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

Taylor Begley and Amiyatosh Purnanandam*

October 30, 2013

Abstract

Using a representative sample of residential mortgage-backed security (RMBS) deals from the pre-crisis period, we show that deals with a higher level of equity tranche have a significantly lower foreclosure rate that cannot be explained away by the underlying loan pool's observable credit risk factors. The effect is concentrated within pools with a higher likelihood of asymmetric information between deal sponsors and potential buyers of the securities. Further, securities that are sold from high-equity-tranche deals command higher prices conditional on their credit ratings. Our study provides the first in-depth analysis of the effectiveness of the equity tranche in mitigating informational frictions in this market.

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

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 Tim Adam, Michael Barr, Sugato Bhattacharyya, Charles Calomiris, John Griffin, Charlie Hadlock, Chris James, Ravi Jagganathan, Pab Jotikasthira, Larry Harris, Han Kim, Gustavo Manso, Andrey Malenko, Greg Nini, Uday Rajan, Amit Seru, Chester Spatt, James Vickery, Joao Santos, Nancy Wallace, and seminar participants at CAREFIN-Milan, Duke University, Federal Reserve Bank of Philadelphia, Financial Intermediation Research Society, NBER Summer Institute, SFS Cavalcade, the University of Michigan, and the University of Southern California Doctoral Finance Conference for helpful comments on the paper. All remaining errors are our own.

1 Introduction

Securitization provides numerous economic benefits to borrowers and lenders such as more favorable terms of credit for borrowers and better liquidity and risk sharing for lenders. However, each step in the securitization process also introduces potentially costly conflicts of interest.[1] At the root of these frictions is the information asymmetry between different agents along the securitization chain. Understanding the various institutional mechanisms and security design solutions that can overcome these problems and facilitate the functioning of these markets has important implications for both the economic theory underlying securitization markets and ongoing policy debates.[2] However, there is surprisingly little empirical work in this area. To fill this gap in the literature, we examine the role of the equity tranche in residential mortgage-backed security (RMBS) deals in mitigating informational frictions between deal sponsors and investors.

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. RMBS sponsors can convey their private information to potential investors by retaining a larger financial interest in the asset's performance (Leland and Pyle, 1977). Motivated by theoretical models such as Gorton and Pennacchi (1990), Boot and Thakor (1993), Riddiough (1997), DeMarzo and Duffie (1999), DeMarzo (2005), and Hartman-Glaser, Piskorski, and Tchistyi (2011), we analyze three main questions in this paper. First, conditional on observable risk metrics of the underlying pool,

[1]See, for example, Keys, Piskorski, Seru, and Vig (2012), Gorton and Metrick (2012) for recent surveys; 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; and Ashcraft and Schuermann (2008) for a detailed analysis of the securitization process.

[2]For example, issues surrounding the equity tranche of securitization deals form an important part of the Dodd-Frank Reform Act. 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 suggestions for the more broad reform of the housing finance system.

does the size of the equity tranche increase with the degree of information asymmetry between deal sponsors and potential buyers of these securities? Second, conditional on the degree of information asymmetry, do pools with a higher level of equity tranche perform better ex-post as compared to observationally similar pools with lower levels of equity tranche? And third, do security buyers pay higher prices for securities sold in high-equity-tranche deals as compared to a similarly rated security in low-equity-tranche deals?

We carefully assemble a representative sample that comprises about 500,000 loans bundled into 196 private-label RMBS deals from 2001-02 and 2005. Our sample covers a wide cross-section of banks and borrowers. We combine tranche-level security data with the underlying pool characteristics at the time of RMBS issuance, and track the default performance of each loan in these pools through December 2011. This comprehensive information on the loan characteristics of the underlying pool, tranche-level security data, and the ex-post foreclosure status of each loan in the pool allows us to examine the three questions posed above.

We use the percentage of no-documentation loans in a pool as a cross-sectional measure of information asymmetry between the deal sponsors and investors. There is no verification of the borrower's income or assets for these loans, 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 sponsor and the investor.[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 such as FICO score and Loan-to-Value ratio (LTV). This finding is consistent with the key idea that investors are likely to have higher

[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.

adverse selection concerns in relatively opaque deals, which in turn motivates the sponsors to create a larger informationally sensitive first-loss equity tranche (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 main question: does the size of the equity tranche serve as a signal of the sponsor's private information about the pool quality? While we do not, by definition, observe the sponsor's private information at the time of RMBS issuance, 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 parts: (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 relationship between private information at issuance and subsequent loan performance. If the level of equity tranche serves as a signal of the sponsor's private information about the pool, then we should find lower abnormal default (i.e., lower residual component) for loans from high-equity-tranche pools. In contrast, if the level of equity tranche is 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

[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).

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. The difference between the actual pool-level default rate we observe and the expected default rate of the pool from the fitted model is our first measure of abnormal default rate. In the second model, we use our sample of about 500,000 loans to create an observationally similar matched pool for each actual pool. We use a loan-by-loan matching algorithm, which is then aggregated to the pool level, that ensures that the actual and matched pools are similar on dimensions such as FICO scores, LTV ratio, loan product type, and the property location. The default rate of the actual pool over and above its 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 the correlation structure of the pool to the extent that it is driven by geographical diversity. Critically, the matched pool, by construction, lacks the private information component that is present in the actual pool. Thus, the difference in the realized default rate of actual and matched pools' default rates provides our second measure of abnormal default.

We find that deals with higher equity tranche have significantly lower abnormal foreclosure rate, and this effect is concentrated among pools with a higher proportion of no-documentation loans. Said differently, for relatively opaque pools, higher equity tranche predicts better performance in future. In economic terms, pools with above-median level of equity tranche have 24-27% lower foreclosure rates that cannot be explained away by observable credit risk characteristics and macroeconomic conditions. The effect of the equity tranche for relatively transparent pools is statistically indistinguishable from zero. These results are consistent with the idea that sponsors create a larger equity tranche in deals with favorable information on unobservable dimensions.

Are our results driven by asymmetric information concerns or simply a lack of information with all agents in the opaque pools? To answer this question more precisely, we conduct an important sub-sample test based on the originator's affiliation with the sponsors. Prior research suggests that sponsors are more likely to possess private information about the underlying loan pools when they are also the loan originators (e.g., see Keys et al., 2010). Consistent with our asymmetric information argument, we find that the effect of equity tranche on ex-post loan performance is stronger in deals where the sponsors are also the top loan originators, i.e., in deals where the sponsors are likely to have better access to private information.

We provide further evidence in support of private information content of equity tranche by exploiting the passage of Anti-Predatory Lending (APL) laws across several states during our sample period. These laws put stricter requirements on the lenders in terms of their lending practices and disclosure policy which, on the margin, made it more difficult for the lenders to originate poor-quality loans. Such a government regulation should reduce the lemons problem in the market, making the use of private contracting mechanisms less important. Therefore, prior to the passage of this law, the equity tranche is likely to serve as a more important signal of private information for loans originated in APL states. At the same time, the states that do not pass such laws should experience no systematic change in the relationship between the equity tranche and abnormal default rate. Consistent with this idea, we show that loans originated in APL states in the pre-passage period default at disproportionately lower rate if they are backed by higher equity tranche.

