Design of Financial Securities: Empirical Evidence from Private-label RMBS Deals

Taylor Begley and Amiyatosh Purnanandam*

March 14, 2013

Abstract

Using a representative sample of RMBS deals from the pre-crisis period, we show that deals with a higher level of equity tranche have significantly lower foreclosure rates that cannot be explained away by observable credit risk factors of the underlying loan pool. Further, securities that are sold from high-equity-tranche deals command higher prices conditional on their credit ratings. These results show that the equity tranche served as a signal of the unobserved pool quality of these deals. In addition, consistent with theoretical models of pooling and tranching, the level of the AAA-rated tranche is significantly higher for pools that bundle loans with commonality in their private information but with uncorrelated risks. Our results highlight the effectiveness of security design solutions in mitigating informational frictions even during the build-up of the subprime mortgage crisis.

Keywords: Security design, Mortgage-backed securities, Subprime mortgage crisis, Equity tranche.

JEL Classification: G20, G30.

---

*Both authors are at Ross School of Business, University of Michigan, 701 Tappan Avenue, Ann Arbor, MI 48109. Phone: (734)764-6886. Email: `tbegley@umich.edu` and `amiyatos@umich.edu`. We are grateful to Sugato Bhattacharyya, John Griffin, Charlie Hadlock, Han Kim, Gustavo Manso, Andrey Malenko, Greg Nini (discussant), Uday Rajan, Amit Seru, and seminar participants at Duke University, Federal Reserve Bank of Philadelphia, and the University of Michigan for helpful comments on the paper. All remaining errors are our own.

1 Introduction

While there is a large body of research on conflicts of interests in the residential mortgage-backed securitization (RMBS) market prior to the financial crisis of 2007-09, there is surprisingly little empirical research on how these financial instruments may have mitigated contracting frictions through security design (see Keys, Piskorski, Seru, and Vig (2012), Gorton and Metrick (2012) for recent surveys).[1] Economic theory suggests that market participants can use the level of equity tranche retention as a costly signal of unobserved asset quality to mitigate the effect of information asymmetry between deal sponsors and investors. However, several market observers have noted that such mechanisms completely lost their relevance during the pre-crisis period because deal sponsors were able to subsequently offload the equity tranche in secondary markets. Whether such contracting devices had any meaningful effect on market outcomes, therefore, remains an empirical issue that has not yet been analyzed. This paper aims to fill this gap in the literature.

We first analyze the role of the equity tranche in mitigating informational frictions in this market and then focus on unique tradeoffs of pooling and tranching that go beyond those present in the sale of a single asset. These issues have important implications for both our understanding of the economic theory underlying the securitization market and the ongoing policy debates regarding this market. For example, issues surrounding the equity tranche of securitization deals form an important part of the Dodd-Frank Reform Act.[2] In addition, our empirical study sheds light on one of the fundamental ideas in economic theory that argues that costly signal can allow high quality sellers to separate themselves from their

[1]See, for example, Keys, Mukherjee, Seru, and Vig (2010), Mian and Sufi (2009), Purnanandam (2011), Demyanyk and Van Hemert (2011), Je, Qian, and Strahan (2012), Loutskina and Strahan (2011), Acharya, Richardson, et al. (2009) for work related to the subprime mortgage crisis. See Ashcraft and Schuermann (2008) for a detailed analysis of the securitization process.

[2]In discussing the effects of risk retention requirements pursuant to the Section 946 of the Dodd-Frank Wall Street Reform and Consumer Protection Act, the treasury secretary stresses the importance of this tool in mitigating some contracting frictions and notes that: "...the academic literature on risk retention with respect to asset-backed securitization is limited." Scharfstein and Sunderam (2011) examine some other recent policy proposals and provide some suggestions for the more broad reform of the housing finance system.

observationally equivalent low quality counterparts.

RMBS sponsors create financial securities by pooling several mortgages together and then issuing marketable tranches against the pool's combined cash flows. Security design, therefore, is at the very core of the existence of this market. As in any market with asymmetric information, sellers (RMBS sponsors in our case) can signal the unobserved quality of assets in question by retaining a larger financial interest in the asset's performance (Leland and Pyle, 1977). Gorton and Pennacchi (1990), Boot and Thakor (2012), Riddiough (1997), DeMarzo and Duffie (1999) and DeMarzo (2005) provide theoretical models of loan sales or securitization with asymmetrically informed sellers and buyers. Motivated by this line of research, we test three fundamental implications of these signaling models. First, conditional on observable risk metrics of the underlying pool, does the size of equity tranche increase with the degree of information asymmetry between the sellers and buyers? Second, do pools with a higher level of equity tranche perform better ex-post as compared to observationally similar pools with low levels of equity tranche? And third, do security buyers pay higher prices for securities sold in a high-equity-tranche deal as compared to a similarly rated security in a low-equity-tranche deal?

We test these predictions using a representative sample of about 190 private-label RMBS deals from 2001-02 and 2005. Our sample includes over half-a-million loans made to a wide cross-section of borrowers across the country. We combine tranche level security data with the underlying loan and pool characteristics at the time of RMBS issuance. Finally, we track the default performance (i.e., foreclosure status) of each loan in the sample through December 2011. Thus, we have comprehensive information on loan characteristics of the underlying pool, tranche-level security data, and the ex-post foreclosure status of each loan in the pool. For our analysis, we group each tranche in a deal into one of the three bins according to its seniority in the payout structure: AAA-rated, mezzanine, and equity tranche.

To test the first implication, we use the percentage of no-documentation loans in a pool as a cross-sectional measure of the information asymmetry between the sellers and buyers. These loans are based on self-reported levels and sources of income, and unlike full-documentation loans, they are not accompanied by key information sources like federal and state income tax filings. This leaves a great degree of discretion with the originating institutions in terms of verifying employment and the level and stability of the borrower's income. Soft pieces of information like these are lost as loans pass through the securitization chain, widening the information gap between the seller and the investor. Deal sponsors also have the advantage of more precise information about the entire distribution and correlation structure of the loan pool. For relatively opaque no-documentation loans, these informational advantages are likely to be higher.[3] We find that deals with a higher proportion of no-documentation loans have significantly higher levels of equity tranche after controlling for the effects of observable pool characteristics. This finding is consistent with the key idea that investors are likely to have higher adverse selection concerns in relatively opaque deals, which in turn motivates the sponsors to create a larger informationally sensitive first-loss piece (DeMarzo and Duffie, 1999).

We also find that measures of observable credit risk, such as FICO score and loan-to-value (LTV) ratio, are unrelated to the size of the equity tranche. However, these variables, and *not* the proportion of no-documentation loans in the pool, drive the division of the sold tranches between AAA and mezzanine groups. These results suggest that concerns about asymmetric information explain the split between sold and initially unsold (equity) tranches, whereas observable and easier to price characteristics of the pool explain the relative distribution between AAA and mezzanine tranches.[4]

We next turn to our second question: did the size of the equity tranche signal sponsor's private information about the unobserved dimensions of the underlying pool's quality? While

[3]The use of this measure is also in the spirit of the "opacity" measure of theoretical papers by Skerta and Veldkamp (2009) and Sangiorgi, Sokobin, and Spatt (2009). Our assumption is that the information asymmetry between sponsors and buyers increases with the loan opacity.

[4]Consistent with this idea, we also find that the hard pieces of information explain the pricing of individual mortgages very well, whereas the extent of no-documentation loans has no effect on pricing measures (see also Rajan, Seru, and Vig, forthcoming).

we do not, by definition, observe the sponsor's private information at the time of security sale, we do observe the ex-post default performance (i.e., the foreclosure status) of every loan in our pools. The ex-post default performance of a loan can be decomposed into three components: (a) a component that is entirely driven by observable information such as the borrower's FICO score, LTV ratio, the geographical location of the property, and the nature of interest rate on the loan; (b) a component that is entirely driven by common macroeconomic shocks affecting all loans in the economy; and (c) a residual component. We relate the level of equity tranche created at the time of security issuance to the residual component to assess the impact of private information on loan quality. If sponsors used the equity tranche as a signal of their private information of the pool, then we should find lower abnormal default (i.e., lower residual component) for loans coming out of high-equity-tranche pools. In contrast, if the level of equity tranche was unrelated to the seller's private information, it should not correlate with the residual component of default.

We implement this idea using two models of default prediction. In the first model, we compute the expected default rate for each loan in the pool by fitting a default prediction model that accounts for the component of default that is driven by observable loan and property characteristics along with the year of loan origination. Based on prior research, we use covariates such as FICO score, LTV ratio, the nature of interest rate on the loan, and the location of the property as predictors of future default. We aggregate the fitted loan default rates to the pool level to obtain a measure of expected default rate for the pool. The difference between the actual default rate we observe and the expected default rate of the pool is our first measure of abnormal default rate.

In our second model, we use our sample of over 500,000 loans to create an observationally similar hypothetical matched pool for each actual pool in our sample. For each loan in a given pool, we find a loan from the entire sample (excluding the actual loan's own pool) with similar risk characteristics. In addition to observables such as FICO scores and LTV ratios, we ensure that the simulated pool is observationally identical along the geographical location of the property. Thus the simulated pool is likely to have similar correlation structure across loans as well. The default rate of the actual pool over and above its hypothetical match is our second measure of abnormal default rate. By comparing the pool with their match, we effectively difference out the effects of observable credit and macroeconomic risks as well as correlation structure of the pool to the extent that it is driven by geographical diversity. The simulated pool, by design, removes the element of the sponsor's private information. Thus the difference in the realized and simulated default rates provides a reasonable measure of private information driven abnormal default.

We find that deals with a higher level of equity tranche have significantly lower abnormal foreclosure rate among pools with higher proportion of no-documentation loans. Said differently, for relatively opaque pools, higher equity tranche predicts better performance in future. In economic terms, such pools have 24% lower foreclosure rates that cannot be explained away by observable credit risk characteristics and macroeconomic conditions. The effect of equity tranche on relatively transparent pools is statistically indistinguishable from zero. These results are consistent with the idea that sponsors created higher equity tranche in deals with favorable information on unobservable dimensions.

The third implication of signaling models we study is the presence of downward sloping demand curve for informationally sensitive securities (see DeMarzo and Duffie (1999)). If the equity tranche works as a signal of sponsors' private information, then the market price of sold tranches should increase with the level of equity tranche in the deal. We test this prediction by estimating the effect of the level of equity tranche on the yield spread of sold tranches after controlling for the credit rating of the security. Since security prices are not directly available, following earlier literature we take yield spread, defined as the markup over a risk-free benchmark rate, as the measure of pricing (see Je et al. (2012)). We find that sold tranches command higher prices (i.e., lower yield spread) for the same credit rating class if they are backed by higher equity tranche. Again, the effect is concentrated within opaque pools. In addition, the effect is stronger for non-AAA rated tranches, i.e., tranches that are expected to have higher information sensitivity. Together, these results show that opaque pools with higher equity tranche have lower abnormal default rate ex-post, and ex-ante they are priced favorably from the issuers perspective. These findings are consistent with the signaling role of equity tranche in RMBS deals.

As mentioned earlier, it has been often argued that equity tranche lost its signaling role during the pre-crisis period because deal sponsors subsequently sold them to other entities. Our empirical tests provide evidence contradicting these claims. While we cannot track the ownership of the equity tranche over time directly, sponsors did have a considerable amount of retained interest in mortgage-backed securities on their balance sheets during our sample period.[5] In addition, the buyers of equity tranches in the secondary market were often active hedge funds or CDO managers whereas the more senior tranches were typically bought by less sophisticated investors such as retirement funds.[6] Such a segmentation in this market is likely to provide incentives to deal sponsors to retain relatively larger portion of better deals since equity tranches are sold to relatively informed buyers. In addition, some of the sales of equity tranches were motivated by regulatory capital arbitrage considerations in which the sponsor retained residual interest in the risk (see Acharya, Schnabl, and Suarez (2013)). Our analysis shows that, despite the possibility of subsequent sale, a higher level of equity tranche predicts better future performance along the unobserved dimension and the securities in these deals commanded a higher price from investors at issuance.

