Adverse or maybe not-so-adverse selection in the Commercial Mortgage-Backed Security Market*

Chenghuan Sean Chu

Federal Reserve Board of Governors

June 2011

Abstract

The commercial mortgage-backed security (CMBS) market is, in theory, subject to adverse selection at the margin between loans that are securitized in-house by the originator and loans sold to competing CMBS underwriters. However, heterogeneous monitoring costs and other causal mechanisms affecting ex post loan performance can also explain why mortgages securitized in-house tend to perform better (as measured by default-adjusted returns) after controlling for observable loan characteristics. Moreover, portfolio diversification incentives may actually lead to selection for in-house loans that have worse ex ante characteristics. I quantify the relative importance of selection versus causal effects. I first estimate the marginal and joint distribution of loan returns. I then use these estimates as inputs to a model of CMBS pool formation in which competing firms buy and sell loans from each other to form bundles of loans to be securitized. The estimation exploits moment inequalities implied by optimality conditions, and considers alternative specifications of the information structure of the game. The results indicate that ex post causal mechanisms more than account for the better performance of in-house loans. The causal effect outweighs non-random selection, which actually has the net effect of selecting for in-house loans with worse ex ante characteristics. The importance of ex post mechanisms suggests that regulatory reform for securitization markets should aim to mitigate moral hazard rather than adverse selection.

*Thanks go to Lamont Black, Brian Bucks, Andrew Cohen, and Joe Nichols for helpful conversations; Maxim Massenkoff for excellent assistance; and Realpoint LLC for supplying the data. Views expressed herein are the author's and do not reflect the stance of the Federal Reserve Board of Governors or Staff.

1 Introduction

In the wake of the recent financial crisis, it has widely been recognized that securitization creates a potential for adverse selection through the transfer of ownership from asset originators to other parties. This paper studies the market for commercial mortgage-backed securities (CMBS), with a focus on explaining differences between returns on loans that were securitized "in-house"—that is, bundled into pools backing CMBS deals underwritten by the same firm as the loan originator—and returns on loans backing deals underwritten by a firm other than the originator. In fact, empirically, in-house loans tend to perform better (after controlling for observable loan characteristics), but this tendency does not by itself imply that adverse selection exists.

A first complication is that differential returns may result from either selection effects or causal mechanisms by which firms can affect the ex post performance of an in-house loan. In particular, the underwriter of a CMBS deal may have stronger incentives to monitor a loan that it has securitized in-house, or may have lower costs for doing so. Stronger incentives would arise if the firm is concerned about its reputation not only as an underwriter but also as a loan originator, rather than just the former. If the market for commercial real estate loans is differentiated and originators have specialized knowledge about the particular niches in which they operate, then in-house securitization could also create a cost advantage. Distinguishing between ex ante selection and ex post hidden action is a familiar empirical challenge from other settings with asymmetric information, such as insurance markets (e.g., Finkelstein and Poterba, 2006; Cohen, 2005). The distinction has important implications for policy. For example, Title IX of the Dodd-Frank Wall Street Reform and Consumer Protection Act (2010) requires loan originators selling loans to securitization pools to retain a financial interest in the securitization deal. While a regulatory floor on retained interest may not necessarily mitigate adverse selection, it could conceivably improve average loan performance if loan originators can causally affect the ex post performance of in-house loans.

A second, less well-recognized complication follows from a feature specific to securitization. Namely, the selection of loans is a *portfolio* decision, with firm profits being determined by the joint distribution of asset returns. Even in the absence of asymmetric information or causal mechanisms, the correlation structure of asset returns can lead to nonrandom selection that confounds identification of adverse selection. Consider two firms that both originate loans and underwrite CMBS deals, each specializing in one of two less-than perfectly correlated markets (California and Illinois). Due to the diversification benefits of holding loans from different geographic markets, the firm based in California would have a greater willingness to pay for a New Jersey loan than for a California loan with the same marginal probability of default, implying that California loans held by the California underwriter are likely to perform better individually.

The classic approach to dealing with an endogeneous explanatory variable such as in-house status is to identify an instrument that is correlated with the endogenous variable but does not enter the equation determining the outcome of interest. However, the economics of CMBS makes it hard to identify such an instrument, as any variable shifting the propensity of a loan to be securitized in-house also tends to be directly correlated with the performance of the loan. For example, an underwriter facing tight short-term funding constraints has more incentive to sell loans it has originated to other underwriters, but if the funding problems are due to the firm being under financial distress, the firm may also have an incentive to undercut its lending standards (see Titman and Tsyplakov, 2010). Similarly, loan characteristics affecting their propensity to be securitized in-house (such as the geographic location of the property serving as the collateral) are also likely to be correlated with monitoring costs and thus the performance of the loan.

To deal with the problem of identification, I estimate a structural model that explicitly specifies CMBS underwriters' decisions regarding which loans to include in their pools. Because an underwriter's pool comprises multiple loans, the problem is a many-to-one matching problem. Firms buy and sell loans from each other, taking into account the effect of trades on the distribution of portfolio returns. Identification comes from two sources. First, if a firm has more available potential trading partners for an originated loan (a concept made precise in the paper), that loan is less likely to be securitized in-house. Note that because the decision to include a loan in a pool is not independent across loans, the number of available potential trading partners would not be a valid instrument in the conventional instrumental variables setting.[1] However, it provides a source of exogenous variation in the structural model.

An identification strategy that exploits variation in the availability of competing firms in a many-to-one match is loosely related to work by Sørensen (2006). A novel feature of the securitization setting is that portfolio effects of competing firms provide an additional source of identification. In particular, if a loan's exogenous characteristics are negatively correlated with pool characteristics for CMBS deals underwritten by firms competing with the loan originator, the loan is less likely to be securitized in-house due to the diversification benefit that it provides to the competing firms. In short, both the number of potential trading partners and the characteristics of other firms' pools

[1] Having more available trading partners increases the *potential* for selection (e.g., consider the extreme case in which a firm has no trading partners and thus no opportunities for adverse selection) and thus the strength of correlation between a loan's unobserved quality and the event of the loan being securitized in house. The number of available trading partners is therefore correlated with loan outcomes independently of its effect on the probability of the loan being securitized in-house.

provide exogenous variation in the propensity of a loan to be securitized in-house without directly affecting the performance of the loan itself.

I estimate the model using data on commercial real estate loans securitized between 2000 and 2007, and exploit the moment inequalities methodology developed by Pakes, Porter, Ho, and Ishii (2006) but with a multistage implementation. The data contain a very rich collection of control variables for payoff-relevant loan characteristics. The high dimensionality of the loan characteristics and complexity of the payoff correlation structure across loans implies that structurally estimating all of the parameters would be cumbersome even if theoretically possible. Because most of the loan characteristics can properly be considered to be exogenous, the multistage approach simplifies estimation by obtaining many of the key parameters directly from the data in the initial stages.

Specifically, I first estimate the marginal distribution of loan returns using a semiparametric mixtures estimator based on Bajari, Fox, Kim, and Ryan (2010). In a second stage, I estimate the joint distribution of loan returns using a copula function. Finally, I take the first- and second-stage estimates of parameters that are associated with exogenous variables, and use them as inputs into the payoff structure of the structural model. The structural model identifies the remaining parameters, including the causal impact of in-house status on loan performance. Because the aggregate effect of selection and causal mechanisms on the performance of in-house versus non–in-house loans is identified in the reduced-form first stage, and the causal effect is identified in the structural model, the net effect can be attributed to selection.

A key advantage of the moment inequalities methodology is that I do not have to make strong assumptions about agents' information sets. However, I explore two alternative assumptions about the information structure of the game. The first case assumes that firms have *symmetric* information about each loan, implying that selection is only over observable loan characteristics. The second case allows for *asymmetric* information but, making a tradeoff necessary for identification, imposes that agents have private signals about loan quality that are specific to originators as opposed to being loan-specific.

The estimates resulting from both specifications imply that the causal channel more than accounts for the observed extent to which in-house loans outperform non–in-house loans. After netting out the estimated causal impact under the arguably more plausible assumption of asymmetric information, the remaining effect due to selection implies that in-house loans have a hazard of default that is twice the non–in-house average. This doubling of the hazard is consistent with firms' incentive to securitize certain loans in-house for idiosyncratic diversification reasons even if those loans have worse individual performance. Estimates under the assumption of symmetric information, which rules out adverse selection, imply an even stronger degree of selection toward worse-performing loans being securitized in-house.

This paper is linked to three distinct literatures. First, it is linked to the theoretical literature on financial intermediation with asymmetric information and bundling of assets, of which DeMarzo and Duffie (1999), DeMarzo (2005), and Glaeser and Kallal (1997) serve as three key examples. These papers consider the optimal design of securities sold to uninformed investors by an intermediary with private signals about a set of assets in its possession. DeMarzo and Duffie (1999) and DeMarzo (2005) show that compared with selling unbundled loans, selling loans as a bundled securitization deal has two competing effects on profits. First, there is an "information destruction" effect that reduces profits, which is driven by the fact that pooling reduces the intermediary's ability to exploit private information about specific loans. Second, if the intermediary can issue "tranched" securities, pooling enables the firm to issue an informationally insensitive debt security, which allows the firm to earn a liquidity premium. Glaeser and Kallal (1997) focus on a different tradeoff. If diversification through pooling improves the worst case of the lemons problem (reducing the range of possible asset values), it increases liquidity and thus the intermediary's profit. On the other hand, if pooling results in a mean-preserving contraction in the distribution of asset valuations, it becomes harder for the signaling equilibrium to enforce truth-telling, which tends to reduce profits. My theoretical framework is complementary to these papers: whereas they focus on a single intermediary with an exogenous asset pool and model equilibrium investor demand, I do not explicitly model investor demand but model the endogenous formation of asset pools as a product of interactions among competing intermediaries.

My paper also joins the budding empirical literature on commercial mortgage-backed securities. Existing works tend to focus on the determinants of default for individual loans backing CMBS deals. Early works by Ambrose and Sanders (2001), Archer et al. (2002), and Deng et al. (2004) relate loan performance to observable underwriting characteristics such as the loan-to-value (LTV) ratio and measures of borrower income. Black, Chu, Cohen, and Nichols (2011) find that loan performance varies systematically with the organizational form of the originator. Titman and Tsyplakov (2010) show that companies undergoing financial distress often lower the quality of their underwriting as they push more marginal loans into securitized pools. Similar to all of these papers, I also examine the determinants of individual loan performance, but I then use individual loan performance to help explain endogenous pool formation. My paper also relates to a small number of descriptive papers on CMBS deal structure. Furfine (2010) finds that an increase in the complexity of CMBS deals over the past decade was accompanied by worsening ex post performance for loans bundled into more complex deals. An et al. (2010) explore differences in security pricing between pools containing loans originated by multiple lenders versus pools containing only loans by a single lender, and find that the former enjoyed a price premium from 1994 to 2000.

Third, this paper finds a new application—empirical finance—for methodology developed by the industrial organization literature on partially-identified games and estimation using moment inequalities. A growing number of papers dealing with this topic have emerged in recent years (Chernozhukov, Hong, and Tamer 2003; Andrews, Berry, and Jia 2004; Shaikh, 2005; Pakes, Porter, Ho, and Ishii, 2006). Yet little work has been done applying the techniques to the field of finance or, in particular, toward understanding strategic interactions among intermediaries in markets for complex financial products. At the same time, my model tries to adhere closely to institutional details underlying the objective functions of practitioners of the securitization industry. For example, to better capture practioners' subjective beliefs, the second stage of my model addresses the joint distribution of asset returns using a copula model, similar to standard industry practice for evaluating credit portfolios.

The rest of the paper proceeds as follows. Section 2 describes the CMBS industry along with stylized facts about the performance of CMBS loans. Section 3 presents the model in three stages. The first stage looks at the performance of individual loans; the second stage examines the joint returns of loans; and the third stage models the game that determines how loans are allocated to deals. Section 4 describes the estimation procedure and the alternative information assumptions about the game. Section 5 presents the first- and second-stage results. Section 6 presents the structural estimation results. Section 7 concludes.

2 The CMBS Industry

Between the 1990s and the recent financial crisis, CMBS grew rapidly to become a significant source of debt financing for commercial mortgages. CMBS currently accounts for approximately one quarter of outstanding commercial real estate (CRE) loans[2] and accounted for almost 40 percent of the CRE loans originated in 2007. Loans securitized in CMBS are typically backed by established, income-generating properties, and have longer maturities than CRE loans held on the originator's balance sheet.

A large number of market participants are involved in each CMBS deal, but most relevant for this paper are the loan originators and the CMBS underwriters. The market participants in loan origination include investment banks, commercial banks, investment companies, specialty finance

[2] Federal Reserve Board's Flow of Funds Table L.220.

companies, and "conduit" firms. Most of these firms originate both loans that remain on their own balance sheets as well as loans that are securitized as CMBS. An exception is the conduits, which hold very little capital and try to sell all loan originations to CMBS deals as quickly as possible. Black et al. (2011) explore differences in loan performance according to the type of originator, and relate these differences to various institutional features of the firms.

CMBS deals are issued by one or more underwriters, one of which acts as the *lead underwriter*, also known as the bookrunner. In some cases, two or three underwriters serve as co-leads. During the sample time period, the lead-underwriting market was dominated by 23 major investment and commercial banks—J. P. Morgan Chase, Bank of America, Credit Suisse, and Goldman Sachs—to name a few. Importantly, these 23 firms also originated 60 percent of the securitized loans. The remaining 40 percent were purchased from other firms that were not engaged in the underwriting business, such as the insurance companies and the conduit lenders.

In a typical securitization, the lead underwriter determines the management structure of the deal (a decision that the rating agencies also weigh in on), finds potential investors, and assembles a pool of loans. Up until some cutoff date, the composition of the pool may change. Pools typically (but not always) include a large number of loans that the underwriter has itself originated, which I refer to as "in-house" loans. For the average CMBS deal in the sample, 54.3 percent of all loans in the pool are in-house. The remaining loans come from other originators, including competing CMBS underwriters as well as the conduits, insurance companies, and finance companies.[3]

At some point before the cutoff date, the underwriter structures the future cashflows from principal and interest payments generated by the loan pool into strictly prioritized claims, and sells the resulting "tranches" as securities. During this process, the rating agencies also exert a considerable degree of control over the tranche structure.[4] Investors pay a premium for more senior tranches, whose capital is shielded from losses by the "credit support" provided by more junior tranches. For example, investors in the "investment-grade" tranches receive principal payments before holders of the "B piece," which incurs the first loss when loans begin to default. Similarly, within the investment-grade tranches, the "Super-senior AAA" bond is paid off before the "Junior AAA" bond, and so on.[5] In practice, there are many complications beyond this basic schema. For

[3]If a CMBS deal included a large number of loans that were originated by another underwriter, the latter firm was often brought on as an additional underwriter, though not necessarily as a co-lead.

[4]In principle, the rating agencies' job is to determine the amount of credit support each tranche needs in order to bring the probability of default below some threshold and thus merit a particular rating: the more credit support (i.e., the thicker the tranches junior to the tranche), the safer the tranche. In practice, a variety of agency conflicts have resulted in ratings that tend to understate the true risk of the bonds. See Cohen (2011) for details.

[5]Note that while more senior bonds are exposed to less credit risk, because they are also the first to be paid off

example, certain "interest-only (IO)" tranches contain only slices of interest payments from other supporting bonds. Moreover, the complexity of the deals tended increase throughout the 2000s, as investors demanded ever more specific claims.[6]

In Figure 1, the vertical axis of the scatterplot shows the residual of the proportion of in-house loans in each deal, relative to the expected proportion if loans were randomly allocated to deals with cutoff dates around the time of the loan origination. As seen in the distribution, the proportion of in-house loans tends to be higher than the expected proportion based on random allocation. Moreover, the residual tends to be negatively correlated with the total amount of loan originations by the underwriter around the time of the deal (the horizontal axis). This correlation is to be expected if, all else equal, underwriters have a desire to diversify their portfolios: because the loans originated by an individual underwriter tend to have somewhat similar characteristics, if the underwriter originates a large number of loans, maintaining pool diversification requires selling a greater proportion of the loans.

The underwriter both retains a stake in and has the ability to affect the performance of the pool after finalizing the deal. Often the retained stake is explicit, for underwriters commonly keep the B piece. Even in cases in which all tranches including the B piece are sold to investors, the underwriter retains an implicit stake insofar as the performance of the securities affects the reputation of the firm. Along with having these incentives, the underwriter has the ability to exercise a degree of control over the ex post performance of the pool through its role in monitoring loans over time. Specifically, the underwriter hires a special servicer to track the performance of individual loans, decide when to place troubled loans into "special servicing," and decide how to work out loans once they are in special servicing.[7] The special servicer's incentives are partly aligned with the underwriter's incentives because of the promise of future business from the underwriter. As well, when the underwriter retains the B piece it may, as the principal, fire and replace the special servicer.

Both endogenous portfolio selection and causal effects of in-house status on ex post loan performance lead to theoretical predictions of performance differences depending upon whether loans are

when loans prepay, they may be exposed to higher "duration risk," namely the risk that the principal on the bond is paid off faster or slower than expected. A bond that pays off faster than expected would be undesireable if the coupon on the bond is higher than prevailing market rates.

