Information Frictions in Securitization Markets: Unsophisticated Investors or Opaque Assets?

David Echeverry *

February 28, 2019

Abstract

Investors in residential mortgage-backed securities have information about the probability that the bond's rating is downgraded, and this is reflected in the security design. In general, a higher level of subordination is predictive of a lower likelihood of downgrade. However, this information content is affected by two possible frictions. Using documentation quality on private label mortgages to measure opacity of a security, I show that opaque deals exhibit less information content. Once opacity is taken into account, the traditional measure of investor sophistication is not the main driver of information content. More precisely, subordination percentages of junior tranches are no more informative than those of AAA tranches within "low-doc" deals, while the latter are no less informative within "full-doc" deals.

JEL classification: G21, G24

Keywords: credit ratings, mortgages, mortgage-backed securities, Gaussian copula, credit risk, bond prices.

*Mendoza College of Business, University of Notre Dame. Email:decheve2@nd.edu. I am grateful to Jiakai Chen, Martijn Cremers, Brendan Daley, Nicolae Gârleanu, Amir Kermani, Ilénin Kondo, Haoyang Liu, Christopher Palmer, Michael Reher, Paul Schultz, James Vickery and Nancy Wallace for their valuable feedback. I thank seminar participants at Banco de la Repúbica de Colombia, Boston University, Universidad de los Andes, Universidad Javeriana, University of Florida and University of Notre Dame for their helpful comments and suggestions. I gratefully acknowledge the generous support from the Fitzgerald Institute for Real Estate and the Fisher Center for Real Estate and Urban Economics, while part of the paper was written. I thank Sanket Korgaonkar and Haoyang Liu for their valuable help.

Information frictions in mortgage-backed security markets are cited among the root causes of the collapse of private label securitization in 2007 and its subsequent stagnation.[1] Yet private investment remains vital to the mortgage market. Finkelstein, Strzodka, and Vickery (2018) document that the government-sponsored entities, Fannie Mae and Freddie Mac, have reduced Federal government exposure to credit risk on close to $1.8tn mortgages by transferring a growing share of it to the private sector. The transfer is accomplished by pooling and tranching loans in order to issue bonds backed by such loans. Yet in spite of the importance of credit risk transfer mechanisms in mortgage-backed securities, much remains to be understood about the information frictions that affect such mechanisms. This paper aims to bridge this gap.

Ashcraft, Goldsmith-Pinkham, Hull, and Vickery (2011) argue that early prices of collateralized mortgage obligations (CMOs) are predictive about subsequent bond downgrades. However, a number of information frictions affect their information content. According to Ashcraft and Schuermann (2008), two frictions take place between the investor and the originator of the securities. The first one is investor unsophistication, whereby junior investors are better informed than senior ones (Boot and Thakor, 1993). According to Gorton and Pennacchi (1990), senior bondholders seek information-insensitive tranches, i.e. AAA-rated bonds, while junior bondholders are better suited to handle information-sensitive tranches, i.e. junior bonds.[2] This can give rise to a principal-agent problem between the uninformed senior investor and the issuer to which the better-informed (junior) investor is not exposed. Accordingly, Adelino (2009) provides evidence that AAA tranches exhibit an information disadvantage relative to the junior ones, suggesting senior bondholders are relatively unsophisticated.

A less explored information friction stems from imprecise knowledge about the quality of the underlying mortgages that compose the pool. When the quality of the securitization deal is opaque this can give rise to adverse selection, as shown by Skreta and Veldkamp (2009).[3] I measure deal opacity for a given securitization deal as the level of documentation completeness on the underlying pool of loans. Thus the opacity issue linked to low documentation, unlike the (lack of) sophistication issue described above, impacts both senior and junior investors equally. This fact allows me to distinguish the information advantage linked to asset opacity (i.e. between high-doc and low-doc deals) from the one linked to investor sophistication as is currently proxied in the literature (i.e. between AAA bonds and others).

The main contribution of this paper is to provide evidence that while asset opacity drives downgrade predictability, lack of sophistication plays a secondary role. In a reduced-form specification using likelihood of a bond downgrade as dependent variable, I find that the subordination percentage (i.e. the percentage of the total capital which is subordinate to the bond in question) of a junior

[1] The market for non-conforming loans is far from its 2007 peak origination of over $2tn despite the recent upwards trend: in 2017, $4.1bn of securities backed by so-called nonprime loans were issued, according to Inside Mortgage Finance, with issuance in Q1 2018 roughly doubling that of Q1 2017).

[2] The efficiency of this arrangement is discussed by Dang, Gorton, and Holmström (2013). In particular, when information is costly this helps the market liquidity (Gorton and Ordonez, 2013).

[3] The adverse selection problem can be related to borrowers fraudulently inflating their income report in "low-doc" loan applications (AMBROSE, CONKLIN, and YOSHIDA, 2016). In terms of the issuer of the loans, Daley, Green, and Vanasco (2018) show that the existence of external agency ratings can reduce the incentives to conduct due diligence on the quality of the assets. This due diligence burden is then passed on to the the bond investors when they purchase the bonds, so I will consider the due diligence problem from their perspective.

tranches appears to be no more predictive of downgrade than that of the AAA tranche for a deal backed by low documentation loans. Conversely, AAA subordination percentage appears no less informative than junior subordination among "full-doc" deals. Among deals with intermediate levels of documentation, I recover evidence in line with Adelino (2009), namely that junior issues are somewhat more informative than senior ones. So while the principal-agent problem discussed in earlier literature does take place, it is mediated by the extent of asset opacity affecting the securities.

The key component in the design of a security, including its agency rating, is the amount of subordination (IOSCO, 2008). When a bond is issued, both its yield at issuance and subordination percentage are determined jointly with the agency rating. Whereas prior literature have focused on either the security design alone (Begley and Purnanandam, 2017) or prices alone (Adelino, 2009; Ashcraft et al., 2011), I study the information content of mortgage-backed security prices by looking at both dimensions of bond price. The results from a reduced-form specification suggest the variation in information content is captured by the subordination structure rather than prices themselves, especially if bonds tend to be issued at par (Adelino, 2009; Ashcraft et al., 2011). In other words, subordination is predictive of downgrades among "high-doc" deals but not among low-doc ones. In line with this, I document in the first part of the paper that the information differential is mainly manifested through subordination percentages.[4]

To reflect the equilibrium relationship between these components, in the second part of the paper I introduce a structural model to infer the default correlation implied by a given bond price and subordination.[5] My model takes into account the probability of default and probability of prepayment of the underlying loans to identify beliefs about default correlations implied from security prices and subordination percentages.

I argue that the main information friction affecting investors in nonagency bond markets is adverse selection, resulting from deal opacity, rather than a principal-agent problem resulting from lack of sophistication. Other papers have considered how the investors mitigate adverse selection.[6] Looking at SEC filings from 234 deals, Begley and Purnanandam (2017) show that investors are aware of the lemons problem they are exposed to, so that opaque deals command more skin in the game, i.e. larger equity tranches, to mitigate the information friction. They focus on the signaling role of the equity tranche about bond outcomes, especially in opaque deals, while I look at how the subordination structure, together with bond price and coupon, is associated with price informativeness and how this is affected by opacity.

[4]Because the equilibrium relationship between price and subordination is not linear, the reduced-form specification I just described does not suffer from collinearity.

[5]Default correlations can be computed ex post from default experience instead of inferring them from bond prices. See Cowan and Cowan (2004); de Servigny and Renault (2002); Geidosh (2014); Gordy (2000); Nagpal and Bahar (2001). Though default-based measures are not directly comparable to mine (Frye, 2008), one study based on default experience worth noting here is Griffin and Nickerson (2016). They infer rating agency beliefs about corporate default correlations by studying collateralized loan obligation (CLO). Their results suggest such beliefs were revised upwards after the crisis, but not sufficiently so when benchmarked against a performance-based estimator accounting for unobserved frailty in the default generating process (Duffie, Eckner, Horel, and Saita, 2009). The results suggest that agency ratings adapted more slowly to the crisis than market prices.

[6]For example, Downing, Jaffee, and Wallace (2009) argue that agency loans sold to special purpose vehicles (SPVs) trade in a market for lemons; bond coupons exhibit a "lemons spread", showing investors are overall aware of the information friction. Adelino, Gerardi, and Hartman-Glaser (2016) argue that investors in the secondary market for loans deal with opacity by *skimming* the underlying loans, deferring the purchase to better observe their quality.

I use loan level information about documentation completeness to construct a deal level opacity index for all the deals in the sample.[7] JEC (2007) documents a relative decline in the number of full documentation subprime loans in the running to the crisis. Keys, Mukherjee, Seru, and Vig (2010) argue that the "low-doc" loans underperformed (in terms of defaults) relative to otherwise similar but better documented loans. This relative underperformance of low-doc loans is confirmed by the results of Begley and Purnanandam (2017) and Kau, Keenan, Lyubimov, and Slawson (2011). Ashcraft, Goldsmith-Pinkham, and Vickery (2010) document the underperformance of low-doc deals. Similarly, Griffin and Maturana (2016) and Piskorski, Seru, and Witkin (2015) provide evidence that misrepresentation of loan features was associated with higher losses.[8] I emphasize investor information about future performance rather than asset performance per se.

The amount of AAA issuance is decreasing in documentation completeness (controlling for deal average probability of default). Prior evidence of rating inflation in CDOs includes Griffin and Tang (2012) who speak of subjective ratings. An, Deng, Nichols, and Sanders (2015) use the number of tranches as a proxy for complexity and argue that more complex CMBS structures see lower subordination levels. Then in RMBS markets, Benmelech and Dlugosz (2010) link rating inflation to rating shopping, but not to asset opacity. My result is in line with the theoretical predictions of Skreta and Veldkamp (2009) that ratings are more likely to be inflated when asset quality is opaque or "complex", to use their term.

Because a central objective of securitization is diversification through pooling, default correlation is essential to the value of the security. Thus prices of structured products that are subject to default risk reflect investors' beliefs about default correlation. [9] I use a single factor Gaussian copula (Li, 2000), which Hull and White (2006) call "the standard market model for valuing collateralized debt obligations and similar instruments".[10] I estimate the probability of default (PD) and loss given default (LGD) from loan performance data, following common practice in CDO pricing models, so that default correlation is implied from the market price. Heterogeneity in the information content of implied correlations means investors disagree about the value of this parameter. By taking default probabilities as fixed and estimating default correlations, the implicit assumption in the Gaussian copula approach is that the main source of disagreement among investors in a given deal is the default correlation. The literature has examined the role of disagreement about other risk attributes such as the probability of a crisis (Simsek, 2013) or the prepayment speed (Carlin, Longstaff, and Matoba, 2014; Diep, Eisfeldt, and Richardson, 2016). The prominence of Gaussian copulas in the CDO literature suggests that the primary source of disagreement across bonds in such a structure is the default correlation.[11]

[7] Begley and Purnanandam (2017) collect information about loan documentation quality from the prospectus of 234 deals, which they retrieve from SEC filings.

[8] The find that misrepresentation is a phenomenon affecting low and high documentation loans alike.

[9] This benefits junior bondholders at the expense of senior ones (Duffie and Gârleanu, 2001)

[10] See Brunne (2006); D'Amato and Gyntelberg (2005); Duffie and Singleton (2012); Elizalde (2005); Hull and White (2004, 2006, 2008); McGinty, Beinstein, Ahluwalia, and Watts (2004); Tzani and Polychronakos (2008). The higher the correlation across defaults, the higher the volatility in the total cashflows, which affects the value of all tranches except the equity piece.

[11] Agency securities do not carry default risk to the investor, only prepayment risk. Schwartz and Torous (1989) and Stanton (1995) are two examples of pricing models that value the prepayment option. Downing, Stanton, and Wallace (2005) propose a two-factor valuation model that distinguishes the separate, competing risks carried by the default and the prepayment options, where default correlations are not modeled. Also Sugimura (2004) develops an intensity model to price RMBS (pass-through) bonds not insured against default risk, and thus exposed to both

Coval, Jurek, and Stafford (2009a) use a Gaussian copula model[12] to show that security prices are sensitive to underlying default correlations, and that this sensitivity compounds along the structured finance chain. As Cordell, Huang, and Williams (2012) show (see Figure A.1) the underlying collateral of cash CDOs is predominantly mezzanine tranches of CMOs, which in turn are composed of mortgages. This makes CDOs very sensitive to loan default correlation, much like the CDO^2 in Coval et al. (2009a).[13] This highlights the importance of CMO default correlations to structured finance markets, while leaving the question open as to which of the investors, the junior or the senior ones, may be miscalculating them. The evidence I provide suggests senior investors were less informed, but only for intermediate levels of opacity.

My focus on early originations leaves out developments that took place over the boom. Recent literature suggests that deal opacity had an increasingly important role in the running to the crisis. Using six measures of deal complexity built from the prospectuses of subprime securities issued between 2002 and 2007, Ghent, Torous, and Valkanov (2016) offer evidence of growing obfuscation between the issuer and the senior investor; the latter didn't price in the higher risks due to security complexity (complexity being measured by the number of tranches in the deal). They argue that complex deals facilitated the collusion between the issuer and the junior investor (Demiroglu and James, 2012) to divert cash flows from senior securities to junior ones. My finding that pre-boom originations across the spectrum (prime, alt-A and subprime) see both senior and junior investors are equally affected by loan opacity, suggesting that collusion became a problem over the boom.

The paper proceeds as follows. Section 1 presents the data. Section 2 presents my empirical strategy based on price, coupon and subordination. Section 3 lays out the copula model from which I infer default correlations and copula model estimates. Section 4 replicates the results from Section 2 using implied correlation as the independent variable. Section 5 concludes.

1 Data

CMOs are traded over the counter. I use data from Thomson Reuters, which records bond prices from January 2004 onwards. I obtain series of prices for CMOs originated before and up to June 2005, i.e., prior to the pre-crisis mortgage boom.[14] ABSNet collects monthly information about private label securitization deals, providing snapshots of all tranches inside a given deal between the time of origination and the end of 2016.[15] I check the consistency between the ABSNet

prepayment and default risk; credit events in his approach are assumed to be uncorrelated. I focus on the implied default correlation as the outcome of the pricing model.

[12] Using their parameters my model replicates their results (see Figure A.2).

[13] Gorton (2009) argues that the information destruction in structured products was caused by their layered structure. Because of this, CMO prices are the closest reflection of the market view on default correlations. I provide a measure of default correlations directly from RMBS prices, which contributes to prior estimates from the CDO pricing literature. Among those, Duffie and Gârleanu (2001) and Duffie and Singleton (2012) discuss the pricing of cash CDOs. Otherwise, the literature has mostly focused on synthetic CDOs and tranches of credit default swap baskets (Andersen and Sidenius, 2004; Andreoli, Ballestra, and Pacelli, 2016; Benešová and Teply, 2010; Brunne, 2006; Buzková and Teplỳ, 2012; Coval, Jurek, and Stafford, 2009b; Elizalde, 2005; D'Amato and Gyntelberg, 2005; Hull and White, 2004, 2006; Longstaff and Rajan, 2008; Schlösser, 2011; Stanton and Wallace, 2011).

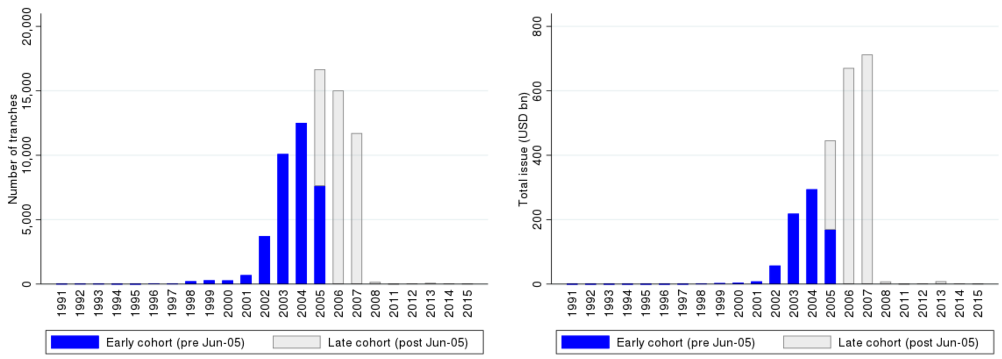
[14] See Echeverry, Stanton, and Wallace (2016).

[15] Starting July 2009, ABSNet also started recording bond prices over time, which allows me to cross-check prices across sources by matching on bond CUSIP, year and month, using the nearest transaction to the rating observation date. The average time difference is 1.83 days, the median being 0 and the 99th percentile 53 days.

price and the mid price in Thomson Reuters. I find a median absolute difference of \$0.06 and a 99th percentile of \$1.51, the gap being consistent with the small time differences in the date of observation across sources.

