NBER WORKING PAPER SERIES

ARE LEMONS SOLD FIRST? DYNAMIC SIGNALING IN THE MORTGAGE MARKET

Manuel Adelino
Kristopher Gerardi
Barney Hartman-Glaser

Working Paper 24180
http://www.nber.org/papers/w24180

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2018

We thank Darren Aiello, Brendan Daley, Stuart Gabriel, Brett Green, Joseph Mason, Christopher Palmer, Anthony Pennington-Cross, Tim Riddiough, Hongfei Tang, Nancy Wallace, Paul Willen, Basil Williams, James Vickery, conference participants at the AFA, AREUEA National Conference, Annual CEPR Symposium, CHUM, FIRS, and NBER Corporate Finance, and seminar participants at Columbia, Duke, MIT (Sloan), Minnesota (Carlson), Northwestern (Kellogg), Pittsburgh (Katz), Virginia (McIntire), University of Colorado at Boulder, UNC Charlotte, and Wharton for their helpful comments and discussions. We thank Valeria Vargas-Sejas for her outstanding research assistance. This paper was previously circulated under the title "A Test of Dynamic Signaling Models: Evidence from Mortgage Securitization." The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Are Lemons Sold First? Dynamic Signaling in the Mortgage Market
Manuel Adelino, Kristopher Gerardi, and Barney Hartman-Glaser
NBER Working Paper No. 24180
January 2018
JEL No. D0,G0

ABSTRACT

A central result in the theory of adverse selection in asset markets is that informed sellers can signal quality and obtain higher prices by delaying trade. This paper provides some of the first evidence of a signaling mechanism through trade delays using the residential mortgage market as a laboratory. We find a strong relationship between mortgage performance and time to sale for privately securitized mortgages. Additionally, deals made up of more seasoned mortgages are sold at lower yields. These effects are strongest in the "Alt-A" segment of the market, where mortgages are often sold with incomplete hard information, and in cases where the originator and the issuer of mortgage-backed securities are not affiliated.

Manuel Adelino
Fuqua School of Business
Duke University
100 Fuqua Drive
Durham, NC 27708
and NBER
manuel.adelino@duke.edu

Kristopher Gerardi
Federal Reserve Bank of Atlanta
1000 Peachtree St. NE
Atlanta, GA 30309
kristopher.gerardi@atl.frb.org

Barney Hartman-Glaser
University of California at Los Angeles
110 Westwood Plaza
Suite C421
Los Angeles, CA 90095
barney.hartman-glaser@anderson.ucla.edu

One of the most widely studied market settings in economics is that of a seller with private information about the quality of an asset facing less-informed buyers. In this kind of setting with adverse selection, sellers can take actions to reveal their private information, as in the classic signaling model of Spence (1973). This notion of signaling has been successfully applied in theoretical models of financial markets to explain a variety of phenomena, from the optimality of debt (DeMarzo and Duffie (1999)) to the temporary freezing of asset markets (Daley and Green (2012)). While many commonly observed behaviors are consistent with signaling, such as the attainment of education or the propensity of underwriters to retain equity in an initial public offering, there is little empirical evidence that agents actually engage in these activities to signal rather than for other reasons. The fundamental challenge for a test of signaling theory is that it requires the econometrician to observe agents' private information or hidden "types." We address this challenge by using unique features of the U.S. mortgage market.

We first present a simple model of mortgage sales to motivate our empirical work. In the model, sellers privately observe mortgage quality, and sellers of high-quality mortgages have a lower cost of waiting because they face lower probabilities of default. We assume that default is publicly observable and eliminates the possibility of a sale. A separating equilibrium emerges in which the time to sale of a mortgage increases in quality, a relationship often referred to as the skimming property. Many recent studies (e.g., Fuchs and Skrzypacz (2013) and Fuchs et al. (2015)) find that the skimming property can emerge in dynamic adverse selection models of financial markets, and a number of others (e.g., Daley and Green (2012) and Daley and Green (2016)) find that the timing of sales in asset markets can serve as a signal of quality. More broadly, the idea that the timing of actions can reveal private information is a central prediction of many adverse selection models.[1]

The mortgage market is well-suited for testing the skimming property and, more generally, trade delay as a signal of quality. Mortgages are durable assets characterized by an objective measure of quality based on the probability of default. Detailed micro data are available to investors, originators, and the econometrician on the characteristics of borrowers and mortgage contracts, which together serve as a good proxy for observable mortgage quality at the time of sale. Crucially, while future default is not known at the time of sale, it is known to the econometrician *ex-post*. These *ex-post* outcomes are (i) unknown to the buyer (ii) known to the econometrician and (iii) correlated with unobserved heterogeneity in asset quality known privately by the seller, as previous studies of the mortgage market show (Demiroglu and James (2012a), Jiang et al. (2014b), Griffin and Maturana (2016), and Piskorski et al. (2015)). The most relevant type of private information that originators collect is knowledge about borrower ability to repay, including future income prospects or stability of employment, or even measures of liquid wealth, none of which are included in data sources available to investors. The distinction between observable and unobservable characteristics is central to our tests and is one of the main reasons that adverse selection models are particularly difficult to test empirically. In fact, most models predict that assets that are observably better

[1] See also Noldeke and Van Damme (1990), Swinkels (1999), Janssen and Roy (2002), Grenadier and Wang (2005), Kremer and Skrzypacz (2007), Guerrieri et al. (2010), Grenadier and Malenko (2011), Chang (2014), and Williams (2016).

should trade faster, not slower.

This paper uses data on mortgages securitized in the non-agency, private-label securitization (PLS) market, which was very active in the middle of the last decade. In this market, investors in mortgage-backed securities (the buyers) purchased claims on large portfolios of mortgages from originators (the sellers). We measure delay of trade from the creation of the asset (the date of origination of each mortgage) up to the issuance of the securities that ultimately receive cash flows on those mortgages. The fact that we have a natural starting point for measuring time to sale is another advantage of using mortgages as a laboratory. While there is a chain of intermediaries between the originators of mortgages and the ultimate buyers of the securities (as Stanton et al. (2014) and Stanton et al. (2015) show), this in general would bias our tests against effectively capturing the role of signaling in transmitting information. We are also able to (imperfectly) observe the prices at which mortgage-backed securities were sold. While most of our analysis focuses on trade delays and mortgage quality, the combination of the availability of observed and unobserved quality measures and prices is rarely available in other contexts.

We find a negative relation between time to sale and the component of mortgage performance that is not explained by observable mortgage characteristics. After conditioning on underwriting characteristics, PLS loans sold five months or more after origination are approximately 5 percentage points less likely to default relative to loans sold immediately after origination. This is an economically meaningful difference, as it is approximately 30% of the average default rate in our sample (16%).

Interpreting these magnitudes through the lens of our model indicates that adverse selection is severe in this market. The difference between the best and worst possible realizations of the originator's private information is almost one third of the average outcome. We provide a quantitative interpretation of our reduced form results by using a simple calibrated version of our model. The cost of signaling is approximately \$540 for a mortgage of \$300,000. This corresponds to a spread of 18 basis points that originators would charge borrowers to compensate for expected signaling costs, a substantial magnitude when compared to the other costs that borrowers pay when taking out a new mortgage.

The results on *ex-post* default are in contrast to those using *ex-ante* measures of credit risk. We construct predicted probabilities of default using only information available to mortgage investors at the time that mortgages are sold into PLS deals. We then explore whether *ex-ante* observable credit risk is related to time to sale. We find a positive relation between *ex-ante* observable risk and time to sale for loans sold in the first 6 months after origination, the opposite of what we find for *ex-post* default, despite the fact that the predicted default measure is highly correlated with observed performance. The relation becomes insignificant for loans sold more than 6 months after origination. Put differently, while unobserved quality is positively related to trade delays, observable risk measures are (weakly) negatively related to time to sale. One interpretation for this finding is that there may be more investor demand for observably safe mortgage-backed securities.

A potential alternative explanation for our findings is that mortgages that do not default in the first months after origination are simply of better quality regardless of the originator's private information. If there is random delay in time to sale, and if delinquent mortgages are less likely to be sold, a longer time to sale may mechanically reflect better quality rather than an intention to signal on the part of the originator. We address this concern by restricting our analysis to mortgages that do not default in the first nine months following origination independently of when they were securitized, so that all mortgages in the sample are current by the time the last mortgage is securitized. In this sample, observing time to sale does not contain additional public information about default history. Our core result is unchanged in this subsample: it is still the case that mortgages with a longer time to sale have lower default rates *ex-post*.

In contrast to the findings in the PLS segment of the market, we find no evidence of a negative relation between time to sale and mortgage default in a large sample of loans sold to the government-sponsored enterprises (GSEs), Fannie Mae and Freddie Mac. This is consistent with the institutional features of the GSE market, in which automated underwriting and the credit guarantees provided by the agencies essentially eliminate asymmetric information on mortgage credit quality (although not necessarily about prepayment risk) between investors in GSE securities and originators.

Using a second loan-level data set (from CoreLogic) that contains information on the identities of originators and security issuers, we separately estimate the correlations between time to sale and default for issuers and originators that are affiliated entities, i.e. they share the same parent company (as in Demiroglu and James (2012a) and Furfine (2014)). This helps distinguish signaling behavior from "unilateral" concerns about warehousing loans on the part of the seller. If our results simply reflected originator reluctance to hold on to bad loans without an intention to signal unobserved quality to buyers, we would expect no differences across affiliated and unaffiliated entities. Instead, we find a significantly weaker negative correlation between time to sale and default risk for the sample of mortgages in which the issuer and originator are affiliated with each other. These results indicate that a key component of information asymmetry leading to a delay in trade is between the originator of the mortgage and the issuer of the security.

The relation between time to sale and default is strongest in the "Alternative-A" (or "Alt-A") segment of the market, which is mostly comprised of low-documentation loans or loans with risk characteristics that prevent them from being securitized in the conforming market.[2] While the subprime segment of the market is riskier than the Alt-A segment, subprime mortgages are more homogeneous in their (potentially unobserved) risk characteristics, lending further credence to an adverse selection, signaling interpretation. While signaling of default risk did not play an important role in the context of subprime loans, there is a significant relation between time to sale and *ex-post* prepayment behavior (consistent with Agarwal et al. (2014), who find that prepayment risk was an important concern in this particular segment of the market).

The Corelogic data also allow us to include originator and issuer fixed effects in the regressions.

[2] Jiang et al. (2014a), Jiang et al. (2014b), Begley and Purnanandam (2017), and Saengchote (2013) discuss the role of private information in low documentation loans.

This helps us account for differences in funding sources across originators, such as the use of very short-term warehouse loans and repurchase agreements that might prevent a signaling mechanism from operating. Stanton et al. (2014) and Ganduri (2016) show that some originators, particularly independent mortgage companies, rely almost exclusively on these types of funding sources. As such, originator fixed effects allow us to investigate within-originator variation in time to sale that is not driven by variation in funding sources. The results are similar to the baseline specifications that do not control for originator and issuer identities.

We then turn to the pricing dimension to determine whether prices rise with time to sale, as predicted by the signaling model. Data on prices paid for individual mortgages are not available (to our knowledge), so we conduct an analysis of mortgage-backed security (MBS) prices. If signaling plays an important role in the market, we should see a positive relation between average time to sale of mortgages included in a deal and security prices. We do not observe prices at origination, so we use spreads of floating rate securities as our measure of pricing (consistent with, among others, Ashcraft et al. (2011), He et al. (2012), and Begley and Purnanandam (2017)). We find that securities backed by loans that take longer to sell (more seasoned loans) are sold at lower yields. One additional month of average seasoning is associated with a 1.5–2.0 basis-point reduction in the yield of triple-A securities (the average spread is 28 basis points). Consistent with the evidence on mortgage performance, the pricing results are non-linear in seasoning and are strongest in the Alt-A segment of the market.

This paper relates to the extant literature on adverse selection and signaling. The seminal work of Akerlof (1970) is the first to show that markets can break down when some participants have valuable private information. In related work, Spence (1973) shows that informed agents can take actions to credibly reveal their private information. These actions then lead to a separating equilibrium where the agent's private information is revealed. Leland and Pyle (1977) are the first to apply this insight to financial markets and show that the issuers of IPOs can signal information by retaining an equity stake in the IPO. DeMarzo and Duffie (1999) use the equilibrium relation between retention and asset quality to show that debt minimizes the costs associated with the separating equilibrium and is hence an optimal security design. DeMarzo (2005) builds on this idea to show that it is optimal to first pool assets to minimize adverse selection and then to create tranches to minimize signaling costs.

This paper also contributes to the empirical literature on the effects of asymmetric information. The seminal work of Genesove (1993) finds weak evidence of adverse selection in the used car market. Garmaise and Moskowitz (2004) use commercial real estate transactions to test a number of theories of asymmetric information, including the prediction that securities issuers retain a stake to signal their information. In contrast to our paper, they find no evidence that informed sellers of commercial real estate signal their information through retention. Downing et al. (2009) also consider retention and find that mortgages sold to special purpose vehicles (SPVs) tend to be of lower quality than mortgages not sold to SPVs. Agarwal et al. (2012) find no systematic differences between subprime mortgages sold in the secondary market and those retained on banks' balance sheets. Closest to our setting, Begley and Purnanandam (2017) find that higher levels of equity tranches in PLS deals (a measure of retention) are associated with lower delinquency rates and higher prices. Aiello (2016) finds evidence that borrower payment behavior during the warehouse period can be a source of private information for originators. An et al. (2011) find that information asymmetries in the secondary commercial mortgage market can lead to market break down. They argue that conduit lenders exist to mitigate asymmetric information. Fuchs et al. (2015) find evidence consistent with the skimming property in the IPO market.

Two studies document misrepresentation in the private mortgage market. Piskorski et al. (2015) finds that lenders often misrepresent loan-to-value ratios when selling mortgages and Garmaise (2015) finds that borrowers often misreport the value of their personal assets on mortgage applications. These studies suggest there is significant scope for private information in the mortgage market.

1 A Model of Signaling through Delayed Trade

To motivate our empirical tests, we present a simple model of adverse selection and delayed trade in the secondary market for mortgages. Time is infinite, continuous, and indexed by t. The model consists of a mortgage originator and a competitive market of mortgage investors. All agents are risk neutral. At time $t = 0$, the seller originates a mortgage for potential sale to the market. This mortgage produces a cash flow of c dollars per unit of time until it defaults at some random time τ. The default time τ is an exponential random variable with parameter λ distributed on the compact interval $[\lambda_\ell, \lambda_h]$ according to the continuous density $f(\lambda)$. The parameter λ is the annualized expected default rate for the mortgage. the While $f(\lambda)$ is common knowledge, the seller privately observes λ at the origination of the mortgage. As is common in such settings, we refer to λ as the seller's type.

While both the seller and potential investors are risk-neutral, gains from trade are generated by a difference in discount rates used by the two classes of agents. Specifically, the seller discounts cash flows at rate γ and the investors discount cash flows at rate $r < \gamma$. This difference in discount rates proxies for the difference in the investment opportunity sets of the seller and the investors. The seller has the technology to originate mortgages. In contrast, investors can only purchase mortgages in a competitive market once they have already been originated. Modeling these gains from trade as a difference in discount rates is convenient for the analysis that follows, but not necessary. Provided the gains from trade between the seller and investors are weakly positive for all seller types, the predictions of the model remain qualitatively unchanged.

We assume that mortgage default is publicly observable such that if the mortgage defaults before the seller has sold it to the investors, no sale occurs. In choosing when to sell the mortgage, the seller takes the market price function $P(t)$ as given. Note that the lowest possible value of a mortgage to investors is

$$p_h = E\left[\int_t^\infty e^{-r(s-t)}\mathbb{1}(s \le \tau)cds|\lambda_h\right] = \frac{c}{r+\lambda_h},$$

while the highest possible value is

$$p_\ell = E\left[\int_t^\infty e^{-r(s-t)}\mathbb{1}(s \le \tau)cds|\lambda_\ell\right] = \frac{c}{r+\lambda_\ell};$$

thus, $P(t) \in [p_h, p_\ell]$.

An outcome of this game is a triple $(\lambda, t, p) \in [\lambda_\ell, \lambda_h] \times [0, \infty) \times [p_h, p_\ell]$, where λ is a realization of the seller's type and t and p respectively correspond to the time and price at which trade takes place if the mortgage has not defaulted by time t. The value for the seller of an outcome of the game is then given by

$$U(\lambda, t, p) = E\left[\int_0^t e^{-\gamma s}\mathbb{1}(s \le \tau)cds + e^{-\gamma t}\mathbb{1}(t \le \tau)p|\lambda\right] = \frac{c}{\gamma+\lambda}\left(1 - e^{-(\gamma+\lambda)t}\right) + e^{-(\gamma+\lambda)t}p.$$

An important feature of the seller's payoff function is the so-called single-crossing property: fixing a price p, it is less costly for better (lower default risk) sellers to delay trade. Intuitively, the lower the default risk, the greater the private value of the cash flows that accrue to the seller from the mortgage before the sale, and the greater the probability that the mortgage will remain current so that it can be sold in the future. This feature of the model gives rise to the common *skimming property*, which is present in much of the literature on dynamic trading and asymmetric information (for example, the literature on sequential bargaining with one-sided incomplete information following Fudenberg and Tirole (1983)) and which is more broadly related to the literature on costly signaling with adverse selection (for example, Spence (1973) and Leland and Pyle (1977)). In our model, the skimming property can be expressed as follows: For a given price function $P(t)$, better sellers wait (weakly) longer to trade, and thus, a trade delay can act as a signal of quality.

An equilibrium of the game is a pair of functions (T, P), where $T(\lambda)$ is the time at which a seller of type λ trades and $P(t)$ is the price of a mortgage sold at time t such that the following conditions hold:

1. Seller optimality: $T(\lambda) \in \arg\max_t U(\lambda, t, P(t),)$.

2. Zero profit for the investors: $P(T(\lambda)) = E\left[\frac{c}{r+\tilde{\lambda}}|T(\lambda)\right]$.

An equilibrium is separating if $P(T(\lambda)) = \frac{c}{r+\lambda}$.

We focus on characterizing a separating equilibrium. Other equilibria, such as pooling equilibria, are eliminated by standard refinement criteria, such as the D1 refinement of Cho and Kreps (1987). The following proposition characterizes the unique separating equilibrium of the game.

Proposition 1. *The unique separating equilibrium of the game is given by*

$$T^*(\lambda) = \frac{\log(r+\lambda_h) - \log(r+\lambda)}{\gamma - r}, \qquad P^*(t) = p_h e^{(\gamma - r)t}. \tag{1}$$

The method for deriving the equilibrium of Proposition 1 is as follows: First, fix some candidate price function $P(t)$ and take a first-order condition for the seller's problem, which yields,

$$c - (\gamma + \lambda)P^*(t) + \frac{d}{dt}P^*(t) = 0. \tag{2}$$

Next, use the fact that for any separating equilibrium,

$$P^*(T(\lambda)) = \frac{c}{r+\lambda},$$

and substitute into equation (2) to obtain the following ordinary differential equation for $P^*(t)$:

$$\frac{d}{dt}P^*(t) = (\gamma - r)P^*(t). \tag{3}$$

Finally, because the highest expected default rate type does not benefit from delaying trades in a separating equilibrium, we must have $T^*(\lambda_h) = 0$ and, hence, $P^*(0) = p_h$. The functions given in Proposition 1 solve Equations (2) and (3) with this initial condition.

To connect the equilibrium given in Proposition 1 to our empirical analysis, it is useful to consider how the type of seller changes with time to sale. We let $\lambda^*(t)$ denote the type of seller that chooses to sell at time t. Applying Proposition 1, we have

$$\lambda^*(t) = (r+\lambda_h)e^{-(\gamma - r)t} - r. \tag{4}$$

Our empirical results relate to the following key properties of the equilibrium given in Proposition 1.

1. The expected default rate of the mortgage decreases with time to sale, $\frac{d}{dt}\lambda^*(t) < 0$. This means that asymmetric information creates a negative relation between time to sale and expected default rate.

2. The price of the mortgage increases with time to sale, $\frac{d}{dt}P^*(t) > 0$. This means that asymmetric information creates a positive relation between price and time to sale.

3. The maximum time to sale for a mortgage is increasing in the difference in default risk between the safest and riskiest mortgage, $\frac{d}{d(\lambda_h - \lambda_\ell)}T^*(\lambda_\ell) > 0$. This means that a more severe adverse selection problem, such as when the uncertainty about mortgage default risk is greater, leads to longer trade delays.

Although the separating equilibrium we detail above is the unique equilibrium selected by D1, a discussion of other possible equilibria is in order. In particular, many pooling equilibria can exist in which all seller types sell at the same time. For example, if investors believe that any mortgage sold after time $t = 0$ is the riskiest type, then all seller types find it optimal to sell their mortgages at $t = 0$ because delaying the sale only leads to forgoing gains from trade and does not increase the sale price. However, imposing D1 refinement eliminates this equilibrium. If investors observe an off-equilibrium-path action, such as a seller delaying a trade when investors expect an immediate sale, then D1 requires that they only place positive weight on those seller types who would gain from deviating given the largest set of prices. This set is always largest for sellers of the least risky mortgages because it is less costly for them to delay trades than for any other seller type. As such, D1 requires that investors must believe that the seller is the least risky type if she even slightly delays a trade. These beliefs thus imply that sellers of the least risky type have a profitable deviation, eliminating the simple pooling equilibrium. Thus, we focus our empirical analysis on the separating equilibrium detailed above.