In our third test, we study the pricing implications of equity tranche. If a higher level of equity tranche conveys the sponsors' positive private information about the pool, then the market should respond to this signal by paying a higher price for the sold tranches of the deal. To separate out the mechanical leverage effect of a higher level of equity tranche, we condition our analysis on the credit ratings of sold tranches. Thus, we estimate the effect 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, giving further support for the interpretation that the result is not driven by the mechanical leverage effect of equity tranche. In addition, the effect is stronger for the more informationally sensitive non-AAA-rated tranches. Together, these results show that opaque pools with a higher level of equity tranche have lower abnormal default rate ex-post, and ex-ante, they command a higher price. These findings are consistent with the idea that the equity tranche serves as a mechanism to convey the sponsor's private information to potential buyers.

Some have argued that the equity tranche lost its signaling role during the pre-crisis period because deal sponsors are free to sell 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 more 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

[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.

at issuance predicts better future performance beyond what can be explained by observed credit risk factors, and markets reflected this quality by paying a higher price for securities from these deals.

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.[7] Our study is related to Demiroglu and James (2012) who show that linkages between syndicate members, namely the originators and sponsors, can result 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).[8] 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.[9] 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.

[7] 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.

[8] 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.

[9] An originate-to-hold model of lending can be viewed as a limiting case of an RMBS deal where the entire stake is kept by the originating bank. From that perspective, our empirical findings are consistent with the basic idea of this literature: as the sellers stake in the deal increases, the underlying loans perform better in future.

Much of the extant literature focuses on the informativeness of ratings, the optimal subordination level, the effect of syndicate structure on deal performance, and the possibility of rating inflation during the years leading up to the crisis. Our paper is the first to provide an in-depth examination of the role of the equity tranche in mitigating informational frictions between deal sponsors and investors in securitization markets. 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 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 have important policy implications for this market.

We do not attempt to test any specific theoretical model in this paper. Instead, we develop our hypotheses based on the collective insight of theoretical models of security sales in the

[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.

presence of asymmetric information. When an uninformed agent buys financial securities from an informed seller, he faces an adverse selection problem which, in turn, imposes a cost on the informed seller. This problem becomes more severe as the fraction of the asset the seller desires to sell increases. However, by selling a higher fraction of assets to outsiders, sponsors are able to redeploy their capital at attractive rates. Optimizing sellers, therefore, face a trade-off between the benefits from selling a larger fraction of assets with the cost of an adverse selection, or "lemons," discount demanded by the buyer.[11] In equilibrium, sellers retain a fraction of the risky assets to signal the quality of the asset (Leland and Pyle, 1977).

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:

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

$\tilde{I}_{ip}$ 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 of the security sold to outside investors depends on the sponsor's

[11]Financial institutions face considerable regulatory capital charge for retaining equity tranche on their balance sheets. This provides a direct justification for the use of equity tranche as a costly signal.

private information. Conditional on the degree of information asymmetry, sponsors sell a relatively smaller fraction of claims on the pool to outsiders if their private information is positive. 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 to outsiders, outsiders rationally infer the sponsor's private information to be worse and demand a liquidity discount (DeMarzo and Duffie, 1999). This forms the basis of our third test that, after controlling for leverage effects of the equity tranche size, the yield spread at issuance is lower for tranches from deals backed by higher equity tranche.

The securitization of a pool of assets adds additional complexity to this standard lemons-discount model. However these basic predictions apply equally well to the sale of securitized assets. 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. In addition to these key hypotheses, his model also provides some novel predictions specific to the pooling and tranching of securities. We postpone the tests of these specific predictions for future work.

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. We hand-collect the security level data from the SEC filings to ensure that we do not miss any tranche in a specific deal. In addition, we hand-collect several important pieces of information such as the proportion of no-documentation loans in a pool and the identity of key players in the securitization chain from the SEC filings that are not easily available from other sources. Our loan-level data contain information on characteristics such as FICO scores and LTV ratios at the time of the deal as well as each loan's ex-post performance. In particular, we have information on whether the property entered into foreclosure any time 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 use a stratified random sampling method to collect private-label RMBS deals covering a wide cross-section of banks and borrowers. We provide detailed description of sample selection criteria and data collection exercise in the Appendix 1a.

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, originators, and servicers. Our main empirical tests are based on a sample of 163 deals that have all the necessary information needed for the analysis. These deals have approximately 3000 tranches issued against cash flows from approximately 500,000 loans. The sample is approximately equally balanced between early and late periods (defined as 2001-02 and 2005, respectively). 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 an economically meaningful sample of deals from the pre-crisis period. It is worth emphasizing that we draw our sample randomly from the universe of all possible deals. This, in turn, provides confidence in the external validity of our results.

Table 1 presents summary statistics. 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. As expected, 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.

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%. Our 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 Table A.1 the Appendix.

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 borrowers with higher FICO scores, lower LTV ratios, and fixed-rate mortgages default at lower rates. Also, loans from the earlier period are about half as likely to end up in foreclosure, showing strong vintage effect. We now describe the construction of our key variables that measure information asymmetry and the level of the equity tranche.

[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.

No-documentation loans

We obtain the percentage of no-documentation (*NoDoc*) loans in a pool directly from the deal prospectus. No-documentation loans are defined as loans that document neither the income nor the assets of the borrowers. Since different originators label these loans differently, we read through all the deal prospectuses to ensure consistency in our definition across deals. Originators classify these loans under various categories such as "stated documentation," "LITE," and "stated income, stated asset." The prospectus provides further details on the originator-specific underwriting criteria and terminologies, including the details on the various documentation classifications and verification undertaken by the originator. Based on this disclosure, we classify a loan under the no-documentation category if the originator has not verified both the borrower's income and assets. We provide an example of these differences in the classification of *NoDoc* loans in Appendix 1b. As shown in the Appendix, the ABFC Mortgage Loan series has three categories of loans in it: "full documentation loans," "stated income, stated asset loans," and "lite documentation loans." Under the full documentation loans, the lender obtains detailed documentations on information such as borrower's employment status, tax returns for the past two years, and pay-stubs. The originator also performs a telephonic verification of employment for salaried employees. No such attempt for income verification is made under the "stated income, stated asset" program, leaving a great deal of discretion with the originator.[14] We classify these loans under the *NoDoc* category. Finally, under the "LITE" category the originator reviews the deposit activity in the borrower's bank account for the past six to twenty-four consecutive months. We classify these loans as "limited documentation" category, and not "no documentation" in our study. We follow such classification strategy for all deals in our sample. Based on this classification scheme, *NoDoc* loans make up about 19% of all loans in the average pool.

[14]Specifically, the prospectus states, "The applicant's income as stated must be reasonable for the applicant's occupation as determined in the discretion of the loan underwriter; however, such income is not independently verified. Similarly the applicant's assets as stated must be reasonable for the applicant's occupation as determined in the discretion of the loan underwriter; however, such assets are not independently verified."

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.