[6]Furfine (2010) provides some evidence that increasing complexity was a profitable strategy for underwriters.

[7]The B piece is more often than not sold to the special servicers, a situation that alleviates some agency problems and creates yet others, as described in Gan and Mayer (2006). Gan and Mayer document that the practice of selling the B piece to the special servicer became less common over time because the special servicers had limited capital and lower risk appetite than the underwriters.

securitized in-house. The obvious causal mechanisms unambiguously imply better performance for in-house loans. Specifically, the underwriter's private information may make it easier to monitor loans that it also originated (and to observe whether the special servicer is shirking). Moreover, the underwriter may have more high-powered incentives to ensure the performance of loans it has originated, which affect not only the firm's reputation as a CMBS underwriter but also as a loan underwriting shop.

In contrast to the causal mechanisms, the consequences of endogenous portfolio selection are theoretically ambiguous. On the one hand, if underwriters have payoff-relevant private information about loans that they themselves originated, adverse selection would result in in-house loans having better characteristics along these dimensions. On the other hand, even under symmetric information, an underwriter may idiosyncratically value certain loans that are individually worse-performing if the loans offer diversification benefits due to correlation patterns with other assets in the underwriter's portfolio. For example, if real estate values in California and Illinois are negatively correlated and lead underwriters i and i' already have portfolios that are heavy in California and Illinois loans, respectively, then i will have a greater marginal valuation for loans from Illinois. In fact, a firm may be willing to tolerate a loan with a higher-than-average marginal probability of default if that loan's returns are negatively correlated with other loans in the portfolio.

Finally, note that adverse selection is unlikely to be *as* severe in CMBS as in the more familiar residential mortgage-backed security (RMBS) market, in particular subprime RMBS. The individual loans in CMBS deals are for income-generating commercial properties as opposed to owner-occupied residential properties. Data on these loans, including rental income history, are widely available to market participants in a standardized form.[8] (In fact, rating agencies and CMBS investors use these loan-level data to evaluate the deals, in contrast to the approach in RMBS where market participants are usually forced to evaluate deals based on aggregate pool-level data.) Nevertheless, the potential for adverse selection still exists, not only because the originator may have "soft" information that is unreported but also because considerable discretion is involved in evaluating key underwriting characteristics.[9]

Data

[8]Commonly known as the CRE Finance Council Investor Reporting Package (IRP).

[9]For example, the two most important indicators of the soundness of a loan are the ratio of the loan amount to the assessed property value (LTV) and the debt-service coverage ratio (DSCR), which measures the borrower's monthly rental and other income relative to payment due on the mortgage. Over time, lenders increasingly used overoptimistic assumptions about property values (implying a lower LTV) and DSCRs based on estimates of future rents rather than actual or historical rental income. For details, see Black et al. (2011).

The data on loans and deals come from Realpoint LLC, a subscription-based CMBS rating agency and data provider. The 60,748 loans in the sample were originated between 1999 and 2007. 99 percent are fixed-rate mortgages. Most are 10-year loans with a 30-year amortization schedule, implying a balloon payment after 10 years.[10]

I observe loan characteristics at origination and the month in which the loan first becomes delinquent (if ever) until the censoring date, June 2010. The loan characteristics include the product type (amortization schedule and maturity), DSCR, LTV, occupancy rate, coupon spread (the contractual interest rate on the loan net of the rate on U.S. Treasuries for the corresponding maturity that were issued in the month of origination[11]), original loan amount, and the name of the originator. DSCR, occupancy rate, and loan amount all influence borrowers' ability to service their debt, while LTV represents their financial interest in the property. The coupon spread proxies for the perceived credit risk of the borrower by the lender.[12] The originator name is used to classify the originators into six types—commercial banks, investment banks, insurance companies, finance companies, foreign entities, and domestic conduit lenders—according to the identity of the originator's topholder parent firm. Due to heterogeneity in business models and institutional structure, the types have somewhat different incentives that may affect underwriting quality, a topic explored in greater depth by Black et al. (2011), who also provide more detailed discussion of how the firms were classified. For our current purposes, it is important merely to note that the CMBS deal underwriters and their loan-originating affiliates were all either commercial banks or investment banks.

The key dependent variable—the timing of default—is derived from the payment history of each loan. Specifically, I consider a loan to be in default as soon as it is 60 or more days delinquent or in special servicing. Another form of default can in principle occur at the maturity date of the loan if the borrower is unable to repay the entire balloon payment. However, in the data I do not observe such "balloon defaults" because practically none of the mortgages matures during the sample period.

The loans are contained in 590 CMBS deals with cutoff dates ranging from 2000 to 2007. For each deal, I observe the number of loans in the pool, the total principal balance of the pool at

[10] The data included one loan originated in 2008. Whenever estimating vintage effects, I treated this loan as part of the 2007 vintage.

[11] I interpolated rates for maturities not offered by the U.S. Treasury.

[12] It is not possible to draw strong conclusions about how risk was priced based solely on the summary statistics on coupon spreads. While coupon spreads depend in large part on the perceived credit quality of the borrower, they also vary according to the overall loan portfolio of the borrower, time-varying risk premia, and other factors affecting the cost of the loan.

the cutoff date, and the "constant prepayment rate (CPR)," which is the underwriter's forecast of the annual rate of prepayment for all loans in the pool. I also observe the tranche structure and ratings by the four main rating agencies for securities issued from the deal,[13] which allows me to construct the total principal balance of all tranches rated AAA.[14] Finally, I construct an indicator for whether each loan is securitized in-house by checking whether the originator's topholder parent matches the topholder parent of any of the CMBS lead underwriters.[15]

After dropping observations with missing data, the final estimation sample comprises 468 deals and 60,688 loans.

Stylized facts

Table 1 provides summary statistics for the observed loan characteristics and outcomes. The average loan is censored at 54 months. 9.3 percent of loans default, with 84 months being the average loan age at default. 16.6 percent of loans prepay, with 84 months being the average loan age at prepayment.

I determine whether a loan is securitized in-house according to whether the ultimate parent firm of the loan originator (in most cases the originator itself) is the same as the ultimate parent of the CMBS lead underwriter. In some cases, an originator may have two co-originators. Likewise, a CMBS deal may have two or three co-bookrunners. For such cases, I consider a loan to be securitized in-house if any of the originators and lead underwriters match. By this definition, 50.7 percent of all loans in the data are securitized in-house.

To descriptively summarize the conditional relationships between various loan characteristics and the propensity to default, Table 2 displays the results from estimating a Cox proportional hazards model in which the dependent variable is the time to default.[16] The specifications control for censoring, with either prepayment or the end of the sample period serving as the censoring event.

[13] The rating agencies are Moody's, Standard and Poor's, Fitch, and DBRS.

[14] Ratings data are missing for some rated tranches, so summing the balances for tranches that are nominally AAA in the data would lead to an underestimate of the total AAA balance. To help mitigate this problem, I assume that the tranches with the maximum reported level of credit support in a deal are always AAA, and take the total AAA balance to be the greater of the balance of these tranches and the balance of all tranches that are nominally AAA in the data.

[15] Topholder data are from the National Information Center. The in-house indicator controls for mergers. For example, a loan originated by Wachovia that was securitized by Wells Fargo after Wells Fargo acquired Wachovia is considered to be in-house.

[16] The proportional hazards assumption implies that at all ages t, the hazard function for each loan i is proportional to a flexible, baseline hazard that varies nonparametrically over time, $\psi(t)$, with the proportion being determined by observable covariates x_i. For details, see Cox (1972).

I include fixed effects for originators and for interactions between the ten geographic regions and the three property categories, which do not appear to affect the results substantially.[17] The three property categories are "office/retail/hotel," "multifamily" (apartments), and "industrial/other."[18]

The underwriting characteristics have the sorts of effects that we would expect. Interest-only loans are more likely to default, and fixed-rate mortgages are less likely to default. Loans for properties with higher DSCR and occupancy ratios and lower LTV, as well as smaller-sized loans, are less likely to default. The percentage of loans lacking reported numbers on DSCR, occupancy ratio, and LTV are 43 percent, 10 percent, and 1 percent, respectively.[19] To avoid having to drop these observations, I included an indicator for whether each of these variables is missing, and set the variable equal to zero for the missing observations. Not surprisingly, a higher coupon spread,[20] which proxies for the originator's perception of the riskiness of the borrower, is associated with a higher risk of default. The vintage effects also indicate that loan quality deteriorated steadily over time: loans originated in the 2007 vintage are about an eight-time greater hazard of default than loans originated in 2000 and before.

The most interesting coefficients are those for the indicator of whether a loan j is in-house and, for Specification II, the interaction of this indicator with the percentage of loans in the pool containing j that are in-house ("% in-house in deal"). The uninteracted effect of in-house status in Specification I is associated with a hazard of default that is approximately 10 percent lower than otherwise. This difference is not statistically significant, and reflects the net effect of both selection and causal factors.

The interaction effect with "% in-house in deal" in Specification II indicates that among deals containing a moderate to high proportion of in-house loans, the in-house loans have a much lower hazard of default than other loans. For deals containing a low proportion of in-house loans, the effect is reversed. A potential explanation for this interaction effect, which is formalized in the structural model, draws from the fact that empirically, loans originated by a particular originator tend to have similar observable characteristics. Specifically, this similarity applies to the set of

[17] The ten regions are "New England," "Mid-Atlantic," "Midwest, Eastern Region", "Midwest, Western Region," "Southern, Atlantic," "South-Central, East," "South-Central, West," "Western, Mountain," "Western, Northern Pacific," and "Other". The last category includes Hawaii, loans for properties outside the United States, and loans for properties in more than one region.

[18] The data distinguish between office, retail, hotel, industrial, and other. However, I aggregate some categories for the sake of parsimony in the structural model.

[19] A certain proportion of these missing cases could be due to imperfections in the data set, but more commonly, the missing values were were simply not reported in the underwriting documents. For example, there is no relevant measure of occupancy rate for most industrial properties.

[20] The spread over Treasuries of comparable maturity to the loan, as of the loan origination date.

in-house loans in a particular deal (which are all underwritten by the same originator, namely the deal underwriter), which presumably generates positive correlation in their returns. Therefore, the marginal benefit to the portfolio of an additional in-house loan diminishes with the number of in-house loans already in the pool. When the share of in-house loans is large, these loans must have a compensatorily lower marginal probability of default in order to rationalize holding them.[21] The purpose of the structural model, described next, is to disentangle the various causal and selection mechanisms embedded in these descriptive facts.

3 Model

The model has an exogenous set of CMBS deals $i = 1, \ldots, I$ and an exogenous set of loans $j = 1, \ldots, J$. Some loans are originated by the deal underwriters, with the remainder originated by lenders that do not underwrite CMBS deals and whose behavior I do not model. Each deal i is backed by a pool of loans, denoted by $\mathcal{J}_i$, which may include both loans that the underwriter of deal i has itself originated as well as loans obtained from other originators. The underwriter maximizes net profit, which depend upon the contents of its chosen portfolio $\mathcal{J}_i$ and the cost basis for the loans.

In the data, each underwriting firm does multiple CMBS deals at different points in time. I assume that underwriters maximize profits independently for each deal. That is, I abstract from profit spillovers across deals such as due to reputation effects. Given this assumption, each deal can be thought of as equivalent to a "firm," and the index i refers to both the deals as well as their underwriters.[22] For example, I model each deal underwritten by Bear Stearns as an independent profit-maximizing entity with its own index i.

I begin by describing a reduced-form model of the distribution of default times for individual loans followed by a model of the joint distribution of default times for different loans. The joint distribution parameters serve as inputs into the payoff structure of the pool-formation game, which I describe last.

[21]The interaction effect may also reflect differences in signaling costs depending upon the total number of loans originated by the underwriter. For example, attempting to sell 6 loans out of 10 sends a different message to potential buyers than attempting to sell 6 out of 100. Without further assumptions, the equilibrium impact on the quality of the retained loans is theoretically ambiguous. This paper does not explicitly model signaling costs, which I leave it to future research.

[22]For deals with co-lead underwriters, the index refers to the set of co-leads.

Distribution of individual loan outcomes

The gross payoff from a mortgage (i.e., excluding the cost basis) depends upon the promised stream of interest payments and the timing distribution for when the loan defaults. Consider a loan j in CMBS deal i underwritten by a firm with the same index i. This loan can be characterized by a vector of exogenous characteristics w_{ij}, the endogenously determined indicator of in-house status v_{ij} equal to one if underwriter i originated j, and characteristics ξ_j that are unobserved by the econometrician. The i subscript of w_{ij} allows for the existence of match-specific characteristics that are exogenous in the sense of being uncorrelated with ξ_j.[23] The market participants commonly observe the loan characteristics w_{ij} and v_{ij}. I defer discussion of market participants' information about ξ_j, which is relevant to the pool-formation game but not the reduced-form model.

The time to default for loan j, T_j (normalized to be in terms of months since the cutoff date of pool i) is distributed as follows:

$$Pr(T_j < t|w_{ij}, \xi_j) = 1 - exp(-\int_0^t \psi(\tau)e^{w'_{ij}\alpha_1+\xi_j})d\tau \tag{1}$$

Conditional on the random term ξ_j, the above function entails the standard proportional hazards assumption. Namely, the hazard of default at time t, $Pr(T_i = t|T_i \geq t)$, is equal to the product of the "baseline" hazard function $\psi(t)$ and the constant proportion $exp(w'_{ij}\alpha_1+\xi_j)$. I allow the distribution of ξ_j to depend upon the in-house status of the loan. Denoting the conditional distribution by $H(\xi_j|v_{ij})$, the distribution of the time to delinquency unconditionally on ξ_j is as follows:

$$Pr(T_j < t|w_{ij}, v_{ij}) = 1 - \int exp(-\int_{\tau=0}^t \psi(\tau)e^{w'_{ij}\alpha_1+\xi_j})d\tau \; dH(\xi_j|v_{ij}) \tag{2}$$

Given that the model of individual loan performance is reduced-form, the dependence of $H(\xi_j|v_{ij})$ on v_{ij} captures the cumulative effects of any selection and causal influences. More formally, we can conceptualize ξ_j as being decomposable into $\xi_j = \alpha_0 v_{ij} + \tilde{\xi}_j$, where α_0 is the *causal* effect of in-house status on the ex post performance of the loan, and the distribution of the residual $\tilde{\xi}_j$

[23] In the empirical implementation, the only match-specific characteristic in w_{ij} is the age of the loan at the deal cutoff date (counting from the origination date)—that is, the "loan seasoning." I rule out that loan seasoning is endogenously determined, based on the fact that in the originate-to-distribute industry model, most loans are securitized as quickly as possible after origination. I include loan seasoning as an explanatory variable to adjust for the fact that I model default time as a distribution over the *loan's* age, whereas I compute portfolio returns for a CMBS deal as of the deal cutoff date. Therefore, the coefficient on loan seasoning will be necessary for adjusting the expected returns on the loan for counterfactual scenarios in which the loan is bundled into an alternative deal.

conditional on v_{ij} captures the effects of selection.[24] The reduced-form distribution $H(\xi_j|v_{ij})$ is identified in the model of individual loan performance described here; the causal parameter α_0 is identified in the structural model described later in the paper; and the selection effect—captured by the distribution of $\tilde{\xi}_j$—is implied by netting α_0 from the distribution of ξ_j.

I make a number of simplifying assumptions. First, I abstract from the effects of prepayment. CMBS loans typically have heavy prepayment penalties designed to compensate the lender for any yield loss caused by prepayment.[25] By ignoring prepayment, the model effectively makes the assumption that when a loan prepays, the the prepaid amount and penalty fees are equivalent to the expected continuation value if the loan did not prepay. This simplification allows us to model delinquency as a simple hazard process while treating prepayment as a censoring event, without having to model the "competing" hazard of a loan prepaying.

Second, I treat default as a terminal event at which point the lender recovers a share $1 - LGD$ of the remaining balance, where LGD represents the loss given default.[26] Thus I abstract from cases in which a delinquent loan becomes current again, or loans that become delinquent multiple times.[27]

Finally, I ignore losses due to balloon default. Because the actual maturity of most loans is much longer than the sample period, the probability of balloon default is not identified by the data. The operating assumption is that the magnitude of losses is, to a first order, captured by default during the term of the loan.

Along with the timing of default, loan returns also depend upon the contractually determined coupon rate and amortization schedule of the mortgage. For simplicity, I treat all loans as having fixed coupon rates and a constant rate of amortization,[28] and denote the implied monthly principal and interest payment by P_j and the remaining balance on the loan at time t by $B_j(t)$. At time 0,

[24] This decomposition implicitly assumes that the causal effect is nonrandom, and captures selection effects through randomness in the distribution of $\tilde{\xi}_j$.

[25] Unlike for residential real estate loans, prepayment penalties for commercial mortgages largely eliminate the incentive of borrowers to refinance when market interest rates decline. In the industry, it is widely recognized that a more common reason for prepayment is if the borrower wants to free up cash for alternative investments. For example, see Gollenberg (1997), pp. 153-154.

[26] Instead of making LGD dependent on j, for simplicity I fix LGD at 0.33, which is a typical value attained in commercial mortgages during this time period.

[27] In the data, only 39.6 percent of loans more than 90 days delinquent ever become current again. Of the loans that recover, 13.1 are delinquent again by the censoring date, which is typically just a few months after the initial default.

