Quantifying Adverse Selection in the Commercial Mortgage-Backed Security Market*

Chenghuan Sean Chu

Federal Reserve Board of Governors

March 2012

Abstract

The commercial mortgage-backed security (CMBS) market is, in theory, subject to adverse selection at the margin between loans securitized "in-house"—that is, pooled into CMBS deals issued by the loan originators themselves—versus loans sold into CMBS deals issued by competing firms. However, the effects of adverse selection on loan quality are confounded by nonrandom selection with respect to common information observed by market participants but not the econometrician. First, an issuer has an incentive to securitize its own loans even if negative market information lowers demand for them by competing issuers, because the outside option for the issuer is to keep ("warehouse") these loans on its own balance sheet. Moreover, issuers may be willing to securitize loans that are lower in quality, viewed individually, when the loans add diversity to the pool. I explicate a model in which competing issuers endogenously form CMBS pools by trading loans with each other, in which variation in potential trading opportunities provides exogenous variation in the propensity for loans to be securitized "in-house." I estimate this model using moment inequalities implied by optimality conditions on firm profits, controlling for selection on unobserved public information, in order to determine the mean effect of adverse selection on the quality of in-house versus non–in-house loans. I find that adverse selection is empirically important and accounts for as much of the observed variation between in-house and non–in-house loan performance as several key reported loan characteristics. These findings have implications for current regulatory reforms concerning securitization markets.

*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 the originators of assets to other parties. Across the spectrum of securitized asset classes—from subprime residential mortgages to car loans and credit card receivables—a pervasive problem is that if the originators of the assets have private information about quality, they have an incentive to sell the assets that are worst in privately observed dimensions while retaining the best.

This paper studies the commercial mortgage-backed securities (CMBS) market. I focus on explaining differences between the performance of loans that were securitized "in-house"—that is, pooled into CMBS deals issued by the firm that originated the loans—as opposed to loans sold into CMBS deals issued by firms other than the originator. In either case, securitization formally transfers ownership of the loans to the investors in the CMBS securities. However, both the issuer of a CMBS deal and the originators of the loans backing the deal (which may or may not be the same firm as the issuer) retain a reputational and sometimes financial interest in the ex post performance of the loan pool. When the originator and issuer are the same firm, they internalize each other's reputational and financial interest in the performance of the loans in the deal, giving loan originators a strong incentive to place loans about which they have the most favorable private information into their own CMBS issuances, while selling the rest to other issuers.

In fact, empirically, loans securitized in-house do *not* perform significantly better on average than other securitized loans: after controlling for observable loan characteristics, in-house loans have a 21-percent *greater* hazard of default. However, this comparison does not by itself prove the absence of adverse selection, because selection on unobserved *public* information may also account for differences in performance between loans securitized in-house and non–in-house. Although we could control for all of the standard reported loan characteristics observed by market participants, certain information such as market rumors about specific loans may not be reflected in the reported data.

There are at least two reasons why selection with respect to public information may potentially be nonrandom. First, the originator of a loan is the residual holder of the loan in the event that it cannot be securitized. Consider a lender that has originated some loans and then faces an unexpected drop in demand for the loans due to a deterioration in investor sentiment either marketwide or about the specific loans. A sufficiently bad shock may result in the loans being "warehoused" (retained) on the originator's balance sheet, which creates an opportunity cost because a loan on the books uses up scarce capital and prevents the originator from making new loans.[1] CMBS issuers therefore have a stronger incentive than their competitors to securitize their own loan originations, because they internalize the cost of warehousing these loans if they are not securitized. For example, the originator of a loan that has a bad reputation may be more willing than other CMBS issuers to securitize that loan, even if doing so results in being penalized by investors through earning a lower spread on the deal.[2] Thus, "warehouse risk" implies that loans should have worse unobserved characteristics, on average, conditional on being securitized in-house.

A second potential source of nonrandom selection with respect to public information arises from the fact that putting together a CMBS deal is a portfolio decision, with firm profits determined by the joint distribution of asset returns. For example, consider two firms that each specialize in lending in a particular geographic market (California or Illinois), both of which also issue CMBS deals. Due to the diversification benefits of combining loans from different geographic markets, the firms have gains to trade: the benefit of adding an Illinois loan to a CMBS pool that is already heavy in California loans is greater than the benefit of adding one more California loan with the same marginal probability of default. Therefore, the unobserved quality of a loan is likely to be negatively correlated with the diversification benefits that the loan brings to the pool. Empirically, loans originated by a single firm—and in particular, loans originated by the CMBS issuer—tend be somewhat similar to each other along some observable dimensions, although the degree of similarity is only slightly higher than among loans from different originators. This fact suggests that loans should have slightly better unobserved quality, on average, conditional on being securitized in-house.

The key question I address is what component of the difference in performance between in-house and non–in-house loans, on average, can be attributed to adverse selection—that is, selection on private information. I should note at the outset that, similar to much of the literature on asymmetric information [cite], I cannot disentangle adverse selection from moral hazard, both of which would be reflected in a quality differential after controlling for public information. In theory, moral hazard arises if CMBS issuers engage in hidden actions that differentially affect the ex post performance of in-house versus non–in-house loans.[3] The facts of the CMBS industry suggest that

[1] The margin between loans that are securitized at all versus retained on the originator's balance sheet creates another obvious opportunity for adverse selection. I abstract from this first margin in order focus on another margin—where the loans end up conditional on being securitized. The first margin can be viewed as determining the quality distribution among loans to be securitized, while the second margin determines the allocation of those loans between in-house and non–in-house deals.

[2] The spread is the difference between the interest revenues generated by the underlying loans and the cost of interest payments that must be paid to the CMBS investors.

[3] For example, originators of CMBS loans are often retained as their primary servicers, and may exercise discretion over their monitoring effort. (Different types of servicers are associated with a CMBS deal. The primary servicer of each loan collects payments from the borrower under ordinary circumstances. The special servicer deals with delinquent loans and "advances" payments owed by the borrowers to the CMBS pool while working out the loans.

ex ante selection is far more important than ex post hidden action [cite], but I cannot rule out moral hazard explaining part of my "adverse selection" effect. The skeptical reader could nevertheless think of my key findings as pertaining to the effect of incentive distortions, more generally, on loan performance.

The classic approach to dealing with the endogeneity problem presented by unobserved public information would be to identify an instrument that is correlated with the propensity of a loan to be securitized in-house but uncorrelated with unobserved loan quality. The effect of selection on private information would then be identified by the manner and degree to which observed loan quality changes with respect to the instrument.

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 unobservables. For example, an issuer facing tight short-term funding constraints has more incentive to sell loans it has originated to other issuers, 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, geographic proximity between the property locations of loans and the location of the deal issuer—which presumably affects the probability of choosing particular loans for the deals—is almost certainly correlated with the quality of the issuer's private information about the loans.

To deal with the problem of identification, I estimate a structural model that explicitly specifies CMBS issuers' decisions regarding which loans to include in their pools. Because the pools comprise 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. 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.[4] However, it provides a source of exogenous variation in the structural model.

The master servicer collects payments from all loans underlying the CMBS pool, and must give consent for any of those loans to be modified.)

[4] Having more available trading partners increases the *potential* for selection effects (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.

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 the portfolio diversification incentives of competing firms provide a second source of identification. In particular, if a given loan's exogenous characteristics are negatively correlated with those of loans originated by the loan originator's trading partners, that 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 the other firms' originations 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 (2011) 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 effect of in-house securitization on loan quality.

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—in other words, ruling out adverse selection. 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 the arguably more plausible assumption of asymmetric information imply that adverse selection accounts for in-house loans having an 84-percent lower hazard of default on average, controlling for other observed loan characteristics. This effect is comparable in magnitude to any of several standard loan characteristics used by market participants to evaluate the quality of loans. The structural estimates thus contrast with the reduced-form finding that in-house loans perform slightly worse overall, indicating the importance of controlling for the endogeneity of the loan portfolios.

Estimates under the assumption of symmetric information are similar and imply an 86-percent lower hazard of default for in-house loans. The model allows for such an effect because it does not impose that privately observed quality must be the same for in-house and non–in-house loans. Nevertheless, the finding is not a plausible equilibrium outcome under the assumption of symmetric information. Moreover, the difference in estimates between the two assumptions is consistent with the direction of the bias if imposing symmetric information results in model misspecification, which reinforces the importance of allowing for asymmetric information in the model.

The empirical magnitude of adverse selection in the CMBS market has important implications for policy. For example, Title IX of the Dodd-Frank Wall Street Reform and Consumer Protection Act (2010) places a floor on the financial interest that loan originators are required to retain in the loans that they sell into securitization pools. The greater the empirical importance of private information about loan quality, the stronger the policy's likely effect on quality differences according to the mode of securitization, and the greater its effect on the quality of securitized loans more generally.

This paper is linked to four distinct literatures. First, it joins the large empirical literature on identifying and estimating adverse selection (e.g., Finkelstein and Poterba, 2006; Cohen, 2005). Second, 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 information about a set of loans or other assets. 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 the CMBS market, many of whose incentive distortions are similar to those in other more well-studied securitization classes such as residential mortgage-backed securities (e.g., Ashcraft, 2007). Existing works on CMBS tend to focus on the determinants of default for individual loans. Early works by Ambrose and Sanders (2001), Archer et al. (2002), and Deng et al. (2004) relate loan performance to observable 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 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. However, I then use individual loan performance as an input to a model of 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.

Finally, 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, 2011). 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. While applying methodology from industrial organization, my specification of firms' objective functions tries to reflect the institutions of the securitization industry by adhering closely to the way in which practitioners actually modeled portfolio returns. 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 and accounted for almost 40 percent of the CRE loans originated in 2007.[5] 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 issuers. The market participants in loan origination include investment banks, commercial banks, investment companies, specialty finance companies, and "conduit" firms. Except for the conduits, most of these firms originate both loans that remain on their own balance sheets as well as loans that are securitized as CMBS.[6]

CMBS deals are put together by one or more issuers, one of which acts as the *lead underwriter*, also known as the bookrunner. In some cases, two or three issuers serve as co-leads. I use the term "issuer" synonymously with "lead underwriter."[7] During the sample period, the CMBS issuing market was dominated by 23 major investment and commercial banks, including 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 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 do not always) include a large number of loans that were originated in-house by the issuer. 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 issuers as well as the conduits, insurance companies, and finance companies.[8]

[5] Federal Reserve Board's December 2011 Flow of Funds Tables L.219, L.220, F.219, and F.220.

[6] The conduits 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.

[7] In the industry jargon, the term "issuer" is often used to refer to any of the originators with loans in the deal, not just the lead udnerwriter.

[8] If a CMBS deal include a large number of loans that were originated by another issuer, the latter firm was often

At some point before the deal cutoff date, the issuer 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.[9] 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.[10] In practice, there are many complications beyond this basic schema. For example, certain "interest-only" (IO) tranches are derived by carving out the interest payments from other tranches. Moreover, the complexity of the deals tended to increase throughout the 2000s, as investors demanded ever more specific claims.[11]

In Figure 1, the vertical axis of the scatterplot shows the proportion of each deal (weighted by loan balance) comprising in-house loans, while the horizontal axis shows the deal issuer's share of all loan originations during a 180-day window around the time of the deal. The proportion of in-house loans varies greatly but is on average significantly higher than the deal issuer's share of overall originations, suggesting that loans have a much higher-than-random chance of being securitized in-house. I discuss later how the in-house share varies conditional on the horizontal axis.

Both the loan originators and the deal issuer retain a stake in the performance of the pool after finalizing the deal. Often the retained stake is explicit, for issuers sometimes keep the B piece.[12]

brought on as an additional issuer, though not necessarily as a co-lead.

[9]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 or expected loss during an adverse credit event 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.

[10]In securitizations other than CMBS, such as RMBS, while more senior bonds are exposed to less credit risk, they are also the first to be paid off when loans prepay. This feature exposes senior bondholders to greater "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. However, duration risk is not a significant factor in CMBS, where prepayment typically takes place by means of defeasance: the borrower is required to substitute other income-producing collateral (typically U.S. Treasuries) to produce the same stream of income as that which would have been generated by the prepaid loan.

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

[12]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 issuers.

Even when all tranches including the B piece are sold to investors, the originators and deal issuer retain an implicit stake insofar as the performance of the securities affects firm reputation.

Vertical integration between originators and issuers generates an incentive for adverse selection, because the originator internalizes the issuer's reputational and financial incentives, and vice versa.[13] Thus, the performance of loans originated and securitized by the same firm has a bigger effect on that firm's payoffs than the performance of loans that it has originated and sold to other issuers. If issuers have private information about loans that they originated, adverse selection would result in in-house loans having better characteristics along these dimensions. However, as noted in the introduction, even in the absence of private information, there are additional selection effects that also contribute to quality differences between in-house versus non–in-house loans.

First, an originator that is unable to securitize a loan must warehouse that loan on its balance sheet and thereby incur an opportunity cost of capital. In that event, a vertically integrated CMBS issuer would internalize the originator's opportunity cost of capital, which increases the issuer's preference for holding in-house loans relative to loans originated by its competitors. The presence of warehouse risk implies that in-house loans should perform worse after controlling for observed loan characteristics, insofar as we cannot observe all payoff-relevant information about a loan.

On the other hand, an issuer may idiosyncratically value loans that are individually worse-performing if those loans offer diversification benefits due to negative correlation in returns with other loans in the issuer's portfolio. For example, if real estate values in California and Illinois are negatively correlated and a CMBS issuers already has a portfolio that is heavy in California loans, then it will have a greater marginal valuation of loans from Illinois, even if it has slightly worse individual quality. Empirically, we would expect diversification incentives to have a small downward effect on the relative quality of in-house loans: observable characteristics tend to be positively correlated among loans originated by the same lender, although the average degree of similarity is only slightly higher than among randomly chosen loans.[14] This effect implies that issuers must be compensated for holding in-house loans through higher unobserved quality, which

[13] This implication presumes that the internal governance of vertically integrated firms is coordinated. If the loan-originating arm of Citigroup operates semi-independently of the CMBS-underwriting arm, which is an empirical question far-removed from the focus of this paper, we would not expect to find especially strong evidence of adverse selection.