Equity Tranche

Our main variable of interest is the level of the equity tranche in a deal. We collect this information from the deal prospectuses that provide detailed security-level data on the notional amount of each tranche in the deal, their credit ratings, and the offered yield spread. We combine all tranches that are rated AAA by at least two rating agencies as the AAA-rated tranche. All tranches that are rated below AAA but above the equity tranche are clubbed together into the mezzanine tranche. Equity tranche is defined as the difference between the principal amount of loans in the pool and the sum of AAA and Mezzanine tranche sold to outside investors. In effect, we create a balance sheet of each deal in our sample and take the difference between the dollar value of assets and debt liabilities as the equity tranche. Thus, our definition of equity tranche represents the residual interest of the sponsors, which is precisely in line with the theoretical papers discussed earlier. In practice, sponsors use two different deal structures for tranching: (i) a six-pack structure, and (ii) an overcollateralization (OC) structure (see Gorton, 2010). In the six-pack structure, the junior most tranche is a well-specified unrated tranche that provides protection to all the senior tranches sold to the investors. In such deals the sum of sold tranches and the equity tranche equals the principal amount of loans in the pool. In the OC structure, the principal amount of loans in the pool exceeds the sum of securities on the liability side. The excess amount – the overcollateralization – provides an additional level of residual interest to the sponsor. Economically, the OC amount is the equity interest of the sponsor.[15] Our construction ensures that we capture the true economic interest of the sponsor, regardless of whether it comes in the form of a well-specified security or by having additional residual interest in the pool. It is worth emphasizing that investors are able to observe this measure of quity tranche at the

[15]As noted by Gorton (2010): "The overcollateralization reverts to an equity claim if it remains at the end of the transaction".

time of deal issuance since this information is readily available in the deal-prospectus. We provide an example from each of these structures and the computation of the equity tranche in each case in Table A.2 in the Appendix.

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. The size of the average AAA-rated tranche drops from 92.56% in 2001-02 to 88.32% in 2005. The level of equity tranche more than doubled from 0.72% to 1.63% over the same time period. 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), who include 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 a signal of private information. 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 reasons 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 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 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. Similarly, Demiroglu and James (2012) provide an example from a deal sponsored by Bear Stearns that shows the sponsor's commitment to initially hold the residual interest: *"The initial owner of the Residual Certificates is expected to be Bear Stearns Securities Corp."*

Second, as suggested by the Citi Bank sponsored deal above, 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. We obtain similar evidence from the annual reports of Goldman Sachs and Merrill Lynch during this period (see footnote 5). 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.

Ultimately, the relationship between the level of the equity tranche and 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 find no correlation between the level of the equity tranche and 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. Our main interest lies in estimating the effect of equity tranche on future default performance of the underlying pool. However, for expositional simplicity, we begin our analysis 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. Our final set of tests examine the ex-ante pricing effect of equity tranche.

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:

$$EquityTranche_p = \alpha + \beta(InfoAsym_p) + \theta(Late_p) + \gamma(Credit_p) + \delta(GeoDiverse_p) + \epsilon_p \quad (2)$$

As discussed earlier, 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.[16] 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* variable. 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. The coefficient estimate on *Late* shows that the extent of equity tranche increased in later periods. In column (2),

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

we include all the control variables and find that the estimate on $\%NoDoc$ remains virtually unaffected. In Column (3) we include sponsor fixed effects in the model. This specification ensures that our results are not driven by sponsor's unobserved characteristics such as its reputation in the market. Our results remain robust to this specification. 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), (5) and (6) in Table 2 present the results.

While $\%NoDoc$ has no effect on the division of sold tranches across Mezzanine and AAA category after controlling for observable measures of credit risk in the full specification in column (5), 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 4 and 5). 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 factors that correlate both with $\%NoDoc$ and the extent of equity tranche. Note that after controlling for FICO score, LTV ratio, $\%ARM$, geographical diversity, time effects, and sponsor fixed effects, $\%NoDoc$ does not have any explanatory power in explaining the division of sold tranches between Mezzanine and AAA categories. If we miss a correlated omitted variable from the model that is observed to the investors, then it is likely to influence both the level of equity tranche and the division of sold tranches across Mezzanine and AAA category. In light of our results on Mezzanine-to-Sold tranches, it is unlikely that our results suffer from any serious omitted variable bias. As an additional test (unreported), we include the weighted average interest rate on mortgages in the pool as an explanatory variable in the regression. Interest rates are likely to capture a bulk of the publicly available information about the credit risk of the borrowers. 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 the underlying pool quality. Does the creation of a larger equity tranche indicate deal sponsors' favorable private information about the underlying loans in the pool? Are these effects mainly concentrated in pools with higher concerns about asymmetric information? We exploit the cross-sectional variation in equity tranche along with data on ex-post performance of mortgages to answer these questions. If sponsors with favorable private information about the underlying pool create a larger equity tranche, then we expect to observe 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 matched pool exercise to create two benchmarks of 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 want to estimate the relationship between the equity tranche and abnormal default rate, conditional on the degree of asymmetric information concerns. We do so by estimating the following empirical model:

$$AbDefault_p = \beta_0 + \beta_1(Opaque_p) + \beta_2(HighEq_p) + \beta_3(Opaque_p \times HighEq_p) + \sum \gamma X_p + \epsilon_p \quad (3)$$

$AbDefault_p$ is the abnormal default rate of pool p. *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 an above-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 default model described later. Our regression model uses a difference-in-differences design to estimate the effect of the equity tranche across opaque and transparent pools. Consistent with theoretical models, this empirical approach estimates the relationship between the level of the equity tranche and abnormal default, conditional on the degree of the buyer's asymmetric information concerns. That is, the model allows us to separately examine the relationship of interest for both high-information-asymmetry pools (*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. 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 pools. 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.

Standard Default Model

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 as follows. We create a benchmark model of loan level foreclosure probability based on publicly available information at the time of issuance. Next, 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:

$$Pr(foreclosure_i = 1) = \frac{1}{1 + e^{-\beta X_i}} \tag{4}$$

$foreclosure_i$ 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. We choose these variables based on economic intuition and previous research in the area (see e.g., Demyanyk and Van Hemert, 2011). They include the borrower's FICO score, LTV ratio, state of the property's location, the purpose of the loan, the year of loan origination, and the type of the loan product. FICO score and LTV ratio are the main drivers of the loan's default risk. The property location and the year of origination control for geographical and temporal variations in the house price appreciation and credit risk of the borrowers. In addition, we include a rich set of control variables based on the purpose of the loan and the type of mortgage product. CoreLogic classifies loans into six buckets based on their purpose. The six categories are: (i) purchase, (ii) refinancing without cash out, (iii) refinancing with cash out, (iv) the remaining refinancing loans (i.e., loans with no information on cash out), (v) second mortgage, and (v) others. Finally, we include the nature of interest rate offered on the loan and its reset terms as additional explanatory variables. Based on CoreLogic data, loans are divided into categories based on the following broad criteria: (i) the nature of interest rate (Fixed Rate, ARM, or Balloon), (ii) the number of years for the reset of loan terms (such as 5 year ARM or 7 year Balloon), and (iii) the special structure of the interest or principal payments (such as interest only loans for the first n number of years). Different combinations of these attributes result in 46 distinct categories of product type, and we include them as fixed effects in the model. In addition to default risk, these variables also control for the cross-sectional differences in the loan's prepayment risk in our sample.