To further explore the role of information in the design of RMBS, we examine DeMarzo's (2005) key prediction about the equilibrium size of the information-insensitive AAA-rated

[5]For example, Goldman Sachs' 2005 annual report states, "During the years ended November 2005 and November 2004, the firm securitized $92.00 billion and $62.93 billion, respectively, of financial assets, including $65.18 billion and $47.46 billion, respectively, of residential mortgage-backed securities." The report also shows the value of their retained interests in mortgage-backed securities to be $2.928 billion and $1.798 billion, respectively, for those time periods. A back of the envelope calculation suggests that (2.928-1.798)/62.93 = 1.73% was retained during this time period. While this is only a rough approximation, it clearly shows that deal sponsors did retain at least a piece of these securities. A similar computation using information from Merrill Lynch's annual reports gives an estimate of 2.84%.

[6]For example, see the representative deal from CitiBank in Financial Crisis Inquiry Commission, Figure 7.2 on page 116.

tranche. His theoretical model explicitly recognizes the unique tradeoffs involved in the pooling and tranching of assets that differentiate this market from the sale of a single asset. He argues that while pooling provides risk-diversification benefits, it also imposes a cost on the seller through its "information destruction" role. When a sponsor sells claims on an entire pool of loans, he forfeits his option to sell the optimal amount of each loan to outside investors on a loan-by-loan basis, thus destroying his ability to make full use of asset-specific information. DeMarzo's (2005) model predicts that pools with common private information but uncorrelated risk across loans obtain maximal benefits through the securitization structure.

For an empirical test of this model, we consider pools with a single originator as pools with relatively homogeneous private information across loans. Lenders are likely to differ in terms of their screening technology, specialization, and organization structure to process information. Therefore, if all loans in a pool come from a single originator, then the pool is likely to be more homogeneous on the private information dimension. In addition, we consider the commonality in the type of loans in a pool as an additional proxy for common private information. We assign each loan to one of four bins based on (a) whether the loan is backed by a single-family or multi-family home and (b) whether the residence is owner occupied or not. We consider pools concentrated with loans in one of these bins to have more common private information, whereas pools that have loans spread over all these four categories are considered to have diffused private information.[7]

We find that deals that are backed by one originator (common information) and have a more geographically diverse pool of loans (risk diversification) have about 2 percentage points larger AAA-rated tranche, after controlling for the effects of key observable credit risk factors. Similarly, pools that are backed by similar types of assets and have higher geographical diversity have significantly larger fraction of AAA-tranche. Together with our earlier analy-

[7]Factors such as proximity to city centers, rising house prices, and demographic changes are likely to affect these categories of homes in different manners. Due to these factors, GSEs and private sector market participants often use different valuation models for these categories of properties.

ses, we conclude that information-based theories are able to explain cross-sectional patterns in RMBS security design even during the build-up to the mortgage crisis period. Whether the level of equity tranche was optimal for the sponsor's incentives is beyond the scope this paper and left for future research. In robustness tests, we also show that our results remain robust to concerns such as sponsor's reputation, servicing rights, and influence over credit rating agencies.

Our study connects to several strands of literature in banking, securitization, and real estate finance. Griffin and Tang (2012) study rating inflation in a large sample of CDOs from 1997 to 2007 and conclude that rating agencies used their subjective assessment to increase the size of AAA-rated tranche beyond the model-implied objective level. Ashcraft, Goldsmith-Pinkham, and Vickery (2010) report a significant decline in RMBS subordination levels between 2005 and mid-2007 and show that the ratings are correlated with ex-ante credit risk measures and they do explain subsequent deal performance.[8] Demiroglu and James (2012) show that the originator's affiliation with the sponsor or servicers results in better ex-post performance of the securitization deals. Hartman-Glaser (2012) studies the effect of seller's reputation capital in these contracts. Je et al. (2012) show the influence of large sponsors on credit rating agencies. An, Deng, and Gabriel (2011) study the role of conduit lenders in mitigating informational problems in CMBS deals. Our work also relates to a growing and large literature regarding the conflicts of interest in the securitization market (see Je et al., 2012; Keys et al., 2010; Purnanandam, 2011; Downing, Jaffee, and Wallace, 2009).[9] Unlike these studies, our paper does not study the motivations behind and differences in securitized versus retained loans, or the possibility of originator moral hazard that comes with securitization. Instead we highlight the effect of informational frictions within the set of securitized deals and the RMBS contract's ability to mitigate some of these frictions.

[8]See Cornaggia and Cornaggia (forthcoming), Becker and Milbourn (2011) and Bongaerts, Cremers, and Goetzmann (2012) for some recent studies on credit ratings for corporate bonds.

[9]See Benmelech, Dlugosz, and Ivashina (2012) on securitization in the case of Collateralized Loan Obligations and Nadauld and Weisbach (2012) for the effect of securitization on the cost of debt.

Much of the current literature focuses on the informativeness of ratings, the optimal subordination level, nature of relationship between syndicate members, and the possibility of rating inflation during the years leading up to the crisis. Our paper provides one of the first pieces of evidence on the signaling role of equity tranche in RMBS deals. In addition, to the best of our knowledge, this is the first paper that tests predictions that are unique to pooling and tranching structure that go beyond signaling models for single asset sale. The rest of the paper is organized as follows. Section 2 discusses the theoretical motivation and develops the main hypotheses of the paper. Section 3 describes the data. Section 4 presents the results and Section 5 concludes the paper.

2 Theoretical Motivation & Hypothesis Development

Absent any market frictions, the pooling and tranching of securities cannot be a value enhancing security design. Theoretical research, therefore, focuses on frictions such as information asymmetries, transactions costs and market incompleteness to explain a financial intermediary's motivations behind asset-backed securitization. At a broad level, the optimal design of financial securities serves as a mechanism to resolve inefficiencies through costly signaling (e.g., Leland and Pyle, 1977; DeMarzo and Duffie, 1999), allocation of cash flow rights (e.g., Townsend, 1979; Gale and Hellwig, 1985), or allocation of control rights (e.g., Aghion and Bolton, 1992).[10] We focus on the asymmetric information-based theories in the paper for two main reasons. First, in recent years there has been considerable discussion and debate among academics, practitioners, and regulators regarding the presence of information problems in this market. Second, information-based theories provide testable cross-sectional hypotheses that are unique to this market.

For expositional simplicity, we introduce some notation in this section as we develop our

[10]This is not a comprehensive list of design solutions. There are other motivations for security design such as transaction costs and market incompleteness. For example, in an incomplete markets setting, Allen and Gale (1988) argue that optimal security design assigns state-contingent cash flows to the agents that values it the most in that state.

hypotheses. Consider a mortgage i in pool p and denote its payoff by a random variable $\tilde{Y}_{ip}$. Let X_{ip} be a set of publicly observable loan characteristics such as FICO score and loan-to-value ratio. We can then express the loan's payoff conditional on observable signals as follows:

$$\tilde{Y}_{ip}|X_{ip} = \tilde{I}_{ip} + \tilde{z}_{ip} \tag{1}$$

I_{ip} is the private information of the sponsor and $\tilde{z}_{ip}$ represents a random shock to the loan's performance. I_{ip} is a known quantity to the sponsor, but remains a random variable to outside investors.

As the distribution of $\tilde{I}_{ip}$ widens, the asymmetric information concerns increase and investors of debt securities issued against this payoff become more concerned about the adverse selection problem (DeMarzo and Duffie, 1999). In such pools, outside investors require the sponsor to hold higher level of equity tranche in equilibrium. Therefore, considering two pools with observationally similar loans (i.e., similar X_{ip}), the pool with wider support of $\tilde{I}_{ip}$ is likely to have a larger equity tranche. This argument forms the basis of our first test that more opaque pools (i.e., those with a higher level of no-documentation loans) should have larger equity tranche.

The optimal quantity ($q^*(I_{ip})$) of security sold to outside investors depends on the sponsor's private information: sponsors sell a relatively smaller fraction of claims on the pool to outsiders if their private information is good. Thus, an implication of the signaling models is that conditional on observable characteristics, pools that are backed by a higher level of equity tranche should perform better ex-post. This forms the basis of our second test that relates the level of the equity tranche to ex-post default performance of loans.

Finally, an important implication of these models is that the demand curve for security is downward sloping: as sponsors sell higher fraction of security of outsiders, outsiders rationally infer the sponsor's private information to be worse and demand a liquidity discount. This forms the basis of our third test that the yield spread paid for the same credit rating tranche is lower for deals backed by higher equity tranche.

The securitization of a pool of assets adds additional complexity to this standard lemons-discount model. DeMarzo and Duffie (1999) show that the quantity of assets retained by the seller serves as a costly signal of the asset's cash flows in a similar manner as in the case of single security sale. DeMarzo (2005) extends this model to address the sponsor's choice between selling assets individually versus selling them as a pool and then studies the optimal tranching decisions. Continuing with our notation from above, the payoff of a pool of two loans with equal proportion in the pool is given as follows:

$$\tilde{Y}_p \equiv \frac{1}{2}\tilde{Y}_{1p} + \frac{1}{2}\tilde{Y}_{2p}|X_{1p}, X_{2p} = \frac{1}{2}(\tilde{I}_{1p} + \tilde{z}_{1p} + \tilde{I}_{2p} + \tilde{z}_{2p}) \tag{2}$$

Assuming that the volatility of information component ($\sigma^2_{I_i}$) and the residual risk ($\sigma^2_{z_i}$) is same for both loans, the volatility of the pool's payoff can be broken down into two components:

$$\Sigma_{Y_p} = \frac{1}{4}[\underbrace{\{\sigma^2_I + cov(I_1, I_2)\}}_{Information} + \underbrace{\{\sigma^2_z + cov(z_1, z_2)\}}_{Risk}] \tag{3}$$

The second piece, which we call the *risk-diversification* component, in the above expression is the well-known diversification benefit of pooling. The first piece, the *information-commonality* component, creates an *information-destruction* cost to the sponsor in DeMarzo's model. Pooling assets with diverse private information (i.e., assets with lower $cov(I_1, I_2)$ in the above model) is not advantageous to the sponsor in equilibrium. This happens because the sponsor loses the flexibility of selling the optimal fraction of each loan on a loan-by-loan basis. Based on loan-specific private information, the sponsor can choose an optimal amount of sale ($q^*(I_1)$ or $q^*(I_2)$) for each loan: by combining them, the sponsor is forced to choose a single amount for the aggregate pool which, in turn, reduces his ability to fully exploit his information.

The information-destruction cost is lowest when the underlying loans have common sources of private information, i.e., when the private information is "general" rather than asset-specific. In such cases the optimal retention amount is likely to be similar across all assets in the pool, which minimizes the information-destruction cost of pooling. Thus, unlike the second component where sponsors benefit by pooling loans with uncorrelated residual risks (lower $cov(z_1, z_2)$), the sponsors benefit by pooling loans with common private information (higher $cov(I_1, I_2)$). From the demand perspective, investors may also prefer loans with similar type of private information because it is easier for them to value a basket of loans that have similar informational concerns than a basket of dissimilar loans. DeMarzo's model predicts that pooling and tranching can be most beneficial if the private information is common, but the risk is asset-specific.[11] Empirically, we test this hypothesis by analyzing whether sponsors are able to sell larger portion of their pool as highly rated AAA-securities in such deals conditional on observable credit risk characteristics. We describe in detail our choice and construction of proxies for $cov(z_1, z_2)$ and $cov(I_1, I_2)$ based on loan and property characteristics in the later sections of the paper.