For each month, ABSNet records rating, subordination, bond maturity and coupon for each tranche. I collect all the snapshots available from each deal in their website. Tranches are organized in a matrix format by increasing subordination levels, which determine the default cushion for a given tranche. From there I derive the detachment point for each tranche, and thus the waterfall of losses for the given deal.[16]

From the early cohorts, i.e., those originated before June 2005, I observe 35,692 tranches -14 tranches per deal on average- for a total \$1,854.8bn of originated securities (see Table 5). In comparison, Adelino (2009) includes boom-time data to obtain 67,412 securities from JP Morgan's MBS database, a total issue of \$4,204.8bn. I follow his data cleaning procedures such as removing Interest Only, Principal Only, Inverse Floater and Fixed to Variable bonds from the sample. Alt-A and subprime deals are the largest classes (see Table 6). Although the growth of these asset classes mostly happened in the running to the crisis (Gorton, 2009) I find that pre-boom originations are also composed mostly of suprime and alt-A bonds.



(a) Number of tranches

(b) Total issue

Figure 1.1: Number of tranches and amount issued by vintage year for private label collateralized mortgage obligations. Source: ABSNet bond data. The counts in our estimation sample (early vintages, prior to June 2005) are recorded in blue, while the numbers for late vintage tranches are illustrated in light grey.

Most bonds are rated AAA at origination (see Figure A.3).[17]

Subordination becomes steeper as the rating increases and for second lien/subprime deals. The tranching structure I observe lines up in general with the one Cordell et al. (2012) obtain from Intex data (see Table 7 for a comparison), apart from relatively thicker AAA tranches in our sample.[18]

[16] Some deals contain more than one structure, each structure giving rise to its own subordination waterfall. I source each structure separately and treat different structures as if they were different deals.

[17] ABSNet provides the Standard & Poor's (S&P) rating. When the security has no S&P rating I use the one issued by Fitch, which uses the same grading scale. Figure 1.2 shows the average subordination percentage by rating at origination.

[18] Rule 144A of the Securities Act of 1933 allows private companies to sell unregistered securities to qualified institutional buyers. Intex contains data on 144A deals, which are not in our sample.

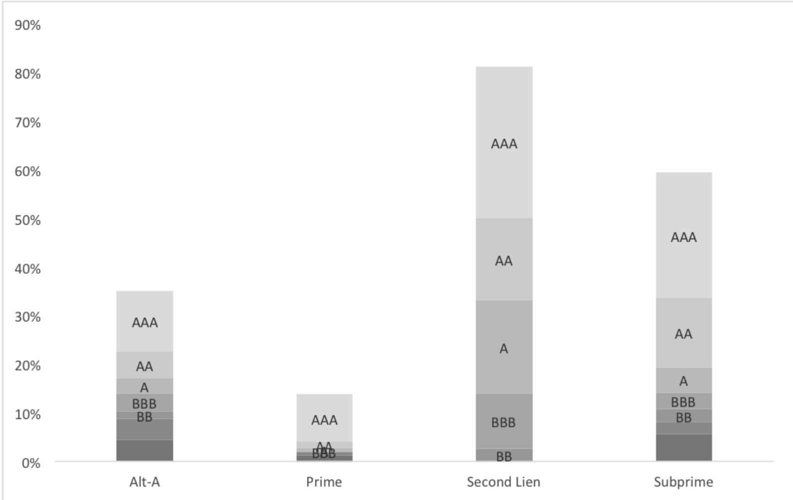


Figure 1.2: Deal structure. Source: ABSNet bond data. For our sample of early vintage deals, I look at the difference in subordination between tranches with consecutive S&P ratings. I then average the outcome by rating and asset type, aggregating at coarse grade level (see mapping in Table 15). This average difference is represented here, stacked by asset type.

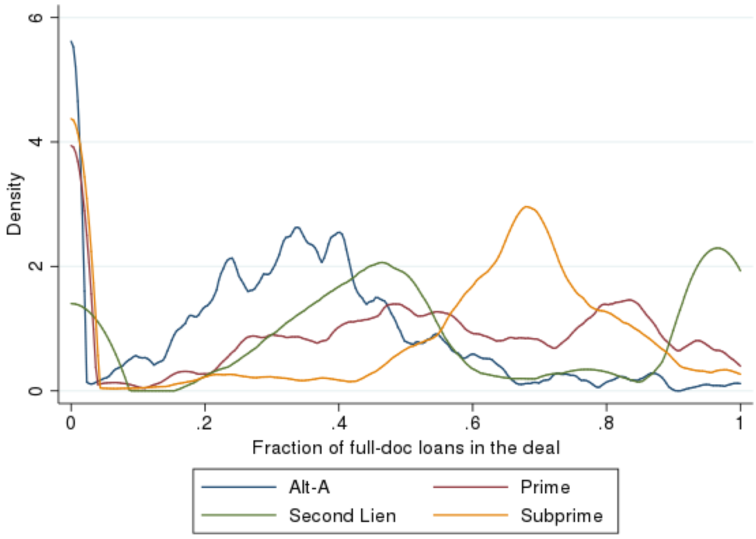
Changes in subordination percentage take place over the cycle, though mostly for subprime deals, reflecting the effect of defaults and prepayments. This is shown in Figure A.5, which depicts the point-in-time difference in average subordination between AAA and BBB tranches. While the difference remains close to constant for alt-A and prime deals, the difference rises for subprime deals in the running to the crisis, with a slight downward trend over time afterwards. In summary, among the tranche-level variables I use for the pricing model, i.e., price, coupon and subordination structure, the first two exhibit more cyclical variation than the latter.

I use origination and monthly performance data on the underlying loans by ABSNet. Loans are linked to their respective deals. I start with a sample of 6,453,799 loans of which 2,944,014 are originated by 2005. I have loan and borrower characteristics such as FICO score, owner occupancy, original loan amount and original LTV, which I will use in Section 3.1 to estimate default and prepayment hazard models.

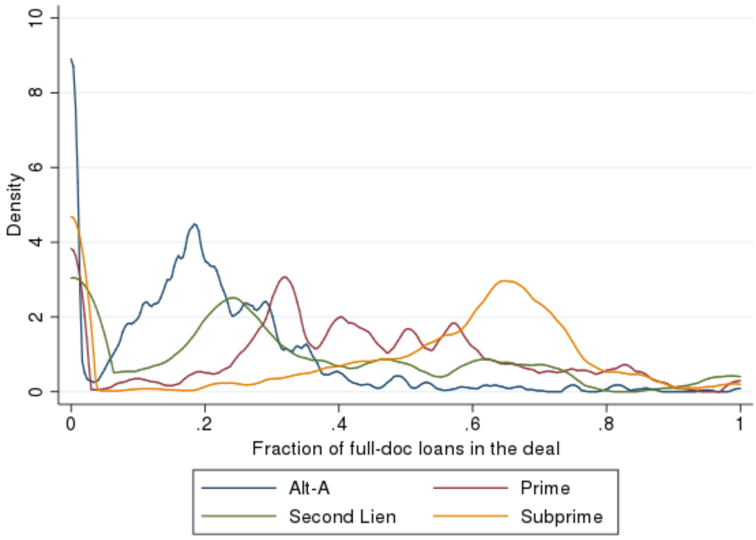
The loan data provides a documentation completeness indicator for each loan. This is categorized as full, limited, alternative or no documentation. Loans with full documentation provide verification of income as well as assets. Loans with limited documentation provide no information about borrower income but do provide some about their assets. "No-documentation" loans provide no information about income or assets. Figure 1.3 shows a distribution of the deal level share of loans with full documentation in our sample of vintages prior to June 2005. It suggests subprime loans were relatively better documented than alt-A deals, with densities peaking around 0.7 and 0.35 approximately. Prime deals show a higher dispersion in terms of documentation completeness.[19] I also check that the distribution of FICO scores does not shift upwards as documentation completeness improves (see Figure A.9) but in fact seems to shift downwards. Something similar happens

[19] In comparison, density plots on post-June 2005 issues suggest that documentation completeness deteriorated more among alt-A, second lien and prime deals in the running to the crisis relative to subprime ones.

to the distribution of LTVs over the documentation spectrum. The evidence thus suggests that low-doc is not a proxy for higher credit risk.



(a) Originated before June 2005



(b) Originated after June 2005

Figure 1.3: Kernel density plot of the distribution of full-documentation loans by deal asset type. For each deal I obtain the percentage of fully documented loans associated to it. The figure represents a kernel density plot of the distribution of deals along this measure. A separate plot on vintages later than June 2005 is provided for comparison.

Including cases of partial and alternative documentation, I assign a documentation score to each loan (no documentation=0; partial=0.1; alternative=0.3; full=1). In comparison Keys et al. (2010) use percentage of completeness, which is equivalent to excluding the intermediate values from my score. I average documentation scores into a deal level opacity index. The higher this index, the better the documentation on the underlying loans and the less opaque the deal. Figure A.6 presents the averages by asset type and vintage year. Note that alt-A deals, which are typically labeled as low-doc, can only be characterized as such from year 2000 onwards. The downward slope in Figure A.6 reflects the decline in lending standards in the running to the crisis observed on subprime loans by Dell'Ariccia, Igan, and Laeven (2012) and Keys et al. (2010).

Other data include dynamic covariates such as CBSA level home price indices from FHFA and interest rate data; I use the difference between the loan original interest rate from ABSNet and the original ten year Treasury rate from FRED. Using Treasury rates I also compute the coupon gap (the difference between the ten year rate at origination and the current ten year rate). From Bloomberg I extract bond contractual maturities and the weighted average life.

2 Empirical strategy

I estimate regressions of the form

$$\mathbb{1}\{downgrade_{i,2009}\} = \begin{cases} 1 \text{ if } \alpha + \beta X_{i0} + \eta_{rating_{i0}} + \varepsilon_i \geq 0 \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

where $\varepsilon_i \mid X_{i0}$ has a logit distribution. The vector X_{i0} contains bond attributes at origination such as price, subordination and coupon, controlling for deal vintage and tranche rating at origination.

Table 1 presents regression results for specification (1). Looking at the full sample in column (1), higher bond price is predictive of a lower probability of downgrade, and a higher percentage subordination has the same effect. Both are significant predictors of downgrades. A higher coupon significantly predicts lower likelihood downgrades (this counterintuitive result only holds for below-AAA bonds as I will show later). Overall Table 1 confirms the findings by Ashcraft et al. (2011) that bond prices contain information about bond performance which is not captured by the agency ratings.

To interact bond seniority with deal opacity I split the sample by value of the documentation index derived in Section 1 for each deal, using four buckets of width 0.25 each. The results, shown in columns (2)-(5) of Table 1, show that the effect most clearly driven by documentation quality is that of subordination percentage: the corresponding regression coefficient decreases monotonically from insignificant for the lowest documentation indices to negative and significant for the highest ones.

Comparing the subsample of AAA bonds and the rest, which I do in Table 2, I find evidence of this monotonicity of the regression coefficient on subordination percentage for both AAA bonds and the rest. So while the effect of price is always negative and significant and that of coupon depends on whether the bond is AAA at origination, the effect of subordination depends on the quality of documentation on the underlying loans as measured by our opacity index. In order to weigh the relative contribution of these different components I will price the bonds using a Gaussian copula model. The outcome of the pricing model, namely the implied correlation, works as a summary statistic of the variables considered so far.

	(1) All	(2) [0, 0.25)	(3) [0.25, 0.5)	(4) [0.5, 0.75)	(5) [0.75, 1]
Price	-0.0187***	-0.0159***	-0.0200***	-0.0110***	-0.0169***
	(0.00151)	(0.00606)	(0.00333)	(0.00267)	(0.00354)
Coupon	-0.142**	-0.123***	-0.0380	-0.117***	-0.0780*
	(0.0178)	(0.0640)	(0.0304)	(0.0441)	(0.0466)
Subordination	-3.130***	0.00163	-1.857***	-4.016***	-5.722***
	(0.268)	(0.864)	(0.657)	(0.489)	(0.943)
Observations	26,242	2,489	5,513	7,073	5,049
Rating at first transaction	Y	Y	Y	Y	Y
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y

Standard errors in parentheses

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

Table 1: Regression results from running logit specification (1) by maximum likelihood, controlling for vintage year, for vintages up to June 2005. The dependent variable is a dummy indicator for whether there was a downgrade by December 2009. Independent variables include price, subordination, coupon and coarse rating dummy indicator at the time of the first transaction. Each column presents the results on a subset of the data corresponding to the documentation index corresponding to the given deal. Errors are clustered at deal level.

The results from Table 2 confirm the results of Adelino (2009) that the information content in coupons is a priori less significant for AAA tranches. However, looking at all pricing measures (bond price, coupon and subordination) shows that the AAA distinction is capturing less about differences in information between investors than the documentation completeness index.

Higher subordination accurately predicts a lower probability of downgrade for deals based on deals backed by well-documented loans. The evidence shows a monotonic increase in significance and magnitude for the role of subordination as deals become less opaque. However, the results on bond coupon seem to suggest the opposite. In particular, higher coupons seem more predictive of a higher probability of downgrade especially for AAA bonds from low-doc deals. This raises a question about the overall direction of results, and motivates the use of a pricing model which I introduce in the following section. The outcome of the pricing model (the implied default correlation) will act as a summary statistic of the three pricing measures considered so far (price, coupon and subordination) and claridy the findings of the empirical strategy.

3 Implied correlation as a summary measure

In this section I derive a pricing equation where the key parameter is the loan default correlation. I use the asymptotic single risk factor model implemented by the IRB approach in Basel II. Credit risk in this basic framework has two components, systematic and idiosyncratic, where correlation is captured by joint dependence on the realization of the systematic factor (Crouhy, Galai, and Mark, 2000). I use a Large Homogeneous Gaussian Copula (LHGC) model (Brunne, 2006; D'Amato and

	(1)	(2)	(3)	(4)	(5)
	All	[0, 0.25)	[0.25, 0.5)	[0.5, 0.75)	[0.75, 1]
			AAA only		
Price	-0.0457***	-0.0352***	-0.0360***	-0.0347***	-0.0539***
	(0.00299)	(0.00900)	(0.00529)	(0.00632)	(0.0127)
Coupon	-0.0365	0.0508***	0.0546	0.0919	0.118*
	(0.0245)	(0.0161)	(0.0451)	(0.0575)	(0.0625)
Subordination	-3.944***	-0.0174	-2.774**	-2.014	-9.907***
	(0.565)	(1.622)	(1.229)	(1.881)	(3.612)
Observations	14,034	1,325	3,073	3,272	2,926
Rating at first transaction	N/A	N/A	N/A	N/A	N/A
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y
			Below AAA		
Price	-0.00932***	-0.0163**	-0.0129***	-0.00786***	-0.0113***
	(0.00149)	(0.00714)	(0.00371)	(0.00250)	(0.00358)
Coupon	-0.184***	-0.367***	-0.167***	-0.201***	-0.156***
	(0.0240)	(0.102)	(0.0475)	(0.0529)	(0.0603)
Subordination	-3.978***	-0.309	-2.648***	-4.501***	-4.193***
	(0.310)	(1.881)	(0.880)	(0.538)	(0.784)
Observations	12,206	1,038	2,248	3,757	2,111
Rating at first transaction	Y	Y	Y	Y	Y
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y

Standard errors in parentheses

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

Table 2: Regression results from running logit specification (1) by maximum likelihood, controlling for vintage year, for vintages up to June 2005. The dependent variable is a dummy indicator for whether there was a downgrade by December 2009. Independent variables include price, subordination, coupon and coarse rating dummy indicator at the time of the first transaction. Each column presents the results on a subset of the data corresponding to the documentation index corresponding to the given deal. Errors are clustered at deal level.

Gyntelberg, 2005; Duffie and Singleton, 2012; Elizalde, 2005; McGinty et al., 2004; Tzani and Polychronakos, 2008).[20][21]

In the LHGC setup two assumptions apply. The first one is that all loans in a given pool have the same probability of default PD and the same loss given default LGD, where both PD and LGD are known. The second assumption is homogeneity, which allows us to normalize individual loan size to one. Consider a pool of N mortgages. Default times $\tau = \tau_1, \ldots, \tau_N$ are correlated random variables, where correlation is captured by the loading on an exogenous, normally distributed factor S. In the one-factor model, the individual default probability is given by

$$p(s, T) := Pr(\tau \leq T | S = s) = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho}s}{\sqrt{1-\rho}}\right). \tag{2}$$

In equation (2), PD is the unconditional default probability and $\rho \in [0, 1)$, the implied default correlation, is the parameter to be estimated from the pricing model. Note that if $\rho = 0$ then the conditional distribution of defaults is exactly the unconditional distribution. As ρ approaches one the opposite happens, namely that the conditional distribution is fully determined by the value of s and so defaults are fully correlated.