Before proceeding further, we situate our model in the extant literature. For the sake of simplicity, we have assumed that the seller can commit to a time to sale, and in that sense our game is essentially static as in the model of Spence (1973) in which students commit to a particular period of education. Swinkels (1999) shows that without commitment, Spence's signaling equilibrium might not exist. However, a number of authors, for example Daley and Green (2012), argue that dynamic concerns can restore delay in trade as a signal of quality. In a dynamic version of our model in the spirit of Fuchs and Skrzypacz (2013) and Fuchs et al. (2015), the qualitative results are unchanged.

1.1 Random Delay, Default, and Prices

To impose further discipline on our empirical analysis, we consider a plausible variation to our model in which a correlation between delayed trades and *ex-post* performance need not be indicative of signaling or adverse selection. Intuitively, if a trade is randomly delayed, then some higher-risk mortgages may default before they can be sold. As a result, mortgages that take longer to sell are positively selected (i.e., they are of higher quality than those that could not be sold). This selection mechanism would then lead to a positive correlation between time to sale and *ex-post* performance (a negative correlation between time to sale and realized default rates). Investors who understand this selection issue believe that mortgages that sell after a longer period of seasoning are of higher quality and so prices increase with seasoning. Importantly, this effect does not arise from signaling, as mortgages are sold randomly into pools by assumption, but rather through a learning process. As such, a simple model with random delay of trade and the associated selection mechanism may appear observationally equivalent to our signaling model. This is a key difficulty in operationalizing models of asymmetric information: they often make similar predictions to models with symmetric information. We can overcome this difficulty in our setting by observing that the selection mechanism can be undone by conditioning the analysis on mortgages that do not subsequently (after origination) default up to a pre-specified period.

To bring some precision to this intuition, suppose that the mortgage seller detailed above has the same information as potential investors. Specifically, she knows that the mortgage she wants to sell has an exponential default time with an intensity $\tilde{\lambda}$ uniformly distributed on $[\lambda_l, \lambda_h]$. When she chooses to sell the mortgage, there is a delay between the point at which she lists the mortgage for sale and the moment at which the transaction is recorded, which is exponentially distributed with parameter μ. If the mortgage defaults before the transaction can be recorded, then no sale takes place. As a result, observing that the mortgage transacts at time t reveals that the mortgage did not default prior to t. Thus, the expected quality of a mortgage that transacts at time t is given by the following expression:

$$E\left[\tilde{\lambda}|\text{sold at time } t\right] = E\left[\tilde{\lambda}|\tau > t\right] = \lambda_h + \frac{1}{t} - \frac{\lambda_h - \lambda_l}{1 - e^{-t(\lambda_h - \lambda_l)}},$$

which is increasing in sale time t. Hence, a model with random delay of trade and symmetric information is consistent with negative correlations between both time to sale and *ex-post* default outcomes and time to sale and *ex-ante* prices. These predictions are essentially the same as properties 1 and 2 of the signaling model described above. In order to empirically test the predictions of the signaling model we need to overcome this selection effect.

A simple way to account for this selection effect is to condition the analysis on loans that do not default until some exogenously specified time s, which needs to be after the period of sale, t. For loans that do not default before s, the fact that the mortgage was sold at time $t < s$ does not contain any additional information about the default risk of the mortgage. The expected quality of a mortgage that has not defaulted by time s and is sold at time $t < s$ is given by the following expression:

$$E\left[\tilde{\lambda}|\text{sold at time } t < s \text{ and } \tau > s\right] = E\left[\tilde{\lambda}|\tau > s\right] = \lambda_h + \frac{1}{s} - \frac{\lambda_h - \lambda_l}{1 - e^{-s(\lambda_h - \lambda_l)}},$$

which is independent of the time of sale t. Thus, in a model with random delay and no signaling mechanism, there is no correlation between time to sale and *ex-post* default outcomes if we condition on a sample of mortgages that do not default before s, where $s > t$. This is in stark contrast to our model of signaling, in which time to sale always reveals information about *ex-post* default risk. We explore whether such a model can explain our results in the empirical tests below.

Finally, it is likely that even if there is delay due to the signaling of private information, there is also some delay that is random and uncorrelated with private information. We do not incorporate this possibility in the model at this stage, as doing so complicates the intuition without leading to new insights that we can benefit from empirically. In Section 5, we use our model to assess the quantitative impact of signaling in the data. To ensure realistic estimates of the magnitudes of the cost of signaling given our empirical findings, we use a version of the model that incorporates random delay.

2 Background on the U.S. Mortgage Market

Our primary focus herein is loans that were sold and then securitized by private financial institutions (or issuers). This segment of the market, often referred to as the PLS (private-label securitization) market, was the initial source of the mortgage foreclosure crisis in 2007, which led to the broader financial crisis and the Great Recession. The PLS market grew rapidly during the housing boom of the mid-2000s, reaching a peak share of approximately 56% of the securitization market in 2006, before completely shutting down in the summer of 2007 when subprime mortgage defaults dramatically increased.

The PLS market is split into segments according to the degree of credit risk. The "Alt-A" segment, also commonly referred to as "near prime," is typically characterized by loans to borrowers with credit scores that are comparable to average credit scores in agency pools, but where borrower income and/or assets are less than fully documented (i.e., low-documentation mortgages). These loans are also more likely to finance investor or vacation properties. The collateral underlying subprime private-label securities is made up of loans given to borrowers with low credit scores and includes a large fraction of cash-out refinance mortgages. The majority of subprime PLS loans did not meet underwriting standards in the agency market and were broadly viewed as low-quality mortgages by market participants. Our primary data set (from McDash Analytics, described in more detail below) includes loans from both segments of the PLS market, while our secondary source of data (CoreLogic's LoanPerformance database, also described below) includes loans from the subprime and Alt-A segments of the market.

There is significant variation in the funding and operational models of mortgage originators in the PLS space, including independent mortgage companies, affiliated mortgage companies, and others. We refer the reader to Stanton et al. (2014) and Ganduri (2016) for detailed descriptions of the structure of this market. Stanton et al. (2014) show that repurchase agreements and warehouse lines of credit with very short maturities are a large funding source in the PLS market. This limits the originators' ability to delay mortgage sales. Our tests require that either originators of mortgages or issuers of PLS (or both) have the ability to hold on to mortgages and delay trades, even if some are limited by contractual features because of their funding sources. Even though we find that the majority of loans in the PLS market were securitized within the first two months after origination, consistent with the evidence provided in Stanton et al. (2014) that warehouse loans and repurchase agreements had 30 to 45 days' maturity, the variation that is most relevant for our tests comes from sales past this time period (up to 9 months after origination).

3 Testing for Dynamic Signaling Using Mortgage Data

We implement empirical tests of Predictions 1 and 2 of the signaling model developed in Section 1. Prediction 1 is that time to sale and default propensities should be negatively related. Although we do not observe the expected default rate of any one particular mortgage, i.e., the λ of the mortgage we describe in the model, we do observe the *ex-post* conditional performance of mortgages. As we argue in the introduction, *ex-post* realized default rates conditional on the observable characteristics that we have available are correlated with the unobserved component of *ex-ante* expected default rates (observed, at least partially, by the seller). Thus, a testable version of Prediction 1 is that conditional realized default rates are negatively associated with time to sale. Prediction 2 is that there should be a positive relationship between time to sale and mortgage prices. Given superior data and the ability to perform much richer cross-sectional tests, we focus primarily on an analysis of default rates. However, we do briefly discuss the setup of the pricing tests in this section and show results in Section 4.6.

3.1 Time to Sale and Mortgage Default

A key issue in implementing an empirical test of the skimming property is distinguishing between observable and unobservable asset quality. Signaling in general and the skimming property in particular refer specifically to quality that the seller is informed about but is unobservable to the buyer.

We implement a strategy similar to Adelino et al. (2017) that uses conditional measures of loan performance to isolate aspects of loan quality that are unobservable to investors at the time of purchase but are correlated with the originators' information set (and which, by virtue of the passage of time, become observable to the econometrician). Specifically, we condition performance on a large set of loan and borrower characteristics used in mortgage underwriting models that are readily available to issuers and institutional investors in the MBS market. Our empirical specifications take the following general form:

$$Default_{ijt} = \alpha + \beta_1 * \text{Months to Sale}_{ij} + \beta_2 * X_{ijt} + \epsilon_{ijt} \tag{5}$$

where i indexes the individual mortgage, j indexes the geographic area in which each mortgage is originated, and t indexes the horizon over which we calculate the realized default rates. X_{ijt} is a vector of control variables. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination).

Months to Sale_{ij} measures the time between mortgage origination and sale into the secondary market (and securitization). The availability of a natural starting point (the date of origination) for measurement of delay of trade is an additional benefit of focusing on the secondary market for mortgages. In many other asset markets there is no such "date 0" to start measuring delay. We show in the Online Appendix (Figure A.4) that the typical prospectus of a private-label deal included average seasoning (average time to sale) in the first table showing the mortgage characteristics included in the deal. We restrict the analysis to loans sold up to 9 months after origination, but Section A.3 of the Online Appendix shows robustness tests for longer sale horizons.

To relate the regression in Equation (5) to the model we present in Section 1, note that the left-hand-side variable, $Default_{ijt}$, is an estimate of the probability (after accounting for observables) that a loan defaults within the first t months. For a mortgage with expected default rate λ, this probability is given by $1 - e^{-\lambda t}$. Prediction 1 of our model is that λ^* is decreasing in time to sale, and hence so is $1 - e^{-\lambda t}$. Thus, in the context of the regression in Equation (5), Prediction 1 is $\beta_1 < 0$. This is a joint test of two hypotheses: that (i) the seller's private information, I_{seller}, is correlated with loan quality, i.e. the expected default rate, after accounting for underwriting characteristics,

$$Corr[(E(Default_i|X_i, I_{seller}) - E(Default_i|X_i)), Default_i)] \neq 0 \tag{6}$$

and that (ii) sellers signal asset quality by delaying trades.

3.1.1 Observable Characteristics and Default Measurement

X_{ijt} in equation 5 is a vector of mortgage-level control variables that includes the relevant observable borrower, loan, and geographic characteristics, including detailed fixed effects. According to Stearns (2006), all issuers and most PLS investors have access to detailed information at the loan level, including original loan balance, FICO score, combined loan-to-value ratio, documentation type, occupancy type, loan purpose (refinance or purchase), property type, loan size, amortization schedule, interest rate, loan type (ARM vs. FRM), and information on the geographic location of the property. This contrasts with the agency market, as the GSEs, partly because they absorb all credit risk, do not disclose as much detailed information about the mortgages that back their securities.

We do not include debt-to-income (DTI) ratios in the regressions because this variable is missing for about 50% of loans in McDash and over one third of loans in Corelogic. In addition, this variable is reported inconsistently across servicers, as it sometimes measures "front-end" DTI (which includes only mortgage expenses), and other times the "back-end" ratio (which includes all debt-related expenses). We show regressions including DTI ratios, as well as a longer discussion of the reasons for exclusion from the analysis, in Section A.4 of the Online Appendix.

Given the poor quality of the DTI variable, and the different definitions used in the industry, it is likely that originators have an information advantage over investors with regard to borrower ability to repay. This can include not just current front- and back-end DTIs, but also future income prospects, occupation, self-employed status, stability of employment, measures of liquid wealth, among others. All of these dimensions are plausibly related to credit quality and are not included in any standard data source that is available to buyers of the mortgages.

Our vector of controls also includes variables that measure conditions in the local housing market, including the county-level unemployment rate and the level and changes of the county-level house price index (normalized by setting the index value for January 2000 to 100 for each county). We include a full set of state-level fixed effects and fixed effects corresponding to the year-quarter of origination and the year-quarter of loan sale. We have also experimented with a specification that includes zip code-level fixed effects and found that the results were largely unaffected. Because including such a large number of fixed effects is parametrically expensive and computationally demanding, we use state fixed effects in all of the tests in the paper. Section A.1 and Table A.9 of the Online Appendix contain a list of all the variables included in our covariate set.

We consider two default horizons, 36 and 60 months, in our primary specifications; these are measured relative to the month of loan origination. We also consider a mortgage to be in default if the borrower is either two payments behind (60+ days delinquent) or three payments behind (90+ days delinquent) at any point between origination and each default horizon.

We use 60-day and 90-day delinquency cutoffs rather than the initiation of foreclosure proceedings so that our default definition reflects borrower behavior that is not confounded by the decisions of mortgage servicers. Unlike other debt markets where monitoring by lenders is common (e.g., bank loans), mortgage servicers do not monitor borrowers prior to serious delinquency. In fact, servicers do not obtain additional information about borrowers after origination other than their payment history, which is observable to any buyer of the mortgage. If servicers did acquire additional information, the tests below might reflect differences in the observable information across borrowers due to a longer horizon for acquiring that information. This is not a plausible mechanism for the findings in the case of the mortgage market.

3.2 Time to Sale and Mortgage Spreads

We do not have access to data on individual mortgage prices, so we conduct our pricing analysis at the security level. The analysis focuses on the average spread (quoted as a spread over the one-month LIBOR rate) of floating rate securities in the PLS market. We calculate a weighted average spread at the deal-level, where spreads are weighted by the face value of securities. If a security is linked to more than one pool, it contributes to each pool's weighted average yield spread.

The analysis focuses primarily on floating rate securities to minimize the possibility that securities were not issued at par. We do not have information on prices at issuance, but floating rate securities were almost always issued at par, in contrast to fixed rate ones. In addition, private-label floating rate securities have very short durations (typically one month), so we can ignore interest rate risk and the negative convexity problem that arises with fixed-rate mortgage-backed securities. We show results using only triple-A securities as well as using all tranches in a deal. Aggregation of spreads over pools of mortgages becomes noisier when we include tranches below triple-A because these were more likely to have claims on more than one pool in the same deal. In addition, the sample becomes smaller because junior tranches were also more likely to be issued at fixed rates (causing deals to be dropped from the analysis).

Our empirical analysis considers the relation between average yield spreads and mortgage seasoning. The seasoning variable, which is calculated as the average months to sale in the pool, and all controls are constructed from loan-level data and aggregated to the pool level. Our specifications take the following form:

$$Spread_i = \alpha + \beta_1 * \text{Seasoning}_i + \beta_2 * X_i + \epsilon_i \tag{7}$$

where i represents a pool and X_i includes the pool averages of all relevant loan and borrower characteristics used in the loan-level tests and the quarter of issuance fixed effects. Prediction 2 of our model is that average seasoning is positively related to price, and hence negatively related to mortgage spreads, so that $\beta_1 < 0$.

3.3 McDash Data

In this section, we describe the two loan-level data sets used in this paper. While both loan-level data sets are similarly structured monthly mortgage panels, there are important differences in the scope of their coverage and in some of the underlying variables that produce advantages and disadvantages in the context of our analysis.

The pricing data at the individual security level were obtained from Bloomberg and cover over 90% of all subprime PLS issued in the U.S. between 2002 and 2007. We are able to combine the CoreLogic and Bloomberg data sets by merging individual security CUSIPs.

3.3.1 Lender Processing Services Data

Our primary data set is from McDash Analytics. The McDash data set covers between 60% and 80% of the U.S. mortgage market and contains detailed information on the characteristics and performance of both purchase-money mortgages and refinance mortgages. The McDash data set is constructed using information from mortgage servicers, financial institutions that are responsible for collecting payments from borrowers. Each loan is tracked at a monthly frequency from the month of origination until it is paid off voluntarily or involuntarily via the foreclosure process. We focus on loans originated during the housing boom, from January 2002 through December 2007.

Importantly, the data set includes a time-varying variable, "investor type," which identifies whether a mortgage is held in a bank's portfolio, is privately securitized, or is securitized by GSEs. This variable allows us to identify if and when a loan is securitized or sold to a GSE.

The main advantage of using McDash data to test the skimming property is the ability to consider sales to both PLS and the GSEs. The GSE market provides us with an important counterfactual exercise because loans are approved based solely on observable characteristics (typically through automated systems). The biggest drawback, however, is the lack of information on the identity of the financial institution that originates the mortgage. In addition, there is some concern that the McDash data set may under-represent the PLS market during our sample period, and that it overweights the Alt-A segment of the market (we discuss this in more detail in Section 3.4 below). For these reasons, we also use data from CoreLogic's LoanPerformance database.

3.3.2 CoreLogic Data

Our second source of mortgage data is CoreLogic's LoanPerformance (CL) database, which covers virtually the entire subprime and Alt-A segments of the private-label securitization market. Like the McDash data set, CL contains detailed information on underwriting characteristics and monthly loan performance, but unlike McDash, it does not have information on portfolio-held loans or loans securitized by the GSEs.

The CL database includes virtually the same mortgage and borrower characteristics (at the time of loan origination) as the McDash database, but, importantly, about 50% of the CL database includes the identity of the originating institution. This allows us to include originator fixed effects in our regressions, and hence purge any time-invariant, unobserved heterogeneity in originator underwriting practices and funding sources from the analysis. In addition to the identity of the originator, CL provides information on the identity of the mortgage servicer and on security identifiers (CUSIPs) and deal identifiers, which allows us to obtain information on the identity of the securitizer (issuer) for most of the loans in the sample and to merge the loan-level data to yield spread data from Bloomberg. Finally, CoreLogic also allows us to distinguish between the subprime and Alt-A segments of the PLS market.

3.4 Summary Statistics

Table 1 shows the distribution of the number of months between origination and sale for the McDash data. It is clear from the table that the majority of both PLS and GSE securitized mortgages are sold very quickly, either immediately or only one month after origination. Few GSE loans (about 8%) are sold more than two months after origination, but a non-trivial fraction of PLS loans are sold later (about 22% are sold more than two months after origination). We impose a maximum threshold of nine months between the origination and sale of a loan to ensure that we have power to identify non-parametric regression specifications by month and to ensure that the loans in the sample were originated with the intention of being sold. Section A.3 of the Online Appendix includes results using higher thresholds of months to sale. This leaves us with a sample of over 5.7 million loans sold to PLS issuers and over 14 million loans sold to the GSEs.

Table 3 displays summary statistics for many of the control variables in the empirical models. In general, PLS loans are characterized by riskier attributes than are GSE loans. For example, there are more interest-only loans, more adjustable-rate loans, more low-documentation loans, more subprime loans, and more loans that carried prepayment penalties in the PLS sample.

Table 2 displays the distribution of months to sale in the CoreLogic data set, while Table 4 provides some basic summary statistics. There are many more PLS loans in CoreLogic than in McDash, reflecting the differences in coverage across the two data sets. The distribution of months to sale in CL is generally similar to McDash, particularly in the case of Alt-A loans.

Note that the McDash sample size of 5.7 million loans listed in the tables understates the total number of PLS loans, as some seasoned mortgages are eliminated from the sample because we only include loans for which we have a full history of performance. In total, there are approximately 8 million PLS loans originated between 2002 and 2007 (inclusive) in the McDash database.

Table 4 shows that the CL sample is characterized by significantly lower credit scores (FICOs), higher interest rates, and lower loan amounts relative to the McDash data set. However, the Alt-A loan characteristics in CL are generally close to the McDash sample.

Table A.8 in the Online Appendix shows summary statistics for all of the pool-level characteristics used in the pricing analysis. The average spread of triple-A securities in the data is 28 basis points, with a standard deviation of 23 basis points. This spread is computed as the pool-level average of all triple-A securities drawing cash flows from a given pool, and the sample is limited to pools with only floating rate triple-A securities. The average pool-level seasoning in the data is 3.3 months, and it is truncated at 9 months following the approach used for the default analysis. Figure 4 shows a histogram and cumulative distribution of the pool-level seasoning variable. Pools are made up of 2,355 loans on average (the median is 1,911), with an average FICO score of 640, and a combined loan-to-value ratio of 84%.

4 Results

In this section, we present results on the relation between time to sale, default, and prices. We first implement tests using linear specifications so that Months to Sale_{ij} (for the loan-level default analysis) and Average Seasoning_i (the pool-level average used in the pricing regressions) take values from 0 to 9. We then consider specifications that allow for potential non-linearities. For the loan-level default regressions, we estimate a non-parametric specification in which we include separate indicator variables for each value of the months to sale variable. Because we cannot distinguish between loans with values of 0 and 1 months to sale, the omitted baseline category for the regressions estimated on McDash data includes both.