3 Sample and Descriptive Statistics

We construct a novel dataset of RMBS pools and tranches using hand-collected data from relevant SEC filings and matching them with loan-level data obtained from CoreLogic, a private data vendor. Our loan-level data contain information on characteristics such as FICO scores and LTV ratios at the time of the deal as well as their ex-post performance. In particular, we have information of whether the property entered into foreclosure any time

[11]In a recent paper, Hartman-Glaser, Piskorski, and Tchistyi (2011) analyze optimal contracts in this market in a setting where the underwriters undertake costly unobservable effort and more crucially the contract depends on the realization of mortgage defaults in the pool. In this setup, investors learn about the true effort of underwriters as mortgages default over time. An interesting feature of their contractual setting is that pooling dominates single mortgage sale through its *information-enhancement* effect. The information-enhancement effect happens due to the ability of investors to learn quickly about the true quality of entire pool as a single mortgage in the pool defaults.

from the deal date through December 31, 2011. Since we do not have data on the entire universe of RMBS deals during the pre-crisis period, we take special care in ensuring that our sample is representative. We provide detailed description of sample selection criteria and data collection in the Appendix. Figure 1 presents a schematic diagram of a representative deal and the relevant data sources. Our random sample begins with 196 securitization deals from 2001-02 and 2005 covering a wide range of sponsors and originators. Our main empirical tests are based on a sample of 163 deals that have all the necessary information needed for the analysis. These deals are backed by approximately a half million loans and there are approximately 3000 tranches issued against their combined cash flows. Our sample represents about 12% of the dollar volume of securities issued in the market during the sample period. Thus, we have a representative as well as economically meaningful sample of deals from the pre-crisis period.

Table 1 presents summary statistics of our sample. We winsorize all variables at 1% from both tails to remove any outlier effects. Panel A of the table presents overall loan-, pool, and tranche-level descriptive statistics. Based on 501,131 loans that enter our full sample, the average loan's FICO score is 656 with an LTV ratio of 77%. These numbers are broadly in line with Keys et al. (2012), who present detailed statistics on this market during 1998-2007. There is considerable cross-sectional heterogeneity in these two key measures of credit risk across loans. About 66% of the loans are classified as Adjustable Rate Mortgages (ARM) and 89% of loans are owner occupied residences. Turning to pool-level statistics, the average pool has $776 million in principal amount and is backed by 3,150 loans. No-documentation (*NoDoc*) loans, obtained directly from the deal prospectus, make up about 19% of all loans in the average pool. There is significant variation in this measure as it ranges from about 3% of the pool in the 25th percentile to 35% of the pool in the 75th percentile. About 52% of pools belong to the *Late* period, defined as deals originated in 2005.

We measure geographical diversification as the complement of one-state concentration of the loan. We first compute the percentage of loans in a pool that comes from each state and then identify the state with maximum share of loans in the pool. Our measure of geographical diversification (*GeoDiverse*) is simply one minus this share.[12] The average pool in our sample has *GeoDiverse* score of 59, representing one-state concentration of 41%.

The sample contains a wide variety of institutional players covering commercial banks, investment banks, and mortgage companies. The full sample contains 22 unique sponsors and 32 unique top originators. We present the list of institutions that are most frequently involved in the deals in our sample in the Appendix. Data on loan origination plays a prominent role in our analysis relating the information destruction effect of pooling. Using data on loan origination, we construct an indicator variable (*OneOrig*) that equals one for pools where all loans were originated by a single originator. About 47% of pools in our sample were originated by a single lender. The key measure of future performance of these loans is their foreclosure status. 16% of the loans in the sample enter foreclosure anytime from the deal origination until December 2011. The dollar-weighted pool level foreclosure rate has a mean of 12% which varies from 3% for the 25th percentile pool to 18% for the 75th percentile.[13]

Panel B in Table 1 provides some basic statistics relating borrower credit risk factors and eventual foreclosure. Consistent with intuition and past literature, we show that borrower's with higher FICO scores, lower LTV ratios, and fixed-rate mortgages defaulted at lower rates. Also, loans from the earlier period are about half as likely to end up in foreclosure, showing strong vintage effect.

Panel C of Table 1 provides descriptive statistics on the tranche structure. Overall, 90.40% of the average deal is tranched into AAA-rated security, while only 1.20% of the average deal is in the equity tranche. Panel C also illustrates the evolution of the average deal structure over our sample period. Relative to the 2001-02 deals, the size of the average

[12] We perform several robustness tests using alternative measures of geographical diversification such as Herfindahl index across states and concentration in top-three states. Our key results remain similar.

[13] The foreclosure information is available for a slightly lower number of deals because it is based on the sample formed by the intersection of our hand-collected data with CoreLogic foreclosure data.

2005 AAA-rated tranche drop from 92.56% to 88.32%, while the equity tranche more that doubled from 0.72% to 1.63%. To give these numbers some perspective, Benmelech and Dlugosz (2009) find that about 71% of CLO pools are rated AAA and 11% are unrated while Stanton and Wallace (2011) find about 84-87% of CMBS pools are rated AAA and 3-4% are unrated equity tranche. Not surprisingly, RMBS tranching structure is closer to the numbers reported by Stanton and Wallace (2011) as compared to the summary statistics of Benmelech and Dlugosz (2009), which comprises several other types of assets in the pool.

We use the level of equity tranche at the time of security sale as the measure of the sponsor's retained interest in the pool. Some observers have argued that if sponsors offload a bulk of this risky tranche in the secondary market, then it has no value as signal. Ideally, we want the amount of securities retained by the sponsors for a long time after the initial deal creation as the measure of retained interest. Unfortunately, this information is not available due to limited disclosure requirements. In the absence of this proxy, the unsold equity tranche at the time of security sale provides the most natural alternative measure. There are several economic reason to support the use of equity tranche for our empirical exercise. First, anecdotal evidence suggests that banks often retained part of this exposure on their balance sheet. For example, the Financial Crisis Inquiry Commission's Report presents a case study of an MBS deal issued by Citi Bank in 2006 called CMLTI 2006-NC2. They provide details on the identity of the holders of different tranches of this deal (see page 116 of the report). The AAA-tranches (78% of the deal) were bought by foreign banks and funds in China, Italy, France, and Germany, the Federal Home Loan Bank of Chicago, the Kentucky Retirement Systems and a few other parties. The Mezzanine tranches comprising 21% of the deal were mostly bought by the sponsors of CDOs. More relevant to our work, Citi Bank did retain a part of the equity tranche in the deal sharing the rest with Capmark Financial Group, a real-estate investment firm.

Second, as suggested by the above deal, the buyers of equity tranches are on average more informed than the buyers of safer tranches. The asymmetric information problem between the buyers and sellers in this market is likely to be relatively lower than the corresponding problem at the time of initial sale. Thus the sponsors' incentive to keep higher proportion of deals with favorable private information remains preserved.

Third, even though the sponsors can subsequently offload this risk in the secondary market in the medium to long run, in the immediate aftermath of the deal the risk remains with the sponsor. Indeed there have been numerous commentaries on the role of warehousing risk in this market during the sub-prime mortgage crisis. Thus the extent of equity tranche at the time of security sale provides a clean proxy for risk exposure during the initial period. Fourth, as shown by Acharya et al. (2013), there are several instances of securitization motivated by regulatory capital arbitrage. In such deals the residual credit risk stayed with the sponsors.

Finally, we check the annual reports of major sponsors in our sample and find significant equity tranche retention on their balance sheets. For example, Lehman Brothers had approximately $2 billion of non-investment grade retained interests in residential mortgaged-backed securitization as of November 30, 2006. Similarly evidence is gathered from the annual reports of Goldman Sachs and Merrill Lynch during this period. While this method does not allow us to get pool level retention amount, it does show that in aggregate the sponsors were holding significant amount of unrated tranches on their balance sheets. Overall, these arguments suggest that equity tranche created at the time of RMBS issuance imposes significant cost on the sponsor consistent with the underlying theoretical assumption of the signaling models.

In the final analysis, the effect of equity tranche on loan quality remains an empirical question. If the deal sponsors did not care about the risk of equity tranche because of the possibility of future sale, then we should not find any effect of equity tranche on future default performance. In contrast, if they did care about this risk, then we expect to observe better performance for deals with high equity tranche. Our empirical analysis allows us to test these competing hypotheses in the paper.

4 Empirical Results

In this section, we present the results of our empirical tests for the key predictions outlined in Section 2. We begin by relating the level of equity tranche to asymmetric information concerns of RMBS buyers. Next, we relate the level of the equity tranche to ex-post foreclosure performance of the entire pool and ex-ante pricing of sold tranches. Our final set of tests examine the unique tradeoffs of pooling and tranching that go beyond those present in single asset sales.

4.1 Cross Sectional Determinants of Tranche Structure

One of the key predictions of information-based models is that the level of the equity tranche should increase with the asymmetric information concerns about the underlying pool. In such deals, debt security buyers are more likely to demand a higher level of equity tranche to mitigate their concerns about adverse selection. We estimate the following pool-level regression model to examine this:

$$EquityTranche_p = \alpha + \beta(InfoAsym_p) + \theta(Late_p) + \gamma(Credit_p) + \delta(GeoDiverse_p) + \epsilon_p \quad (4)$$

We use the percentage of *NoDoc* loans in the pool as the proxy for the extent of asymmetric information ($InfoAsym_p$), or opacity of the underlying pool, faced by the investors.

We separate out the effect of observable risk factors in this regression model by including several pool-specific measures of credit risk, $Credit_p$, as explanatory variables. These variables include the weighted average FICO score, the weighted average LTV ratio, and the fraction of adjustable rate mortgages (ARM) in the pool. The first two variables directly measure the credit risk and leverage of the deal, and hence are predictors of future default by the borrower. We include percentage of ARM in the pool as an additional control variable for both credit and interest rate risks of the pool. We control for the time effect by including an indicator variable *Late* that equals one for deals from 2005, and zero for the earlier period.[14] Inclusion of this variable in the regression model allows us to separate the effect of aggregate macroeconomic shocks such as the level of interest rate and the demand of such securities from the outside investors. We include a measure of geographical diversification ($GeoDiverse_p$) of the pool as an additional variable to capture the effect of correlations of loans within the pool.

Columns (1) and (2) of Table 2 present the results. In column (1), which only includes *Late* as a control variable, we find a positive and significant (at 1%) coefficient on the $\%NoDoc$. In economic terms, one standard deviation increase in no-documentation loans (17.8 percentage points) is associated with an increase of about 0.45 percentage points, or a 60% increase in the equity tranche level for the median deal. Further, the coefficient estimate on *Late* shows that the extent of equity tranche increased in later periods. In column (2), we include all the control variables and find that the estimate on $\%NoDoc$ remains virtually unaffected. Overall, these estimates show that the opacity of the loan pool is a key driver of the size of the equity tranche. Observable credit risk characteristics of the pool such as FICO score and LTV ratio do not explain significant variation in equity tranche across deals. These results are consistent with our first prediction that the level of equity tranche increases with the size of the wedge between sponsors' and buyers' information sets.

We next turn to the division of sold tranches (i.e., the complement of the equity tranche) into AAA and Mezzanine categories. The dependent variable in these specifications measures the ratio of Mezzanine tranche to the sum of AAA-rated and mezzanine tranche in the deal. The *Mezzanine-to-Sold* ratio is 8.57% for the average deal in our sample with significant cross-sectional variation. Using the same modeling approach as above, we regress explanatory variables capturing credit risk and information concerns on this dependent variable. Columns (4) and (5) in Table 2 present the results.

[14] In unreported regressions we control for even finer time-periods such as the month or quarter of the deal. Our results do not change.

While $\%NoDoc$ has no effect on the division of sold tranches across Mezzanine and AAA category after controlling for observable measures of credit risk, this division is explained well by observable credit risk factors such as FICO score and LTV ratio. As expected, pools with lower FICO score and higher LTV ratio have relatively higher proportion of Mezzanine (lower AAA) tranche within the sold portion of the deal. Loan pools with more geographical diversity have relatively higher proportion of AAA-rated tranche. These results show that pools with lower observable credit risk and higher risk diversification have relatively higher AAA-rated tranche.

Taken together with the earlier results, we find that concerns about private information drive the cross-sectional dispersion in the level of the equity tranche, whereas hard pieces of information such as FICO score, LTV ratio, and geographical diversification drive the division of the sold tranche into AAA and mezzanine categories. In addition to the slope coefficients, the R^2 of the models provides an interesting insight as well. For the equity tranche regression, inclusion of observable credit risk variables improves the model's R^2 from 26.8% to a marginally higher 31.8% (columns 1 and 2), whereas the corresponding R^2 improves from 33.4% to 85.7% for the *Mezzanine-to-Sold* regression (columns 5 and 6). Hard pieces of information are easier to price and therefore can be incorporated in the security pricing relatively easily. In contrast concerns about information asymmetry are harder to price and the level of the equity tranche emerges as an additional contracting tool in such settings. Our results provide evidence in support of these arguments.