Defaults are conditionally independent given the realization of the systematic factor S, i.e.

$$Pr(\tau_1 \leq t, \ldots, \tau_N \leq t | S = s) = \prod_{k=1}^{N} Pr(\tau_k \leq t | S = s)$$

which greatly simplifies pricing computations. Total losses from the pool accumulate over time to

$$l(t) = \frac{1}{N} \sum_{k=1}^{N} LGD \mathbb{1}_{(\tau_k \leq t)}.$$

Losses are distributed as they happen along the tranches from the deal, affecting first the junior ones and then the senior ones. A tranche's position in the deal is characterized by its attachment and detachment points a and b, where $0 \leq a < b \leq 1$. The tranche notional is a proportion $b - a$ of the total pool notional N. The losses borne by this tranche are given by

$$l_{[a,b]}(t) = \frac{[l(t) - a]^+ - [l(t) - b]^-}{b - a}.$$

The normality assumption yields the following estimate of expected losses within the $[a, b]$ tranche by payment date T_i:

$$E[l_{[a,b]}(T_i)] = \frac{1}{b-a} \int_{-\infty}^{\infty} \frac{e^{-s^2/2}}{\sqrt{2\pi}} \left([LGD\, p(s, T_i) - a]^+ - [LGD\, p(s, T_i) - b]^+\right) ds \tag{3}$$

[20]Following Li (2000) the Gaussian copula offered a conceptually simple framework for pricing structured securities, which made its use widespread. The model was also used for risk management, which Jarrow (2011) shows is inappropriate. The inappropriate use of copulas is blamed for a surge in investor overconfidence and eventually set the stage for the financial crisis in 2007. See Felix Salmon, *Recipe for Disaster: the Formula that Killed Wall Street* (`https://www.wired.com/2009/02/wp-quant/`).

[21]Duffie and Gârleanu (2001) and Coval et al. (2009a) look at the sensitivity of expected recovery to default correlation in a Gaussian copula model. Figure A.2 replicates the exercise in Coval et al. (2009a) by plotting expected recovery for each value of ρ, normalized by the value corresponding to $\rho = 20\%$.

Using payment dates $0 < T_1 < \cdots < T_m = T$ (where T is the maturity of the security), write

$$\frac{V_{[a,b]}}{N(b-a)} = c\sum_{i=1}^{m} B(0,T_i)\Delta(T_{i-1},T_i)(1-l_{[a,b]}(T_i)). \tag{4}$$

Formula (4) equates security value to the sum of two terms: the discounted cash flows from coupon payments and the residual value (after accounting for defaults) of principal outstanding. Here $B(t_1,t_2)$ discounts a payoff at t_2 to t_1, c denotes the tranche coupon and $\Delta(T_{i-1},T_i)$ is the time difference between two payment dates (for mortgage bonds I use $\Delta(T_{i-1},T_i) \equiv 1/12$). The bond pricing equation is then $pN(b-a) = E[V_{[a,b]}]$. Writing $e_i^{[a,b]} = E[1 - l_{[a,b]}(T_i)]$ the following holds at origination:[22]

$$p_0 = c\sum_{i=1}^{m} B(0,T_i)\Delta(T_{i-1},T_i)e_i^{[a,b]} \tag{5}$$

The pool is exposed to prepayment risk. As prepayments take place the coupon rate is applied to the balance outstanding, while the prepaid amount is allocated across tranches. In the absence of prospectus information about the order of the cashflows for each specific deal, I make the simplifying assumption that prepayments are uniformly distributed across tranches.[23] This yields

$$p_t = \sum_{i=t+1}^{m} B(t,T_i)e_i^{[a,b]} \prod_{k=t+1}^{i-1}(1-SMM_k)\left(\underbrace{c\Delta(T_{i-1},T_i)(1-SMM_i)}_{\text{coupon payment}} + \underbrace{SMM_i}_{\text{prepaid principal}}\right) \tag{6}$$

where SMM_k is the single month mortality rate at time k, and is given by the prepayment speed model. Given the unconditional default probability PD, the recovery rate RR and prepayment rate SMM_k, pricing equation (6) pins down a value of ρ, the market estimate of default correlation for the given pool of loans. Note that expression (2) is only defined for $\rho \in [0,1)$ and thus the existence of a solution to equation (6) is not guaranteed to fall within the unit interval for an arbitrary choice of p and c. So instead of solving equation (6), I solve the minimization problem

$$min_{\rho\in[0,1)} \left| p_t - \sum_{i=t+1}^{m} B(t,T_i)e_i^{[a,b]} \prod_{k=t+1}^{i-1}(1-SMM_k)\left(c\Delta(T_{i-1},T_i)(1-SMM_i) + SMM_i\right)\right| \tag{7}$$

Solving equation 7 gives the market estimate of default correlations that I now compute on the panel of security prices.

[22] Note that formula (6) implies that default occurs immediately after the following period payment.

[23] Duffie and Singleton (2012) discuss two prioritization schemes, uniform and fast. Both imply prepayment cash flows are sequential over seniorities. In the absence of deal-level information about the allocation of prepayments, I assume no prioritization. However, many subprime-backed MBS structures had prepayment cashflows going first to the senior tranches, then to junior ones. The purpose is to give the senior tranches a higher credit quality. Because of prepayment penalties, this a priori loss of precision on prepayment cashflow allocation is actually not significant.

3.1 Model parameters: default and prepayment

The pricing model focuses on expected losses (EL). Equation (3) uses the identity $EL = PD \times LGD$, which implies both factors must be based on the same definition of default. Since recoveries in our data are based on liquidated values, I use of liquidation as the default event. Figure A.7 shows an increase in cumulative liquidation rates in the running to the crisis, though the trend is only upward sloping from 2005 vintages onward. Alt-A default rates were roughly half those of subprime deals until early 2005, when both rates soared in the running to the crisis. By 2008, securitization issuance have collapsed. One difference is that while the 90+ delinquency rate they report remains lower for alt-A deals, I find that their cumulative liquidation rate, initially similar to that of prime deals, caught up with that of subprime ones in the running to the crisis.

From loan loss event data I compute LGDs at the deal level (see Figure A.8 for a count of observations by vintage and asset type). Figure A.11 shows that LGD was nearly monotonically increasing from 1990 to 2007, except for a peak in 1996, so that investors may have been adjusting their expectations of LGD over the cycle. However, for LGDs to be computed the full post-workout must have been observed, which usually takes a substantial observation time after default.[24] I apply the common assumption of constant LGD using the long run average on our sample of 60% that is also typically assumed in the literature (Altman, 2006; Brunne, 2006; Coval et al., 2009b; Hull and White, 2004, 2008).

Investors' beliefs about default rates are elicited with a regression model establishing the likelihood of default as a function of loan covariates and estimated on default history. For prepayment speeds I use a proportional hazard model over the first 60 months of the life of the loan. I use a separable hazard model, treating observations representing default as censored as in Palmer (2015) and Liu (2016). Default and prepayment are termination reasons happening at a random time τ^{term}, for termination cause $term \in \{default, prepayment\}$. The intensity of τ is given by equation (8).

$$\lambda_i^{term}(t) = \lim_{\epsilon \to 0} \frac{Pr_i(t - \epsilon < \tau^{term} \leq t \mid t - \epsilon < \tau^{term},\ X)}{\epsilon}. \tag{8}$$

In equation 8 i denotes loan and t denotes time after origination. The density function in equation (8) is modeled as

$$\frac{\lambda_i^{term}(t)}{\lambda_0^{term}(t)} = \exp(X'_{it}\beta^{term}) \tag{9}$$

where $\lambda_0^{term}(t)$is the baseline hazard function that depends only on the time since origination t. Covariates in X_{it} include loan attributes like loan amount, coupon gap relative to 10 year constant maturity Treasury, LTV and prepayment penalty indicator. It also includes borrower characteristics like FICO score and owner occupancy, as well as variables at the CBSA level such as home price appreciation and unemployment rate.

[24]Models of LGD with incomplete workouts, like Rapisarda and Echeverry (2013), were far from the norm, especially in the running to the crisis.

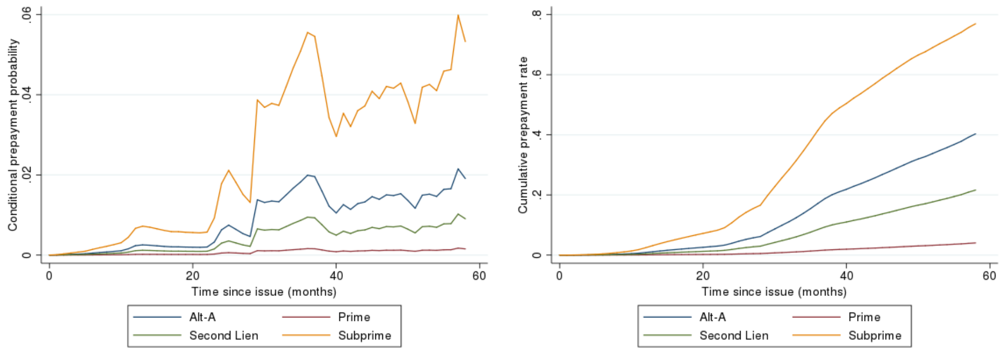
To estimate the continuous time model specified in equation 8 on discrete time data, I accumulate the intensity process λ over time intervals per equation (10).

$$Pr_i(t < \tau^{term} \mid t-1 < \tau^{term}) = \exp\left(-\int_{t-1}^{t} \lambda_i^{term}(u)du\right) \tag{10}$$

This leads to the complementary log-log specification

$$Pr_i(t < \tau^{term} \mid t-1 < \tau^{term}) = \exp(-\exp(X_{it}'\beta^{term})\lambda_0^{term}(t)). \tag{11}$$

I estimate specification (11) with months since origination fixed effects to obtain the hazard functions. I document the results in Table 9 and plot the resulting prepayment rates on Figure 3.1. I find that adjustable rate mortgages are both more likely to default and to prepay than fixed rate types. Subprime loans are the asset type most likely to default. In terms of prepayment hazard, there is no significant difference across asset types other than prime loans being less subject to prepayment than other types.



(a) Prepayment hazard

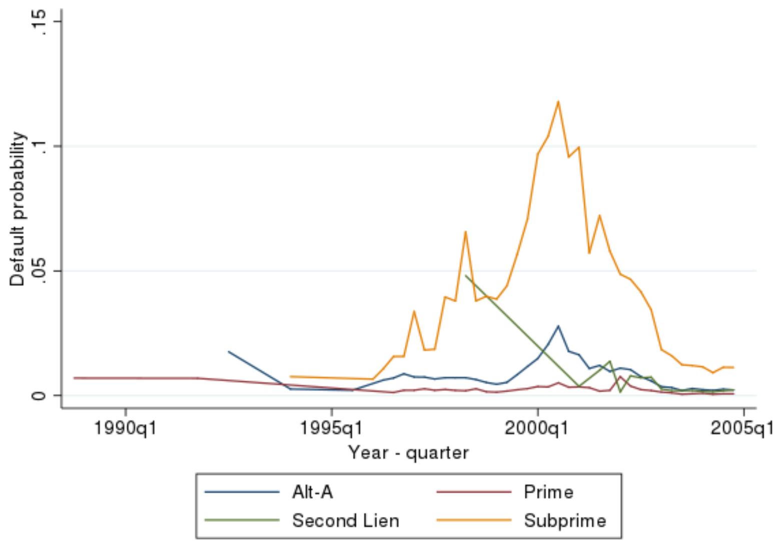
(b) Cumulative prepayment rate

Figure 3.1: Marginal and cumulative prepayment rates implied from the model (11), as summarized in Table 10. Using loan covariates at origination, prepayment hazard rates are computed at the loan level. Averages are computed by asset type and month after origination and plotted here.

I now compare the results from Table 9 with the ones obtained by Liu (2016), who uses the same model to estimate default and prepayment hazard rates on agency loans.[25] On one hand, we document the same sign for the effect of FICO score, the difference between the original loan interest rate and the original 10 year rate, and the unemployment rate. In terms of default hazard we find similar effects of LTV and home price appreciation.

[25] Adding late originations, i.e., up to 2007, I find a number of similarities. The main difference that arises is that now subprime loans can be seen to be prepaying significantly more than other types and significantly more than early vintages. This suggests that the link between subprime origination and home prices through prepayments was specific to the pre-crisis boom rather than a design feature of subprime loans. Macroeconomic factors, such as home price appreciation and unemployment, exhibit a similar effect on defaults and prepayments when adding late vintages. Instead, for coupon gap there is a change compared to the early sample. The coupon gap, i.e., the change in 10 year rates between origination and present, reflects stronger incentives to refinance. The expectation is that this leads to a higher probability of prepayment and a lower probability of default, which I see once I add late cohorts, but not in the early sample.

On the other hand we document a few differences, mostly about the link between home prices and prepayment rates. Liu (2016) finds that home price appreciation increases prepayment hazard while I find the opposite. Similarly, he finds that higher LTV reduces prepayment hazard while I find no clear link. As discussed by Gorton (2009), while the prepayment option is always valuable to a conforming loan borrower, i.e., when house prices rise borrowers build up equity, for subprime loans lenders hold an implicit option to benefit from house price changes. Table 9 corroborates the deterring effect of prepayment penalties.



(a) Probability of default

Figure 3.2: Probability of default implied from the complementary log-log model, estimates of which are in Table 10. Using loan covariates at origination, default probabilities are computed at the loan level. Averages are computed by asset type and month after origination, and plotted here.

The break-even probabilities of a crisis computed by Beltran, Cordell, and Thomas (2017) from CDO prices show a decrease from pre-2006 cohorts to later ones, which suggests a relatively high risk premium was charged in early cohorts. Though there are no studies on risk premia in mortgage markets, I can benchmark the ones I use against the corporate market. Berndt, Douglas, Duffie, Ferguson, and Schranz (2005) derive actual and risk-neutral probabilities from CDS market quotes. They find that the corresponding coverage factor, i.e. the ratio of risk neutral to the real probability, oscillates between 1.5 and 3.5 from 2002 to 2003. I use a coverage ratio of 3.[26]

Using the model in Table 10 I predict prepayment hazards and default probabilities at the loan level and average them at the deal level, so that both the default probability and the hazard rate are estimated deal by deal. For the prepayment hazard, I use the full schedule in order to estimate the average prepayment speed for the given deal over the first 60 months. As Figure 3.1 illustrates, subprime loans have the highest prepayment rates, followed by alt-A loans. They also have the

[26] Heynderickx, Cariboni, Schoutens, and Smits (2016) quantify coverage factors from CDS quotes of European corporates and find that they range between 1.27 for Caa (Moody's) ratings to 13.51 for Aaa ones on pre-crisis data. Like Heynderickx et al. (2016), Denzler, Dacorogna, Müller, and McNeil (2006) argue that risk spreads exhibit a scaling law, whereby risk premia are decreasing in the probability of default. The results in Table 8 imply coverage ratios between 2.03 for subprime deals and 3.27 for Alt-A ones, in line with the literature.

highest default probabilities, as shown in Figure 3.2. I use the model-implied PDs from Table 10 (see Figure 3.2) and include them as controls.

I source contractual maturity from Bloomberg, which for most bonds is close to 30 years. These figures are high compared with realized maturity, defined as the first observation where the tranche balance is zero, the difference being 16.27 years on average on a sample of 5,507 tranches. Figure A.12 also suggests that bonds do not live that long on average. Adelino (2009) uses weighted average life (WAL) instead of contract maturity, which is closer to the realized maturity. I also source WAL for a sample of our loans where I could find it, but found that WALs are low compared to realized maturities in the data (the average difference is 6.77 years on a sample of 16,894 tranches, see Figure A.13 for a further breakdown of the difference). In the final model I use contractual maturity, relying on the prepayment speed model to achieve an accurate reduction of tranche balance over time.

The model in Table 10 incorporates all observations. In reality, agents' expectations about default evolve over time, especially as the business cycle unfolds. As an example, take home prices, which fluctuate over the cycle. As Table 11 shows, home price appreciation is the variable whose effect on defaults changes the most over the cycle. In particular, the negative relationship between price appreciation and defaults documented in Table 10 is an average between the positive effect recorded in the early years of the sample (up to 2002) and the negative effect in subsequent years. I expect this to have a modest impact on the pricing model, given that over the times of the prices I am interested in, mostly 2004 and 2005, the coefficients in Table 11 tend to be close to those in Table 10.