4.1 Default and Time to Sale

Panel A of Table 5 displays results for the linear regression estimated on loans in the McDash data set. We use two different default definitions (60+ DQ and 90+ DQ, 60- and 90-day delinquency, respectively) and two different default horizons (36 months and 60 months relative to the month of origination). The results show a negative, statistically significant relation between default risk and time to sale. The magnitude of the coefficient in the linear specification is approximately -0.01, which implies that a one-month increase in time to sale is associated with a 1-percentage-point decrease in the average realized default rate. The results are consistent over different horizons and default definitions. Table A.3 in the Online Appendix presents results for when the maximum horizon for time to sale is extended from 9 months to 12 and 24 months. The results are consistent with the main analysis (for details see the discussion in Section A.3 of the Online Appendix). Table A.9 in the Online Appendix displays coefficient estimates for all of the variables in our covariate set.

Table 6 explores whether there is a non-linear relationship between time to sale and default. Columns 1–2 ("Full Sample") show that average realized default rates are decreasing in time to sale until the 5th month after origination, at which point these rates begin to rise moderately. Mortgages sold in the 5th month after origination have realized default rates that are approximately 5 percentage points lower than loans sold in either the month of origination or the month after origination, while mortgages sold in the 9th month after origination have realized default rates that are lower by 3–4 percentage points on average. Again, the estimation results are consistent across alternative default definitions and across different horizons (but are omitted for brevity). Panel A of Figure 1 plots the estimated coefficients from Column 1 in Table 6.

Overall, the results in Tables 5 and 6 provide evidence of a negative relation between time to sale and conditional *ex-post* default risk, which supports the presence of signaling in the PLS market.

4.2 Accounting for "Mechanical" Effects of Random Delays

One potential concern in the default analysis above is the role of early payment defaults in generating a mechanical relation between time to sale and *ex-post* default risk due to institutional features of the PLS market. We discuss this possibility in Section 1.1 in the context of our model. In short, delinquent loans may be harder to sell into a securitized pool of loans, which could create a selection effect of loans sold late relative to loans sold immediately. This in turn could create a negative relation between time to sale and default that is independent from private information and signaling. Random delays would mean that loans sold quickly would be representative of the population of eligible loans, whereas loans that take longer to sell would be of higher average quality than the population of eligible loans.

To address this issue, we undo this potential selection effect by including in the analysis only those loans that do not default within the first 9 months of origination. Put differently, we drop all loans that are securitized between months 0 and 9 and become delinquent by month 9. We refer to this sample as the "restricted sample." This restriction forces the sample of sold loans to be homogeneous in terms of early payment defaults irrespective of when they were sold into securitization pools. Thus, a negative relation between time to sale and default risk in this restricted sample cannot be explained by the selection effect described above.

While this correction directly addresses potential selection bias, it is possible that signaling behavior is precisely about the likelihood of early-payment default. That is, if most of the private information on loan quality concerns the likelihood of default within the first few years of origination, this "correction" would effectively eliminate the variation of greatest interest. For this reason, we choose to display the correction as a robustness check rather than to adopt it as our baseline specification.

Panel B of Table 5 and columns 3–4 in Table 6 display the same set of results for our restricted sample, where we exclude all loans that default within 9 months of origination. We find that the effects are virtually unchanged for the linear specification of the Months to Sale_{ij} variable, but there are a few subtle differences for the non-linear specifications. This sample restriction marginally mitigates the negative relation between time to sale and default for loans sold within 4 months, but it has the opposite effect for loans sold later. Overall, the sample correction appears to have a minor effect on the results, which suggests that sample selection bias is not an important issue.

4.3 Default and Time to Sale: Agency Loans

We next turn to the sample of loans sold to government-sponsored enterprises (GSEs). We view this as an important counter-factual exercise. The GSE market is dominated by automated underwriting systems in which agencies pre-commit to funding loans based on hard (observable) characteristics of the loans, so that originators have no need to signal loan quality through sale delays.

Table 7 displays results for loans sold to the GSEs. We find essentially no evidence of a relation between time to sale and *ex-post* default risk in the GSE segment of the market. The estimates are all close to zero and generally statistically insignificant. We plot the coefficients from a non-parametric specification in Panel B of Figure 1 (the same specification used to construct the PLS graph in the top left panel). There is a stark difference in this pattern compared to the PLS graph. While there is a clear downward trend in the PLS estimates that flattens out toward the end of the time to sale distribution, the GSE coefficients are marginally positive until the end of the distribution, when they begin to fall (although the sample size becomes significantly reduced in these later months).

The GSE results are consistent with our hypothesis that private information does not play a substantive role in the agency market compared to the PLS market.

4.4 Predicted Default and Time to Sale

In this section, we estimate the empirical relation between time to sale and *ex-ante* credit risk; that is, credit risk measured at the time of issuance based on observable characteristics only. As with the regressions using GSE loans, we view this as another important counter-factual exercise. If we found the same type of relation for *ex-ante* risk as we do for *ex-post* default (described above), this would call into question our ability to isolate the observable and unobservable components of risk.

We construct *ex-ante* default probabilities for each loan using the data available in McDash and using only performance information available at the time of origination in a manner similar to the method used in Ashcraft et al. (2010). We focus on 36- and 60-month horizons for the default forecasts. For each quarter in our sample, we take all loans that were originated between 48 months and 36 months prior and track those mortgages over the subsequent 36 months. We then estimate a discrete choice model (linear probability and logit, both shown in Table 8) using variables that are available in McDash to predict 36-month defaults for these loans. We use an analogous strategy for the 60-month horizon (i.e., we take all loans originated between 72 and 60 months prior and track them over the subsequent 60 months). The regressions include the same set of covariates that are included in the *ex-post* default regressions above. We take the estimated coefficients from each quarterly credit risk model and apply them to the characteristics of the loans originated in the current quarter to create 36-month, loan-level, predicted default probabilities. This leaves us with an *ex-ante* credit risk measure that uses only information available at the time of issuance.

We take the predicted default probabilities and substitute them into Equation 5 to estimate the relation between time to sale and observable default risk. Table 8 shows the results. We find positive coefficients for all models in Panel A (PLS loans), which is consistent with the intuition that observably better loans tend to transact faster. We observe no relation between *ex-ante* risk and time to sale for GSE loans (Panel B). Figure 1 plots the non-parametric coefficients where *ex-ante* risk is used as the outcome variable, and this again shows that loans sold later have a (weakly) higher predicted default probability (not lower, as in the case of *ex post* default). These differences could be driven by the demand side, in particular if there is higher demand for observably safer securities. This would also be consistent with the fact that observably riskier PLS loans take longer to sell than GSE loans.

These patterns are in stark contrast to the estimated relation between *ex-post* default rates and time to sale in Section 4.1, and they confirm that the distinction between observable and unobservable risk is crucial for our tests.

4.5 Default and Time to Sale in the CoreLogic Data Set

Table 9 displays results concerning the relation between *ex-post* default risk and time to sale using the sample of PLS loans in CoreLogic. One of the main reasons for using CL data is the ability to control for the identity of mortgage originators and security issuers. Panel A focuses on the effect of controlling for originator heterogeneity, while Panel B focuses on issuer heterogeneity. Issuer information corresponding to the private financial institution responsible for pooling and securitizing the mortgages is obtained from Bloomberg. In each panel, we display results for the full sample of PLS loans (columns (1)–(3)), the sample of Alt-A PLS loans (columns (4)–(6)), and the sample of subprime PLS loans (columns (7)–(9)), and use a default definition of 60+ days delinquency over a 36 month horizon.

In columns (1), (4), and (7) of Panel A, we do not control for originator heterogeneity, so these results are directly comparable to the McDash results displayed in Table 5. In columns (2), (5), and (8) we include a full set of originator fixed effects. Information on the originators is available for slightly more than half of the loans in the CL data set, so we focus our analysis on this subsample across all specifications. Finally, columns (3), (6), and (9) display results from a specification that includes originator-by-year-quarter-of-origination fixed effects. This is a reasonably demanding test, as it uses variation on months to sale and default from loans originated by the same institution in the same year-quarter to estimate the regressions.

In the full sample of PLS loans, the estimates are negative and statistically significant, but smaller in magnitude than the corresponding estimates obtained using McDash data. The coefficient in column (1), which corresponds exactly to our McDash specification (i.e., no originator fixed effects) is -0.36 percentage points, roughly one-third of the magnitude of the estimate in Table 5 (-0.94 percentage points). We return to this comparison below when we separate loans into the Alt-A and subprime segments of the market. In column (2), we observe that the inclusion of originator fixed effects slightly decreases (in absolute magnitude) the coefficient associated with months to sale, while the inclusion of originator-by-year-quarter-of-origination fixed effects (column (3)) further decreases the magnitude, although the estimate remains negative and statistically significant.

In column (1) of Panel B, we display results from a specification that includes both originator and issuer fixed effects. Compared to the specification with only originator fixed effects (column (2) in Panel A), the estimated effect increases (in absolute magnitude) from -0.28 to -0.41 percentage points. In column (2), we add issuer-by-year-quarter-of-issuance fixed effects, which approximately halves the magnitude of the coefficient. Finally, column (3) displays results from a specification that includes both originator-by-year-quarter-of-origination and issuer-by-year-quarter-of-issuance fixed effects. The estimated effect remains negative and statistically significant, and an additional month of delay is associated with a 0.17-percentage-point decrease in the likelihood of default, all else equal.

4.5.1 Alt-A PLS vs. Subprime PLS

In addition to the information on the identities of originators, an advantage of using CL data is the ability to analyze different segments of the PLS market. *A priori*, we expect signaling unobservable mortgage quality to have a larger role in the Alt-A segment of the PLS market because it is largely comprised of low-documentation mortgages. Table 4 shows that over 70% of Alt-A mortgages were less than fully documented, compared to 35% of subprime loans. Industry sources suggest that at least some of the Alt-A loans that appear as "fully documented" may also suffer from documentation issues that prevent them from being sold in the GSE (conforming) market.

Columns (4)–(9) in Table 9 display results from separately estimating regressions for the subprime and Alt-A segments of the PLS market, and Figure 2 plots results for the non-parametric specifications.[3] The differences between the subprime and Alt-A results are fairly striking, and they help to explain the source of the differences between the McDash and CL results. There is a significantly weaker relation between *ex-post* default risk and time to sale among subprime loans, while for Alt-A loans the magnitudes are similar to those estimated in the McDash sample. As discussed above, when we compare summary statistics between McDash and CL (Tables 3 and 4), it appears that the McDash sample of PLS loans is more similar to the Alt-A mortgage sample than the subprime sample in CL. This can rationalize the differences in the quantitative magnitudes of the estimates derived from each subsample in CL. In addition, the Alt-A results are much less sensitive to the inclusion of originator-by-time and issuer-by-time fixed effects. In contrast, results for the subprime sample are highly sensitive and even disappear in some specifications.

In the Online Appendix (Table A.10) we present results following restriction of the sample of Alt-A and subprime loans to those that have not become delinquent 9 months after origination. This is the same correction that we implemented and discussed in Section 4.2. We find that the coefficient on Alt-A loans is slightly reduced in magnitude relative to the whole sample case (for example, without originator or issuer fixed effects, the coefficient moves from 0.072 in Table 9 to 0.054. All specifications (with and without originator and issuer fixed effects) are still highly statistically significant. The results for subprime loans are similarly affected.

These results suggest that signaling of default risk did not play an important role in the subprime

[3]In the Online Appendix we also show results for estimations where we extend the maximum horizon for time to sale to 12 and 24 months (Section A.3). The results are robust to these changes.

segment of the PLS market (as opposed to the Alt-A segment). However, there is evidence from the literature that adverse selection with respect to prepayment risk may have played a role in this market (Agarwal et al. (2014)) and that prepayment risk was an important concern for PLS investors in the pre-crisis period. For example, in a 2006 primer on mortgage-backed securities, the American Securitization Forum wrote, "Prepayment risk is the key source of cash flow uncertainty in RMBS [Residential Mortgage Backed Securities]. " (ASF (2006)). As a final test, we regress prepayment risk for hybrid adjustable rate mortgages (ARMs) on our time to sale variable. We discuss these results in detail in Section A.6 the Online Appendix. A negative prepayment event is defined as a borrower prepaying more than six months before the date on which the mortgage resets. We find a negative relation between time to sale and the likelihood of an early prepayment (Table A.7 of the Online Appendix). The majority of the hybrid ARMs that we consider are in the subprime segment of the market, which suggests that asymmetric information on prepayment risk may have been a more relevant factor in that market rather than that on credit risk.

4.5.2 Affiliation Results

In this section, we test whether an affiliation between the originator (seller) and the issuer (buyer) plays a role in the relation between time to sale and default risk. Many issuers and originators in the PLS market share direct relationships. In some cases, the originator and issuer are the same institution, while in others they are part of the same vertically integrated corporation (in which case the originator is typically a subsidiary of the issuer). *A priori*, we expect that the scope for private information between an originator and issuer who are affiliated is less than that between an originator and issuer that are independent entities (Demiroglu and James (2012b), Furfine (2014)). Thus, if signaling is driving our results, we expect a weaker negative relation between time to sale and default risk for the sample of loans in which the issuer and originator are affiliated with each other.

We obtain information on the identity of issuers from Bloomberg and supplemented this with hand-collected data from the pooling and service agreements (PSA) associated with the PLS deals.[4] There is some uncertainty about whether the originator field in the CoreLogic database actually corresponds to the lender of record (i.e., the institution that underwrote and originated the loan) or to another institution in the intermediation chain (Stanton et al. (2014)). In Section A.2.2.1 in the Online Appendix we discuss an external check we performed to ensure that the originator we observe is, indeed, the lender of record.

We focus only on loans that are in deals in which either all of the loans were made by affiliated originators or all of the loans were made by unaffiliated originators. Table 10 and Figure 3 display the results. As in our analysis of documentation status above, we stratify our sample of all PLS loans and our separate Alt-A and subprime samples by affiliation status.

The negative correlation between time to sale and default risk does not appear to be sensitive to affiliation status in the full sample (columns 1–2 in Table 10). However, the difference between

[4] PSAs are available at the SEC's EDGAR website: http://www.sec.gov/edgar/searchedgar/companysearch.html.

affiliated and unaffiliated issuers and originators is significant in the Alt-A segment of the market. Loans sold six months after origination by affiliated originators are approximately 3 percentage points less likely to default compared to loans sold in the month of origination (column 3 of Panel B in Table 10), while this effect increases to almost 9 percentage points for loans originated by unaffiliated originators. Panel A in Figure 3 shows that this difference is highly statistically significant over the entire distribution of time to sale.

There is some ambiguity regarding the exact places in the mortgage funding chain in which asymmetric information might play an important role. One possibility is that it occurs between the originator and the issuer, while a second possibility is that it occurs between the issuer and the ultimate PLS investors. The affiliation results shed light on this issue because the two possibilities yield different predictions. If asymmetric information is present between the originator and issuer, we would expect to find a weaker relationship between time to sale and default for affiliated institutions. However, if asymmetric information is present between the issuer and PLS investors, then we might expect to find a stronger relationship for affiliated originators and issuers because investors may perceive that issuers are more likely to obtain private information on mortgage quality when they have an affiliation with the originators. Our finding of a weaker relationship shown in Table 10 suggests that asymmetric information between originators and issuers plays a more important role.

4.5.3 Documentation Results

We further explore the role of documentation standards by stratifying our PLS sample into loans with full documentation of income and assets and loans with less than full documentation ("low doc"). We stratify by documentation type for the full sample of PLS loans and for our separate subprime and Alt-A samples. The results are displayed in Table 11, with Panel A containing the results for the linear specifications and Panel B containing the results for the non-parametric specifications.

The results are mixed. In the sample of all PLS loans (subprime and Alt-A combined), there appears to be a stronger negative relation between time to sale and default for low-documentation loans compared to full-documentation loans. This negative relation is approximately 50% larger (in absolute value) in the sample of low-documentation PLS loans (columns 1–2). However, the results in columns 3–6 (breaking down loans into Alt-A and subprime) show that there are essentially no differences between full-documentation and low-documentation loans within each of the two subsamples.

4.6 Security Spreads and Time To Sale

We now turn to the empirical relation between time to sale and security prices. The unit of observation for this analysis is a pool, i.e., a group of loans from which different securities in each private-label deal derive cash flows. We focus our attention on floating rate securities and show results using two security samples: (i) only triple-A tranches, and (ii) all tranches, including lower-rated securities. Including securities below triple-A makes aggregation less straightforward and reduces the sample used in the regressions because many of these securities were issued at fixed rates (see the discussion in Section 3.2).

Table 12 displays results from regressing the average pool-level spreads on average pool-level seasoning. We include pool-level averages of FICO and CLTV as covariates, as well as the fraction of loans in each pool that fall into various FICO and CLTV categories (displayed in Table A.8). Panel A shows the results when we focus on triple-A securities and we include only a linear term for average seasoning; Panel B also focuses on triple-A securities and includes a quadratic term to capture potential non-linearities; and Panel C extends the analysis to all securities, including those in lower-rated tranches. The results on *ex-post* realized default rates discussed above were significantly different in the sample of mortgages that collateralized Alt-A securities than in the sample of loans that backed subprime securities. Thus, in all panels, we show results for the full sample of floating-rate securities (columns 1–3) and results for Alt-A (columns 4–6) and subprime (columns 7–9) securities separately to determine whether similar patterns emerge on the pricing dimension.

In Table 12, we display results for three different regression specifications. The first specification includes the quarter of issuance fixed effects but no other control variables. The second specification includes the list of pool-level controls listed in the Online Appendix (Table A.8) along with the quarter of issuance fixed effects. The third specification includes a full set of issuer fixed effects in addition to the pool-level controls and month of issuance fixed effects.

Column (1) in Panel A shows that one additional month of average mortgage seasoning is associated with a 1.5-basis-points lower yield spread, which is about 5% of the average spread in the sample (28 basis points). When pool-level controls and both issuer and month of issuance fixed effects are included (column (3)), the coefficient estimate declines slightly but remains statistically significant. Similar to our findings in the default analysis above, we see in columns (4)–(9) that this effect is concentrated in the Alt-A sample. For Alt-A securities, one additional month of average mortgage seasoning is associated with a 2.4-basis-points lower yield spread. The results in Panel C using all tranches are consistent with those in Panel A, although we find a larger magnitude of the effects for the Alt-A sample (to 3.4 basis points per month of seasoning).

For the non-linear specification results reported in Panel B, both the linear and the quadratic terms are significant in the full sample and the Alt-A sample. The linear terms are negative and the quadratic terms are positive, which implies a non-linear relation between time to sale and security spreads similar to the relation that we documented above between time to sale and mortgage default. Figure 5 displays the predicted security spreads as a function of the average pool-level seasoning calculated using the estimation results from the specification reported in column (6) in Panel B. The figure includes 95% confidence intervals calculated using the delta method. The plot reveals several notable points. First, the minimum spread as a function of average seasoning is achieved between four and five months. Second, after five months, the spread begins to increase in seasoning; however, the confidence bands show that we begin to lose precision for seasoning greater than five months because there are so few securities in the data set with high values of average seasoning (Figure 4).

5 Quantifying the cost of signaling

In this subsection, we use a calibrated version of our theoretical model to quantify the economic magnitude of the signaling costs borne by originators. The model we present in Section 2 is somewhat unrealistic in two important respects. First, the model assumes that there is zero recovery in the event of default. Second, the baseline model assumes that there is no delay in trade that is unrelated to signaling. However, we observe that even in the case of GSE mortgages for which there is limited or no scope for private information, mortgages do not always sell immediately. We make two main changes to our model. First, to address recovery in default, we assume that, in the event of default, the mortgage returns $\theta < 1$ of its initial face value B_0. Second, to accommodate the possibility of delay in trade that is unrelated to private information, we assume that there is an exogenous random delay between the time at which an originator offers a mortgage to the market and when a trade actually occurs. For simplicity, we assume that this delay, denoted ν, is an exponential random variable with parameter η and is publicly observable. One interpretation of this element of delay is that, once the originator offers the mortgage to the market, a random amount of time passes before she is matched with a set of buyers. Importantly, these buyers observe the time at which the originator chose to offer the mortgage to the market. In the Online Appendix, we show that, under these assumptions, the equilibrium strategy of the originator remains the same as in Proposition 1. The price for a mortgage is now pinned down by the date at which the originator first offers the mortgage to the market and not the date at which the actual sale takes place.

5.1 Calibration

This subsection describes the procedure used to calibrate the parameters of the model. First, we assume the payment c is set so that

$$c = r_m B_0 \tag{8}$$

where r_m is the annualized percentage rate (APR) of the mortgage. Calculating the payment this way ignores amortization, but given that the effects we document are in the first nine months of the mortgage or less, amortization will not have a large quantitative impact on the results. The parameters B_0, r_m and α are all directly measured in the data. We set B_0 and r_m to match the summary statistics reported in Table 4. We set θ to be consistent with observed recovery rates on foreclosed loans. An and Cordell (2017) estimate recovery rates on foreclosed agency (GSE) loans of between 75% and 95% in the 2002–2007 period (see Figure 2 in their paper). Figure 1 in Adelino et al. (2013) shows that self-cure rates were between 50% and 60% for PLS loans that became 60 days delinquent between 2005 and 2007, which implies a probability of foreclosure conditional on 60-day delinquency of about 50%. Given that our default event is 60+ days delinquent and not foreclosure, we choose a value of θ of 0.90. This is the combination of a 50% likelihood that a delinquent loan terminates in foreclosure and the average recovery rate conditional on foreclosure of about 80%. Finally, we set the rate of delay unrelated to signaling to be $\eta = 9.73$ (in annual units) to be consistent with the average delay in trade of 1.2 months found in the GSE market implied by the distribution of time to sale in Table 1 (using loans sold up to month 9 to match our empirical exercise).