A potential concern with our analysis is the omission of some observable credit risk metrics that correlates both with our measure of loan opacity and the extent of equity tranche. In column (3) of Table 2, we include the weighted average interest rate on mortgages in the pool as an additional explanatory variable in the regression. Interest rates are likely to capture all the publicly available information about the credit risk and the priced component of the originator's private information. Thus the inclusion of interest rate in the model provides a reasonable control for the measures of credit risk that may be known to the investors, but not to us as econometricians. The estimate shows that the coefficient on $\%NoDoc$ remains unaffected. We repeat the same exercise for the division between AAA and mezzanine tranche in column (6) and show that our results remain unchanged for that model as well.

As an alternative estimation technique, we also estimate a seemingly unrelated regression model for the proportion of AAA, mezzanine, and equity tranche in a deal, which we do not tabulate for brevity. Our key results are stronger for this specification. We also perform our tests with standard errors clustered at the sponsor level and find that our inferences are unaffected. However, we need a sufficiently large number of clusters to obtain consistent standard errors using this method. Since we only have 22 clusters, we present our results without clustering.

4.2 Ex-Post Performance of Pools

We have shown that more opaque pools have a relatively larger equity tranche. While consistent with the broad idea behind adverse selection models, this test is not conclusive in terms of evaluating the role of the equity tranche as a signal of underlying pool quality. Did the creation of a larger equity tranche indicate deal sponsors' relatively favorable private information about the pool assets? We exploit the cross-sectional variation in equity tranche within the opaque pools along with data on ex-post performance of mortgages to answer this question.

It is an extremely challenging task to tease out the unobserved private information of sponsors solely based on publicly available information at the time of security sale. However, we do observe the ex-post default performance (i.e., incidence of foreclosure) of every loan in the pools. If sponsors with favorable private information about the underlying pool create a larger equity tranche as a signal of good private information, then we expect relatively better ex-post default performance by such pools after conditioning on observable pool characteristics. In other words, we expect *abnormal* default performance of high equity tranche pools to be better, where *abnormal* default performance measures the actual default rate of the pool against a benchmark default rate based on ex-ante observable information. We use a standard default model and then a simulation exercise to create two benchmark expected default rates to test these predictions. We first describe the empirical design and then discuss the construction of abnormal default performance measures in greater detail.

We relate the cross-sectional variation in equity tranche to the abnormal default performance of opaque pools, as proxied by the proportion of no-documentation loans in the pool, using the following empirical model:

$$AbDefault_p = \beta_0 + \beta_1(Opaque_p) + \beta_2(HighEq_p) + \beta_3(Opaque_p \times HighEq_p) + \sum \gamma X_p + \epsilon_p \quad (5)$$

$AbDefault_p$ is the abnormal default rate of pool p. Our regression model uses a difference-in-differences design to estimate the effect of the equity tranche across opaque and transparent pools. For easier economic interpretation, we use indicator variables for opaque and transparent pools as well as high and low equity tranche pools in the regression. *Opaque* equals one for pools that have an above median percentage of no-documentation loans, and zero otherwise. *HighEq* equals one for pools that have higher than median level of equity tranche, and zero otherwise. X_p measures some pool level control variables such as the pool's weighted average FICO scores and the LTV ratio and the year of the deal. We include them in the regression model to capture any remaining pool specific variation that does not get captured by the loan-level matching strategy. The regression coefficients in this model estimate the abnormal default rate across different pools as shown below:

	Transparent Pool	Opaque Pool
Low Equity Tranche	β_0	$\beta_0 + \beta_1$
High Equity Tranche	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$
Difference	β_2	$\beta_2 + \beta_3$

Our interest lies in both $\hat{\beta}_3$, the difference-in-differences estimator, and the sum of coefficients $\hat{\beta}_2+\hat{\beta}_3$. The sum of these coefficients provides an estimate of the difference in abnormal default rates across high and low equity tranche deals for opaque pools. If sponsors used equity tranche as a tool to signal their favorable private information, then we expect the sum of these coefficients to be negative. $\hat{\beta}_3$ provides an estimate of the differential effect of equity tranche on default rate for opaque pools as compared to the corresponding difference for the transparent pools. In other words, $\hat{\beta}_3$ differences out the effect of equity tranche on abnormal default rate that we observe within the relatively transparent pool. We expect this difference-in-difference estimate to be negative as well. If the equity tranche correlates with abnormal default rate of pools for reasons unrelated to information asymmetry at the time of issuance, then this estimator provides a useful way to separate out those effects. For example, factors such as economy wide abundance of capital or investment opportunity set of the sponsors can potentially affect both the level of equity tranche and the ex-post default rates for all the pools. Our estimator is able to remove all such effects as long as they affect opaque and transparent pools in similar manner.

Default Models

Our goal is to parse out the effect of observable loan and property characteristics from the default performance (i.e., foreclosure rate) of the loans, which we do in two steps. First, we create a benchmark model of loan level foreclosure probability based on publicly available information at the time of issuance. Second, we aggregate this at the pool level to compute the expected default rate of the pool. We then take the ratio of actual foreclosure rate we observe ex post to the expected foreclosure rate as the measure of abnormal default.

We estimate a benchmark model of foreclosure probability for every loan based on the following logistic regression model:

$$Pr(foreclosure_i = 1) = \frac{1}{1 + e^{-\beta X_i}} \tag{6}$$

$foreclosure_i$ equals one for loans that enter foreclosure any time up to December 31, 2011. X_i is a set of observable loan and property characteristics that are likely to predict the loan's default rate. They include the borrower's FICO score, LTV ratio, state of the property's location, the purpose of the loan (refinancing versus purchase loans), the year of loan origination, and the type of the loan product (i.e., ARM, balloon or fixed rate loans).[15] We choose these variables based on economic intuition and previous research in the area (see e.g., Demyanyk and Van Hemert, 2011). The estimated default model uses roughly 500,000 loans originated primarily during the years of 2002 to 2005. Of these loans, about 16% enter foreclosure during our sample period. As noted earlier, the foreclosure rates of loans based on observable metrics such as FICO score, LTV ratio, loan interest rate type and time period correlate in the expected directions as shown in Panel B of Table 1.

Since the estimates of the logistic regression are consistent with previous findings in the literature, we do not tabulate them for brevity. Borrowers with higher FICO scores and lower LTV ratios are significantly less likely to get into foreclosure. Loans used for home purchase are significantly less likely to get into foreclosure as compared to refinancing and cash-out loans. All other categorical variables (state, year of origination, and the product type) have significant predictive power in explaining the foreclosure rate as well.

After estimating the model, we use the fitted values from the model to obtain the predicted default likelihood $\widehat{foreclosure}_{ip}$ of each loan i in pool p. The predicted foreclosure rate provides us with an in-sample benchmark for the expected default rate of the loan conditional on key observable characteristics. We aggregate this measure at the pool level to compute a predicted foreclosure rate of the pool. The abnormal default rate for the pool ($AbDefault_p$) with N_p loans in it is then calculated as follows:

$$AbDefault_p = \frac{\sum_{i=1}^{N_p} w_i(foreclosure_{ip})}{\sum_{i=1}^{N_p} w_i(\widehat{foreclosure}_{ip})} \tag{7}$$

[15] The results we present are based on a pooled model using all available data. Our results are unchanged if we compute our benchmark default model separately for the early and late time periods as well as using different performance horizons (e.g., two years, five years, etc.).

Our measure computes the dollar-weighted ratio (weights w_i with $\sum_{i=1}^{N_p} w_i = 1$) of number of loans in a pool that actually defaulted to the number of loans that were expected to default based on observable characteristics at the time of issuance.[16] We plot the kernel density of *AbDefault* measure in Figure 2a. As expected, the average number is centered around one with significant cross-sectional variation. The $75th$ percentile pool has abnormal default ratio of 1.18, indicating that the pool's actual default rate is 18% higher than the expected default rate based on observable characteristics. In contrast, the $25th$ percentile pool has a ratio of 0.47 indicating 53% lower default rate than its benchmark.

We estimate regression equation (5) based on this measure of abnormal default and report the results in column (1) of Table 3. Since signaling should be most useful when there is more opacity about the underlying asset, the key variable of interest is the interaction of the strength of the signal, *HighEq*, and pool opacity, *Opaque* ($\hat{\beta}_3$ from the earlier discussion). Column (1) reports an estimate of $\hat{\beta}_3 = -0.244$, which is significant at 1% level. Opaque pools with a higher level of equity tranche had significantly lower abnormal default rates than those with lower equity tranche, as compared to the corresponding difference for the transparent pools. The difference-in-differences coefficient translates into a lower default rate of 24.4% for higher equity tranche pool. The second estimate of interest is the difference in abnormal default between deals with higher and lower levels of equity tranche within opaque pools, which is $\hat{\beta}_3 + \hat{\beta}_2 = -0.134$ and significant at the 6% level (not tabulated). Thus, pools with higher level of equity tranche have 13% lower abnormal default rate within opaque pools. These results show that equity tranche predicts better future performance conditional of ex-ante loan characteristics.

To see if this effect is driven by some effect of the size of the mezzanine tranche, We re-estimate these models with *HighMezz*, an indicator variable that equals one when the deal has a higher than median size mezzanine tranche, and its interaction with *Opaque* as additional regressors and present the results in column (2). The coefficients on these addi-

[16] Alternatively, we compute our default benchmark measures of abnormal default based on the *number* of loans that enter foreclosure (i.e., equal weighting) and find similar results for our tests.

tional variables are not significant, while the estimates on $HighEq * Opaque$ are strengthened both economically and statistically. This result contrasts the difference between equity and mezzanine tranche and emphasizes the importance of equity tranche in predicting future performance of the entire pool.

Simulation Based Default Benchmark

One of the basic rationales behind the creation of mortgage-backed securities is the benefit of diversification that can be achieved by pooling several loans together. Indeed, a key input to the RMBS pricing models is the underlying correlation matrix of the loans in a pool. Our default risk model in the previous section ignores the within pool correlation of default risk of loans. We now account for this effect as well the effects of macroeconomic shocks through a simulation exercise which we describe below.[17]

For every loan in a given pool, we find a matching loan with similar observable characteristics from the universe of all loans in our sample *excluding* the loans in the loan's own pool. The matched loan is similar on key dimensions of default and interest rate risk such as FICO score, LTV ratio, loan amount, year of origination, type of interest rate on the loan (e.g., ARM, balloon or fixed rate) and geographical location. We outline the precise matching algorithm in the Appendix. The key idea is to match the actual pool created by the informed sponsor to a hypothetical pool that is, by construction, from an uninformed sponsor. Our random selection of loans into the pool gets rid of the sponsor's pool-specific private information, while retaining the observable similarity of the actual and hypothetical pool. Loans in the hypothetical pool are likely to have similar correlation structure as the actual pool, especially since we match these loans based on the geographical location as

[17] An alternative approach to parse out the effect of latent macroeconomic shocks is to use a frailty correlated default model. Duffie, Eckner, Horel, and Saita (2009) propose such a model and estimate it for a sample of U.S. nonfinancial firms. They find strong evidence for the presence of common latent factor even after controlling for commonly used firm specific default predictors. In unreported robustness test, we estimate a maximum likelihood based frailty model and obtain similar results. We prefer the simulation based approach for our exercise as it allows us to account for correlation structure of the loans in a relatively straightforward manner.

well. We compute the foreclosure rate of the hypothetical pool as a benchmark for the actual pool's default rate. In essence, we are creating an uninformed benchmark for the actual pool that is created by the informed intermediary. Since the hypothetical pool is observationally similar and the loans in the pool are subjected to similar macroeconomic shocks as the actual pool, the foreclosure rate on hypothetical pool provides us with a benchmark that account for ex-ante loan characteristics, macroeconomic shocks and the correlation structure of the loans in a non-parametric way. As before, we take the ratio of the actual pool's default rate to its matched hypothetical pool's default rate as the measure of abnormal default.

A kernel density of the abnormal default rate based on this measure is provided in Figure 2b. Like our first measure, the average performance is centered around 1 with a large cross-sectional variation. The ratio ranges from 0.67 to 1.21 as we move from the $25th$ to the $75th$ percentile of the distribution.