Loan performance data gives a basis for consensus about probability of default, loss given default and prepayment speed. As discussed in the introduction, default correlation is the parameter on which market participants are more likely to disagree [27]. Seeing these disagreements as the starting point for differential information, I will use the pricing model from Section 2 to generate a summary statistic that signals future downgrades, and study how asset opacity drives the informativeness of the signal as opposed to bond seniority.

3.2 Deriving implied default correlations from CMO prices

For a given bond I compute its implied correlation ρ using the coupon rate c, market price p, attachment point and detachment point $a \geq 0$ and $a < b \leq 1$. The probability of default and prepayment speed are estimated per Section 3.1. I use the discount rate $r = 4.27\%$, the average 10-year constant maturity treasury (annual) rate between 1995 and 2015. The numerical computations of loss probability are evaluated using a trapezoidal rule, which Brunne (2006) deems faster than Gauss-Legendre and Gauss-Hermite methods. Figure 3.3 provides a summary of observations.

Implied correlations were uniformly increasing across seniorities in the running to the crisis, as can be seen from Figure 3.3. The overall distribution of individual outcomes is strongly bimodal (see

[27] "Currently, the weakest link in the risk measurement and pricing of CDOs is the modeling of default correlation." Duffie (2008)

Figure A.14). Tzani and Polychronakos (2008) find that in CDS markets model correlations would sometimes have had to exceed 100% in order to price supersenior tranches, which is suggested by Figure A.15. The extreme values suggest there is a role for market incompleteness as in Andreoli et al. (2016) and Stanton and Wallace (2011).

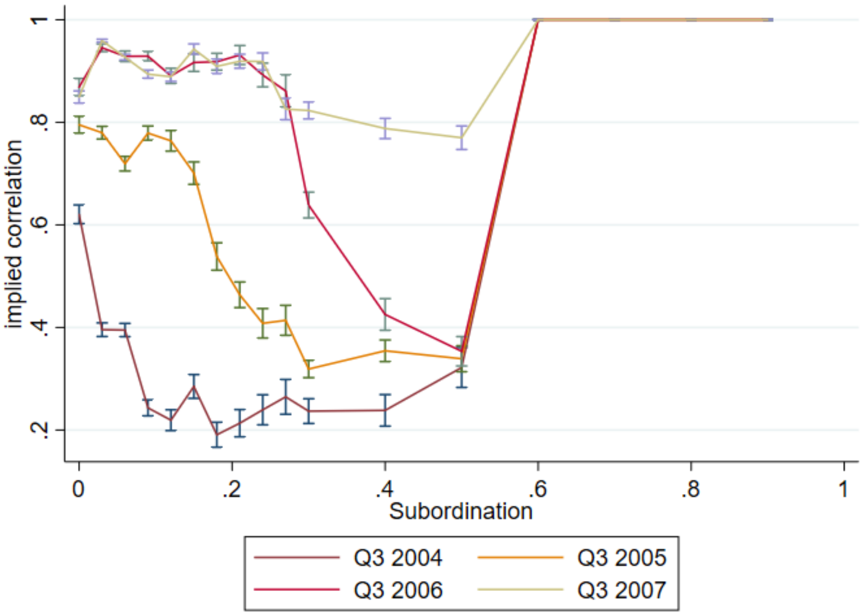


Figure 3.3: Average correlation plotted against tranche subordination percentage on four given quarters. I use the sample of bonds originated prior to June 2005. Subordinations are assigned to 10 bins. Within each subordination bin I plot the average correlation, along with vertical whiskers representing the standard error of the average.

By the third quarter of 2006, implied correlations had reached crisis levels except for upper seniorities. Taking into account the highest seniorities, Figure 3.3 also shows evidence of a correlation smile.[28]

Implied correlations increased further during the crisis. I observe a statistically significant increase, from 0.89 in September 2007 to 0.93 in February 2009.[29] Using also a one factor Gaussian copula model, Buzková and Teplỳ (2012) analyze prices of the 5-year, North American investment grade CDX (V3) index. They report that for synthetic CDOs, implied correlations see a large increase on average, from 0.15 to 0.55. In terms of seniorities, the difference observed by Buzková and Teplỳ

[28]The correlation smile is an artifact from the compound correlation method (O'Kane and Livesey, 2004). A method that is used to derive increasing correlations is the base correlation, which is computed as follows: let the attachment points in the full waterfall be given by $(b_1, \ldots, b_n)$, where $b_n = 1$. First, solve 6 for the tranche $[0, b_k]$, $k = 1 \ldots n$. This gives an estimate of $e_i^{[0,b_k]}$. Using the identity

$$(b-a)e_i^{[a,b]} = be_i^{[0,b]} - ae_i^{[0,a]},$$

the expected losses in tranche $[a, b]$ can be sequentially computed along the waterfall: once the $[b_{k-1}, b_k]$ tranche has been priced, the following one can be priced using

$$(b_{k+1} - b_k)e_i^{[b_k,b_{k+1}]} = b_{k+1}e_i^{[0,b_{k+1}]} - b_k e_i^{[0,b_k]}.$$

Base correlations price all tranches in a deal simultaneously. Because simultaneous transactions do not coincide frequently in time, I use the compound correlations and price each tranche separately.

[29]Breaking the change by asset type I document an increase for alt-A and subprime deals, from 0.81 to 0.97 and from 0.85 to 0.89 respectively (at 99% confidence), but no change for prime tranches (0.93). The upward adjustment was thus the largest among alt-A issues (see Figure A.16).

(2012) over the crisis is mainly driven by mezzanine tranches (7%-10% and 10%-15%). Figure 3.3 also suggests the increase in correlation is larger among intermediate seniorities.

Ratings were mostly stagnant in the running to the crisis, especially for AAA tranches, in comparison with default correlations (see Figure A.17). BBB tranches even see an improvement in ratings ahead of the crisis while correlations are increasing (except for subprime deals, which see both downwards and upwards changes. The sharpness of rating downgrades suggests this is a concern for BBB tranches. Griffin and Tang (2012) argue that AAA ratings were inflated in CDO securities, with optimistic ratings applied to a large share of bonds issued. Because CDOs are mainly composed of mezzanine CMO tranches, a potential channel for rating inflation in AAA CDO tranches is rating upgrades of BBB tranches.[30] This gives a possible channel for ratings inflation over the cycle other than that of boom time originations.

The graphic evidence presented is suggestive that market prices lead ratings. Whether this means investors learn faster than ratings agencies will be assessed by whether ratings are sufficient for implied correlations regarding bond outcomes, which I consider in the next section.

4 The information content of implied correlations

This section will focus on assessing the findings of section 2 using implied correlations as a summary statistic for prices, coupons and subordinations. I start with the same logit specification as equation 1, using implied correlation as the independent variable. More specifically,

$$\mathbb{1}\{downgrade_{i,2009}\} = \begin{cases} 1 \text{ if } \alpha + \beta\rho_{i0} + \eta_{rating_{i0}} + \varepsilon_i \geq 0 \\ 0 \text{ otherwise} \end{cases} \tag{12}$$

where $\varepsilon_i \mid \rho_{i0}$ has a logit distribution. The independent variable of interest is the implied correlation at the first transaction, ρ_{i0}.

As shown in Figure A.2, expected losses for the senior tranche are monotonically increasing in default correlation ρ while the opposite happens for the junior tranche. The mezzanine tranche behaves like a senior tranche for low correlations and like a junior tranche for high ones (Ashcraft and Schuermann, 2008; Duffie and Gârleanu, 2001; Duffie, 2008). Due to their subordination percentages, the bonds I observe behave more like a mezzanine or a senior tranche, so that a higher implied correlation should predict a more likely downgrade. I control for rating at origination using dummy indicators and for vintage year. Also, I cluster standard errors in all tests at the deal level to control for the fact that classes in the same deal are often (down)graded at the same time.

The results in Table 3 reconstruct those of Table 1, using implied correlation as a summary statistic of price, coupon and subordination. Ratings at origination are not statistically sufficient for implied correlations in predicting subsequent bond downgrades, similar to the results in Table 1 where rating is not sufficient for price, coupon and subordination. This is shown in column (1). Just as

[30] Stanton and Wallace (2018) find that upgrades were more pervasive among commercial mortgage-backed securities than among RMBS, the upgrades being attributed to regulatory capital arbitrage.

a lower price is reflective of more risk and thus of more likely downgrades, the implied correlation follows a similar intuition. I find a positive, significant coefficient, so that higher implied correlation increases the likelihood of downgrades.

Again I arrange the sample by bins decreasing in the level of opacity and present the result in columns (2)-(5). I find a ranking along the opacity index similar to the one discussed in Section 2, whereby the coefficient on implied correlations is monotonically increasing in the value of the opacity index, from insignificant at 10% for tranches below 0.25 to positive and significant at 1% for tranches above 0.75.

	(1) All	(2) [0, 0.25)	(3) [0.25, 0.5)	(4) [0.5, 0.75)	(5) [0.75, 1]
Correlation at first transaction	0.414***	0.243	0.605***	0.476***	0.569***
	(0.0629)	(0.250)	(0.200)	(0.102)	(0.135)
Model-implied PD	2.294**	0.381	13.60	4.331	4.225*
	(0.922)	(1.675)	(10.51)	(3.000)	(2.521)
Observations	28,991	2,723	6,285	7,808	5,565
Rating at first transaction	Y	Y	Y	Y	Y
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Regression results from running logit specification (12) by maximum likelihood, controlling for vintage year (vintages up to June 2005) and model-implied probability of default as estimated in subsection 3.1. The dependent variable is a dummy indicator for whether there was a downgrade by December 2009. Independent variables include implied correlation and coarse rating dummy indicator at the time of the first transaction. Each column presents the results on a subset of the data corresponding to the value of the documentation index corresponding to the given deal. Errors are clustered at deal level.

Table 4 breaks down the result between bonds initially rated AAA and the rest. While the coefficient for correlation at first transaction remains significant for grades below AAA, implied correlations seem to have no predictive power in terms of bond downgrades, similar to the findings in Adelino (2009). Comparing AAA tranches and others, a similar pattern across the two rating categories emerges; as opacity falls, implied correlation becomes more informative about subsequent bond downgrades. Note that on the left hand side extreme, i.e., for deals close to "no-doc", both AAA and non-AAA implied correlations are uninformative. On the opposite side of the asset opacity distribution, both are informative. For tranches where the documentation index is above 0.5 implied correlation is predictive of bond downgrades. Seen together, the results suggest that uninformed investors are not so much those in AAA tranches as those subject to low-doc deals. They also suggest that the agency problem between junior and senior investors obtains for intermediate levels of asset opacity.

Figure 3.3 shows that for the most data relevant region (subordinations up to 0.4) the relationship between subordination and implied correlation is downward sloping. Higher seniorities see lower implied correlations, especially for early transactions. Accordingly, the results in Tables 3 and 4 echo those of Tables 1 and 2. However, the overall relationship is neither trivial (because the link

	(1) All	(2) [0, 0.25)	(3) [0.25, 0.5)	(4) [0.5, 0.75)	(5) [0.75, 1]
			AAA only		
Correlation at first transaction	0.299	1.018	0.430	1.647***	0.842***
	(0.201)	(0.703)	(0.599)	(0.627)	(0.321)
Model-implied PD	4.308	47.95	-13.93	12.91***	3.301
	(3.648)	(48.60)	(45.34)	(4.648)	(2.970)
Observations	16,618	1,529	3,765	3,975	3,429
Rating at first transaction	N/A	N/A	N/A	N/A	N/A
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y
			Below AAA		
Correlation at first transaction	0.268***	0.0485	0.370**	0.314***	0.353**
	(0.0644)	(0.283)	(0.155)	(0.109)	(0.158)
Model-implied PD	1.503	-2.323	26.14**	1.906	5.083
	(1.023)	(2.661)	(10.85)	(2.607)	(3.180)
Observations	12,371	1,045	2,289	3,787	2,124
Rating at first transaction	Y	Y	Y	Y	Y
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y

Standard errors in parentheses

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

Table 4: Regression results from running logit specification (12) by maximum likelihood, controlling for vintage year (vintages up to June 2005) and model-implied probability of default as estimated in subsection 3.1. The dependent variable is a dummy indicator for whether there was a downgrade by December 2009. Independent variables include implied correlation and coarse rating dummy indicator at the time of the first transaction. Each column presents the results on a subset of the data corresponding to the value of the documentation index corresponding to the given deal. Errors are clustered at deal level.

between subordination and implied correlation cannot be derived in closed-form, and in particular is not linear) not monotonic (because of the presence of correlation smiles). These observations support the use of implied correlations as non-redundant summary measure of the pricing measures discussed in Section 2.

A going concern is that ratings do not appear to be statistically sufficient sufficient simply reflects the finer granularity of price and subordination relative to a rating notch. As a robustness check, I run the same set of regressions as before, using the deal level average correlation instead of the individual one. The results, shown in column (1) of Table 12, show that correlation loses its predictive power overall when averaged across the deal. The average at rating level retains some predictive power about subsequent downgrades. However, breaking the results down by opacity index in columns (2)-(5), I find the same monotonicity in predictiveness of implied correlations, though the coefficient becomes significant only for the highest values of the documentation index. However, once I break down the results between AAA and sub-AAA tranches in Table 13, only AAA tranche implied correlations are predictive but only in the highest documentation index values. In all these tables, the monotonicity property observed before is best represented by the subordination amount.

Low-doc assets should in principle require a form of compensation for the lemons problem; all else constant, a sophisticated investor requires more subordination when the underlying asset is opaque. Instead, Skreta and Veldkamp (2009) predict that rating inflation can be worse when assessing the true value of the asset is difficult and so ratings are noisy. For their result to hold, investors must be unable to infer the rating selection bias. Similarly in our case, investors who are unaware of the deficiency in documentation are more likely subject to inflated ratings. Table 14 provides evidence that AAA share at origination is increasing in opacity, controlling for the model-implied probability of default.[31] Because the equity tranche size is also likely to be increasing in opacity, as shown by Begley and Purnanandam (2017), the larger AAA tranche might be precisely an equilibrium response to the strong quality signal. This aspect of security design deserves further attention.

5 Summary and discussion

This paper weighs the relative importance of two key information frictions that take place between the private investor and the issuer of mortgage-backed securities. Though there is a role for what the literature calls investor unsophistication, proxied by a AAA rating at origination, asset opacity predominates as a friction. I measure opacity using a deal-level index of documentation completeness. I observe less of a differential in information content across seniorities than across low-doc assets and full-doc ones. The latter exhibit better information content across the rating spectrum. In particular, AAA prices are no less predictive than the rest when the bond comes from a deal with a high standard of documentation. In addition, the results suggest that errors in computing default correlations in the running to the crisis were not a problem of AAA investors per se, but rather a problem of low-doc investors.

[31]Note that Begley and Purnanandam (2017) find the opposite result on a smaller sample of deals, but the significance disappears once controlling for risk attributes.

The results suggest that a regulation intervention focusing on the agency problem, such as risk retention in the form of skin in the game, does not address the main information friction. Therefore it should be complemented by market transparency initiatives aiming towards better documentation on the underlying loans. To the extent that incomplete information is easier to tackle than differential sophistication, such transparency initiatives can be an effective instrument to help price informativeness in private label securitization markets. A blanket reduction of low-doc lending in mortgage markets could address the widespread increase of fraud that was observed during the boom, but would likely result in rationing against self-employed borrowers as shown by AMBROSE et al. (2016). My results, which focus on low-doc bonds instead of the underlying low-doc loans, have no direct implications about the issuance of the loans. Instead, they suggest that the securitization of these loans results in prices that are less informative than those arising from other asset classes. Linking skin in the game requirements in securitization deals to documentation completeness might give an incentive to weed out fraudulent alt-A issuance while preserving credit issuance to the self-employed.

The literature has largely attributed default clustering to joint dependence on a systematic shock (Bisias, Flood, Lo, and Valavanis, 2012; Chan-Lau, Espinosa, Giesecke, and Sole, 2009; Bullard, Neely, Wheelock, et al., 2009; Khandani, Lo, and Merton, 2013). Using a Gaussian copula I find that implied correlations are large in subprime deals compared to other asset classes, which reflects a design feature that made subprime loans jointly dependent on house price indices. Recent literature distinguishes two additional sources of default clustering: unobserved frailty (Duffie et al., 2009; Kau, Keenan, and Li, 2011; Griffin and Nickerson, 2016) and contagion (see appendix B).[32] In particular Azizpour, Giesecke, and Schwenkler (2016); Gupta (2016); and Sirignano, Sadhwani, and Giesecke (2016) suggest the contagion channel is important. Whether these additional sources of default clustering were priced remains an open question.

