2014

Essays on Mortgage Securitizations

Hong Lee
Louisiana State University and Agricultural and Mechanical College

ESSAYS ON MORTGAGE SECURITIZATIONS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Interdepartmental Program in Business Administration

(Finance)

by
Hong Lee
B.A., Korea University, 2004
M.A., Brown University, 2007
August 2014

ACKNOWLEDGMENTS

First of all, I would like to thank my dissertation advisor, Prof. Joseph Mason. I could never develop this research work without his kind guidance and unconditional support for data sets, software programs, and travels to conferences. I am greatly grateful to Prof. Carter Hill, Prof. Kelley Pace, and Prof. Rajesh Narayanan in my dissertation committee for their invaluable comments and suggestions. I owed a huge debt to Prof. Carlos Slawson who unsparingly allocated departmental resources for my research. I also want to express my gratitude to Prof. Ji-Chai Lin, Prof. Don Chance, Prof. Wei-Ling Song, and Prof. Shan He who helped me to set up and achieve my career goals.

I dedicate this dissertation to my family. My parents raised me into a person who appreciates the pleasure of learning. My lovely wife, Clara, always identified and brought out the best in me. My son, William, made every moment in the last six months at LSU sparkling.

Finally, I am thankful to my colleagues Shuang Zhu and Meagan McCollum for their continuous encouragement and reviews of my works.

TABLE OF CONTENTS

ABSTRACT

This dissertation studies how the affiliation between different entities in securitization process make different market outcomes, and how this estimation of affiliation effects is susceptible to limitations in securitized residential mortgage data. Three essays constitute the dissertation.

The first essay illustrates the limitations and potential bias in the loan-level trustee data. Substantial amounts of loan attributes and risk factors are missing. The patterns of data omissions dramatically vary across different risk factors, sponsors, and trustees, and over the time. Missing of one risk factor is in general positively correlated with missing of the other. Omissions of loan attributes are systematically associated with intermediate level of ex-post default risk. These findings suggest that if any data is sliced and diced based on the availability of loan attributes, the sample for default regression model may not be random.

The second essay examines how default risk is associated with the affiliation between the loan provider and the sponsor. The identity of loan provider is, however, selectively disclosed for riskier mortgages. Without consideration of this selective disclosure, the affiliation is seemingly linked to higher ex-ante and ex-post default risk. In contrast, if the affiliation is correctly calculated by backfilling loan provider identity, or if the sample selection problem is explicitly addressed, then loan providers cherry-pick mortgages with better ex-ante risk characteristics for their affiliated securitizations. Also, with more complete sample where missing and erroneous loan provider identities are backfilled and corrected, the affiliation between the loan provider and the sponsor is shown to significantly decrease the likelihood of default.

The third essay examines why sponsors are concerned about the performance of mortgages even after they are securitized and sold to investors in the form of bonds. Without any empirical tests, previous studies assume that sponsors have “skin in the game” because they retain the certificates

backed by the residual tranches. However, I show that sponsors with their own servicing platform increase their servicing quality even after the most junior tranche has dried up. This result implies reputational concerns may make sponsors care about performance of securitized mortgages.

CHAPTER 1. LOAN-LEVEL DISCLOSURE OF RISK FACTORS IN RESIDENTIAL MORTGAGE SECURITIZATIONS

1.1 Introduction

Loan-level analysis has become important more and more in the examination of mortgage industry.[1] Academics usually have access to the loan-level data in the web sites of trustees who serve as analytics providers for the investors in mortgage-backed securities (MBS).[2] There are two tiers of loan-level data. The first is a group of loan underwriting characteristics sometimes called loan "attributes," loan "characteristics," or the "loan tape." That data set consists of a variety of factors used for loan credit quality assessments such as mortgagor credit scores, the amount of home equity, the interest rate, and other items at the time of origination, reported for 1st and 2nd lien closed end loans and home equity lines of credits (HELOCs). Those attributes are originally produced by the originator.[3] The originator transfers its loan-level information to the servicer who makes monthly collections of principals and interests, and manages defaults. The sponsor[4] and underwriters[5] of MBS use the characteristics of collateralized mortgages as the basis for representations made in the prospectus supplements and other offering documents. Through the trustee's website, investors are informed of the attributes of mortgages back their mortgage securities. I explore the limitations and potential bias arising from reporting practice of these attributes in the paper.

The second tier is the "remittance" data. Unlike cross-sectional loan attributes, the remittance data has a panel structure. The servicer tracks and generates time-varying information for a mortgage

---

[1] The number of papers found with the keyword of "loan-level" in Google Scholar has monotonically increased from 134 in 2006 to 495 in 2013.
[2] The trustee owns the collateralized mortgages for the benefit of MBS investors. See American Bankers Association White Paper (2010) for the basic duties of the trustee.
[3] The originator is defined as a lending institution that extends mortgage credit to borrowers.
[4] Sponsors organize and initiate MBS transactions by purchasing and pooling mortgages to back the certificates issued by their mortgage trusts (SPVs).
[5] Underwriters of MBS are also called lead- and co-managers.

including monthly payments, outstanding balance, adjustment of interest rates, delinquency and payoff status. This remittance data is also posted and available from the trustee's website.

Many academic studies have used the trustee data in recent research. In their default or prepayment models, the response variable constructed with the remittance data is explained by a variety of loan characteristics. Mian and Sufi (2009) was one of the first to emphasize, generally, the importance of micro-level data to explore the origin of the subprime mortgage crisis. Keys, Mukherjee, Seru, and Vig (2010a), Demyanyk and Van Hermert (2011) and Mayer, Piskorski, and Tchistyi (2013) use LoanPerformance from First American CoreLogic[6]. Agarwal, Chang, and Yavas (2012), Bubb and Kaufman (2009), and Keys, Seru, Mukherjee, and Vig (2010b), Keys, Seru, and Vig (2012), Piskorski, Seru, and Vig (2010), and Ghent and Kudlyak (2011) use Lender Processing Services[7] (LPS or formerly known as McDash Analytics). Demiroglu and James (2012) [8] use ABSnet Loan from Lewtan[9]. Piskorski, Seru, and Witkin (2013) and the present paper and use BBx from BlackBox Logic LLC.[10]

Two basic assumptions imposed in all the studies cited above are that their loan-level data is correct, and that the samples they use appropriately represent the characteristics of population. As academic research has progressed, however, other authors have explored the possibility that a naïve use of loan-level data may be misleading. For instance, Piskorski, Seru, and Witkin (2013) present the evidence that lien and occupancy types have been misreported for substantial portion of private-label mortgages. Moreover, any omissions of data fields for certain loans may not be random,

[6] See http://www.corelogic.com/solutions/loan-performance-secondary-market-analytics-for-capital-markets.aspx#home-Datasets.
[7] See http://www.lpsvcs.com/Pages/default.aspx.
[8] Demiroglu and James (2012) note that risk may not be correctly measured in the deal-level analysis, but do not analyze the potential for such data shortcomings.
[9] See http://www.lewtan.com/products/absnet.html.
[10] See http://www.bbxlogic.com/bbx-logic-US-RMBS-non-agency-solutions.php.

leading to a sample selection bias. It is the second limitation in the loan-level data that I deal with in the paper.

It is natural to be concerned about non-random samples in micro-level mortgage analyses, since every loan-level information is voluntarily disclosed with little or no regulatory, accounting, or legal guidance. While the U.S. Securities and Exchange Commission (SEC) adopted Regulation AB (Reg AB) in 2005 in order to govern generally the disclosure regarding the securitization of assets including residential mortgages, Reg AB provides specific guidance only for deal level disclosures.

The financial crisis in the late 2000's was the opportunity to question the validity of the current reporting regime. In April 2010, the SEC released Proposed Rules (33-9117) to revise the existing Reg AB (colloquially referred to as "Reg AB II"). As summarized by Cadwalader[11] at that time, "To ensure that investors receive sufficient information to evaluate an investment in ABS, the Proposed Rules require ABS issuers to disclose granular asset-level data (or with respect to credit card receivables, grouped account data) relating to the terms, obligor characteristics, and underwriting of each asset backing an ABS (or group of assets, as applicable.[12] *In proposing rules for standardized asset-level disclosure in ABS transactions, the SEC is embracing an approach that it declined to take in the original adoption of* Reg *AB, where it concluded that it would not be 'practical or effective to draft detailed disclosure guides for each asset type that may be securitized.'*[13]

[11] "SEC Proposes Significant Enhancements to Regulation of Asset-Backed Securities," Apr 20, 2010, at http://www.cadwalader.com/resources/clients-friends-memos/sec-proposes-significant-enhancements-to-regulation-of-asset-backed-securities

[12] See Proposed Rule 17 C.F.R. § 229.1111(h) at 383. The SEC is proposing to exempt ABS backed by stranded costs from the obligation to provided asset-level data. Stranded costs are certain capital costs incurred by public utilities which are permitted, by action of a state legislature or other regulatory authority, to be recouped over time from rate payers.