[14] For example, the Herfindahl-Hirschmann Indices (HHIs) for property type and location (MSA), by share of the total loan balance of all loans in the sample, are .249 and .028, respectively. The corresponding HHIs for loans originated by a particular originator are on average .264 and .063 for originators that have at least least 500 loans in the dataset. (Comparison of the overall-sample HHIs with HHIs for originators with many fewer than 500 loans is not valid, because the HHIs are mechanically driven higher when the sample size is small). The HHI for loan size quartile , by share of the total number of loans in the sample, is .25 by construction, while the corresponding HHI for a specific originator is on average .274.

leads to non-random selection in the opposite direction to selection induced by warehouse risk.

The lack of a strong relationship between the in-house proportion of loans in a deal and the issuer's share of overall loan originations during the surrounding time period, which varies along the horizontal axis of Figure 1, also provides suggestive evidence of diversification incentives. We would expect this relationship to be increasing if loans were allocated randomly to deals. However, to the extent that issuers are purchasing loans from other originators for diversication purposes, when the issuer is behind a larger share of the loans originated in a given time period, maintaining pool diversification will tend to require selling a greater proportion of the loans it originated. To summarize, the overall effect of nonrandom selection is ambiguous: conditioning on observed loan characteristics, adverse selection implies better performance for in-house loans; the warehouse risk effect implies worse better performance; and diversification incentives could theoretically go either way but probably imply somewhat better performance insofar as loans originated by the same firm tend to have positively correlated returns.

Finally, note that adverse selection is unlikely to be *as* severe in CMBS as in the more familiar case of residential mortgage-backed securities (RMBS), including the now largely defunct subprime RMBS market. 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.[15] (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 loan underwriting characteristics.[16]

Data

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,

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

[16] 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).

implying a balloon payment after 10 years.[17]

I observe loan characteristics at origination and the month in which the loan first becomes delinquent (if ever) until the censoring date, June 2011. These characteristics include all of the standard ones used by market participants to evaluate loan quality, including the product type (amortization schedule and maturity), DSCR, LTV, property type and location, 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[18]), 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.[19] 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 issuers 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 identities of the loans in the pool, the tranche structure of the securities issued from the deal, and ratings of the securities by the four main rating agencies,[20] which allows me to construct

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

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

[19]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.

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

the total principal balance of all tranches rated AAA.[21] After dropping observations with missing data, the final estimation sample comprises 441 deals containing 57,353 loans.

Stylized facts

Table 1 provides summary statistics for the observed loan characteristics and outcomes. The average loan is censored at 54 months. 13.9 percent of loans default, with 35 months being the average loan age at default. 23.9 percent of loans prepay, with 84 months being the average loan age at prepayment. For about 22 percent of loans, no occupancy information is available, either due to missing data or because occupancy was never reported for those loans in the deal prospectuses. To avoid having to drop such a large number of observations, I include an indicator for whether occupancy information is missing, and set occupancy to zero for the missing observations.

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 issuer (lead underwriter).[22] In some cases, an originator may have two co-originators. Likewise, a CMBS deal may have two or three co-lead underwriters. For such cases, I consider a loan to be securitized in-house if any of the originators matches with any of the co-leads. By this definition, 54.1 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 (Cox, 1972). The hazards model controls for censoring, with either prepayment or the end of the sample period serving as the censoring event. I include fixed effects for originators and for interactions between the ten geographic regions and the three property categories, which do not substantially affect the estimates of the remaining parameters.[23] The three property categories are "office/retail/hotel,"

[21] 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.

[22] Firm ownership 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.

[23] The regions are "New England," "Mid-Atlantic," "Midwest, Eastern," "Midwest, Western," "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 "cross-collateralized" by multiple properties spanning more than one region.

"multifamily" (apartments), and "industrial/other."[24]

The loan characteristics have the sorts of effects that we would expect. For example, interest-only loans are more 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 occupancy ratio is 22 percent.[25] To avoid having to drop these observations, I included an indicator for missing occupancy ratio information and set occupancy ratio equal to zero for the missing observations. Not surprisingly, a higher coupon spread,[26] 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 default at an intensity roughly three times that of loans originated in 2000 and before.

The most interesting coefficients are those for 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 21 percent higher than otherwise. This difference reflects the net effect of both adverse selection and selection on unobserved public information. The overall direction of the effect suggests that adverse selection is outweighed by selection on unobserved public information. The goal of the structural model is to disentangle these two effects.

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. This interaction is consistent with the possibility that originators securitize in-house loans that have bad publicity to avoid warehousing them, but only when the number of such loans is small. Alternatively, the interaction is also consistent with selection effects driven by diversification incentives. If the marginal benefit to the portfolio of an additional in-house loan diminishes with the number of in-house loans in the pool (due to positively correlated returns), when the share of in-house loans is large, there must be unobserved factors that rationalize holding a large proportion of in-house loans. The interaction effect may also reflect differences in signaling

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

[25]These cases could either be due to missing data or because occupancy information was simply not reported in the loan prospectuses. For example, an industrial property with a single occupant may not have a reported occupancy ratio.

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

costs depending upon the total number of loans originated by the issuer. 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.

Specification 3 naïvely applies an instrumental variables approach. In a first stage, I regress in-house status on time dummies and the number of CMBS deals around the time of the loan origination and the total number of loan originations by the originator around the time of the loan origination. I then estimate the effect of the fitted value of in-house status in the hazard regression. For reasons described in the introduction, the first-stage regressors are not valid instruments, and I include this specification primarily for comparison purposes with the structural estimates.

3 Model

Consider 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 issuers, with the remainder originated by lenders that do not underwrite 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 issuer of deal i has itself originated as well as loans obtained from other originators. The issuer maximizes net profit, which depends upon the contents of $\mathcal{J}_i$ and the cost basis for the loans.

In the data, each issuer does multiple CMBS deals at different points in time. However, for simplicity I assume that issuers maximize profits independently for each deal. That is, I abstract from profit spillovers across deals such as due to dynamic reputation effects. Given this assumption, we can think of each deal as equivalent to a "firm," and I use the index i to refer to both the deals as well as their issuers.[27] 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, which I then extend to a joint distribution of default times across loans. The joint distribution parameters serve as inputs into the payoff structure of the pool-formation game, which I describe last.

[27] 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 distribution of the random time at which the loan defaults. A loan j in deal i can be characterized by a vector of exogenous characteristics w_{ij}; the endogenously determined indicator of in-house status v_{ij}, which equals 1 if issuer i originated j and equals 0 otherwise; and a random, unobserved (to the econometrician) quality ξ_j. 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.[28] 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_{ij} (normalized to be in terms of months since the cutoff date of pool i) is distributed as follows:

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

Conditional on the random term ξ_j, the above function entails the standard proportional hazards assumption. Namely, the hazard of default at time t, $Pr(T_{ij} = t | T_{ij} \geq t)$, is equal to the product of the "baseline" hazard function $\psi(t)$ and the constant proportion $exp(\alpha_1' w_{ij} + \xi_j)$. The distribution of ξ_j may 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:

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

Given the reduced-form nature of the model so far, the dependence of $H(\xi_j | v_{ij})$ on v_{ij} captures the cumulative effects of selection on unobserved quality (both private and public information). To help set things up for the structural model, we can decompose ξ_j—which can be thought of as j's

[28] 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 assume that loan seasoning is exogenously determined, based on the fact that in the originate-to-distribute industry model, firms try to securitize loans as quickly as possible after origination. I include loan seasoning as an explanatory variable to adjust for the fact that default is a process that occurs over the course of the *loan's* life, but I compute portfolio returns as of the *deal's* cutoff date, which typically comes at some point following the date of origination for the loan. We will need the coefficient on loan seasoning to compute expeted returns under counterfactual scenarios in which the loan is bundled into deals other than the actual one.

true unobserved quality—into unobserved public information about j's quality, ξ_j^{pub}; and j's true quality net of what is commonly observed by all firms, ξ_j^{npub}.

$$
\begin{aligned}
&\xi_j = \xi_j^{npub} + \xi_j^{pub} = \\
&\quad \mathcal{E}[\xi_j^{npub}|v_{ij}=0] + (\mathcal{E}[\xi_j^{npub}|v_{ij}=1] - \mathcal{E}[\xi_j^{npub}|v_{ij}=0]) \cdot v_{ij} + \xi_j^{npub} - \mathcal{E}[\xi_j^{npub}|v_{ij}] + \xi_j^{pub} \equiv \\
&\quad \mathcal{E}[\xi_j^{npub}|v_{ij}=0] + \alpha_0 v_{ij} + \xi_j^{npub} - \mathcal{E}[\xi_j^{npub}|v_{ij}] + \xi_j^{pub} \equiv \\
&\quad \mathcal{E}[\xi_j^{npub}|v_{ij}=0] + \alpha_0 v_{ij} + \tilde{\xi}_j^{npub} + \xi_j^{pub} = \alpha_0 v_{ij} + \tilde{\xi}_j^{npub} + \xi_j^{pub}
\end{aligned}
\tag{3}
$$

The term $\alpha_0 \equiv \mathcal{E}[\xi_j^{npub}|v_{ij}=1] - \mathcal{E}[\xi_j^{npub}|v_{ij}=0]$ captures the mean effect of selection on private information for in-house loans, and $\tilde{\xi}_j^{npub} \equiv \xi_j^{npub} - \mathcal{E}[\xi_j^{npub}|v_{ij}]$ denotes the residual. The final equality comes from normalizing $\mathcal{E}[\xi_j^{npub}|v_{ij}=0]$ to zero, because it is not separately identified from the baseline hazard function. This decomposition shows that if v_{ij} has an unknown correlation with ξ_j^{pub}, then α_0 cannot be identified from the reduced-form model.

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.[29] 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.[30] Thus I abstract from cases in which a delinquent loan recovers and becomes current again, or loans that become delinquent multiple times.[31]

Finally, I ignore losses due to balloon default. Because the actual maturity of most loans is

[29]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.

[30]Instead of making LGD dependent on j, for simplicity I fix LGD at 0.4, which is a typical value attained in commercial mortgages during this time period.

[31]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.

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 interest rate and amortization schedule of the mortgage. I treat all loans as paying a fixed coupon and having a constant rate of amortization,[32] 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, the distribution of the return on loan j in deal i, denoted by the random variable R_{ij}, is thus

$$Pr(R_{ij} < r) = \inf_t \left\{ Pr(T_{ij} < t) \quad \middle| \quad \sum_{\tau=0}^{t} \left(\prod_{\tau'=0}^{\tau-1} \delta_{\tau'} \right) P_j + \delta^t (1 - LGD) B_j(t) < r \right\}. \tag{4}$$

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 valid 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.[33] 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[34]—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 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 $F_j(T_{ij}) \equiv Pr(T_{ij} < t | w_{ij}, v_{ij})$ and consider the loans in portfolio $\mathcal{J}_i$,

[32] In the data, almost all mortgages are fixed-rate.

[33] 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).

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

indexed by $j = 1, 2, \ldots, J_i$. The joint distribution of $T_1, T_2, \ldots, T_{J_i}$ is given by

$$F(T_1, T_2, \ldots, T_{J_i}) = \Phi(\Phi^{-1}(F_1(T_1))\Phi^{-1}(F_2(T_2)), \ldots \Phi^{-1}(F_{J_i}(T_{J_i}))\ ;\ \Omega), \tag{5}$$

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 correlation in loan returns via equation (4).

Issuer utility

Each deal issuer i maximizes its utility with respect to a set of *feasible* portfolios, which I define as follows. Let t_i^0 denote the cutoff date for deal i and t_j^0 the origination date for loan j, both of which are exogenously determined. I assume that any loan j can feasibly be matched with any deal i such that $t_j^0 \leq t_i^0 < t_j^0 + d$ for $d > 0$. In other words, loans can be matched with deals with cutoff dates just after the origination date of the loan, but not more than d days after.[35] I denote the set of loans that can feasibly be matched with firm i by $\mathcal{D}_i$.[36] The powerset $\mathcal{P}(\mathcal{D}_i)$ then denotes the set of feasible portfolios for i. Firm i's gross utility is a function of its chosen portfolio, $\mathcal{J}_i \in \mathcal{P}(\mathcal{D}_i)$, and is expressed as follows:

$$\begin{aligned} u_i \quad &= \quad E_{\{R_{ij}\}_{j\in\mathcal{J}_i}|\{w_{ij}\}_{j\in\mathcal{J}_i},\{v_{ij}\}_{j\in\mathcal{J}_i},\alpha_0,\hat{\alpha}_1,\hat{\Omega}} \left[\min\left\{ b_i(\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i})\ ,\ \sum_{j\in\mathcal{J}_i} R_{ij} \right\} + \right. \\ &\qquad \left. \beta_t \cdot \max\left\{ 0\ ,\ \sum_{j\in\mathcal{J}_i} R_{ij} - b_i(\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i}) \right\} \right] \quad + \quad \sum_{j\in\mathcal{J}_i} z_{ij} \\ &\equiv \quad U_i(\mathcal{J}_i) \quad + \quad \sum_{j\in\mathcal{J}_i} z_{ij}. \end{aligned} \tag{6}$$

The utility function has both an observed component and an unobserved component, because the econometrician cannot directly observe issuer i's private information or certain public information

[35] 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, if we observe a loan j in deal i such that $t_i^0 \geq t_j^0 + d$, I assume that set of deals that j could feasibly be matched with comprises all deals i' such that $t_j^0 < t_{i'}^0 \leq t_i^0$.

[36] 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.

about each loan j. The expectation term above, which I denote by $U_i(\mathcal{J}_i)$ as shorthand, specifies an approximation of the gross utility based on observables. The inside of the expectation term is an approximation of the payoff corresponding to a specific realization of returns for loans in the portfolio, $\{R_{ij}\}_{j\in\mathcal{J}_i}$, whose functional form I explain at length below. The expectation is taken over a return distribution generated conditional on exogenous loan characteristics $\{w_{ij}\}_{j\in\mathcal{J}_i}$ and in-house indicators $\{v_{ij}\}_{j\in\mathcal{J}_i}$. Specifically, I set the loghazard of default for loan j to $\alpha_0 v_{ij} + \hat{\alpha}_1' w_{ij}$, where $\hat{\alpha}_1$ are the estimated parameters from the first stage.[37] The loghazard for in-house loans is shifted by an amount α_0, which captures the mean effect of selection on private information. The joint distribution of returns on loans in the portfolio is determined by the marginal returns, as implied by the hazards for individual loans, in conjunction with the estimated copula parameters $\hat{\Omega}$. Constructing $U_i(\mathcal{J}_i')$ for a counterfactual portfolio $\mathcal{J}_i'$ requires additional consideration of how to treat loans that are securitized in-house in the counterfactual case but not in the data, or vice versa. I discuss this matter in Appendix C.