[32]For a review of recent literature on contagion see Bai, Collin-Dufresne, Goldstein, and Helwege (2015).

A Supporting tables and figures

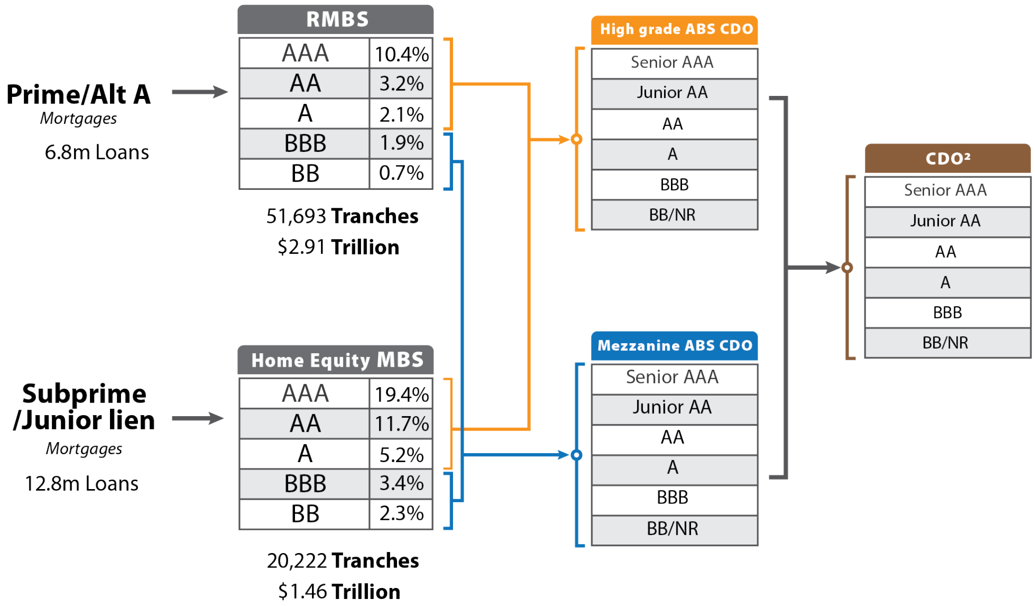


Figure A.1: Diagram: from loans to RMBS CMO, from CMO to CDO, from CDO to CDO^2. Details are reported on the total number of loans recorded by ABSNet, the universe of securities issued and the average subordination percentage by Standard & Poor's rating, as explained in Section 1

Year	ABSNet sample		Adelino (2009)	
	Origination ($bn)	Count	Origination ($bn)	Count
≤2002	319.3	5,438		
2003	470.5	10,120	496.5	8,574
2004	677.4	12,519	767.3	11,460
2005	904.5	16,684	1,058.5	17,135
2006	1,038.0	15,022	1,080.4	18,206
2007	939.4	11,716	802.1	12,037
≥2008	31.2	177		
Total	4,380.3	71,676	4,204.8	67,412

Table 5: Origination amounts and counts at origination, by vintage year, compared to the sample in Adelino (2009).

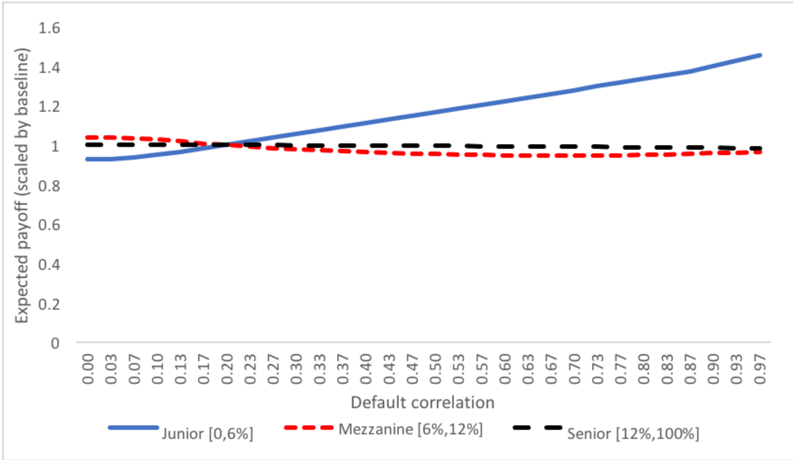
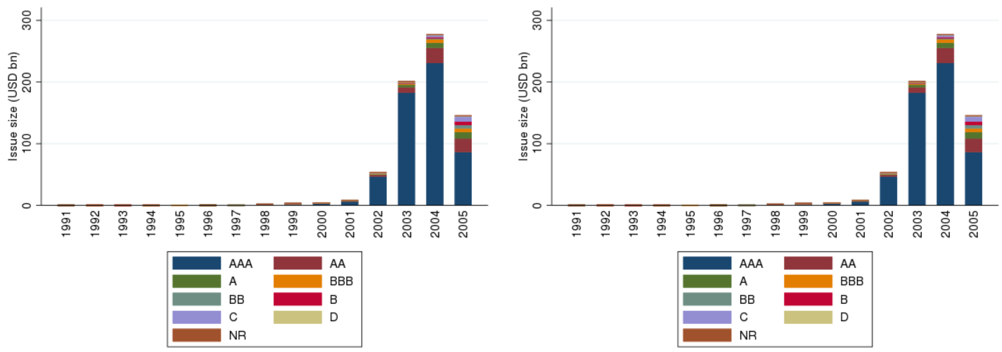


Figure A.2: Sensitivity of a simulated CMO structure to default correlations. We plot the expected payoff within a given tranche against the value of the underlying correlation ρ (parameters are PD=5% and LGD=50% as in Coval et al. (2009a)). The results are normalized by baseline estimate, based on the same parameters and a correlation $\rho = 20\%$. No prepayments are incorporated (i.e. SMM=0%) for comparability of outcomes.

Asset type	After Jun-05		Before Jun-05	
	Origination ($bn)	Count	Origination ($bn)	Count
Alt-A	1,179.0	16,837	557.7	11,000
Prime	621.7	9,097	557.9	14,759
Second Lien	64.7	478	19.0	408
Subprime	660.0	9,811	720.2	9,525
Total	2,525.4	36,223	1,854.8	35,692

Table 6: Issued amounts and counts by asset type.



(a) Number of tranches

(b) Total issue

Figure A.3: Number of tranches and amount issued by vintage year for private label collateralized mortgage obligations. Source: ABSNet bond data. For our sample of early vintages (prior to June 2005) we provide the distribution by (coarse, see Table 15) initial rating.

rating	Our sample Prime/Alt-A	Our sample Second Lien/Subprime	Cordell et al. (2012) Prime/Alt-A	Cordell et al. (2012) Second Lien/Subprime
AAA	10.8%	25.7%	6%	23%
AA	3.4%	14.3%	3%	13%
A	3.0%	5.9%	2%	8%
BBB	2.9%	4.0%	1%	4%

Table 7: Subordination percentage by tranche rating - comparison. The figures computed using ABSNet data are derived by aggregating the subordination percentages at orgination as given in Table 1.2. Our sample contains only early vintages (prior to June 2005) while Cordell et al. (2012) use late vintages as well.

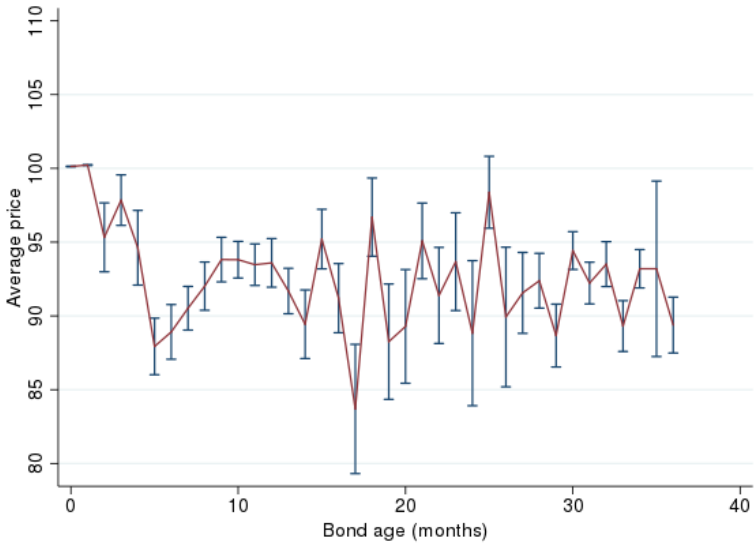


Figure A.4: Average tranche price by age of the bond in months. For our sample of bonds originated in 2004 and 2005 we compute the average price by the time elapsed (in months) since the bond issue. Vertical whiskers show the standard errors.

Asset type	(1) Early vintages	(2) Late vintages	(3) Model PD
Alt-A	7.5%	19.5%	24.5%
Prime	2.3%	6.6%	6.4%
Second Lien	7.2%	25.8%	21.1%
Subprime	14.8%	30.5%	30%
Observations	4,060,698	631,793	2,112

Table 8: Liquidation rates from the loan sample, and PD used for baseline estimation. Column (1) calculates the percentage of loans linked to early vintage deals (before June 2005) that are liquidated. Column (2) calculates the same ratio for late vintage loans. Column (3) shows the PD parameters used for the pricing model, calculated as the average of the deal level liquidation rates for both early and late deals.

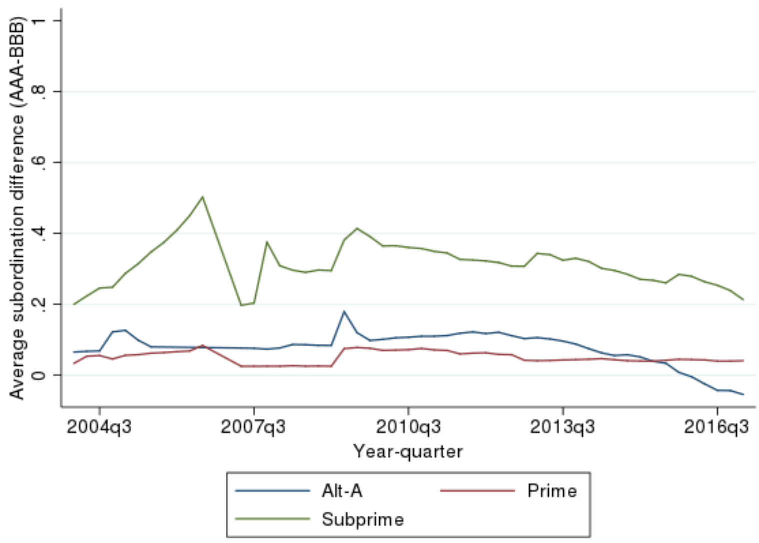


Figure A.5: Average subordination difference between AAA and BBB bonds. Source: ABSNet bond data.The figure presents the difference between the average AAA and average BBB subordination over trading time (for early vintages, prior to June 2005) using the rating at the given trading time. The difference is computed by asset type.

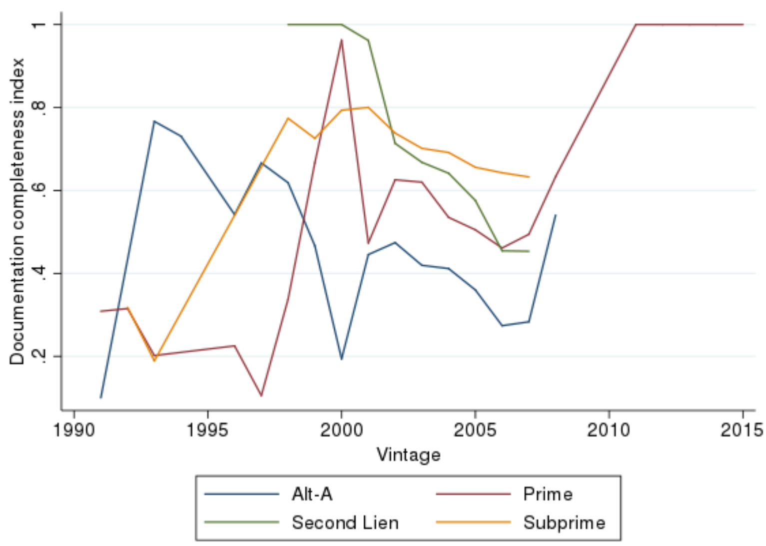


Figure A.6: Average documentation index by vintage year. Source: ABSNet loan data. We assign a documentation score to each loan (no documentation=0; partial=0.1; alternative=0.3; full=1). Then for a given deal we compute the average documentation score and present the averages by asset type and vintage year.

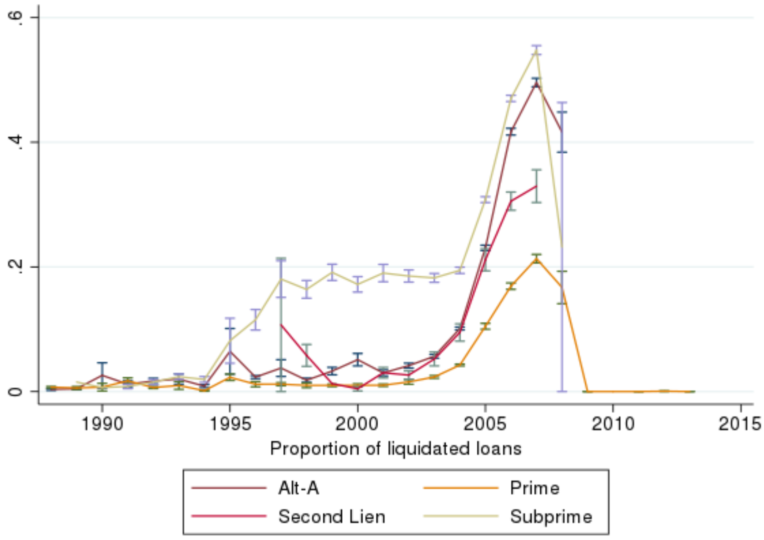


Figure A.7: Probability of default by vintage year. We compute the default rate for each of the deals that compose our population, and then average by vintage year and asset type. The results are presented here along with standard error bands around the average.

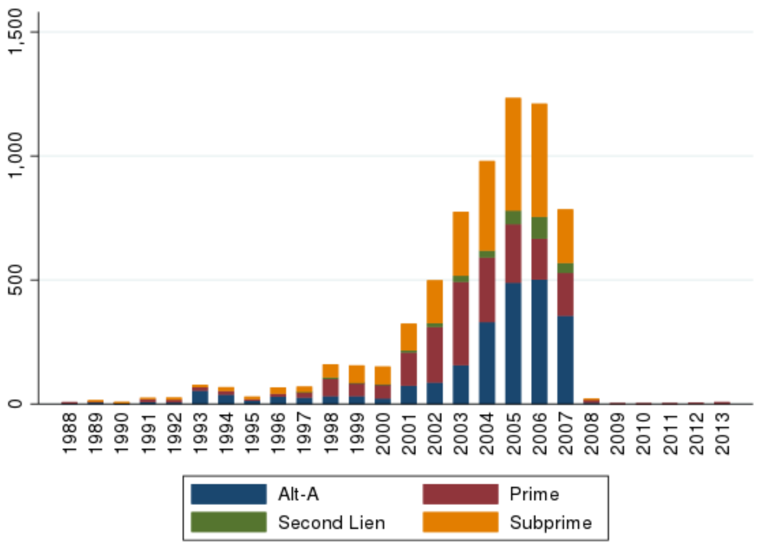


Figure A.8: Number of deals originated by asset type and vintage year.

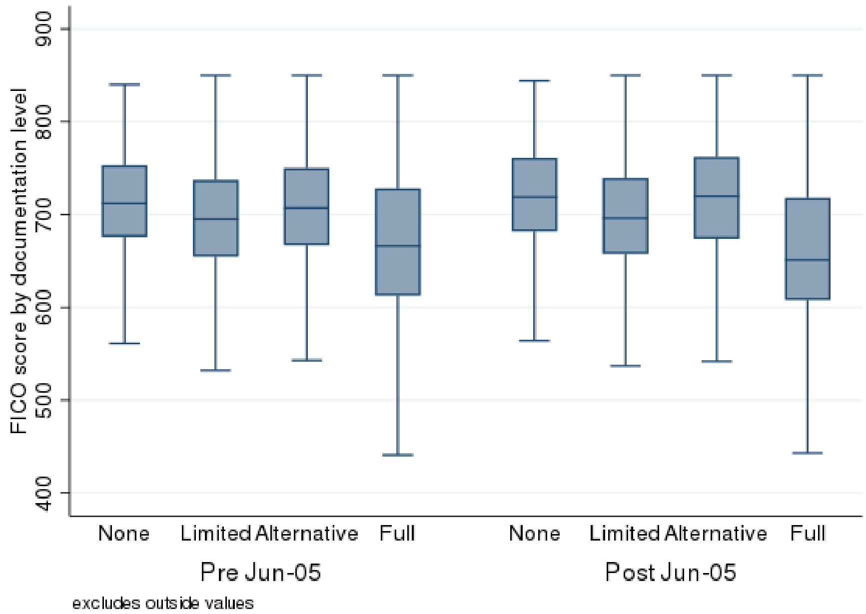


Figure A.9: Box plot exhibiting the median, interquartile range, lower and higher adjacent values of FICO scores at origination over categories of documentation completeness. The plot excludes loans originated after June 2005.