[13] See page 1509 of the adopting release for Regulation AB (Release Nos. 33-8518 and 34-50905).

Under the Proposed Rules for Reg AB II, 28 items must be disclosed at the loan level regardless of asset types, and 137 additional items are required to be disclosed for residential mortgage loans. Nonetheless, the proposed regulation is currently in its third comment period, due in part to concerns about how much individual loan data can be disclosed without violating consumer privacy laws. Thus, loan-level attribute data remains voluntarily disclosed.

Absence of formal reporting requirement for loan-level information implies that the source data used commonly in academic studies and cited in policy development may be affected by systematic biases from missing loan-level data. This paper is designed to make three unique contributions to the literature of loan-level analysis. First, I provide an extensive illustration of the disclosure practices of loan-level information across different major risk factors, different time, and different institutions engaged in the securitization process. I show data fields are often missing for a substantial amount of loans[14] although loan-level information has been available generally more and more over time. Also, the disclosure rates for major risk factors are shown to dramatically vary across different sponsors and trustees.[15] These findings suggest the missing of loan-level data fields is not trivial at all, and we may need to carefully control for the variation of omissions along a variety of dimensions.

Second, I examine how the missing of one risk factor affects the omission of the other. I show loan attributes generally tend to be reported together and missing together, suggesting that attrition of loan-level information may be more severe for a particular group of mortgages that were poorly handled in the securitization process.

[14] For example, the information for negative amortization is missing for approximately 90% of the mortgages in BBx.
[15] According to BBx, Washington Mutual provided borrowers' credit scores only for 40.9% of mortgages for which it served as the trustee while Wells Fargo provided for 96.1%.

Third, to the best of my knowledge, this is the first paper that explores the relation between loan-level disclosures (or lack thereof) and the performance of securitized mortgages. I find evidence that omissions of loan attributes are systematically linked to intermediate level of *ex post* default risk. For example, mortgages whose FICO credit scores are missing default more than those with FICO scores higher than 680 by 5.75% to 8.02% depending on model specifications. The increase in defaults associated with omissions of FICO is economically and statistically significant, however it is smaller than the effects of FICO less than 680 (11.08% to 20.44%). This suggests the possibility that credit scores are less likely to be reported for the mortgages with intermediate level of FICO scores (approximately with the mean of 680). This may imply the convexity of the curve on the coordinates whose X- and Y-axes represent FICO scores and the corresponding disclosure rate respectively. This is an arguably reasonable disclosure pattern for two reasons: 1) sponsors may be more willing to convey the FICO score if it is high; 2) mortgages for borrowers with low FICO scores may have riskier other loan attributes than those with higher FICO scores, and MBS investors would demand the disclosure of FICO more aggressively if other loan attributes indicate higher risk. Failure to account for this relation between missing data and default risk can, therefore, lead to biased inferences in loan-level analyses and poor policy recommendations.

The rest of the paper is structured as follows. Section 2 introduces the loan characteristic data reported by vendors and how that has been implemented in key academic studies. Section 3 shows which data is commonly missing and how missing data may be related to certain securitization sponsors and trustees. Section 4 illustrates patterns among missing data fields, particularly how missing data fields for one characteristic are commonly related to missing data fields for others. Section 5 explores empirically the relationship between disclosure and *ex post* default risk. The last section concludes.

1.2 Loan characteristics disclosed in securitization data

Prior research relies crucially on several key data fields included in the trustee data. Table 1.1 presents the loan characteristics relied upon in 11 key papers in the literature.[16] The LPS data was used in six of the papers including Agarwal, Chang, and Yavas (2012), Bubb and Kaufman (2009), Keys, Seru, Mukherjee, and Vig (2010b), Keys, Seru, and Vig (2012), Piskorski, Seru, and Vig (2010), and Ghent and Kudlyak (2011). Keys, Mukherjee, Seru, and Vig (2010a), Demyanyk and Van Hermert (2011) and Mayer, Piskorski, and Tchistyi (2013) use LoanPerformance. Demiroglu and James (2012) and Piskorski, Seru, and Witkin (2013) use ABSnet Loan and BBx respectively.

The LPS data is completely different from ABSNet, BBx, and LoanPerformance in that it has mortgages that are both securitized and retained on banks' balance sheets while the other three datasets contain only securitized mortgages. ABSNet, BBx, and LoanPerformance all retrieve the data from the same trustees: the base data is, therefore, all the same.

Mason, Imerman, and Lee (2014) identify 22 loan attributes used for loan-level estimations in the 11 key papers. The studies used 11 data fields, on average, ranging from Keys, Mukherjee, Seru, and Vig (2013) with four fields to Keys, Mukherjee, Seru, and Vig (2010) with fifteen fields. Every paper uses Loan Type to screen first mortgages from second liens of various types. FICO credit score is the most frequently used data field in the loan-level analysis, followed by Loan Amount, Doc Type, Loan Purpose, and loan-to-value (LTV) ratio. Indicators associated with Adjustable-Rate Mortgage (ARM) Margin, Investment Bank Underwriter, and Negative Amortization, and Loan Originator are the least used.

[16] The 11 key papers meet three requirements: 1) It should be published in one of the qualified finance journals (Journal of Finance, Journal of Financial Economics, Review of Financial Studies, and Journal of Financial and Qualitative Analysis); 2) It should rely on loan-level analysis; and 3) It should use one of the loan-level data including ABSNet, BBx, LoanPerformance, and LPS.

I use the trustee data covered by BlackBox Logic (BBx) to examine how the loan-level disclosure practices vary across different major loan attributes during the pre-crisis period. As of March 2014, BBx contains data on 21,898,192 closed-end loans and HELOCs with original balances of more than $4.8 trillion.

BBx reports loan characteristics in the name of "Loans_Chars." This consists of 189 variables that fall into 36 categories. Blackbox collects information for 127 fields from trustees, which are standardized into 61 variables.11 This paper does not cover variables: 1) if they are associated with post-securitization characteristics rather than with loan underwriting; 2) if the coverage for a cleansed variable is substantially larger than a raw variable (hence the cleansed variable does not properly represent the reporting practices by lenders or securitizers); or 3) if they are rarely employed in loan-level analysis.

As a result of this selection rule, I focus on 16 variables that are "Cleansed and Standardized Across Data Providers." They are categorical variables related to loan types and numerical attributes at the time of origination such as Negative amortization (NEGAMSTATUSIND), Loan Originator Identity (ORIGINATORNAME), Balloon (BALLOONSTATUSIND), Combined-lien LTV ratio (COMBINEDLIENLTVCALC), Documentation (DOCTYPESUMMARY), Interest Only (IOSTATUSIND), Lien (LIENTYPE), FICO credit score (FICOSCOREORIGINATIONCALC), Property appraisal value (ORIGAPPRAISALVALUECALC), Purpose (PURPOSETYPE), Occupancy (OCCTYPE), Interest rate (ORIGINTRTCALC), Property (PROPTYPE), Simple LTV (ORIGLTVRATIOCALC), interest rate adjustability (INTRTTYPESUMMARY), and Loan amount (ORIGINALBALCALC).[17]

[17] BBx data field names are in parentheses.

Of those 16, 13 overlap directly with typical variables used in key studies, 1 (appraisal value) overlaps indirectly with those key variables (since it is the basis for LTV), and two are subsumed into the summary measure "Mortgage Type" (BALLOONSTATUSIND and IOSTATUSIND) in the list of fields used in academic studies.

The data fields are a combination of numeric data about the loans and qualitative data about the loans. In terms of numeric data:

- COMBINEDLIENLTVCALC considers every loan on the property, and is thought to be a more relevant risk factor than ORIGLTVRATIOCALC for second lien mortgages whose lender has claims subordinate to senior lenders.[18]
- FICOSCOREORIGINATIONCALC is a borrower's credit score issued by credit bureaus such as TransUnion, Experian, and Equifax based on the software licensed from the Fair Isaac Corporation. FICOSCOREORIGINATIONCALC is deemed to be one of the summary measures for default risk because it captures the probability of negative events in two years.[19]
- ORIGAPPRAISALVALUECALC is the property value estimated and provided by trustees. While this field is not used directly in the studies, it is the basis for computing the LTV fields.
- ORIGINTRTCALC is the coupon rate charged to the mortgagor for the first month after origination. All these numeric variables are measured at the time of loan closing.