We estimate regression equation (5) based on this measure of abnormal default and report the results in Column (3) of Table 3. The estimates on our variables of interest largely mirror our findings from our first measure of abnormal default. Column (3) reports an estimate of $\hat{\beta}_3 = -0.221$, which indicates that opaque pools with a higher equity tranche have a 22.1% lower default rate as compared to the corresponding difference for relatively transparent pools. The difference in abnormal default between deals with higher and lower levels of equity tranche within opaque pools, which is $\hat{\beta}_3 + \hat{\beta}_2 = -0.188$ and significant at the 2% level (not tabulated). Thus, pools with higher level of equity tranche have an 18.8% lower abnormal default rate within opaque pools. We re-estimate these models with *HighMezz* and its interaction with *Opaque* as additional regressors and present the results in column (4). We find that the difference-in-differences estimator, as was the case with the default model specification in column (2), is strengthened with $\hat{\beta}_3 = -0.270$. These findings indicate that equity tranche created at the time of security sale forecasts better than expected foreclosure outcomes for loans in the underlying pool. Overall, these findings are consistent with equity tranche being an indicator of favorable private information of the sponsor. But did it act as a credible signal to the security buyers? To answer this question, we look at the pricing effect of the level of equity tranche in the following section.

4.3 Pricing Effect of Equity Tranche

An important prediction of signaling models is the presence of a downward sloping demand curve: as sponsors sell more of their assets, investors demand lower prices (e.g., see DeMarzo and Duffie, 1999). Sponsors trade off the resulting liquidity discount from selling more of their assets with the cost of retaining higher equity tranche. Since pricing data for sold tranches is unavailable, following the prior literature we use yield spread on these securities to test this prediction (Je et al., 2012). It is relatively straightforward to compute yield spread for floating rate coupons – It is estimated as the spread over LIBOR benchmark reported in the deal prospectus. For the fixed rate tranches, we need to know the duration of these securities to be able to compute the benchmark rate. Absent this information, we only focus on floating rate tranches for this part of the analysis. Despite this limitation, we are able to cover about 70% of tranches in our sample.[18]

For this exercise, we divide all tranches into broad credit rating classes: AAA, AA, A, and BBB.[19] For deals with multiple tranches within one rating class, we compute a dollar-weighted average coupon rate and consolidate them into one observation. This aggregation leads to 549 sold tranches in our sample, out of which 384 are floating rate. We break all pools into two categories based on whether they have above or below the median level of equity tranche. Table 4 presents the cross-tabulation of the average coupon rate of sold tranches across high and low equity tranche groups for every credit rating category. There is a clear pattern in the data: within each credit rating class, the coupon rate is lower for pools with higher equity tranche. As sponsors sell more of their pool's cash flows to outside

[18] In a robustness exercise, we include fixed rate tranches as well, and obtain similar results. For this analysis, we subtract the 5-year risk-free treasury rate from the fixed coupon rate of the tranche.

[19] There are a very small number of sold tranches below the BBB rating. We include them in the BBB category.

investors, the price decreases (coupon rate increases).

We estimate a regression model relating yield spread to level of the equity tranche in the deal after controlling for the credit rating fixed effects. The regression coefficients, therefore, estimate the difference in coupon rate across high and low equity tranche deals for the same credit rating class. Columns (1)-(2) of Table 5 present the results. We find a significant negative coefficient on *HighEq*: for the same credit rating class high equity tranche deals have 27 basis points lower coupon rate. More important, the effect comes entirely from the *Opaque* deals. This is precisely the group where we find a considerably lower foreclosure rate in our earlier tests. We further break our analysis down to AAA-rated and non-AAA rated securities and report the results in columns (5)-(6). The effect is concentrated among the non-AAA rated tranche backed by opaque deals. With their higher informational sensitivity, we expect signaling effect to be higher for them and the empirical results confirm this intuition.

Taken together with the abnormal default rate results, our results support the signaling role of equity tranche. A higher level of equity tranche among opaque pools predicts lower foreclosure rates ex-post, and the market prices reflect this ex-ante.

4.4 Information-destruction and risk-diversification

We now turn to the determinants of AAA-rated tranche in RMBS deals based on DeMarzo (2005). As discussed in Section 2, the pooling of loans with diffused private information has a value destroying effect on the sponsors. The risk diversification effect, in contrast, creates an advantage to pooling. Therefore, pools that contain loans with common sources of private information and uncorrelated sources of risks are likely to be the most attractive pools in the securitization market. A key prediction of DeMarzo's (2005) model is that the sponsor will benefit most from securitization in such deals by selling highly rated AAA tranche at attractive price. We test this prediction by analyzing the effect of these pool characteristics on the fraction of AAA-rated tranche sold by the sponsor.

We need reasonable measures of commonality in private information and risk diversification to test this prediction. We propose state level geographical diversification of the pool as a natural measure of the risk diversification benefits. Local economic conditions play a large role in the price movement in the housing market. Thus pools that are backed by loans from different states are more likely to provide risk-diversification benefits as compared to loans that come from one or few states. The pools in our sample span a wide spectrum of geographical diversification, ranging from pools with all loans originated in a single state to pools containing loans from over 40 states. For the average pool, the state that is most well represented comprises about 40% of the pool (which would translate into our measure of geographical diversity, *GeoDiverse*, as 100-40=60). In our empirical tests, we use an indicator variable *HiGeoDiverse* that equals one for pools that have above median geographical diversification and zero otherwise. Our results are robust to other measures of geographical diversification including a Herfindahl index or the continuous variable *GeoDiverse*. We present results based on the indicator variable measure for ease of interpretation.

We create measures of information commonality based on the originator concentration and the concentrations of specific types of loans within the pools. First, we take pools that have a single loan originator (*OneOrig*) as a proxy for commonality in private information. As discussed earlier, all loans in these pools are likely to come from similar screening technology and are therefore likely to share common components of private information. For example, the nature of private information is likely to be more homogeneous for pools where all loans come from either Option One Mortgage or Bank of America, as compared to pools where each originator has 50% share in the pool. For the second measure, we divide loans into four bins based on whether they are single family or not and whether they are owner occupied or not. We then create the Herfindahl index for each pool based on this distribution. Finally, we construct an indicator variable *HiHerf*, which equals one for deals with above median Herfindahl index, and zero otherwise. In line with DeMarzo's theoretical model, our key economic construct is *commonality*, or positive correlation, in the private information across loans, not necessarily the level of private information.

Table 6 presents the tests of this hypothesis. The dependent variable for these specifications is the percent of the deal with a AAA rating. As expected, and consistent with our earlier results, pools backed by loans with better observable credit risk have a higher level of AAA tranche. The coefficients on FICO scores and LTV ratio are both economically and statistically significant. The coefficient of interest in these tests is the interaction of common information (*OneOrig* and *HiHerf*) and risk diversification (*HiGeoDiverse*). Using both proxies of commonality of information, we find that the pools with common information and diffused risk have economically and statistically higher level of AAA-rated tranche. Conditional on credit risk, the estimates of the key interaction term in columns (2) and (4) indicate that these deals get a AAA-rated tranche that is about two percentage points larger.[20] In columns (5)-(6), we use the combination of our two measures of common information and then interact this with *HiGeoDiverse*. There are 31 of these "most desirable deals" in our sample when we combine both measures of informational commonality. The coefficient of 2.73 on *HiGeoDiverse* ∗ *HiHerf* ∗ *OneOrig* in column (6) shows that these deals get an economically and statistically larger portion of AAA-rated tranche. This estimate represents an increase of approximately $21 million in the portion of AAA-rated tranche for the average deal, which has an aggregate pool amount of $775 million. Further, in unreported tests we find that the AAA-rated tranche backed by deals with common private information and higher diversification are priced no differently from the AAA-rated tranches of remaining deals. Said differently, sponsors are able to sell higher fraction of these deals as AAA-rated security without any price discount. Thus the sponsors get higher proceeds from the AAA-rated tranche in these deals after controlling for key observable characteristics.

Overall, the results in Table 6 provide evidence that corroborates the key prediction of DeMarzo (2005) – deals backed by pools with common private information and diversified residual risk have higher AAA-rated tranche. While we are unable to comment on whether

[20]The sample size drops from 163 deals in columns (1)-(2) to 147 deals in columns (3)-(4) because the latter model requires loan level data that comes from CoreLogic dataset.

the level of AAA or other tranches were optimal, our evidence shows that the cross sectional relationship between the pool characteristics and level of these tranches are consistent with the theoretical predictions of information based models.

4.5 Alternative Channels

It has been recognized in the literature that in addition to tranche structure, concerns such as sponsor's reputation, servicing rights, and influence over credit rating agencies can play important roles in the way participants contract in this market. These considerations could potentially interact with the retention of equity tranche, creation of AAA-tranche and other related features of the RMBS design. While we do not explore these interactions in detail, this section presents several tests to establish the robustness of our analysis even in the presence of these competing influences.

We first consider the possibility that our results are driven by deals where sponsors and originators have more "skin in the game" by holding servicing contracts.[21] In addition to earning fees from the origination of loans, lenders sometimes retain servicing rights on loans that provide them with an additional stream of income for the life of the loan. This income averages about 37 basis points per year for the deals in our sample. If the sponsors hold servicing rights on the loans, this implicit equity stake may provide stronger incentives for them to ensure that the pool is populated with higher quality loans. If deals with higher servicing "skin in the game" coincide with those with higher equity tranche, then our inferences maybe contaminated. To empirically separate out this alternative channel, we collect data on the identity of primary servicer for the loans in the pool. We create a dummy variable that indicates if the sponsor is also the servicer (*SellAndService*) and a dummy that indicates if the top originator for the pool is also the servicer (*TopOrigAndService*).[22]

[21]See Piskorski, Seru, and Vig (2010) for a detailed discussion of several interesting incentives issues arising out of servicing rights.

[22]We perform the same tests using a dummy variable that indicates if the servicer is any of the top four originators and get qualitatively identical results.

Another mechanism that can potentially confound our results is the reputational concerns of the members of the syndicate. If the originator is also the sponsor, who often has its name in the title of the securities, then poorly performing loans and thus poorly performing securities may damage its reputation (see also Demiroglu and James (2012) and Hartman-Glaser (2012)). Thus, we hypothesize that deals for which the same entity is the top originator and the sponsor should have relatively higher reputational concerns. To empirically investigate this effect, we create a dummy variable that indicates if the sponsor is also the top originator (*SellAndTopOrig*) in the pool.

Additionally, we consider the heterogeneity in the sponsor-type to control for the reputational concerns. For example, one might expect that long-lived and established commercial banks (e.g., JP Morgan) have different concerns about protecting their franchise values than specialized mortgage originating institutions such as Ameriquest. Also, large commercial and investment banks may be able to exert more influence over the credit rating agencies to receive inflated ratings relative to smaller purely mortgage lenders (Je et al., 2012). To address these issues, we include institution-type fixed effects in our specification. We classify each sponsor as a commercial bank, investment bank, savings and loan institution, or mortgage lender and then include dummy variables for these categories in the specification.

We re-estimate our key regression models after including each one of these variables in the model. Table 7 reproduces the main results from earlier sections of the paper alongside a specification that includes the variables mentioned above. All our key results remain qualitatively similar. Among the additional control variables, we do find some effect consistent with servicer's "skin in the game" hypothesis. However, inclusion of this control variable does not change any of our results.

5 Discussions and Conclusion

This paper empirically examines the design of residential mortgage-backed securities during the period before the sub-prime mortgage crisis. These securities are created by pooling multiple loans in one basket and then issuing collateralized securities against the pooled cash flows. In addition to the general problem of adverse selection faced by uninformed buyers, pooling and tranching presents unique tradeoffs that are unexplored by the prior empirical literature.

We document that RMBS sponsors use the equity tranche as a signal of their private information in opaque pools. Within such pools, those with higher levels of equity tranche experience significantly lower future foreclosure rates as compared to benchmark models that account for ex-ante loan characteristics, macroeconomic shocks and the correlation structure of loans in the pool. Further, investors paid higher prices for sold securities in such deals. These pieces of evidence provide support for the basic signaling models such as Leland and Pyle (1977) and DeMarzo and Duffie (1999). We extend our analysis to test predictions of a model specific to the pooling and tranching of information sensitive assets. Consistent with the theoretical work of DeMarzo (2005), we find that the sponsors are able to sell a significantly higher proportion of deals as AAA-rated when the underlying information asymmetry about the loans is common, but the loan-specific risk is relatively uncorrelated.