The term z_{ij} is the deviation of firm i's subjective beliefs about the return to having loan j in the deal, relative to our observable measure of the payoff. Recall from equation 3 that the unobserved quality of loan j in deal i can be expressed as $\xi_j = \alpha_0 v_{ij} + \tilde{\xi}_j^{npub} + \xi_j^{pub}$, which combines unobserved public information (ξ_j^{pub}), the mean effect of adverse selection ($\alpha_0 v_{ij}$), and the residual of after netting out public information and the mean effect of adverse selection. z_{ij} can thus be loosely thought of as a signal combining public information with i's subjective beliefs about the quality of j, net of expected adverse selection.[38] The summation $\sum_{j\in\mathcal{J}_i} z_{ij}$ is the cumulative effect of firm i's signals about individual loans on its anticipated payoff from the deal, where the functional form presumes that i's signals about individual loans have an additive effect.[39] 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 beliefs.

The remainder of this subsection details i's returns conditional on a particular realization of loan returns (the inside of the expectation term). For simplicity I aggregate the returns to all tranches subordinate to the AAA bond, which I collectively denote as the "B piece." In the utility function, the "min" and "max" terms are the values of the AAA bond and the B piece, respectively. In other words, I differentiate between a dollar of principal or interest going to the AAA bondholder versus

[37] I zero out the unobserved heterogeneity embodied in the random distribution of ξ_j.

[38] The analogy is loose because, strictly speaking, the effect on payoffs of loan j's unobserved quality, ξ_j, may be dependent on what other loans are in the portfolio, a result of the correlation in loans returns. As expressed, the effect z_{ij} does not depend upon the remaining loans in the portfolio, which is slight simplification.

[39] This assumption does not preclude i's signals about loans j and j' from being correlated, but merely states that i's unobserved utility from loan j does not depend upon its signal about loan j'.

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.[40]

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 issuer as caring about the *revenues* that it can generate 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.[41] 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 $\sum_{j \in \mathcal{J}_i} R_{ij}$ net of payments to the AAA bondholders. The time-varying parameter β_t captures the valuation 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 an opportunity cost of capital that is time-varying and exogenous to the model. 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 issuers selling any bonds at all as opposed to retaining the entire portfolio suggests a positive opportunity cost of capital ($\beta_t < 1$). On the other hand, if the firms have a very high revealed preference for risk, β_t could be higher than 1.

For simplicity, the size of the AAA tranche, $b_i(\{w_{ij}, v_{ij}\}_{j \in \mathcal{J}_i})$, is treated as an exogenous function of the observable portfolio characteristics. Specifically, I assume that the AAA tranche for deal i is

[40] CMBS deal structures are considerably simpler than analogous structures for 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.

[41] 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 issuer, and if issuers do not have market power, then a standard model would imply that revenues from bond sales would equal the expected value of the bond.

required to have enough credit support such that conditional on $\{w_{ij}, v_{ij}\}_{j \in \mathcal{J}_i}$, the probability of it incurring losses of any size is equal to some probability p_i, which is a parameter.[42] 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).[43] 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 issuer 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).[44] Moreover, at least in principle, the amount of credit protection may signal the issuer's private information and thus affect the demand for the securities.[45] 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. Prior to making any trades, firm i observes exogenous characteristics w_{ij} and forms subjective beliefs about the quality of all loans $j \in \mathcal{D}_i$.[46]

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, $c_j^{ii'} = B_j^0 \cdot exp(\gamma' w_j) + \zeta_j^{ii'}$, where B_j^0 is the initial principal on loan j, w_j is the vector of exogenous and non–match-specific

[42] For any given size of the AAA tranche, this probability is implied by the first- and second-stage estimates, similar to the expected values of the AAA tranche and the B piece.

[43] 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).

[44] 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, issuers 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 issuer's business by offering looser standards.

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

[46] Recall from footnote 28 that the only match-specific characteristic in $\{w_{ij}\}$ is the loan seasoning at the cutoff date conditional on j being matched to i. Without this characteristic, we could omit the i subscript.

loan characteristics[47], γ are parameters, and $\zeta_j^{ii'}$ is an unobserved error.[48] Furthermore, for any pair of firms i and i', firm i incurs a fixed cost c_f in order to hold any positive amount of loans in the feasible set of firm i', $D_{i'}$. Thus, selling all loans in the feasible set of firm i' results in a savings of c_f by firm i.[49]

Optimality requires that firm i's net profit given its actual portfolio $\mathcal{J}_i$, taking into account transfer payments, must be weakly higher than the net profit given any alternative feasible portfolio $\mathcal{J}_i'$. 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 following are necessary optimality conditions:

$$\begin{aligned} U_i(\mathcal{J}_i) - U_i(\mathcal{J}_i') \; - \; \textstyle\sum_{j \in \mathcal{J}_i \backslash \mathcal{J}_i'} c_j^{ii'} \; + \; \textstyle\sum_{j \in \mathcal{J}_i \backslash \mathcal{J}_i'} z_{ij} \; + \; \textstyle\sum_{j \in \mathcal{J}_i' \backslash \mathcal{J}_i} c_j^{ii'} \; - \; \textstyle\sum_{j \in \mathcal{J}_i' \backslash \mathcal{J}_i} z_{ij} \\ - \; c_f \cdot \mathbf{1}\{\mathcal{J}_i \cap D_{i'} \neq \emptyset\} \cdot \mathbf{1}\{\mathcal{J}_i' \cap D_{i'} = \emptyset\} \; + \; c_f \cdot \mathbf{1}\{\mathcal{J}_i \cap D_{i'} = \emptyset\} \cdot \mathbf{1}\{\mathcal{J}_i' \cap D_{i'} \neq \emptyset\} \\ \geq 0, \quad \forall i, i' \quad \forall \mathcal{J}_i' \in \mathcal{P}(\mathcal{J}_i \cup (\mathcal{J}_{i'} \cap \mathcal{D}_i)). \end{aligned} \tag{7}$$

4 Estimation and Identification

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

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 (Heckman and Singer, 1984) and to 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 distribution

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

[48] The exponential functional form constrains the observed component to be positive.

[49] Denoting the set of loans in the intersection of the observed portfolio of firm i with the feasible set for firm i' by $A(i') \equiv \mathcal{J}_i \cap \mathcal{D}_{i'}$, the portfolio of firm i can be expressed as $\mathcal{J}_i = A(1) \cup A(2) \ldots \cup A(i') \ldots \cup A(I)$ for $i' = 1, 2, \ldots I$. If $A(1) = A(2) = A(I)$, then firm i's total fixed costs associated with holding portfolio $\mathcal{J}_i$ equal c_f. Otherwise, the fixed costs equal c_f times the number of subsets in the partion of $\mathcal{J}_i$ defined by $A(1), A(2) \ldots A(I)$ that belong to $A(i')$ for exactly one i'.

over the support, conditional on 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

$$Pr(T_{ij} < t|w_{ij}, v_{ij}) = 1 - \sum_{r=1}^{R} \theta(r|v_{ij}) exp\left[- \int_{\tau=0}^{t} \psi(\tau) d\tau \; e^{\alpha_1' w_{ij} + \xi^r} \right]. \tag{8}$$

The semiparametric specification is both more flexible than the traditional parametric approach as well as computationally less burdensome. On the one hand, the set of parameters is larger 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 compute the likelihood without taking simulation draws, an advantage that outweighs the larger number of parameters unless the grid is chosen to be extremely fine.[50]

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 sample. The estimation procedure can be stated as the following constrained maximization.

$$\begin{aligned}
\max_{\alpha_1, \theta, \psi} \sum_j \sum_{r=1}^{R} \Bigg[\theta(r|v_{ij}) \quad & \mathbf{1}(t_j > t_j^c) \cdot exp\left[- \int_{\tau=0}^{t_j^c} \psi(\tau) d\tau \; e^{\alpha_1' w_{ij} + \xi^r} \right] \\
+ \quad & \mathbf{1}(t_j = t_j^c) \cdot exp\left[- \int_{\tau=0}^{t_j} \psi(\tau) d\tau \; e^{\alpha_1' w_{ij} + \xi^r} \right] \psi(t_j) e^{\alpha_1' w_{ij} + \xi^r} \Bigg] \\
s.t. \quad & \sum_{r=1}^{R} \theta(r|v_{ij} = 0) \;=\; 1, \\
& \sum_{r=1}^{R} \theta(r|v_{ij} = 1) \;=\; 1, \\
& 0 \;\leq\; \theta(r|v_{ij}) \;\leq\; 1, \qquad \forall\, r,\; v_{ij}.
\end{aligned} \tag{9}$$

[50]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 necessarily globally concave in α_1, which enters nonlinearly. However, I suspect that 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.

The parameters 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 (w_{ij} and v_{ij}) and the propensity 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 have better observable characteristics on average than loans with low ξ_j. For the set of surviving loans, the quality of unobservable and observable characteristics becomes increasingly negatively correlated as t increases. Therefore, greater dispersion in ξ_j implies that as t goes up, there is a greater reduction 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 computational burden.[51]

Estimation and identification of joint loan outcome distribution

For tractability, I impose certain restrictions on Ω, the correlation structure of the copula function (5). 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. 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 loan j belongs. For a loan j in CMBS deal i, I assume that the distribution of the time to default, $Pr(T_{ij}<t|w_{ij},v_{ij},i)$, depends upon the idiosyncratic factor η_{ij} and the category-specific factor $\varepsilon_{ik(j)}$ in the following way.

$$\begin{gathered} F(t|w_{ij},i)=Pr(T_{ij}<t|w_{ij},v_{ij},i)=\Phi(\varepsilon_{ik(j)}+\eta_{ij}), \\ \text{with} \quad (\varepsilon_{i1}\ldots\varepsilon_{iK})\sim N(\mathbf{0},\Omega_\varepsilon), \quad \eta_{ij}\sim N(0,\omega_\eta), \quad \text{and} \quad \mathcal{E}[\eta_{ij}\varepsilon_i]=\mathbf{0}. \end{gathered} \tag{10}$$

The category-specific factors are jointly distributed $N(0,\Omega_\varepsilon)$ 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 attempt to specify or

[51]Specifically, I assume that $\psi(t)=(a+bt+ct^2)^2$ for parameters a, b, and c. Restricting $\psi(t)$ to be the square of a quadratic function ensures that the baseline hazard is always positive.

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.[52] To get around this problem, I instead estimate Ω_ε and ω_η using an hierarchical Markov-Chain Monte Carlo algorithm that combines Gibbs sampling with a Metropolis-Hastings algorithm that draws from the posterior for the latent random variables $(\varepsilon_{i1} \ldots \varepsilon_{iK})$ and η_{ij}. This procedure is detailed in Appendix A.

Estimation and identification of structural parameters

The effects of the exogenous covariates w_{ij} on issuer's payoffs, α_1, and the correlation structure of the copula, Ω, are identified by the previous stages of the model. However, the parameters α_0, β_t, and γ must be identified from the necessary optimality conditions on 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 computational 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 4 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 necessary conditions (7) were observed, then these inequalities would provide a set of conditions defining the possible values for the structural parameters. However, the deviations 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 this endogeneity problem, I use a strategy first developed by Pakes et al. (2006). Namely, I find particular linear combinations of the necessary conditions such that the contribution

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

of unobservables either: (1) cancels out altogether; or (2) is uncorrelated or negatively correlated with the observable part of the necessary condition. These linear combinations satisfy moment inequalities that set-identify the parameters. I form these moment inequalities under two alternative assumptions about the information structure of the game. The casual reader may wish to skip ahead to the discussion of identification (page 33), after perusing Assumption 1 and Assumption 2 (page 29).

Assumption 1 *(Symmetric information):* $z_{ij} = z_{i'j}$, *for all* i, i'.

Assumption 1 implies that firms have symmetric information. The equality of z_{ij} and $z_{i'j}$ implies that when the buyer and seller transact over a loan j, the selling firm's forgone value as predicted by j's unobserved characteristics equals the gain to the buying firm. Together with the fact that the transfer payment paid by the buyer necessarily equals the transfer payment 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.

Consider any subset of loans $\mathcal{J}$ in the portfolio of firm i whose constituent loans can all feasibly be matched to firm i'. The necessary conditions imply that the firms are each better off with their observed portfolios than if the loans in $\mathcal{J}$ were switched from i to i'. Thus, we have:

$$\begin{aligned} \Delta\pi_i(\mathcal{J}_i, \mathcal{J}_i \backslash \mathcal{J}) \quad &\equiv \quad U_i(\mathcal{J}_i) - U_i(\mathcal{J}_i \backslash \mathcal{J}) + -\textstyle\sum_{j\in\mathcal{J}}(-B_j^0 \cdot exp(\gamma' w_j) - \zeta_j^{ii'} + z_{ij}) \\ &\quad\; - \; c_f \cdot \mathbf{1}\{\mathcal{J} = \mathcal{J}_i \cap D_{i'}\} \geq 0 \\ \Delta\pi_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \cup \mathcal{J}) \quad &\equiv \quad U_{i'}(\mathcal{J}_{i'}) - U_{i'}(\mathcal{J}_{i'} \cup \mathcal{J}) + \textstyle\sum_{j\in\mathcal{J}}(B_j^0 \cdot exp(\gamma' w_j) + \zeta_j^{ii'} - z_{i'j}) \\ &\quad\; + \; c_f \cdot \mathbf{1}\{\mathcal{J}_{i'} \cap D_i = 0\} \geq 0. \end{aligned} \tag{11}$$

In constructing the counterfactual utilities $U_i(\mathcal{J}_i \backslash \mathcal{J})$ and $U_{i'}(\mathcal{J}_{i'} \cup \mathcal{J})$ there are two complications. First, we must determine how to specify the observed component of the counterfactual utilities when the in-house status of loans switches between the actual and counterfactual cases. The exact specification affects the interpretation of Assumptions 1 and 2, and I do so in a way that is consistent with the characterization of Assumption 1 as signifying symmetric information, a detail discussed in Appendix C. Second we must specify the size of the AAA tranche under the counterfactual scenario. (To construct the firm i's utility under the actual scenario, we of course set $b_i(\{w_{ij}, v_{ij}\}_{j\in\mathcal{J}_i})$ to the empirical tranche size, b_i^*.) 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 scenarios. Thus if we let $G_{\mathcal{J}_i}(\cdot)$ and $G_{\mathcal{J}_{i'}}(\cdot)$ represent the return distributions for the actual portfolios of firms i and i', $\mathcal{J}_i$ and $\mathcal{J}_{i'}$; and let $G_{\mathcal{J}_i \setminus \mathcal{J}}(\cdot)$ and $G_{\mathcal{J}_{i'} \cup \mathcal{J}}(\cdot)$ represent the distributions for the counterfactual portfolios; then the tranche sizes under the counterfactual scenarios are given by $G^{-1}_{\mathcal{J}_i \setminus \mathcal{J}}(G_{\mathcal{J}_i}(b_i^*))$ and $G^{-1}_{\mathcal{J}_{i'} \cup \mathcal{J}}(G_{\mathcal{J}_{i'}}(b_{i'}^*))$. In the estimation procedure, I must simulate the portfolio distribution under each scenario. The simulation procedure is complicated and is described in Appendix D.