[18] According to the prospectus supplement for ACE 2006-ASP2. "… Mortgage Loans secured by second liens that have high combined loan-to-value ratios because it is comparatively more likely that the Servicer would determine foreclosure to be uneconomical in the case of such Group II Mortgage Loans. The rate of default of second lien Group II Mortgage Loans may be greater than that of mortgage loans secured by first liens on comparable properties."

[19] Piskortski, Seru, and Vig (2010) documented FICO score, loan-to-value ratio, and interest rate type are three basic components for the loan contracts.

- ORIGLTVRATIOCALC is defined as a simple ratio of a single primary loan amount and the property value.
- ORIGINALBALCALC means the dollar amount of principal. This variable is the numerator of ORIGLTVRATIOCALC and COMBINEDLIENLTVCALC.

In terms of qualitative data:

- Negative amortization mortgage (, or payment option ARM) indicated with NEGAMSTATUSIND in BBx allows borrowers to defer monthly payments increasing their principals.
- ORIGINATORNAME for each loan is "the name of the entity that originated the mortgage."
- BALLOONSTATUSIND indicates loans for which a large lump-sum payment is scheduled near the maturity at the expense of low coupon payments during the first several years.[20]
- DOCTYPESUMMARY means documentation type. This is a standardized code whose values indicate full, reduced[21], low, no, or unknown amount of income[22] documentation provided by mortgagors.

---

[20] According to the prospectus supplement for FFML 2006-FF17, "Balloon loans pose a special payment risk because the borrower must make a large lump sum payment of principal at the end of the loan term."

[21] The prospectus for RALI 2006-QH1 documents that "Certain of the mortgage loans have been originated under "reduced documentation" or "no stated income" programs, which require less documentation and verification than do traditional "full documentation" programs. Generally, under a "reduced documentation" program, no verification of a mortgagor's stated income is undertaken by the originator. Under a "no stated income" program, certain borrowers with acceptable payment histories will not be required to provide any information regarding income and no other investigation regarding the borrower's income will be undertaken. Under a "no income/no asset" program, no verification of a mortgagor's income or assets is undertaken by the originator. The underwriting for those mortgage loans may be based primarily or entirely on an appraisal of the mortgaged property and the LTV ratio at origination."

[22] Mortgagors are usually required to provide their "assets, liabilities, income, credit history, employment history and personal information." (See the section of "underwriting guideline" in the prospectus supplement

- Borrowers may also be allowed not to pay principals for a certain period of time if they get interest-only (or IO) mortgages indicated with IOSTATUSIND in BBx.
- LIENTYPE is an indicator for lenders' relevant claim positions on the collateralized properties.
- PURPOSETYPE is associated with the reason for loans among purchase, refinancing, and cash-out refinancing.
- OCCTYPE shows whether the mortgagor uses the property for primary residence, investment, or second home. Investor loans are said to be riskier than owner-occupied or second home loans.[23]
- PROPTYPE indicates which the property belongs to among single family, planned urban development, condominium, etc.
- INTRTTYPESUMMARY indicates whether the monthly coupon amount varies across time. If the interest rate is periodically variable depending on the value of an index, those mortgages are called an adjustable rate mortgages (ARMs) while the others are called fixed rate mortgages (FRMs).

Financial researchers seem to at least partially recognize the potential for data errors and screen the data for obvious problems. Piskorski, Seru, and Vig (2010), for instance, use an algorithm to remove outliers and errors by excluding loans with non-traditional maturities and extreme values of FICO and LTV. But researchers still tend to ignore the magnitude and importance of missing data. Piskorski, Seru, and Witkin (2013), for instance, drop loans with missing values of CLTV or those

for RALI 2006-QH1.) However, the most recent cycle of BBx restricts the content of "documentation" only to income in its data definition file.

[23] See https://www.fanniemae.com/content/guide/selling/b3/2/02.html#Occupancy.20Type

with large difference between LTV and CLTV. But academic studies do not check to see how much data they are losing and how missing observations affect their sample and their research results. In fact, studies stand to lose substantial information and/or induce significant observation bias if they just winsorize or exclude loans with unknown or invalid entries of attributes.

Table 1.1. Summary of unknown loan characteristic values reported by BBx

	% undisclosed
NEGAMSTATUSIND	94.8
ORIGINATORNAME	84.6
BALLOONSTATUSIND	61.5
COMBINEDLIENLTVCALC*	47.3
DOCTYPESUMMARY	41.1
IOSTATUSIND	43.9
LIENTYPE	26.5
FICOSCOREORIGINATIONCALC*	29.3
ORIGAPPRAISALVALUECALC*	18.5
PURPOSETYPE	13.6
OCCTYPE	10.2
PROPTYPE	8.6
ORIGLTVRATIOCALC*	6.3
ORIGINTRTCALC*	5.5
INTRTTYPESUMMARY	3.9
ORIGINALBALCALC*	0.6

*Based on authors' calculation

Supplementary reports accompanying vendor data sets often show the importance of the problem. Each month, BBx sends to customers a "documentation package" summarizing all available data fields, additions, and changes to its reporting. The BBx documentation packages contain two key files relating to our research. One, Loans_Chars_Frequency_201403.pdf, reports directly the number of entries for selected data fields with a value of "U", or unknown. The other, Loans_Chars_QA_201403.pdf, reports the same measures for every non-numeric (categorical) loan characteristic data field. Both show directly that there exist "unknown" entries for a substantial number of loans. Table 1.1 illustrates the results shown in those table for our 16 data fields.[24]

The number of loans whose characteristics are not reported varies dramatically across different data fields. INTRTTYPESUMMARY and PROPTYPE are reported for most of the population, with only 3.9% and 8.6% of observations unknown, respectively. In contrast, NEGAMSTATUSIND is unknown 94.8% of the time, or for about 21 million out of 22 million loans. ORIGINATORNAME is missing 84.6% of the time.

BALLOONSTATUSIND and IOSTATUSIND are unknown in 61.5% and 43.9% of the cases reported. COMBINEDLIENLTVCALC is unknown in 47.3% of the cases reported. DOCTYPE – which 8 of the 11 cited studies rely upon – is unknown in 41.1% of the cases reported. LIENTYPE is unknown for 26.5% of the cases. FICOSCOREORIGINATIONCALC – used in every study cited above – is unknown in 29.3% of the cases reported. ORIGAPPRAISALVALUECALC is

[24] Table 1.2 presents the percentages of loans whose attributes are reported as "U" in Loans_Chars_Frequency_201403.pdf and the reported (one minus) percentage of valid observations in Loans_Chars_QA_201403.pdf. Neither file provides the distributions for numerical variables (COMBINEDLIENLTVCALC, FICOSCOREORIGINATIONCALC, ORIGAPPRAISALVALUECALC, ORIGINTRTCALC, ORIGLTVRATIOCALC, and ORIGINALBALCALC) and ORIGINATORNAME. Hence, I calculate the missing rates for those variables from Dec 2013 BBx. The "percent undisclosed" and (one minus) "percent valid" from both files are identical for fields that are reported by BBx.

unknown in 18.5% of the cases reported. OCCTYPE (used in 3 of the 11 studies) and PURPOSETYPE (used in 7 of the 11 studies) are unknown for 10.2% and 13.6% of the cases, respectively.

The statistics reported by BBx, however, cover the entire history of the data set and use all loans, including HELOCs. Still, the disclosures cited by BBx should alert researchers to the potential for selection bias. Since most research of interest focuses on some period associated with the recent financial crisis and focuses only on closed-end loans, it is more illustrative to look only at first- and second-lien loans securitized in 2005-2007. The next sections examine our 16 characteristics limited to those loans to get a better idea of how the lack of reporting could affect recent work.

1.3 Major observations for loan-level disclosure

1.3.1 Disclosure rates for major risk factors

I start the description of loan-level reporting practices with the disclosure rates of 16 key attributes for the 11,956,563 first mortgages reported in BBx associated with securitizations issued during 2005, 2006, and 2007. The key attributes are chosen on the basis of their prevalence of missing values in the BBx data. As noted above, there is a significant overlap with the data fields typically used in academic research.