Overall, our findings show that market participants understood informational frictions in the RMBS market to some extent and incorporated them in the design of these securities. In other words, the design of mortgage-backed securities was able to mitigate some of the contracting frictions as predicted by extant theoretical models in the literature. Our study is focused on understanding the drivers of the cross-sectional variation in the construction of securities in the RMBS securitization market. Therefore, we are only able to comment on the ability of these models in explaining outcomes in a relative sense. It may, for example, be true that the level of equity tranche supporting these deals was too low even though the relative distribution across the deals is explained well by concerns about adverse selection. Indeed, Stanton and Wallace (2011) show that in the period leading up to the crisis, the rating agencies allowed subordination levels in CMBS markets to fall to suboptimal levels. The key contribution of our paper is to show that cross-sectional pattern in securitization design does follow the predictions of signaling models as well models more specific to the RMBS market that incorporate the tradeoff of information-destruction and risk-diversification. This finding has important implications for the development of future theoretical models in this area as well as for informing policy debates surrounding this market.

References

Viral Acharya, Matthew Richardson, et al. *Restoring Financial Stability: How to Repair a Failed System*, volume 542. Wiley, 2009.

Viral V. Acharya, Philipp Schnabl, and Gustavo Suarez. Securitization without risk transfer. *Journal of Financial Economics*, 107(3):515 – 536, 2013.

Philippe Aghion and Patrick Bolton. An incomplete contracts approach to financial contracting. *The review of economic Studies*, 59(3):473–494, 1992.

Franklin Allen and Douglas Gale. Optimal security design. *Review of Financial Studies*, 1 (3):229–263, 1988.

Xudong An, Yongheng Deng, and Stuart A Gabriel. Asymmetric information, adverse selection, and the pricing of cmbs. *Journal of Financial Economics*, 100(2):304–325, 2011.

Adam Ashcraft, Paul Goldsmith-Pinkham, and James Vickery. *MBS Ratings and the Mortgage Credit Boom*. DIANE Publishing, 2010.

Adam B Ashcraft and Til Schuermann. *Understanding the securitization of subprime mortgage credit*. Now Pub, 2008.

Bo Becker and Todd Milbourn. How did increased competition affect credit ratings? *Journal of Financial Economics*, 101(3):493–514, 2011.

Efraim Benmelech and Jennifer Dlugosz. The alchemy of CDO credit ratings. *Journal of Monetary Economics*, 56(5):617–634, 2009.

Efraim Benmelech, Jennifer Dlugosz, and Victoria Ivashina. Securitization without adverse selection: The case of clos. *Journal of Financial Economics*, 2012.

Dion Bongaerts, KJ Cremers, and William N Goetzmann. Tiebreaker: Certification and multiple credit ratings. *The Journal of Finance*, 67(1):113–152, 2012.

Arnoud WA Boot and Anjan V Thakor. Security design. *The Journal of Finance*, 48(4): 1349–1378, 2012.

Jess Cornaggia and Kimberly Cornaggia. Estimating the cost of issuer-paid credit ratings. *Review of Financial Studies*, forthcoming.

Peter DeMarzo and Darrell Duffie. A liquidity-based model of security design. *Econometrica*, 67(1):65–99, 1999.

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

Cem Demiroglu and Christopher James. 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, 2012.

Yuliya Demyanyk and Otto Van Hemert. Understanding the subprime mortgage crisis. *Review of Financial Studies*, 24(6):1848–1880, 2011.

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

Darrell Duffie, Andreas Eckner, Guillaume Horel, and Leandro Saita. Frailty correlated default. *The Journal of Finance*, 64(5):2089–2123, 2009.

Douglas Gale and Martin Hellwig. Incentive-compatible debt contracts: The one-period problem. *The Review of Economic Studies*, 52(4):647–663, 1985.

Gary Gorton and Andrew Metrick. Securitization. Working Paper 18611, National Bureau of Economic Research, December 2012.

Gary Gorton and George Pennacchi. Financial intermediaries and liquidity creation. *Journal of Finance*, pages 49–71, 1990.

John M Griffin and Dragon Yongjun Tang. Did subjectivity play a role in CDO credit ratings? *The Journal of Finance*, 67(4):1293–1328, 2012.

Barney Hartman-Glaser. Reputation and signaling. *Working paper*, 2012.

Barney Hartman-Glaser, Tomasz Piskorski, and Alexei Tchistyi. Optimal securitization with moral hazard. *Journal of Financial Economics*, 104(1):186–202, 2011.

Jie Je, Jun Qian, and Philip Strahan. Are all ratings created equal? the impact of issuer size on the pricing of mortgage-backed-securities. *Journal of Finance*, 2012.

Benjamin Keys, Tomasz Piskorski, Amit Seru, and Vikrant Vig. Mortgage financing in the housing boom and bust. In *Housing and the Financial Crisis*. University of Chicago Press, 2012.

Benjamin J Keys, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. Did securitization lead to lax screening? evidence from subprime loans. *The Quarterly Journal of Economics*, 125(1):307–362, 2010.

Hayne Ellis Leland and David H Pyle. Informational asymmetries, financial structure, and financial intermediation. *Journal of Finance*, 32(2):371–87, 1977.

Elena Loutskina and Philip E Strahan. Informed and uninformed investment in housing: The downside of diversification. *Review of Financial Studies*, 24(5):1447–1480, 2011.

Atif Mian and Amir Sufi. The consequences of mortgage credit expansion: Evidence from the us mortgage default crisis. *The Quarterly Journal of Economics*, 124(4):1449–1496, 2009.

Taylor D Nadauld and Michael S Weisbach. Did securitization affect the cost of corporate debt? *Journal of financial economics*, 105(2):332–352, 2012.

Tomasz Piskorski, Amit Seru, and Vikrant Vig. Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis. *Journal of Financial Economics*, 97 (3):369–397, 2010.

Amiyatosh Purnanandam. Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies*, 24(6):1881–1915, 2011.

Uday Rajan, Amit Seru, and Vikrant Vig. The failure of models that predict failure: distance, incentives and defaults. *Journal of Financial Economics*, forthcoming.

Timothy J Riddiough. Optimal design and governance of asset-backed securities. *Journal of Financial Intermediation*, 6(2):121–152, 1997.

David Scharfstein and Adi Sunderam. The economics of housing finance reform. In Martin Neil Baily, editor, *The Future of Housing Finance*, pages 146–198. Brookings Institution Press, 2011.

Richard Stanton and Nancy Wallace. CMBS subordination, ratings inflation, and regulatory-capital arbitrage. *Working paper, UC Berkeley*, 2011.

Robert Townsend. Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293, 1979.

Appendix

Appendix 1: Sample Construction and Data Collection We use a stratified random sampling method to select private label (i.e., non-agency backed) RMBS deals for inclusion in our study. We choose two time periods for our sample selection: an "early period" that covers deals from 2001-02 and a "late period" that covers deals from 2005. This stratification strategy allows us to separate out time-specific effects from our main cross-sectional results. It also allows us to investigate the time variation in the functioning of this market. Ashcraft and Schuermann (2008) report that the issuance of non-agency mortgage-backed securities increased eight-fold from $99 billion in 2001 to $797 billion in 2005 in the sub-prime and Alt-A segment. Thus our sample covers both an early/nascent period and a relatively matured period of RMBS market. We also stratify the sample along the prime-subprime dimension, slightly over-sampling the subprime pools to make sure that portion of the sample is large enough to make statistically meaningful inference. Since there is no clear delineation between prime and subprime deals, we use the weighted average FICO score of 660 as the cutoff between the two types of deals. Our random sample begins with 196 deals. Due to variation in the data items included in the filings, our main regression specifications include 163 deals that have full data on all variables of interest.

We collect data on mortgage pools and their tranches from Form 424(b)(5) filings which are submitted to the SEC pursuant to SEC Rule 424(b)(5). While the detail of the information provided varies slightly from deal to deal, the form typically contains data on all the major participants in the deal (e.g., sponsor, originators), pool level characteristics and tranche/security level data. Among other items, these data specifically include the loan originators and the share of the deal they originated, weighted average loan-to-value (LTV) ratio, weighted average FICO score, and a breakdown of loan types, geography and loan documentation levels within the pool. A complete list of the variables we use and their definition is presented in the Appendix.

Form 424(b)(5) also provides a listing of each tranche in the pool along with its principal amount and credit rating. For our analysis, we aggregate the tranches into three bins: AAA-rated tranches, mezzanine tranches and equity tranches. We define the equity tranche as the piece that is not offered to the public and is not rated by the credit rating agencies. The AAA tranche is self explanatory and the mezzanine tranche is simply the subordinated tranche that lies between the AAA and equity tranches. The publicly offered tranches (AAA and mezzanine) include ratings from at least two major credit rating agencies. While disagreements in ratings among the ratings agencies are not common and rare for the senior tranches, we use the lower of the ratings when conflicts occur.

We match these deals with detailed loan-level data obtained from CoreLogic. Pools in our sample cover over half a million individual mortgages. We obtain key information for each loan in a given pool from CoreLogic. These data include the loan amount, FICO score, LTV ratio, and loan type along with location of the property and various other characteristics. Finally, we obtain the ex-post performance of these loans from CoreLogic as well. We obtain information on the incidence of foreclosure anytime from the origination of the deal till December 2011. This information allows us to conduct our test relating tranche structure to ex-post loan performance. Our sample size drops slightly to 151 deals for which we are able to match our pool level data with CoreLogic database. These deals cover 501,131 underlying loans in total.

Appendix 2: Simulated Pool Construction

We construct a hypothetical pool of loans that look observationally similar to loans in actual pools. As described in Section 4.2, our goal is to create a random pool of loans that is likely to have similar foreclosure performance as the actual pool in terms of observable loan and property characteristics, macroeconomic shocks, and correlation structure of loans with the pool. For every loan i in pool p, we start with all other loans in our sample, excluding the pool where loan i resides, and follow the following matching algorithm:

1. Drop potential matches that were not originated in the same (early or late) as loan i.

2. Drop potential matches that are not sufficiently close to loan i in terms of two most important observable characteristics of this market: FICO scores and LTV ratio. We ensure that potential control loans are within one-tenth of the standard deviation of FICO and LTV of the loan being matched. This criteria ensures that LTV ratio of matched firms fall within 1.4 percentage points and FICO score within 11.2 points of loan i.

3. Drop potential matches that are not located in the same state as loan i.

4. We break all loans into three groups based on the nature of interest rate: fixed rate loans, ARM, and Balloons. Drop potential matches that do not have the same interest rate type as loan i.

5. Drop potential matches that are not within 25% of the principal loan amount of loan i.

6. Drop potential matches that whose origination date is not within ± 90 days of loan i.

7. From the remaining set of potential matches, assign the loan with LTV ratio closest to loan i as the matched loan.

We repeat this exercise for all loans in a pool. We are able to obtain matches for 401,228 loans based on this criteria. This leaves us with approximately 100,000 loans that remains unmatched after the first iteration. For loans without a match, we continue as follows:

8. Return to Step (2) above, but drop the requirement that the matched loan be within 1.4 percentage points of loan i in terms of LTV ratio.

This iteration yields another 101,963 matches and almost completes the matching. For a very small number of loans (19,079) that remain unmatched, we continue as follows:

9. Return to Step (2), dropping the LTV caliper requirement as in Step (8), and widen the range of FICO scores to be within one-fifth of the standard deviation and allow the loan origination date to be within ± 180 days of that of loan i.

With less than 4% of loans matched based on the looser criteria of Step (9), our results do not change if we drop these loans altogether from the sample. Based on this matching procedure, we are able to create a hypothetical pool that has loans with extremely similar characteristics on observable dimensions (with exact matches for state, loan type, and early/late period). In Figure A.1, we plot the kernel density of FICO and LTV ratios across actual and simulated pools to illustrate how similar the matched pools are on these dimension.