[28] In the data, 95.6 percent of all mortgages are fixed-rate.

the distribution of the return on loan j, denoted by the random variable Y_j, is thus

$$Pr(Y_j < y_j) = \inf_{t_j} \left\{ Pr(T_j < t_j) \quad \middle| \quad \sum_{\tau=0}^{t} \left(\prod_{\tau'=0}^{\tau-1} \delta_{\tau'} \right) P_j + \delta^t (1 - LGD) B_j(t) < y_j \right\}. \tag{3}$$

Note that the exogenous single-period discount rate, δ_t, differs across periods, in general.

Joint distribution of loan outcomes

Payoffs to different tranches of a CMBS pool are determined by the joint distribution of returns on the constituent loans. Given any joint distribution of default times (and by extension a joint distribution of returns), a copula function exists that links the marginal default-time distributions to the joint distribution.[29] During the pre–financial-crisis period, securitization industry practitioners typically assumed that the copula is multivariate normal. Since the crisis, many have observed that normal copulae–as opposed to certain alternative distribution families[30]—understate the degree of dependence in the tails of the distributions even if they correctly capture the *correlation* structure of default. Thus, the normal-copula assumption tends to understate the risk of extreme events in which all loans perform poorly. However, I maintain the normality assumption for two reasons. First, because it was most commonly used by market practitioners, it correctly captures at least the *subjective* payoffs perceived by the agents whose incentives my model tries to capture. Second, as discussed below, the problem of censoring makes it extremely difficult to estimate the copula using classical approaches, whereas Bayesian estimation with a normal prior is straightforward due to the self-conjugacy property.

More formally, define $F_j(T_j) \equiv Pr(T_j < t | w_{ij}, v_{ij})$ and consider the loans in deal i's portfolio $\mathcal{J}_i$, indexed by $j = 1, 2, \ldots, J_i$. The joint distribution of $T_1, T_2, \ldots, T_{J_i}$ is given by

$$F(T_1, T_2, \ldots, T_{J_i}) = \Phi(\Phi^{-1}(F_1(T_1))\Phi^{-1}(F_2(T_2)), \ldots \Phi^{-1}(F_{J_i}(T_{J_i})) \; ; \; \Omega), \tag{4}$$

where Φ^{-1} is the inverse of the standard normal distribution and $\Phi(u_1, u_2, \ldots, u_{J_i} \; ; \; \Omega)$ is the multivariate normal distribution with covariances Ω. Positive or negative correlation in the time to default for any two loans —captured by the off-diagonal terms of Ω implies—implies correlation in loan returns via equation (3).

[29]By Sklar's theorem, the copula is unique if the joint distribution is continuous. For details on the estimation of copulae, see Trivedi and Zimmer, (2005).

[30]Examples include Student's t and Archimedian.

Underwriter utility

Each deal underwriter i maximizes its utility with respect to a set of *feasible* portfolios, which I define as follows. The cutoff date for each deal i, t_i, and the origination date for each loan j, t_j, are exogenously determined. I assume that each loan j can feasibly be matched with any deal i such that such that $t_i \geq t_j$ and $t_i < t_j + D$ days. I denote the set of loans that can feasibly be matched with firm i by $\mathcal{D}_i$.[31] The powerset $\mathcal{P}(\mathcal{D}_i)$ then denotes the set of feasible portfolios for i. This definition of feasibility implies that loans can be matched with deals with cutoff dates just after the origination date of the loan, but not too long after. In other words, underwriters can foresee future deals, but the foresight may be limited to the near future.[32]

Firm i's gross utility is a function of its chosen portfolio, $\mathcal{J}_i \in \mathcal{P}(\mathcal{D}_i)$. For simplicity, the utility function aggregates returns for all tranches subordinate to the AAA bond, which I collectively denote as the "B piece." In other words, I differentiate between a dollar of principal or interest going to the AAA bondholder versus a dollar going to the holder of any security subordinate to the AAA bond, but do not differentiate among the junior tranches. This simplification is borne of necessity: deal structures below the AAA tranche are typically very complex and idiosyncratic, preventing direct comparison of most junior tranches across deals.[33] The expected gross utility function (not including payments for the sale or purchase of loans) is as follows:

$$
\begin{aligned}
u_i \quad = \quad & E_{t_i|w,v}\Big[\min\Big\{b_i(\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i})\ ,\ \textstyle\sum_{j\in\mathcal{J}_i} y_{ij}\Big\} + \\
& \quad \beta_t \cdot \max\Big\{0\ ,\ \textstyle\sum_{j\in\mathcal{J}_i} y_{ij} - b_i(\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i})\Big\}\Big] \ + \ \textstyle\sum_{j\in\mathcal{J}_i} z_{ij}.
\end{aligned} \tag{5}
$$

The "min" and "max" terms are the values of the AAA bond and the B piece, respectively. The expectation is taken with respect to the distribution of portfolio returns, $\sum_{j\in\mathcal{J}_i} y_{ij}$, conditional on observed data and model parameters. The expectation is not conditioned on the unobserved

[31] The model identification is robust to the actual feasible set being a superset of the set defined here, so long as the distribution of unobservables–described later in the section—has the same mean for both the smaller and larger sets.

[32] In practice, I choose $D = 90$ days. We observe a small number of loans contained in deals whose cutoff dates are more than 90 days after the loan origination date. As an alternative to choosing D large enough so that exceptions never occur, I instead rationalize the data by modifying the feasible set for these cases. Specifically, I assume that the set of deals that could feasibly be matched with such loans j comprises the set of all deals i' with $t_{i'} : t_j < t_{i'} \leq t_i$.

[33] CMBS deal structures are considerably simpler than analogous structures for residential mortgage-backed securities (RMBS), but are still complicated. The promised streams of payments to bondholders typically entail payments that depend upon both the magnitude and timing of principal and interest shortfalls. Deals also became more complex for later vintages, a fact documented by Furfine (2010) and also reflected in the fact that the average number of tranches increased from 14.5 in 2000 to 26.4 in 2007.

heterogeneity in loans' propensity to default, ξ_j. Rather, firm i private *beliefs* about ξ_j are reflected by the term z_{ij}. More precisely, z_{ij} captures the discrepancy between i's private expectations about the return on the loan (given w_{ij} and v_{ij}) and the econometrician's expectations based on the unconditional distribution of ξ_j. The term $\sum_{wj\in\mathcal{J}_i} z_{ij}$ is the cumulative effect of i's private shocks on i's valuation of the portfolio, where the additivity is based on the assumption that that firm i's private shocks are independent. A key advantage of the empirical approach detailed in the next section is that we do not have to make stronger assumptions about the information sets of the players. In particular, the model does not need to specify firm i's beliefs about competing firms' private signals.

The term $b_i(\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i})$ is the principal amount of the AAA bond tranche. Conceptually, we should think of the underwriter as deriving utility from *revenues* from selling the AAA bond. However, modeling investor demand is constrained by the limited availability of price data, so the utility function instead specifies that the utility attributable to the AAA bond is proportional to i's expected value of the cashflows from the bond.[34] Equating security prices with the expected value of promised cashflows is entirely standard in the asset-pricing literature. However, a natural agenda for future research is to obtain better price data that would allow us to avoid having to impose this equality.

The value of the B piece is the expected value of the portfolio returns, $\sum_{j\in\mathcal{J}_i} y_{ij}$, minus payments to the AAA bondholders. The time-varying parameter β_t captures the utility of B-piece returns relative to AAA bond returns (which are normalized to one). The parameter β_t can be thought of as reflecting either risk-aversion or the opportunity cost of capital. Risk aversion implies that the utility function discounts expected returns on the B piece, which has more volatile returns compared with the AAA bond. An opportunity cost of capital arises because in contrast to proceeds from the sale of the AAA bond, which are realized at the cutoff date, cashflows from the B piece are realized over the life of the deal. While I do not constrain the value of β_t in estimation, the fact that we observe underwriters selling any bonds at all as opposed to retaining the entire portfolio suggests that $\beta_t < 1$.

The size of the AAA tranche, $b_i(\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i})$, is an exogenous function of the observable portfolio characteristics. Specifically, I assume that the AAA tranche for deal i is required to have enough credit support such that conditional on $\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i}$, the probability of it incurring any

[34]Data on the prices at which bonds were offered initially offered to investors are available for only 53 percent of the tranches, largely because many of the deals were not publicly registered. If investors are risk-neutral and have the same information set as the underwriter, and if underwriters do not have market power, then a standard model would imply that revenues from bond sales would equal the expected value of the bond.

losses is equal to some probability p_i. This assumption holds true as a first-order approximation in the CMBS industry, where the size of the AAA tranche is generally prescribed by the credit rating agency (or agencies).[35] At least in principle, the size of the AAA tranche is supposed to reflect the rating agency's assessment of how much credit support is needed to keep the risk of principal losses below what it deems to be the "AAA" threshold. This assumption holds only approximately, because in practice, the deal underwriter may exercise a degree of control over the AAA bond size through the ability to engage in "rating shopping." For details on the rating shopping phenomenon, see Cohen (2011).[36] Moreover, at least in principle, the amount of credit protection may signal the underwriter's private information and thus affect the demand for the securities.[37] However, endogenizing the choice of bond size is beyond the scope of this paper, where the focus is on the portfolio decision. However, it would be an important extension for future work.

Trading and necessary conditions for equilibrium

After a loan j is originated, it may be bought and sold an arbitrary number of times between pairs of firms that can feasibly hold j. Firm i observes the exogenous characteristics $\{w_{ij}\}$ and the private signal z_{ij} before buying or selling loan j.[38] I assume that loans are indivisible, though this assumption is for ease of exposition and does not drive any of the model implications.

When a firm i sells a loan j to firm i', in exchange, firm i' makes a transfer payment $c_j^{ii'}$ to i. The transfer payment depends upon the identities of the loan j and the transacting firms (i and i'), but I assume that it is independent across transactions, implying that a firm's total transfer payments are additive across loans. I express $c_j^{ii'}$ as a function of observables, $c_j^{ii'} = A_j \cdot \gamma' w_j + \zeta_j^{ii'}$, where A_j is the principal amount on loan j, w_j is the vector of exogenous and non–match-specific loan characteristics[39], γ are parameters, and $\zeta_j^{ii'}$ is an unobserved error.

[35] If a deal ends up being rated by two or more rating agencies, the chosen size of the AAA tranche generally reflects the most conservative structure among the structures prescribed by the respective agencies. For details on the rating process, see Cohen (2011).

[36] Investors typically require only one or two ratings on a deal. Because there are multiple competing rating agencies whose assessments of the pool quality may exhibit some degree of randomness, underwriters can typically lower the level of credit support required for AAA simply by obtaining preliminary quotes from more rating agencies than they ultimately plan to use, and then hiring the rating agency with the laxest requirements. Rating shopping exploits two phenomena: (1) the fact that the required credit support decreases in the number of agencies shopped due to a purely mechanical order-statistic effect; (2) rating agencies have an incentive to compete for the underwriter's business by offering looser standards.

[37] Theory papers by DeMarzo and Duffie (1999), DeMarzo (2005), and Glaeser and Kallal (1997) provide theoretical insights on the role of signaling.

[38] Recall from footnote 23 that the only match-specific characteristic in $\{w_{ij}\}$ is the loan seasoning at the cutoff date conditional on j being matched to i.

[39] w_j does not include loan seasoning at the cutoff date, which is the only thing that maes it different from w_{ij}.

Optimality requires that firm i's net profit given its chosen portfolio $\mathcal{J}_i$, taking into account transfer payments, must be weakly higher than the net profit given any alternative feasible portfolio. In particular, i's actual profit must be weakly higher than that delivered by any alternative feasible portfolio that can be obtained by modifying the chosen portfolio through trades with firm i'. $\mathcal{J}_{i'} \cap \mathcal{D}_i$ is the set of all loans in the portfolio of i' that firm i can feasibly hold. The powerset $\mathcal{P}(\mathcal{J}_i \cup (\mathcal{J}_{i'} \cap \mathcal{D}_i))$ comprises all alternative portfolios that i could feasibly choose based on all loans in the actual portfolios chosen by i and i'. Letting $A \backslash B$ denote the set of elements of A that are not in B, the optimality condition is as follows. Let $U_i(\mathcal{J}_i)$ represent utility net of transfer payments and private information (i.e., everything in equation 5 other than $\sum_{j \in \mathcal{J}_i} z_{ij}$). Formally, the optimality conditions can be stated as follows:

$$E_{t_i|w,v}\left[U_i(\mathcal{J}_i) - U_i(\mathcal{J}_i')\right] \; - \; \sum_{j \in \mathcal{J}_i \backslash \mathcal{J}_i'} c_j^{ii'} \; + \; \sum_{j \in \mathcal{J}_i \backslash \mathcal{J}_i'} z_{ij} \; + \; \sum_{j \in \mathcal{J}_i' \backslash \mathcal{J}_i} c_j^{ii'} \; - \; \sum_{j \in \mathcal{J}_i' \backslash \mathcal{J}_i} z_{ij} \geq 0$$
$$\forall i, i' \quad \forall \mathcal{J}_i' \in \mathcal{P}(\mathcal{J}_i \cup (\mathcal{J}_{i'} \cap \mathcal{D}_i)). \tag{6}$$

4 Estimation and Identification

This section discusses estimation and identification for each stage of the model, in sequence.

Estimation and identification of individual loan outcome distributions

I estimate the loan default process for individual loans using a semiparametric mixtures estimator, related to the nonparametric maximum likelihood approach (see Heckman and Singer, 1984) and more recent work by Bajari et al. (2010). Rather than specify a parametric distribution for the unobserved heterogeneity in the hazard function (2), I assume that the random coefficient ξ_j is drawn from a discrete grid $\{\xi^r\}_{r=1,\ldots,R}$. I use $\theta(r|v_{ij}) \equiv Pr(\xi_j = \xi^r | v_{ij})$ to denote the conditional distribution over the support, given the in-house status of the loan, v_{ij}. (Recall from Section 3 that the conditioning on v_{ij} reflects the reduced-form impact of both causal factors and selection). With the discretized distribution support, equation 2 can be reexpressed as

$$Pr(T_j < t | w_{ij}, v_{ij}) = 1 - \sum_{r=1}^{R} \theta(r|v_{ij}) exp\left[-\int_{\tau=0}^{t} \psi(\tau) d\tau \; e^{w_{ij}'\alpha_1 + \xi^r}\right]. \tag{7}$$

The semiparametric specification is both more flexible than the traditional parametric approach as well as computationally less burdensome. One the one hand, the set of parameters is larger in the semiparametric case because it includes the frequencies $\theta(r,|v_{ij})$. On the other hand, not having to simulate over the possible values of ξ_j means that we can estimate all of the parameters using maximum likelihood without taking simulation draws, an advantage that outweighs the larger number of parameters unless the grid is chosen to be extremely fine.[40]

For all loans j, let t_j denote the default time (possibly unobserved) and let t_j^c denote the lesser of the default time and the censoring time defined by the data sample. The estimation procedure can be stated as the following constrained maximization.

$$
\begin{aligned}
\max_{\alpha_1,\theta,\psi} \sum_j \sum_{r=1}^{R} \Bigg[\theta(r|v_{ij}) \quad & \mathbf{1}(t_j > t_j^c) \cdot exp\left[-\int_{\tau=0}^{t_j^c} \psi(\tau) d\tau \; e^{w_{ij}'\alpha_1 + \xi^r} \right] \\
+ \quad & \mathbf{1}(t_j = t_j^c) \cdot exp\left[-\int_{\tau=0}^{t_j} \psi(\tau) d\tau \; e^{w_{ij}'\alpha_1 + \xi^r} \right] \psi(t_j) e^{w_{ij}'\alpha_1 + \xi^r} \Bigg] \\
s.t. \quad & \sum_{r=1}^{R} \theta(r|v_{ij}=0) \;=\; 1, \\
& \sum_{r=1}^{R} \theta(r|v_{ij}=1) \;=\; 1, \\
& 0 \;\leq\; \theta(r|v_{ij}) \;\leq\; 1, \qquad \forall\, r,\ v_{ij}.
\end{aligned}
\tag{8}
$$

The parameters of this model are identified as follows. The non-random parameters, α_1, and the conditional means (i.e., the "location") of the random coefficient, $E(\xi_j|v_{ij}=0)$ and $E(\xi_j|v_{ij}=1)$, are identified by correlation between the covariates and the propensity for loans to default by a given maturity. The distribution of ξ_j around the conditional means is identified by violations of strict proportionality in the hazard. In particular, loans with worse unobservable characteristics (higher ξ_j) are less likely to survive, implying that conditional on survival to a given maturity, loans with high ξ_j must have better observable characteristics on average than loans with low ξ_j. Conditional on survival to t, the quality of unobservable and observable characteristics becomes increasingly negatively correlated with as t increases. Therefore, greater dispersion in ξ_j implies a

[40] If *all* of the regression coefficients, including α_1, were random and distributed over a grid, then the loglikelihood function would be globally concave, ensuring convergence to the global optimum. (The estimation objective function would also be globally concave if we instead used a constrained linear least squares approach, as described in Bajari, Fox, Kim, and Ryan (2010)). For our chosen specification, the loglikelihood function is not globally concave in α_1, which enters nonlinearly. However, the loglikelihood is still relatively well-behaved because it is globally concave in θ. In practice, it is easiest to first estimate the model with nonrandom coefficients for all explanatory variables in order to establish a good starting value, before estimating the full model with random effects.

greater degree of weakening over time in the observed correlation between the covariates and the intensity of default.