Figure 1.1 presents the proportion of mortgages that reported major risk factors. Among the 16 loan characteristics, the disclosure rates are lower than 90% for 10 factors including NEGAMSTATUSIND (9.8%), ORIGINATORNAME (24.8%), BALLOONSTATUSIND (40.0%), COMBINEDLIENLTVCALC (66.6%), DOCTYPESUMMARY (68.9%), IOSTATUSIND (70.5%), LIENTYPE (80.9%), FICOSCOREORIGINATIONCALC (85.0%), ORIGAPPRAISALVALUECALC (85.0%), and PURPOSETYPE (89.3%). The least reported factor is NEGAMSTATUSIND, or the indicator for negative amortization, which is followed by ORIGINATORNAME, or the institution that sold the mortgage into the pool. The low disclosure

rates for key data fields used in academic studies in Table 1.1 suggest that academic researchers may lack significant information upon which to base conclusions.

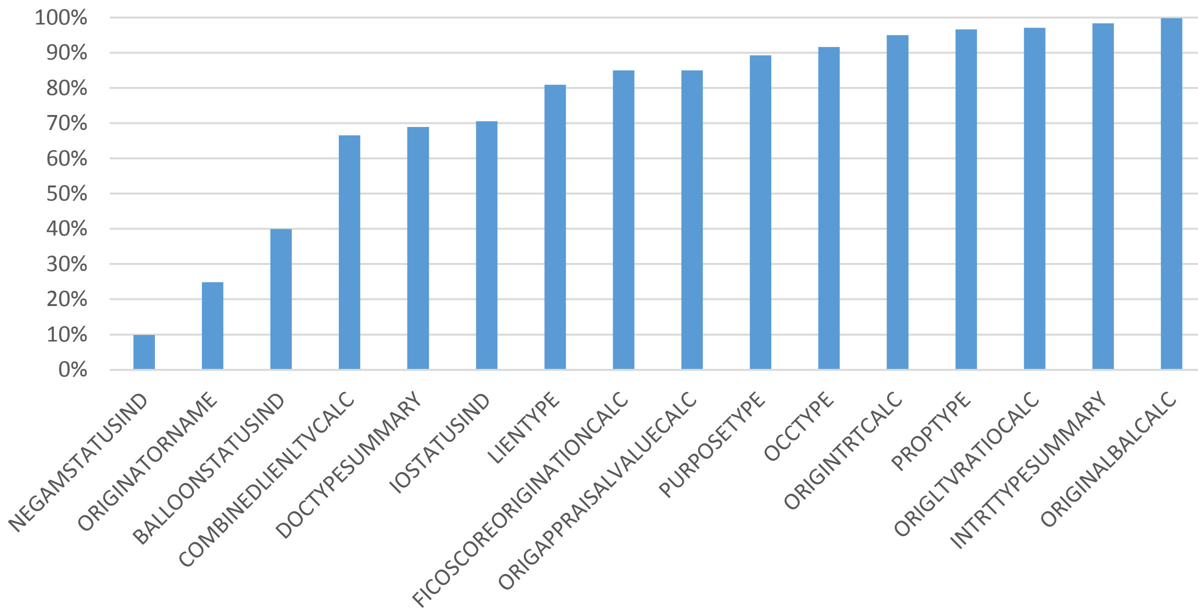

Figure 1.1. The disclosure rate by major risk factors in 2005-2007 (first liens only)

Figures 1.2.1 and 1.2.2 show how the loan-level disclosure rates in BBx have changed over time for major numeric and categorical loan characteristics. I restricted the sample period until 2008 because the private-label mortgage market was practically frozen after that. While disclosure rates increase over time for most of the loan characteristics, there is a substantial decrease in the disclosure of FICOSCOREORIGINATIONCALC (96.1% to 78.2%), ORIGINTRTCALC (93.5% to 73.0%) and ORIGAPPRAISALVALUECALC (83.2% to 65.5%) between 2007 and 2008. Among the categorical loan attributes, BALOONSTATUSIND has also decreased except for 2005 and 2006.[25] On the other hand, it is notable that the disclosure of originator identities dramatically increased from 2005 when Reg AB was adopted and the originators who provide 10% or more of the pool

[25] According to The Mortgage Market Statistical Annual, option ARMs had increased from $145 billion in 2004 to $255 billion in 2006, and 40-year Balloon mortgages had increases from 0 to $90 billion during the same periods.

assets were required to be summarized in aggregate in the offering documents, even though those were not required to be disclosed at the loan level. [26]

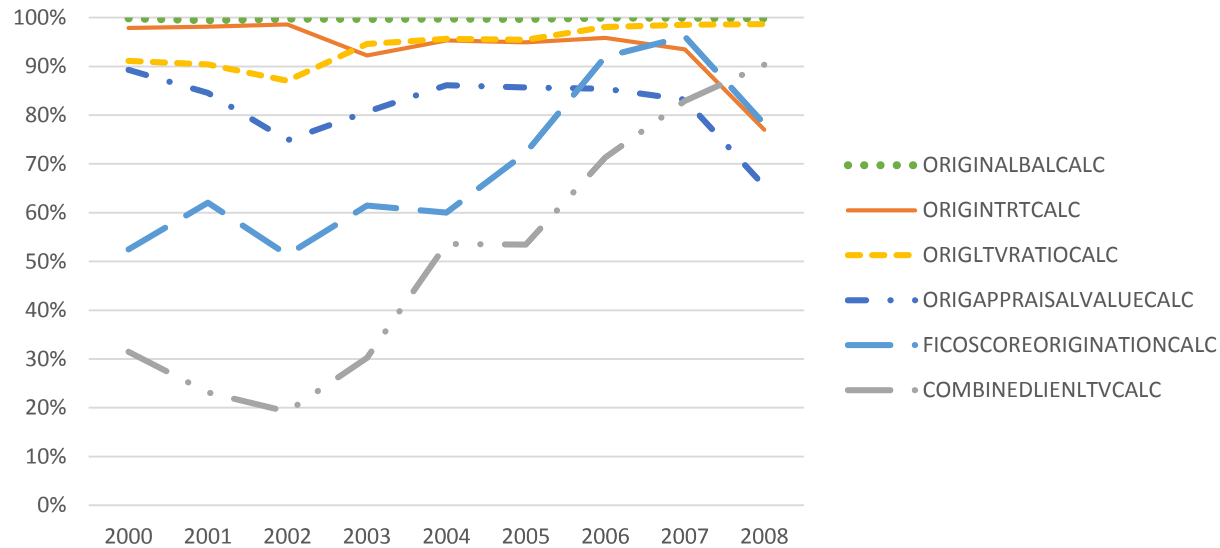

Figure 1.2.1. The yearly time-variation in the disclosure rate for numeric risk factors

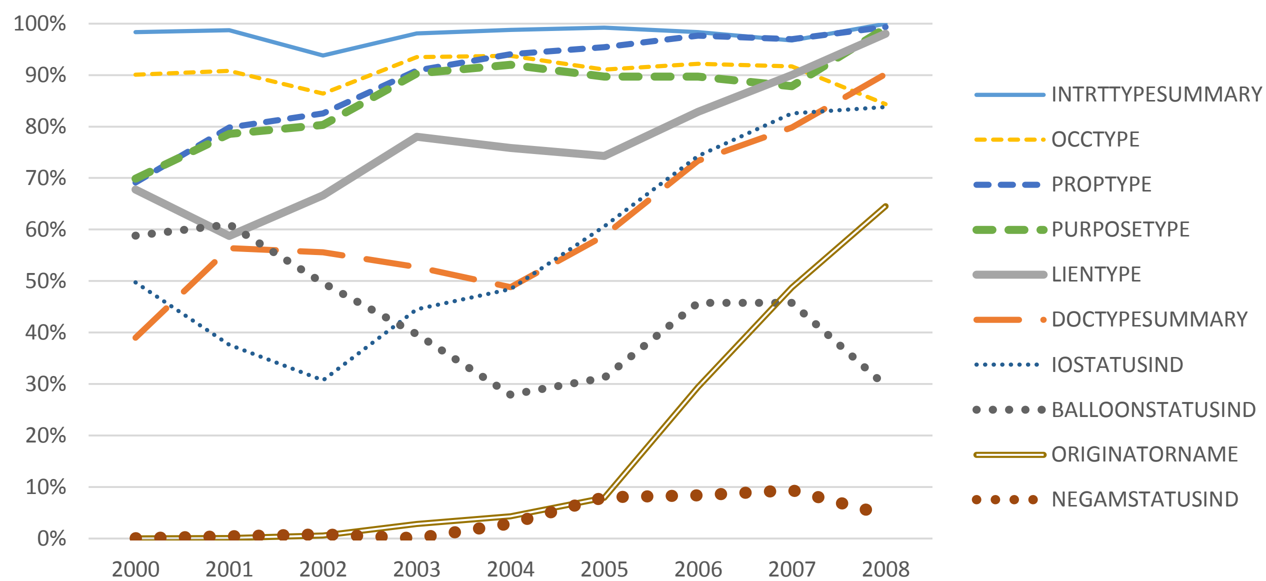

Figure 1.2.2. The yearly time-variation in the disclosure rate for categorical risk factors

[26] While an originator may be reported at the loan level, there is no regulatory guidance or rule on the definition of the originator that is reported in this field. Thus, it appears that institutions reported to be the "originator" in BBx can actually be any institution prior to the sponsor's purchase in the securitization process.