Table A.1: **Variable Definitions**

This table presents definitions of variables collected and constructed from CoreLogic and SEC Form 424(b)(5) filings.

PrincipalPoolAmount	Total principal outstanding balance of the mortgage pool (in millions of dollars).
NumLoans	Total number of loans in the mortgage pool.
LTV	Weighted average loan-to-value ratio for the pool.
FICO	Weighted average FICO credit score for the pool.
% ARM	Percent of the mortgage pool composed by adjustable rate mortgages.
Single Family Residence	Dummy equal to one if the loan is for a single family residence (as opposed to multiple family residence, condominium, etc).
Owner Occupied	Dummy equal to one if the loan is for a residence that is to be the primary residence of the borrower (as opposed to a second home or investment property).
Foreclosure	Weighted average foreclosure rate for the pool.
%NoDoc	Percent of the mortgage pool with no documentation as indicated in the deal prospectus. Each deal was carefully examined to confirm this figure as different originators use slightly different terminology for their documentation classifications.
GeoDiverse	100 - percent of largest one state origination concentration of the mortgage pool.
Late	Dummy equal to one if the deal is issued in 2005 (0 if issued in 2001-02).
Subprime	Dummy equal to one if the weighted average credit score is less than 660.
%AAA Tranche	Percent of the tranches (by dollar amount) within a pool that rated AAA.
%Mezzanine Tranche	Percent of the tranches (by dollar amount) within a pool that are subordinate to the senior tranches and publicly offered.
%Equity Tranche	Percent of the tranches (by dollar amount) within a pool listed as not publicly offered.
Mezzanine-to-Sold	The ratio of principal dollar amount of the mezzanine tranche to the total principal dollars amount publicly offered (mezzanine plus AAA).

Table A.2: **Institutions and their Various Roles**

This table presents the most common institutions in the sample and the frequency in which they participated in various roles.

Institution	Seller	Top Originator	Type
Ace	5	0	Mortgage Lender
Ameriquest	14	15	Mortgage Lender
Bear Stearns	17	0	Investment Bank
Bank of America	28	23	Commercial Bank
Citi	8	4	Commercial Bank
Credit Suisse	16	10	Investment Bank
Countrywide	6	10	Savings and Loan
Deutsche Bank	5	0	Commercial Bank
Goldman Sachs	16	0	Investment Bank
HSBC	3	0	Commercial Bank
IndyMac	10	11	Savings and Loan
JP Morgan	9	5	Commercial Bank
Lehman Brothers	6	4	Investment Bank
Merrill Lynch	8	1	Investment Bank
Option One	8	13	Mortgage Lender
Stanwich	3	0	Mortgage Lender
UBS	6	0	Commercial Bank
Washington Mutual	11	14	Savings and Loan
Wells Fargo	12	24	Commercial Bank
Other	5	62	

Figure A.1: **Actual Pools and their Match**

This figure presents kernel densities of the weighted average FICO scores and loan-to-value (LTV) ratios of pools in the sample along with the kernel densities of matched pools to illustrate the comparability of the two for these two observable primary drivers of credit risk.

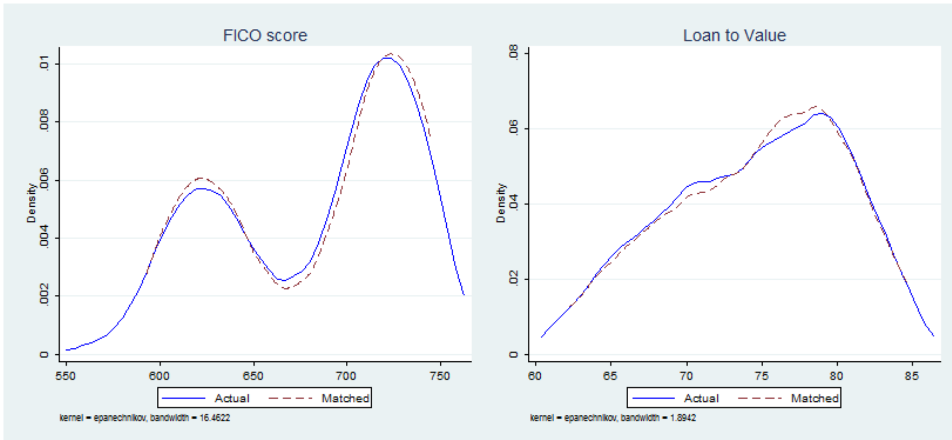


Table 1: **Full Sample Summary Statistics**

This table presents summary statistics for our sample. Panel A presents various loan level, pool level, and tranching structure characteristics, Panel B presents ex-post foreclosure rates, divided by various loan characteristics and Panel C presents the tranche structure of deals in our sample across time periods. The variables are defined in the appendix.

Panel A: Loan, Pool, and Tranche Structure Summary Statistics

	Mean	Std Dev	Min	25%	50%	75%	Max	N
Loan Level:								
Loan Amount	259781.43	206639.42	3150.00	110000.00	196000.00	365000.00	4350000.00	501131
FICO	656.26	76.92	496.00	599.00	657.00	716.00	799.00	501131
LTV	77.27	13.59	31.25	71.93	80.00	85.00	100.00	501131
ARM	0.66	0.47	0.00	0.00	1.00	1.00	1.00	501126
Single Family Residence	0.76	0.42	0.00	1.00	1.00	1.00	1.00	501131
Owner Occupied	0.89	0.30	0.00	1.00	1.00	1.00	1.00	501131
Foreclosure	0.16	0.37	0.00	0.00	0.00	0.00	100.00	501131
Pool Level:								
PrincipalPoolAmount (mil)	775.85	507.28	151.84	422.34	664.12	1000.08	3267.41	196
NumLoans	3150.46	2535.52	340.00	1343.50	2269.00	4409.75	12202.00	196
% NoDoc	18.77	17.84	0.00	2.94	14.34	34.68	79.13	172
GeoDiverse	59.47	17.26	0.00	49.48	61.31	74.15	87.54	196
OneOrig	0.47	0.50	0.00	0.00	0.00	1.00	1.00	196
Late	0.52	0.50	0.00	0.00	1.00	1.00	1.00	196
Subprime	0.36	0.48	0.00	0.00	0.00	1.00	1.00	194
Foreclosure (dollar weighted)	0.12	0.10	0.00	0.03	0.10	0.18	0.41	152
Tranche Structure:								
% AAA Tranche	90.40	7.17	72.40	82.80	93.52	96.51	98.75	196
% Mezzanine Tranche	8.40	6.70	0.00	2.77	5.48	15.67	27.60	196
% EquityTranche	1.20	1.27	0.00	0.50	0.75	1.70	7.43	196
Mezzanine-to-Sold	8.57	6.83	0	2.80	5.49	15.86	21.99	196

Panel B: Ex-post Default Probabilities Across Risk Factors

	No – mean/[loan count]	Yes – mean/[loan count]
Above median FICO	0.22	0.11
	[251,350]	[249,781]
Above median LTV	0.15	0.20
	[346,616]	[154,515]
Fixed-rate Mortgage	0.19	0.11
	[353,342]	[147,789]
Late period (2005)	0.09	0.19
	[135,474]	[365,657]

Panel C: Tranche Structure Across Time

Piece	All	Early (2001-02)	Late (2005)
AAA	90.36	92.59	88.32
Mezzanine	8.44	6.69	10.05
Equity	1.20	0.72	1.63
Observations	196	94	102

Table 2: **Cross Sectional Determinants of Deal Structure**
This table presents OLS estimates from regressions of *%Equity Tranche* (models (1)-(3)) and *Mezzanine-to-Sold* (models (4)-(6)) on loan pool characteristics. *%Equity Tranche* is the percent of the principal pool amount that is not publicly offered, *Mezzanine-to-Sold* is a computed as the ratio of principal dollar amount of the mezzanine tranche to the total principal dollars amount publicly offered (mezzanine plus AAA), *Late* is a dummy variable equal to 1 for deals from 2005, *% NoDoc* is the percent of the loan pool with no documentation loans, *FICO* is pool's the weighted average FICO score, *LTV* is pool's the weighted average loan-to-value ratio, *% ARM* is the percent of the loan pool with adjustable rate mortgage loans, and *GeoDiverse* measures the geographic diversity and is 100 - (percent of largest one state origination concentration) in the mortgage pool. All standard errors are heteroskedasticity robust.

	%Equity			Mezzanine-to-Sold		
	(1)	(2)	(3)	(4)	(5)	(6)
Late	0.886***	0.943***	0.645**	2.751***	3.181***	3.757***
	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
% NoDoc	0.025***	0.023***	0.024***	0.200***	-0.007	-0.012
	(0.00)	(0.01)	(0.01)	(0.00)	(0.55)	(0.34)
FICO		-0.004	-0.006**		-0.101***	-0.097***
		(0.16)	(0.04)		(0.00)	(0.00)
LTV		-0.025	-0.011		0.301***	0.281***
		(0.23)	(0.64)		(0.00)	(0.00)
% ARM		0.005**	0.003		-0.015***	-0.012*
		(0.01)	(0.11)		(0.01)	(0.06)
GeoDiverse		-0.008	-0.005		-0.054***	-0.060***
		(0.27)	(0.45)		(0.00)	(0.00)
Mortgage Rate			-0.212			0.367
			(0.18)			(0.45)
Constant	0.295***	5.268*	7.293**	3.847***	58.015***	54.463***
	(0.01)	(0.08)	(0.02)	(0.00)	(0.00)	(0.00)
Observations	163	163	160	163	163	160
R^2	0.268	0.318	0.322	0.334	0.857	0.859

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: **Ex-Post Outcomes: Abnormal Default**

This table presents OLS estimates from regressions of *AbDefault* on loan pool characteristics. In models (1) and (2), *AbDefault* is the ratio of the actual ex-post pool default rate to a predicted default rate based on a default model calibrated using the full sample. In models (3) and (4), *Abnormal Default* is the ratio of the actual ex-post pool default rate to the default rate on a pool of loans that are matched, loan by loan, to the actual pool based on observable characteristics. *Opaque* is a dummy variable equal to 1 for deals with *%NoDoc* greater than that of the median deal, *HighEq* is a dummy variable equal to 1 for deals with *%Equity Tranche* greater than that of the median deal, *HighMezz* is a dummy variable equal to 1 for deals with *%Mezzanine Tranche* greater than that of the median deal, *Late* is a dummy variable equal to 1 for deals from 2005, *FICO* is pool's the weighted average FICO score, and *LTV* is pool's the weighted average loan-to-value ratio. All standard errors are heteroskedasticity robust.

	Default Model		Simulation	
	(1)	(2)	(3)	(4)
Late	0.321***	0.293***	-0.006	0.006
	(0.00)	(0.00)	(0.95)	(0.94)
FICO	0.001	0.002*	0.002**	0.002*
	(0.38)	(0.08)	(0.05)	(0.06)
LTV	0.052***	0.051***	0.069***	0.070***
	(0.00)	(0.00)	(0.00)	(0.00)
Opaque	0.163*	0.109	0.099	0.045
	(0.06)	(0.26)	(0.49)	(0.78)
HighEq	0.110	0.139*	0.033	0.067
	(0.13)	(0.06)	(0.72)	(0.46)
HighEq * Opaque	-0.244**	-0.263***	-0.221*	-0.270**
	(0.01)	(0.01)	(0.07)	(0.02)
HighMezz		0.080		-0.112
		(0.51)		(0.32)
HighMezz * Opaque		0.115		0.161
		(0.31)		(0.22)
Constant	-3.783***	-4.477***	-5.588***	-5.649***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	151	151	151	151
R^2	0.650	0.659	0.440	0.445

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: **Equity Tranche and Coupon Rates Cross-tabulation**

This table presents the mean coupon rate spread over LIBOR for variable rate tranches in the sample according to the size of the equity tranche as a percentage of total pool size and the tranche's rating class. There is a unit of observation for each deal-rating class combination. For deals with multiple tranches within a rating class, the observation is the dollar-weighted average of the coupons. *High Equity* indicates that the pool under consideration has *%Equity Tranche* greater than that of the median deal.