The unobservables $\zeta_j^{ii'}$, z_{ij} and $z_{i'j}$ are positively correlated with the event that the above inequalities hold, implying that the expectation of the observable part is not necessarily ≥ 0, conditional on the entire inequality holding. However, we can construct three types of moment conditions based on linear combinations of the inequalities. First, we consider perturbations where $\mathcal{J}$ is a singleton comprising just one loan j. 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 $\Delta\pi_i(\mathcal{J}_i, \mathcal{J}_i \setminus j)$ and $\Delta\pi_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \cup j)$ by $\Delta r_i(\mathcal{J}_i, \mathcal{J}_i \setminus j) = U_i(\mathcal{J}_i) - U_i(\mathcal{J}_i \setminus j) - B_j^0 \cdot exp(\gamma' w_j) - c_f \cdot \mathbf{1}\{j = \mathcal{J}_i \cap D_{i'}\}$ and $\Delta r_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \cup j) = U_{i'}(\mathcal{J}_{i'}) - U_{i'}(\mathcal{J}_{i'} \cup j) + B_j^0 \cdot exp(\gamma' w_j) + c_f \cdot \mathbf{1}\{\mathcal{J}_{i'} \cap D_i = 0\}$. 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:

$$\begin{gathered}\chi(i, i', j)[\Delta\pi_i(\mathcal{J}_i, \mathcal{J}_i \setminus j)] + \chi(i', i, j)[\Delta\pi_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \setminus j)] = \\ \chi(i, i', j)[\Delta r_i(\mathcal{J}_i, \mathcal{J}_i \setminus j) + z_{ij} - \zeta_j^{ii'}] + \chi(i', i, j)[\Delta r_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \setminus j) + z_{i'j} - \zeta_j^{ii'}] \geq 0.\end{gathered} \tag{12}$$

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 $j \in \{\mathcal{J}_i \cap D_{i'}\} \cup \{\mathcal{J}_{i'} \cap D_i\}$ (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 21:

$$\mathcal{E}\left[M_1(i, i', j) | x_j^{ii'}\right] \equiv \mathcal{E}\left[\chi(i, i', j)\Delta r_i(\mathcal{J}_i, \mathcal{J}_i \setminus j) + \chi(i', i, j)\Delta r_{i'}(\mathcal{J}_{i'}, \mathcal{J}_{i'} \setminus j) \ \ | x_j^{ii'}\right] \geq 0, \tag{13}$$

Similarly, we can construct moment conditions based on switching *all* loans in $\{\mathcal{J}_i \cap D_{i'}\}$ from firm i to i' and switching all loans in $\{\mathcal{J}_{i'} \cap D_i\}$ from firm i' to i:

$$\begin{aligned}&\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash D_{i'})+\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash D_i)=\\&\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash D_{i'})+\textstyle\sum_{j\in\mathcal{J}_i\cap\mathcal{D}_{i'}}[z_{ij}-\zeta_j^{ii'}]+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash D_i)+\sum_{j\in\mathcal{J}_{i'}\cap\mathcal{D}_i}[z_{i'j}-\zeta_j^{ii'}]\geq 0.\end{aligned}\tag{14}$$

Because the unobservables in the above inequality are summed over all loans that could potentially be switched from one firm to the other, again they drop out in expectation, implying the following:

$$\mathcal{E}\left[M_1(i,i')|x_j^{ii'}\right]\equiv\mathcal{E}\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash D_{i'})+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash D_i)\ \ |x_j^{ii'}\right]\geq 0,\tag{15}$$

Note that moment conditions based on trading larger sets of loans as opposed to individual loans are particularly useful for identifying the fixed cost c_f. The reason is that in the data, firms usually hold more than one loan that they could feasibly sell to a trading partner, in which case only selling all of these loans will affect the incurred fixed cost.

By a symmetric argument, we can establish a second type of moment condition based on the optimality of each firm's observed portfolio relative to adding individual loans or sets of loans from the other firm:

$$\begin{aligned}\mathcal{E}\left[M_2(i,i',j)|x_j^{ii'}\right]&\equiv\mathcal{E}\left[\chi(i',i,j)\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\cup j)+\chi(i,i',j)\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)\ \ |x_j^{ii'}\right]\geq 0\\ \mathcal{E}\left[M_2(i,i')|x_j^{ii'}\right]&\equiv\mathcal{E}\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\cup(\mathcal{J}_{i'}\cap\mathcal{D}_i))+\Delta r_i(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup(\mathcal{J}_i\cap\mathcal{D}_{i'}))\ \ |x_j^{ii'}\right]\geq 0,\end{aligned}\tag{16}$$

A third type of moment inequality is constructed by summing the inequalities for firms i and i' in 11 conditional on firm i holding the loan(s) in $\mathcal{J}$, which expresses the total gains to trade. The transfer payments net out and by Assumption 1, z_{ij} and $z_{i'j}$ cancel each other out:

$$\begin{aligned}&\chi(i,i',j)\left[\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)\right]=\chi(i,i',j)\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)\right]\geq 0,\\&\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash D_{i'})+\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup(\mathcal{J}_i\cap\mathcal{D}_{i'}))=\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash D_{i'})+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup(\mathcal{J}_i\cap\mathcal{D}_{i'}))\geq 0.\end{aligned}\tag{17}$$

Taking expectations,

$$\begin{aligned}\mathcal{E}\left[M_3(i,i',j)|x_j^{ii'}\right]&\equiv\mathcal{E}\left[\chi(i,i',j)\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)\right]|x_j^{ii'}\right]\geq 0,\\ \mathcal{E}\left[M_3(i,i')|x_j^{ii'}\right]&\equiv\mathcal{E}\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash D_{i'})+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup(\mathcal{J}_i\cap\mathcal{D}_{i'}))|x_j^{ii'}\right]\geq 0,\end{aligned}\tag{18}$$

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 for a firm i, z_{ij} is the same for all loans j within a given category. More formally, denoting the category of loan j by $k(j)$,

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

A key modeling issue is how to define the categories, which must be sufficiently fine so that an issuer's signals can be plausibly be thought of as being the same for all loans within a category. Specifically, for each issuer i, I consider all loans in $\mathcal{D}_i$ that are by a particular originator as comprising a category. For example, I assume that if Bear Stearns issues a deal, Bear Stearns values all loans originated by the lender New Century according to their exogenous characteristics and a private signal that is common to all loans by New Century. This assumption also implies that Bear Stearns has the same private information about all of its own loan originations for a particular deal.[53] Essentially, my model implies that Bear Stearns' having better private information about its own loans would lead to more loans being securitized in-house, but does not look at differences in private information within the set of in-house loans. This assumption is admittedly an abstraction, but in the results section, I argue that the direction of the bias is unclear even if this assumption were violated.

Assumption 2 allows me to treat the choice over which loans from each category to include in a portfolio as being loosely similar to a choice over the number of such loans, with the difference being that I allow for loan-specific covariates. The *ordered choice* model (Ishii, 2005; Smith, 2011) 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).

Note that Assumptions 1 and 2 are non-nested: the former imposes symmetric information across *issuers*; the latter imposes symmetric information about *loans* originated by a given firm. Appendix C describes how I specify the observed component of the counterfactual utilities under Assumption 2 when the in-house status of loans switches between the actual and counterfactual cases, which I do somewhat differently from the specification under Assumption 1 so as to be

[53]For loans with two or more co-lead underwriters, the set of loans originated by each co-issuer comprises a separate category.

consistent with this interpretation. 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 $z_{i'j}$. On the other hand, by finding linear combinations of the inequalities such that each unique value of $z_{ik(j)}$ receives equal weight, we can construct moment conditions that are unbiased by selection over the unobservables.

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 $z_{ik(j)} - \zeta^{ii'}_{k(j)}$ is negative conditional on firm i holding zero loans from category k. Likewise, 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 $\mathcal{J}^{K(i)}_{i'}$ denote the set of loans in the portfolio of firm i' belonging to categories observed in firm i's portfolio. Thus, $\mathcal{J}_{i'} \backslash \mathcal{J}^{K(i)}_{i'}$ refers to the set of loans in the portfolio of firm i' belonging to categories not found in firm i's portfolio. Let $\tilde{\chi}(i,i',j)$ be an indicator function equal to 1 if $j \in \mathcal{J}_{i'} \backslash \mathcal{J}^{K(i)}_{i'}$ and 0 otherwise. The necessary conditions imply the following:

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

K_j is a random variable which is known to be sufficiently positive such that the inequalities $K_j + z_{ik(j)} - \zeta^{ii'}_{k(j)} \geq 0$ and $K_j + z_{i'k(j)} - \zeta^{ii'}_{k(j)} \geq 0$ hold.[54] The above expression contains a weighted sum over the necessary conditions for all loans matched to firm i along with an expression known to be

[54] Because the moment conditions take expectations over the inequalities, it is actually sufficient for K_j to merely satisify $\mathcal{E}\left[K_j + z_{ik(j)} - \zeta^{ii'}_{k(j)} \mid \tilde{\chi}(i,i',j) = 1\right] \geq 0$ and $\mathcal{E}\left[K_j + z_{i'k(j)} - \zeta^{ii'}_{k(j)} \mid \tilde{\chi}(i',i,j) = 1\right] \geq 0$. I set K_j to be 5 percent of the original loan balance of loan j. The more conservative the choice of K_j, the larger the identified set. I have found that my estimates are robust to larger choices of K_j.

positive for all loan categories held by firm i' but not i. Similarly, it contains a weighted sum over the necessary 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 unobservables $z_{ik(j)}$, $z_{i'k(j)}$, and $\zeta_{k(j)}^{ii'}$ enter unconditionally and therefore their expectation over i, i', and $j \in \{\mathcal{J}_i \cap D_{i'}\} \cup \{\mathcal{J}_{i'} \cap D_i\}$ is equal to zero. Therefore, the unobservables drop out of the expectation of inequality 19:

$$\begin{aligned}\mathcal{E}\left[M_1(i,i',j)|x_j^{ii'}\right] \quad \equiv \quad \mathcal{E}\Big[&\tfrac{\chi(i,i',j)}{n_{ik}}\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j) + \tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}}K_j + \\ &\tfrac{\chi(i',i,j)}{n_{i'k}}\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash j) + \tfrac{\tilde{\chi}(i',i,j)}{n_{ik}}K_j|x_j^{ii'}\Big] \quad \geq 0.\end{aligned} \tag{20}$$

As is the case under Assumption 1, we can also construct moment conditions based on switching multiple loans from one firm to another. Specifically, I consider random subsets of loans $\mathcal{J}_{1ii'}^{random}$ drawn from $\mathcal{J}_i \cap \mathcal{D}_{i'}$) and $\mathcal{J}_{2ii'}^{random}$ drawn from $\mathcal{J}_{i'}\backslash J_{i'}^{K(i)}$, where the probability of including a particular loan in $\mathcal{J}_{1ii'}^{random}$ or $\mathcal{J}_{2ii'}^{random}$ is chosen in a way such that each category k is weighted equally.[55]

$$\begin{gathered}\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash\mathcal{J}_{1ii'}^{random}) + \textstyle\sum_{j\in\mathcal{J}_{2ii'}}[K_j + z_{ik(j)} - \zeta_{k(j)}^{ii'}] + \\ \Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash J_{1i'i}^{random}) + \textstyle\sum_{j\in\mathcal{J}_{2i'i}}[K_j + z_{i'k(j)} - \zeta_{k(j)}^{ii'}] = \\ \Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash\mathcal{J}_{1ii'}^{random}) + \textstyle\sum_{j\in\mathcal{J}_{1ii'}}[z_{ik(j)} - \zeta_{k(j)}^{ii'}] + \textstyle\sum_{j\in\mathcal{J}_{2ii'}}[K_j + z_{ik(j)} - \zeta_{k(j)}^{ii'}] + \\ \Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash J_{1i'i}^{random}) + \textstyle\sum_{j\in\mathcal{J}_{1i'i}}[z_{i'k(j)} - \zeta_{k(j)}^{ii'}] + \textstyle\sum_{j\in\mathcal{J}_{2i'i}}[K_j + z_{i'k(j)} - \zeta_{k(j)}^{ii'}] \geq 0.\end{gathered} \tag{21}$$

By construction, the unobservables are weighted equally across loan categories and drop out in expectation, implying the following:

$$\mathcal{E}\left[M_1(i,i')|x_j^{ii'}\right] \equiv \mathcal{E}\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash\mathcal{J}_{1ii'}^{random}) + \Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\backslash\mathcal{J}_{1i'i}^{random}) + \sum_{j\in\mathcal{J}_{2ii'}\cup\mathcal{J}_{2i'i}} K_j \;\; |x_j^{ii'}\right] \geq 0, \tag{22}$$