The remaining parameters of the model are not directly measurable, so we calibrate them using simulated data and our model. We first assume that the distribution of expected default rates, $f(\lambda)$, is given by

$$f(\lambda) = \frac{1}{(\lambda_h - \lambda_\ell)^{1+\alpha+\beta}} \frac{(\lambda - \lambda_\ell)^\alpha (\lambda_h - \lambda)^\beta}{\mathrm{B}(\alpha, \beta)} \tag{9}$$

where B is the Beta function. This distribution is flexible enough to allow us to closely match the empirical distribution of time to sale. We set

$$\lambda_\ell = \frac{\bar{\lambda} - \frac{\alpha}{\alpha+\beta}\lambda_h}{1 - \frac{\alpha}{\alpha+\beta}}, \tag{10}$$

where $\bar{\lambda}$ is the annualized average expected default rate of the mortgages in our sample (i.e. the private-label loans used to construct the estimates shown in Table 8). Equation (10) implies that $E[\lambda] = \bar{\lambda}$, so that $\bar{\lambda}$ pins down the level of the distribution, and thus guarantees that the average quality of mortgages in our calibration matches the average quality in the data.

We then proceed with the following steps:

1. Specify initial parameter estimates r, γ, λ_h, α and β.

2. Simulate a sample of two million draws of λ and of ν. Each draw of (λ, ν) represents a mortgage with the expected default rate λ that will face a delay of ν between when the originator offers the mortgage for sale and the actual sale date.

3. For each mortgage, we simulate a default time τ given its expected default rate.

4. Apply the equilibrium strategy for the originator for each mortgage given its expected default rate and add ν to get a proposed sale date for the mortgage $S = T(\lambda) + \nu$.

5. Form a sample of mortgages for which the proposed sale date is prior to the simulated default time $S \leq \tau$. We call this the sample of sold mortgages.

6. Within the sample of sold mortgages, replicate the regression that is reported in Table 9, Panel A, Column (4). In other words, regress a dummy for whether a mortgage has defaulted in the first 36 months and the date at which it is sold (rounded down to the nearest month).

7. Within the sample of sold mortgages, calculate the distribution of sale times.

8. Within the entire sample of mortgages calculate the average of the gross proceeds to the originator.

9. Repeat steps until the regression coefficient calculated in Step 6 roughly matches that in Table 9, the distribution calculated in Step 7 matches that in Table 2, and the value to the originator calculated in Step 8 matches B_0. This comparison is equivalent to assuming that originators break-even on average.

Table 13 (Panel A) reports the calibrated parameter values.

Although the procedure simultaneously calibrates r, γ, λ_h, α, β, it is useful to emphasize that each parameter is largely determined by a particular moment (or set of moments) in the data. The difference between r and γ is primarily pinned down by the regression coefficient in Table 9, Panel A, Column (4). Note that while this difference is large, it likely does not represent the difference between the long term discount rates between the originator and the issuer, which could be much smaller. The parameters α and β are a function of the time-to-sale distribution in Table 2. Finally, the parameters λ_h and γ are estimated assuming that originators break even on average.

5.2 Model output

Table 13 (Panel B) and Figure 6 report how closely our simulations match the data. Importantly, we match the regression coefficient of default on time to sale from Table 9 and the face value of the mortgage almost exactly. Figure 6 displays the simulated distribution of time to sale compared to the actual distribution, which also matches quite closely. Table 13 (Panel B) also reports the regression coefficient of default on time to sale in the unrestricted sample. Here our calibrated model implies a somewhat larger coefficient than the actual one, suggesting that the difference between r and γ might not be as large as we calibrate. To account for this, we examine the sensitivity of our quantification of the loss due to asymmetric information to the rate r. Finally, we simulate our pricing regression by forming 400 pools of 2000 mortgages each and regress the pool yield on time to sale with results reported in Table 13 (Panel B). Here our model underestimates the effect somewhat suggesting that the difference between r and γ is underestimated.

To evaluate the economic magnitude of the cost of signaling, we calculate the expected loss in surplus attributable to the incremental delay in trade due to signaling. Using our simulated data, we calculate the average proceeds to the originator and compare this with the average proceeds that the originator would collect without asymmetric information. Note that this latter quantity still accounts for random delay ν. Table 13 (Panel B) reports the results of this exercise for our calibrated parameters and for a 100 bps increase and decrease in γ relative to the main calibration. The loss in value due to signaling is \$536 per mortgage, or about 18 bps of average principal balance of the mortgages in our sample. One way to interpret this number is as the amount the originator would be willing to pay to transact with full information instead of asymmetric information and signaling. To put this quantity in perspective, if we suppose that this cost is passed on to borrowers, then the originator would charge 18 bps just to cover signaling costs.

Interestingly, Table 13, Panel C, shows the percentage loss in surplus is somewhat insensitive to the originator's discount rate. This is because changing the magnitude of gains from trade has two effects. First, increasing the magnitude of the gains from trade, i.e., increasing γ, allows the originator to signal quality using a shorter holding period. Second, when gains from trade are larger, any given delay in trade is associated with larger loss in surplus. In a neighborhood of $\gamma = 7.92\%$, these two effects cancel out.

6 Implications for the allocation of credit

In addition to leading to a loss in efficiency in the secondary market for mortgages, asymmetric information between mortgage originators and MBS issuers can lead to a distortion in the allocation of credit in the primary market. Vanasco (2017) identifies two externalities that trading in the secondary market can have on primary market outcomes when originators of new assets can exert effort that both improves asset quality and screens for good assets. On the one hand, an originator who anticipates (at least partially) selling mortgages to a secondary market does not fully internalize the effect that her screening effort will have on mortgage quality. This effect lowers the originator's effort relative to first-best, and as a result, mortgages with too high an expected default rate will be originated in equilibrium. On the other hand, if the originator anticipates gaining some private information in the screening process that she must later signal to the market, then she will seek to minimize her signaling costs by improving screening. This effect increases the originator's effort relative to first-best, and as a result, too few low quality mortgages will be originated. Vanasco shows that the first effect dominates the second when gains from trade are large. Our numerical calibration suggests that gains from trade in the secondary mortgage market are quite large. Thus, viewed through the lens of Vanasco's model, our findings are consistent with an under-provision of originator effort in the primary mortgage market.

Chemla and Hennessy (2014) identify another channel through which secondary market illiquidity can affect primary market efficiency. In their model, expected signaling costs decrease the difference between the payoff for good and bad type assets. This in turn lowers an originator's incentives to originate good quality assets. In our simple calibration above, signaling costs are substantial in the market between mortgage originators and MBS issuers. These costs imply another reason why there is an under-provision of screening effort in the primary market for mortgages.

Central to the allocative inefficiency in the primary market in Vanasco (2017) and Chemla and Hennessy (2014) is the asymmetric information problem between the originator of an asset and potential buyers. Thus, a natural question to ask is whether it is possible to mitigate this information asymmetry to restore efficiency in the primary market. Daley et al. (2017) show that public information revealed after sellers have committed to an action can reduce signaling costs *ex ante*, i.e., mitigate information asymmetries, while decreasing primary market efficiency. The intuition is that public information allows some pooling to occur in equilibrium, which in turn decreases originator incentives to screen. In our setting, the interim performance of a mortgage prior to sale is a form of public information, provided it is credibly released to investors. One implication of the work of Daley et al., is that releasing this type of information to investors could actually decrease the efficiency of the primary mortgage market.

7 Conclusion

A general feature of models of asymmetric information and delayed trade is that the prices and (unobserved) quality of goods increase over time. This paper provides some of the first empirical evidence of this prediction in the context of the residential mortgage market. Using detailed loan-level data on privately securitized mortgages, we find a statistically significant and economically meaningful positive correlation between conditional *ex-post* mortgage performance and time to sale. This finding is robust to different ways of measuring performance and, importantly, is not generated by the component of mortgage performance that is predictable by buyers using *ex-ante* observable information on underwriting characteristics. Using a calibrated version of our model, we find a substantial loss in surplus due to asymetric information of around 18 bps of the average mortgage's face value. Furthermore, the positive relation between time to sale and mortgage performance is not present in the agency securitization market, in which adverse selection between originators and issuers along the credit dimension cannot take place. This estimated correlation is stronger for deals where the originator and the issuer are not affiliated, and it is strongest in the Alt-A segment of the market.

Taken together, the results both confirm the importance of private information in the non-agency securitization market and provide evidence consistent with a signaling mechanism by which lenders in the market are able to reveal the quality of their loans by delaying trades.

References

Adelino, M., W. S. Frame, and K. Gerardi (2017). The effect of large investors on asset quality: Evidence from subprime mortgage securities. *Journal of Monetary Economics 87*, 34–51.

Adelino, M., K. Gerardi, and P. S. Willen (2013). Why don't lenders renegotiate more home mortgages? redefaults, self-cures and securitization. *Journal of Monetary Economics 60*(7), 835–853.

Agarwal, S., Y. Chang, and A. Yavas (2012). Adverse selection in mortgage securitization. *Journal of Financial Economics 105*(3), 640–660.

Agarwal, S., Y. Chang, and A. Yavas (2014). Adverse selection in mortgage securitization. *Journal of Financial Economics 105*(3), 640–660.

Aiello, D. (2016). Information exploitation? a pre-crisis rmbs issuer's private information. Working paper, University of California Los Angeles.

Akerlof, G. A. (1970). The market for lemons: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

An, X. and L. Cordell (2017). Regime shift and the post-crisis world of mortgage loss severities. Technical report, Federal Reserve Bank of Philadelphia.

An, X., Y. Deng, and S. A. Gabriel (2011). Asymmetric information, adverse selection, and the pricing of cmbs. *Journal of Financial Economics 100*(2), 304–325.

ASF (2006). American securitization institute: Residential mortgage-backed securities.

Ashcraft, A., P. Goldsmith-Pinkham, P. Hull, and J. Vickery (2011). Credit ratings and security prices in the subprime mbs market. *The American Economic Review 101*(3), 115–119.

Ashcraft, A. B., P. Goldsmith-Pinkham, and J. I. Vickery (2010). Mbs ratings and the mortgage credit boom. *FRB of New York Staff Report* (449).

Begley, T. A. and A. Purnanandam (2017). Design of financial securities: Empirical evidence from private-label rmbs deals. *Review of Financial Studies 30*(1), 120–161.

Chang, B. (2014). Adverse selection and liquidity distortion. Working paper, University of Wisconsin.

Chemla, G. and C. A. Hennessy (2014). Skin in the game and moral hazard. *The Journal of Finance 69*(4), 1597–1641.

Cho, I. and D. Kreps (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 179–221.

Daley, B. and B. Green (2012). Waiting for news in the market for lemons. *Econometrica 80*(4), 1433–1504.

Daley, B. and B. Green (2016). An information-based theory of time-varying liquidity. *Journal of Finance 71*(2), 809–870.

Daley, B., B. S. Green, and V. Vanasco (2017). Securitization, ratings, and credit supply. *Working Paper*.

DeMarzo, P. M. (2005). The pooling and tranching of securities: A model of informed intermediation. *Review of Financial Studies 18*(1), 1–35.

DeMarzo, P. M. and D. Duffie (1999). A liquidity-based model of security design. *Econometrica 67*(1), 65–100.

Demiroglu, C. and C. James (2012a). How important is having skin in the game? originator-sponsor affiliation and losses on mortgage-backed securities. *Review of Financial Studies 25*(11), 3217–3258.

Demiroglu, C. and C. James (2012b). How important is having skin in the game? originator-sponsor affiliation and losses on mortgage-backed securities. *Review of Financial Studies 25*, 3217–3258.

Downing, C., D. Jaffee, and N. Wallace (2009). Is the market for mortgage-backed securities a market for lemons? *Review of Financial Studies 22*(7), 2457–2494.

Fuchs, W., B. S. Green, and D. Papanikolaou (2015). Adverse selection, slow moving capital and misallocation. *Journal of Financial Economics*.

Fuchs, W. and A. Skrzypacz (2013). Costs and benefits of dynamic trading in a lemons market. *Working Paper*.

Fudenberg, D. and J. Tirole (1983). Sequential bargaining with incomplete information. *The Review of Economic Studies*, 221–247.

Furfine, C. H. (2014). Complexity and loan performance: Evidence from the securitization of commercial mortgages. *Review of Corporate Finance Studies 2*, 154–187.

Ganduri, R. (2016). Repo regret? Working paper, Emory University.

Garmaise, M. J. (2015). Borrower misreporting and loan performance. *The Journal of Finance 70*(1), 449–484.

Garmaise, M. J. and T. J. Moskowitz (2004). Confronting information asymmetries: Evidence from real estate markets. *Review of Financial Studies 17*(2), 405–437.

Genesove, D. (1993). Adverse selection in the wholesale used car market. *Journal of Political Economy*, 644–665.

Grenadier, S. R. and A. Malenko (2011). Real options signaling games with applications to corporate finance. *Review of Financial Studies 24*(12), 3993–4036.

Grenadier, S. R. and N. Wang (2005). Investment timing, agency, and information. *Journal of Financial Economics 75*(3), 493–533.

Griffin, J. M. and G. Maturana (2016). Who facilitated misreporting in securitized loans? *The Review of Fianncial Studies 29*(2), 384–419.

Guerrieri, V., R. Shimer, and R. Wright (2010). Adverse selection in competitive search equilibrium. *Econometrica 78*(6), 1823–1862.

He, J. J., J. Q. Qian, and P. E. Strahan (2012). Are all ratings created equal? the impact of issuer size on the pricing of mortgage-backed securities. *The Journal of Finance 67*(6), 2097–2137.

Janssen, M. C. and S. Roy (2002). Dynamic trading in a durable good market with asymmetric information. *International Economic Review 43*, 257–282.

Jiang, W., A. A. Nelson, and E. Vytlacil (2014a). Liar's loan? effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics 96*(1), 1–18.

Jiang, W., A. A. Nelson, and E. Vytlacil (2014b). Securitization and loan performance: A contrast of ex ante and ex post relations in the mortgage market. *Review of Financial Studies 27*(2), 454–483.

Kramer, B. and G. Sinha (2006). Bear stearns quick guide to non-agency mortgage-backed securities. Technical report, Bear Stearns.

Kremer, I. and A. Skrzypacz (2007). Dynamic signaling and market breakdown. *Journal of Economic Theory 133*(1), 58–82.

Leland, H. E. and D. H. Pyle (1977). Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance 32*(2).

Noldeke, G. and E. Van Damme (1990). Signalling in a dynamic labour market. *The Review of Economic Studies 57*(1), 1–23.

Piskorski, T., A. Seru, and J. Witkin (2015). Asset quality misrepresentation by financial intermediaries: evidence from the rmbs market. *The Journal of Finance 70*(6), 2635–2678.

Saengchote, K. (2013). Soft information in the subprime mortgage market. In *26th Australasian Finance and Banking Conference*.

Sengupta, R. (2010). Alt-a: The forgotten segment of the mortgage market. *Federal Reserve Bank of St. Louis Review 92*(1), 55–71.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 355–374.

Stanton, R., J. Walden, and N. Wallace (2014). The industrial organization of the us residential mortgage market. *Annu. Rev. Financ. Econ. 6*(1), 259–288.

Stanton, R., J. Walden, and N. Wallace (2015). Mortgage loan-flow networks and endogenous financial norms.

Stearns, B. (2006). Bear stearns quick guide to non-agency mortgage-backed securities. *Residential Mortgage Criteria Report*.

Swinkels, J. M. (1999). Education signalling with preemptive offers. *The Review of Economic Studies 66*(4), 949–970.

Vanasco, V. (2017). The downside of asset screening for market liquidity. *The Journal of Finance 72*(5), 1937–1982.

Williams, B. (2016). Search, liquidity, and retention: Signaling multidimensional private information. Technical report, Working Paper, NYU.

Table 1: Distribution of Time to Sale in the McDash Sample

	PLS Loans		**GSE Loans**	
Months to Sale	# Loans	Cumulative %	# Loans	Cumulative %
0/1	3,278,163	54.8	8,250,788	58.30
2	1,414,587	78.4	4,771,199	92.01
3	584,061	88.2	682,223	96.83
4	210,834	91.7	171,949	98.04
5	93,901	93.3	74,302	98.57
6	62,138	94.3	37,419	98.83
7	44,893	95.0	23,329	99.00
8	32,973	95.6	18,788	99.13
9	26,204	96.0	15,845	99.24
≥ 10	237,954	100.0	107,432	100.00

This table displays the distribution of the number of months between the time of origination and the time of sale (months to sale) for privately securitized mortgages in the McDash data set. The McDash sample includes only first-lien mortgages originated between January 2002 and December 2007. The sample is further restricted to mortgages that enter the data set in either the month of origination or the month following origination.

Table 2: Distribution of Time to Sale in the CoreLogic PLS Sample

	All PLS		**Subprime**		**Alt-A**	
Months to Sale	# Loans	Cum. % of Sample	# Loans	Cum. % of Sample	# Loans	Cum. % of Sample
0	2,446,106	17.9	1,079,646	12.4	1,366,460	27.7
1	3,675,646	44.8	2,296,307	38.7	1,379,339	55.6
2	2,952,576	66.4	2,026,277	62.0	926,299	74.3
3	2,064,585	81.6	1,521,350	79.4	543,235	85.3
4	1,149,410	90.0	861,916	89.3	287,494	91.1
5	571,103	94.2	415,989	94.1	155,114	94.3
6	286,959	96.3	201,827	96.4	85,132	96.0
7	140,231	97.3	86,683	97.4	53,548	97.1
8	87,131	97.9	51,849	98.0	35,282	97.8
9	56,839	98.3	32,197	98.4	24,642	98.3
≥ 10	228,536	100.0	85,146	100.0	143,390	100.0

This table displays the distribution of the number of months between the time of origination and the time of sale (months to sale) for privately securitized mortgages in the CoreLogic data set. The CoreLogic sample includes only first-lien mortgages backing subprime and Alt-A PLS that were originated between January 2002 and December 2007. The time of sale corresponds to the month in which the PLS security was issued.

Table 3: Summary Statistics: McDash Sample

	PLS Loans		**GSE Loans**	
	Mean	SD	Mean	SD
Loan/Borrower Characteristics (continuous variables)				
Term	351	52	325	71
Original Rate	5.91	1.90	6.05	0.80
Original Amount	297,898	201,098	171,454	87,557
LTV Ratio	72.7	15.1	73.4	18.5
FICO	702	67	714	63
Loan/Borrower Characteristics (dummy variables)				
	Mean		Mean	
Purchase (d)	0.471		0.463	
Cash Out Refinance (d)	0.196		0.156	
Arm (d)	0.497		0.119	
Balloon (d)	0.007		0.004	
Interest Only (d)	0.222		0.052	
"B" or "C" Grade (d)	0.166		0.011	
Jumbo (d)	0.304		0.004	
Low Doc (d)	0.151		0.134	
Prepay Penalty (d)	0.261		0.099	
Primary Residence (d)	0.871		0.892	
Single Family (d)	0.823		0.853	
Geographic Characteristics				
	Mean	SD	Mean	SD
Unemployment rate (county-level)	4.9	1.5	5.1	1.5
36 month unemployment growth (%)	0.7	40.9	14.8	46.4
Price Index (county-level)	184	53	158	44
36 month HPA (%)	42.9	26.1	30.7	22.0
Default Rates				
	Mean		Mean	
60+ DQ, 36-month horizon	0.160		0.090	
60+ DQ, 36-month horizon	0.149		0.078	
60+ DQ, 60-month horizon	0.210		0.114	
90+ DQ, 36-month horizon	0.127		0.060	
90+ DQ, 60-month horizon	0.189		0.094	
# Loans	5,747,722		14,045,839	

This table displays summary statistics for both privately securitized mortgages (PLS) and mortgages acquired by the housing GSEs (Fannie Mae and Freddie Mac) in the McDash data set. The McDash sample includes only first-lien mortgages originated between January 2002 and December 2007 that were sold to either PLS issuers or the GSEs within nine months of origination (inclusive). All variables in the table are included as covariates in the regressions below. For a full list of covariates, see Section A.1 of the Online Appendix.