1.3.2 The relationship between loan-level disclosure and deal-level disclosure

Curiously, the comparison of the deal prospectuses and the loan-level BBx data reveals that some loan-level information summarized in prospectus supplements was not made available to investors before the crisis. For example, FICO score distributions are almost always presented in prospectuses. However, data fields for FICO scores can be missing in the loan-level data made available by trustees. Table 1.2 shows the summary statistics on FICO presented in the prospectus supplements for AMSI 2006-R1, sponsored by Ameriquest. The table shows FICO calculated for all 9,046 loans in the deal. However, the loan-level data from trustee Deutsche Bank reports FICO score for zero loans.

Table 1.2. A summary table in the prospectus for AMSI 2006-R1

Collateral Type								
COLLATERAL TYPE	NUMBER OF MORTGAGE LOANS	PRINCIPAL BALANCE AS OF THE CUT-OFF DATE ($)	% OF PRINCIPAL BALANCE AS OF THE CUT-OFF DATE	REMAINING TERM TO MATURITY (months)	DEBT-TO-INCOME (%)	MORTGAGE RATES (%)	FICO	OLTV (%)
2 YR/6MO LIB	4,674	758,768,587.55	50.58	358	42.38	8.904	578	77.53
2 YR/6MO LIB - 5YR IO	420	113,694,769.36	7.58	359	43.29	7.811	652	82.43
3 YR/6MO LIB	2,220	327,749,792.99	21.85	357	41.14	8.349	589	79.10
3 YR/6MO LIB - 5YR IO	327	79,825,955.26	5.32	359	40.73	7.382	673	83.09
5 YR/6MO LIB	18	3,384,003.36	0.23	359	39.03	7.230	734	87.12
5 YR/6MO LIB - 5YR IO	39	9,951,654.18	0.66	359	39.31	7.123	729	82.50
FIXED RATE	1,184	163,666,196.05	10.91	338	41.06	8.271	646	83.17
FIXED RATE - 5YR IO	164	42,972,083.92	2.86	359	41.96	7.434	690	78.79
Total:	**9,046**	**1,500,013,042.67**	**100.00**	**356**	**41.91**	**8.492**	**603**	**79.25**

Table 1.3 breaks down the disclosure rate for each loan attribute by sponsors for securitizations issued in 2006. Since sponsor is not a legal designation, and it is not an item provided in BBx, it must be hand-assembled, limiting our present analysis to loans securitized in 2006. The reporting of certain common loan characteristics is correlated with the deal sponsor.

Table 1.3. Disclosure rate for major loan characteristics by 10 largest sponsors in 2006

Largest 10 sponsors	CW	LB	EMC	RFC	GS	MS	DB	DLJ	BOA	JPM	Aggregate Disclosure Rate by # of Loans
# of loans	659,703	389,962	318,784	314,100	243,423	198,536	192,101	181,725	170,279	159,244	N/A
NEGAMSTATUSIND	8.80%	11.00%	18.10%	9.60%	4.60%	0.00%	0.80%	0.00%	0.20%	0.00%	7.14%
ORIGINATORNAME	0.50%	54.20%	61.40%	0.00%	46.80%	45.50%	55.60%	7.70%	42.40%	42.10%	30.93%
BALLOONSTATUSIND	22.00%	62.60%	59.70%	100.00%	32.40%	36.40%	22.80%	76.00%	8.80%	65.60%	47.60%
COMBINEDLIENLTVCALC	99.50%	72.90%	56.70%	48.10%	91.50%	58.10%	44.80%	93.40%	76.30%	67.30%	74.38%
DOCTYPESUMMARY	98.30%	70.40%	54.30%	79.70%	77.50%	39.40%	82.60%	98.00%	90.80%	89.70%	79.48%
IOSTATUSIND	92.00%	64.80%	97.40%	99.20%	88.80%	43.90%	91.40%	34.70%	92.30%	50.90%	79.99%
ORIGAPPRAISALVALUECALC	98.40%	59.50%	91.20%	100.00%	93.40%	94.40%	80.60%	38.70%	88.00%	63.40%	84.05%
PURPOSETYPE	98.40%	97.10%	57.00%	80.30%	92.00%	92.90%	90.70%	100.00%	100.00%	100.00%	90.37%
LIENTYPE	99.70%	97.20%	93.50%	100.00%	85.50%	48.70%	96.90%	100.00%	98.10%	43.40%	90.45%
OCCTYPE	99.50%	97.10%	59.50%	93.70%	99.60%	92.90%	93.00%	100.00%	100.00%	100.00%	93.21%
FICOSCOREORIGINATIONCALC	99.30%	86.10%	99.20%	100.00%	95.60%	84.10%	99.60%	97.40%	98.90%	83.90%	95.17%
PROPTYPE	99.50%	99.80%	99.50%	80.70%	99.80%	99.70%	99.70%	99.80%	99.90%	99.90%	97.57%
ORIGLTVRATIOCALC	99.50%	97.80%	97.20%	100.00%	99.20%	98.80%	94.20%	93.10%	99.40%	99.30%	98.20%
INTRTTYPESUMMARY	98.90%	99.90%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	99.73%
ORIGINTRTCALC	100.00%	100.00%	100.00%	99.60%	100.00%	100.00%	100.00%	99.90%	100.00%	100.00%	99.95%
ORIGINALBALCALC	100.00%	99.80%	100.00%	100.00%	100.00%	100.00%	99.90%	100.00%	100.00%	100.00%	99.97%
Total Disclosure Rate for 16 Fields	82.14%	79.39%	77.79%	80.68%	81.67%	70.93%	78.29%	77.42%	80.94%	75.34%	

Notes: Largest 10 sponsors in BBx are Countrywide, Lehman Brothers, EMC Mortgage Corp, Residential Funding Corp, Goldman Sachs, Morgan Stanley, Deutsch Bank, DLJ Mortgage Capital, Bank of America, and JP Morgan Chase.

Table 1.3 shows that NEGAMSTATUSIND is reported the least, for 7.14% of loans, with EMC Mortgage Corp. (EMC) reporting the field the most among sponsors (18.1% of the time) and Morgan Stanley (MS), DLJ Mortgage Capital (DLJ), and JP Morgan Chase (JPM) reporting 0.00% of the time. Bank of America (BOA) reports NEGAMSTATUSIND 0.20% of the time and Deutsch Bank (DB) reports the field 0.80% of the time, and Goldman Sachs (GS) 4.60% of the time. Residential Funding Corp (RFC), Countrywide (CW), and Lehman Brothers (LB) all report around 10% of the time.

ORIGINATORNAME is reported 30.93% of the time, although RFC reports the field 0.00% of the time, CW only 0.50% of the time, and DLJ only 7.70% of the time. The others report ORIGINATORNAME about half the time, by number of loans.

Along with NEGAMSTATUSIND and ORIGINATORNAME, ORIGAPPRAISALVALUECALC, IOSTATUSIND, DOCTYPESUMMARY, COMBINEDLIENLTVCALC, and BALLOONSTATUSIND were all disclosed by the top ten sponsors in BBx less than 90% of the time by loan count in the aggregate. BALLOONSTATUSIND was reported on average for 47.60% of loans, with a low of 8.80% (BOA) and a high of 100.00% (RFC). COMBINEDLIENLTVCALC was reported on average for 44.805% of loans, with a low of 99.50% (DB) and a high of 100.00% (CW). DOCTYPESUMMARY was reported on average for 79.48% of loans, with a low of 39.40% (MS) and a high of 98.30% (CW). IOSTATUSIND was reported on average for 79.48% of loans, with a low of 34.70% (DLJ) and a high of 99.20% (RFC). ORIGAPPRAISALVALUECALC was reported on average for 84.05% of loans, with a low of 38.70% (DLJ) and a high of 100.00% (RFC).

ORIGINALBALCALC, ORIGINRTCALC, INTRTTYPESUMMARY, and ORIGLTVRATIOCALC were all disclosed by the top ten sponsors for more than 98.00% of loans in 2006.

Among our 16 characteristics, MS loans have the most missing data, with 70.93% of coverage for our 16 attributes, followed by JPM (75.34%), DLJ (77.42%), EMC (77.79%) and DB (78.29%). The highest coverage is from CW with 82.14%. With such uneven coverage among deal sponsors, it seems necessary to control for the variations in data omissions across different sponsors.