By a symmetric argument, we can establish a second type of moment condition based on the optimality of each firm's observed portfolio relative to adding individual loans or sets of loans from the other firm:

[55] For all loans $j \in \mathcal{J}_i \cap \mathcal{D}_{i'}$) or $j \in \mathcal{J}_{i'}\backslash J_{i'}^{K(i)}$, I set the probability of j's inclusion in $J_{1ii'}^{random}$ or $J_{2ii'}^{random}$, respectively, to be n_{min}/n_{ik}. I define n_{min} as the number of elements in the category with the smallest number of loans included in $\mathcal{J}_i \cap \mathcal{D}_{i'}$ or $\mathcal{J}_{i'}\backslash J_{i'}^{K(i)}$. This probability ensures equal weighting for all categories and also maximizes the expected number of loans that are shifted in the counterfactual scenario.

$$\begin{aligned}
\mathcal{E}\left[M_2(i,i',j)|x_j^{ii'}\right] &\equiv \mathcal{E}\left[\tfrac{\chi(i',i,j)}{n_{ik}}\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\cup j)+\tfrac{\tilde{\chi}(i',i,j)}{n_{i'k}}K_j+\right. \\
&\qquad \left.\tfrac{\chi(i,i',j)}{n_{i'k}}\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)+\tfrac{\tilde{\chi}(i,i',j)}{n_{ik}}K_j|x_j^{ii'}\right] \geq 0, \\
\mathcal{E}\left[M_2(i,i')|x_j^{ii'}\right] &\equiv \mathcal{E}\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\cup \mathcal{J}_{1i'i}^{random})+\right. \\
&\qquad \left.\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup \mathcal{J}_{1ii'}^{random})+\textstyle\sum_{j\in\mathcal{J}_{2i'i}\cup\mathcal{J}_{2ii'}}K_j \;\; |x_j^{ii'}\right] \geq 0.
\end{aligned} \tag{23}$$

Finally, similar to the case for Assumption 1, we can construct a third type of moment inequality based on the total gains to trade:

$$\begin{gathered}
\tfrac{\chi(i,i',j)}{n_{ik}}[\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)]+\tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}}[L_j+z_{ik(j)}-z_{i'k(j)}]= \\
\tfrac{\chi(i,i',j)}{n_{ik}}[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)+z_{ik(j)}-z_{i'k(j)}]+\tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}}[L_j+z_{ik(j)}-z_{i'k(j)}]\geq 0, \\
\Delta\pi_i(\mathcal{J}_i,\mathcal{J}_i\backslash\mathcal{J}_{1ii'}^{random})+\Delta\pi_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup J_{1ii'}^{random})+\textstyle\sum_{j\in\mathcal{J}_{2ii'}}[L_j+z_{ik(j)}+z_{i'k(j)}]= \\
\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash\mathcal{J}_{1ii'}^{random})+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup J_{1ii'}^{random})+\textstyle\sum_{j\in\mathcal{J}_{1ii'}}[z_{ik(j)}+z_{i'k(j)}]+\sum_{j\in\mathcal{J}_{2ii'}}[L_j+z_{ik(j)}+z_{i'k(j)}]\geq 0.
\end{gathered} \tag{24}$$

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.[56] By construction, the expectation over $z_{ik(j)} - z_{i'k(j)}$ in the expression is zero, yielding the following expressions:

$$\begin{gathered}
\mathcal{E}\left[M_3(i,i',j)|x_j^{ii'}\right]\equiv\mathcal{E}\left[\tfrac{\chi(i,i',j)}{n_{ik}}[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash j)+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup j)]+\tfrac{\tilde{\chi}(i,i',j)}{n_{i'k}}L_j \;\; |x_j^{ii'}\right]\geq 0, \\
\mathcal{E}\left[M_3(i,i')|x_j^{ii'}\right]\equiv\mathcal{E}\left[\Delta r_i(\mathcal{J}_i,\mathcal{J}_i\backslash\mathcal{J}_{1ii'}^{random})+\Delta r_{i'}(\mathcal{J}_{i'},\mathcal{J}_{i'}\cup J_{1ii'}^{random})+\textstyle\sum_{j\in\mathcal{J}_{2ii'}}L_j \;\; |x_j^{ii'}\right]\geq 0.
\end{gathered} \tag{25}$$

Instruments

The instruments $x_j^{ii'}$ must be mean-independent of the unobservables z_{ij} and $\zeta_j^{ii'}$ and be in the information sets of firms i and i'. I include four types of instruments: (1) the exogenous characteristics of loan j;[57] (2) the mean exogenous 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

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

[57]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.

deals i and i'. The instruments that are not dummy variables must be transformed so that they are non-negative 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 $h(x_j^{ii'})$.

Let Θ denote the parameter space and θ an element therein. I construct the estimator using the empirical analogues to the moment conditions, summing over all possible combinations of firms i and i' that have feasible trades. I define the empirical moments as $m(\theta) = [m_1(\theta), m_2(\theta)]$, where

$$m_1(\theta) = \sum_i \sum_{i'} \sum_{j \in \{\mathcal{J}_i \cap D_{i'}\} \cup \{\mathcal{J}_{i'} \cap D_i\}} \begin{bmatrix} M_1(i,i',j)h(x) \\ M_2(i,i',j)h(x) \\ M_3(i,i',j)h(x) \end{bmatrix}, \quad m_2(\theta) = \sum_i \sum_{i'} \begin{bmatrix} M_1(i,i')h(x) \\ M_2(i,i')h(x) \\ M_3(i,i')h(x) \end{bmatrix}. \tag{26}$$

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 the vector γ, scalars c_f and α_0, β_t for all deal vintages t in the sample, and p_i for all deals 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 γ, c_f, and α_0 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 originated by competing firms. More negative correlation implies that loan j produces diversification benefits for the competitors, increasing their demand for j.

The necessary exclusion restrictions are that for all loans j and all deals i such that $j \in \mathcal{D}_i$, the set of competing firms that can feasibly hold j, the exogenous characteristics of loans originated by those firms, and the exogenous characteristics of loan j (w_{ij}) be mean-independent of the unobservables (z_{ij} and ζ_{ij}). In other words, these "instruments" do not shift the distribution of unobservables for the set of loans that could feasibly be matched with i, but only affect the propensity of loans to be securitized in-house. Given how I define z_{ij}, this restriction rules out correlation between the instruments and average unobserved public information about loans in D_i. For example, we cannot have correlation between the number of currently active CMBS issuers and the degree of "market optimism" about loans in general. Because I control for both year and month dummies in my first-stage model of loan-performance, this seems like a tenable assumption. Under Assumption 2, which allows for private information, this restriction also rules out correlation between the instruments and the distribution of firm i's idiosyncratic beliefs about loan category $k(j)$.

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', but there is more exogenous demand for j than 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 holding loan j'. Similarly, the parameters in γ for the non-constant terms are identified the correlations between the covariates and the degree of sensitivity to exogenous changes in demand.

Similar to the variable transfer payments, the fixed cost c_f is also identified by the sensitivity of a firm i's probability of holding loans in response to exogenous shifts in demand from potential trading partners. Whereas γ asks how much the probability of i holding an individual loan j decreases with respect to exogenous increases in competitors' demand for j, c_f asks how much probability of i holding any positive amount of loans from a particular set $D_{i'}$ changes with respect to exogenous increases in competitors' demand for loans in $D_{i'}$. The higher is c_f, the lower the threshold level of demand from competitors at which this amount jumps to zero.

The parameter α_0 captures the average quality of in-house loans relative to non–in-house loans that is not captured by exogenous covariates or by public information. It is identified by the sensitivity of i's propensity to hold loans that it originated itself relative to all other loans, in response to an exogenous shift in demand for loans from i's competitors. If private information has a big effect on payoffs (α_0 is negative and has a large magnitude) and i has an adverse (favorable) private signal about the loans it originated, it would hold few (many) of those loans both before and after an exogenous increase in demand for loans for i's competitors. On the other hand, if private information is unimportant (α_0 has a small magnitude), then the proportion of i's own loans that i would securitize in-house would decline rapidly with respect to an exogenous increase in demand from 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.

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 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 expected loan returns starting from the cutoff date. However, I include the loan seasoning at the cutoff date (cutoff date minus loan origination date) as a control.

For the random coefficients, I specify $\xi^r = \xi_1^r + v_{ij}\xi_2^r$—so ξ_2^r is latent unless $v_{ij} = 1$—and estimate the joint density of (ξ_1^r, ξ_2^r) over a grid.[58] 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.

The estimated distribution of ξ_j conditional on in-house status (v_{ij}) indicates that in-house loans perform worse on average (similar to the reduced-form estimates without random effects in Table 2) but have a lower variance in outcome, mainly due to there being a longer left-hand tail in the distribution of random coefficients for loans that are not in-house. The mean of ξ_j is .756

[58]I set the support for both ξ_1^r and ξ_2^r $\{-\infty, -5, -4, -3, -2, -1, [-.75 : .05 : .75], 1, 2, 3, 4, 5\}$ and set the grid to the Cartesian product of the supports. 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 are smaller in absolute magnitude.

conditional on $v_{ij} = 0$ and .916 conditional on $v_{ij} = 1$, almost exactly reproducing the 21-percent higher hazard of default for in-house loans implied by the reduced-form regression without random effects (Table 2). The variance is .611 or .049, respectively, conditional on $v_{ij} = 0$ or $v_{ij} = 1$.

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. To summarize, 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 B.5 and B.6 display the structural estimates based on Assumptions 1 and 2, respectively. I obtain point estimates in both cases, which is not surprising given the large number of inequality moments. For the main set of estimates, I exclude CMBS deals from the 2007 vintage. Spreads on CMBS bonds widened dramatically beginning in the middle of 2007, reflecting the sudden increase in market turmoil.[59] Excluding data from 2007 is a way to rule out the confounding influence of any change in the structural parameters after the onset of the crisis. Estimates including the 2007 vintage are similar and are reported in Appendix E.

I compute 95-percent confidence intervals using the "conservative" methodology described in Pakes et al. (2006).[60] There are 50,168 pairs of firms i and i' such that i's observed portfolio

[59]The spread on previously issued 10-year AAA-rated CMBS bonds was 23bp at the beginning of 2007, 28bp the first week of June, 57bp the first week of September, and 103bp the first week of December. (Data from JP Morgan Chase via the Federal Reserve Board of Governors.)

[60]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

contains one or more loans that could feasibly be held by i'.[61] Among these 50,168 pairs, there are more than 1.3 million feasible trades involving individual loans. I construct moment conditions based on the first firm in each pair selling a set of multiple loans to the second firm, as described in the previous section, resulting in 50,168 "observations". In addition, I construct moment conditions based on feasible trades involving individual loans. Because there are many more such potential individual-loan trades than are needed to obtain very precise estimates, I use a 10-percent random sample in order to reduce the computational time (resulting in 132,392 "observations").

The two sets of estimates are qualitatively similar. Beginning with the key parameter estimate of interest, $\hat{\alpha}_0$, both specifications indicate that an exogenous increase in the probability of a loan being securitized in-house is associated with a reduction in the loan's hazard of default, suggesting the presence of adverse selection. Under Assumption 1, $\hat{\alpha}_0 = -2.00$, implying a hazard ratio of .136. Under Assumption 2, $\hat{\alpha}_0 = -1.83$, implying a hazard ratio of .161. The magnitude of this effect on the hazard of default is comparable to that of one standard deviation change in the DSCR, occupancy rate, or LTV. The magnitude is about 5 times as big as that of one standard deviation change in loan seasoning, and about 1/5 as big as that of one standard deviation change in loan size.

Recall that the reduced-form relationships indicate that, all else equal, loans securitized in-house actually tend to perform worse. Consider the distribution of the residuals after we net out $\hat{\alpha}_0$ from the estimated distribution of random effects for in-house loans, $H(\xi_j|v_{ij} = 1)$ (see Table 2). The distribution of the residuals implies a hazard of default that for in-house loans is higher on average by a factor of 26.8 under the Assumption 1 and 22.7 under Assumption 2. This difference reflects the impact of non-random selection with respect to unobserved public information, suggesting that the mitigation of warehouse risk is an important incentive for issuers to securitize in-house loans for which low demand from competitors would otherwise force the them to keep the loans on their balance sheets.

Note that although the two alterative specifications are non-nested and there is no straightforward way to directly test the identifying assumptions, the fact that $\hat{\alpha}_0 < 0$ makes Assumption 2 (asymmetric information) more appealing than Assumption 1. The implication that in-house loans have better quality after controlling for unobserved public information is clearly mechanically possible under either assumption (since we are not solving for the equilibrium, and thus do not impose that privately observed quality must be the same for in-house and non–in-house loans). However, given equilibrium behavior, the actual estimate of α_0 suggests that Assumption 2 is more compelling

the structural estimates.

[61] Note that this relationship is not necessarily commutative.

on conceptual grounds.[62] Furthermore, the difference in $\hat{\alpha}_0$ between the two model specifications is consistent with the direction of the estimation bias that we would expect if the Assumption 1 is misspecified. To see why, suppose firm i has a more favorable signal than firm i' about the quality of loan j originated by i. Mistakently imposing equality between firm i and firm i's signals fails to control for correlation between the propensity for j to be securitized in-house and i's private signal.

The parameter on expected B-piece cashflows (the β_t parameters) is less than 1, for the early vintages (2000–2004), indicating that firms place greater value on expected cashflows from the AAA tranche. In other words, holding everything else equal, firms have a low revealed preference for volatility in the early years. In fact, the estimated values are 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.[63]

Under Assumption 2, β_t reaches a low of -2.2 in 2004, increases to 2.2 in 2005, then skyrockets to 11.1 just before the crisis in 2006. Estimates including data from the 2007 vintage (Appendix E) indicate an even higher value of β_t for that year. The increase in β_t after 2004 indicates an increase in the revealed preference for risk, reflecting 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 that would have reduced the volatility of the portfolios but were not in fact carried out.

Estimates of β_t under Assumption 1 are extremely similar but in general slightly higher. This difference between the specifications is driven by the fact that when we allow for asymmetric information under Assumption 2, firms' failure to carry out feasible, volatility-reducing trades can be explained by idiosyncratic preferences in addition to preference for volatility.