Table 4: Summary Statistics: CoreLogic Sample

	All PLS		**Subprime**		**Alt-A**	
	Mean	SD	Mean	SD	Mean	SD
Loan/Borrower Characteristics (continuous variables)						
Term	356	37	355	33	358	47
Original Rate	7.48	1.57	7.87	1.33	6.26	1.62
Original Amount	214,855	150,813	190,628	125,503	291,003	192,566
LTV Ratio	83.0	14.3	83.7	14.0	80.8	15.1
FICO	639	70	617	61	710	48
Loan/Borrower Characteristics (dummy variables)						
	Mean		Mean		Mean	
Purchase (d)	0.395		0.363		0.495	
Cash Out Refinance (d)	0.500		0.552		0.339	
Arm (d)	0.741		0.763		0.669	
Balloon (d)	0.070		0.090		0.009	
Interest Only (d)	0.184		0.117		0.391	
Jumbo (d)	0.129		0.089		0.257	
Low Doc. (d)	0.442		0.351		0.728	
Prepay Penalty (d)	0.661		0.745		0.394	
Primary Residence (d)	0.870		0.919		0.716	
Single Family (d)	0.743		0.782		0.622	
Geographic Characteristics						
	Mean	SD	Mean	SD	Mean	SD
Unemployment rate (county-level)	5.23	1.58	5.34	1.59	4.88	1.49
36-month unemployment growth (\%)	5.4%	39.3%	8.8%	40.3%	-5.2%	33.7%
36-month HPA (\%)	42.3%	26.3%	40.7%	26.1%	47.1%	26.3%
Default Rates						
	Mean		Mean		Mean	
60+ Days Delinquent, 36-month horizon	0.304		0.333		0.215	
60+ Days Delinquent, 60-month horizon	0.372		0.390		0.318	
90+ Days Delinquent, 36-month horizon	0.251		0.272		0.186	
90+ Days Delinquent, 60-month horizon	0.327		0.339		0.291	
# Loans	7,868,492		5,969,285		1,899,207	

This table displays summary statistics for loans backing subprime and Alt-A PLS in the CoreLogic data set. The CoreLogic sample includes only first-lien mortgages originated between January 2002 and December 2007. In addition, the sample only includes loans that were sold to PLS issuers within nine months of origination (inclusive). All variables in the table are included in the regressions below.

Table 5: Baseline Parametric Results for the Sample of PLS Loans in McDash

Panel A: Full Sample

Default Horizon:	36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale	-0.0094	-0.0098	-0.0097	-0.0108
	(10.75)	(11.69)	(11.47)	(13.13)
# Loans	5,747,722	5,747,722	5,747,722	5,747,722
Adjusted R^2	0.24	0.22	0.26	0.25
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

Panel B: Restricted Sample

Default Horizon:	36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale	-0.0095	-0.0092	-0.0101	-0.0104
	(11.47)	(11.56)	(12.34)	(13.03)
# Loans	5,574,463	5,574,463	5,574,463	5,574,463
Adjusted R^2	0.2	0.19	0.24	0.23
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 on PLS loans in the McDash data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that default over a 36-month horizon (columns 1 and 2) and over a 60-month horizon (columns 3 and 4). Default is defined as a loan that is 60+ days delinquent (columns 1 and 3) and 90+ days delinquent (columns 2 and 4). Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, and the detailed list of covariates described in the text and listed in Section A.1 of the Online Appendix. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination). The restricted sample only includes loans that do not default within 9 months of origination.

Table 6: Baseline Non-Parametric Results for the Sample of PLS Loans in McDash

Default Horizon:	36 Months			
	Full Sample		**Restricted Sample**	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale = 2	-0.018	-0.018	-0.011	-0.011
	(14.65)	(14.69)	(9.07)	(9.39)
Months to Sale = 3	-0.034	-0.034	-0.024	-0.023
	(15.63)	(16.03)	(11.17)	(11.42)
Months to Sale = 4	-0.051	-0.053	-0.043	-0.041
	(14.57)	(15.69)	(12.24)	(12.22)
Months to Sale = 5	-0.053	-0.054	-0.049	-0.047
	(11.79)	(12.48)	(11.32)	(11.28)
Months to Sale = 6	-0.046	-0.048	-0.052	-0.049
	(8.19)	(8.99)	(9.97)	(10.10)
Months to Sale = 7	-0.038	-0.040	-0.047	-0.045
	(6.95)	(7.94)	(9.90)	(10.17)
Months to Sale = 8	-0.022	-0.027	-0.042	-0.041
	(2.63)	(3.42)	(6.06)	(6.21)
Months to Sale = 9	-0.026	-0.031	-0.046	-0.045
	(2.86)	(2.93)	(5.18)	(5.27)
# Loans	5,747,722	5,747,722	5,574,463	5,574,463
Adjusted R^2	0.24	0.22	0.20	0.19
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 on PLS loans in the McDash data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that default over a 36-month horizon. Default is defined as a loan that is 60+ days delinquent (columns 1 and 3) and 90+ days delinquent (columns 2 and 4). Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, and the detailed list of covariates described in the text and listed in Section A.1 of the Online Appendix. The first row for each variable shows the regression coefficient estimates and the second row presents the t-statistics. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination). The restricted sample only includes loans that do not default within 9 months of origination.

Table 7: Baseline Results for the Sample of GSE Loans in McDash

Panel A: Linear Specification

Default Horizon:	36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale	0.0009	-0.0002	0.0009	-0.0003
	(2.78)	(0.80)	(2.50)	(1.01)
# Loans	14,045,839	14,045,839	14,045,839	14,045,839
Adjusted R^2	0.14	0.14	0.17	0.16
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

Panel B: Non-Parametric Specification

Default Horizon:	36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale = 2	-0.004	-0.004	-0.005	-0.005
	(8.40)	(7.83)	(9.33)	(9.11)
Months to Sale = 3	0.007	0.005	0.007	0.005
	(11.47)	(9.65)	(8.93)	(7.32)
Months to Sale = 4	0.023	0.007	0.024	0.008
	(19.53)	(7.38)	(17.39)	(6.73)
Months to Sale = 5	0.014	0.007	0.014	0.008
	(8.54)	(5.13)	(7.45)	(4.40)
Months to Sale = 6	0.006	0.001	0.009	0.003
	(2.50)	(0.61)	(3.44)	(1.45)
Months to Sale = 7	0.003	-0.001	0.005	0.002
	(0.93)	(0.40)	(1.56)	(0.53)
Months to Sale = 8	-0.013	-0.012	-0.005	-0.007
	(3.89)	(4.22)	(1.41)	(2.05)
Months to Sale = 9	-0.025	-0.026	-0.019	-0.021
	(3.54)	(3.82)	(2.90)	(3.29)
# Loans	14,045,839	14,045,839	14,045,839	14,045,839
Adjusted R^2	0.14	0.14	0.17	0.16
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 on GSE loans in the McDash data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that default over a 36-month horizon (columns 1-2) and over a 60-month horizon (columns 3-4). Default is defined as a loan that is 60+ days delinquent (columns 1 and 3) and 90+ days delinquent (columns 2 and 4). Months to sale is defined as the number of months that elapse between origination and sale to a GSE. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, and a detailed list of the covariates described in the text and listed in Section A.1 of the Online Appendix. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination).

Table 8: Ex-Ante Default Risk Results

Panel A: PLS Loans

Model:	Linear Probability				Logit			
Default Horizon:	36 Months		60 Months		36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale	0.0058	0.0045	0.0057	0.0040	0.0047	0.0015	0.0031	0.0010
	(8.15)	(8.67)	(8.74)	(8.72)	(2.80)	(0.77)	(6.08)	(2.15)
# Loans	3,672,426	3,672,426	3,672,426	3,672,426	3,660,474	3,660,474	3,613,121	3,613,121
Adjusted R^2	0.26	0.24	0.30	0.36	0.41	0.42	0.59	0.67
Orig Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: GSE Loans

Model:	Linear Probability				Logit			
Default Horizon:	36 Months		60 Months		36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale	0.0004	0.0002	0.0021	0.0013	0.0001	-0.0004	0.0007	0.0002
	(2.39)	(2.08)	(6.58)	(6.90)	(0.10)	(0.62)	(1.42)	(0.39)
# Loans	7,378,891	7,378,891	7,378,891	7,378,891	7,121,472	7,121,458	7,378,462	7,377,410
Adjusted R^2	0.29	0.30	0.52	0.56	0.26	0.25	0.51	0.46
Orig Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y

This table shows results for loan-level, OLS, and logit regressions, where the dependent variables are the 36-month and 60-month *ex-ante* default rates at the time the loan was originated, where the *ex-ante* default rates are calculated using data on loan and borrower characteristics at the time of origination for the previous three years for the 36-month *ex-ante* rates and five years for the 60-month *ex-ante* rates. Default is defined as a loan being 60 days and 90 days delinquent or more at any point since origination. The independent variable of interest is months to sale, which is defined as the number of months that elapse between origination and sale to a PLS issuer or GSE. All regressions include origination year-quarter fixed effects and year-quarter of sale fixed effects. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination). The first row for each variable shows the regression coefficient and the second row shows the t-statistic.

Table 9: Baseline Parametric Results for the Sample of CoreLogic PLS Loans

Panel A: Including Originator Fixed Effects

	Default Definition: 60+ DQ over 36 Months								
	All PLS			**Alt-A**			**Subprime**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Months to Sale	-0.0036	-0.0028	-0.0015	-0.0072	-0.0063	-0.0057	-0.0020	-0.0015	0.0005
	(8.57)	(6.41)	(3.66)	(10.55)	(10.54)	(11.46)	(5.15)	(3.65)	(1.11)
# Loans	7,860,499	7,858,236	7,855,810	1,895,618	1,894,861	1,893,617	5,964,881	5,963,091	5,961,433
Adjusted R^2	0.21	0.21	0.22	0.25	0.26	0.27	0.19	0.19	0.20
Orig YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Originator FE	N	Y	Y	N	Y	Y	N	Y	Y
Originator x Orig-YQ FE	N	N	Y	N	N	Y	N	N	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Including Issuer Fixed Effects

	Default Definition: 60+ DQ over 36 Months								
	All PLS			**Alt-A**			**Subprime**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Months to Sale	-0.0041	-0.0019	-0.0017	-0.0063	-0.0047	-0.0051	-0.0027	-0.0003	0.0005
	(8.29)	(4.15)	(3.79)	(11.83)	(9.28)	(9.94)	(6.16)	(0.70)	(1.09)
# Loans	7,725,370	7,725,369	7,722,948	1,849,905	1,849,904	1,848,674	5,875,179	5,875,179	5,873,522
Adjusted R^2	0.21	0.22	0.22	0.26	0.27	0.27	0.19	0.20	0.20
Orig YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Originator x Orig-YQ FE	N	N	Y	N	N	Y	N	N	Y
Issuer FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer x Issue -YQ FE	N	Y	Y	N	Y	Y	N	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 on PLS loans in the CoreLogic data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that default over a 36-month horizon. Default is defined as a loan that is 60+ days delinquent. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, and the detailed list of covariates described in the text and in Section A.1 of the Online Appendix. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination).

Table 10: Affiliation Results for the Sample of CoreLogic PLS Loans

Panel A: Parametric Results

Default Definition:	60+ DQ over 36 Months					
	All PLS		**Alt-A**		**Subprime**	
	Affiliation	No Affiliation	Affiliation	No Affiliation	Affiliation	No Affiliation
Months to Sale	-0.0046	-0.0049	-0.0049	-0.0100	-0.0029	-0.0031
	(6.93)	(8.02)	(7.12)	(10.84)	(4.96)	(5.21)
# Loans	3,176,715	3,473,338	603,234	735,374	2,573,481	2,737,861
Adjusted R^2	0.20	0.21	0.24	0.26	0.19	0.20
Orig Qtr FE	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y
Issuer FE	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y

Panel B: Non-Parametric Results

Default Definition:	60+ DQ over 36 Months					
	All PLS		**Alt-A**		**Subprime**	
	Affiliation	No Affiliation	Affiliation	No Affiliation	Affiliation	No Affiliation
Months to Sale = 1	-0.0039	-0.0313	-0.007	-0.0442	-0.0014	0.0021
	(2.12)	(8.66)	(3.40)	(13.42)	(1.11)	(0.68)
Months to Sale = 2	-0.0071	-0.0396	-0.0114	-0.0513	-0.0027	-0.0065
	(3.34)	(11.78)	(4.91)	(14.97)	(1.46)	(2.16)
Months to Sale = 3	-0.0153	-0.0468	-0.0206	-0.0615	-0.0082	-0.0136
	(5.77)	(12.26)	(4.34)	(15.34)	(3.64)	(4.15)
Months to Sale = 4	-0.0197	-0.0482	-0.0243	-0.0712	-0.0112	-0.0142
	(6.56)	(11.14)	(6.76)	(13.69)	(4.22)	(3.77)
Months to Sale = 5	-0.0271	-0.0523	-0.02	-0.075	-0.018	-0.0185
	(6.58)	(11.04)	(4.24)	(14.32)	(5.37)	(4.51)
Months to Sale = 6	-0.024	-0.054	-0.0221	-0.0817	-0.0147	-0.0192
	(5.12)	(9.74)	(4.11)	(11.91)	(3.37)	(3.83)
Months to Sale = 7	-0.0326	-0.0555	-0.0263	-0.0923	-0.0274	-0.0178
	(5.03)	(9.52)	(3.97)	(12.01)	(4.20)	(3.31)
Months to Sale = 8	-0.0316	-0.0507	-0.0491	-0.0971	-0.0212	-0.009
	(4.49)	(8.68)	(6.31)	(10.61)	(2.83)	(1.49)
Months to Sale = 9	-0.0152	-0.052	-0.0428	-0.1128	-0.0035	-0.0021
	(1.81)	(8.11)	(4.24)	(12.40)	(0.32)	(0.31)
# Loans	3,176,715	3,473,338	603,234	735,374	2,573,481	2,737,861
Adjusted R^2	0.20	0.21	0.24	0.26	0.19	0.20
Orig Qtr FE	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y
Issuer FE	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 on PLS loans in the CoreLogic data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that default over a 36-month horizon. Default is defined as a loan that is 60+ days delinquent. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, originator fixed effects, and the detailed list of covariates described in the text and in Section A.1 of the Online Appendix. "Affiliated" PLS deals are those in which the originator of all of mortgages in the deal is affiliated with the issuer (either the same company or part of the same vertical corporation). The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination.

Table 11: Documentation Results for the Sample of CoreLogic PLS Loans

Panel A: Linear Results

Default Definition:	60+ DQ over 36 Months					
	All PLS		**Alt-A**		**Subprime**	
	Full Doc	Low Doc	Full Doc	Low Doc	Full Doc	Low Doc
Months to Sale	-0.0033	-0.0051	-0.0061	-0.0051	-0.0025	-0.0034
	(7.78)	(7.35)	(9.56)	(8.71)	(5.94)	(4.81)
# Loans	4,275,516	3,408,451	493,756	1,344,859	3,781,606	2,063,379
Adjusted R^2	0.18	0.25	0.16	0.28	0.17	0.24
Orig Qtr FE	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y
Issuer FE	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y

Panel B: Non-Parametric Results

Default Definition:	60+ DQ over 36 Months					
	All PLS		**Alt-A**		**Subprime**	
	Full Doc	Low Doc	Full Doc	Low Doc	Full Doc	Low Doc
Months to Sale = 1	-0.0032	-0.0112	-0.0125	-0.0181	-0.0008	0.0025
	(2.55)	(5.35)	(6.80)	(10.59)	(0.64)	(1.35)
Months to Sale = 2	-0.0068	-0.0173	-0.0194	-0.0227	-0.0025	-0.0005
	(4.96)	(7.74)	(8.46)	(11.39)	(1.78)	(0.24)
Months to Sale = 3	-0.0145	-0.0244	-0.0291	-0.0272	-0.009	-0.0083
	(8.23)	(8.72)	(9.76)	(11.19)	(5.14)	(3.10)
Months to Sale = 4	-0.0173	-0.0273	-0.0343	-0.0319	-0.0119	-0.0093
	(8.40)	(7.78)	(9.81)	(10.42)	(5.80)	(2.75)
Months to Sale = 5	-0.0199	-0.0347	-0.0359	-0.0313	-0.0148	-0.0196
	(8.18	(9.15)	(8.43)	(9.74)	(6.14)	(4.94)
Months to Sale = 6	-0.0194	-0.0358	-0.0332	-0.0365	-0.0141	-0.0188
	(6.42)	(7.18)	(7.01)	(8.05)	(4.73)	(3.51)
Months to Sale = 7	-0.0208	-0.0339	-0.0428	-0.0381	-0.0151	-0.0194
	(5.29)	(5.61)	(7.04)	(6.53)	(3.79)	(3.11)
Months to Sale = 8	-0.0118	-0.0275	-0.0464	-0.0455	-0.0049	-0.0033
	(2.82)	(4.08)	(7.63)	(6.49)	(1.08)	(0.51)
Months to Sale = 9	-0.0033	-0.032	-0.0535	-0.0526	0.0097	-0.0004
	(0.64)	(5.08)	(7.24)	(9.48)	(1.70)	(0.06)
# Loans	4,275,516	3,408,451	493,756	1,344,859	3,781,606	2,063,379
Adjusted R^2	0.18	0.25	0.16	0.28	0.17	0.24
Orig Qtr FE	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y
Issuer FE	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y

This table displays the results from the estimation of Equation 5 on PLS loans in the CoreLogic data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that default over a 36-month horizon. Default is defined as a loan that is 60+ days delinquent. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, originator fixed effects, and the detailed list of covariates described in the text and in Section A.1 of the Online Appendix. "Full Doc" loans correspond to those in which the borrower's income and assets were not fully documented at the time of origination, while "Low Doc" loans are those in which either the borrower's income or assets (or both) were not fully documented. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination).

Table 12: Pricing Analysis Results

Panel A: Linear Specification, Triple A securities

Dependent Variable: Pool-level Average Yield Spread (Triple-A Securities Only)

	All Securities			Alt-A Securities			Subprime Securities		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Seasoning	-0.015***	-0.004	-0.011***	-0.023*	-0.024	-0.025**	0.003*	0.002	0.002
	(0.004)	(0.001)	(0.004)	(0.014)	(0.017)	(0.012)	(0.002)	(0.002)	(0.002)
Pool Covariates	N	Y	Y	N	Y	Y	N	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer FE	N	N	Y	N	N	Y	N	N	Y
Observations	3,532	3,532	3,513	909	909	909	2,623	2,615	2,615
Adjusted R^2	0.17	0.30	0.44	0.08	0.15	0.34	0.67	0.68	0.71

Panel B: Panel A: Non-linear Specification, Triple A securities

Dependent Variable: Pool-level Average Yield Spread (Triple-A Securities Only)

	All Securities			Alt-A Securities			Subprime Securities		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Seasoning	-0.103***	-0.056***	-0.037***	-0.199***	-0.051	-0.185***	-0.003	-0.006	-0.010
	(0.022)	(0.021)	(0.014)	(0.055)	(0.037)	(0.061)	(0.006)	(0.007)	(0.008)
Seasoning^2	0.011***	0.007**	0.003**	0.023***	0.003	0.021**	0.001	0.001	0.001
	(0.003)	(0.003)	(0.002)	(0.008)	(0.005)	(0.009)	(0.001)	(0.001)	(0.001)
Pool Covariates	N	Y	Y	N	Y	Y	N	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer FE	N	N	Y	N	N	Y	N	N	Y
Observations	3,532	3,532	3,513	909	909	909	2,623	2,615	2,615
Adjusted R^2	0.19	0.31	0.44	0.12	0.34	0.17	0.67	0.69	0.71

Panel C: Linear specification, All securities

Dependent Variable: Pool-level Average Yield Spread

	All Securities			Alt-A Securities			Subprime Securities		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Seasoning	-0.001	-0.003	-0.002	-0.041***	-0.034**	-0.034**	0.006**	0.004	0.004
	(0.004)	(0.004)	(0.005)	(0.014)	(0.015)	(0.015)	(0.003)	(0.004)	(0.004)
Pool Covariates	N	Y	Y	N	Y	Y	N	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer FE	N	N	Y	N	N	Y	N	N	Y
Observations	2,254	2,254	2,247	453	453	453	1,799	1,795	1,795
Adjusted R^2	0.35	0.44	0.56	0.10	0.18	0.18	0.53	0.67	0.67

This table displays results from the estimation of Equation 7. The sample includes triple-A, floating rate subprime, and Alt-A securities issued between January 2002 and December 2007. The dependent variable is the weighted average spread over the 1-month LIBOR of all triple-A securities with claims on cash flows for a given mortgage pool. Seasoning is the average age (# months) of all mortgages in a pool at the time of issuance. All regressions include month-of-issue fixed effects. The set of pool-level covariates corresponds to the variables included in Table A.8, which are all pool-level averages. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered at the deal-level. *** p <0.01, ** p <0.05, * p<0.1.

Table 13: Quantitative magnitudes of signaling costs.