In Table 1.4, I calculate the disclosure rate for our 16 key loan characteristics by the top 10 trustees for 2005 to 2007 securitizations. As described above, the individual loan characteristics are initially produced by the originator. The originator conveys those to the servicer so that the servicer knows what types of loans they are servicing. If the loans are securitized (as are all the loans used here), the loan characteristics would be conveyed to the sponsor and investment bank as the basis for representation made in the prospectus supplements and other offering documents. Those attributes may also be posted on the trustee's web site in order to inform investors about the loans. The hypothesis, therefore, is that the trustee may somehow be associated with reporting.

Here, I find that there is even *greater* heterogeneity in reporting of certain loan common attributes at the loan level among trustees, so that reporting is highly correlated with the deal trustee. Again, coverage of NEGAMSTATUSIND and ORIGINATORNAME is lowest, with 8.13% and 25.02% coverage, in the aggregate. DB, RFC, LaSalle, JPM and Citi all report *de minimis* coverage of NEGAMSTATUSIND, ranging from 0.00% to 2.20%, while WaMu covers 51.90%. BONY, RFC, IndyMac, and WaMu all provide *de minimis* coverage of ORIGINATORNAME, ranging from 0.20% to 2.90%, while LaSalle covers up to 50.40%.

Table 1.4. Disclosure rate for major loan characteristics by top ten trustees, 2005 to 2007

Top 10 Trustees	WF	BONY	DB	RFC	LaSalle	US Bank	JPM	IndyMac	Citi	WaMu	Aggregate Disclosure Rate by # of Loans
Number of loans	5163810	1652638	1447918	808550	751323	670973	538501	301478	296150	230533	N/A
NEGAMSTATUSIND	8.28%	10.86%	1.70%	1.64%	0.34%	19.45%	0.00%	19.35%	2.19%	51.91%	8.13%
ORIGINATORNAME	36.49%	0.25%	19.58%	2.91%	50.36%	44.22%	11.27%	2.06%	9.64%	0.86%	25.02%
BALLOONSTATUSIND	15.14%	11.33%	17.41%	97.54%	99.31%	99.95%	100.00%	95.98%	90.97%	89.66%	39.86%
COMBINEDLIENLTVCALC	76.17%	91.05%	46.89%	52.42%	7.58%	86.22%	30.00%	95.94%	55.17%	62.46%	66.91%
DOCTYPESUMMARY	78.82%	91.00%	19.27%	84.91%	1.27%	77.82%	96.61%	88.18%	39.60%	97.60%	69.12%
IOSTATUSIND	82.84%	87.18%	11.69%	94.69%	100.00%	38.77%	24.00%	95.98%	29.95%	89.66%	70.30%
LIENTYPE	90.50%	90.92%	18.63%	98.84%	99.29%	90.04%	49.40%	95.96%	87.35%	89.91%	81.06%
FICOSCOREORIGINATIONCALC	96.12%	86.57%	47.99%	99.90%	92.86%	82.66%	62.40%	95.80%	67.74%	40.92%	84.88%
ORIGAPPRAISAISALVALUECALC	86.68%	92.07%	99.46%	99.97%	81.30%	30.66%	86.02%	96.16%	21.71%	99.98%	85.26%
PURPOSETYPE	97.47%	89.29%	99.82%	84.91%	0.23%	99.39%	100.00%	95.74%	87.70%	100.00%	89.60%
OCCTYPE	97.61%	93.07%	99.90%	99.97%	7.04%	99.54%	99.81%	98.32%	95.05%	100.00%	91.89%
ORIGINTRTCALC	99.98%	99.80%	99.98%	99.60%	99.99%	99.94%	99.99%	100.00%	100.00%	100.00%	99.94%
PROPTYPE	98.72%	93.12%	99.90%	84.91%	100.00%	99.63%	95.19%	95.74%	93.71%	100.00%	96.92%
ORIGLTVRATIOCALC	96.82%	93.04%	100.00%	99.97%	97.75%	96.63%	99.96%	96.17%	95.95%	99.98%	97.10%
INTRTTYPESUMMARY	99.99%	96.82%	100.00%	89.82%	100.00%	99.95%	100.00%	88.25%	92.19%	100.00%	98.36%
ORIGINALBALCALC	99.96%	99.42%	100.00%	99.97%	100.00%	99.99%	99.99%	99.66%	99.83%	100.00%	99.90%
	76.85%	76.61%	61.39%	80.74%	64.84%	79.06%	72.17%	84.96%	66.79%	82.69%	

Some trustees report OCCTYPE, PURPOSETYPE, ORIGAPPRAISALVALUECALC, LIENTYPE, IOSTATUSIND, DOCTYPESUMMARY, COMBINEDLIENLTVCALC, and BALLOONSTATUSIND at levels of 95.00% and above, but others report those same fields at rates down to 0.20%,with no one trustee exceeding 21.70%. Reporting for PURPOSETYPE ranges from 0.20% (LaSalle) to 100.00% (JPM, WaMu), a 99.80% spread; DOCTYPESUMMARY ranges from 1.30% (LaSalle) to 97.60% (WaMu), a 96.30% spread; OCCTYPE ranges from 7.00% (Lasalle) to 100.00% (RFC, WaMu), a 93.00% spread; BALLOONSTATUSIND ranges from 11.30% (BONY) to 100% (USBank, JPM), an 88.70% spread; COMBINEDLIENLTVCALC ranges from 7.60% (LaSalle) to 95.90% (IndyMac), an 88.30% spread; IOSTATUSIND ranges from 11.70% (DB) to 100.00% (Lasalle), an 88.30% spread; LIENTYPE ranges from 18.60% (DB) to 99.30% (LaSalle), an 80.70% spread; ORIGAPPRAISALVALUECALC ranges from 21.70% (Citi) to 100.00% (RFC, WaMu), a 78.30% spread; finally, FICOSCOREORIGINATIONCALC ranges from 40.90% (WaMu) to 99.90% (RFC), a 59.00% spread for a variable that is almost universally relied upon in academic studies.

ORIGINRTCALC, ORIGINALBALCALC, INTRTTYPESUMMARY, ORIGLTVRATIOCALC, and PROPTYPE are reported more than roughly 85% of the time by all trustees, although only Interest rate, Loan amount, and Interest rate type are reported more than 98.00% of the time.

Among our 16 characteristics, loans in which deals for which DB is trustee have the worst coverage (61.39%), followed by Lasalle (64.84%), Citi (66.79%), JPM (72.17%) and Bony (76.61%). The top trustee is WaMu (82.69%), although WaMu is also the smallest trustee of the top ten in terms of the number of loans. With even more uneven coverage among deal trustees than deal sponsors, it seems mandatory to carefully control for variation in missing loan-level information across different trustees.

1.3.3 The possibility of inferring loan-level disclosure from other sources

In this section I explore whether (1) those missing loan-level observations are trivial, i.e., whether a missing value is zero because the data is reported elsewhere, like the prospectus supplement, (2) whether they are truly missing and irreconcilable, and (3) whether there are apparent discrepancies among reporting sources. I establish that while in some cases one can backfill data using inferential methods, there is no single rule that can be implemented to do so without careful attention to deal-level and loan-level reporting.

First, I examine the possibility of using prospectus supplement disclosures using the Balloon Type data field, whose loan-level disclosure rate is the third lowest (39.97%) following NEGAMSTATUSIND (9.84%) and ORIGINATORNAME (24.84%). Then, I explore the more difficult case of the Loan Originator field.

1.3.3.1 Inferring Balloon Type from deal-level data

As a first consideration, prospectus supplement disclosures need not exactly represent the composition of the loan pool that is finally securitized. Rather, it is common for prospectus supplement to contain language to the effect of, "Prior to the Closing Date… we may remove mortgage loans from the mortgage pool and we may substitute other mortgage loans for the mortgage loans we remove. The depositor believes that the information set forth in this prospectus supplement with respect to the mortgage pool as presently constituted is representative of the characteristics of the mortgage pool as it will be constituted on the Closing Date, although certain characteristics of the mortgage loans in the mortgage pool may vary."[27]

[27] See, for instance, BSARM 2006-4 Prospectus Supplement at 26.

Still, in some cases one can use prospectus supplement disclosures to fill in missing data. For example, a prospectus supplement may document that the pools consist of 100% or 0% balloon mortgages although the value for BALLOONSTATUSIND is "U" for every loan in those deals.