The transfer-payment parameters γ indicate the following. Payments tend to be higher on average for more seasoned loans. Also, payments tend to be higher for riskier loans as measured by DSCR and LTV, but lower with respect to the coupon spread on the loan (another indicator of risk). Because riskier loans generally have a higher coupon to compensate for the greater risk of default, the signs of the three coefficients are somewhat hard to interpret directly, but suggest a

[62] If we are willing to accept that moral hazard (hidden action) is an important factor, then a negative estimate of α_0 could also be reconciled with symmetric ex ante information about the loans.

[63] The model imposes linearity of utility with respect to the expected returns on each tranche. Higher volatility decreases the expected returns on the AAA tranche but increases the expected returns on the B piece. If increasing volatility reduces overall utility at a faster rate than linearly with respect to the expected AAA returns, the implied parameter on the B piece can be negative.

complex relationship among the three variables in shaping perceived risk and return. Finally, the signs of the two parameters on occupancy rate depend on the model specification.

Based on the large coefficients in $\hat{\gamma}$, the predicted transfer payments have higher variance across observations than we would expect for the actual transfer payments, suggesting a high degree of variance in the unobservables $\zeta_j^{ii'}$ that is not captured by the model. (It seems reasonable to expect the transfer payment per dollar of loan principal to be on the same order of magnitude as about a dollar.) Note that I could have included a larger set of covariates, but chose to include just a few key variables because the observable determinants of transfer payments are not a key focus of the paper.

The implied fixed cost of trading with a particular competing issuer, $\hat{c}_f$, is \$299 million under Assumption 1 and \$224 million under Assumption 2—about 17.2 percent or 12.8 percent of the average pool size—respectively. The greater importance of fixed costs under Assumption 1 is instructive.[64] Under Assumption 2, which allows for asymmetric information, the failure of trades to occur between potential trading partners can also be explained by unobserved, firm-specific preferences. By contrast, when we impose symmetric information, the model relies to a greater extent on fixed costs to rationalize the absence of certain trades.

I also estimated the model imposing zero fixed costs under Assumption 2. The estimates (6), which are in general extremely similar to the baseline estimates under Assumption 2, imply slightly higher variable costs (by 6.1 percent, on average) and slightly stronger adverse selection ($\hat{\alpha}_0 =$ -2.15). The difference in $\hat{\alpha}_0$ relative to the baseline makes sense, because when we rule out fixed costs, rationalizing the large number of loans that are securitized in-house requires a stronger advantage to doing so.

The probability of the AAA tranche experiencing losses (p_i) increases dramatically for later vintages of CMBS deals, on average. The mean probability is no higher than about 0.002 for the 2000–2003 vintages, compared with approximately 0.04, 0.3, and 0.75 for the 2004, 2005, and 2006 vintages, respectively.

Finally, I consider possible violations of Assumption 2 that could lead to estimation bias. The direction of the bias is far from clear. On the one hand, because the set of loans originated by each originator comprises a category, if there is unobserved within-category heterogeneity in *public* information, deal issuers' incentive to mitigate warehouse risk would tend to cause loans with lower unobserved quality to be securitized in-house (biasing α_0 upward), implying that our estimate

[64] The variable transfer payments are almost the same between the two specifications, on average.

of the degree of adverse selection is actually an underestimate. On the other hand, unobserved within-category heterogeneity in *private* signals about loans originated by the issuer would create an incentive for in-house securitization of loans that have higher unobserved quality, which leads to the opposite bias.

While there is no way to measure the degree of within-category heterogeneity in the unobservables, the following heuristic exercise is informative. Note that one potential driver of within-category heterogeneity in the unobservables is if the true joint distribution of loan returns exhibits dependencies not captured by my model's correlation structure. Assuming that market participants are aware of these dependencies, diversification incentives would tend to result in lower unobserved quality for loans that offer greater diversification benefits not captured by my model. While capturing every moment of the joint distribution is empirically infeasible, we can get an informal idea of how well the baseline specification captures the dependency structure by comparing it with an alternate specification in which loan returns are assumed to be independent.[65] The results (not reported) are extremely similar to the baseline results under Assumption 3 (with $\alpha_0 = -1.86$), which suggests that there is probably not a high degree of unobserved within-category heterogeneity on account of imperfect measurement of the dependency structure.

7 Conclusion

There has been little work quantifying the magnitudes of different incentive distortions in securitization markets, in large part because of the difficulty of separately identifying adverse selection from other confounding factors that affect the ex post performance of loans. In the case of CMBS, higher privately observed quality for loans securitized in-house—when CMBS issuer is the same firm as the loan originator—may be masked by selection based on information that is common knowledge to market participants but unobserved to the researcher. For instance, issuers have a strong incentive to securitize their own loans in-house when there is a negative demand shock for the loans among competing CMBS issuers. Furthermore, because issuers' decisions over asset bundles involve jointly choosing loans in a portfolio decision, we must control for the dependence structure of loans' returns. Because of these challenges, previous empirical research on CMBS has been limited to studying the performance of individual loans, without providing compelling evidence on selection effects.

[65] I accomplish this by zeroing out the covariance terms in the joint distribution estimated in the second stage of the model).

I estimate a model of CMBS pool formation 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 issuers, which is uncorrelated with unobserved public information about particular loans, allows me to identify the mean effect of adverse selection. In the most plausible specification of my model, I find that adverse selection accounts for loans that are securitized in-house having an 84-percent lower hazard of default. The magnitude of this effect on the hazard of default is comparable to that of several key loan underwriting variables, including debt-service coverage ratio (DSCR), the occupancy rate of the underlying property, and the loan-to-value (LTV) ratio. The fact that in-house loans actually perform worse overall, after controlling for observable characteristics, suggests that warehouse risk is an important competing incentive. That is, loan originators have an incentive to securitize their own loans—as opposed to keep them on their balance sheet—when negative public information (which is imperfectly observed by the econometrician) drives down overall demand for those loans from competing CMBS issuers.

The empirical importance of adverse selection in the CMBS market has potential implications for regulatory policy. For example, there is currently a policy debate over whether firms should be compelled to retain an interest in securitization deals backed by assets that they have originated and, if so, the precise form that the retention requirement should take. The finding that adverse selection is important implies that this policy is likely to have a material effect on the quality of loans that are securitized as well as the pricing of the securities, though the precise magnitude of the effect will depend on how the regulatory floor on the retained share changes the equilibrium.

Finally, this paper illustrates the usefulness of applying techniques for estimating partially identified games to 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.

Ashcraft, A. and T. Schuermann (2006): *Understanding the Securitization of Subprime Mortgage Credit*, Foundations and Trends in Finance, vol 2. no. 3, pp. 191-309.

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

Black, L., C. S. Chu, A. Cohen, and J. Nichols (2011): "Differences Across Originators in CMBS Loan Underwriting," Federal Reserve Finance and Economics Discussion Series (FEDS) 2011-05.

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

Cohen, A. (2011): "Rating Shopping in the CMBS Market," *Mimeo.*

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 (2011): "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.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and Their Application," *Mimeo*.

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

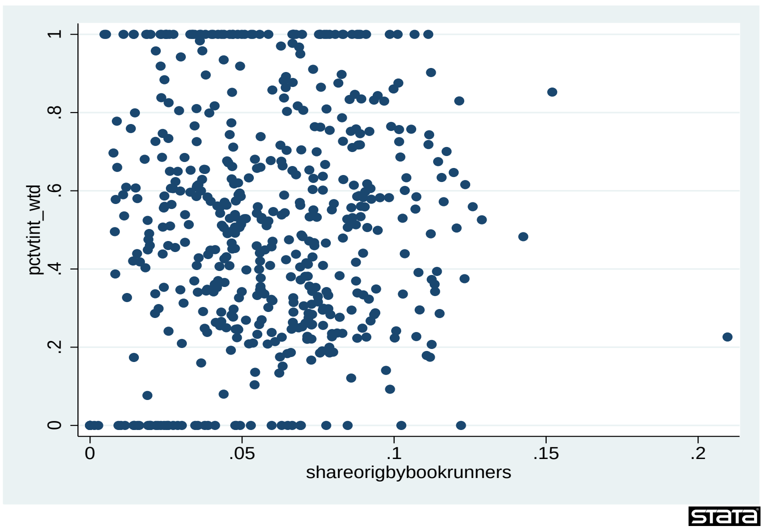
Smith, J. (2011): "The Effect of College Applications on Enrollment," *Mimeo*, Boston University.

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: Percentage of In-House Loans in Each Deal vs. Share of All Originations by Deal Issuer Around Time of Deal



Each point corresponds to a deal. The vertical axis plots the share of in-house loans in the deal, weighted by loan size. The horizontal axis plots the number of loans originated by the issuer 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	Mean	Std. Dev.	Loan-specific variables	Mean	Std. Dev.
cutoff date	2004.71	(2.15)	original loan amount ($M)	10.3	(15.9)
cutoff # loans	143.0	(66.0)	in-house	.541	
cutoff pool balance ($M)	1741.6	(1191.1)	DSCR at issuance	1.46	(0.40)
% loans in-house	.564		Occupancy at issuance	.94	(.08)
AAA tranche proportion	0.830	(0.066)	No occupancy data	0.223	
vintage = 2000	0.07		Original LTV	0.690	(0.113)
vintage = 2001	0.09		loan seasoning at cutoff	4.12	(9.43)
vintage = 2002	0.08		coupon spread	1.56	(0.72)
vintage = 2003	0.11		IO loan	0.103	
vintage = 2004	0.16		Fixed-rate mortgage	0.998	
vintage = 2005	0.16		vintage = 2000	0.053	
vintage = 2006	0.17		vintage = 2001	0.067	
vintage = 2007	0.16		vintage = 2002	0.061	
			vintage = 2003	0.089	
			vintage = 2004	0.121	
			vintage = 2005	0.179	
			vintage = 2006	0.204	
			vintage = 2007	0.172	
			Commercial-bank loan	0.429	
			Insurance co. loan	0.071	
			I-bank loan	0.214	
			Conduit loan	0.029	
			Finance co. loan	0.057	
			Foreign conduit loan	0.200	
			Ever default	0.139	
			Ever prepay	0.239	
			Age at default	35.1	
			Age prepayment	84.1	
			Age at default, prepayment or censoring date	53.5	
N	468		N	57,353	

Table 2: Reduced-form Hazard Regressions

	(I)		(II)	
in-house	1.21	(0.07)	1.44	(0.13)
% in-house in deal	-	-	1.50	(0.12)
(in-house)*(% in-house in deal)	-	-	0.65	(0.07)
DSCR at issuance	0.81	(0.05)	0.80	(0.05)
Occupancy at issuance	0.12	(0.02)	0.12	(0.02)
No occupancy data	0.14	(0.02)	0.14	(0.02)
Original LTV	32.99	(5.90)	31.93	(5.72)
coupon spread	1.51	(0.03)	1.51	(0.03)
original loan amount	2.37	(0.15)	2.36	(0.15)
IO loan	1.11	(0.05)	1.12	(0.05)
Fixed-rate mortgage	2.06	(0.58)	1.98	(0.56)
Vintage				
2000	0.60	(0.04)	0.59	(0.04)
2001	0.51	(0.03)	0.50	(0.03)
2002	0.31	(0.02)	0.30	(0.02)
2003	0.43	(0.03)	0.42	(0.03)
2004	0.75	(0.04)	0.73	(0.04)
2005	1.20	(0.07)	1.17	(0.06)
2006	1.66	(0.09)	1.61	(0.09)
2007	1.81	(0.11)	1.76	(0.10)
Originator fixed effects?	Yes		Yes	
Dummies for origination month?	Yes		Yes	
Region–property-type interactions?	Yes		Yes	
N	58,309		58,309	

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

	Loan characteristics (hazard ratio)	
Loan seasoning at cutoff	1.0704	(0.0014)
DSCR at issuance	0.7489	(0.0512)
Occupancy at issuance	0.2191	(0.0304)
No occupancy data	0.2401	(0.0319)
Original LTV	20.717	(3.964)
coupon spread	1.5528	(0.0383)
Original loan amount	2.3518	(0.1754)
IO loan	1.0614	(0.0477)
Fixed-rate mortgage	23.512	(7.221)
Insurance co. loan	0.7775	(0.0468)
I-bank loan	1.1351	(0.0365)
Domestic conduit loan	1.3986	(0.0928)
Finance co. loan	1.0032	(0.0591)
Foreign conduit loan	1.2761	(0.0411)
Vintage fixed effects?	Yes	
Dummies for origination month?	Yes	
Region–property-type interactions?	Yes	
	Square root of baseline hazard	
Constant	1.756e-05	(4.195e-06)
t (months)	-5.091e-06	(1.082e-06)
t^2	2.722e-07	(5.745e-08)
	θ: Density of (ξ_1^r, ξ_2^r), where $\xi_j = \xi_1^r + v_{ij}\xi_2^r$	
$(-4, 5)$	0.0159	(0.0085)
$(-1, 2)$	0.0420	(0.0406)
$(-0.55, -0.05)$	0.0237	(0.0130)
$(-0.10, 1.00)$	0.0398	(0.0634)
$(1.00, -0.05)$	0.8785	(0.0372)
N	57,353	

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 issuer 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)	-1.9974	[-1.9974	-1.9974]
Transfer payment parameters			
Constant	4.8659	[4.8659	4.8659]
Loan seasoning at cutoff	0.0377	[0.0377	0.0377]
DSCR at issuance	-5.7756	[-5.7756	-5.7756]
Occupancy at issuance	-3.7515	[-3.7515	-3.7515]
No occupancy data	-3.7716	[-3.7716	-3.7716]
Original LTV	9.0647	[9.0647	9.0647]
Coupon Spread	-1.4676	[-1.4676	-1.4676]
Fixed cost of trading			
c_f	298.9529	[298.9528	298.9532]
Utility of "B-piece" cashflows			
β_{2000}	-9.9485	[-9.9485	-9.9485]
β_{2001}	-0.8446	[-0.8446	-0.8446]
β_{2002}	-0.7103	[-0.7103	-0.7102]
β_{2003}	-1.2554	[-1.2554	-1.2553]
β_{2004}	-2.4392	[-2.4392	-2.4392]
β_{2005}	3.4335	[3.4335	3.4336]
β_{2006}	10.9375	[10.9375	10.9375]
Mean of p_i (Prob. of AAA tranche experiencing credit losses), by vintage			
2000 vintage	0.0038	[0.0038	0.0038]
2001 vintage	0.0039	[0.0039	0.0039]
2002 vintage	0.0040	[0.0040	0.0040]
2003 vintage	0.0056	[0.0056	0.0056]
2004 vintage	0.0425	[0.0425	0.0425]
2005 vintage	0.2932	[0.2932	0.2932]
2006 vintage	0.7469	[0.7469	0.7469]