Panel A: Parameter Values

Parameter	Value	Source
B_0	$\$291,003$	Table 4
r_m	6.26%	Table 4
α	90%	An and Cordell (2017)
		and Adelino et al. (2013)
η	9.73	Average time to sale for GSE Loans Table 1
$\bar{\lambda}$	0.0621	Annualized expected default rates for sample of loans in Table (8)
λ_ℓ	0.0487	Simulation
λ_h	0.0642	Simulation
α	1.75	Simulation
β	.275	Simulation
r	5.58%	Simulation
γ	7.92%	Simulation

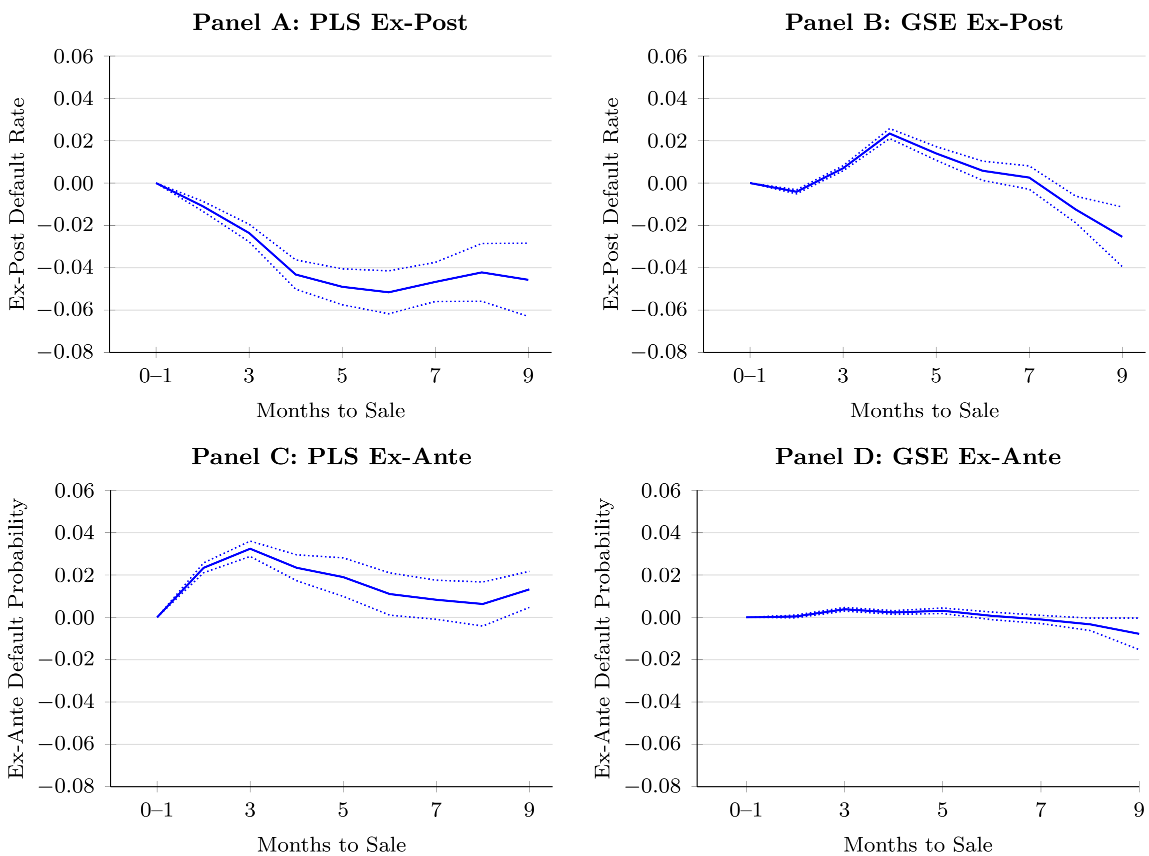
Panel B: Simulated vs Empirical Data

	Simulated	Actual
Coefficient of Default on Time to Sale	-0.0072	-0.0072
Mortgage Value to Originator	$\$291,002$	$\$291,003$
Coefficient of Yield Spread on Time to Sale	-0.0070	-0.034

Panel C: Loss in Surplus

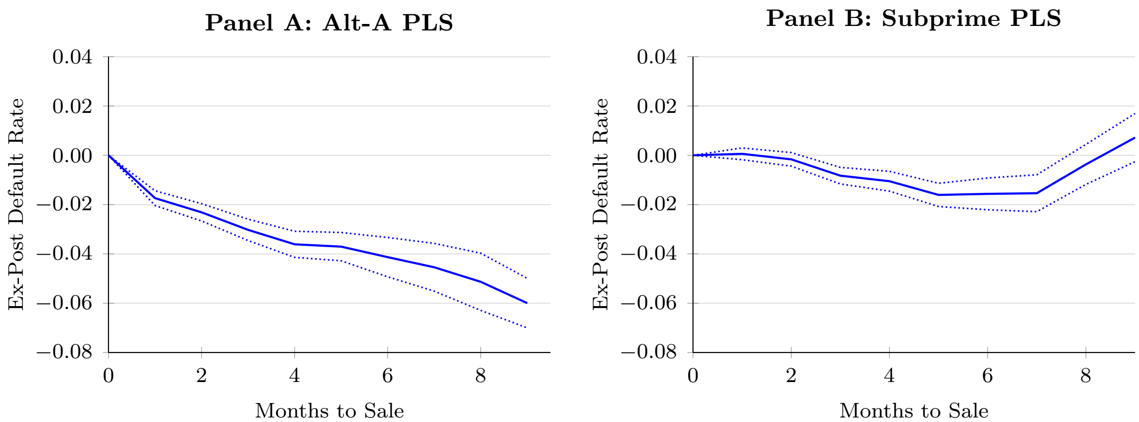
	Average Proceeds with Signaling	Average Proceeds without Signaling	Percentage Loss
Main Parameters	$\$291,002$	$\$291,538$	18.4 bps
$\gamma = .0842$	$\$290,854$	$\$291,391$	18.5 bps
$\gamma = .0742$	$\$291,152$	$\$291,686$	18.3 bps

Figure 1: Ex-Ante vs. Ex-Post McDash Results



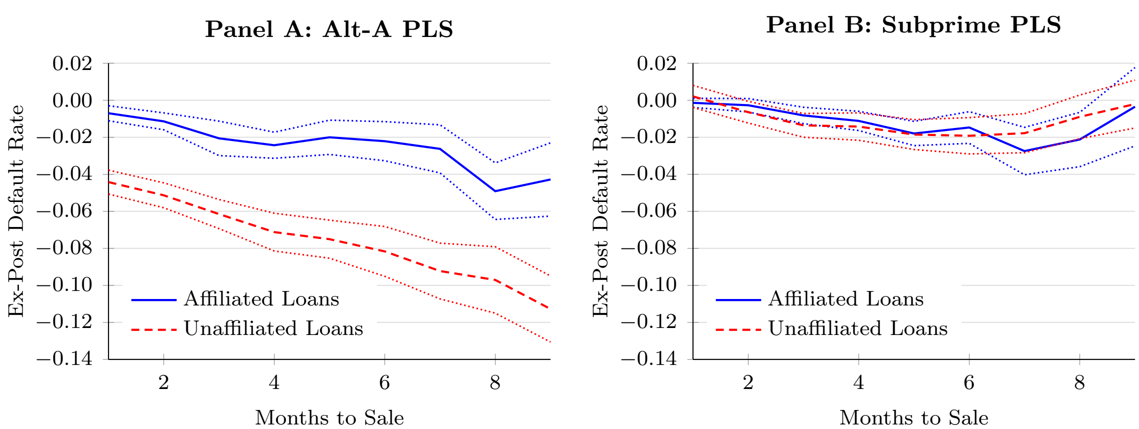
This figure displays results from the estimation of the non-parametric version of Equation 5 for both the PLS and GSE loans in the McDash data set originated in the 2002–2007 period. Panels A and B correspond to the *ex-post* default rates, while panels C and D correspond to *ex-ante* predicted default rates. Default is defined as a loan that becomes 60 days delinquent over a 36-month horizon measured from origination. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. The dotted lines correspond to 95% confidence intervals.

Figure 2: CoreLogic PLS Results



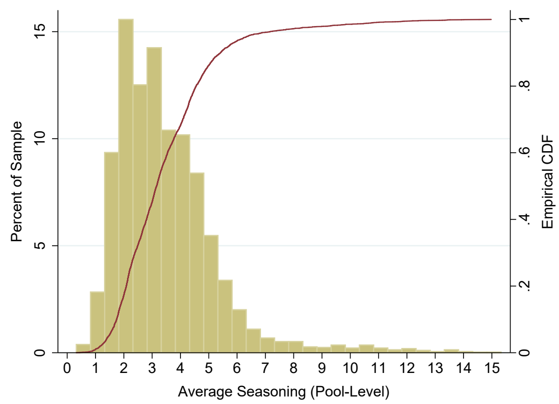
This figure displays results from the estimation of the non-parametric version of Equation 5 for PLS loans in CoreLogic. Panel A corresponds to Alt-A PLS loans and panel B corresponds to subprime loans. Default is defined as a loan that becomes 60 days delinquent over a 36-month horizon measured from origination. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. The dotted lines show 95% confidence intervals.

Figure 3: CoreLogic Affiliation Results



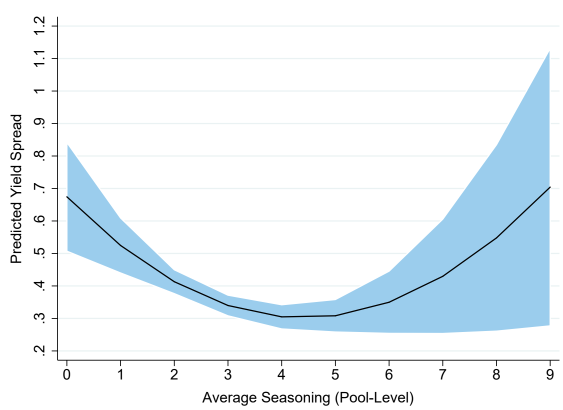
This figure displays results from the estimation of the non-parametric version of Equation 5 for PLS loans in the CoreLogic data set. Panel A corresponds to Alt-A PLS loans and panel B corresponds to subprime loans. Default is defined as a loan that becomes 60 days delinquent over a 36-month horizon measured from origination. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. The dotted lines show 90% confidence intervals.

Figure 4: Distribution of Pool-Level Seasoning



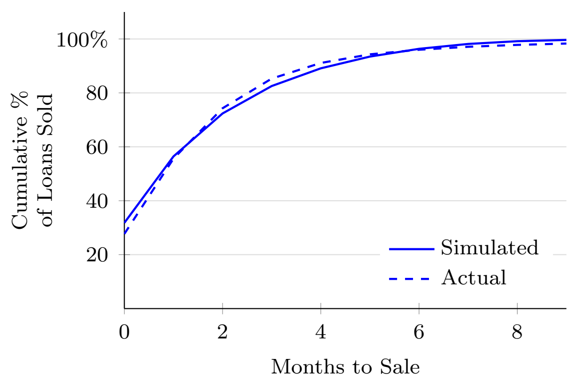
Notes: This figure displays the density and cumulative distribution of the average months of seasoning in the sample of floating-rate, triple-A, Subprime and Alt-A securities issued between January 2002 and December 2007 used in the pricing analysis in section 4.6.

Figure 5: Predicted Yield Spread as a Function of Seasoning



Notes: This figure displays the predicted security spreads (over the 1-month LIBOR) as a function of the average pool-level seasoning calculated using the estimation results from the specification reported in column (6) in panel B of Table 12. The shaded area corresponds to 95% confidence intervals calculated using the delta method.

Figure 6: Simulated vs Actual Distribution of Months to Sale



Notes: This figure displays the distribution of time to sale in the simulated data vs the same distribution in the actual data for the CoreLogic sample taken from Table 2.

Online Appendix (Not for Publication)

This appendix supplements the empirical analysis and provides an additional proof for the model used in the calibration exercise in "Are Lemons Sold First? Dynamic Signaling in the Mortgage Market" by Adelino, Gerardi, and Hartman-Glaser. Below is a list of the sections contained in this appendix.

Contents

A.1 Variable Definitions

ARM: An indicator variable that takes the value of 1 if the mortgage has an adjustable rate and 0 if it has a fixed rate.

Balance: The natural logarithm of the principal balance of the loan at origination.

Balloon: An indicator variable that takes the value of 1 if the mortgage is characterized by a balloon payment at the end of its term and 0 if it is a fully amortizing mortgage.

Condo: An indicator variable that takes the value of 1 if the property is a condominium or a townhouse and 0 otherwise.

FICO: The credit score of the borrower at origination. All models include both the continuous FICO variable and a set of indicator variables corresponding to 5 FICO intervals: FICO < 580, $580 \leq$FICO < 620, $620 \leq$FICO < 660, $660 \leq$FICO < 700, and FICO ≥ 700.

House Prices: County-level house price indices from CoreLogic. We include both price level in the county in the month of origination and the cumulative growth in prices from the month of mortgage origination calculated over the default horizon.

Interest-Only: An indicator variable that takes the value of 1 if the loan requires interest-only payments for a specified period of time and 0 otherwise.

Jumbo: An indicator variable that takes the value of 1 if the loan amount at origination exceeds the conforming loan limit set by statute that limits the size of mortgages eligible to be insured by the GSEs (during the vast majority of our sample period, the limit was $417,000 for mortgages on single-family properties) and 0 otherwise.

Loan-to-Value (cumulative): The loan-to-value ratio at origination computed using information on the first and second liens. All models include both the continuous LTV variable and a set of indicator variables corresponding to 5 LTV intervals: LTV < 70, $70 \leq$ LTV < 80, $80 <$ LTV < 90, $90 \leq$LTV< 100, and LTV ≥ 100. An indicator variable for the LTV ratios exactly equal to 80 is also included as a proxy for unreported second liens.

Low Documentation: An indicator variable that takes the value of 1 if the borrower's income and assets are not fully documented in the underwriting process and 0 if they are fully documented.

Month to Sale: The number of months after the date of origination in which a loan is sold to a PLS issuer or acquired by one of the GSEs. In the McDash data set, the variable is based on a field that is updated monthly and shows the current holder of the loan. In the CoreLogic LoanPerformance database, the variable is based on the length of time between the month of origination and the month in which the corresponding PLS security is issued.

Multi-family: An indicator variable that takes the value of 1 if the property is a 2–4-family house and 0 otherwise.

Negative Amortization: An indicator variable that takes the value of 1 if the loan requires payments of less than interest and principal for a specified period of time and 0 otherwise.

Prepayment Penalty: An indicator variable that takes the value of 1 if the mortgage contains a prepayment penalty and 0 otherwise.

Primary Residence: An indicator variable that takes the value of 1 if the property is the primary residence of the borrower and a value of 0 if the property is either an investment or a second home.
Purchase Loan: An indicator variable that takes the value of 1 if the loan is used to purchase property and 0 otherwise.
Refinance (traditional): An indicator variable that takes the value of 1 if the loan is used to refinance previous mortgage debt without converting any equity into cash and 0 otherwise.
Refinance (cashout): An indicator variable that takes the value of 1 if the loan is used to refinance previous mortgage debt with a portion of the equity converted to cash and 0 otherwise.
Single Family: An indicator variable that takes the value of 1 if the property is a detached single-family home and 0 otherwise.
Term: The maturity length of the mortgage in months.
Unemployment: County-level unemployment rates from the Bureau of Labor Services (BLS). We include both the rates in the county in the month of origination and the cumulative growth in the unemployment rate from the month of mortgage origination calculated over the default horizon.

A.2 Further Detail on Data Sets

A.2.1 Lender Processing Services (McDash)

Our primary data set is sourced from Lender Processing Services (McDash). We adopt standard sample restrictions in our analysis of the McDash data. We consider only first lien mortgages originated in the 2002–2007 period that were sold to PLS issuers or to the GSEs, so we eliminate loans kept in the portfolios of mortgage originators and never sold. In addition, a small number of loans in the data set were sold to the Federal Home Loan Banks (FHLBs), which we also eliminate from the sample. We only retain loans originated in the 50 United States and restrict the sample to loans that enter the data set in either the same month of origination or in the month following origination. We also address outliers in the data by winsorizing the distributions of credit scores, original loan balances, LTV ratios at origination, and interest rates at origination at the 1st and 99th percentiles of each distribution. We also explored trimming these variables instead of winsorizing and found that this change had little effect on the results.

A.2.2 Corelogic

The second mortgage data set used is sourced from Corelogic. In addition to the differences noted in the main text, the timing for when a loan enters each data set is also different across the McDash and CL data sets. In McDash, we observe most loans from the month of origination, and we can directly observe the month in which they are sold out of banks' portfolios to PLS issuers or the GSEs. In CL, however, we compute time to sale as the difference between the date of issuance of the mortgage-backed security in which the loan is included and the reported month of origination of the mortgage. Loans enter the CL data set on the issue date, so we do not observe the performance history of loans before they are securitized.

A.2.2.1 Lender Identity in Corelogic

There is some uncertainty about whether the originator field in the CoreLogic database actually corresponds to the lender of record (i.e., the institution that underwrote and originated the loan) or to what is sometimes referred to as the "aggregator" or "seller," which is the institution responsible for purchasing loans from various lenders to fill the PLS mortgage pools and then selling those loans to the issuer (Stanton et al. (2014)). This is a potentially important distinction because it could be more likely that private information is obtained by the lender of record because it has more interaction with the mortgage borrower.

To verify that the originator field in CoreLogic corresponds to the lender of record, we match our CoreLogic mortgage data to a database of public mortgage filings that contains the identity of the lender of record. This database covers the universe of all residential mortgages in the state of Massachusetts during our sample period and comes from county deed registries that record information on property transactions. We compare the lender of record with the originator listed in the CoreLogic database for the sample of matched Massachusetts mortgages. We find that for 83% of the matched sample, the lender of record matched the CoreLogic originator field. The remaining 17% are either cases in which CoreLogic is reporting an entity other than the lender of record (most likely the aggregator) or cases that are bad matches (there is the potential for significant matching errors because we are not able to perform a precise match using loan account numbers or social security numbers). Thus, we view the 17% figure as an upper bound on the severity of the potential for misidentifying the true originator in the CoreLogic data.

A.2.3 Bloomberg

Pricing data is sourced from Bloomberg. The data fields consist of security identifiers (including CUSIP and ticker), issuer name, issuance date, the identification of the loan pool that the security has claims on, the spread over one-month LIBOR at origination, and the weighted average life as advertised in the prospectus.

A.3 Robustness of the Month to Sale Threshold Value

In this section we conduct a sensitivity analysis of our choice of nine months for the maximum threshold between origination and sale. Imposing a threshold is important for our analysis for three reasons. First, we want to ensure that the loans in our estimation sample were originated with the intention of being sold. The distributions of months to sale in Tables 1 and 2 clearly show that the vast majority of loans are sold relatively quickly. More than 94% of all loans (including loans sold to PLS issuers and the GSEs) are sold within 6 months of origination. We are concerned that loans sold more than a year (or several years) after origination may be different in ways that are unobservable to us. Second, there was a fairly robust market for loans that were in delinquency early in their lives, but that were then rehabilitated at some point and sold. These loans were sold on a separate secondary market, which the industry referred to as "scratch and dent." We are unable to explicitly identify "scratch and dent" (S&D) loans in our McDash and CoreLogic data sets, but believe that many of the loans in the right-tail of the months to sale distribution may have been sold on the S&D market. MBS issuers and investors would have known that these were previously problematic loans, and including these loans in our sample would contaminate our empirical tests. A final reason for imposing a sale threshold is that it is necessary for implementing our sample selection correction discussed in Section 2.1.

While a months to sale threshold is necessary in our context, the choice of nine months is somewhat arbitrary. Thus, we display results below for our baseline specifications using longer thresholds. We consider 12-month and 24-month thresholds in both the CoreLogic and McDash data sets, and a 36-month threshold in our McDash data set. In CoreLogic, we do not observe loans seasoned more than 26 months, so the 24-month threshold includes essentially all loans. In contrast, we do see loans in our McDash data set that are sold more than two years after origination and that we suspect are S&D loans. Thus, when we consider longer horizons in the McDash data set we attempt to identify and eliminate them in a manner described in further detail below. CoreLogic, in contrast, contains loans sold exclusively in the subprime and Alt-A segments, and since S&D loans were typically pooled together for explicit S&D securities (not included in CoreLogic), we do not believe that S&D is an important issue for our analysis using this data set.[1]

Tables A.1 and A.2 below show the baseline results for the linear specification of months to sale in the CoreLogic data set using a months to sale threshold of 12 months and 24 months, respectively. The tables are exact counterparts to Table 9 in the main text. The top panels show the effect of including lender and lender-by-year-quarter fixed effects while the bottom panels show the effect of including issuer and issuer-by-year-quarter fixed effects for the sample of all PLS loans and the sample of Alt-A and subprime loans. The results are very similar to those displayed in Table 9, which assume a 9-month threshold. Figure A.2 shows results for the non-parametric months to sale specification for the Alt-A and subprime PLS samples assuming a 24-month threshold. This is the exact counterpart to Figure 2 in the main text. We group all loans sold between 12 and 24 months after origination into the same category due to the small number of loans sold after 12 months. The patterns for both the Alt-A and subprime samples are very similar to those in Figure 2.