The baseline hazard is identified by variation over time in the unconditional propensity to default. In general, the baseline hazard $\psi(t)$ is nonparametrically identified (see Cox, 1984), but because estimating the structural model requires simulating portfolio outcomes, I specify $\psi(t)$ as a polynomial function in order to reduce the computation burden.[41]

Estimation and identifation of joint loan outcome distribution

For tractability, I impose certain restrictions on Ω, the correlation structure of the copula function (4). Specifically, I assume that the variation in default times can be decomposed in terms of an idiosyncratic factor and factors for ad hoc loan categories based on observable loan characteristics. Specifically, I categorize all loans into $K = 30$ different categories corresponding to each unique combination of the three property types ("office/retail/hotel," "multifamily," and "industrial/other") and ten regions (e.g., "Mid-Atlantic," "Western, Northern Pacific"). Let $k(j)$ denote the category to which j belongs. For a loan j in CMBS deal i, I assume that the distribution of the time to default, $Pr(T_j < t|w_{ij}, v_{ij}, i)$, depends upon the idiosyncratic factor η_{ij} and the category-specific factor $\varepsilon_{ik(j)}$ in the following way.

$$\begin{aligned} F(t|w_{ij}, i) = Pr(T_j < t|w_{ij}, v_{ij}, i) = \Phi(\varepsilon_{ik(j)} + \eta_{ij}), \\ \text{with} \quad \varepsilon_i \equiv (\varepsilon_{i1} \ldots \varepsilon_{iK}) \sim N(\mathbf{0}, \Omega_\varepsilon), \quad \eta_{ij} \sim N(0, \omega_\eta), \quad \text{and} \quad E[\eta_{ij}\varepsilon_i] = \mathbf{0}. \end{aligned} \tag{9}$$

The category-specific factors are jointly distributed $N(0, \omega_\epsilon)$ and the idiosyncratic factor is distributed $N(0, \omega_\eta)$. The off-diagonal components of Ω_ε capture factor correlations across categories. Default times may also be correlated *between* deals i and $i' \neq i$, but I do not specify or attempt to estimate the correlation structure across deals.

In theory, the covariance structure can be estimated jointly with the hazard model. However, estimating Ω_ε and ω_η in a separate step is preferable because of the difficulty of computing the "full-information" likelihood. The censoring problem makes the regions of integration for the likelihood function extremely complex, even with the restrictions implied by the factors.[42] To get around this problem, I instead use a Bayesian approach and estimate Ω_ε and ω_η using an hierarchical

[41]Specifically, I assume that $\psi(t)$ is the square of a quadratic function of t, which is a 4th-degree polynomial. Restricting $\psi(t)$ to be the square of a quadratic function ensures that the baseline hazard is always positive.

[42]If there were no censoring, Ω_ε and ω_η can be consistently estimated by the correlation structure of the marginal probabilities $Pr(T_j < t_j|\hat{\alpha}, \hat{\theta}, \hat{\psi})$, where t_j are the empirical default times and the estimates $\hat{\alpha}, \hat{\theta}, \hat{\psi}$ are the hazard-model estimates.

Markov-Chain Monte Carlo algorithm that combines Gibbs sampling with a Metropolis-Hastings algorithm that draws from the posterior for the latent random variables ε_{ik} and η_{ij}. This procedure is detailed in the appendix.

Estimation and identifation of structural parameters

The effects of the exogenous covariates w_{ij} on the distribution of payoffs and the correlation structure of payoffs are identified by the previous stages of the model. However, the parameters α_0, β_t, and γ must be identified from optimality conditions on the firms' decisions. For each proposed value of these remaining structural parameters, the estimation procedure evaluates a set of moment conditions by simulating over distributions of portfolio returns, using as inputs the parameter estimates from the previous two steps. The fact that most of the parameters are already identified by the initial steps implies considerable savings in computation burden when estimating the structural model.

The payoff structure of the structural model hinges on the distribution of portfolio returns. Given any distribution of default times, equation 3 produces the implied distribution of returns expressed as a present-discounted value. For loans in deal i, I set the discount rate for period t, δ_t, to be the discount rate implied by the Treasury yield curve as of the cutoff date for deal i.

A key issue is how to construct valid moment conditions. If all of the terms in the Nash inequalities (equation 6) were observed, then these inequalities would provide a set of conditions restricting the possible values for the structural parameters. However, the private signals z_{ij} and transfer payments $c_j^{ii'}$ are unobserved and have an unknown correlation with the match between loans and firms. In particular, it is reasonable to think that z_{ij} and $c_j^{ii'}$ are positively and negatively correlated, respectively, with the event that loan j is matched with firm i.

To handle the foregoing endogeneity problem, I use a strategy first developed by Pakes et al. (2006). Namely, I find particular linear combinations of the Nash inequalities such that the contribution of unobservables either: (1) cancels out altogether; or (2) is uncorrelated or negatively correlated with the differences in utility due to the observable components of the model. These linear combinations satisfy moment inequalities that either fully identify or at least set-identify the parameters.

I form these moment inequalities under two alternative sets of assumptions about the information structure of the game.

Assumption 1 $z_{ij} = z_{i'j}$, *for all* i, i'.

Assumption 1 specifies that firms have symmetric information about the loans. This restriction implies that when the buyer and seller transact over a loan j, the selling firm's forgone private valuation due to unobserved characteristics of loan j equals the buying firm's marginal private valuation of j due to unobservable characteristics. Together with the fact that the transfer payment paid by the buyer necessarily equals the transfer paymnet received by the seller, Assumption 1 implies that we can "difference out" the two unobservables, z_{ij} and $c_j^{ii'}$, by summing the inequalities for the two firms. This approach is analogous to the approach taken by Ho's (2009) work on insurer-provider networks in the medical care market.

Without loss of generality suppose that we observe loan j matched to firm i rather than another firm i' to which j could also feasibly be matched. The Nash condition implies that the firms are each better off with their observed portfolios than if loan j were matched to i' instead of i. Let $\mathcal{J}_i \backslash j$ be the set of all loans in i's porfolio other than j, and let $\mathcal{J}_{i'} \cup j$ be the portfolio formed by adding j the observed portfolio of firm i'.

For all firms i and i' and loans $j \in \mathcal{J}_i$:

$$\begin{aligned} \Delta\pi_i(\mathcal{J}_i, \mathcal{J}_i\backslash j) &\equiv E_{t_i|w,v}\left[U_i(\mathcal{J}_i) - U_i(\mathcal{J}_i\backslash j)\right] - A_j \cdot \gamma' w_j - \zeta_j^{ii'} + z_{ij} \geq 0 \\ \Delta\pi_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \cup j) &\equiv E_{t_{i'}|w,v}\left[U_{i'}(\mathcal{J}_{i'}) - U_{i'}(\mathcal{J}_{i'} \cup j)\right] + A_j \cdot \gamma' w_j + \zeta_j^{ii'} - z_{i'j} \geq 0. \end{aligned} \tag{10}$$

In computing the actual utilities $U_i(\mathcal{J}_i)$ and $U_{i'}(\mathcal{J}_{i'})$ the obvious choice for the size of the AAA tranche $b_i(\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i})$ is to set it to the empirical tranche size, b_i^*. A problem we have to deal with is is how to determine the AAA tranche size in computing utilities under the alternative scenarios. I assume that that the probability of losses to each deal i's AAA tranche, conditional on observable portfolio characteristics, is held to an exogenous probability p_i under all alternative choices. Thus if we let $G_{\mathcal{J}_i}(\cdot)$ and $G_{\mathcal{J}_i'}(\cdot)$ represent the return distributions for the actual portfolios $\mathcal{J}_i$ and $\mathcal{J}_{i'}$; and let $G_{\mathcal{J}_i,\mathcal{J}_i\backslash j}(\cdot)$ and $G_{\mathcal{J}_i,\mathcal{J}_i\cup j}(\cdot)$ represent the distributions for the counterfactual portfolios; then the tranche sizes under the counterfactual scenarios are given by $G^{-1}_{\mathcal{J}_i,\mathcal{J}_i\backslash j}(G_{\mathcal{J}_i}(b_i^*))$ and $G^{-1}_{\mathcal{J}_i,\mathcal{J}_i\cup j}(G_{\mathcal{J}_i}(b_{i'}^*))$. In the estimation procedure, I must simulate the portfolio distribution under each scenario. The simulation procedure is complicated and is described in Appendix C.

In expression 10, the terms $\zeta_j^{ii'}$, z_{ij} and $z_{i'j}$ are correlated with the observable parts of the inequalities, implying that the latter are not necessarily ≥ 0 in expectation. However, we can construct three types of moment conditions based on linear combinations of the inequalities. Doing so requires some additional notation. Let $\chi(i,i',j)$ be an indicator function equal to 1 if firm i holds loan j and 0 if i' holds loan j. Denote the observed components of $\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)$ and $\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)$ by $\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)=E_{t_i|w,v}\left[U_i(\mathcal{J}_i)-U_i(\mathcal{J}_i\backslash j)\right]-A_j\cdot\gamma' w_j$ and $\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)=E_{t_{i'}|w,v}\left[U_{i'}(\mathcal{J}_{i'})-U_{i'}(\mathcal{J}_{i'}\cup j)\right]+A_j\cdot\gamma' w_j$. Let $x_j^{ii'}$ be a set of instruments that are in the information sets of firms i and i' when they are deciding whether to transact. First consider the following linear combination of optimality conditions relative to dropping loan j:

$$\begin{gathered}\chi(i,i',j)[\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)]+\chi(i',i,j)[\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash j)]=\\ \chi(i,i',j)[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+z_{ij}-\zeta_j^{ii'}]+\chi(i',i,j)[\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash j)+z_{i'j}-\zeta_j^{ii'}]\geq 0.\end{gathered}\tag{11}$$

Because $z_{ij}=z_{i'j}$ by Assumption 1, the above expression contains the same combination of unobservables regardless of whether i holds the loan ($\chi(i,i',j)=1$) or i' ($\chi(i',i,j)=1$) holds the loan. Therefore, the expectation of the unobservables over i, i', and $j\in\{\mathcal{J}_i\cup\mathcal{J}_{i'}\}\cap\mathcal{D}_i\cap\mathcal{D}_{i'}$ (the set of loans in the firm's actual portfolios that could feasibly be matched with either firm), conditional on instruments $x_j^{ii'}$, is equal to zero: $\mathcal{E}[\chi(i,i',j)\cdot(z_{ij}-\zeta_j^{ii'})+\chi(i',i,j)\cdot(z_{i'j}-\zeta_j^{ii'})\ \mid\ x_j^{ii'}]=0$. Therefore, the unobservables drop out of the expectation of inequality 11:

$$\mathcal{E}\left[R_1(i,i',j)|x_j^{ii'}\right]\equiv\mathcal{E}\left[\chi(i,i',j)\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\chi(i',i,j)\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash j)|x_j^{ii'}\right]\geq 0,\tag{12}$$

By a symmetric argument, we can establish a second type of moment condition based on the optimality of the observed portfolios relative to adding a loan j:

$$\mathcal{E}\left[R_2(i,i',j)|x_j^{ii'}\right]\equiv\mathcal{E}\left[\chi(i',i,j)\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\cup j)+\chi(i,i',j)\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)|x_j^{ii'}\right]\geq 0,\tag{13}$$

The third type of moment inequality is constructed by summing the inequalities for firms i and i' in 10 conditional on firm i holding a loan j ($\chi(i,i',j)=1$), which expresses the total gains to trade. The transfer payment nets out and by Assumption 1, z_{ij} and $z_{i'j}$ cancel each other out:

$$\chi(i,i',j)\left[\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)\right]=\chi(i,i',j)\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)\right]\geq 0.\tag{14}$$

Taking expectations,

$$\mathcal{E}\left[R_3(i,i',j)|x_j^{ii'}\right] \equiv \mathcal{E}\left[\chi(i,i',j)\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)\right]|x_j^{ii'}\right] \geq 0, \tag{15}$$

The second alternative assumption allows firms to have asymmetric information. However, it assumes that the loans belong to observed categories $k = 1, ..., K$, and imposes that each firm i's private information is the same for all loans within a given category. Denoting the category of loan j by $k(j)$,

Assumption 2 *$\forall i$ and $\forall j, j'$, if $k(j) = k(j')$, then $z_{ij'} = z_{ij} \equiv z_{ik(j)}$ and $\zeta_j^{ii'} \equiv \zeta_{k(j)}^{ii'}$.*

Note that Assumptions 1 and 2 are non-nested: the former imposes symmetric information across *underwriters*; the latter imposes symmetric information about *loans* within a category. A key modeling issue is how to define the categories, which must be sufficiently fine so that each underwriter can reasonably be thought of as having homogeneous private information about all loans within a category. Specifically, for each underwriter i, I consider all loans in $\mathcal{D}_i$ that are by a particular originator as comprising a category. For example, I assume that for a deal i underwritten by Bear Stearns, the underwriter may have private information based on its general impression of all loans originated by the loan originator New Century around the time of deal i, but rules out private information about individual loans. Assumption 2 seems intuitively weaker than Assumption 1, so in the Results section, I place greater emphasis on findings based on Assumption 2.

Similar to the case for Assumption 1, we can exploit the symmetry of the transfer payment for the buyer and the seller. However, we can no longer exploit symmetry between z_{ij} and $z_{i'j}$. Instead, my approach is analogous to the *ordered choice* approach (Ishii, 2005; Smith, 2011), which is adapted to models in which the choice set can be logically ordered in some way (in this case, by the number of loans in a portfolio from a particular originator) and the marginal contribution of unobservables to agents' utility is constant over that ordering (which in this case is implied by Assumption 2). By finding linear combinations of the inequalities such that each unique value of z_{ik} receives equal weight, we can construct moment conditions that are unbiased by correlation between unobservables and the observed number of loans chosen from each category.

A slight complication is that for any pair of firms i and i', there may possibly be categories of loans that are held by one firm but not the other. This creates a "boundary problem." For example, if only i' but not i holds any loans from category k, we cannot compare i's utility before and after dropping a loan from this category. Simply ignoring the comparison creates a potential for bias, because the expectation of $z_{ik(j)} - \zeta_{k(j)}^{ii'}$ is negative conditional on firm i holding zero loans from category k. Similarly, we cannot compare the utility of firm i' before and after adding a loan from category k. The solution is to treat the inequalities at the boundaries as missing observations. We substitute these missing observations with a random variable with a known inequality relationship with the unobservables, and then average over the full sample.

Specifically, let n_{ik} represent the observed number of loans in category k held by firm i. Let $\mathcal{J}_{i'}^{K(i)}$ denote the set of loans in the portfolio of firm i' that are in categories observed in the portfolio of firm i. Thus, $\mathcal{J}_{i'} \backslash J_{i'}^{K(i)}$ refers to the set of loans in the portfolio of firm i' that belong to categories not found in the portfolio of firm i. Let $\tilde{\chi}(i, i', j)$ be an indicator function equal to 1 if $j \in \mathcal{J}_{i'} \backslash \mathcal{J}_{i'}^{K(i)}$ and 0 otherwise. The Nash conditions imply the following:

$$
\begin{gathered}
\tfrac{\chi(i,i',j)}{n_{ik}} \Delta\pi_i(\mathcal{J}_i, \mathcal{J}_i \backslash j) + \tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}} [K_j + z_{ik(j)} - \zeta_{k(j)}^{ii'}] + \\
\tfrac{\chi(i',i,j)}{n_{i'k}} \Delta\pi_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \backslash j) + \tfrac{\tilde{\chi}(i',i,j)}{n_{ik}} [K_j + z_{i'k(j)} - \zeta_{k(j)}^{ii'}] = \\
\tfrac{\chi(i,i',j)}{n_{ik}} [\Delta r_i(\mathcal{J}_i, \mathcal{J}_i \backslash j) + z_{ik(j)} - \zeta_{k(j)}^{ii'}] + \tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}} [K_j + z_{ik(j)} - \zeta_{k(j)}^{ii'}] + \\
\tfrac{\chi(i',i,j)}{n_{i'k}} [\Delta r_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \backslash j) + z_{i'k(j)} - \zeta_{k(j)}^{ii'}] + \tfrac{\tilde{\chi}(i',i,j)}{n_{ik}} [K_j + z_{i'k(j)} - \zeta_{k(j)}^{ii'}] \geq 0.
\end{gathered} \tag{16}
$$

K_j is a random variable which is known to be sufficiently positive such that the inequalities $K_j + z_{ik(j)} - \zeta_{k(j)}^{ii'} \geq 0$ and $K_j + z_{i'k(j)} - \zeta_{k(j)}^{ii'} \geq 0$ hold.[43] The above expression contains a weighted sum over the Nash conditions for all loans matched to firm i along with an expression known to be positive for all loan categories held by firm i' but not i. Similarly, it contains a weighted sum over the Nash conditions for all loans held by firm i' along with an expression known to be positive for all loan categories held by firm i but not i'. The terms $z_{ik(j)}$, $z_{i'k(j)}$, and $\zeta_{k(j)}^{ii'}$ enter the expression unconditionally and therefore their expectation over i, i', and $j \in \{\mathcal{J}_i \cup \mathcal{J}_{i'}\} \cap \mathcal{D}_i \cap \mathcal{D}_{i'}$ is equal to zero. Therefore, the unobservables drop out of the expectation of inequality 16:

$$
\begin{aligned}
\mathcal{E}\left[R_1(i, i', j) | x_j^{ii'}\right] \quad \equiv \quad & E\Big[\tfrac{\chi(i,i',j)}{n_{ik}} \Delta r_i(\mathcal{J}_i, \mathcal{J}_i \backslash j) + \tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}} K_j + \\
& \tfrac{\chi(i',i,j)}{n_{i'k}} \Delta r_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \backslash j) + \tfrac{\tilde{\chi}(i',i,j)}{n_{ik}} K_j | x_j^{ii'}\Big] \quad \geq 0.
\end{aligned} \tag{17}
$$

Note that the more conservative the choice of K_j, the larger the set of parameters satisfying the inequality. Thus, in the estimation, we want to be sufficiently conservative but not more so.

