependent variable is the annualized return of houses between two arms length transactions. Controls are as in Table II. I include single-family residences sold by a developer in the 2000 to 2007 period in developments with an integrated lender that were built on expansive soil. In column (3) I only include houses whose owner applied for a mortgage from the integrated lender, and in column (4) I only include houses whose owner applied for a mortgage from a non-integrated lender. Standard errors are clustered at the developer level. Significance levels: * ($p<0.10$), ** ($p<0.05$), *** ($p<0.01$).

	P(Foreclosure)	Return - Period A		
	(1)	(2)	(3)	(4)
Integrated Lender	-0.053**	0.271		
	(0.021)	(0.387)		
Non-Integrated Lender + Rejection by IL	0.119**	-0.780		
	(0.056)	(1.413)		
Non-Integrated Lender + No Application to IL	-0.033	0.106		
	(0.023)	(0.534)		
Rejected by Integrated Lender			-2.231***	
			(0.745)	
Rejected by Non-Integrated Lender				0.022
				(0.867)
Controls	H,B,F, T,D1	H,B,F, T,D1	H,B,F, T,D1	H,B,F, T,D1
Sample - Applied to:			Integrated Lender	Non-Integrated Lender
R^2		0.888	0.891	0.883
$\bar{y}$	0.045	8.010	8.402	7.596
N	1,342	2,655	1,764	1,019

Notes

[1]This model abstracts from a number of important housing and mortgage market frictions, such as search frictions (Wheaton (1990), Piazzesi, Schneider and Stroebel (2013)), and frictions related to the tax treatment of residential real estate (e.g., Floetotto et al. (2014)).

[2]A standard feature of these models is that the equilibrium bidding strategies of individual non-integrated lenders are indeterminate. What is determinate is the minimum of all non-integrated lenders' bids. Hence solving an equilibrium with many uninformed lenders is equivalent to solving the equilibrium of competition between the integrated lender and one representative non-integrated lender (Engelbrecht-Wiggans et al. (1983)).

[3]The setup analyzed here is similar to the first-price sealed-bid common value auction analyzed by Milgrom and Weber (1982), who show that when the information set of the less-informed competitor is less finely partitioned, the less-informed lenders will make zero profit in equilibrium.

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

[5]A formal definition of the different types of soil is provided in the USDA's National Engineering Handbook, Part 630 (Hydrology): `http://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=17757.wba`.

[6]The Federal Reserve's Regulation C, which governs the HMDA, applies to most depository institutions with a branch office in a metropolitan area. Banks below \$39 million in assets are exempt from reporting requirements, as are nondepository institutions with assets below \$10 million.

[7]I exclude repeat sales pairs for which the time difference between the two sales is less

than 270 days. Such sales often precede or follow the redevelopment of a property. For similar reasons, the Case-Shiller house price index excludes transaction pairs with less than six months' time difference.

[8]This results in primarily dropping observations from Pima county (city of Tuscon), which does not usually provide lot size and building size in the assessment records. See the discussion in Section II.E of the Internet Appendix.

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

[10]A robustness check shows that the results are unaffected when including observations from Pima county and dropping the control variables with incomplete field population from the empirical model.

[11]Kurlat and Stroebel (2015) construct a different measure of structure share in total value for Los Angeles. They exploit that the tax assessor provides a separate valuation for the land and the structure. Unfortunately, the Arizona assessor data do not allow me to construct a similar measure.

References

Bayer, Patrick, Robert McMillan, Alvin Murphy, and Christopher Timmins, 2011, A dynamic model of demand for houses and neighborhoods, NBER Working Paper 17250.

Elul, Ronel, Nicholas Souleles, Souphala Chomsisengphet, Dennis Glennon, and Robert Hunt, 2010, What "triggers" mortgage default?, *American Economic Review* 100, 490-494.

Engelbrecht-Wiggans, Richard, Paul R. Milgrom, and Robert J. Weber, 1983, Competitive bidding and proprietary information, *Journal of Mathematical Economics* 11, 161-169.

Floetotto, Max, Michael Kirker, and Johannes Stroebel, 2014, Government intervention in the housing market: Who wins, who loses?, Working Paper, New York University.

Hauswald, Robert, and Robert Marquez, 2006, Competition and strategic information acquisition in credit markets, *Review of Financial Studies* 19, 967-1000.

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

Milgrom, Paul R., and Robert J. Weber, 1982, The value of information in a sealed-bid auction, *Journal of Mathematical Economics* 10, 105-114.

Piazzesi, Monika, Martin Schneider, and Johannes Stroebel, 2013, Segmented housing search, Working Paper, Stanford University.

The Arizona Republic, 2009, Appeals of property valuations soar in Maricopa County, Published on November 10, 2009.

von Thadden, Ernst L., 2004, Asymmetric information, bank lending and implicit contracts: The winner's curse, *Finance Research Letters* 1, 11-23.

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