	Tranche Rating			
Equity Tranche Size	AAA	AA	A	$\leq$ BBB
Low Equity	0.44	1.27	1.50	2.43
	(0.06)	(0.30)	(0.25)	(0.22)
High Equity	0.34	0.77	1.25	2.24
	(0.03)	(0.10)	(0.11)	(0.12)

Standard errors in parentheses

Table 5: **Price Response (Coupon) to Equity Tranche**

This table presents OLS estimates from regressions of the coupon spread (in percentage points) on loan pool characteristics. For each observation represents a *Pool* × *Rating Class* dollar-weighted spread for variable rate tranches, where we define *Rating Class* as *AAA*, *AA*, *A*, and *BBB and below*. *Late* is a dummy variable equal to 1 for deals from 2005, *Opaque* is a dummy variable equal to 1 for deals with *%NoDoc* greater than that of the median deal, and *HighEq* is a dummy variable equal to 1 for deals with *%Equity Tranche* greater than that of the median deal. All standard errors are heteroskedasticity robust.

	(1) All	(2) All	(3) non-AAA	(4) AAA	(5) non-AAA	(6) AAA
Late	-0.49***	-0.58***	-0.62***	-0.24***	-0.76***	-0.25***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
HighEq	-0.27**		-0.38**	-0.09		
	(0.01)		(0.02)	(0.13)		
Opaque		-0.18			-0.35	-0.07
		(0.44)			(0.38)	(0.48)
HighEq * Opaque		-0.34***			-0.46***	-0.09
		(0.00)			(0.00)	(0.34)
HighEq * Not Opaque		-0.08			-0.21	-0.06
		(0.74)			(0.58)	(0.49)
Rating Class FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	379	379	262	117	262	117
R^2	0.43	0.45	0.30	0.15	0.34	0.17

p-values in parentheses

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

Table 6: **Information Destruction and Risk Diversification**

This table presents OLS estimates from regressions of *PercentAAA* on loan pool characteristics. *Late* is a dummy variable equal to 1 for deals from 2005, *FICO* is pool's the weighted average FICO score, *LTV* is pool's the weighted average loan-to-value ratio, *% ARM* is the percent of the loan pool with adjustable rate mortgage loans, *HiGeoDiverse* is a dummy variable equal to 1 for deals with geographic diversity greater than that of the median deal, *HiHerf* is a dummy variable equal to 1 for deals where the product of *%SingleFamilyResidence* and *%OwnerOccupied* is higher than that of the median deal, and *OneOrig* is a dummy variable equal to 1 for deals where all loans are from a common originator. All standard errors are heteroskedasticity robust.

	(1)	(2)	(3)	(4)	(5)	(6)
Late	-3.88***	-3.80***	-3.73***	-3.59***	-3.80***	-3.67***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LTV	-0.28***	-0.26***	-0.30***	-0.28***	-0.30***	-0.25***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
FICO	0.10***	0.10***	0.10***	0.10***	0.10***	0.12***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
% NoDoc	-0.01	-0.00	-0.00	0.00	0.00	0.01
	(0.69)	(0.98)	(0.97)	(0.86)	(0.95)	(0.69)
% ARM	0.01	0.00	0.01	0.01	0.01	0.00
	(0.18)	(0.43)	(0.20)	(0.44)	(0.22)	(0.44)
HiGeoDiverse	1.40***	0.38	1.08**	-0.06	1.11**	0.20
	(0.01)	(0.55)	(0.05)	(0.93)	(0.04)	(0.70)
OneOrig	0.18	-0.64				
	(0.71)	(0.19)				
HiGeoDiverse * OneOrig		2.01**				
		(0.03)				
HiHerf (Common Info)			0.50	-0.42		
			(0.40)	(0.51)		
HiGeoDiverse * HiHerf				2.10**		
				(0.03)		
HiHerf * OneOrig					1.01	0.00
					(0.13)	(1.00)
HiGeoDiverse * HiHerf * OneOrig						2.73**
						(0.02)
Constant	44.00***	40.47***	44.36***	40.63***	41.00***	30.58**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
Observations	163	163	147	147	147	147
R^2	0.83	0.84	0.82	0.83	0.82	0.83

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: **Robustness – Alternate Channels**
This table presents OLS estimates from regressions of *PercentAAA* on loan pool characteristics. *Late* is a dummy variable equal to 1 for deals from 2005, *%NoDoc* is the percent of the loan pool with no documentation loans, *FICO* is pool's the weighted average FICO score, *LTV* is pool's the weighted average loan-to-value ratio, *% ARM* is the percent of the loan pool with adjustable rate mortgage loans, *OneOrig* is a dummy variable equal to 1 for deals where all loans are from a common originator, *GeoDiverse* measures the geographic diversity and is 100 - (percent of largest one state origination concentration) in the mortgage pool, *HiGeoDiverse* is a dummy variable equal to 1 for deals with geographic diversity greater than that of the median deal, *Opaque* is a dummy variable equal to 1 for deals with *%NoDoc* greater than that of the median deal, *HighEq* is a dummy variable equal to 1 for deals with *%Equity Tranche* greater than that of the median deal, and *SellAndService* is a dummy variable equal to 1 for deals where the issuer is also the primary servicer. *TopOrigAndService* is a dummy variable equal to 1 for deals where the top originator in the pool is also the primary servicer. *SellAndTopOrig* is a dummy variable equal to 1 for deals where the issuer is also the top originator in the pool. Institution-Type effects refers to the inclusion of a set of dummy variables that identify sellers as a commercial bank, investment bank, savings and loan or strictly mortgage lender. All standard errors are heteroskedasticity robust.

	%Equity		Mezzanine-to-Sold		Abnormal Default Sim		AAA - DeMarzo	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Late	0.943***	0.848***	3.181***	3.233***	-0.006	-0.059	-3.805***	-3.731***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.95)	(0.45)	(0.00)	(0.00)
%NoDoc	0.023***	0.017**	-0.007	-0.011				
	(0.01)	(0.05)	(0.55)	(0.40)				
FICO	-0.004	0.005	-0.101***	-0.103***	0.002**	0.001	0.103***	0.094***
	(0.23)	(0.17)	(0.00)	(0.00)	(0.05)	(0.28)	(0.00)	(0.00)
LTV	-0.025	0.009	0.301***	0.288***	0.069***	0.063***	-0.264***	-0.281***
	(0.32)	(0.75)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ARM	0.005**	0.006***	-0.015***	-0.016***			0.004	0.003
	(0.01)	(0.01)	(0.01)	(0.01)			(0.43)	(0.57)
GeoDiverse	-0.008	-0.008	-0.054***	-0.058***				
	(0.27)	(0.21)	(0.00)	(0.00)				
OneOrig							-0.644	-0.855*
							(0.20)	(0.09)
HiGeoDiverse							0.379	0.329
							(0.55)	(0.62)
OneOrig * HiGeoDiverse							2.010**	2.156**
							(0.03)	(0.02)
Opaque					0.099	0.124		
					(0.49)	(0.33)		
HighEq					0.033	0.085		
					(0.72)	(0.40)		
HighEq * Opaque					-0.221*	-0.288**		
					(0.07)	(0.02)		
SellAndService		-0.071		-0.747		0.158		0.653
		(0.88)		(0.21)		(0.17)		(0.40)
TopOrigAndService		-0.158		0.180		-0.366***		-0.296
		(0.64)		(0.78)		(0.00)		(0.62)
SellAndTopOrig		-0.639**		0.191		0.103		0.587
		(0.02)		(0.76)		(0.23)		(0.40)
Constant	5.268	-3.040	58.015***	60.322***	-5.588***	-4.651***	40.475***	47.627***
	(0.11)	(0.52)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Institution Type FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	163	163	163	163	151	151	163	163
R^2	0.318	0.518	0.857	0.860	0.440	0.544	0.838	0.844

p-values in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: **Example Deal: Fremont Home Loan Trust Series 2002-1**

This figure provides an example deal from our sample to illustrate the construction of a typical deal and the sources of our data. Loan specific characteristics such as FICO score, loan amount, loan type, LTV, etc. are from CoreLogic. Aggregate deal statistics, including the tranche structuring of the deal, were hand collected from the Form 424(b)(5) filings to the SEC.

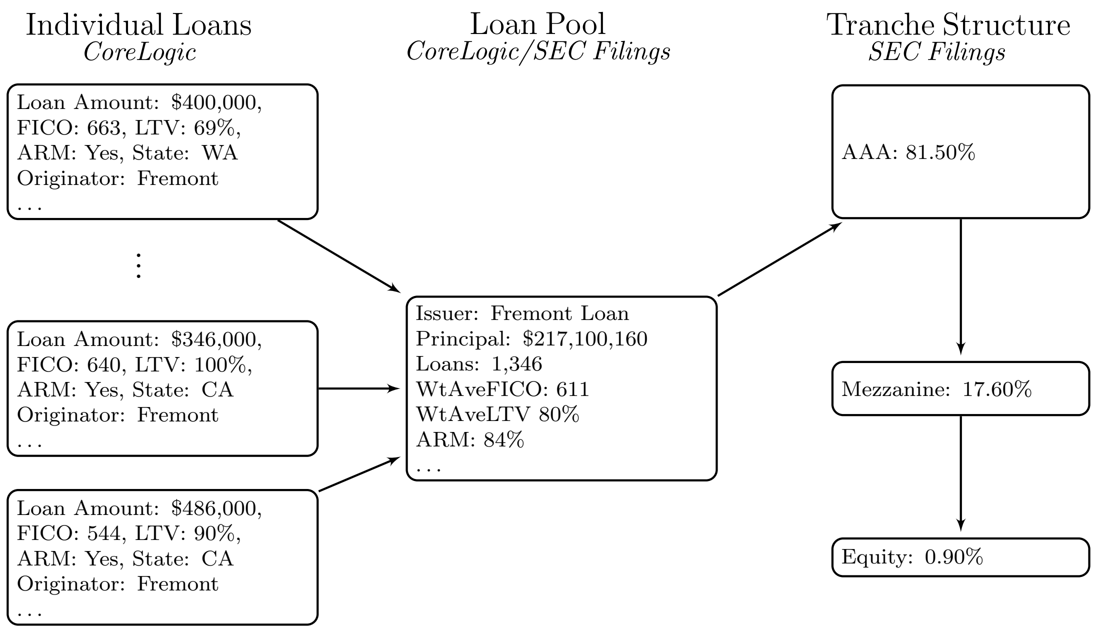
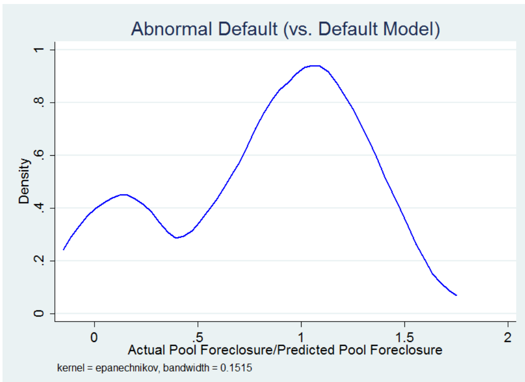
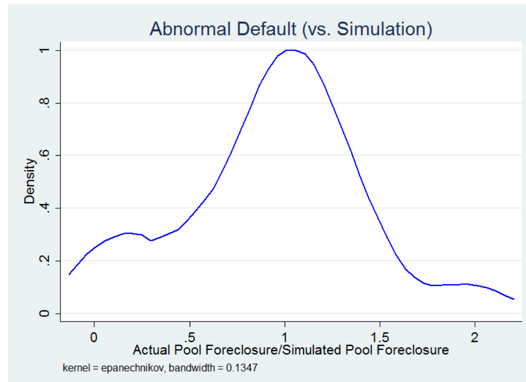


Figure 2: **Measures of Abnormal Default**

This figure presents kernel densities of our measures of abnormal default. Panel 2a presents a kernel density of our first measure of abnormal default which we calculate as the ratio of the actual ex-post pool default rate to a predicted default rate based on a default model calibrated using the full sample. Panel 2b presents our second measure of abnormal default which we calculate as the ratio of the actual ex-post pool default rate to the default rate on a pool of loans that are matched, loan by loan, to the actual pool based on observable characteristics.



(a) Default Model



(b) Simulation