The estimated default model uses roughly 500,000 loans originated primarily during the years of 2001-2005.[17] 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.

The estimates of the logistic regression are consistent with previous findings in the lit-

[17]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.).

erature (see Appendix 3 for the estimates). Borrowers with higher FICO scores and lower LTV ratios are significantly less likely to get into foreclosure. Loans with features such as ARM and Balloon payments are significantly more likely to default. All other categorical variables (state, year of origination, and the product type) have significant predictive power in explaining the foreclosure rate as well.[18]

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:

$$AbDefault_p = \frac{\sum_{i=1}^{N_p} w_i(foreclosure_{ip})}{\sum_{i=1}^{N_p} w_i(\widehat{foreclosure}_{ip})} \tag{5}$$

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.[19] 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

[18] We consider several alternative models of default prediction in unreported robustness tests. In particular, we include additional variables such as: (a) the second LTV ratio of the loan (if any), (b) whether the loan is for single-family housing or other assets such as condominiums or manufactured housing, (c) whether the loan has a negative amortization feature, (d) the debt-to-income ratio of the borrower, and (e) the margin on the interest rate. Our results remain robust to these alternative specifications. Some of the additional variables are not available for all loans in the sample. Thus, when we include these co-variates in the model, we lose some observations. More important, addition of these variables do not considerably improve the fit of our model as compared to the base case analysis that captures all the key drivers of default. Therefore, we only report results based on the base case model to save space.

[19] 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.

ratio of 0.47 indicating 53% lower default rate than its benchmark.

We estimate regression equation (3) based on this measure of abnormal default and report the results in column (1) of Table 3. Since the signaling value of the equity tranche 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 have 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.

In Column (2) of the Table we show that these results are driven by the level of equity tranche, and not by the level of Mezzanine tranche in the same deal. We do so by including an indicator variable *HighMezz* that equals one for deals with higher than median level of Mezzanine tranche and its interaction with *Opaque* as additional regressors in the model. The coefficients on these additional 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 tranches and emphasizes the importance of equity tranche, not just the level of AAA subordination, in predicting the future performance of the entire pool. Finally, in Column (3) we include the sponsor fixed effects in the model, and show that our results remain practically unchanged, both in statistical and economic terms.

Matched Sample Benchmark Model of Default

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 matching exercise which we describe below.[20]

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 Appendix 2. 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 matched pool lacks the sponsor's pool-specific private information component, while retaining the similarity along observable dimensions. 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 well. 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 accounts for ex-ante loan characteristics, macroeconomic shocks and the correlation structure of the loans in a non-parametric way. In particular, the matched sample approach allows us to difference out the effect of house price appreciation and the prepayment risk of the loans

[20] 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 matching based approach for our exercise as it allows us to account for correlation structure of the loans in a relatively straightforward manner.

on default rates, since these factors are likely to be driven by geographical location of the property, the timing of the loan origination, and the nature of interest rate on the loan.[21] 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 (3) based on this measure of abnormal default and report the results in Column (4) 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 rate between deals with higher and lower levels of equity tranche within opaque pools is (i.e., $\hat{\beta}_3 + \hat{\beta}_2$) -0.188 which is 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 (5). We find that the difference-in-differences estimator is strengthened with $\hat{\beta}_3 = -0.270$. Finally, in Column (6) we include the sponsor fixed effects in the model, and show that our results remain practically unchanged, both in statistical and economic terms. 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.

Our results also highlight an important distinction across pools with varying degree of no-documentation loans. The effect of equity tranche on future loan performance is

[21] In unreported tests, we also consider matching the property location at the zip-code level. Our results remain similar. However, we are unable to find a match in the same zip-code for several loans in our sample. Hence we prefer the state-level matching for our main analysis.

concentrated within the opaque pools, i.e., pools with higher proportion of no documentation loans. Prior literature has argued that soft information is more important for loans with poor documentation (e.g., see Keys et al. (2010) and Demiroglu and James (2012)). Our results extend the literature by showing that in such deals, the equity tranche conveys important information about the underlying loan quality.

4.3 The Channel of Private Information

Where do sponsors get information about the underlying loan quality? As a part of the securitization chain, sponsors are likely to have access to much more detailed documents from the originators as compared to the buyers in addition to other informal channels of information exchange. If the loans are originated by the sponsors themselves, the information advantage over potential RMBS investors increases even more (e.g., see Keys et al. (2010)). We collect the identity of top originators for each pool in our sample from the deal prospectus. In almost half the cases, sponsors are also the top originators of the loan pool. In such cases, sponsor's information advantage over the buyers is likely to be higher, and we expect equity tranche to play an even more meaningful role here. This sub-sample test allows us to establish that our results are driven by asymmetric information concerns, and not by a pure absence of information with the sponsors. Thus, it establishes an important economic channel of private information.

In Table 4, we re-estimate our main regression specification (3) across these two groups. For expositional clarity, we reproduce the full sample results in Column (1), and the sub-sample results in Columns (2) and (3). While the estimated coefficient on ($Opaque_p \times HighEq_p$) remains negative for both subgroups, it is statistically and economically significant only in the sub-group where sponsors are also the top originators. In fact the economic magnitude of the coefficient increases by almost 70% for this sub-group. When the information advantage of sponsors is likely to be higher, equity tranche forecasts the default rate better. This is consistent with our main assertion that the equity tranche captures the sponsor's private information in RMBS deals.[22]

4.4 Identification Using Anti-Predatory Lending Laws

A potential concern with our analysis may be that opaque pools with a higher level of equity tranche are systematically *better* on observable dimensions that we, as econometricians, are unable to control for. If that be the case, we would find lower ex-post default for such pools even without any private information component of the equity tranche. Note that we have already controlled for some of the most important observable loan characteristics such as FICO score, LTV ratio, the purpose of the loans, the nature of interest rate, year of origination, and geographical location of the property in our default model. Earlier research has shown that these variables explain most of the variation in ex-post default of mortgages. Second, our matched sample exercise eliminates the effect of any observable characteristics that remains similar across the actual and matched pool. Third, we focus on the difference-in-differences coefficient in our empirical tests. This design ensures that we eliminate the effect of any differences in missing observables that correlate with the levels of loan opacity and equity tranche on an unconditional basis. Therefore, it is unlikely that our results are simply an artifact of missing observable characteristics. To further support this claim, we exploit the passage of state-level Anti-Predatory Lending Laws (APLs) as a source of exogenous variation in concerns about lenders' private information.

Several states passed these laws during our sample period to protect homeowners from predatory lending practices. These laws are structured along the lines of Federal Home Ownership and Equity Protection Act (HOEPA), and they typically impose more stringent restrictions on lending practices at the state level as compared to the Federal Act. APLs vary across states in terms of the type of loans they cover and the restrictions they impose on the lenders in terms of required lending practices and information disclosure rules. For example,

[22]In unreported tests, we estimate this model in a triple-interaction framework as well and obtain similar results. We do not report them for brevity.

some of these restrictions include limits on allowable prepayment penalties and balloon payments, borrower counseling requirements, and restriction on mandatory arbitration. Ho and Pennington-Cross (2005, 2006) provide detailed explanations of these laws and the timing of their passage by different states.