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

	Estimate	Conservative 95% CI	
α_0 (in-house effect on hazard)	-1.8287	[-1.8287	-1.8287]
Transfer payment parameters			
Constant	-45.0393	[-45.0397	-45.0390]
Loan seasoning at cutoff	0.0292	[0.0292	0.0292]
DSCR at issuance	-2.1743	[-2.1743	-2.1742]
Occupancy at issuance	34.6973	[34.6970	34.6976]
No occupancy data	33.6297	[33.6294	33.6300]
Original LTV	19.0825	[19.0824	19.0825]
Coupon Spread	-1.0784	[-1.0784	-1.0783]
Fixed cost of trading			
c_f	223.7895	[223.7893	223.7923]
Utility of "B-piece" cashflows			
β_{2000}	-10.1967	[-10.1967	-10.1966]
β_{2001}	-0.7038	[-0.7039	-0.7037]
β_{2002}	-0.5619	[-0.5620	-0.5618]
β_{2003}	-0.9326	[-0.9327	-0.9325]
β_{2004}	-2.2396	[-2.2397	-2.2396]
β_{2005}	2.2375	[2.2374	2.2376]
β_{2006}	11.0857	[11.0855	11.0858]
Mean of p_i (Prob. of AAA tranche experiencing credit losses), by vintage			
2000 vintage	0.0038	[0.0038	0.0038]
2001 vintage	0.0039	[0.0039	0.0039]
2002 vintage	0.0040	[0.0040	0.0040]
2003 vintage	0.0060	[0.0060	0.0060]
2004 vintage	0.0484	[0.0484	0.0484]
2005 vintage	0.3083	[0.3083	0.3083]
2006 vintage	0.7606	[0.7606	0.7606]

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 20, 23, and 25 are set to 5 percent and 2.5 percent of the principal amount of loan j, respectively.

Table 6: Structural estimates: originator-specific private beliefs, no fixed costs of trading

	Estimate	Conservative 95% CI	
α_0 (in-house effect on hazard)	-2.1520	[-2.1520	-2.1520]
Transfer payment parameters			
Constant	-37.5179	[-37.5181	-37.5179]
Loan seasoning at cutoff	0.0276	[0.0276	0.0276]
DSCR at issuance	-3.6248	[-3.6248	-3.6247]
Occupancy at issuance	28.5740	[28.5739	28.5742]
No occupancy data	27.4379	[27.4378	27.4381]
Original LTV	19.7724	[19.7724	19.7724]
Coupon Spread	-1.2299	[-1.2299	-1.2299]
Utility of "B-piece" cashflows			
β_{2000}	-10.3722	[-10.3722	-10.3722]
β_{2001}	-0.8715	[-0.8716	-0.8715]
β_{2002}	-0.7682	[-0.7682	-0.7682]
β_{2003}	-2.4493	[-2.4493	-2.4492]
β_{2004}	-3.5355	[-3.5355	-3.5354]
β_{2005}	0.1639	[0.1638	0.1639]
β_{2006}	11.9921	[11.9920	11.9921]
Mean of p_i (Prob. of AAA tranche experiencing credit losses), by vintage			
2000 vintage	0.0038	[0.0038	0.0038]
2001 vintage	0.0038	[0.0038	0.0038]
2002 vintage	0.0040	[0.0040	0.0040]
2003 vintage	0.0055	[0.0055	0.0055]
2004 vintage	0.0375	[0.0375	0.0375]
2005 vintage	0.2807	[0.2807	0.2807]
2006 vintage	0.7324	[0.7324	0.7325]

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, and setting the fixed cost of trading (c_f) to zero. K_j and L_j for the "boundary" cases in equations 20, 23, and 25 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 $\varepsilon_i \equiv (\varepsilon_{i1} \ldots \varepsilon_{iK})$ and normal random variable η_{ij} from Equation 10. 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 $y_{ij} \equiv Pr(T_{ij} < t_{ij}|w_{ij}, v_{ij}) = \varepsilon_{ik(j)} + \eta_{ij}$, and define $y_j^c \equiv Pr(T_{ij} < t_j^c|w_{ij}, v_{ij})$. 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.[66]

Let $n = 1 \ldots N_i$ 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 $k = 1 \ldots K$, $j = 1 \ldots 2K$, and $n = 1 \ldots N$. 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.

[66] 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 $\{\eta_{nj}^{t-1}\}$:

$$\begin{gathered} K(\omega_\eta|\{\eta_{nj}^{t-1}\}_{n,j}) \text{ is } IG(v_0 + ND, \tfrac{v_0 s_0 + ND\bar{s}_\eta}{v_0 + ND}), \\ K(\Omega_\varepsilon|\{\varepsilon_n^{t-1}\}_n) \text{ is } IW(v_0 + N, \tfrac{v_0 S_0 + N\bar{S}_\varepsilon}{v_0 + D}), \text{ where} \\ \bar{s}_\eta = (1/ND)\textstyle\sum_{n,j}(\eta_{nj}^{t-1})^2 \quad , \quad \bar{S}_\varepsilon = (1/N)\sum_n \varepsilon_n^{t-1}\varepsilon_n^{t-1'}. \end{gathered} \tag{27}$$

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

$$\frac{L(y_n|\{\tilde{\eta}^t_{nj}\}_{n,j}, \{\tilde{\varepsilon}^t_n\}_n)\phi(\{\tilde{\varepsilon}^t_n\}_n|\Omega^t_\varepsilon)\phi(\{\tilde{\eta}^t_{nj}\}_{n,j}|\omega^t_j)}{L(y_n|\{\eta^{t-1}_{nj}\}_{n,j}, \{\varepsilon^{t-1}_n\}_n)\phi(\{\varepsilon^{t-1}_n\}_n|\Omega^t_\varepsilon)\phi(\{\eta^{t-1}_{nj}\}_{n,j}|\omega^t_j)} > \upsilon_n \tag{28}$$

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 10 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	0.0011	-0.0031	-0.0013	-0.0026	-0.0007	0.0004	-0.0026	0.0002	0.0009	-0.0010	-0.0009	0.0011	-0.0006	-0.0020
OTH*FAM	0.0011	0.5009	0.0047	-0.0000	0.0011	0.0028	-0.0004	-0.0015	-0.0027	-0.0008	0.0028	-0.0023	0.0026	-0.0007	-0.0021
OTH*COM	-0.0031	0.0047	0.5022	0.0008	-0.0008	0.0018	0.0007	-0.0002	-0.0019	0.0017	-0.0011	0.0024	0.0016	0.0011	-0.0006
MAT*OTH	-0.0013	-0.0000	0.0008	0.5009	0.0004	0.0014	0.0003	-0.0047	0.0018	0.0003	-0.0019	-0.0003	0.0004	0.0004	-0.0005
MAT*FAM	-0.0026	0.0011	-0.0008	0.0004	0.5060	0.0007	0.0013	-0.0001	-0.0006	0.0014	0.0014	-0.0034	0.0002	-0.0009	0.0002
MAT*COM	-0.0007	0.0028	0.0018	0.0014	0.0007	0.5037	0.0002	0.0013	0.0013	0.0006	-0.0007	0.0012	0.0026	-0.0008	-0.0011
MWE*OTH	0.0004	-0.0004	0.0007	0.0003	0.0013	0.0002	0.5012	-0.0004	-0.0005	-0.0017	-0.0005	-0.0023	-0.0001	0.0003	-0.0030
MWE*FAM	-0.0026	-0.0015	-0.0002	-0.0047	-0.0001	0.0013	-0.0004	0.5031	-0.0019	-0.0005	0.0015	-0.0006	-0.0025	-0.0008	-0.0003
MWE*COM	0.0002	-0.0027	-0.0019	0.0018	-0.0006	0.0013	-0.0005	-0.0019	0.5040	-0.0009	-0.0006	0.0014	0.0006	-0.0005	-0.0010
MWW*OTH	0.0009	-0.0008	0.0017	0.0003	0.0014	0.0006	-0.0017	-0.0005	-0.0009	0.5054	0.0029	0.0006	0.0008	-0.0003	-0.0007
MWW*FAM	-0.0010	0.0028	-0.0011	-0.0019	0.0014	-0.0007	-0.0005	0.0015	-0.0006	0.0029	0.5003	0.0002	-0.0002	-0.0008	0.0011
MWW*COM	-0.0009	-0.0023	0.0024	-0.0003	-0.0034	0.0012	-0.0023	-0.0006	0.0014	0.0006	0.0002	0.5077	-0.0003	-0.0003	-0.0031
NGL*OTH	0.0011	0.0026	0.0016	0.0004	0.0002	0.0026	-0.0001	-0.0025	0.0006	0.0008	-0.0002	-0.0003	0.5030	-0.0010	0.0003
NGL*FAM	-0.0006	-0.0007	0.0011	0.0004	-0.0009	-0.0008	0.0003	-0.0008	-0.0005	-0.0003	-0.0008	-0.0003	-0.0010	0.5085	-0.0015
NGL*COM	-0.0020	-0.0021	-0.0006	-0.0005	0.0002	-0.0011	-0.0030	-0.0003	-0.0010	-0.0007	0.0011	-0.0031	0.0003	-0.0015	0.5046
SAT*OTH	-0.0015	0.0028	-0.0013	0.0008	-0.0008	-0.0024	-0.0001	-0.0004	-0.0013	0.0008	0.0001	0.0017	0.0001	0.0009	0.0016
SAT*FAM	0.0030	-0.0025	-0.0008	-0.0010	0.0006	-0.0006	0.0002	-0.0009	-0.0013	0.0008	-0.0009	0.0002	0.0004	-0.0009	0.0019
SAT*COM	-0.0020	0.0001	0.0003	-0.0021	0.0001	-0.0002	-0.0006	0.0002	-0.0022	-0.0013	-0.0006	0.0002	0.0030	-0.0004	-0.0015
SEC*OTH	-0.0018	0.0008	0.0028	0.0010	0.0001	-0.0005	-0.0016	-0.0008	-0.0007	0.0023	-0.0021	0.0010	-0.0006	0.0028	0.0011
SEC*FAM	-0.0035	-0.0006	-0.0007	-0.0020	-0.0008	-0.0011	0.0002	-0.0020	-0.0030	-0.0012	-0.0016	0.0023	0.0007	0.0016	-0.0009
SEC*COM	-0.0027	0.0023	0.0018	-0.0001	0.0006	-0.0006	0.0010	0.0004	0.0003	-0.0025	-0.0012	-0.0024	-0.0010	0.0009	-0.0017
SWC*OTH	-0.0002	0.0012	-0.0020	0.0004	-0.0001	0.0007	0.0001	0.0010	0.0002	0.0008	-0.0015	-0.0009	0.0022	0.0010	0.0000
SWC*FAM	-0.0022	-0.0005	-0.0006	-0.0008	0.0001	-0.0002	0.0018	-0.0002	-0.0004	-0.0019	-0.0011	-0.0022	0.0008	-0.0001	0.0020
SWC*COM	-0.0004	0.0037	0.0011	-0.0017	0.0007	-0.0030	-0.0002	-0.0018	-0.0021	-0.0004	-0.0017	0.0008	-0.0002	-0.0019	-0.0008
MTN*OTH	0.0003	0.0025	0.0009	0.0006	0.0003	0.0005	0.0016	-0.0001	0.0024	-0.0008	0.0000	-0.0038	-0.0001	-0.0019	-0.0001
MTN*FAM	0.0024	0.0019	0.0003	0.0000	0.0006	0.0001	0.0006	-0.0006	-0.0017	-0.0004	0.0002	-0.0008	-0.0015	-0.0002	0.0008
MTN*COM	-0.0018	0.0002	0.0017	-0.0006	0.0007	-0.0014	-0.0008	-0.0024	0.0015	0.0020	0.0003	0.0002	0.0014	-0.0007	-0.0014
PNW*OTH	0.0004	0.0002	0.0006	0.0007	0.0013	-0.0004	0.0013	-0.0017	0.0026	-0.0020	-0.0018	0.0008	0.0007	-0.0007	0.0026
PNW*FAM	0.0008	-0.0011	-0.0011	0.0005	0.0021	-0.0027	-0.0026	-0.0008	-0.0004	0.0015	-0.0012	-0.0001	0.0002	0.0009	-0.0019
PNW*FAM	0.0006	0.0007	0.0002	-0.0010	0.0014	0.0001	0.0005	-0.0014	-0.0012	0.0004	-0.0013	0.0016	-0.0023	0.0018	-0.0020