[43] Because the moment conditions take expectations over the inequalities, it is actually sufficient for K_j to merely satisify $\mathcal{E}\left[K_j + z_{ik(j)} - \zeta_{k(j)}^{ii'} \mid \tilde{\chi}(i, i', j) = 1\right] \geq 0$ and $\mathcal{E}\left[K_j + z_{i'k(j)} - \zeta_{k(j)}^{ii'} \mid \tilde{\chi}(i', i, j) = 1\right] \geq 0$. For the base specification, I set K_j to be 5 percent of the original balance of loan j.

By a symmetric argument, we can establish a second type of moment condition based on the optimality of the observed portfolios relative to adding a loan j:

$$\begin{aligned}\mathcal{E}\left[R_2(i,i',j)|x_j^{ii'}\right] \quad \equiv \quad & E\Big[\tfrac{\chi(i',i,j)}{n_{ik}}\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\cup j)+\tfrac{\tilde{\chi}(i',i,j)}{n_{i'k}}K_j+ \\ & \tfrac{\chi(i,i',j)}{n_{i'k}}\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)+\tfrac{\tilde{\chi}(i,i',j)}{n_{ik}}K_j|x_j^{ii'}\Big] \quad \geq 0.\end{aligned} \tag{18}$$

Finally, similar to the case for Assumption 1, we can construct a third type of moment inequality that is based on the total gains to trade conditional on firm i holding loan j:

$$\begin{aligned}&\tfrac{\chi(i,i',j)}{n_{ik}}[\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)]+\tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}}[L_j+z_{ik(j)}-z_{i'k(j)}]= \\ &\tfrac{\chi(i,i',j)}{n_{ik}}[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)+z_{ik(j)}-z_{i'k(j)}]+\tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}}[L_j+z_{ik(j)}-z_{i'k(j)}]\geq 0.\end{aligned} \tag{19}$$

In the above expression, the transfer payment nets out as in the case for Assumption 1. L_j is a random variable which is known to be sufficiently positive such that the inequality $L_j + z_{ik(j)} - z_{i'k(j)} \geq 0$ holds.[44] The term $z_{ik(j)} - z_{i'k(j)}$ enters the expression unconditionally, so when we take expectations, the contribution of the unobservables is zero. Therefore, taking expectations over i,i', and $J_i \cup \mathcal{J}_{i'}$:

$$\mathcal{E}\left[R_3(i,i',j)|x_j^{ii'}\right] \equiv \mathcal{E}\left[\tfrac{\chi(i,i',j)}{n_{ik}}[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)]+\tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}}L_j|x_j^{ii'}\right]\geq 0, \tag{20}$$

Instruments

The instruments $x_j^{ii'}$ must be mean-independent of the unobservables z_{ij} and $\zeta_j^{ii'}$ and in the information sets of the firms i and i' when deciding whether to transact on loan j. I include four types of instruments: (1) the underwriting characteristics of loan j;[45] (2) the mean underwriting characteristics for loans other than j in the portfolio of firm i and the means for loans other than j in the portfolio of firm i'; (3) dummies for the vintages (by year) of CMBS deals i and i' (4) the amount of time between the cutoff dates for deals i and i'. The instruments must be transformed

[44] Because the moment conditions take expectations over the inequalities, it is actually sufficient for L_j to merely satisify $\mathcal{E}\left[L_j + z_{ik(j)} - z_{i'k(j)} \;\mid\; \tilde{\chi}(i,i',j) = 1\right]$ For the base specification, I choose L_j to be 2.5 percent of the original balance of loan j.

[45] I use the components of w_{ij} from equation 1, excluding the only component that depends upon the identity of the deal, namely loan seasoning at the cutoff date.

so that they are all positive in order to ensure that none of the inequalities is reversed by its interaction with an instrument. Specifically, for each element of $x_j^{ii'}$, I include an indicator equal to 1 if the instrument exceeds its mean, and another indicator equal to 1 if the instrument is less than or equal to its mean. I denote the transformed vector of instruments by $h(x_j^{ii'})$.

Each observation consists of a pair of firms i and i' and a loan $j \in \{\mathcal{J}_i \cup \mathcal{J}_{i'}\} \cap \mathcal{D}_i \cap \mathcal{D}_{i'}$ (the set of loans in the firm's actual portfolios that could feasibly be matched with either firm). Let Θ denote the parameter space and θ an element therein. I construct the empirical moment functions as the vector

$$m(\theta) = \sum_i \sum_{i'} \sum_{j \in J_i \cup \mathcal{J}_{i'}} \begin{bmatrix} R_1(i,i',j)h(x) \\ R_2(i,i',j)h(x) \\ R_3 w(i,i',j)h(x) \end{bmatrix} \tag{21}$$

The identified set, $\Theta_0 = \{\theta \in \Theta : \mathcal{E}m(\theta) \geq 0\}$, is the set of parameters satisfying the moment inequalities, and may in general be a non-singleton. My estimate of Θ_0 minimizes the norm of negative component of the empirical moment function: $\hat{\Theta}_0 = \arg\min_{\theta \in \Theta} \parallel \min(0, m(\theta)) \parallel$.

Identification

The structural parameters are α_0, $\{\beta_t\}$, the vector γ, and $\{p_i\}$. What variation in the data serves to identify them? Consider a loan $j \in D_i$ (that is, a loan j that can feasibly be matched to i). The parameters α_0 and γ are identified by exogenous variation in demand for loan j by i's competitors, which comes from two sources. First, there is variation in the set of competing firms that could feasibly hold loan j, which depends upon which deals are close in time to the origination date of j. The more competing deals there are, the greater the total demand for j by i's competitors. Second, there is variation in the degree of correlation between loan j's exogenous characteristics and the exogenous characteristics of the loans in competing deals' pools. More negative correlation implies that loan j produces diversification benefits for the competitors, increasing their demand for j. Because these two sources of heterogeneity are uncorrelated with unobservable determinants of loan j's propensity to default, they provide valid identification.

For example, the constant term in the vector γ (the overall level of transfer payments) is identified by the degree of sensitivity in the propensity of loans to be securitized by i in response to an exogenous increase in demand from i's competitors. Suppose that firm i can feasibly be matched with either of two loans j and j' that differ in that there is more demand for j than for j' among i's competitors. The greater the transfer payments, the less likely j is to be matched to

i compared with j'. The reason is that higher transfer payments increase the opportunity cost of firm i holding loan j relative to the opportunity cost of holding loan j'. Similarly, the non-constant parameters in γ are identified the correlations between the covariates and the degree of sensitivity to exogenous changes in demand.

The causal impact of in-house securitization, α_0, is identified by exogenous variation in firm i's competitors' demand for loans originated by firm i. Suppose j is a loan originated by i. The more negative is α_0 (i.e., the more in-house status reduces default risk), the less sensitive is the propensity of j to be securitized in-house to an exogenous increase in demand from the competing underwriters. In particular, if α_0 is extremely negative, firm i will securitize j in-house regardless of the level of demand from its competitors.

The utility of cashflows from the B piece relative to the AAA bond, β_t, is identified by firms' revealed preference for portfolio volatility. As the volatility of a portfolio increases, the expected return on the B piece (per dollar of principal) generally increases while the return on the AAA bond generally decreases. Thus, a stronger preference at time t for loans whose returns are negatively (or positively) correlated with the rest of the portfolio implies greater volatility reduction and thus lower (higher) preferences for returns on the B piece versus returns on the AAA bond, which in turn implies a lower value of β_t. In estimation, I allow for a separate parameter β_t for each deal vintage.

Finally, p_i, the probability of losses to the AAA bond issued by firm i, is identified by the observed size of the bond, b_i^*, in conjunction with the simulated distribution of portfolio returns based on the first- and second-stage estimates and the in-house parameter α_0.

5 First- and second-stage Results

Table 3 presents the maximum-likelihood results for the first-stage hazard model. Estimates for all non-random coefficients are qualitatively similar to estimates from the reduced form model presented in Table 2. A minor difference is that here the dependent variable is the default time counting from the deal cutoff date rather than the date of loan origination, because these estimates will serve as inputs to the structural model, in which I model loan returns counting from the cutoff date. However, I include the loan seasoning at cutoff (cutoff date minus loan origination date) as a control.

For the random coefficients, I specify that $\xi^r = \xi_1^r + v_{ij}\xi_2^r$—where ξ_2^r is latent unless loan j in deal i is in-house—and estimate probabilities over the Cartesian product of the one-dimensional grids that form the support for ξ_1^r and ξ_1^r.[46] Table 3 shows the density of (ξ_1^r, ξ_2^r) at points in the support with positive estimated density. Although I use a fairly fine grid, only a few points have positive estimated density, which Bajari et al. (2010) have found to be typical for nonparametric mixtures estimators in general. Note that $-\infty$ is in the estimated support, implying that some proportion of loans have zero hazard of default. This outcome is clearly possible given a finite sample even if the true hazard of default is never exactly zero.

The estimated distribution of ξ_j conditional on loan j being in-house is "close to being" stochastically dominated by the distribution conditional on being non–in-house, except at the highest range of the support due to a small estimated density at $(\xi_1^r, \xi_2^r) = (5, 10)$. The expected impact of in-house status on the hazard of default (i.e., the expectation of $e^{\xi_2^r}$taken over the estimated density) is .949, which is similar to the reduced-form estimate from Table 2, and indicates that selection and causal effects have the cumulative effect of making in-house loans slightly less likely to default.

Bayesian estimates of the correlation structure of default times are reported in Appendix B. Comparing the variance of the idiosyncratic factor (ω_η) and the variances of the category-specific factors (the diagonal terms of Ω_ε) indicates that slighty more than half of the total variance is idiosyncratic. However, default times within each of the 30 loan categories are strongly correlated. By contrast, correlations across categories (the off-diagonal terms of Ω_ε) are weak. The weak cross-category correlation is unsurprising given that the model of individual loan default times already controls for loan vintage. Thus, very little time-series variation that is common across all loan categories remains to be explained by the copula model. In sum, the estimated correlation structure implies that holding a portfolio of loans within the same category achieves some degree of diversification (reducing idiosyncratic risk), but that much more diversification is achieved by holding loans from different categories.

6 Structural Estimation Results

Tables 4 and 5 display the structural estimates based on Assumptions 1 and 2, respectively. I obtain a point estimate under Assumption 1 (symmetric private information), which is not surprising given

[46] I set the support for both ξ_1^r and ξ_2^r to be $\{-\infty, -5, -4, -3, -2, -1, [-.75 : .05 : .75], 1, 2, 3, 4, 5\}$. The middle part of the grid is more finely spaced (ranging from -.75 to .75 in increments of .05): because the effect of ξ_j on the hazard is exponential, the densities for adjacent points of the support can better be separately identified for values of ξ_j that smaller in absolute magnitude.

the large number of inequality moments. Under Assumption 2 (asymmetric private information that is homogeneous for all loans by a given originator) the estimate is a set, reflecting the fact that the controlling for the boundary problem increases the size of the identified set. For each parameter, the table reports the value of that parameter at the extreme points of the identified set in that dimension. That is, denoting the kth parameter by θ_k, I report $\min_{\theta_k}\{\theta_k \in \hat{\Theta}_0\}$ and $\max_{\theta_k}\{\theta_k \in \hat{\Theta}_0\}$. [47]

I compute conservative 95-percent confidence intervals using the methodology described in Pakes et al. (2006).[48] There are more than 15 million potential trades (with a potential trade being between a pair of firms i and i' for a loan $j \in \{\mathcal{J}_i \cup \mathcal{J}_{i'}\} \cap \mathcal{D}_i \cap \mathcal{D}_{i'}$), which is many more than are needed to obtain very precise estimates. Therefore, in order to reduce the computational time I construct the objective function using a random 1-percent sample of the potential trades. Even random sampling leaves a large number of observations, resulting in very narrow confidence intervals.

For the most part, the two sets of estimates in Tables 4 and 5 are very similar. In addition to these two baseline estimates, I also estimated the model under Assumption 2 but without accounting for the boundary problem.[49] The resulting estimates (not shown) are, in fact, much closer to the estimates under Assumption 1 than to the baseline estimates under Assumption 2. We can thus conclude that the difference between the two baseline estimates is primarily driven by whether or not we allow for asymmetric private information across CMBS underwriters, rather than by whether we allow for heterogeneity in private information for different loans originated by the same lender.

The estimated causal effect of in-house status reduces the hazard of default. However, the

[47] I find these extreme points by first finding one point in the set $\hat{\Theta}_0$. Starting from this point, I then peform a line search in some dimension and direction until I hit the boundary of the set. Whenever I hit the boundary, I perform a line search in a different dimension and direction, and repeat this process for a very large number of iterations. Also, before switching dimensions at the boundary, I perturb the current parameter vector by a small random amount in order to avoid the (unlikely) situation in which the search gets caught in an infinite loop. While searching, I store the current elements of $\hat{\Theta}_0$ that deliver the greatest and least values of θ_k in each dimension k.

[48] I have not adjusted the confidence intervals to account for the fact that the parameters from the first and second stages of the model are estimated. Due to the recentness of the literature on moment-inequality estimation, it is not clear whether there is an analytical way to make such an adjustment. While it is in theory possible to take bootstrap draws of the data and reestimate from the beginning, doing so is computationally very burdensome because estimating the Gaussian copula takes several days. While not adjusting the confidence intervals is less than ideal, the high precision of the unadjusted estimates at least partially mitigates potential concerns about the precision of the structural estimates.

[49] That is, I set K_j and L_j to zero.

magnitude of this effect differs across specifications. Under Assumption 1, $\hat{\alpha}_0 = -3.34$, implying a hazard ratio of .035. Under the more plausible Assumption 2, $\hat{\alpha}_0 = -.765$, implying a hazard ratio of .465. What explains this difference? Assumption 1 imposes symmetry across all firms' beliefs about any loan j, ruling out positive correlation between the event of j being matched to firm i and i's private information about j. Therefore, by ruling out selection based on private information, the model relies to a greater extent on the ex post channel to explain the observed propensity for loans to be securitized in-house.

Furthermore, the estimate of α_0 under both specifications implies that the causal impact of in-house status *more* than accounts for the observed difference between the hazard for loans that are securitized in-house versus loans not securitized in-house. Suppose we subtract $\hat{\alpha}_0$ from the random effects for in-house loans (on whose distribution, $H(\xi_j | v_{ij} = 1)$, I report estimates in Table 2). The distribution of the remaining residuals implies a hazard of default that for in-house loans is higher on average by a factor of 26.9 under the assumption of symmetric information and by a factor of 2.0 under the more plausible assumption of asymmetric information. This factor reflects the impact of selection effects, which in the case of asymmetric information includes both adverse selection and idiosyncratic diversification benefits. Because adverse selection necessarily decreases the tendency of in-house loans to default (relative to other loans), the fact that selection actually implies a greater hazard of default for in-house loans suggests the importance of selection due to portfolio effects, which may go in either direction.

For most vintages and across both specifications, the implied utility of B-piece cashflows (the β_t parameters) is less than 1, indicating that firms value cashflows from the AAA tranche relatively higher. In other words, holding everything else equal, firms have a distaste for volatility. In fact, some of the estimated values show up as negative, reflecting the empirical fact that the size of the B piece is very small compared with the size of the AAA tranche. The thinness of the B piece causes the β_t parameters to be sensitive to nonlinearities in the valuation of cashflows that are not explicitly captured by the model.[50]

The β_t estimates differ in a few respect between the two specifications. Under Assumption 1, the β_t parameters generally move in tandem with the empirical hazard of default for loans originated in each vintage, reaching a low of .095 in 2005 and then skyrocketing to 1.26 in 2006 and then

[50] Recall from Figure 1 that as CMBS underwriters originate more loans, the observed share of those loans securitized in-house increases more slowly than the proportion that they represent out of total originations made over a particular window of time. This empirical regularity forces the model to produce decreasing returns on the number of in-house loans in the pool. The model accomplishes this through a strong aversion to volatility, because loans originated by the underwriter (as with all loans originated by a single lender) have positively correlated observed characteristics and thus thus increase the volatility of a pool as their share increases.

to 24.4 at the height of the crisis in 2007. An exceptionally high estimate for 2007 reflects a combination of the dramatic rise in the hazard of default for late-vintage loans (reflected in the first-stage estimates) together with fact that there were many feasible trades (as defined by my model) that would have reduced the volatility of the portfolios but were not carried out.