The passage of the law is likely to decrease the lenders' ability to originate and package predatory or abusive loans at the margin (see Agarwal, Amromin, Ben-David, Chomsisengphet, and Evanoff, 2012). Such a government regulation should make the use of private contracting mechanisms less important. Therefore, prior to the passage of this law, the equity tranche is likely to serve as a more important signal of private information. Said differently, if the equity tranche indeed conveys private information, then it should have a higher impact in the pre-APL period, i.e., during the period with relatively less government regulation on information disclosure rules.

Ho and Pennington-Cross (2005, 2006) provide an index of the strength of APLs across states as well as the date of the law passage. Their index varies from 4 to 17 with a median score of 10, where a higher index level indicates stronger laws in the state. Based on this measure, we classify all states with index value of 10 or above as the states with strong APL. These states are California, Colorado, Connecticut, Georgia, Illinois, Indiana, Massachusetts, New Jersey, New Mexico, North Carolina, and Washington DC. Of these states, all but Massachusetts and Connecticut, passed their law during our sample period (i.e., between 2002 and 2004), providing us with data on both before and after the law passage. For our test, we create an indicator variable APL that takes a value of one for states with strong APL, and zero otherwise. We create an indicator variable *Before* that equals one for loans that belong to states before the passage of law, and zero after that. As in our earlier tests, we create an indicator variable for high equity tranche (*HighEq*) based on the median level of this variable in our sample. With these three variables, namely APL, *HighEq*, and *Before*, we estimate a triple-difference model to estimate the difference in the effect of high equity tranche on future loan foreclosure rate for the APL states before and after the passage of the law as compared to the corresponding difference for states without the law. In this estimation strategy, we separate out the unconditional level effects of each one of these variables on the foreclosure rate as well as all the double-interaction effects. The coefficient on the triple-interaction term presents us with the estimate of interest.

We first estimate the model at loan-level. In this specification, we fit a loan level logistic regression model with foreclosure status as the dependent variable. Column (1) of Table 5 presents results estimated with the entire sample. Since *HighEq* is a pool-level variable, we cluster all standard errors at the pool level.[23] We find a negative and significant coefficient on *APL * HighEq * Before* indicating that equity tranche conveys stronger information about the future loan performance for APL states before the passage of the law. This is consistent with our prediction outlined above. In column (2), we restrict our sample to only *Opaque* pools and find the point estimate on the triple interaction term to be over twice as large, which indicates that the effect is especially strong for pools with higher concern about information asymmetry.

In columns (3) and (4), we collapse the data at the pool level. In the process we lose the loan-level variation since loans in a pool often come from different states. We compute the fraction of loans that comes from APL states and classify pools as *HighAPL* if the fraction of loans from APL states in the pool exceeds the sample median (59%). With this definition of pool-level APL, we estimate the triple-interaction model using an OLS approach. Despite the loss in variation arising out of pool-level aggregation, we find a negative and significant coefficient on the triple-interaction term.

[23] The use of clustered standard errors for logistics regressions requires a caveat. The estimates produced by standard maximum likelihood estimates (i.e., the ones produced by statistical packages such as STATA) may not be the true estimates when we have clustered observations. This happens because the observations are no longer independent within clusters. Hence the joint distribution function for the sample may no longer be the product of the distribution functions for each observation. Without a precise knowledge of the correlation structure within clusters, one cannot write down the true likelihood of the sample. Thus the estimates are consistent only under special cases. Considering these limitations, we also estimate the model by collapsing loan-level observations to pool level. While we lose the loan-level granularity in this approach, we are able to avoid the econometric concerns with clustered logistics regression model. Our main results remain similar across these modeling approaches.

Overall, these findings are consistent with equity tranche being an indicator of sponsors' favorable private information. Did investors recognize and respond to this indicator? To answer this question, we look at the pricing effect of the level of equity tranche in the following section.

4.5 Pricing Effect of Equity Tranche

Did investors pay higher prices for securities backed up by higher 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 more precisely. 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.[24]

We want to estimate the effect of equity tranche on the pricing of sold tranches in the same deal. An immediate implication of higher equity tranche is that there is less leverage in the deal. In such deals, superior tranches that are sold to the investors are safer and therefore they should command attractive prices. This effect is independent of any information revelation via the equity tranche that we are interested in. To separate out the leverage effect, we condition our analysis on the credit rating of sold tranches. We compare the pricing of two similarly rated tranches coming from deals with different levels of equity tranche. We

[24] 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.

maintain our basic empirical design that estimates the effect of equity tranche separately across opaque and transparent pools. If the effect of equity tranche on prices come entirely due to the leverage effect, then we should find no difference across opaque and transparent pools. On the other hand, if the effect comes via the revelation of private information, then we expect to see higher prices for tranches backed by higher equity tranche only in the opaque pools.

We divide all tranches into broad credit rating classes: AAA, AA, A, and BBB.[25] For deals with multiple tranches within one rating class, we compute a dollar-weighted average yield spread and consolidate them into one observation. This aggregation leads to 549 sold tranches in our sample, out of which 379 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 6 presents the cross-tabulation of the average yield spread 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 yield spread is lower for pools with higher equity tranche. As sponsors sell more of their pool's cash flows to outside investors, the price decreases (yield spread 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. Columns (1)-(2) of Table 7 present our base results. The significant negative coefficient on *HighEq* indicates that after controlling for the credit rating class, high-equity-tranche deals have 27 basis points lower yield spread. 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 the pricing effect to be higher for these pools and the empirical results confirm this intuition. Taken together

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

with the abnormal default rate results, our results show that equity tranche did contain the sponsor's private information, and market prices reflect this ex ante in a cross-sectional sense.

4.6 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 (e.g., Piskorski, Seru, and Vig, 2010; Demiroglu and James, 2012). 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 variable that indicates if the top originator for the pool is also the servicer (*TopOrigAndService*).[26]

[26] 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 (Hartman-Glaser, 2012). In our main tests, we already include sponsor fixed effects in the estimation exercise. This ensures that we are able to separate out time-invariant reputational effect of the sponsors. Since we consider a short time-period (2002-2005) for our analysis, it is reasonable to assume that a sponsor's reputation remained practically constant during the sample period. As an alternative test in a similar spirit, we consider the heterogeneity in the sponsor-type to control for the reputational concerns. We expect that long-lived and established commercial banks such as JP Morgan Chase have different concerns about protecting their franchise values as compared to 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 stand-alone mortgage lenders (Je et al., 2012). To address these issues, 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 regression model.

Table 8 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 "skin in the game" hypothesis as deals where the top originator is also the servicer have better ex-post performance. However, inclusion of this control variable does not change any of our results. In unreported tests, we repeat these analyses for other tests of the paper as well, and our main results remain robust to these controls. Overall, our results are unlikely to be affected by these alternative channels.