Table A.3 displays results for the linear specification of months to sale in the McDash sample of private-label loans for the longer thresholds. These results are directly comparable to those reported in Panel A of Table 5 in the main text. The top panel of Table A.3 displays results for a 12-month threshold for both the 36-month and 60-month default horizons and for our 60+ DQ and 90+ DQ default definitions. The results are consistent with those in Table 5. The bottom panel of the table shows results for 24-month and 36-month thresholds, where we focus on the 60+ DQ default definition and the 60 month default horizon. We focus on the 60-month horizon to ensure

[1] The fact that we do not see as many highly seasoned loans in CoreLogic supports this position.

that we have a reasonable length of time to measure default for loans that are sold between 24 and 36 months after origination. As we discussed briefly above, we are especially concerned about the presence of S&D loans in the McDash sample. Since many of these loans default before being sold, it is probably the case that they would be more likely to default after being sold compared to a loan that did not default before sale. Inclusion of S&D loans would thus contaminate our test and likely bias us against finding better ex-post performance for loans sold later. Since we track loans in McDash from origination we can identify loans that default before they are sold. Thus, in an attempt to purge the sample of S&D loans, we eliminate any mortgage that defaults before sale for the analysis using longer sale thresholds (24 and 36 months).[2] In the bottom panel of Table A.3 we show results for the longer thresholds with and without this S&D correction. It is clear that the correction does have a significant effect on the results and confirms our intuition that the presence of S&D loans among the sample of loans sold long after origination produces an upward bias in our estimate of the relationship between time-to-sale and default.

[2]The vast majority of loans sold within a year of origination in our McDash data set do not default before they are sold. However, the fraction of loans that default before sale increases fairly dramatically for loans sold after a year. For example, about one-quarter of loans sold between 24 and 36 months defaulted before sale in our McDash sample.

Table A.1: Baseline Parametric Results for the Sample of CoreLogic PLS Loans: 12-month Sale Threshold

Panel A: Including Originator Fixed Effects

	Default Definition: 60+ DQ over 36 Months								
		All PLS			**Alt-A**			**Subprime**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Months to Sale	-0.0031	-0.0024	-0.0013	-0.0070	-0.0061	-0.0060	-0.0014	-0.0010	0.0010
	(3.40)	(2.76)	(1.09)	(6.62)	(4.73)	(5.18)	(1.60)	(1.19)	(1.13)
# Loans	7,911,165	7,908,898	7,906,364	1,916,628	1,915,876	1,914,573	5,994,537	5,992,733	5,990,980
Adjusted R^2	0.21	0.21	0.22	0.25	0.26	0.27	0.19	0.19	0.20
Orig YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lender FE	N	Y	Y	N	Y	Y	N	Y	Y
Lender x Orig-YQ FE	N	N	Y	N	N	Y	N	N	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Including Issuer Fixed Effects

	Default Definition: 60+ DQ over 36 Months								
		All PLS			**Alt-A**			**Subprime**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Months to Sale	-0.0033	-0.0025	-0.0014	-0.0063	-0.0046	-0.0054	-0.0022	-0.0004	0.0012
	(3.34)	(2.97)	(1.35)	(5.30)	(4.59)	(4.44)	(2.53)	(0.52)	(1.37)
# Loans	7,770,236	7,770,236	7,767,706	1,870,072	1,870,071	1,868,782	5,899,872	5,899,872	5,898,121
Adjusted R^2	0.21	0.22	0.22	0.26	0.27	0.27	0.19	0.19	0.20
Orig YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lender FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lender x Orig-YQ FE	N	N	Y	N	N	Y	N	N	Y
Issuer FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer x Issue -YQ FE	N	Y	Y	N	Y	Y	N	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 (in the main text) on PLS loans in the CoreLogic data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that default over a 36-month horizon. Default is defined as a loan that is 60+ days delinquent. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, and the detailed list of covariates described in the text. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination).

Table A.2: Baseline Parametric Results for the Sample of CoreLogic PLS Loans: 24-month Sale Threshold

Panel A: Including Originator Fixed Effects

	Default Definition: 60+ DQ over 36 Months								
		All PLS			**Alt-A**			**Subprime**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Months to Sale	-0.0027	-0.0016	-0.0007	-0.0064	-0.0052	-0.0053	-0.0014	-0.0004	0.0014
	-2.77	-1.65	-0.59	-5.55	-3.71	-3.81	-1.25	-0.34	1.13
# Loans	7,989,106	7,986,805	7,984,021	1,938,373	1,937,617	1,936,208	6,050,732	6,048,891	6,046,946
Adjusted R^2	0.22	0.22	0.23	0.28	0.29	0.29	0.20	0.21	0.21
Orig YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lender FE	N	Y	Y	N	Y	Y	N	Y	Y
Lender x Orig-YQ FE	N	N	Y	N	N	Y	N	N	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

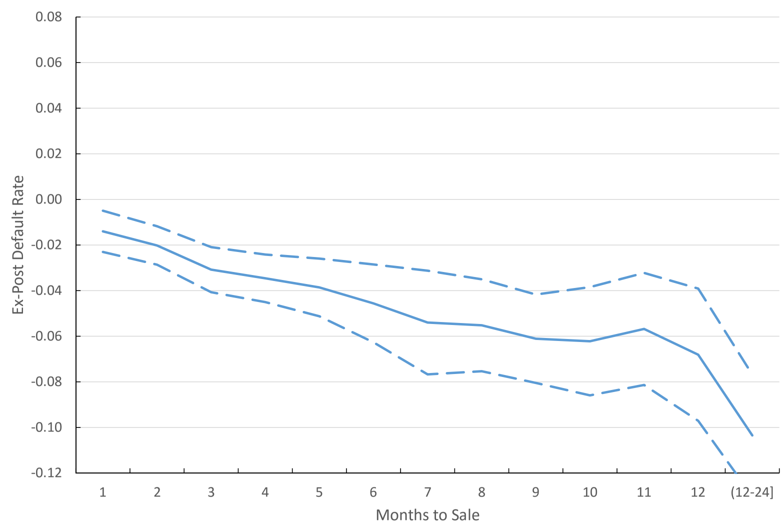
Panel B: Including Issuer Fixed Effects

	Default Definition: 60+ DQ over 36 Months								
		All PLS			**Alt-A**			**Subprime**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Months to Sale	-0.0018	-0.0009	0.0001	-0.0047	-0.0039	-0.0045	-0.0008	0.0013	0.0028
	-1.66	-0.98	0.09	-3.51	-2.94	-3.88	-0.64	1.46	2.96
# Loans	7,834,853	7,834,853	7,832,072	1,889,307	1,889,306	1,887,908	5,945,246	5,945,246	5,943,302
Adjusted R^2	0.22	0.23	0.23	0.29	0.29	0.29	0.21	0.21	0.21
Orig YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lender FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lender x Orig-YQ FE	N	N	Y	N	N	Y	N	N	Y
Issuer FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer x Issue -YQ FE	N	Y	Y	N	Y	Y	N	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

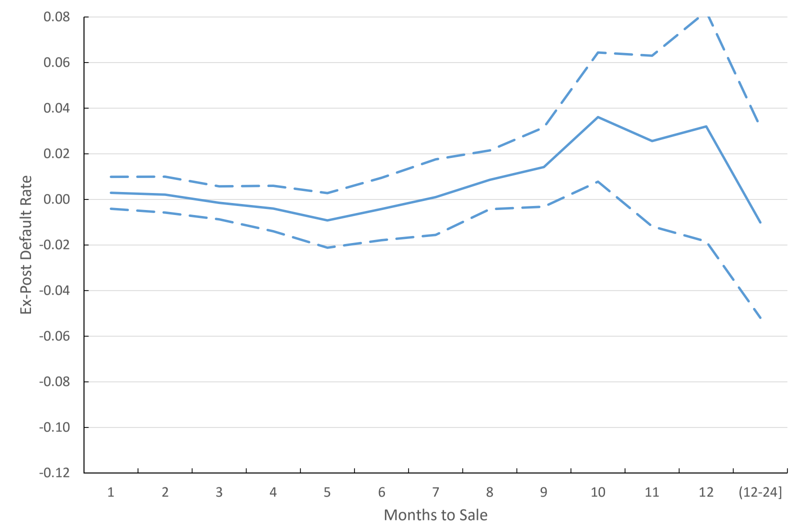
This table displays results from the estimation of Equation 5 (in the main text) on PLS loans in the CoreLogic data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that default over a 36-month horizon. Default is defined as a loan that is 60+ days delinquent. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, and the detailed list of covariates described in the text. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination).

Figure A.1: CoreLogic PLS Results: 24-Month Sale Threshold

Panel B: Alt-A PLS



Panel C: Subprime PLS



Notes: This figure displays results from the estimation of the non-parametric version of Equation 5 for PLS loans in CoreLogic. Panel A corresponds to Alt-A PLS loans and panel B corresponds to subprime loans. Default is defined as a loan that becomes 60 days delinquent over a 36-month horizon measured from origination. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. The dotted lines show 90% confidence intervals.

Table A.3: McDash Private-Label Results: Higher Sale Thresholds

Panel A: 12-Month Sale Threshold

Default Horizon:	36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale	-0.0067	-0.0072	-0.0068	-0.0078
	(8.40)	(9.43)	(9.46)	(11.23)
# Loans	5,811,639	5,811,639	5,811,639	5,811,639
Adjusted R^2	0.24	0.22	0.26	0.25
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

Panel B: Higher Sale Thresholds

Default Horizon:	60 Months			
Default Definition:	60+ DQ			
Sale Threshold:	24 Months		36 Months	
Scratch & Dent Correction	No	Yes	No	Yes
Months to Sale	-0.0028	-0.0076	-0.0021	-0.0059
	(4.31)	(12.39)	(4.09)	(11.05)
# Loans	5,879,757	5,860,471	5,910,105	5,885,068
Adjusted R^2	0.26	0.26	0.26	0.26
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

A.4 Robustness to the Inclusion of Debt-to-Income Ratios in the Covariate Set

In this section we conduct a sensitivity analysis to the inclusion of the debt-to-income (DTI) ratio as a control variable in our empirical analysis. We chose to leave the DTI ratio out of the covariate set in the baseline specifications reported in the main text for two reasons. First, the variable has poor coverage as it is missing for approximately 50 percent of the loans in our McDash sample and approximately 35 percent in our CoreLogic sample of Alt-A mortgages.[3] Thus, including it in our set of control variables results in a loss of a significant fraction of our sample of loans. Second, the variable includes information on the "front-end" DTI ratio (only mortgage debt divided by income at origination) for some loans, but for other loans it includes information on the "back-end" DTI ratio (all outstanding debt, including mortgage, credit card, auto, education debts, divided by income at origination). Unfortunately, McDash does not provide us with the necessary information to distinguish between the two different types of DTI ratios, which results in significant measurement error.[4]

In Table A.4 below we display estimation results of our primary McDash specifications in which we include the DTI ratio in the covariate set. Despite losing approximately half of the sample, the results are broadly similar to those reported in Tables 5 and 6 in the main text.

In Table A.5 below we display estimation results of our primary CoreLogic specifications in which we include the DTI ratio in the covariate set. The results are broadly similar to those reported in Table 9 in the main text.

[3] It is better populated in our sample of CoreLogic subprime loans (about 90 percent).

[4] We are slightly more confident that the variable reflects the "front-end" DTI ratio in the CoreLogic data set.

Table A.4: The Effect of Including Debt-to-Income Ratios: PLS Loans in McDash

Panel A: Parametric Results

Default Horizon:	36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale	-0.0076	-0.0078	-0.0117	-0.0122
	(6.82)	(7.21)	(7.35)	(7.84)
# Loans	2,968,692	2,968,692	2,968,692	2,968,692
Adjusted R^2	0.24	0.23	0.25	0.25
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

Panel B: Non-Parametric Results

Default Horizon:	36 Months		60 Months	
Default Definition:	60+ DQ	90+ DQ	60+ DQ	90+ DQ
Months to Sale = 2	-0.0079	-0.0068	-0.0148	-0.0133
	(3.26)	(2.88)	(4.24)	(3.87)
Months to Sale = 3	-0.0222	-0.0215	-0.0339	-0.0326
	(6.56)	(6.73)	(6.15)	(6.23)
Months to Sale = 4	-0.0404	-0.0406	-0.0595	-0.0611
	(10.51)	(11.87)	(9.32)	(10.36)
Months to Sale = 5	-0.0383	-0.0377	-0.0578	-0.0589
	(7.27)	(7.49)	(8.22)	(8.52)
Months to Sale = 6	-0.0379	-0.0385	-0.0629	-0.0675
	(5.77)	(6.28)	(7.50)	(8.27)
Months to Sale = 7	-0.0207	-0.0236	-0.0461	-0.0509
	(2.92)	(3.38)	(4.62)	(5.13)
Months to Sale = 8	-0.0255	-0.0295	-0.0411	-0.0469
	(2.31)	(2.89)	(3.38)	(4.00)
Months to Sale = 9	-0.041	-0.0461	-0.0429	-0.0521
	(2.32)	(2.72)	(2.99)	(3.69)
# Loans	2,968,692	2,968,692	2,968,692	2,968,692
Adjusted R^2	0.24	0.23	0.25	0.25
Orig Qtr FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 on PLS loans in the McDash data set originated in the 2002–2007 period. The specifications are identical to those in Table 5 with the addition of the debt-to-income ratio in the covariate set. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination).

Table A.5: The Effect of Including Debt-to-Income Ratios: PLS Loans in CoreLogic

Panel A: Including Originator Fixed Effects

	Default Definition: 60+ DQ over 36 Months								
		All PLS			**Alt-A**			**Subprime**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Months to Sale	-0.0034	-0.0027	-0.0012	-0.0066	-0.007	-0.0053	-0.0023	-0.0013	0.0005
	(3.33)	(3.04)	(1.22)	(5.12)	(4.85)	(5.00)	(2.52)	(1.55)	(0.51)
# Loans	6,470,565	6,469,485	6,467,766	1,196,111	1,195,640	1,194,611	5,274,454	5,273,683	5,272,583
Adjusted R^2	0.2	0.21	0.21	0.25	0.25	0.26	0.19	0.19	0.2
Orig YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Originator FE	N	Y	Y	N	Y	Y	N	Y	Y
Originator x Orig-YQ FE	N	N	Y	N	N	Y	N	N	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Including Issuer Fixed Effects

	Default Definition: 60+ DQ over 36 Months								
		All PLS			**Alt-A**			**Subprime**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Months to Sale	-0.0044	-0.0020	-0.0015	-0.0072	-0.0052	-0.0047	-0.0028	-0.0003	0.0006
	(8.29)	(4.15)	(3.79)	(4.03)	(4.22)	(3.83)	(2.91)	(0.22)	(0.55)
# Loans	6,354,517	6,354,517	6,352,805	1,158,779	1,158,779	1,157,760	5,195,576	5,195,575	5,194,476
Adjusted R^2	0.21	0.21	0.21	0.25	0.26	0.26	0.19	0.20	0.20
Orig YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Originator x Orig-YQ FE	N	N	Y	N	N	Y	N	N	Y
Issuer FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer x Issue -YQ FE	N	Y	Y	N	Y	Y	N	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 on PLS loans in the CoreLogic data set originated in the 2002–2007 period. The specifications are identical to those in Table 9 with the addition of the debt-to-income ratio in the covariate set. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by state-quarter (of origination).

A.5 Pricing Analysis – Detail on Including Lower Rated Securities

Table A.6 shows the effect of including lower-rated securities in our pricing analysis. The table is structured identically to Table 13 in the main text, with the top panel containing estimation results when we include only a linear term for average seasoning and the bottom panel containing results from a quadratic specification. The results are broadly consistent with those in Table 13. The negative relationship between pool-level seasoning and yield spreads is stronger for the sample of Alt-A pools (columns (4)–(6)), but is weaker for the full sample of PLS pools. Figures A.2 and A.3 display the predicted security spreads as a function of the average pool-level seasoning calculated using the estimation results in column (6) in Panels A and B of Table A.6, respectively. Figure A.3 is extremely similar to Figure 5 in the main text. Average yield spreads associated with pools with average seasoning of 4–5 months are about 40 basis points lower than those associated with pools with very little seasoning.

Table A.6: Pricing Analysis Results

Panel A: Linear Specification

Dependent Variable: Pool-level Average Yield Spread

	All Securities			Alt-A Securities			Subprime Securities		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Seasoning	-0.001	-0.003	-0.002	-0.041***	-0.034**	-0.034**	0.006**	0.004	0.004
	(0.004)	(0.004)	(0.005)	(0.014)	(0.015)	(0.015)	(0.003)	(0.004)	(0.004)
Pool Covariates	N	Y	Y	N	Y	Y	N	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer FE	N	N	Y	N	N	Y	N	N	Y
Observations	2,254	2,254	2,247	453	453	453	1,799	1,795	1,795
Adjusted R^2	0.35	0.44	0.56	0.10	0.18	0.18	0.53	0.67	0.67

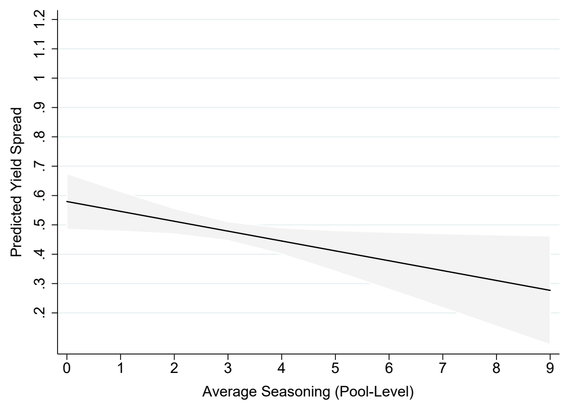
Panel B: Non-Linear Specification

Dependent Variable: Pool-level Average Yield Spread

	All Securities			Alt-A Securities			Subprime Securities		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Seasoning	-0.037**	-0.031**	-0.014	-0.185***	-0.177***	-0.177***	0.003	-0.012	-0.005
	(0.016)	(0.015)	(0.017)	(0.043)	(0.053)	(0.053)	(0.011)	(0.011)	(0.014)
Seasoning^2	0.005**	0.004*	0.002	0.019***	0.019***	0.019***	0.001	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.005)	(0.007)	(0.007)	(0.001)	(0.001)	(0.002)
Pool Covariates	N	Y	Y	N	Y	Y	N	Y	Y
Issue Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer FE	N	N	Y	N	N	Y	N	N	Y
Observations	2,254	2,254	2,247	453	453	453	1,799	1,799	1,795
Adjusted R-squared	0.35	0.44	0.56	0.13	0.20	0.20	0.53	0.61	0.67

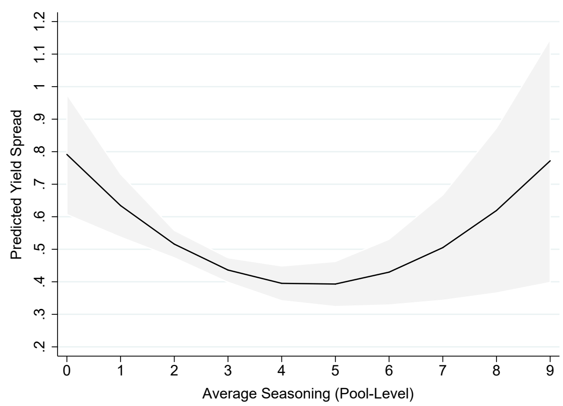
This table displays results from the estimation of Equation (7) in the main text. The sample includes triple-A and lower-rated, floating rate subprime, and Alt-A securities issued between January 2002 and December 2007. The dependent variable is the weighted average spread over the 1-month LIBOR of all securities with claims on cash flows for a given mortgage pool. Seasoning is the average age (# months) of all mortgages in a pool at the time of issuance. All regressions include month-of-issue fixed effects. The set of pool-level covariates corresponds to the variables included in Table A.5, which are all pool-level averages. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered at the deal-level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A.2: Predicted Yield Spread as Function of Seasoning



Notes: This figure displays the predicted security spreads (over the 1-month LIBOR) as a function of the average pool-level seasoning calculated using estimation results from the specification reported in column (6) in panel B of Table A.6. The shaded area corresponds to 95% confidence intervals calculated using the delta method.

Figure A.3: Predicted Yield Spread as Function of Seasoning



Notes: This figure displays the predicted security spreads (over the 1-month LIBOR) as a function of the average pool-level seasoning calculated using estimation results from the specification reported in column (6) in panel B of Table A.6. The shaded area corresponds to 95% confidence intervals calculated using the delta method.

A.6 Early Prepayment Analysis

While default is clearly undesirable from the perspective of an MBS investor, the risk of early prepayment is another potentially negative outcome for mortgage investors. Residential mortgages contain a prepayment option that allows the borrower to fully repay the outstanding principal balance of the loan before it reaches full maturity. Early prepayment risk was an important consideration for investors in the period before the housing bust and financial crisis, especially given the low levels of default rates that prevailed during that period.