Identification of the β_t parameters is much weaker under Assumption 2 (asymmetric information), which is reflected their estimates being wide ranges rather than points. These ranges are generally less than 1, similar to the symmetric-information estimates, with which the ranges overlap for most vintages. Also similarly, the upper end of the range spikes in 2006 and 2007. However, the low end of the range remains low. This difference between specifications is driven by the fact that when we allow for asymmetric information, firms' failure to carry out feasible, volatility-reducing trades may reflect idiosyncratic preferences as opposed to their taste for volatility.

The transfer-payment parameters γ mostly have sensible interpretations.[51] Transfer payments are higher for loans with a higher coupon spread (i.e., higher gross returns) and lower for loans with high LTV, no reported LTV, low DSCR, or no reported occupancy ratio (which all predict a higher risk of default). Prices are slightly lower for more seasoned loans, which makes sense given that the transfer payments are scaled to be a proportion of the original loan balance (A_j), and that more seasoned loans may already be partially amortized. The occupancy ratio has a negative coefficient conditional on it being reported, and having no reported DSCR data has a positive coefficient, which is somewhat surprising but entirely possible given that these two variables may be correlated with omitted payoff-relevant variables.

Finally, the probability of the AAA tranche experiencing losses (p_i) increases dramatically for later vintages of CMBS deals, on average. The mean probability is 0.01 for the 2000 vintage, compared with 0.90 or 0.91 for the 2007 vintage, depending upon the specification.

7 Conclusion

There has been little work quantifying the magnitudes of different incentive distortions in securitization markets, largely because of the difficulty of separately identifying adverse selection from the effects of moral hazard on ex post asset performance. Furthermore, there are additional confounding selection effects because deal underwriters' decisions over asset bundles involve jointly

[51] I could have included a larger set of covariates, I chose to include just a few key variables because the observable determinants of transfer payments are not a key focus of the paper.

choosing loans in a portfolio decision. For these reasons, much of the theory literature on financial intermediation with asset bundling treats the bundles as exogenous, and existing empirical research on CMBS focuses on the behavior of individual loans.

I estimate a structural model of asset pool formation in the CMBS market, in which I endogenize the match between CMBS deals and individual loans, taking into account the correlation structure of loan returns. Exogenous variation in demand for loans from competing CMBS deal underwriters allows me to identify the causal impact of in-house securitization (defined as being originated and securitized by the same firm) on the ex post performance of loans. If we allow for asymmetric private information about loan quality, the implied impact of ex post hidden action by underwriters lowers in-house loans' hazard of default by a factor of .465. After netting out the impact of ex post hidden action from the overall empirical effect of in-house status on loan performance, the remaining effect—that attributable to the endogenous matching of loans to deals—is associated with an increase in the hazard of default by a factor of 2.0. The fact that endogenous selection actually *increases* the propensity of in-house loans to default suggests that adverse selection (which unambiguously reduces the propensity of in-house loans to default) is outweighed by selection due idiosyncratic diversification incentives (which can go in either direction).

The relative importance of ex post hidden action in explaining the better performance of in-house loans has implications for regulatory policy. Specifically, it suggests that measures that mitigate moral hazard would be more effective than measures designed to reduce adverse selection. For example, there is currently a policy debate on whether firms should be compelled to retain an interest in securitization deals backed by assets that they have originated. The importance of ex post incentives suggests that it may be best to require underwriters to hold a pro rata, "vertical," slice of the entire deal, so that the underwriter continues to have a direct interest in the deal no matter how much of the principal has already been lost. Likewise, regulators may want to give rating agencies stronger incentives to monitor CMBS deals over the course of their lives, not just prior to issuance.

Finally, this paper illustrates the usefulness of estimation techniques based on partially identified games for studying financial product markets in which assets are bundled or aggregated. Because of the high dimensionality of the bundling problem, it is in general infeasible to fully solve for the equilibrium. Through the use of moment inequalities, we can proceed without having to abstract from the endogeneity of bundle choice.

References

Ambrose, B. and A. Sanders (2003): "Commercial Mortgage (CMBS) Default and Prepayment Analysis," *Journal of Real Estate Finance and Economics*, Vol. 26, pp. 179-196.

An, X., D. Yongheng, and S. Gabriel (2009): "Is Conduit Lending to Blame? Asymmetric Information, Adverse Selection, and the Pricing of CMBS," *Mimeo*, University of Southern California.

Archer, W., P. Elmer, M. Harrison, and D. Ling (2002): "Determinants of Multifamily Mortgage Default," *Real Estate Economics*, Vol. 30, pp. 445-473.

Bajari, P., J. Fox, K. Kim, and S. Ryan (2010): "A Simple Nonparametric Estimator for the Distribution of Random Coefficients," *Mimeo.*

Cohen, A. (2005): "Asymmetric Information and Learning in the Automobile Insurance Market," *Review of Economics and Statistics*, Vol. 87, pp. 197-207.

Cox, D. R. (1972): "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, 34(2): 187-220.

DeMarzo, P. (2005): "The Pooling and Tranching of Securities: a Model of Informed Intermediation," *Review of Financial Studies*, Vol. 18, pp. 1-35.

DeMarzo, P. and D. Duffie (1999): "A Liquidity-Based Model of Securitiy Design," *Econometrica*, Vol. 67, pp. 65-99.

Deng, Y., J. Quigley, and A. Sanders (2004): "Commercial Mortgage Terminations: Evidence from CMBS," USC Working Paper.

Finkelstein, A. and J. Poterba (2006): "Testing for Adverse Selection with 'Unused Observables,' " NBER Working Paper 12112.

Furfine, Craig (2010): "Deal Complexity, Loan Performance, and the Pricing of Commercial Mortgage Backed Securities," *Mimeo*, Northwestern University.

Gan, Y. H. and C. Mayer (2006): "Agency Conflicts, Asset Substitution, and Securitization," NBER Working Paper 12359.

Glaeser, E. and H. Kallal (1997): "Thin Markets, Asymmetric Information, and Mortgage-Backed Securities," *Journal of Financial Intermediation*,Vol. 6, pp. 64-86.

Gollenberg, K. (1997): "An Investor's Perspective on Commercial Mortgage-Backed Coupon Strips," Chapter 8 in *Commercial Mortgage-Backed Securities* (F. Fabozzi and D. Jacob, eds.), Frank J. Fabozzi Associates and Nomura Securities International, Inc.

Heckman, J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models," *Econometrica* Vol. 52, pp. 271-320.

Ho, Katherine (2009): "Insurer-Provider Networks in the Medical Care Market," *American Economic Review* Vol. 99, pp. 393-430.

Ishii, Joy (2005): "Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry," *Mimeo*, Stanford University.

Kiff, J. and M. Kisser (2011): "A Shot at Regulating Securitization," *Mimeo*, International Monetary Fund and Norwegian School of Economics and Business Administration.

Sørensen, M. (2007): "How Smart is Smart Money? A Two-Sided Matching Model of Venture Capital," *Journal of Finance* Vol. 62, pp. 2725-2762.

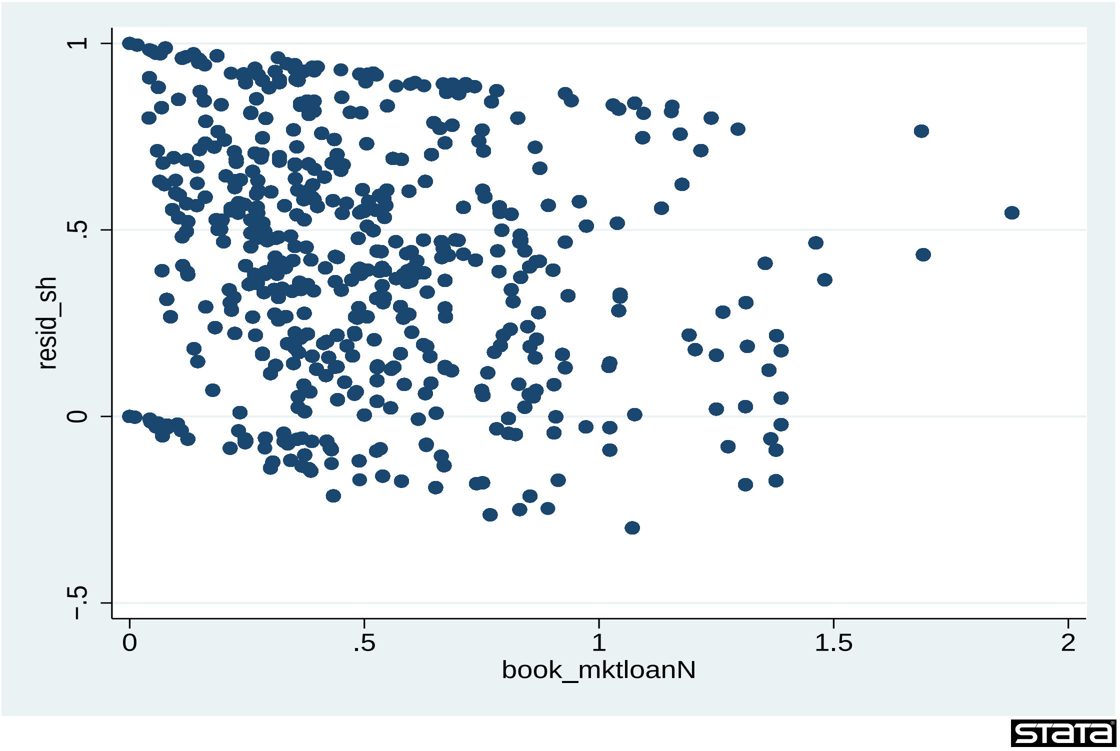
Smith, J. (2011): "The Effect of College Applications on Enrollment," *Mimeo*, Boston University.

Titman, S. and S. Tsyplakov (2010): "Originator Performance, CMBS Structures, and the Risk of Commercial Mortgages," *Review of Financial Studies*, Vol. 23, pp. 3558-3594.

Train, Kenneth (2001): "Discrete Choice Methods with Simulation," Cambridge, United Kingdom, Cambridge University Press.

Trivedi, P. and D. Zimmer (2005): "Copula Modeling: an Introduction for Practitioners," *Foundations and Trends in Econometrics*, Vol. 1, pp. 1-111.

Figure 1: Residual of Percentage of In-House Loans in Each Deal vs. Share of All Originations by Deal Underwriter Around Time of Deal



Each point corresponds to a deal. The vertical axis plots the residual of the share of in-house loans in the deal, relative to the share that would be expected if loans were allocated at random across all deals in the window of time 90 days before and after the origination date of each loan. The horizontal axis plots the number of loans originated by the underwriter of the deal (in thousands) in the window of time 90 days before and after the cutoff date of the deal.

Table 1: Summary Statistics

Deal-specific variables			Loan-specific variables		
	Mean	Std. Dev.		Mean	Std. Dev.
cutoff date	2004.92	(2.23)	original loan amount ($M)	10.0	(15.7)
cutoff # loans	143.3	(83.5)	in-house	.507	
cutoff pool balance ($M)	1685.5	(1185.6)	DSCR at issuance	1.31	(0.58)
% loans in-house	.538		No DSCR data	0.101	
constant prepayment rate	3.65	(14.93)	Occupancy at issuance	0.536	(0.473)
AAA tranche proportion	0.827	(0.073)	No occupancy data	0.435	
vintage = 2000	0.08		Original LTV	0.670	(0.135)
vintage = 2001	0.10		No LTV data	0.010	
vintage = 2002	0.09		loan seasoning at cutoff	4.54	(10.25)
vintage = 2003	0.11		coupon spread	1.60	(0.75)
vintage = 2004	0.15		IO loan	0.099	
vintage = 2005	0.16		Fixed-rate mortgage	0.988	
vintage = 2006	0.17		vintage = 2000	0.055	
vintage = 2007	0.15		vintage = 2001	0.073	
			vintage = 2002	0.065	
			vintage = 2003	0.087	
			vintage = 2004	0.119	
			vintage = 2005	0.174	
			vintage = 2006	0.197	
			vintage = 2007	0.174	
			in-house	0.507	
			Commercial-bank loan	0.432	
			Insurance co. loan	0.081	
			I-bank loan	0.184	
			Conduit loan	0.053	
			Finance co. loan	0.056	
			Foreign conduit loan	0.194	
N	468		N	60,688	

Table 2: Reduced-form Hazard Regressions

	(I)		(II)	
in-house	0.91	(0.09)	1.20	(0.15)
% in-house in deal	-	-	1.46	(0.13)
(in-house)*(% in-house in deal)	-	-	0.55	(0.07)
DSCR at issuance	0.79	(0.06)	0.78	(0.06)
No DSCR data	0.84	(0.09)	0.86	(0.10)
Occupancy at issuance	0.16	(0.04)	0.17	(0.04)
No occupancy data	0.20	(0.04)	0.20	(0.04)
Original LTV	18.32	(3.31)	17.37	(3.14)
No LTV data	8.38	(1.34)	8.52	(1.37)
coupon spread	1.59	(0.03)	1.60	(0.03)
original loan amount	2.37	(0.18)	2.35	(0.18)
IO loan	1.10	(0.06)	1.10	(0.06)
Fixed-rate mortgage	0.51	(0.04)	0.48	(0.04)
Vintage				
2000	0.32	(0.02)	0.32	(0.02)
2001	0.36	(0.02)	0.34	(0.02)
2002	0.31	(0.02)	0.30	(0.02)
2003	0.45	(0.03)	0.44	(0.03)
2004	0.85	(0.05)	0.83	(0.05)
2005	1.49	(0.09)	1.46	(0.08)
2006	2.24	(0.13)	2.18	(0.13)
2007	2.58	(0.18)	2.51	(0.18)
Originator fixed effects?	Yes		Yes	
Dummies for origination month?	Yes		Yes	
Region–property-type interactions?	Yes		Yes	
N	60,748		60,748	

All reported estimates are hazard ratios based on a standard Cox proportional hazard regression for which the dependent variable is the time to default for loan j counting from the origination date of the loan. Figures in parentheses are standard errors.

Table 3: First-Stage Hazard Regressions

	Underwriting characteristics (hazard ratio)	
Loan seasoning at cutoff	1.0662	(0.0017)
DSCR at issuance	0.7607	(0.0548)
No DSCR data	0.7952	(0.0928)
Occupancy at issuance	0.2237	(0.0598)
No occupancy data	0.2631	(0.0662)
Original LTV	12.039	(2.371)
No LTV data	4.0579	(0.7789)
coupon spread	1.6550	(0.0394)
Original loan amount	2.3325	(0.2081)
IO loan	1.0293	(0.0605)
Fixed-rate mortgage	0.5500	(0.0466)
I-bank loan	1.1410	(0.0486)
Insurance co. loan	0.5054	(0.0371)
Domestic conduit loan	1.4447	(0.0856)
Finance co. loan	1.0325	(0.0721)
Foreign conduit loan	1.3735	(0.0542)
Vintage fixed effects?	Yes	
Dummies for origination month?	Yes	
Region–property-type interactions?	Yes	
	Square root of baseline hazard	
Constant	1.000247	(0.000080)
t (months)	0.999295	(0.000134)
t^2	1.037595	(0.006879)
	θ: Density of (ξ_1^r, ξ_2^r), where $\xi_j = \xi_1^r + v_{ij}\xi_2^r$	
$(-\infty, 0)$	0.0089596	(0.0436127)
$(-5, -\infty)$	0.0469685	(0.0451106)
$(-0.5, -0.05)$	0.1413134	(0.6385605)
$(-0.5, 0)$	0.8027422	(0.6384472)
$(5, 5)$	0.0000163	(0.0000743)
N	60,688	

Table reports maximum-likelihood estimates of the hazard model with random effects, in which the dependent variable is the time to default counting from the deal cutoff date (in contrast to Table 2, in which the dependent variable is the time to default counting from the loan origination date). The random effects are specified as $\xi_j = \xi_1^r + v_{ij}\xi_2^r$, where v_{ij}=1 if loan j is securitized in-house by underwriter i, and $= 0$ otherwise. The support for (ξ_1^r, ξ_2^r) is specified over the grid defined by the Cartesian product of $\{-\infty, -5, -4, -3, -2, -1, [-.75 : .05 : .75], 1, 2, 3, 4, 5\}$. All parameters other than θ expressed as hazard ratios. Figures in parentheses are standard errors.