5 Discussions and Conclusion

This paper empirically examines the role of equity tranche in residential mortgage-backed securities during the build-up to the sub-prime mortgage crisis. We document that the level of equity tranche conveys the sponsor's private information in opaque pools. Within such pools, higher levels of equity tranche is associated with significantly lower future foreclosure rates after parsing out the effects of 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 some of the fundamental predictions of security design models based on asymmetric information (e.g., Leland and Pyle (1977) and DeMarzo and Duffie (1999)).

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. By design, our study is cross-sectional in nature. Therefore, we are able to comment on the ability of equity tranche in explaining economic outcomes only in a relative sense. Our study does not rule out the possibility that the absolute level of equity tranche supporting these deals was too low during the sample period. 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 asymmetric information models. 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

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

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

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas Evanoff, 2012, Predatory lending and the subprime crisis, *Working Paper* .

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

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

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

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

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

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

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

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

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

Boot, Arnoud WA, and Anjan V Thakor, 1993, Security design, *The Journal of Finance* 48, 1349–1378.

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

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

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

Demiroglu, Cem, and Christopher James, 2012, How Important is Having Skin in the Game? Originator-Sponsor Affiliation and Losses on Mortgage-backed Securities, *Review of Financial Studies* 25, 3217–3258.

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

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

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

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

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

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

Gorton, Gary B, 2010, Slapped by the invisible hand: The panic of 2007 .

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

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

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

Ho, Giang, and Anthony Pennington-Cross, 2005, The impact of local predatory lending laws on the flow of subprime credit, *Federal Reserve Bank of St. Louis Working Paper No. 2005-049B. Available at: http://research.stlouisfed.org/wp/2005/2005-049.pdf* .

Ho, Giang, and Anthony Pennington-Cross, 2006, The impact of local predatory lending laws on the flow of subprime credit, *Journal of Urban Economics* 60, 210–228.

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

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

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

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

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

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

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

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

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

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

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

Scharfstein, David, and Adi Sunderam, 2011, The economics of housing finance reform, in Martin Neil Baily, ed., *The Future of Housing Finance*, 146–198 (Brookings Institution Press).

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

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

Appendix

Appendix 1a: 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 and exploit changes in anti-predatory-lending laws. 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. 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-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.

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 present a detailed discussion of the equity tranche in Section 3. 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 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 500,000 individual mortgages. We obtain key information for each loan in a given pool from CoreLogic such as 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 through 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.

Appendix 1b: Example of Documentation Description from a Deal Prospectus

Series Name: ABFC Mortgage Loan Asset-Backed Certificate, Series 2002-WF2

The Originator's subprime mortgage loan programs include a full documentation program, a "stated income, stated asset" program and a "lite" documentation program. Under the full documentation program, loans to borrowers who are salaried employees must be supported by current employment information in the form of one current pay-stub with year-to-date information and W-2 tax forms for the last two years (a complete verification of employment may be substituted for W-2 forms). The Originator also performs a telephone verification of employment for salaried employees prior to funding. In some cases, employment histories may be obtained through V.I.E., Inc., an entity jointly owned by the Originator and an affiliated third party, that obtains employment data from state unemployment insurance departments or other state agencies. Under the full documentation program, borrowers who are self-employed must provide signed individual federal tax returns and, if applicable, signed year-to-date income statements and/or business federal tax returns. Evidence must be provided that the business has been in existence for at least one year. If the business has been in existence less than two years, evidence must be provided that the applicant had previously been in the same line of work for at least one year. Under the full documentation program, at certain loan-to-value ratio levels and under certain circumstances not all sources of funds for closing are verified as the borrowers.

Under the Originator's "Stated Income, Stated Asset" program, the applicant's employment, income sources and assets must be stated on the initial signed application. The applicant's income as stated must be reasonable for the applicant's occupation as determined in the discretion of the loan underwriter; however, such income is not independently verified. Similarly the applicant's assets as stated must be reasonable for the applicant's occupation as determined in the discretion of the loan underwriter; however, such assets are not independently verified. Except under the Stated Asset Program, verification of funds sufficient to close the mortgage loan is performed. Under the "LITE" Documentation program, the Originator reviews the deposit activity reflected in the most recent six or twenty-four consecutive months of the applicant's bank statements as an alternative method of establishing income. Maximum loan-to-value ratios within each credit level are lower under the stated income, stated asset program than under the full documentation program.

Appendix 2: Matched 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).

Table A.1: **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	

Table A.2: **The Equity Tranche**

This table presents examples of two common tranche structures used for RMBS and how the equity tranche is computed for each case.

Panel A: Six-pack

Offered	Class	Principal ($)	Rate	Rating S&P	Rating Moody's
Y	A	399,181,000	LIBOR+0.34	AAA	Aaa
Y	M-1	35,789,000	LIBOR+0.65	AA+	Aa2
Y	M-2	27,530,000	LIBOR+1.20	A+	A2
Y	M-3	23,776,000	LIBOR+2.00	BBB	Baa2
Y	M-4	6,757,000	LIBOR+2.30	BBB-	Baa3
N	CE	7,508,765		NR	NR
	Sum	500,541,765			
Pool:	Mortgages	3,737			
	Principal	500,541,765			
	Equity Tranche =	7,508,765/500,541,765 = 1.50%			

Panel B: Overcollateralization

Offered	Class	Principal ($)	Rate	Rating S&P	Rating Moody's
Y	A	154,414,000	5.50	AAA	Aaa
Y	M-1	27,440,000	LIBOR+0.50	AA	Aa2
Y	M-2	12,267,000	LIBOR+0.75	A	A2
Y	M-3	4,196,000	LIBOR+0.80	A-	A3
Y	B-1	5,058,000	LIBOR+1.25	BBB+	Baa1
Y	B-2	3,336,000	LIBOR+1.30	BBB	Baa2
Y	B-3	6,564,000	LIBOR+2.15	BBB-	Baa3
	Sum	213,275,000			
Pool:	Mortgages	1,039			
	Principal	215,212,063			
	Equity Tranche =	(215,212,063–213,275,000)/215,212,063 = 0.90%			

Table A.3: **Default Model**

This table presents the results of the default model. We use the estimated coefficients of this model to predict the loan-by-loan probability of foreclosure to construct our measure of *Abnormal Default* used for the estimates in Table 4. Following prior literature (e.g., see Demyanyk and Van Hemert, 2011), we include the borrower's FICO score, the loan-to-value ratio, loan purpose (e.g., Refinancing with Cash-Out), loan type, (e.g., 5-year Interest Only), state fixed effects, and year fixed effects. The results below show the key drivers of default risk, with the point estimates on the other variables in the estimation omitted in the interest of space.

	Prob(Foreclosure)	
	b	se
FICO	-0.0059***	(0.000)
LTV	0.0180***	(0.000)
Refinancing with Cash-Out	-0.2455***	(0.000)
Refinancing w/o Cash-out	-0.3458***	(0.000)
5-year Interest Only	0.8339***	(0.000)
10-year Interest Only	0.8071***	(0.000)
Adjustable Rate Mortgage	0.2906***	(0.000)
5-year I.O. ARM	0.6705***	(0.000)
10-year I.O. ARM	0.6598***	(0.000)
7-year I.O.ARM	0.0436	(0.441)
2-year I.O. ARM	0.8834***	(0.000)
7-year Balloon	1.9788	(0.185)
15-year Balloon	-1.0840***	(0.000)
ARM Balloon	0.8550***	(0.000)
Balloon-Other	1.0936***	(0.000)
Arizona	0.3898**	(0.019)
California	0.4543***	(0.006)
Florida	0.8219***	(0.000)
Georgia	1.0356***	(0.000)
Nevada	1.2092***	(0.000)
State Fixed Effects	Yes	
Year Fixed Effects	Yes	
Other Controls	Yes	
Observations	497367	

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

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.