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	-0.0015	0.0030	-0.0020	-0.0018	-0.0035	-0.0027	-0.0002	-0.0022	-0.0004	0.0003	0.0024	-0.0018	0.0004	0.0008	0.0006
OTH*FAM	0.0028	-0.0025	0.0001	0.0008	-0.0006	0.0023	0.0012	-0.0005	0.0037	0.0025	0.0019	0.0002	0.0002	-0.0011	0.0007
OTH*COM	-0.0013	-0.0008	0.0003	0.0028	-0.0007	0.0018	-0.0020	-0.0006	0.0011	0.0009	0.0003	0.0017	0.0006	-0.0011	0.0002
MAT*OTH	0.0008	-0.0010	-0.0021	0.0010	-0.0020	-0.0001	0.0004	-0.0008	-0.0017	0.0006	0.0000	-0.0006	0.0007	0.0005	-0.0010
MAT*FAM	-0.0008	0.0006	0.0001	0.0001	-0.0008	0.0006	-0.0001	0.0001	0.0007	0.0003	0.0006	0.0007	0.0013	0.0021	0.0014
MAT*COM	-0.0024	-0.0006	-0.0002	-0.0005	-0.0011	-0.0006	0.0007	-0.0002	-0.0030	0.0005	0.0001	-0.0014	-0.0004	-0.0027	0.0001
MWE*OTH	-0.0001	0.0002	-0.0006	-0.0016	0.0002	0.0010	0.0001	0.0018	-0.0002	0.0016	0.0006	-0.0008	0.0013	-0.0026	0.0005
MWE*FAM	-0.0004	-0.0009	0.0002	-0.0008	-0.0020	0.0004	0.0010	-0.0002	-0.0018	-0.0001	-0.0006	-0.0024	-0.0017	-0.0008	-0.0014
MWE*COM	-0.0013	-0.0013	-0.0022	-0.0007	-0.0030	0.0003	0.0002	-0.0004	-0.0021	0.0024	-0.0017	0.0015	0.0026	-0.0004	-0.0012
MWW*OTH	0.0008	0.0008	-0.0013	0.0023	-0.0012	-0.0025	0.0008	-0.0019	-0.0004	-0.0008	-0.0004	0.0020	-0.0020	0.0015	0.0004
MWW*FAM	0.0001	-0.0009	-0.0006	-0.0021	-0.0016	-0.0012	-0.0015	-0.0011	-0.0017	0.0000	0.0002	0.0003	-0.0018	-0.0012	-0.0013
MWW*COM	0.0017	0.0002	0.0002	0.0010	0.0023	-0.0024	-0.0009	-0.0022	0.0008	-0.0038	-0.0008	0.0002	0.0008	-0.0001	0.0016
NGL*OTH	0.0001	0.0004	0.0030	-0.0006	0.0007	-0.0010	0.0022	0.0008	-0.0002	-0.0001	-0.0015	0.0014	0.0007	0.0002	-0.0023
NGL*FAM	0.0009	-0.0009	-0.0004	0.0028	0.0016	0.0009	0.0010	-0.0001	-0.0019	-0.0019	-0.0002	-0.0007	-0.0007	0.0009	0.0018
NGL*COM	0.0016	0.0019	-0.0015	0.0011	-0.0009	-0.0017	0.0000	0.0020	-0.0008	-0.0001	0.0008	-0.0014	0.0026	-0.0019	-0.0020
SAT*OTH	0.5006	-0.0014	-0.0002	0.0004	0.0005	-0.0008	-0.0002	-0.0006	-0.0008	0.0019	0.0005	-0.0003	0.0010	-0.0007	-0.0003
SAT*FAM	-0.0014	0.5043	-0.0024	0.0019	-0.0014	-0.0004	0.0007	-0.0004	0.0004	0.0008	0.0000	0.0038	-0.0010	0.0008	-0.0010
SAT*COM	-0.0002	-0.0024	0.5008	0.0002	-0.0024	0.0003	0.0031	0.0003	0.0009	-0.0035	0.0020	0.0001	-0.0008	-0.0006	0.0005
SEC*OTH	0.0004	0.0019	0.0002	0.5010	0.0010	0.0009	-0.0006	-0.0009	-0.0005	0.0021	0.0017	-0.0007	-0.0008	0.0042	0.0012
SEC*FAM	0.0005	-0.0014	-0.0024	0.0010	0.5047	0.0007	-0.0014	-0.0037	-0.0008	0.0007	0.0011	-0.0004	0.0001	-0.0034	0.0011
SEC*COM	-0.0008	-0.0004	0.0003	0.0009	0.0007	0.5043	0.0020	-0.0010	0.0004	-0.0013	0.0005	0.0018	-0.0016	0.0004	-0.0015
SWC*OTH	-0.0002	0.0007	0.0031	-0.0006	-0.0014	0.0020	0.5040	0.0011	-0.0004	0.0021	-0.0005	-0.0017	-0.0016	0.0010	0.0003
SWC*FAM	-0.0006	-0.0004	0.0003	-0.0009	-0.0037	-0.0010	0.0011	0.5010	0.0001	-0.0016	0.0005	-0.0014	-0.0017	-0.0002	-0.0003
SWC*COM	-0.0008	0.0004	0.0009	-0.0005	-0.0008	0.0004	-0.0004	0.0001	0.5020	0.0010	0.0008	-0.0007	0.0019	-0.0002	-0.0020
MTN*OTH	0.0019	0.0008	-0.0035	0.0021	0.0007	-0.0013	0.0021	-0.0016	0.0010	0.5018	0.0008	0.0006	0.0008	0.0027	-0.0006
MTN*FAM	0.0005	0.0000	0.0020	0.0017	0.0011	0.0005	-0.0005	0.0005	0.0008	0.0008	0.5024	0.0000	-0.0005	0.0003	0.0023
MTN*COM	-0.0003	0.0038	0.0001	-0.0007	-0.0004	0.0018	-0.0017	-0.0014	-0.0007	0.0006	0.0000	0.5040	0.0013	0.0016	0.0006
PNW*OTH	0.0010	-0.0010	-0.0008	-0.0008	0.0001	-0.0016	-0.0016	-0.0017	0.0019	0.0008	-0.0005	0.0013	0.5028	0.0018	0.0000
PNW*FAM	-0.0007	0.0008	-0.0006	0.0042	-0.0034	0.0004	0.0010	-0.0002	-0.0002	0.0027	0.0003	0.0016	0.0018	0.5019	0.0019
PNW*COM	-0.0003	-0.0010	0.0005	0.0012	0.0011	-0.0015	0.0003	-0.0003	-0.0020	-0.0006	0.0023	0.0006	0.0000	0.0019	0.5064
Idiosyncratic factor (ω_η)	0.6295														

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	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013
OTH*FAM	0.0013	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
OTH*COM	0.0013	0.0013	0.0019	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0014	0.0013
MAT*OTH	0.0013	0.0013	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
MAT*FAM	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013
MAT*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	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	0.0018	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	0.0013	0.0013	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	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MWW*OTH	0.0013	0.0013	0.0014	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013
MWW*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013
MWW*COM	0.0014	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	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	0.0013	0.0019	0.0014	0.0013
NGL*FAM	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	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	0.0019
SAT*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	0.0013	0.0013	0.0013
SAT*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
SAT*COM	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	0.0013
SEC*OTH	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	0.0013
SEC*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	0.0014	0.0013	0.0013
SEC*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
SWC*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	0.0013	0.0014	0.0013
SWC*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
SWC*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
MTN*OTH	0.0013	0.0014	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
MTN*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
MTN*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
PNW*OTH	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
PNW*FAM	0.0013	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
PNW*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	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	0.0013	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	0.0013
OTH*FAM	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	0.0014	0.0013
OTH*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
MAT*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	0.0013	0.0013	0.0013
MAT*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
MAT*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	0.0014	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	0.0013	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	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
MWW*OTH	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	0.0013	0.0013
MWW*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
MWW*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	0.0014
NGL*OTH	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	0.0013	0.0013
NGL*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
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	0.0013
SAT*OTH	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	0.0013
SAT*FAM	0.0013	0.0019	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
SAT*COM	0.0013	0.0013	0.0018	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013
SEC*OTH	0.0013	0.0013	0.0013	0.0018	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SEC*FAM	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SEC*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
SWC*OTH	0.0013	0.0014	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014
SWC*FAM	0.0013	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
SWC*COM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0018	0.0014	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	0.0014	0.0018	0.0013	0.0013	0.0014	0.0014	0.0013
MTN*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0019	0.0013	0.0013	0.0013	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	0.0018	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	0.0014	0.0013	0.0013	0.0019	0.0013	0.0013
PNW*FAM	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0014	0.0013	0.0013	0.0013	0.0018	0.0013
PNW*COM	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	0.0019
Idiosyncratic factor (ω_η)	0.0003														

Appendix C: Technical Detail on Construction of Counterfactual Utilities

For counterfactual cases, the "observed" component of a firm's utility is constructed slightly differently under Assumptions 1 and 2. Suppose that we observe loan j in the portfolio of firm i, and wish to construct the observed utility for a different firm i' for a counterfactual case in which the counterfactual portfolio of firm i', $J'_{i'}$, contains loan j. $U_{i'}(J'_{i'})$ denotes the observed utility of firm i' under this counterfactual case, and v_{ij} and $v_{i'j}$ refer to loan j's in-house status in the actual and counterfactual scenarios, respectively.

Under Assumption 1, I generate $U_{i'}(\mathcal{J}'_{i'})$ setting j's default hazard to $\alpha_0 v_{ij} + \hat{\alpha}'_1 w_{ij}$—that is, to the expected hazard given j's *actual* in-house status, v_{ij}. With this construction, firms i and i' derive approximately the same observed utility from loan j in the actual and counterfactual scenarios, respectively. (The equality is only approximate, because correlation in returns across loans implies that loan j may offer differing amounts of diversification benefits to two different portfolios.) Combined with the maintained assumption about the unobservables, $z_{ij} = z_{i'j}$, this construction therefore implies that firms i and i' value loan j approximately the same in the actual and counterfactual scenarios, respectively. Thus, we can correctly interpret Assumption 1 as implying that the firms have symmetric information.

Under Assumption 2, I generate $U_{i'}(\mathcal{J}'_{i'})$ setting j's default hazard to $v_{i'j} \cdot \alpha_0 + \hat{\alpha}'_1 w_{ij}$—that is, to the expected hazard given j's *counterfactual* in-house status, $v_{i'j}$. Suppose that the actual portfolio of firm i' contains a loan j' in the same category as loan j. With this construction, firm i' derives approximately the same observed utility from loan j in the counterfactual scenario as it does from loan j' in the actual scenario. (Again, the equality is only approximate, because correlation in returns across loans implies that loans j and j' may offer differing amounts of diversification benefits to the portfolio.) Combined with the maintained assumption about the unobservables, $z_{i'j} = z_{i'j'}$, this construction therefore implies that firm i' values loans j and j' approximately the same after controlling for their exogenous characteristics, $w_{i'j}$ and $w_{i'j'}$. Thus, we can correctly interpret Assumption 2 as implying that after controlling for exogenous loan characteristics, unobserved values to firm i' of all loans within a category are the same.

Appendix D: 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 10.
- 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 $p^{(s)} = Pr(T_{ij} < t_j^{(s)})$, where $Pr(T_{ij} < t_j^{(s)})$ 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 4.

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 $p = Pr(T_{ij} < t_j)$, 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 $(v_{ij}\alpha_0 + \alpha_1' w_{ij})$ 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.

Appendix E: Estimates Including 2007 Data

Table B.5: Structural estimates: imposing symmetric information

	Estimate	Conservative 95% CI	
α_0 (in-house effect on hazard)	-2.3722	[-2.3722	-2.3722]
Transfer payment parameters			
Constant	0.9654	[0.9654	0.9654]
Loan seasoning at cutoff	0.0448	[0.0448	0.0448]
DSCR at issuance	-1.3703	[-1.3703	-1.3702]
Occupancy at issuance	-4.2840	[-4.2840	-4.2840]
No occupancy data	-4.1840	[-4.1840	-4.1840]
Original LTV	6.4596	[6.4596	6.4596]
Coupon Spread	-0.9731	[-0.9731	-0.9731]
Fixed cost of trading			
c_f	253.3499	[253.3499	253.3502]
Utility of "B-piece" cashflows			
β_{2000}	-10.0030	[-10.0030	-10.0029]
β_{2001}	-0.4056	[-0.4057	-0.4056]
β_{2002}	-0.4040	[-0.4041	-0.4040]
β_{2003}	-0.4136	[-0.4136	-0.4136]
β_{2004}	-2.0096	[-2.0097	-2.0096]
β_{2005}	1.9881	[1.9881	1.9881]
β_{2006}	6.6410	[6.6410	6.6410]
β_{2007}	32.7186	[32.7185	32.7187]
Mean of p_i (Prob. of AAA tranche experiencing credit losses), by vintage			
2000 vintage	0.0038	[0.0038	0.0038]
2001 vintage	0.0038	[0.0038	0.0038]
2002 vintage	0.0040	[0.0040	0.0040]
2003 vintage	0.0054	[0.0054	0.0054]
2004 vintage	0.0325	[0.0325	0.0325]
2005 vintage	0.2607	[0.2607	0.2607]
2006 vintage	0.7117	[0.7116	0.7117]
2007 vintage	0.9360	[0.9360	0.9360]

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

Table B.6: Structural estimates: originator-specific private beliefs

	Estimate	Conservative 95% CI	
α_0 (in-house effect on hazard)	-1.7835	[-1.7835	-1.7835]
Transfer payment parameters			
Constant	-35.6436	[-35.6437	-35.6434]
Loan seasoning at cutoff	0.0399	[0.0399	0.0399]
DSCR at issuance	-2.4487	[-2.4487	-2.4487]
Occupancy at issuance	32.8961	[32.8958	32.8962]
No occupancy data	31.5831	[31.5829	31.5832]
Original LTV	8.6592	[8.6592	8.6593]
Coupon Spread	-0.5034	[-0.5034	-0.5034]
Fixed cost of trading			
c_f	106.5110	[106.5109	106.5116]
Utility of "B-piece" cashflows			
β_{2000}	-5.0475	[-5.0475	-5.0474]
β_{2001}	-0.4264	[-0.4264	-0.4264]
β_{2002}	-1.1076	[-1.1077	-1.1076]
β_{2003}	-1.3909	[-1.3909	-1.3909]
β_{2004}	-3.0158	[-3.0158	-3.0158]
β_{2005}	1.9296	[1.9296	1.9297]
β_{2006}	5.5954	[5.5954	5.5955]
β_{2007}	59.2245	[59.2243	59.2249]
Mean of p_i (Prob. of AAA tranche experiencing credit losses), by vintage			
2000 vintage	0.0038	[0.0038	0.0038]
2001 vintage	0.0039	[0.0039	0.0039]
2002 vintage	0.0040	[0.0040	0.0040]
2003 vintage	0.0060	[0.0060	0.0060]
2004 vintage	0.0502	[0.0502	0.0502]
2005 vintage	0.3113	[0.3113	0.3113]
2006 vintage	0.7635	[0.7635	0.7635]
2007 vintage	0.9655	[0.9655	0.9655]

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 20, 23, and 25 are set to 5 percent and 2.5 percent of the principal amount of loan j, respectively.