Table 4: Structural estimates: imposing symmetric information

	Estimate	Conservative 95% CI	
α_0 (in-house effect on hazard)	-3.34417	[-3.34417	-3.34417]
Transfer payment parameters			
Constant	1.36928	[1.36928	1.36928]
Loan seasoning at cutoff	-0.00462	[-0.00462	-0.00462]
DSCR at issuance	0.16209	[0.16209	0.16209]
No DSCR data	0.17528	[0.17528	0.17528]
Occupancy at issuance	-0.80745	[-0.80745	-0.80744]
No occupancy data	-0.57617	[-0.57617	-0.57616]
Original LTV	-0.24835	[-0.24835	-0.24835]
No LTV data	-0.13106	[-0.13106	-0.13106]
Coupon Spread	0.04385	[0.04385	0.04385]
Utility of "B-piece" cashflows			
β_{2000}	-0.19559	[-0.19559	-0.19559]
β_{2001}	0.18816	[0.18816	0.18816]
β_{2002}	0.06661	[0.06661	0.06661]
β_{2003}	-0.02625	[-0.02625	-0.02624]
β_{2004}	-0.10049	[-0.10050	-0.10049]
β_{2005}	0.69916	[0.69916	0.69917]
β_{2006}	1.25806	[1.25805	1.25806]
β_{2007}	24.38078	[24.38074	24.38080]
Mean of p_i (Prob. of AAA tranche experiencing credit losses), by vintage			
2000 vintage	0.01001	[0.01001	0.01001]
2001 vintage	0.00369	[0.00369	0.00369]
2002 vintage	0.00386	[0.00385	0.00386]
2003 vintage	0.00434	[0.00434	0.00434]
2004 vintage	0.03373	[0.03372	0.03373]
2005 vintage	0.27108	[0.27106	0.27109]
2006 vintage	0.70701	[0.70700	0.70703]
2007 vintage	0.92230	[0.92229	0.92231]

This table shows structural estimates based on Assumption 1, which imposes that firms have symmetric information about the quality of each loan.

Table 5: Structural estimates: originator-specific private beliefs

	Estimated set		Conservative 95% CI	
α_0 (in-house effect on hazard)	[-0.76515	-0.76515]	[-0.76515	-0.76515]
Transfer price parameters (γ				
Transfer pmnt Constant	[1.41351	1.43801]	[1.41350	1.43808]
Cutoff seasoning	[-0.00150	0.00051]	[-0.00150	0.00051]
DSCR at issuance	[0.16020	0.16063]	[0.16020	0.16064]
No DSCR data	[0.41328	0.49205]	[0.41328	0.49211]
Occupancy at issuance	[-0.73135	-0.70820]	[-0.73135	-0.70819]
No occupancy data	[-0.68167	-0.68071]	[-0.68167	-0.68062]
Original LTV	[-0.36974	-0.36920]	[-0.36974	-0.36902]
No LTV data	[-0.27813	-0.26391]	[-0.27813	-0.26382]
Coupon Spread	[0.02435	0.02444]	[0.02435	0.02444]
Discount factor for "B-piece" cashflows				
β_{2000}	[-0.18219	0.05334]	[-0.18219	0.05947]
β_{2001}	[-0.09829	0.35459]	[-0.09829	0.36074]
β_{2002}	[-0.21023	0.35425]	[-0.21023	0.36053]
β_{2003}	[-1.17997	-0.08341]	[-1.17998	-0.07723]
β_{2004}	[-1.07939	0.99750]	[-1.07940	1.01011]
β_{2005}	[0.38461	2.25475]	[0.38460	2.26164]
β_{2006}	[2.86174	4.75258]	[2.86174	4.75936]
β_{2007}	[-1.08626	44.08948]	[-1.08632	44.09618]
Mean of p_i (Prob. of AAA tranche experiencing credit losses), by vintage				
2000 vintage	[0.01023	0.01023]	[0.01022	0.01023]
2001 vintage	[0.00396	0.00396]	[0.00396	0.00397]
2002 vintage	[0.00415	0.00415]	[0.00415	0.00416]
2003 vintage	[0.01053	0.01053]	[0.01052	0.01053]
2004 vintage	[0.20949	0.20949]	[0.20948	0.20950]
2005 vintage	[0.55487	0.55487]	[0.55486	0.55488]
2006 vintage	[0.79986	0.79986]	[0.79985	0.79987]
2007 vintage	[0.93208	0.93208]	[0.93207	0.93209]

This table shows structural estimates based on Assumption 2, which allows for asymmetric information but imposes that firm i's private signals are the same across all loans j by a particular originator. K_j and L_j for the "boundary" cases in equations 17, 18, and 20 are set to 5 percent and 2.5 percent of the principal amount of loan j, respectively.

Appendix A: MCMC Estimation of Joint Default Times

This appendix describes the procedure for estimating the correlation structure of the multivariate normal vector $\varepsilon_i \equiv (\varepsilon_{i1} \dots \varepsilon_{iK})$ and normal random variable η_{ij} from Equation 9. For background information on specific steps, see Train (2002).

Data

The raw data are the default times for uncensored loans and censoring times for the remaining loans. Let t_j denote the (possibly censored) default time for loan j and let t_j^c denote the lesser of the default time and the censoring time (if there is any). Let $k(j)$ denote category for loan j, and define $y_{ij} \equiv Pr(T_j < t_j | w_{ij}, v_{ij}) = \varepsilon_{ik(j)} + \eta_{ij}$, and define $y_j^c \equiv Pr(T_j < t_j^c | w_{ij}, v_{ij})$. In principle, we can form the posteriors based on the entire set of loans. However, the posteriors are difficult to calculate (even up to a proportionality constant) due to the presence of constraints on the correlation structure. Instead, for each deal i, I simulate N_i random draws of vectors of loans using a procedure described in the next paragraph, and treat these vectors as independent observations. This drawing procedure yields a consistent estimate of the posterior mean, although the posterior variances in principle need to be adjusted.[52]

Let $n = 1 \dots N_i$ index the draws of vectors of loansq from pool i, with $N \equiv \sum_i N_i$ denoting the overall number of draws. The vectors are drawn independently, but the selection of loans for a particular draw is not independent. Specifically, draw n comprises up to $D = 2K$ loans selected at random from the pool of loans for a particular deal i, with the restriction that (1) loans are chosen without replacement for that draw (so that the draw cannot include the same loan multiple times), and (2) the draw contains no more than two loans from any single category k. Including more than one loan from each category is necessary for identifying the correlation structure across loans within a category. Limiting the number of loans from each category to no more than two is not strictly necessary, but for a given number of observations, allows for more variation identifying the cross-category correlations.

Note that in general, an observation vector may contain fewer than $2K$ elements, either because not all loan categories are represented in all pools, or because (at least in principal) certain categories may have only one loan in a pool. However, it is convenient to regard the set of parameters to be estimated as ω_η, Ω_ε, and a full set of $3KN$ latent variables ε_{nk} and η_{nj} for $k = 1 \dots K$, $j = 1 \dots 2K$, and $n = 1 \dots N$. The elements of $\{\varepsilon_{nk}\}_{n,k}$ and $\{\eta_{nj}\}_{n,j}$ corresponding to "missing" data (which can equivalently be thought of as being censored at $t = 0$) enter the posterior density conditional on ω_η and Ω_ε, but do not interact with data in the likelihood function.

[52] Adjusting the standard errors is necessary due to the randomness in sampling. In practice, I set N_i to be proportional to the number of loans in deal i (such that all observations are equally weighted) and such that the overall number of vector draws, N, is approximately equal to the total number of loans in the data divided by D.

Prior distributions

The estimation procedure simulates the posterior distribution of Ω_ε, ω_η, and the latent variables, $\{\varepsilon_{nk}\}_{n,k}$ and $\{\eta_{nj}\}_{n,j}$. The MCMC chain uses Gibbs sampling with a nested Metropolis-Hastings step. The Gibbs sampler alternates between drawing from the posterior distribution of Ω_ε and ω_η conditional on $\{\varepsilon_{nk}\}_{n,k}$ and $\{\eta_{nj}\}_{n,k}$ and drawing from the posterior distribution of ε_{nk} and η_{nj} conditional on Ω_ε, ω_η, and data.

The prior distribution for ω_η is inverted gamma with degrees of freedom $v_0 = 1$ and scale $s_0 = 1$, denoted $IG(v_0, s_0)$. The prior distribution for Ω_ε is K-dimensional inverted Wishart with degrees of freedom $v_0 = K$ and scale $S_0 = I$ (the identity K-by-K matrix), denoted $IW(v_0, S_0)$. The prior distributions for $\varepsilon_n \equiv \{\varepsilon_{nk}\}_k$ and η_{nj} are $N(\mathbf{0}, I)$ and $N(0,1)$, respectively, for all draws n and all loans j. This choice of priors is motivated by the fact that they are self-conjugate.

MCMC iterations

Initialize Ω_ε^0 to the identity matrix and initialize ω_j^0 to 1. Draw $\{\varepsilon_{nk}^0\}$ from the multivariate normal density $\phi(\varepsilon|\Omega_\varepsilon^0)$. For censored observations, draw $\{\eta_{nj}^0\}$ from the univariate normal density $\phi(\varepsilon|\omega_j^0)$. For uncensored observations, set $\eta_{nj}^0 = y_{nj} - \varepsilon_{nk(j)}^0$.

Each subsequent MCMC iteration $t = 1 \ldots S$ involves the following steps. For efficiency, the procedure should be vectorized for the entire set of draws $n = 1 \ldots N_i$.

1. Draw ω_η^t and Ω_ε^t from their posterior distributions conditioning on $\{\varepsilon_{nk}^{t-1}\}$ and $\{\eta_{nj}^{t-1}\}$:

$$\begin{gathered} K(\omega_\eta|\{\eta_{nj}^{t-1}\}_{n,j}) \text{ is } IG(v_0 + ND, \tfrac{v_0 s_0 + ND\bar{s}_\eta}{v_0 + ND}), \\ K(\Omega_\varepsilon|\{\varepsilon_n^{t-1}\}_n) \text{ is } IW(v_0 + N, \tfrac{v_0 S_0 + N\bar{S}_\varepsilon}{v_0 + D}), \text{ where} \\ \bar{s}_\eta = (1/ND)\textstyle\sum_{n,j}(\eta_{nj}^{t-1})^2 \quad , \quad \bar{S}_\varepsilon = (1/N)\sum_n \varepsilon_n^{t-1}\varepsilon_n^{t-1'}. \end{gathered} \tag{22}$$

2. Create proposal Markov draws for $\{\varepsilon_n^t\}_n$ and $\{\eta_{nj}\}_{n,j}$ as $\{\tilde{\varepsilon}_n^t\}_n$ and $\{\tilde{\eta}_{nj}^t\}_{n,j}$. For each n, begin by drawing the K-by-1 standard normal random vector ν_1 from $\phi(\varepsilon|\Omega_\varepsilon^t)$ and drawing the standard normal random scalar ν_{2j} from $\phi(\varepsilon|\omega_\eta^t)$. Then set $\tilde{\varepsilon}_n^t = \varepsilon_n^{t-1} + \rho\nu_1$, where the step-size ρ is an estimation parameter. If loan jn is uncensored and appears in the data, set $\tilde{\eta}_{nj}^t = y_{nj} - \tilde{\varepsilon}_{nk(j)}^t$. If loan jn is censored or is "missing" from the data, set $\tilde{\eta}_{nj}^t = \eta_{nj}^{t-1} + \rho\nu_{2j}$.

3. For each n, draw a random variable υ_n from a uniform distribution.

4. Calculate the ratio of the quasi-posterior density evaluated at the proposal draw to the quasi-posterior density evaluated at the previous draw. The first component of the quasi-posterior density evaluated at $\{\tilde{\varepsilon}_n^t\}$ and $\{\tilde{\eta}_{nj}^t\}$ is the density $\phi(\{\tilde{\varepsilon}_n^t\}_n|\Omega_\varepsilon^t)\phi(\{\tilde{\eta}_{nj}^t\}_{mn,j}|\omega_j^t)$.

   The second component is $L(y_n|\{\tilde{\eta}_{nj}^t\}_{n,j}, \{\tilde{\varepsilon}_n^t\}_n)$, the likelihood of the data conditional on $\{\varepsilon_n^t\}_n$ and $\{\eta_{nj}^t\}_{n,j}$. Notice that for all uncensored loans, $\tilde{\varepsilon}_{nk(j)}^t + \tilde{\eta}_{nj}^t = y_{nj}$ by construction.

Therefore, the likelihood $L(y_n|\{\tilde{\eta}^t_{nj}\}_{n,j}, \{\tilde{\varepsilon}^t_n\}) = 1$ if $\tilde{\varepsilon}^t_{nk(j)} + \tilde{\eta}^t_{nj} > t^c_j$ for all censored loans, and $= 0$ otherwise.

5. Accept the proposal draw if

$$\frac{L(y_n|\{\tilde{\eta}^t_{nj}\}_{n,j}, \{\tilde{\varepsilon}^t_n\}_n)\phi(\{\tilde{\varepsilon}^t_n\}_n|\Omega^t_\varepsilon)\phi(\{\tilde{\eta}^t_{nj}\}_{n,j}|\omega^t_j)}{L(y_n|\{\eta^{t-1}_{nj}\}_{n,j}, \{\varepsilon^{t-1}_n\}_n)\phi(\{\varepsilon^{t-1}_n\}_n|\Omega^t_\varepsilon)\phi(\{\eta^{t-1}_{nj}\}_{n,j}|\omega^t_j)} > \upsilon_n \tag{23}$$

If the proposal draws is accepted, then for all n and all j, set $\varepsilon^t_n = \tilde{\varepsilon}^t_n$ and $\eta^t_{nj} = \tilde{\eta}^t_{nj}$. Otherwise, set $\varepsilon^t_n = \varepsilon^{t-1}_n$ and $\eta^t_{nj} = \eta^{t-1}_{nj}$.

6. Return to the first step.

Appendix B: Joint Default Time Estimates

The tables on the following pages display the estimates of the correlation structure for the joint distribution of default times, as described in equation 9 in the text. The estimates were obtained using the MCMC procedure described in Appendix A.

Tables B.1 and B.2 show the means of the posterior distributions for the parameters Ω_ε (the covariance matrix for the category-specific factors) and ω_η (the variance of the idiosyncratic factor). Tables B.3 and B.4 show the standard deviations of the posterior distributions for the same parameters.

The 30 category-specific factors are defined for interactions between 10 regions and 3 property types. The table abbreviates the region names and property type name, with the first three letters of each name corresponding to the region and the second three letters corresponding to the property type.

Regions are abbreviated as follows:

"MAT"= Mid-Atlantic Region; "MWE"= Midwest, Eastern Region; "MWW"= Midwest, Western Region; "NGL"= New England; "SAT"= Southern, Atlantic; "SEC"= South-Central, East; "SWC"= South-Central, West; "MTN"= Western, Mountain; "PNW"= Western, Northern Pacific (including California); "OTH"= Other (including Hawaii, places outside the United States, and loans for properties in more than one region).

Property types are abbreviated as follows: "COM" (for "commercial")=office/retail/hotel; "FAM"=multifamily (apartments); "OTH"=industrial/other.