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
Late	0.52	0.50	0.00	0.00	1.00	1.00	1.00	196
Subprime (FICO<660)	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 (loan counts in brackets)

	No	Yes
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* (columns (1)-(3)) and *Mezzanine-to-Sold* (columns (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 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, *FICO* is the pool's weighted average FICO score, *LTV* is the pool's weighted average loan-to-value ratio, *% ARM* is the percent of the loan pool with adjustable rate mortgage loans, *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***	0.761**	2.751**	3.181***	3.167**
	(0.00)	(0.00)	(0.01)	(0.02)	(0.00)	(0.01)
% NoDoc	0.025***	0.023**	0.016*	0.200***	-0.007	-0.007
	(0.00)	(0.04)	(0.06)	(0.00)	(0.67)	(0.77)
FICO		-0.004	0.004		-0.101***	-0.101***
		(0.23)	(0.29)		(0.00)	(0.00)
LTV		-0.025	0.034		0.301***	0.275***
		(0.32)	(0.20)		(0.00)	(0.01)
% ARM		0.005	0.006**		-0.015***	-0.013**
		(0.14)	(0.04)		(0.01)	(0.04)
GeoDiverse		-0.008	-0.005		-0.054***	-0.056***
		(0.42)	(0.62)		(0.00)	(0.00)
Constant	0.295**	5.268	-3.850	3.847***	58.015***	61.686***
	(0.04)	(0.11)	(0.35)	(0.00)	(0.00)	(0.00)
Sponsor FE	No	No	Yes	No	No	Yes
Observations	163	163	163	163	163	163
R^2	0.268	0.318	0.577	0.334	0.857	0.869

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 columns (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 columns (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 the pool's weighted average FICO score, and *LTV* is the pool's weighted average loan-to-value ratio. All standard errors are heteroskedasticity robust.

	Default Model			Matched Pool		
	(1)	(2)	(3)	(4)	(5)	(6)
Late	0.321***	0.293***	0.293***	-0.006	0.008	0.037
	(0.00)	(0.00)	(0.00)	(0.95)	(0.92)	(0.68)
FICO	0.001	0.002*	0.002	0.002**	0.002*	0.001
	(0.38)	(0.08)	(0.13)	(0.05)	(0.10)	(0.58)
LTV	0.052***	0.051***	0.046***	0.069***	0.069***	0.060***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Opaque	0.163*	0.109	0.140	0.099	0.057	0.123
	(0.06)	(0.26)	(0.20)	(0.49)	(0.73)	(0.47)
HighEq	0.110	0.139*	0.202**	0.033	0.062	0.138
	(0.13)	(0.06)	(0.02)	(0.72)	(0.50)	(0.14)
HighEq * Opaque	-0.244**	-0.263***	-0.241**	-0.221*	-0.265**	-0.267**
	(0.01)	(0.01)	(0.02)	(0.07)	(0.03)	(0.05)
HighMezz		0.080	0.155		-0.115	-0.027
		(0.51)	(0.27)		(0.36)	(0.86)
HighMezz * Opaque		0.115	0.113		0.141	0.088
		(0.31)	(0.32)		(0.32)	(0.54)
Constant	-3.783***	-4.477***	-4.315***	-5.588***	-5.473***	-4.130**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Sponsor FE	No	No	Yes	No	No	Yes
Observations	151	151	151	151	151	151
R^2	0.650	0.659	0.723	0.440	0.444	0.518

p-values in parentheses

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

Table 4: **Ex-Post Outcomes: The Channel of Private Information**

This table presents OLS estimates from regressions of *AbDefault* on loan pool characteristics. *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 the pool's weighted average FICO score, and *LTV* is the pool's weighted average loan-to-value ratio. *Sponsor is Top Originator* indicates deals where the deal sponsor originated more loans in pool than any other originator. All standard errors are heteroskedasticity robust.

	(1) All	(2) Sponsor not Top Originator	(3) Sponsor is Top Originator
Late	0.008	0.035	-0.050
	(0.92)	(0.74)	(0.71)
FICO	0.002*	0.000	0.003**
	(0.10)	(0.92)	(0.05)
LTV	0.069***	0.070***	0.080***
	(0.00)	(0.00)	(0.00)
Opaque	0.057	0.021	0.027
	(0.73)	(0.93)	(0.90)
HighEq	0.062	-0.062	0.126
	(0.50)	(0.71)	(0.21)
HighEq * Opaque	-0.265**	-0.074	-0.421***
	(0.03)	(0.72)	(0.01)
HighMezz	-0.115	-0.370	-0.060
	(0.36)	(0.19)	(0.65)
HighMezz * Opaque	0.141	0.241	0.149
	(0.32)	(0.33)	(0.34)
Constant	-5.473***	-4.368**	-7.325***
	(0.00)	(0.05)	(0.00)
Observations	151	73	78
r2	0.444	0.398	0.538

p-values in parentheses

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

Table 5: **Anti-Predatory Lending Laws, Equity Tranche, and Ex-Post Outcomes**
Columns (1) and (2) present loan-level logistic regression estimates where the dependent variable is an indicator variable equal to one if the loan ends up in foreclosure. Columns (3) and (4) present pool-level OLS regression estimates where the dependent variable is the pool level abnormal default, *AbDefault*, which 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. We compute heteroskedasticity robust standard errors for these specifications. Columns (2) and (4) present estimates from regressions including only *Opaque* pools, which are those 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, *APL* is a dummy variable equal to 1 for loans from states that enact anti-predatory lending laws, *HighAPL* is a dummy variable equal to 1 for pools where the proportion of loans from APL states is greater than sample median, *Before* is a dummy variable equal to 1 for the time period prior to the passage of APL laws. *FICO* is the FICO score and *LTV* is the loan-to-value ratio, where we use dollar-weighted averages for the pool-level specifications. The loan level regressions also include dummies for different loan purposes and loan types (coefficients not reported).