It is well established in the mortgage literature that interest rate movements largely drive the prepayment behavior of borrowers with fixed-rate mortgages. In contrast, the prepayment of adjustable-rate mortgages is typically driven by life events that are unrelated to interest rate movements, such as new housing purchases driven by employment changes or changes in household size due to the birth of a child or death of a family member. In the PLS market, however, in addition to responses to life events, prepayments of adjustable-rate mortgages were often driven by specific contractual features. In particular, the prepayment behavior of 2/28 and 3/27 hybrid ARMS, the most common types of PLS ARMs (accounting for about 75% of the market), was highly correlated with the duration of the period in which the interest rate was frozen: two years for the 2/28s and 3 years for the 3/27s. After the initial period, the interest rate would reset to a new level and track a market interest rate (such as the 6-month LIBOR or the 10-year Treasury rate). Because the interest rate typically reset to a higher level, many borrowers prepaid either right at or shortly after the reset period. In addition, many ARMs in the PLS market contained prepayment penalties that expired at the same time of the interest rate reset, providing further incentive for borrowers to wait until the reset date to exercise their prepayment options.[5]

We focus on the sample of 2/28 and 3/27 ARMs that did not default and define a negative outcome to be an ARM that was prepaid several months before the interest rate reset month.[6] We consider two cutoffs, six and nine months before the reset date, in defining our early prepayment indicator variables, as the most common type of prepayment penalty associated with these mortgages was six months of interest on 80% of the principal amount prepaid. An ARM that carried this prepayment penalty and prepaid more than six months before the reset date would generate lower cash flows for investors than a loan that prepaid at the reset date, and prepayment can thus be considered as a negative outcome for a PLS investor.

Table A.7 contains the results of the early prepayment analysis. Panel A displays results for parametric (quadratic) specifications, while Panel B displays results for the non-parametric specifications. We show results pertaining to various corrections for the potential "mechanical" selection issue discussed in Section 4.2 above. Specifically, we exclude from the sample loans that prepay within three, six, and nine months from origination. Just as in the case of default, however,

[5] For an excellent reference on the PLS market in general and for empirical analyses on the prepayment and default behavior of various types of PLS loans in particular, see Kramer and Sinha (2006). See Sengupta (2010) for a detailed discussion of the composition of loans in the Alt-A and subprime PLS markets.

[6] We eliminate defaults from our analysis to isolate voluntary prepayment risk.

this may be an "over-correction" to the extent that investors may be especially concerned with prepayments within the first year or so after origination, and such a restriction could eliminate the true signaling effect rather than simply correct the sample selection bias.

Table A.7 clearly shows a negative relation between time to sale and early prepayment risk. As months to sale increase, the likelihood of early prepayment decreases in a relatively monotonic manner. Focusing on the first two columns in the table (no correction), PLS loans sold six months after origination are approximately 6–7% less likely to prepay early compared to loans sold immediately, while loans sold nine months after origination are about 10–11% less likely to prepay early. The negative relation remains significant when we exclude prepayments that occur in the first few months after origination, but the non-parametric specification shows that the relation flattens for five months in columns (5) through (8).

In general, results on the correlation between time to sale and early prepayment are consistent with the default analysis and support the mechanism of using sales delays to signal quality. While PLS investors were likely concerned about significant credit risk in the case of a large downturn (which, of course, occurred), prepayment risk is present in both good and bad economic conditions, and it was thus an important consideration for mortgage investors. In addition, while our results suggest that asymmetric information on default risk did not play an important role in the subprime PLS market, asymmetric information on prepayment risk may have been important as the vast majority of 2/28 and 3/27 hybrid-ARMs were placed in subprime securities.[7] These findings are consistent with Agarwal et al. (2014), who find evidence of adverse selection with respect to prepayment risk but not default risk in the PLS market.

[7]In our CL sample, approximately 96% of 2/28s and 79% of 3/27s were in subprime securities.

Table A.7: Early Prepayment Results

Panel A: Parametric Specification

Correction:	None		$\leq$ 3 months		$\leq$ 6 months		$\leq$ 9 months	
Reset Month - Prepay Month	$\geq$ 6 Months	$\geq$ 9 Months	$\geq$ 6 Months	$\geq$ 9 Months	$\geq$ 6 Months	$\geq$ 9 Months	$\geq$ 6 Months	$\geq$ 9 Months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Months to Sale	-0.0129	-0.0152	-0.0089	-0.0105	-0.0111	-0.0131	-0.0144	-0.0169
	(6.20)	(6.28)	(4.11)	(4.15)	(4.76)	(4.75)	(5.66)	(5.57)
Months to Sale ^{2}	0.0007	0.0009	0.0004	0.0005	0.0012	0.0015	0.0019	0.0023
	(2.56)	(2.83)	(1.36)	(1.58)	(3.75)	(4.03)	(5.07)	(5.36)
# Loans	4,024,361	4,024,361	3,968,227	3,968,227	3,701,607	3,701,607	3,302,260	3,302,260
Adjusted R^2	0.13	0.13	0.13	0.13	0.12	0.11	0.10	0.08
Orig Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Non-parametric Specification

Correction:	None		$\leq$ 3 months		$\leq$ 6 months		$\leq$ 9 months	
Reset Month - Prepay Month	$\geq$ 6 Months	$\geq$ 9 Months	$\geq$ 6 Months	$\geq$ 9 Months	$\geq$ 6 Months	$\geq$ 9 Months	$\geq$ 6 Months	$\geq$ 9 Months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Months to Sale = 1	-0.024	-0.027	-0.023	-0.025	-0.024	-0.027	-0.025	-0.028
	(4.90)	(4.87)	(4.41)	(4.34)	(4.51)	(4.39)	(4.07)	(3.87)
Months to Sale = 2	-0.033	-0.038	-0.028	-0.032	-0.030	-0.034	-0.031	-0.035
	(6.90)	(6.88)	(5.70)	(5.63)	(5.81)	(5.69)	(5.17)	(4.98)
Months to Sale = 3	-0.039	-0.045	-0.030	-0.035	-0.032	-0.037	-0.034	-0.039
	(7.09)	(7.07)	(5.19)	(5.13)	(5.36)	(5.25)	(5.13)	(4.89)
Months to Sale = 4	-0.043	-0.049	-0.034	-0.038	-0.029	-0.033	-0.030	-0.033
	(7.24)	(7.48)	(5.36)	(5.47)	(4.51)	(4.53)	(4.50)	(4.38)
Months to Sale = 5	-0.049	-0.056	-0.040	-0.045	-0.026	-0.028	-0.028	-0.030
	(9.32)	(9.35)	(7.06)	(7.02)	(4.43)	(4.21)	(4.69)	(4.26)
Months to Sale = 6	-0.059	-0.066	-0.049	-0.055	-0.024	-0.024	-0.027	-0.027
	(8.59)	(8.93)	(6.93)	(7.15)	(3.03)	(2.88)	(3.24)	(3.02)
Months to Sale = 7	-0.064	-0.072	-0.054	-0.060	-0.027	-0.028	-0.014	-0.012
	(7.97)	(7.83)	(6.65)	(6.54)	(3.22)	(3.01)	(1.50)	(1.14)
Months to Sale = 8	-0.082	-0.090	-0.073	-0.078	-0.046	-0.047	-0.017	-0.011
	(10.65)	(11.38)	(8.99)	(9.56)	(5.57)	(5.63)	(1.91)	(1.22)
Months to Sale = 9	-0.096	-0.108	-0.085	-0.097	-0.059	-0.065	-0.011	-0.008
	(9.67)	(9.07)	(8.58)	(8.00)	(5.84)	(5.44)	(1.01)	(0.58)
# Loans	4,024,361	4,024,361	3,968,227	3,968,227	3,701,607	3,701,607	3,302,260	3,302,260
Adjusted R^2	0.13	0.13	0.13	0.13	0.12	0.11	0.10	0.08
Orig Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Sale Qtr FE	Y	Y	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y	Y	Y

This table displays results from the estimation of Equation 5 on adjustable-rate PLS loans in the CoreLogic data set originated in the 2002–2007 period. The dependent variable is an indicator variable for loans that prepay more than three months or six months before the month in which the interest rate resets from a fixed rate to an adjustable rate. All loans that are prepaid within three months of origination are eliminated from the sample. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, originator fixed effects, and the detailed list of covariates described in the text. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by year-quarter of origination.

A.7 Additional Summary Statistics and Tests

Table A.8 below displays summary statistics for the sample of triple-A securities included in the pricing analysis discussed in Section 4.6 in the main text.

Table A.9 displays the full set of estimation results (for all covariates) corresponding to the specification reported in Column (1), Panel A of Table 5 in the main text.

Finally, Table A.10 displays results from the estimation of Equation 5 on Alt-A and Subprime PLS loans in the CoreLogic data set that do not default within 10 months of origination. The specifications are identical to those in Table 9 in the main text.

Table A.8: Pricing Analysis Summary Statistics

	Mean	Standard Dev.	Minimum	25th Perc.	Median	75th Perc.	Maximum
Yield Spread	0.28	0.23	0.04	0.16	0.23	0.32	2.09
Months to Sale	3.3	1.4	0.3	2.2	3.1	4.2	9.0
# Loans	2,355	1,833	55	1,108	1,911	3,078	18,190
Log Loan Balance	12.2	0.4	11.0	11.9	12.1	12.4	14.9
FICO	640	43	413	609	624	682	764
FICO $<$ 580	0.20	0.15	0.00	0.01	0.22	0.31	0.87
580 $\leq$ FICO $<$ 620	0.19	0.12	0	0.05	0.22	0.27	0.67
620 $\leq$ FICO $<$ 660	0.23	0.08	0	0.19	0.24	0.28	0.68
660 $\leq$ FICO $<$ 700	0.18	0.09	0.01	0.11	0.15	0.25	0.72
FICO $\geq$ 700	0.20	0.21	0	0.06	0.10	0.35	0.92
CLTV	84	6	39	80	84	88	102
CLTV $<$ 70	0.13	0.08	0	0.08	0.12	0.17	0.87
70 $\leq$ CLTV $<$ 80	0.15	0.07	0	0.10	0.14	0.19	0.49
80 $\leq$ CLTV $<$ 90	0.28	0.13	0	0.20	0.27	0.36	0.92
90 $\leq$ CLTV $<$ 100	0.24	0.10	0	0.18	0.23	0.29	0.97
CLTV $\geq$ 100	0.20	0.20	0	0.02	0.16	0.32	0.96
LTV = 80	0.16	0.12	0	0.08	0.12	0.20	0.91
Term	359	15	120	356	359	360	480
Purchase Loan	0.42	0.20	0	0.27	0.40	0.57	1
Cashout Refinance	0.48	0.19	0	0.33	0.50	0.62	1
Primary Residence	0.87	0.13	0	0.85	0.91	0.95	1
Single-Family Property	0.73	0.11	0	0.68	0.75	0.80	0.99
Condominium	0.08	0.04	0	0.05	0.07	0.09	0.36
ARM	0.83	0.18	0	0.76	0.85	1	1
Interest-Only	0.21	0.28	0	0	0.10	0.26	1
Negative Amortization	0.10	0.30	0	0	0	0	1
Low Documentation	0.47	0.23	0	0.31	0.41	0.61	1
Balloon	0.08	0.15	0	0	0	0.05	1
Jumbo	0.19	0.24	0	0	0.10	0.27	1
Prepayment Penalty	0.69	0.21	0	0.65	0.74	0.81	1
Fraction in CA	0.26	0.17	0	0.13	0.23	0.34	1
Unemployment Rate	5.14	0.61	1.73	4.66	5.06	5.63	6.83
Predicted WAL	2.59	0.61	0	2.23	2.52	2.90	6.61
Subordination	0.86	0.14	0	0.81	0.85	0.91	2.65
# Securities				3,532			

Notes: This table displays summary statistics for the variables included in the pricing analysis presented in section 4.6. All mortgage characteristics correspond to averages that are calculated at the pool-level in the sample of CoreLogic loans, which includes mortgages backing subprime and Alt-A triple-A floating rate securities issued between January 2002 and December 2007. Yield Spread is the weighted average spread over the 1-month LIBOR of all triple-A securities with claims on cash flows for a given mortgage pool. Seasoning is the average age (# months) of all mortgages in a pool at the time of issuance. Predicted WAL is a model-based calculation of the expected weighted average life. Subordination is calculated as the ratio of the total face value of all triple-A securities associated with the deal to the sum of the remaining principal balances of all of the loans in the deal in the month of issuance.

Table A.9: Coefficient Estimates for Control Variables

Dependent Variable: Indicator for 60+ DQ within 36 months of origination		
	Coefficient	*t-statistic*
Months to Sale	-0.0094	(10.75)
Primary Residence (d)	-0.0012	(0.49)
Prepayment Penalty (d)	0.0687	(7.70)
ARM (d)	0.0281	(2.24)
Balloon Payment (d)	0.0890	(4.74)
Low Documentation (d)	0.0515	(9.74)
Missing Documentation (d)	0.0119	(1.80)
B or C Grade Mortgage (d)	0.1091	(9.38)
Single Family Property (d)	-0.0010	(0.69)
Missing Property Type (d)	0.0302	(7.12)
Interest-Only (d)	0.0130	(1.44)
Purchase Loan (d)	0.0015	(0.22)
Refinance (cash-out) (d)	0.0141	(3.04)
Missing Loan Type (d)	0.0141	(3.04)
Term	0.0001	(2.81)
LTV	0.0010	(3.96)
Missing LTV (d)	0.1632	(4.23)
$70 \leq$ LTV < 80 (d)	0.0352	(4.19)
LTV $= 80$ (d)	0.0257	(7.33)
$80 <$ LTV < 90 (d)	0.0443	(4.75)
$900 \leq$ LTV < 100 (d)	0.0608	(5.72)
LTV ≥ 100 (d)	0.0459	(4.04)
FICO	-0.0011	(8.59)
Missing FICO (d)	-0.8955	(8.54)
FICO < 580 (d)	-0.0614	(3.22)
$580 \leq$ FICO < 620 (d)	-0.0482	(4.53)
$620 \leq$ FICO < 660 (d)	-0.0149	(5.86)
$660 \leq$ FICO < 700 (d)	-0.0128	(2.72)
Interest Rate (at origination)	0.0110	(6.53)
Jumbo (d)	0.0217	(2.55)
Unemployment Rate (at origination)	0.0041	(7.63)
Cumulative Change in Unemployment Rate (36 months)	0.0244	(5.75)
House Price Level (at origination)	0.0016	(12.36)
Cumulative Change in House Prices (36 months)	-0.1583	(7.65)
# Loans	5,747,722	
Adjusted R^2	0.24	
Orig Qtr FE	Y	
State FE	Y	
Sale Qtr FE	Y	
Originator FE	N	

This table displays the coefficients for all variables included as controls in the regression shown in the first column of Panel A, Table 5 in the paper (Baseline parametric results for the sample of PLS Loans in McDash). The dependent variable is an indicator variable for loans that default over a 36-month horizon. Default is defined as a loan that is 60+ days delinquent. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, and year-quarter of sale fixed effects. The standard errors are heteroskedasticity-robust and are clustered by year-quarter of origination.

Table A.10: Correcting for Potential Selection Bias: Alt-A and Subprime PLS Loans

Panel A: Effect of Lender Fixed Effects

	Default Definition: 60+ DQ over 36 Months					
	Alt-A			**Subprime**		
Months to Sale	-0.0054	-0.0042	-0.0035	0.0020	0.0026	0.0039
	(7.62)	(6.86)	(6.92)	(4.83)	(5.72)	(8.60)
# Loans	1,848,602	1,847,871	1,846,633	5,426,811	5,425,136	5,423,582
Adjusted R^2	0.24	0.25	0.26	0.17	0.18	0.18
Orig YQ FE	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y
Lender FE	N	Y	Y	N	Y	Y
Lender x Orig-YQ FE	N	N	Y	N	N	Y
Other Controls	Y	Y	Y	Y	Y	Y

Panel B: Effect of Issuer Fixed Effects

	Default Definition: 60+ DQ over 36 Months					
	Alt-A			**Subprime**		
Months to Sale	-0.0040	-0.0023	-0.0026	0.0020	0.0040	0.0043
	(7.37)	(4.41)	(5.09)	(4.00)	(7.82)	(8.03)
# Loans	1,803,941	1,803,940	1,802,714	5,344,226	5,344,226	5,342,673
Adjusted R^2	0.25	0.26	0.26	0.18	0.18	0.18
Orig YQ FE	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y
Issue YQ FE	Y	Y	Y	Y	Y	Y
Lender FE	Y	Y	Y	Y	Y	Y
Lender x Orig-YQ FE	N	N	Y	N	N	Y
Issuer FE	Y	Y	Y	Y	Y	Y
Issuer x Issue -YQ FE	N	Y	Y	N	Y	Y
Other Controls	Y	Y	Y	Y	Y	Y

This table displays the results from the estimation of Equation 5 on Alt-A and Subprime PLS loans in the CoreLogic data set that do not default within 10 months of origination. The specifications are identical to those in Table 9 in the text. The dependent variable is an indicator variable for loans that default over a 36-month horizon. Default is defined as a loan that is 60+ days delinquent. Months to sale is defined as the number of months that elapse between origination and sale to a PLS issuer. All regressions include origination year-quarter fixed effects, state fixed effects, year-quarter of sale fixed effects, and the detailed list of covariates described in the text. The first row for each variable shows the regression coefficient and the second row shows the t-statistic. The standard errors are heteroskedasticity-robust and are clustered by year-quarter of origination.

Figure A.4: Example of Prospectus: Summary Mortgage Data Table

swap agreement will either increase or reduce the amount available to make payments on the certificates, as described under "*Description of the Certificates—Supplemental Interest Trust*" in this prospectus supplement. The interest rate

The mortgage loans have original terms to maturity of not greater than 360 months, have a weighted average remaining term to scheduled maturity of 358 months and have the following approximate characteristics as of the cut-off date:

Selected Mortgage Loan Pool Data[1]

	Group I		Group II		
	Adjustable-Rate	**Fixed Rate**	**Adjustable-Rate**	**Fixed Rate**	**Aggregate**
Scheduled Principal Balance:					
Number of Mortgage Loans:					
Average Scheduled Principal Balance:					
Weighted Average Gross Interest Rate:					
Weighted Average Net Interest Rate:[2]					
Weighted Average Original FICO Score:					
Weighted Average Original LTV Ratio:					
Weighted Average Combined LTV with Silent Seconds[3]:					
Weighted Average Stated Remaining Term (months):					
Weighted Average Seasoning (months):					
Weighted Average Months to Roll:[4]					
Weighted Average Gross Margin:[4]					
Weighted Average Initial Rate Cap:[4]					
Weighted Average Periodic Rate Cap:[4]					
Weighted Average Gross Maximum Lifetime Rate:[4]					
Weighted Average % of Silent Seconds[5]:					
Weighted Average Debt to Income Ratio at Origination%:					

Redacted

(1)
(2) **Redacted**
(3)

(4)
(5)

S-11

A.8 Solving the model with additional random delay in trade

In this appendix, we show that the equilibrium strategies and prices are unaffected by random delay in trade as specified in Section 5. In that section, we assume the selling protocol for a mortgage is as follows. The originator chooses to list a mortgage for sale at some publically observable date t. Once the originator lists the mortgage for sale, there is a random interval of time ν between the listing date and the date of sale distributed as an exponential random variable with parameter η. At the date of sale, buyers Bertrand compete for the asset all with common knowledge of the listing date t. We also assume that if the mortgage defaults, a lump sum recovery of αB_0 is paid to the holder of the mortgage. Given an originator of type λ, a listing date t, and a price p, the originators value is

$$U(\lambda, t, p) = \frac{(r_m + \alpha\lambda)B_0}{\gamma + \lambda}\left(1 - \frac{\eta e^{-(\gamma+\lambda)t}}{\gamma + \lambda + \eta}\right) + \frac{\eta e^{-(\gamma+\lambda)t} p}{\gamma + \lambda + \eta}.$$

Note that prices here will depend on the listing date and not the actual sale date. Thus, the first-order condition for listing date is now

$$(r_m + \alpha\lambda)B_0 - (\gamma + \lambda)P^*(t) + \frac{d}{dt}P^*(t) = 0$$

but for any separating equilibrium

$$P^*(T(\lambda)) = \frac{(r_m + \alpha\lambda)B_0}{r + \lambda},$$

which in turns gives the following ODE for the price of a mortgage listed for sale at time t

$$\frac{d}{dt}P^*(t) = (\gamma - r)P^*(t).$$

We restrict attention to equilibria in which the worst type does not delay, which in turn implies the following initial condition

$$P^*(0) = p_h = \frac{(r_m + \alpha\lambda_H)B_0}{r + \lambda_H}.$$

This ODE has the following solution

$$P^*(t) = p_h e^{(\gamma - r)t}.$$

Note this solution is the same as the equilibrium given in Proposition 1 up to the difference in p_h that results from including some recovery. As such, the equilibrium strategy for the originator is unaffected by publicly observable random delay.