Table B.1: Posterior means of covariance parameters

Category-specific factors (Ω_ε)	OTH*OTH	OTH*FAM	OTH*COM	MAT*OTH	MAT*FAM	MAT*COM	MWE*OTH	MWE*FAM	MWE*COM	MWW*OTH	MWW*FAM	MWW*COM	NGL*OTH	NGL*FAM	NGL*COM
OTH*OTH	0.5030	-0.0009	-0.0014	0.0014	-0.0017	0.0018	-0.0015	-0.0004	-0.0032	0.0047	0.0004	0.0005	0.0016	0.0005	-0.0008
OTH*FAM	-0.0009	0.4986	0.0029	0.0005	-0.0019	-0.0004	-0.0017	0.0018	-0.0006	0.0013	0.0028	-0.0001	0.0012	-0.0008	0.0007
OTH*COM	-0.0014	0.0029	0.5029	0.0015	-0.0015	-0.0003	0.0001	-0.0008	0.0002	-0.0012	-0.0016	-0.0011	0.0034	-0.0005	0.0028
MAT*OTH	0.0014	0.0005	0.0015	0.4999	-0.0017	-0.0006	-0.0001	-0.0018	0.0001	-0.0011	0.0003	0.0004	-0.0001	0.0008	-0.0006
MAT*FAM	-0.0017	-0.0019	-0.0015	-0.0017	0.5038	-0.0002	0.0027	-0.0002	0.0018	0.0001	-0.0006	-0.0017	0.0034	-0.0003	0.0011
MAT*COM	0.0018	-0.0004	-0.0003	-0.0006	-0.0002	0.5029	-0.0009	0.0000	0.0004	0.0005	0.0020	-0.0012	0.0002	-0.0002	-0.0004
MWE*OTH	-0.0015	-0.0017	0.0001	-0.0001	0.0027	-0.0009	0.5054	0.0007	-0.0004	0.0023	-0.0010	0.0004	-0.0005	0.0013	-0.0002
MWE*FAM	-0.0004	0.0018	-0.0008	-0.0018	-0.0002	0.0000	0.0007	0.4999	-0.0007	0.0010	-0.0002	-0.0024	0.0005	-0.0021	-0.0025
MWE*COM	-0.0032	-0.0006	0.0002	0.0001	0.0018	0.0004	-0.0004	-0.0007	0.5088	0.0016	0.0001	-0.0015	0.0003	0.0005	-0.0008
MWW*OTH	0.0047	0.0013	-0.0012	-0.0011	0.0001	0.0005	0.0023	0.0010	0.0016	0.5005	-0.0010	0.0015	-0.0001	0.0001	-0.0012
MWW*FAM	0.0004	0.0028	-0.0016	0.0003	-0.0006	0.0020	-0.0010	-0.0002	0.0001	-0.0010	0.5032	-0.0017	-0.0020	-0.0031	-0.0013
MWW*COM	0.0005	-0.0001	-0.0011	0.0004	-0.0017	-0.0012	0.0004	-0.0024	-0.0015	0.0015	-0.0017	0.5049	-0.0010	0.0020	-0.0006
NGL*OTH	0.0016	0.0012	0.0034	-0.0001	0.0034	0.0002	-0.0005	0.0005	0.0003	-0.0001	-0.0020	-0.0010	0.5003	-0.0006	0.0028
NGL*FAM	0.0005	-0.0008	-0.0005	0.0008	-0.0003	-0.0002	0.0013	-0.0021	0.0005	0.0001	-0.0031	0.0020	-0.0006	0.5088	-0.0005
NGL*COM	-0.0008	0.0007	0.0028	-0.0006	0.0011	-0.0004	-0.0002	-0.0025	-0.0008	-0.0012	-0.0013	-0.0006	0.0028	-0.0005	0.4992
SAT*OTH	0.0004	0.0014	0.0025	0.0007	-0.0026	-0.0026	-0.0005	0.0001	0.0020	0.0005	0.0008	-0.0013	-0.0007	0.0001	0.0011
SAT*FAM	0.0018	-0.0010	-0.0014	0.0001	0.0002	0.0010	-0.0026	0.0011	-0.0017	0.0014	-0.0017	0.0000	0.0012	-0.0022	-0.0033
SAT*COM	-0.0010	0.0016	-0.0012	-0.0006	-0.0001	-0.0004	0.0016	-0.0003	-0.0019	-0.0010	-0.0005	0.0002	0.0009	0.0012	-0.0013
SEC*OTH	-0.0004	-0.0017	0.0013	0.0012	-0.0013	0.0000	0.0009	-0.0005	-0.0022	-0.0015	-0.0017	-0.0001	0.0003	0.0045	-0.0004
SEC*FAM	-0.0008	-0.0041	-0.0030	-0.0017	-0.0020	0.0006	0.0001	-0.0024	0.0014	0.0012	-0.0016	0.0005	0.0000	0.0024	-0.0007
SEC*COM	-0.0010	-0.0023	0.0004	0.0021	0.0026	-0.0022	-0.0014	-0.0026	-0.0008	0.0007	-0.0014	0.0006	0.0013	0.0014	-0.0005
SWC*OTH	-0.0002	-0.0001	-0.0003	-0.0008	-0.0003	0.0002	-0.0014	-0.0009	-0.0030	0.0003	-0.0016	-0.0003	-0.0005	-0.0005	-0.0008
SWC*FAM	-0.0003	0.0000	0.0015	-0.0033	0.0009	-0.0008	-0.0018	0.0007	-0.0019	0.0013	-0.0025	0.0001	0.0017	0.0005	-0.0010
SWC*COM	-0.0002	0.0005	0.0003	-0.0016	0.0001	0.0004	-0.0008	0.0005	-0.0016	-0.0030	0.0005	-0.0030	-0.0010	-0.0014	-0.0011
MTN*OTH	0.0004	-0.0014	0.0009	-0.0006	0.0005	-0.0013	0.0026	-0.0001	0.0000	0.0006	-0.0001	-0.0024	0.0005	-0.0013	-0.0013
MTN*FAM	0.0017	0.0002	0.0005	-0.0002	0.0009	-0.0011	-0.0017	0.0003	-0.0008	-0.0022	-0.0039	-0.0015	-0.0027	-0.0001	-0.0009
MTN*COM	-0.0002	0.0017	0.0017	-0.0026	-0.0031	-0.0006	-0.0009	-0.0029	-0.0033	-0.0003	-0.0003	-0.0010	0.0004	-0.0010	-0.0007
PNW*OTH	0.0015	0.0002	-0.0011	0.0027	0.0014	0.0002	-0.0015	-0.0013	-0.0025	-0.0019	-0.0011	-0.0021	-0.0006	-0.0014	0.0010
PNW*FAM	0.0003	0.0008	0.0004	0.0013	0.0022	-0.0011	-0.0016	-0.0025	-0.0018	0.0007	0.0001	0.0009	0.0008	-0.0007	-0.0019
PNW*FAM	0.0006	-0.0007	0.0016	0.0006	-0.0010	0.0010	-0.0002	-0.0011	-0.0021	0.0005	-0.0011	-0.0002	-0.0008	0.0042	-0.0013

Table B.2: Posterior means of covariance parameters (continued)

Category-specific factors (Ω_ε)	SAT*OTH	SAT*FAM	SAT*COM	SEC*OTH	SEC*FAM	SEC*COM	SWC*OTH	SWC*FAM	SWC*COM	MTN*OTH	MTN*FAM	MTN*COM	PNW*OTH	PNW*FAM	PNW*COM
OTH*OTH	0.0004	0.0018	-0.0010	-0.0004	-0.0008	-0.0010	-0.0002	-0.0003	-0.0002	0.0004	0.0017	-0.0002	0.0015	0.0003	0.0006
OTH*FAM	0.0014	-0.0010	0.0016	-0.0017	-0.0041	-0.0023	-0.0001	0.0000	0.0005	-0.0014	0.0002	0.0017	0.0002	0.0008	-0.0007
OTH*COM	0.0025	-0.0014	-0.0012	0.0013	-0.0030	0.0004	-0.0003	0.0015	0.0003	0.0009	0.0005	0.0017	-0.0011	0.0004	0.0016
MAT*OTH	0.0007	0.0001	-0.0006	0.0012	-0.0017	0.0021	-0.0008	-0.0033	-0.0016	-0.0006	-0.0002	-0.0026	0.0027	0.0013	0.0006
MAT*FAM	-0.0026	0.0002	-0.0001	-0.0013	-0.0020	0.0026	-0.0003	0.0009	0.0001	0.0005	0.0009	-0.0031	0.0014	0.0022	-0.0010
MAT*COM	-0.0026	0.0010	-0.0004	0.0000	0.0006	-0.0022	0.0002	-0.0008	0.0004	-0.0013	-0.0011	-0.0006	0.0002	-0.0011	0.0010
MWE*OTH	-0.0005	-0.0026	0.0016	0.0009	0.0001	-0.0014	-0.0014	-0.0018	-0.0008	0.0026	-0.0017	-0.0009	-0.0015	-0.0016	-0.0002
MWE*FAM	0.0001	0.0011	-0.0003	-0.0005	-0.0024	-0.0026	-0.0009	0.0007	0.0005	-0.0001	0.0003	-0.0029	-0.0013	-0.0025	-0.0011
MWE*COM	0.0020	-0.0017	-0.0019	-0.0022	0.0014	-0.0008	-0.0030	-0.0019	-0.0016	0.0000	-0.0008	-0.0033	-0.0025	-0.0018	-0.0021
MWW*OTH	0.0005	0.0014	-0.0010	-0.0015	0.0012	0.0007	0.0003	0.0013	-0.0030	0.0006	-0.0022	-0.0003	-0.0019	0.0007	0.0005
MWW*FAM	0.0008	-0.0017	-0.0005	-0.0017	-0.0016	-0.0014	-0.0016	-0.0025	0.0005	-0.0001	-0.0039	-0.0003	-0.0011	0.0001	-0.0011
MWW*COM	-0.0013	0.0000	0.0002	-0.0001	0.0005	0.0006	-0.0003	0.0001	-0.0030	-0.0024	-0.0015	-0.0010	-0.0021	0.0009	-0.0002
NGL*OTH	-0.0007	0.0012	0.0009	0.0003	0.0000	0.0013	-0.0005	0.0017	-0.0010	0.0005	-0.0027	0.0004	-0.0006	0.0008	-0.0008
NGL*FAM	0.0001	-0.0022	0.0012	0.0045	0.0024	0.0014	-0.0005	0.0005	-0.0014	-0.0013	-0.0001	-0.0010	-0.0014	-0.0007	0.0042
NGL*COM	0.0011	-0.0033	-0.0013	-0.0004	-0.0007	-0.0005	-0.0008	-0.0010	-0.0011	-0.0013	-0.0009	-0.0007	0.0010	-0.0019	-0.0013
SAT*OTH	0.5056	-0.0026	0.0020	0.0001	-0.0020	-0.0005	-0.0004	-0.0026	0.0005	0.0025	-0.0001	-0.0006	-0.0008	-0.0011	0.0007
SAT*FAM	-0.0026	0.5016	-0.0010	-0.0012	0.0003	-0.0013	0.0002	-0.0017	-0.0046	0.0000	0.0005	-0.0029	0.0007	0.0000	-0.0013
SAT*COM	0.0020	-0.0010	0.5020	0.0005	-0.0017	-0.0009	-0.0031	-0.0026	-0.0062	0.0004	-0.0043	-0.0015	-0.0016	-0.0003	0.0002
SEC*OTH	0.0001	-0.0012	0.0005	0.5090	0.0015	0.0016	-0.0001	-0.0028	-0.0004	-0.0006	-0.0012	-0.0024	-0.0025	0.0005	0.0007
SEC*FAM	-0.0020	0.0003	-0.0017	0.0015	0.5037	0.0001	-0.0014	-0.0007	-0.0022	-0.0008	0.0015	-0.0024	-0.0013	-0.0017	-0.0020
SEC*COM	-0.0005	-0.0013	-0.0009	0.0016	0.0001	0.5073	0.0001	-0.0014	-0.0003	0.0010	-0.0009	-0.0021	0.0032	0.0010	-0.0011
SWC*OTH	-0.0004	0.0002	-0.0031	-0.0001	-0.0014	0.0001	0.5020	-0.0007	0.0005	-0.0002	0.0007	0.0012	0.0014	-0.0001	-0.0026
SWC*FAM	-0.0026	-0.0017	-0.0026	-0.0028	-0.0007	-0.0014	-0.0007	0.5021	-0.0011	-0.0027	-0.0023	-0.0027	-0.0021	-0.0009	-0.0013
SWC*COM	0.0005	-0.0046	-0.0062	-0.0004	-0.0022	-0.0003	0.0005	-0.0011	0.5014	0.0007	-0.0007	-0.0015	0.0000	-0.0019	-0.0016
MTN*OTH	0.0025	0.0000	0.0004	-0.0006	-0.0008	0.0010	-0.0002	-0.0027	0.0007	0.5042	-0.0004	-0.0010	-0.0038	-0.0014	-0.0033
MTN*FAM	-0.0001	0.0005	-0.0043	-0.0012	0.0015	-0.0009	0.0007	-0.0023	-0.0007	-0.0004	0.5038	-0.0006	-0.0009	0.0008	0.0007
MTN*COM	-0.0006	-0.0029	-0.0015	-0.0024	-0.0024	-0.0021	0.0012	-0.0027	-0.0015	-0.0010	-0.0006	0.5033	0.0025	-0.0010	-0.0002
PNW*OTH	-0.0008	0.0007	-0.0016	-0.0025	-0.0013	0.0032	0.0014	-0.0021	0.0000	-0.0038	-0.0009	0.0025	0.5001	-0.0009	-0.0019
PNW*FAM	-0.0011	0.0000	-0.0003	0.0005	-0.0017	0.0010	-0.0001	-0.0009	-0.0019	-0.0014	0.0008	-0.0010	-0.0009	0.5007	-0.0011
PNW*COM	0.0007	-0.0013	0.0002	0.0007	-0.0020	-0.0011	-0.0026	-0.0013	-0.0016	-0.0033	0.0007	-0.0002	-0.0019	-0.0011	0.5017
Idiosyncratic factor (ω_η)	0.6068														

Table B.3: Posterior standard deviations of covariance parameters

Category-specific factors (Ω_ε)	OTH*OTH	OTH*FAM	OTH*COM	MAT*OTH	MAT*FAM	MAT*COM	MWE*OTH	MWE*FAM	MWE*COM	MWW*OTH	MWW*FAM	MWW*COM	NGL*OTH	NGL*FAM	NGL*COM
OTH*OTH	0.0019	0.0014	0.0014	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
OTH*FAM	0.0014	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013
OTH*COM	0.0014	0.0013	0.0019	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0014	0.0013	0.0014	0.0013
MAT*OTH	0.0013	0.0013	0.0014	0.0019	0.0013	0.0013	0.0014	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
MAT*FAM	0.0014	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
MAT*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MWE*OTH	0.0014	0.0013	0.0013	0.0014	0.0013	0.0014	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MWE*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	0.0014	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
MWE*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MWW*OTH	0.0013	0.0013	0.0013	0.0014	0.0014	0.0013	0.0013	0.0014	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013
MWW*FAM	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	0.0014	0.0013	0.0014	0.0013
MWW*COM	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0019	0.0014	0.0014	0.0013
NGL*OTH	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0019	0.0014	0.0013
NGL*FAM	0.0013	0.0014	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0014	0.0014	0.0019	0.0013
NGL*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0018
SAT*OTH	0.0014	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0014	0.0014	0.0013	0.0013	0.0013	0.0013
SAT*FAM	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013
SAT*COM	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013
SEC*OTH	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014
SEC*FAM	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SEC*COM	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0014	0.0014
SWC*OTH	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SWC*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0014	0.0013	0.0013	0.0013	0.0014	0.0013
SWC*COM	0.0014	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MTN*OTH	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0014	0.0013	0.0013	0.0014	0.0013
MTN*FAM	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013
MTN*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
PNW*OTH	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
PNW*FAM	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
PNW*COM	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013

Table B.4: Posterior standard deviations of covariance parameters (continued)

Category-specific factors (Ω_ε)	SAT*OTH	SAT*FAM	SAT*COM	SEC*OTH	SEC*FAM	SEC*COM	SWC*OTH	SWC*FAM	SWC*COM	MTN*OTH	MTN*FAM	MTN*COM	PNW*OTH	PNW*FAM	PNW*COM
OTH*OTH	0.0014	0.0013	0.0013	0.0013	0.0014	0.0014	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
OTH*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
OTH*COM	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0014
MAT*OTH	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013
MAT*FAM	0.0013	0.0014	0.0014	0.0014	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013
MAT*COM	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MWE*OTH	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013
MWE*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MWE*COM	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0014	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MWW*OTH	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
MWW*FAM	0.0014	0.0013	0.0014	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
MWW*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
NGL*OTH	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
NGL*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0014	0.0014	0.0013	0.0013	0.0013	0.0013
NGL*COM	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SAT*OTH	0.0019	0.0014	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013
SAT*FAM	0.0014	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SAT*COM	0.0014	0.0013	0.0019	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SEC*OTH	0.0013	0.0013	0.0013	0.0019	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013
SEC*FAM	0.0014	0.0013	0.0013	0.0013	0.0018	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013
SEC*COM	0.0013	0.0013	0.0014	0.0014	0.0013	0.0019	0.0013	0.0013	0.0013	0.0014	0.0014	0.0013	0.0013	0.0013	0.0013
SWC*OTH	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0018	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SWC*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0018	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
SWC*COM	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0018	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013
MTN*OTH	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013
MTN*FAM	0.0014	0.0013	0.0013	0.0014	0.0014	0.0014	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013
MTN*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0018	0.0013	0.0013	0.0013
PNW*OTH	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0014
PNW*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0018	0.0013
PNW*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0019
Idiosyncratic factor (ω_η)	0.0003														

Appendix C: Simulation of Portfolio Returns

To compute the estimation objective function for the structural model, I simulate the distribution of portfolio returns for deal i given actual and alternative portfolio contents using 200 simulation draws. Suppose there are N_i loans in the data that can feasibly be matched with deal i. Each simulation draw s is performed as follows:

- Draw N_i+30 uniformly distributed random variables using Halton sequences. The first N_i random variables are used to generate idiosyncratic factors for each loan j, and the latter 30 to generate the 30 category-specific factors in the factor model described by equation 9.
- Using the estimated parameters of the factor model (see Appendix B), transform the uniform random variables into draws of the idiosyncratic factors $\eta_{ij}^{(s)}$ and the category-specific factors $\varepsilon_i^{(s)}$.
- For each loan j, compute $p^{(s)} = \Phi(\varepsilon_{ik(j)}^{(s)} + \eta_{ij}^{(s)})$.
- For each loan j, obtain a simulation draw for $t_j^{(s)}$ by inverting $p^{(s)} = Pr(T_j < t_j^{(s)})$, where $Pr(T_j < t_j^{(s)})$ is defined by the estimated parameters for equation 2 and the current value of α_0.
- For each loan j, simulate the return $y_j^{(s)}$ by setting t_j to $t_j^{(s)}$ in equation 3.

To compute the simulated portfolio returns under various alternatives, simply aggregate the individual loan returns. This procedure produces a simulated distribution that is a step function. To make the simulated distribution function invertible (necessary for computing the AAA bond sizes under the alternative portfolios) and to ensure a smooth objective function, I use interpolation to convert the step function into a locally linear function with nodes passing through each of the jumps in the step function.

Note 1: following standard practice, the same Halton draws are reused for each evaluation of the objective function.

Note 2: There is no closed-form solution for the inverse of the function $p = Pr(T_j < t_j)$, so the inversion must be performed using a nonlinear solver. To avoid having to repeatedly resolve a large number of nonlinear equations for eachevaluation of the objective function, I instead estimate a sieve approximation for t_j, with powers of p and powers of $(v_{ij}\alpha_0 + w'_{ij}\alpha_1)$ serving as the sieve basis functions. I only estimate the sieve approximation once, and then use the fitted values of $t_j^{(s)}$ for each subsequent evaluation of the objective function.