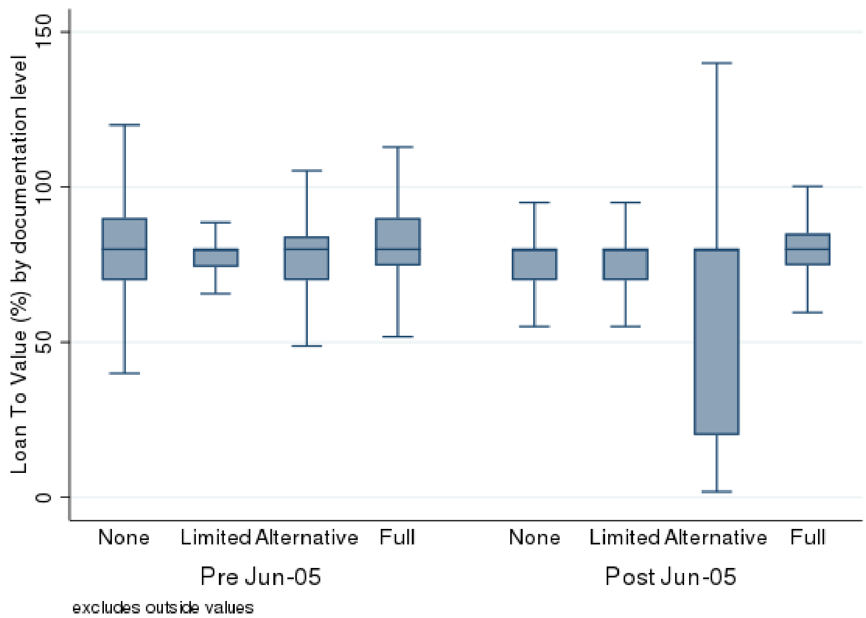


Figure A.10: Box plot exhibiting the median, interquartile range, lower and higher adjacent values of LTV at origination for each category of documentation completeness. The plot excludes loans originated after June 2005.

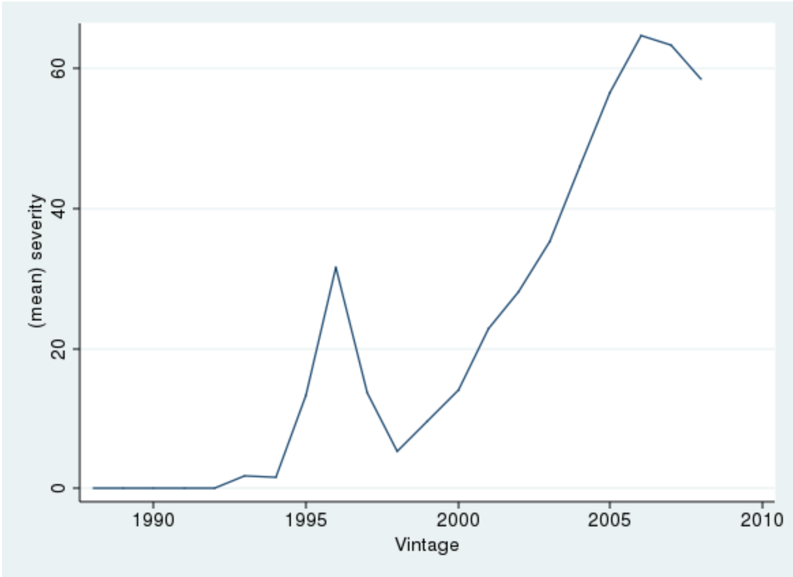


Figure A.11: Percentage loss given default by vintage year. The aggregate loss given default is computed from the sample of loans associated to the deals that compose our population of CMOs.

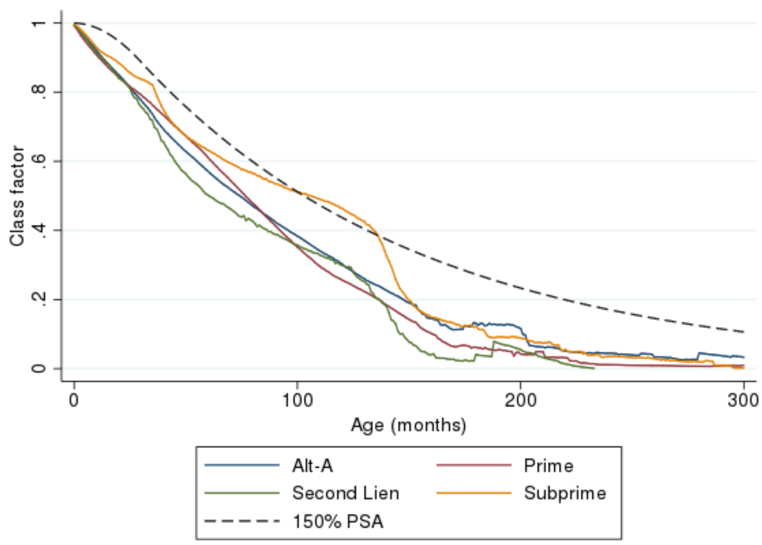
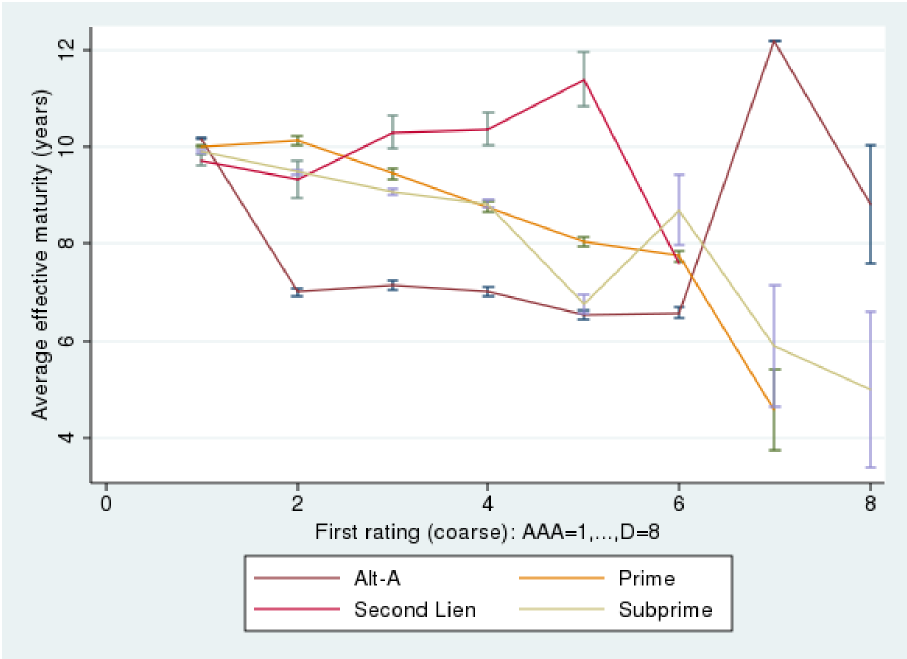
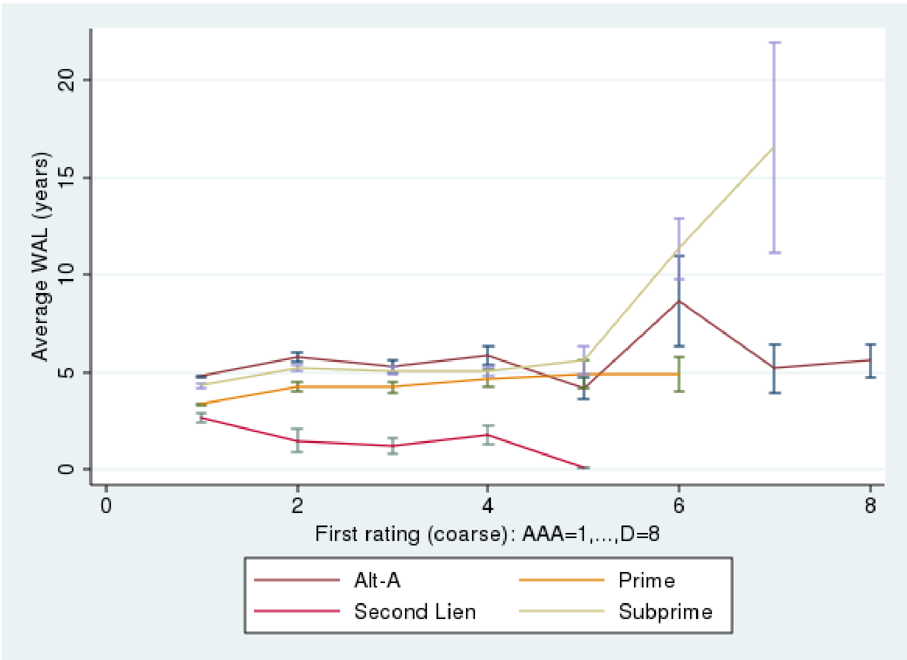


Figure A.12: Average class balance factor by asset class over tranche age. Alongside the averages, we compute the balance factor that results from a 150% payment schedule alone (excluding planned amortization).



(a) Average realized



(b) WAL

Figure A.13: Average realized and weighted average life by coarse rating and asset type. The second panel includes observations where we found a matching WAL in Bloomberg.

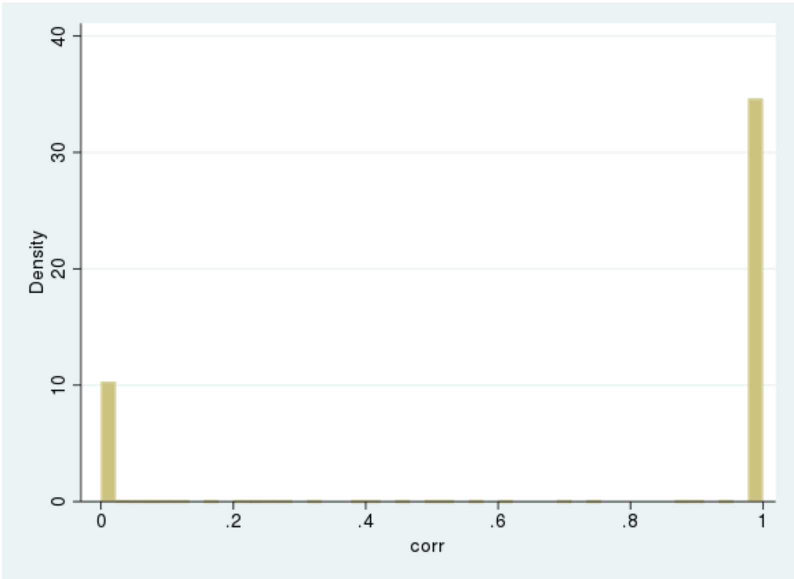


Figure A.14: Histogram plotting all outcomes from the pricing model.

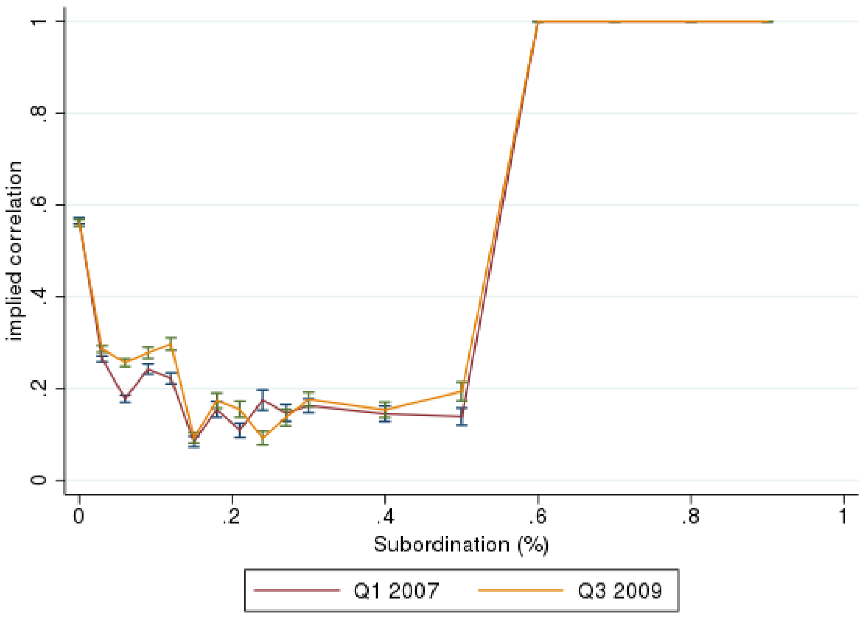


Figure A.15: Average correlation plotted against tranche subordination percentage, on two given dates. We use the sample of early vintage bonds (originated prior to June 2005). Subordinations are assigned to 10 equally spaced bins. Within each subordination bin we plot the average correlation, along with vertical whiskers representing the standard error of the average.

	with data up to 2004		with data up to 2007	
	(1)	(2)	(3)	(4)
	Default	Prepayment	Default	Prepayment
log(FICO)	-1.468***	1.408***	-2.076***	0.305**
	-0.157	-0.155	-0.199	-0.12
owner occupied	0.039	-0.024	-0.098*	0.024
	-0.05	-0.02	-0.054	-0.02
original r - original 10 year rate	0.475***	0.249***	0.252***	0.066***
	-0.01	-0.017	-0.011	-0.006
log(original amount)	0.421***	0.257***	0.143***	0.02
	-0.043	-0.031	-0.041	-0.026
log(original LTV)	0.439***	-0.007	0.183***	0.069***
	-0.043	-0.036	-0.033	-0.02
prepayment penalty	-1.866***	-1.034***	-0.914***	-0.950***
	-0.08	-0.073	-0.031	-0.025
adjustable rate mortgage	0.655***	0.493***	0.367***	0.467***
	-0.062	-0.047	-0.038	-0.015
log(Cumulative HPA)	-8.398***	-7.780***	-6.482***	-2.474***
	-1.041	-0.963	-0.652	-0.41
coupon gap	0.400***	0.120*	-0.255***	-0.144**
	-0.05	-0.062	-0.04	-0.06
unemployment	0.330***	0.320***	0.201***	0.319***
	-0.072	-0.075	-0.068	-0.075
Asset type: Prime	-1.008***	-0.147***	-1.130***	-0.603***
	-0.078	-0.027	-0.078	-0.033
Asset type: Second Lien	-0.580***	0.124	0.843***	0.385***
	-0.142	-0.079	-0.064	-0.028
Asset type: Subprime	0.504***	-0.021	1.113***	0.201***
	-0.053	-0.05	-0.037	-0.02
CBSA FE	Y	Y	Y	Y
Month since origination FE	Y	Y	Y	Y
Observations	68,634,789	76,206,672	121,236,208	126,625,633

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: This table shows estimates using the maximum likelihood estimation of the complementary log-log specification in (11), using a nonparametric baseline hazard, on the loan level data available from ABSNet for private label loans (purchases only). The model treats competing risks independently, indicating 1 for failure and 0 for censoring. Each coefficient is the effect of the corresponding variable on the log hazard rate for either the default or prepayment of a mortgage. The conditional hazard is captured by performance month dummies, where performance is tracked over the first 60 months of the sample. The sample is truncated at December 2004 for columns (1) and (2), and at June 2007 for columns (3) and (4). Errors are clustered at CBSA level.