For instance, the prospectus supplement for MSAC 2006-NC1 documents, "Approximately 0.00% of the mortgage loans will not be fully amortizing over their terms to maturity and, thus, will require substantial principal payments, i.e., balloon payments, at their stated maturity." Alternatively, according to the representations and warranties section in the prospectuses for HVMLT 2006-3, 5, and 9, "(z) No mortgage loan has a balloon payment feature." For both of those deals, BALLOONSTAUSIND is unknown for 100% of the cases. In both cases, one could use the prospectus supplement disclosures to fill in the proper data (or a reasonably accurate approximation of such data).

In other cases, information for balloon type is unavailable at the loan-level but partially available at the deal-level. In this case, however, prospectuses provide no information about balloon type other than a simple hint that balloon mortgages may be included in the pool. For instance, the value for balloon type variable is "unknown" in BBx for every mortgage in BAFC 2006-1 while the prospectus for BAFC 2006-1 states merely that "A trust may include one or more of the following types of mortgage loans… [,including] balloon loans."[28]

Similarly, CMLTI 2006-WFH4 provides only partial reporting. In CMLTI 2006-WFH4, only 0.58% of the loans are reported to have balloon payments in the BBx data while the prospectus supplement documents 304 such loans out of a total of 5,782, for a proportion of 5.26%. (6.07% by principal

---

[28] *Prospectus*, Banc of America Funding Corporation, Depositor, Bank of America, National Association, Sponsor, Mortgage Pass-Through Certificates, January 27, 2006, p. 2

balance).[29] The difference implies that approximately 270 of the balloon loans in this deal are reported in loan-level data as "unknown."

In yet other cases, there may exist an apparent discrepancy for balloon type reporting between loan-level data and the deal-level prospectus supplements. For instance, in BSABS 2006-IM1, 4,320 out of 4,321 loans (99.98%) have balloon feature according to the loan-level BBx data. However, the BSABS 2006-IM1 prospectus supplement lists only 666 out of 4,321 loans possessing balloon characteristics, or 15.41% of the pool (4.42% by principal balance).[30]

While I cannot remedy the discrepancy cited above without substantial additional investigation, I can see that in some cases "unknown" balloon types can be ruled out if deal prospectus supplements do not allow such loans. In those cases, at least some of the missing values can be back-filled, therefore potential bias can be alleviated through further investigation.

1.3.4. Disclosure of loan providers

Reg AB requires securitizers to report the distribution of material originators (defined as those that originate more than 10% of the mortgages in the pool) at the deal level.[31] In contrast, there exists no guidance for reporting the originator at the *loan*-level. In fact, there appears to be significant discrepancy in such reporting.

Sponsors often acquire mortgages from initial loan originators or entities that purchase loans from those originators. Given the absence of any regulation for loan-level disclosure of loan originators, sponsors seem to have chosen among institutions along the chain of origination to be disclosed at the loan level. Figure 1.3 shows the distribution of the number of loan originators disclosed at the loan-level in 2005 to 2007. The vast majority of deals in this period, 2,235 deals in total, reported

[29] See CMLTI 2006-WFH4 prospectus supplement at 33.
[30] See BSABS 2006-IM1 prospectus supplement at p. 162.
[31] See the section for originators on p. 1538 of 33-8518FR.

zero originators at the loan-level. Around 175 deals reported a single originator name for all loans in the deal, and almost 250 reported two originator names. On the other end of the scale, around 125 deals reported fifty or more originator names associated with the deal.

Additionally, as shown in Figures 1.2.1 and 1.2.2, there has been substantial variation across time in the reporting of different loan characteristics. Except for COMBINEDLIENLTVCALC, NEGAMSTATUSIND, BALLOONSTATSIND, and, ORIGINATORNAME all other attributes have historically been consistently disclosed for 30% or more of loans in securitized pools. Over time, however, the loan-level disclosure rate for ORIGINATORNAME has increased dramatically, from 5.5% in 2005 to 64.6% in 2008.

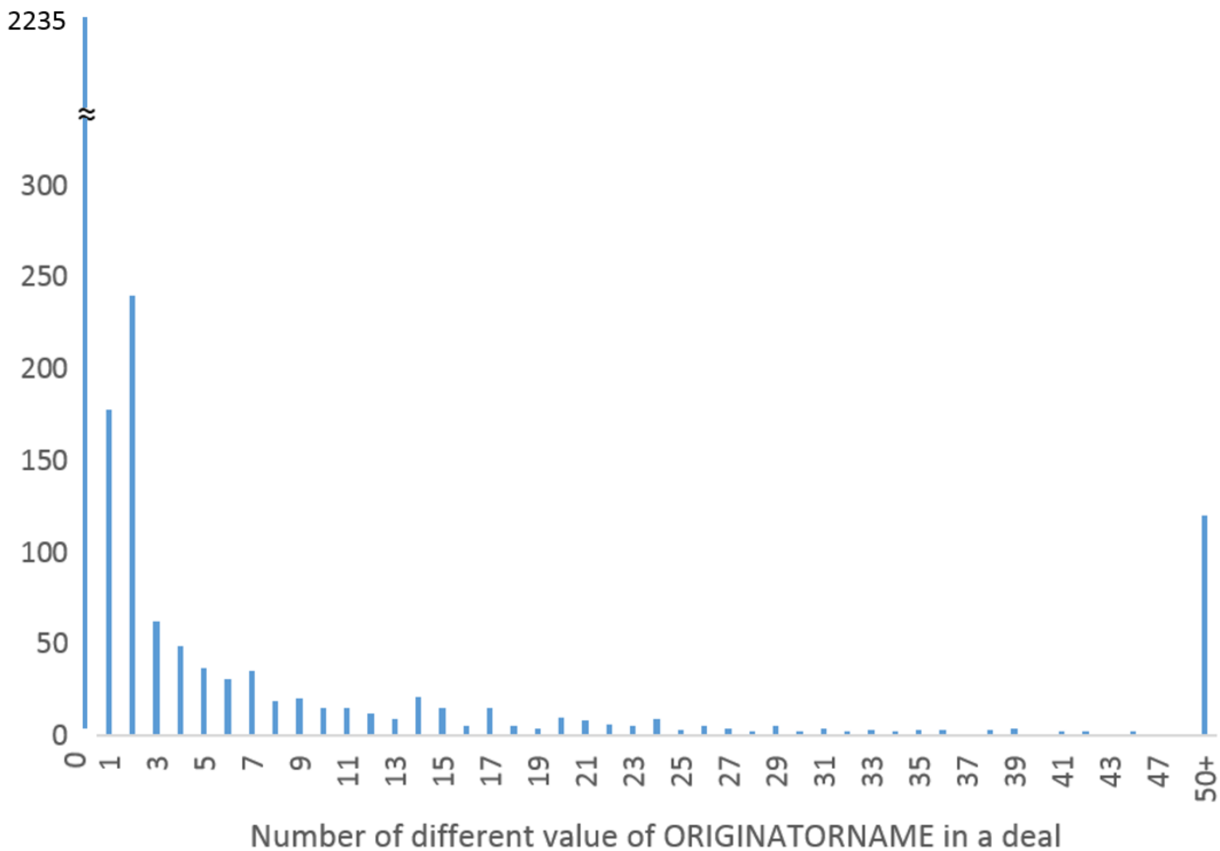

Figure 1.3. Histogram of the number of loan originators disclosed in loan-level data[32]

[32] The deals with more than 100 of originators are treated to have 100 of originators. (Some deals (BNP2006005) have 1,179 originators.)

Still, that increase in ORIGINATORNAME reporting masks substantial changes in the nature of reporting both within and across deals. For example, Taylor, Bean, and Whitaker (TBW) always purchases mortgages from brokers and correspondents as a wholesale lender and a sponsor of its own securitizations. In BBx, it reported 1,114 brokers as the originators for 3,001 mortgages in TBW 2006-5. However, TBW reported itself as the sole originator for 2,565 mortgages in TBW 2006-6. Moreover, while the originator identity in the loan-level data varies across different series of TBW securitizations, TBW is always reported to be the single originator in their prospectus supplements.

Obviously no inferential approach could be used to add originator detail to TBW's disclosure of itself as originator when it is obvious that many others are involved. On the other hand, detail on every single person that originated loans in each deal would not meaningfully help researchers evaluate risk. Somewhere in between is an economically meaningful balance of whose guidelines the origination was supposed to have adhered to, whether because they were the originator or because they provided funding to the originator contingent upon such guidelines being followed. But because of the multiplicity of potential originators to be reported with each loan in the present data, further work is needed to meaningfully disentangle the reporting relationships in this data field.[33]

1.4 Patterns of disclosure among multiple data fields

In this section, I examine the empirical relation between loan-level disclosures across different characteristics to see if there are patterns of non-disclosure among the different data fields. A simple correlation matrix looking at the pairwise correlations between disclosure dummy variables

[33] See, for instance, Mason and Lee 2014.

that are 1 if the data item is reported and 0 otherwise provides intriguing evidence that loans without information on one data field are likely to lack data on others.