	Loan Level		Pool Level	
	(1) All	(2) Opaque	(3) All	(4) Opaque
FICO	-0.006*** (0.00)	-0.005*** (0.00)	0.002* (0.06)	0.002 (0.19)
LTV	0.018*** (0.00)	0.009*** (0.00)	0.054*** (0.00)	0.067*** (0.00)
HighEq	-0.153 (0.10)	-0.208** (0.02)	-0.077 (0.29)	-0.162 (0.11)
Before	-0.053 (0.86)	-0.086 (0.74)	-0.342** (0.02)	-0.407** (0.02)
HighEq * Before	0.611*** (0.00)	0.434** (0.01)	0.346** (0.04)	0.396** (0.05)
APL	0.337 (0.22)	0.451 (0.15)		
APL * HighEq	0.256*** (0.00)	0.293*** (0.00)		
APL * Before	-0.297*** (0.01)	-0.021 (0.86)		
APL * HighEq * Before	-0.226* (0.09)	-0.463*** (0.00)		
HighAPL			-0.050 (0.60)	0.061 (0.55)
HighAPL * HighEq			0.036 (0.78)	-0.234 (0.12)
HighAPL * Before			-0.085 (0.62)	0.027 (0.91)
HighAPL * HighEq * Before			-0.443** (0.04)	-0.486* (0.09)
Constant	-13.685 (0.21)	0.621* (0.07)	-4.034*** (0.00)	-5.041*** (0.00)
Observations	497367	254774	151	72
Pseudo R^2	0.110	0.092		
R^2			0.688	0.569

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

Table 6: **Equity Tranche and Yield Spreads Cross-tabulation**

This table presents the mean yield spread for variable rate tranches in the sample according to the size of the equity tranche and the tranche's rating class. 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 7: **Price Response to Equity Tranche**

This table presents OLS estimates from regressions of the yield spread (in percentage points) on loan pool characteristics. 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 8: **Robustness – Alternate Channels**

This table presents our main results from earlier tables alongside specification that include other variables that capture the roles and connections of the various agents in the securitization chain. *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 the pool's weighted average FICO score, *LTV* is the pool's weighted average loan-to-value ratio, *% ARM* is the percent of the loan pool with adjustable rate mortgage loans, *GeoDiverse* measures the geographic diversity and is computed as 100 - (percent of largest one state origination concentration) in the mortgage pool, *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. Institution-Type effects refers to the inclusion of a set of dummy variables that identify sponsors 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 Match	
	(1)	(2)	(3)	(4)	(5)	(6)
Late	0.943***	0.840***	3.181***	3.235***	-0.006	-0.044
	(0.00)	(0.00)	(0.00)	(0.00)	(0.95)	(0.56)
NoDoc	0.023***	0.019**	-0.007	-0.012		
	(0.01)	(0.04)	(0.55)	(0.38)		
FICO	-0.004	0.004	-0.101***	-0.102***	0.002**	0.001
	(0.16)	(0.29)	(0.00)	(0.00)	(0.05)	(0.25)
LTV	-0.025	0.009	0.301***	0.288***	0.069***	0.062***
	(0.23)	(0.74)	(0.00)	(0.00)	(0.00)	(0.00)
ARM	0.005**	0.006**	-0.015***	-0.016***		
	(0.01)	(0.01)	(0.01)	(0.01)		
GeoDiverse	-0.008	-0.010	-0.054***	-0.058***		
	(0.27)	(0.10)	(0.00)	(0.00)		
Opaque					0.099	0.138
					(0.49)	(0.29)
HighEq					0.033	0.074
					(0.72)	(0.46)
HighEq * Opaque					-0.221*	-0.300**
					(0.07)	(0.02)
SellAndService		-0.453*		-0.632		0.218**
		(0.10)		(0.29)		(0.02)
TopOrigAndService		-0.210		0.196		-0.361***
		(0.41)		(0.70)		(0.00)
Constant	5.268*	-2.111	58.015***	60.044***	-5.588***	-4.546***
	(0.08)	(0.59)	(0.00)	(0.00)	(0.00)	(0.00)
Institution Type FE	No	Yes	No	Yes	No	Yes
Observations	163	163	163	163	151	151
R^2	0.318	0.494	0.857	0.860	0.440	0.540

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.

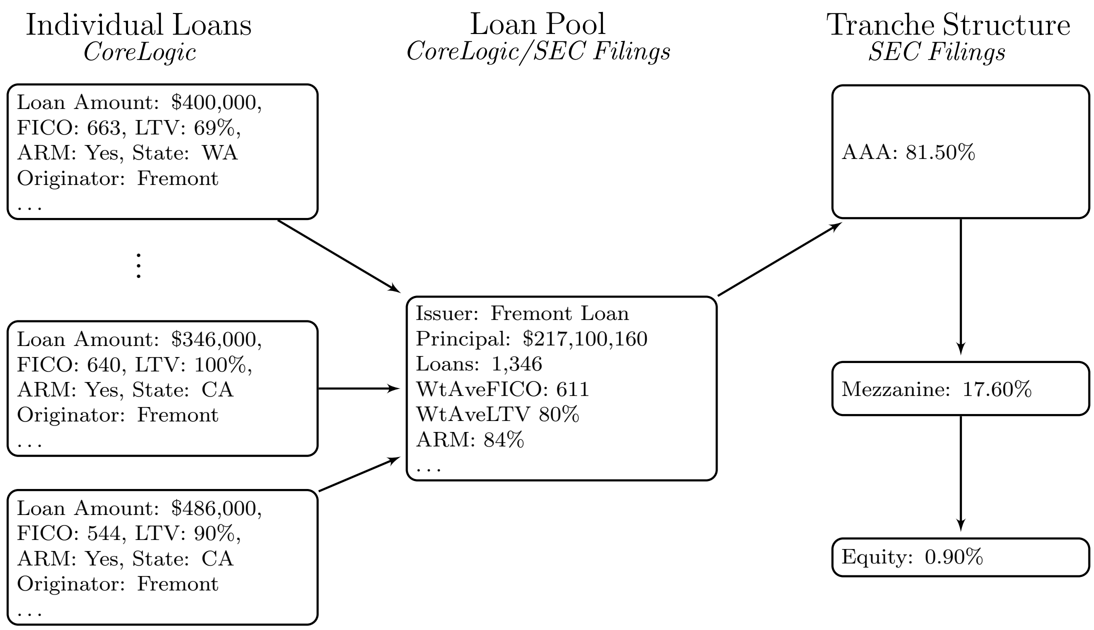
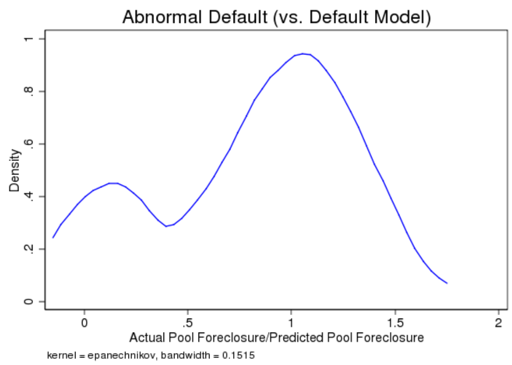
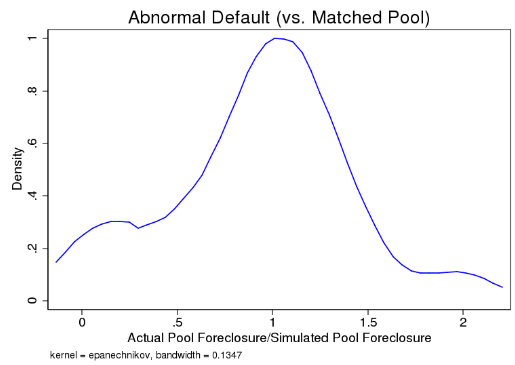


Figure 2: **Measures of Abnormal Default**

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) Matched Pools