	(1) default	(2) prepayment
log(FICO)	-2.481***	0.448***
	-0.064	-0.018
owner occupied	0.025*	0.372***
	-0.014	-0.005
original r - original 10 year rate	0.429***	-0.011***
	-0.004	-0.001
log(original amount)	0.137***	0.324***
	-0.01	-0.003
log(original LTV)	0.572***	0.183***
	-0.012	-0.005
adjustable rate mortgage	0.487***	0.579***
	-0.016	-0.004
log(Cumulative HPA)	-1.826***	-1.581***
	-0.051	-0.011
coupon gap	0.848***	-0.261***
	-0.007	-0.002
unemployment	0.080***	0.001
	-0.004	-0.001
Asset type: Prime	-0.808***	-2.719***
	-0.044	-0.014
Asset type: Second Lien	-0.794***	0.298***
	-0.038	-0.011
Asset type: Subprime	0.402***	1.079***
	-0.025	-0.005
CBSA FE	N	N
Month since origination FE	N	Y
Observations	2,630,290	76,374,400

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

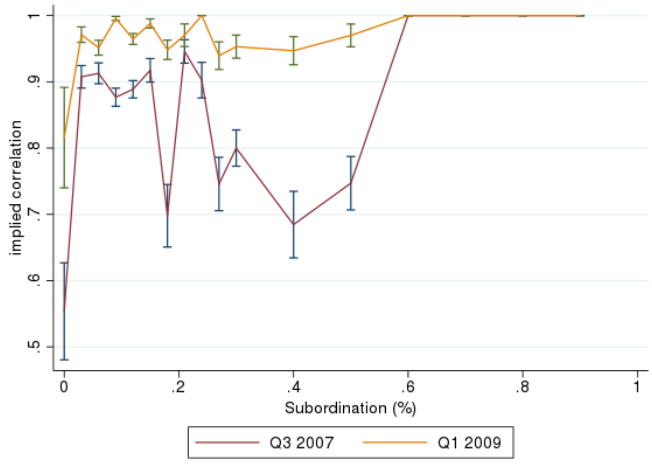
Table 10: This table shows estimates using the maximum likelihood estimation of a complementary log-log specification, using a hazard specification for prepayments and a dummy indicator for default, on the loan level data available from ABSNet for private label loans (purchases only). The hazard model treats default risk as censored. Each coefficient is the effect of the corresponding variable on the log hazard rate for prepayment or the log probability of default of a mortgage. The conditional hazard is captured by performance month dummies, where performance is tracked over the first 60 months of the sample. The sample is truncated at December 2004.

	default indicator (by the end of the given year)									
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
log(FICO)	-2.485***	-3.599***	-4.816***	-2.583***	-3.141***	-3.682***	-4.561***	-4.160***	-3.029***	-2.059***
	(0.494)	(0.257)	(0.151)	(0.101)	(0.071)	(0.052)	(0.038)	(0.027)	(0.018)	(0.014)
owner occupied	-0.318**	0.037	0.263***	0.133***	-0.214***	-0.329***	-0.333***	-0.247***	-0.097***	-0.148***
	(0.139)	(0.069)	(0.041)	(0.024)	(0.016)	(0.011)	(0.008)	(0.006)	(0.004)	(0.003)
original r - original	-0.052	0.277***	0.199***	0.431***	0.459***	0.343***	0.164***	0.178***	0.158***	0.102***
10 year rate	(0.038)	(0.018)	(0.011)	(0.006)	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
log(original amount)	-0.053	-0.038	-0.235***	0.125***	0.150***	-0.026***	-0.297***	-0.126***	-0.029***	-0.013***
	(0.084)	(0.043)	(0.025)	(0.016)	(0.011)	(0.007)	(0.005)	(0.003)	(0.002)	(0.002)
log(original LTV)	0.828***	0.698***	0.585***	0.772***	0.682***	0.548***	0.445***	0.178***	0.124***	0.078***
	(0.266)	(0.099)	(0.030)	(0.019)	(0.014)	(0.010)	(0.007)	(0.004)	(0.003)	(0.002)
adjustable rate	-0.707***	0.145**	0.305***	0.335***	0.261***	0.269***	0.291***	-0.045***	-0.130***	0.001
mortgage	(0.104)	(0.058)	(0.035)	(0.023)	(0.016)	(0.011)	(0.008)	(0.006)	(0.004)	(0.003)
log(cumulative HPA)	1.921***	2.981***	4.548***	-3.303***	-1.878***	-0.877***	0.412***	-1.998***	-5.796***	-4.319***
	(0.676)	(0.248)	(0.122)	(0.103)	(0.054)	(0.030)	(0.018)	(0.017)	(0.011)	(0.007)
coupon gap	-1.930***	0.216***	-0.591***	1.234***	0.998***	0.832***	0.170***	-1.057***	-0.810***	0.889***
	(0.062)	(0.037)	(0.019)	(0.013)	(0.009)	(0.006)	(0.005)	(0.004)	(0.002)	(0.002)
unemployment	0.137***	-1.052***	-0.342***	-0.080***	0.011**	0.126***	0.176***	0.004***	-0.183***	-0.309***
	(0.037)	(0.026)	(0.014)	(0.009)	(0.006)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
Asset type: Prime	0.000	-1.048***	-0.805***	-0.748***	-0.621***	-0.299***	-0.248***	-0.669***	-1.456***	-1.640***
	(.)	(0.245)	(0.100)	(0.068)	(0.049)	(0.035)	(0.028)	(0.024)	(0.018)	(0.011)
Asset type: Second Lien	0.000	-3.509***	-5.936***	-4.410***	-1.984***	-0.213***	0.471***	0.877***	0.639***	0.616***
	(.)	(1.012)	(1.002)	(0.271)	(0.062)	(0.027)	(0.018)	(0.012)	(0.007)	(0.005)
Asset type: Subprime	2.872***	0.939***	0.213***	0.147***	0.274***	0.438***	0.742***	1.027***	0.777***	0.710***
	(0.311)	(0.128)	(0.062)	(0.040)	(0.028)	(0.019)	(0.014)	(0.010)	(0.005)	(0.004)
Obs	230,631	516,866	865,545	1,435,035	2,630,290	4,307,739	5,766,680	6,014,866	6,014,866	6,014,866

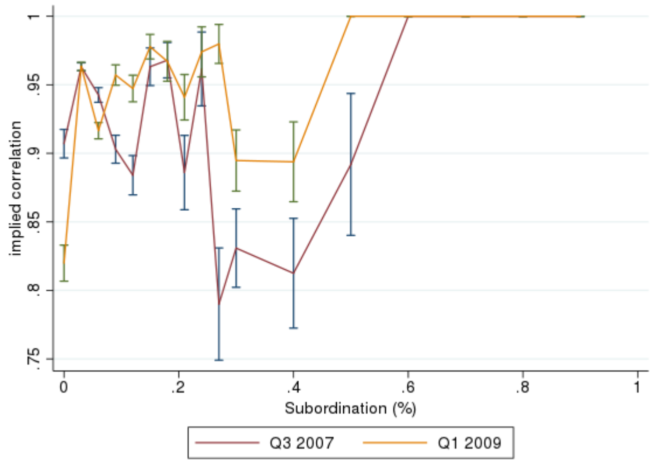
Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

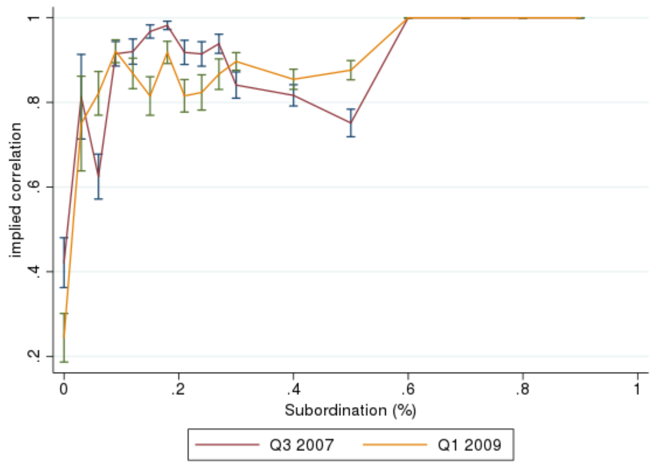
Table 11: This table shows estimates using the maximum likelihood estimation of a complementary log-log specification, using a dummy indicator for default, on the loan level data available from ABSNet for private label loans (purchases only). For each year, variables are taken at the measurement point, either default time, if defaulted, or observation time, which is the end of the given year.



(a) Alt-A

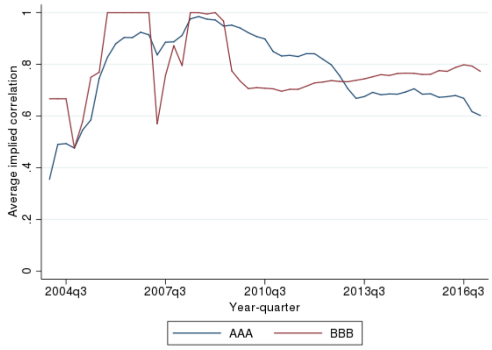


(b) Prime

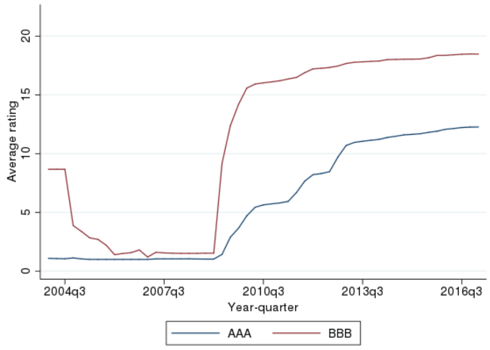


(c) Subprime

Figure A.16: Average correlation plotted against tranche subordination percentage, on two given dates. Subordination values are assigned to 10 equally spaced bins. Within each subordination bin we plot the average correlation, along with vertical whiskers representing the standard error of the average.

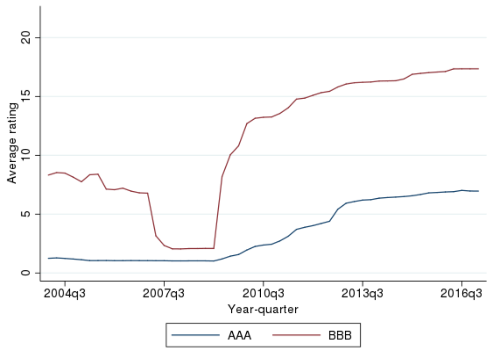


(a) Implied correlation - Alt-A

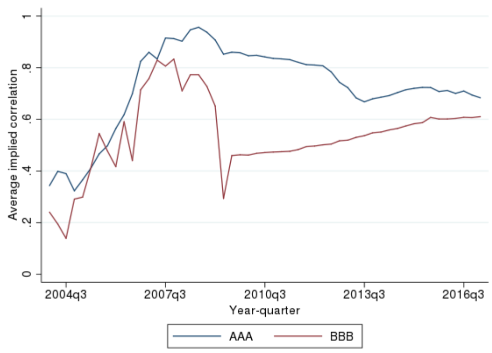


(b) Rating - Alt-A

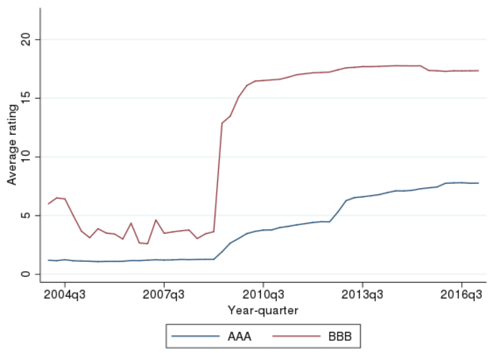
(c) Implied correlation - prime



(d) Rating - prime



(e) Implied correlation - subprime



(f) Rating - subprime

Figure A.17: Performance of early vintage tranches: average implied correlation and average rating for bonds originated before June 2005. For a given tranche we compute the implied correlation at each point in time. The average is taken by transaction period, by coarse rating at origination (AAA=1,..., BBB=4,..., D=8).

	(1)	(2)	(3)	(4)	(5)
	All	[0, 0.25)	[0.25, 0.5)	[0.5, 0.75)	[0.75, 1]
	Downgrade indicator				
Deal average correlation	0.211	-0.581	-0.155	-0.0162	0.901**
	(0.189)	(0.750)	(0.522)	(0.367)	(0.395)
Price	-0.0185***	-0.0165***	-0.0202***	-0.0110***	-0.0167***
	(0.00152)	(0.00631)	(0.00343)	(0.00269)	(0.00356)
Coupon	-0.123***	-0.134**	-0.0369	-0.117***	-0.0749
	(0.0178)	(0.0649)	(0.0310)	(0.0442)	(0.0465)
Subordination	-3.168***	-0.0512	-1.869***	-4.013***	-5.858***
	(0.271)	(0.852)	(0.653)	(0.489)	(0.959)
Observations	26,242	2,489	5,513	7,073	5,049
Rating at first transaction	Y	Y	Y	Y	Y
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y

Standard errors in parentheses

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

Table 12: Regression results from running logit specification 12 by maximum likelihood, controlling for vintage year (vintages up to June 2005) and model-implied probability of default as estimated in section 3.1. The dependent variable is a dummy indicator for whether there was a downgrade by December 2009. Independent variables include deal level average implied correlation and coarse rating dummy indicator at the time of the first transaction. Each column presents the results on a subset of the data corresponding to the value of the documentation index corresponding to the given deal. Errors are clustered at deal level.

B Additional causes of default clustering: frailty and contagion

Following Azizpour et al. (2016), defaults are driven by three factors: systemic risk[33] as captured by macroeconomic variables (Bullard et al., 2009; Khandani et al., 2013)[34], an unobserved frailty factor (Duffie et al., 2009; Kau et al., 2011) and a contagion factor, which captures the extent to which more defaults increase the conditional intensity of default arrival.

A given loan n has a default time T_n. Defaults have a conditional mean of arrival λ given by

$$\lambda_t = \exp\left(a_0 + \sum_{i=1}^{d} a_i X_{i,t}\right) + Y_t + Z_t$$

[33]Bisias et al. (2012) provides a survey of systemic risk measures. See also Chan-Lau et al. (2009). Other approaches include macro measures (costly asset-price boom/bust cycles, property-price, equity-price, credit-gap indicators), forward-looking measures (e.g. absorption rate as in Kritzman, Li, Page, and Rigobon (2010)), cross-sectional measures (CoVaR, Co-Risk, marginal and systemic expected shortfall, see Acharya, Pedersen, Philippon, and Richardson (2012)), stress tests (e.g. Duffie (2011)), illiquidity and insolvency (e.g. Brunnermeier, Gorton, and Krishnamurthy (2011)). Giglio, Kelly, Pruitt, and Qiao (2013) use predictive quantile regression to provide an empirical assessment of 17 of them. Their main finding is that, overall, the compendium of systemic risk measures contains useful predictive information. Instead individual measures tend to fail in capturing systematic risk.

[34]The characterization of systemic risk as deterioration of macroeconomic indicators leaves aside the widely discussed view that the pre-crisis mortgage system was systemically vulnerable (Hellwig, 2009; Poitras and Zanotti, 2016).

	(1)	(2)	(3)	(4)	(5)
	All	[0, 0.25)	[0.25, 0.5)	[0.5, 0.75)	[0.75, 1]
	Downgrade indicator - AAA only				
AAA average correlation	0.721*	5.310	0.701	1.074	3.019***
	(0.427)	(3.508)	(1.594)	(0.962)	(0.857)
Price	-0.0446***	-0.0309***	-0.0351***	-0.0328***	-0.0535***
	(0.00305)	(0.00991)	(0.00542)	(0.00655)	(0.0118)
Coupon	-0.0402	0.0452***	0.0544	0.0898	0.149**
	(0.0245)	(0.0173)	(0.0453)	(0.0580)	(0.0726)
Subordination	-4.144***	0.610	-2.791**	-2.283	-11.79***
	(0.601)	(1.530)	(1.260)	(1.851)	(4.377)
Observations	14,034	1,325	3,073	3,272	2,926
Rating at first transaction	N/A	N/A	N/A	N/A	N/A
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y
	Downgrade indicator - not AAA				
Below-AAA average correlation	0.369*	-1.266	0.527	-0.0727	0.195
	(0.217)	(0.886)	(0.548)	(0.431)	(0.574)
Price	-0.00937***	-0.0154**	-0.0127***	-0.00783***	-0.0115***
	(0.00150)	(0.00704)	(0.00370)	(0.00249)	(0.00362)
Coupon	-0.183***	-0.357***	-0.174***	-0.201***	-0.155***
	(0.0240)	(0.103)	(0.0496)	(0.0529)	(0.0599)
Subordination	-4.113***	0.133	-2.715***	-4.482***	-4.278***
	(0.316)	(1.905)	(0.878)	(0.541)	(0.852)
Observations	12,206	1,038	2,248	3,757	2,111
Rating at first transaction	Y	Y	Y	Y	Y
Vintage year	Y	Y	Y	Y	Y
Asset type	Y	Y	Y	Y	Y

Standard errors in parentheses

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

Table 13: Regression results from running logit specification 12 by maximum likelihood, controlling for vintage year (vintages up to June 2005) and model-implied probability of default as estimated in section 3.1. The dependent variable is a dummy indicator for whether there was a downgrade by December 2009. Independent variables include AAA average correlation (upper panel), sub-AAA average correlation (lower panel) and coarse rating dummy indicator at the time of the first transaction. Each column presents the results on a subset of the data corresponding to the value of the documentation index corresponding to the given deal. Errors are clustered at deal level.

	(1)	(2)	(3)	(4)
	AAA balance at origination as share of deal issuance			
Opacity index	-0.104***	-0.0835***	-0.101***	-0.0259*
	(0.0154)	(0.0153)	(0.0151)	(0.0151)
Observations	1,902	1,902	1,902	1,902
Model-implied PD	N	Y	Y	Y
Vintage year	N	N	Y	Y
Asset type	N	N	N	Y

Standard errors in parentheses

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

Table 14: Regression results from running a linear regression at deal level of AAA origination (as share of total) on the deal opacity index. Controls include model-implied PD, vintage year (we include vintages up to June 2005) and asset type.

S&P rating	Code	Coarse rating	Code
AAA	1	AAA	1
AA+	2	AA	2
AA	3	AA	2
AA-	4	AA	2
A+	5	A	3
A	6	A	3
A-	7	A	3
BBB+	8	BBB	4
BBB	9	BBB	4
BBB-	10	BBB	4
BB+	11	BB	5
BB	12	BB	5
BB-	13	BB	5
B+	14	B	6
B	15	B	6
B-	16	B	6
CCC	17	C	7
CCC-	18	C	7
CC	19	C	7
C	20	C	7
D	21	D	8
NR	-	NR	-

Table 15: Mapping of ratings - fine and coarse level (with numbering code)

where X represents a vector of macroeconomic variables. Unobservable frailty Z_t follows the CIR process

$$dZ_t = k(z - Z_t)dt + \sigma\sqrt{Z_t}dW_t$$
$$Z_0 \sim \Gamma\left(\frac{2kz}{\sigma^2}, \frac{\sigma^2}{2k}\right)$$

Defaults are self-exciting, in the sense that the mass of defaults at a given time increases the rate of arrival. This is captured by means of a contagion factor Y such that

$$Y_t = b \sum_{n:T_n \leq t} e^{-\kappa(t-T_n)} U_n$$
$$U_n = \max(0, \log u_n)$$

where u_n is the sum of defaulted debt at time T_n. This implies that larger defaults are followed by more defaults.

The estimation of $\theta = (a, k, z, \sigma, b, \kappa)$ is a filtered likelihood problem (the likelihood is a posterior mean of the complete-data likelihood), and can be solved following Giesecke and Schwenkler (2016). The likelihood is written as a product of two terms, one that depends on event data (defaults) and one that depends on factor data. The decomposition is based on a change of measure, which resolves the interaction between the point process and the factors of λ.[35] One of the terms is a point process filter, which makes the computation difficult. Giesecke and Schwenkler (2016)

[35] An alternative is to apply the expectation maximization (EM) algorithm. Giesecke and Schwenkler (2016) compare the two approaches.

propose an approximation based on a quadrature method, from which the posterior mean can be computed. They write an algorithm and derive conditions for convergence.

References

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. P. Richardson (2012, Feb). Measuring systemic risk.

Adelino, M. (2009). Do investors rely only on ratings? the case of mortgage-backed securities.

Adelino, M., K. Gerardi, and B. Hartman-Glaser (2016). Are lemons sold first? dynamic signaling in the mortgage market.

Altman, E. I. (2006). Default recovery rates and lgd in credit risk modeling and practice: an updated review of the literature and empirical evidence. *New York University, Stern School of Business*.

AMBROSE, B. W., J. CONKLIN, and J. YOSHIDA (2016). Credit rationing, income exaggeration, and adverse selection in the mortgage market. *The Journal of Finance 71*(6), 2637–2686.

An, X., Y. Deng, J. B. Nichols, and A. B. Sanders (2015). What is subordination about? credit risk and subordination levels in commercial mortgage-backed securities (CMBS). *The Journal of Real Estate Finance and Economics 51*(2), 231–253.

Andersen, L. and J. Sidenius (2004). Extensions to the gaussian copula: Random recovery and random factor loadings. *Journal of Credit Risk 1*(1), 05.

Andreoli, A., L. V. Ballestra, and G. Pacelli (2016). From insurance risk to credit portfolio management: a new approach to pricing CDOs. *Quantitative Finance 16*(10), 1495–1510.

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

Ashcraft, A., P. Goldsmith-Pinkham, and J. Vickery (2010). Mbs ratings and the mortgage credit boom.

Ashcraft, A. B. and T. Schuermann (2008). Understanding the securitization of subprime mortgage credit. *Foundations and Trends® in Finance 2*(3), 191–309.

Azizpour, S., K. Giesecke, and G. Schwenkler (2016). Exploring the sources of default clustering.

Bai, J., P. Collin-Dufresne, R. S. Goldstein, and J. Helwege (2015). On bounding credit-event risk premia. *The Review of Financial Studies 28*(9), 2608.

Begley, T. and A. Purnanandam (2017, 01). Design of financial securities: Empirical evidence from private-label rmbs deals. *30*, 120–161.

Beltran, D. O., L. Cordell, and C. P. Thomas (2017). Asymmetric information and the death of ABS CDOs. *Journal of Banking & Finance 76*, 1 – 14.

Benešová, P. and P. Teply (2010, January). Main Flaws of The Collateralized Debt Obligations: Valuation Before And During The 2008/2009 Global Turmoil. Working Papers IES 2010/01, Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

Benmelech, E. and J. Dlugosz (2010). The credit rating crisis. *NBER Macroeconomics Annual 24*(1), 161–208.

Berndt, A., R. Douglas, D. Duffie, M. Ferguson, and D. Schranz (2005). Measuring default risk premia from default swap rates and edfs.

Bisias, D., M. Flood, A. W. Lo, and S. Valavanis (2012). A survey of systemic risk analytics. *Annual Review of Financial Economics 4*(1), 255–296.

Boot, A. W. and A. V. Thakor (1993). Security design. *The Journal of Finance 48*(4), 1349–1378.

Brunne, T. (2006). Implied correlation of synthetic CDOs with liquid markets.

Brunnermeier, M. K., G. Gorton, and A. Krishnamurthy (2011). *Risk Topography*, pp. 149–176. National Bureau of Economic Research, Inc.

Bullard, J., C. J. Neely, D. C. Wheelock, et al. (2009). Systemic risk and the financial crisis: a primer. *Federal Reserve Bank of St. Louis Review 91*(5 Part 1), 403–18.

Buzková, P. and P. Teplỳ (2012). Collateralized debt obligations valuation using the one factor gaussian copula model. *Prague Economic Papers 21*(1), 30–49.

Carlin, B. I., F. A. Longstaff, and K. Matoba (2014). Disagreement and asset prices. *Journal of Financial Economics 114*(2), 226 – 238.

Chan-Lau, J. A., M. Espinosa, K. Giesecke, and J. A. Sole (2009). Assessing the systemic implications of financial linkages. *IMF Global Financial Stability Report 2*.

Cordell, L., Y. Huang, and M. Williams (2012). Collateral damage: Sizing and assessing the subprime CDO crisis.

Coval, J., J. Jurek, and E. Stafford (2009a). The economics of structured finance. *The Journal of Economic Perspectives 23*(1), 3–25.

Coval, J. D., J. W. Jurek, and E. Stafford (2009b). Economic catastrophe bonds. *The American Economic Review 99*(3), 628–666.

Cowan, A. M. and C. D. Cowan (2004). Default correlation: An empirical investigation of a subprime lender. *Journal of Banking & Finance 28*(4), 753 – 771. Retail Credit Risk Management and Measurement.

Crouhy, M., D. Galai, and R. Mark (2000). A comparative analysis of current credit risk models. *Journal of Banking & Finance 24*(1), 59–117.

Daley, B., B. S. Green, and V. Vanasco (2018). Security design with ratings.

D'Amato, J. and J. Gyntelberg (2005). CDS index tranches and the pricing of credit risk correlations. *BIS Quarterly Review*.

Dang, T. V., G. Gorton, and B. Holmström (2013). The information sensitivity of a security. *Unpublished working paper, Yale University*.

de Servigny, A. and O. Renault (2002). Default correlation: empirical evidence. *Standard and Poor's*.

Dell'Ariccia, G., D. Igan, and L. Laeven (2012, 03). Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market. *Journal of Money, Credit and Banking 44*, 367–384.

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

Denzler, S. M., M. M. Dacorogna, U. A. Müller, and A. J. McNeil (2006). From default probabilities to credit spreads: Credit risk models do explain market prices. *Finance Research Letters 3*(2), 79–95.

Diep, P., A. L. Eisfeldt, and S. A. Richardson (2016). Prepayment risk and expected mbs returns.

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

Downing, C., R. Stanton, and N. Wallace (2005). An empirical test of a two-factor mortgage valuation model: How much do house prices matter? *Real Estate Economics 33*(4), 681–710.

Duffie, D. (2008, July). Innovations in credit risk transfer: Implications for financial stability. *BIS Working Paper* (255).

Duffie, D. (2011, August 2011). Systemic risk exposures: A 10-by-10-by-10 approach. *National Bureau of Economic Research Working Paper Series No. 17281.*

Duffie, D., A. Eckner, G. Horel, and L. Saita (2009). Frailty correlated default. *The Journal of Finance 64*(5), 2089–2123.

Duffie, D. and N. Gârleanu (2001). Risk and valuation of collateralized debt obligations. *Financial Analysts Journal 57*(1), 41–59.

Duffie, D. and K. J. Singleton (2012). *Credit risk: pricing, measurement, and management.* Princeton University Press.

Echeverry, D., R. Stanton, and N. Wallace (2016). Funding fragility in the residential-mortgage market.

Elizalde, A. (2005). Credit risk models iv: Understanding and pricing cdos. *CEMFI and Universidad Publica de Navarro, download: www. cemfi. es/ elizalde.*

Finkelstein, D., A. Strzodka, and J. Vickery (2018). Credit risk transfer and de facto gse reform. Staff Reports 838, Federal Reserve Bank of New York.

Frye, J. (2008). Correlation and asset correlation in the structural portfolio model. *The Journal of Credit Risk 4*(2), 75–96.

Geidosh, M. (2014, 06). Asset correlation in residential mortgage-backed security reference portfolios. *Journal of Credit Risk 10*(2), 71–95.

Ghent, A. C., W. N. Torous, and R. Valkanov (2016). Complexity in structured finance. *The Review of Economic Studies*.

Giesecke, K. and G. Schwenkler (2016). Filtered likelihood for point processes.

Giglio, S., B. Kelly, S. Pruitt, and X. Qiao (2013). Systemic risk and the macroeconomy: An empirical evaluation. *Chicago Booth Research Paper* (12-49).

Gordy, M. B. (2000). A comparative anatomy of credit risk models. *Journal of Banking & Finance 24*(1), 119–149.

Gorton, G. (2009). The subprime panic*. *European Financial Management 15*(1), 10–46.

Gorton, G. and G. Pennacchi (1990). Financial intermediaries and liquidity creation. *The Journal of Finance 45*(1), 49–71.

Gorton, G. B. and G. Ordonez (2013). The supply and demand for safe assets. Technical report, National Bureau of Economic Research.

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

Griffin, J. M. and J. Nickerson (2016). Debt correlations in the wake of the financial crisis: What are appropriate default correlations for structured products?

Griffin, J. M. and D. Y. Tang (2012). Did subjectivity play a role in cdo credit ratings? *The Journal of Finance 67*(4), 1293–1328.

Gupta, A. (2016). Foreclosure contagion and the neighborhood spillover effects of mortgage defaults. Technical report, Working Paper.

Hellwig, M. F. (2009). Systemic risk in the financial sector: An analysis of the subprime-mortgage financial crisis. *De Economist 157*(2), 129–207.

Heynderickx, W., J. Cariboni, W. Schoutens, and B. Smits (2016). The relationship between risk-neutral and actual default probabilities: the credit risk premium. *Applied Economics 48*(42), 4066–4081.

Hull, J. and A. White (2008). Dynamic models of portfolio credit risk: A simplified approach. *Journal of Derivatives 15*(4), 9–28.

Hull, J. C. and A. D. White (2004). Valuation of a CDO and an n-th to default CDS without monte carlo simulation. *The Journal of Derivatives 12*(2), 8–23.

Hull, J. C. and A. D. White (2006). Valuing credit derivatives using an implied copula approach. *The Journal of Derivatives 14*(2), 8–28.

IOSCO (2008, March). The role of credit rating agencies in structured finance markets. Technical report, International Organization of Securities Commissions.

Jarrow, R. A. (2011). Risk management models: construction, testing, usage. *The Journal of Derivatives 18*(4), 89–98.

JEC (2007). The subprime lending crisis: The economic impact on wealth, property values and tax revenues, and how we got here. *World Wide Web page¡ http://www. jec. senate. gov/Documents/Reports/10.25. 07OctoberSubprimeReport. pdf¿(accessed January 15, 2008)*.

Kau, J. B., D. C. Keenan, and X. Li (2011). An analysis of mortgage termination risks: a shared frailty approach with msa-level random effects. *The Journal of Real Estate Finance and Economics 42*(1), 51–67.

Kau, J. B., D. C. Keenan, C. Lyubimov, and V. C. Slawson (2011). Subprime mortgage default. *Journal of Urban Economics 70*(2), 75–87.

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig (2010). Did securitization lead to lax screening? evidence from subprime loans. *The Quarterly journal of economics 125*(1), 307–362.

Khandani, A. E., A. W. Lo, and R. C. Merton (2013). Systemic risk and the refinancing ratchet effect. *Journal of Financial Economics 108*(1), 29–45.

Kritzman, M., Y. Li, S. Page, and R. Rigobon (2010). Principal components as a measure of systemic risk.

Li, D. X. (2000). On default correlation: A copula function approach. *The Journal of Fixed Income 9*(4), 43–54.

Liu, H. (2016). Do government guarantees inhibit risk management? evidence from Fannie Mae and Freddie Mac.

Longstaff, F. A. and A. Rajan (2008). An empirical analysis of the pricing of collateralized debt obligations. *The Journal of Finance 63*(2), 529–563.

McGinty, L., E. Beinstein, R. Ahluwalia, and M. Watts (2004). Credit correlation: A guide. Technical report.

Nagpal, K. and R. Bahar (2001). Measuring default correlation. *Risk 14*(3), 129–132.

O'Kane, D. and M. Livesey (2004). Base correlation explained. *Lehman Brothers, Fixed Income Quantitative Credit Research*.

Palmer, C. (2015). Why did so many subprime borrowers default during the crisis: Loose credit or plummeting prices?

Piskorski, T., A. Seru, and J. Witkin (2015, 12). Asset quality misrepresentation by financial intermediaries: Evidence from the rmbs market. *The Journal of Finance 70*(6), 2635–2678.

Poitras, G. and G. Zanotti (2016). Mortgage contract design and systemic risk immunization. *International Review of Financial Analysis 45*, 320–331.

Rapisarda, G. and D. Echeverry (2013). A nonparametric approach to incorporating incomplete workouts into loss given default estimates. *The Journal of Credit Risk 9*(2), 47.

Schlösser, A. (2011). *Pricing and risk management of synthetic CDOs*, Volume 646. Springer Science & Business Media.

Schwartz, E. S. and W. N. Torous (1989). Prepayment and the valuation of mortgage-backed securities. *The Journal of Finance 44*(2), 375–392.

Simsek, A. (2013). Belief disagreements and collateral constraints. *Econometrica 81*(1), 1–53.

Sirignano, J., A. Sadhwani, and K. Giesecke (2016). Deep learning for mortgage risk.

Skreta, V. and L. Veldkamp (2009). Ratings shopping and asset complexity: A theory of ratings inflation. *Journal of Monetary Economics 56*(5), 678–695.

Stanton, R. (1995). Rational prepayment and the valuation of mortgage-backed securities. *The Review of Financial Studies 8*(3), 677.

Stanton, R. and N. Wallace (2011). The bear's lair: Index credit default swaps and the subprime mortgage crisis. *Review of Financial Studies 24*(10), 3250–3280.

Stanton, R. and N. Wallace (2018, 1). Cmbs subordination, ratings inflation, and regulatory capital arbitrage. *Financial Management 47*(1), 175–201.

Sugimura, T. (2004). Valuation of residential mortgage-backed securities with default risk using an intensity-based approach. *Asia-Pacific Financial Markets 11*(2), 185–214.

Tzani, R. and A. Polychronakos (2008, 12). Correlation breakdown, copula credit default models and arbitrage. *Global Association of Risk Professionals Magazine*.