Pearson's Phi coefficients are used to measure the association between two dichotomous disclosure indicators in Table 1.5. Every correlation coefficient is significant at the 1% level except for the correlation between NEGAMSTATUSIND and OCCTYPE. The sample is restricted to the loans securitized in 2005 to 2007. Second lien loans are excluded, as are loans that disclose lien as unknown.

First, note that most of the correlation coefficients are positive indicating that many of the data items are reported together and missing together. Among 105 non-diagonal elements in the upper triangular matrix presented in Table 1.5, only 26 elements have negative values. Therefore, it is not likely that one risk factor is missing at the expense of disclosing the other. However, disclosure of a few variables do have frequently negative correlations with others. Negative correlations are the most frequent for disclosure of BALLOONSATUSIND (9), which is followed by ORIGINTRTCALC (7). In contrast, disclosures of ORIGLTVRATIOCALC and ORIGINALBALCALC have no negative associations with disclosures of other loan attributes.

The highest correlation coefficient is 0.78 which occurs between OCCTYPE and PURPOSETYPE. Thus, disclosing whether a mortgage is for an owner-occupied home is likely to occur with reporting whether the mortgage is for a purchase, refinance, or cash-out refinance. Next, information regarding whether or not a loan was interest only, or IOSTATUSIND, is highly correlated with the documentation provided with correlation coefficients greater than 0.40. This is very interesting considering that at the center of the crisis was the higher rate of low-doc/no-doc loans that were made as well as the increased number of exotic mortgages. The fact that, empirically, these two

critical risk factors were likely to be reported together or missing together suggests any analysis made regarding their relative contribution to the performance of the loan would be severely biased.

OCCTYPE and PURPOSETYPE have high correlation with COMBINEDLIENLTVCALC (0.34 and 0.32), which is a bit puzzling. The interpretation here is that a loan not having disclosed the type of occupancy or purpose was more likely to also omit the COMBINEDLIENLTVCALC. One would think that these two variables would be independent of home equity, but the substantial correlation coefficients indicate that there may be something else going on.

In fact, COMBINEDLIENLTVCALC disclosure has a correlation coefficient higher than 0.1 with six out of the remaining 15 variables; four out of those six are qualitative variables. Considering the importance of COMBINEDLIENLTVCALC in risk analysis, it is worth looking into why it would be so highly correlated with these other factors.

Lastly, let us turn our attention to correlations between FICOSCOREORIGINATIONCALC and the other variables. Four out of the 15 other variables have a correlation coefficient greater than 0.10. This suggests that when one of those data items is missing – which include DOCTYPESUMMARY, IOSTATUSIND, ORIGAPPRAISALVALUECALC, and ORIGINATORNAME – it is likely to also be missing the FICOSCOREORIGINATIONCALC. This too would have an impact on the results of any risk analysis being performed on the mortgages collateralizing the RMBS of interest.

Table 1.5. Correlations among disclosures for verifiable first-lien loans securitized 2005-2007

	NEGAMSTATUSIND	ORIGINATORNAME	BALLOONSTATUSIND	COMBINEDLIENLTVCALC	DOCTYPESUMMARY	IOSTATUSIND	FICOSCOREORIGINATIONCALC	ORIGAPPRAISALVALUECALC	PURPOSETYPE	OCCTYPE	PROPTYPE	ORIGLTVRATIOCALC	INTRTTYPESUMMARY	ORIGINALBALCALC	ORIGINTRTCALC
NEGAMSTATUSIND	1	0.05	0.04	0.06	0.06	0.03	-0.08	-0.04	0.02	0.00	-0.11	0.00	0.02	0.00	0.08
ORIGINATORNAME		1	-0.01	0.16	0.08	0.15	0.14	-0.09	-0.07	-0.04	0.06	0.01	0.02	0.00	-0.02
BALLOONSTATUSIND			1	-0.23	-0.11	-0.10	-0.08	-0.14	-0.24	-0.24	-0.04	0.04	0.01	0.00	0.15
COMBINEDLIENLTVCALC				1	0.23	0.04	0.06	0.07	0.32	0.34	0.06	0.14	-0.02	0.02	0.09
DOCTYPESUMMARY					1	0.40	0.17	0.08	0.45	0.45	0.22	0.02	-0.01	0.01	-0.01
IOSTATUSIND						1	0.11	0.24	-0.08	-0.08	0.10	0.03	0.06	0.01	-0.05
FICOSCOREORIGINATIONCALC							1	0.11	0.04	0.06	0.04	0.03	0.00	0.02	-0.05
ORIGAPPRAISALVALUECALC								1	0.01	0.01	0.02	0.28	0.06	0.03	0.02
PURPOSETYPE									1	0.78	0.22	0.03	0.10	0.02	0.01
OCCTYPE										1	0.24	0.04	0.01	0.02	-0.01
PROPTYPE											1	0.09	0.03	0.04	-0.02
ORIGLTVRATIOCALC												1	0.01	0.08	0.03
INTRTTYPESUMMARY													1	0.02	-0.01
ORIGINALBALCALC														1	0.01
ORIGINTRTCALC															1

I can also test the institutions' unwillingness to disclose risky values with the credit category assigned by issuers. While FICOSCOREORIGINATIONCALC is missing for just under 15% of mortgages, those missing values are, again, not random.

Table 1.6. FICO disclosure rate by credit category

	Prime	Alt-A	Subprime	Unknown
	5,728,732	2,155,354	3,670,613	401,860
NEGAMSTATUSIND	14.30%	7.60%	**0.20%**	4.30%
ORIGINATORNAME	25.60%	26.70%	**24.90%**	4.20%
BALLOONSTATUSIND	38.20%	46.30%	**36.40%**	63.90%
COMBINEDLIENLTVCALC	72.30%	69.00%	**60.40%**	28.90%
DOCTYPESUMMARY	77.30%	72.90%	**57.50%**	32.10%
IOSTATUSIND	82.10%	77.00%	**53.20%**	29.80%
LIENTYPE	90.10%	89.50%	**64.30%**	56.80%
FICOSCOREORIGINATIONCALC	95.80%	87.90%	**73.40%**	21.60%
ORIGAPPRAISALVALUECALC	87.10%	**82.90%**	85.50%	62.30%
PURPOSETYPE	**89.70%**	85.50%	91.80%	82.70%
OCCTYPE	92.50%	**87.90%**	93.00%	86.20%
ORIGINTRTCALC	**93.40%**	95.40%	96.80%	98.70%
PROPTYPE	96.70%	**96.00%**	98.10%	86.10%
ORIGLTVRATIOCALC	98.20%	**95.40%**	97.10%	91.90%
INTRTTYPESUMMARY	**98.50%**	99.20%	**98.50%**	91.70%
ORIGINALBALCALC	100.00%	99.90%	**99.80%**	99.5 %

Table 1.6 shows, based upon lender classifications, that the disclosure rate for FICOSCOREORIGINATIONCALC is the lowest (73.4%) for subprime loans and highest (95.8%) for prime loans. The other fields that are missing the most are also concentrated among subprime loans. For 10 out of 16 variables, the disclosure rate is the lowest for subprime mortgages. The spread between the highest and lowest disclosure is the largest (28.9%) for IOSTATUSIND, which is followed by LIENTYPE (25.8%). There is the smallest difference in disclosure rate between

prime and subprime sectors for ORIGINATORNAME (0.7%). However, generally speaking, the disclosure is the most likely for prime and the least likely for subprime mortgages. Such patterns support the notion that reporting is associated with credit risk. I explore that hypothesis more formally in the next section.

1.5 Initial evidence of selection bias in disclosure and ex-post mortgage performance

I next investigate how disclosure affects the post-securitization performance. Table 1.7 presents information on each risk factor used in my analysis.

Table 1.7. Summary statistics

Variable	N	N Miss	Min	Max	Mean	Std Dev
FICO	7,135,077	612,616	302	850	669.276	73.127
Combined-Lien LTV	5,734,628	2,013,065	1	149.7	80.982	16.567
Simple LTV	7,631,887	115,806	1	150	76.957	13.660

Interest rate type	
ARM	65.87%
FRM	33.94%
Unknown	0.19%

Lien type	
First	100.00%
Junior	0.00%
Unknown	0.00%

Doc type	
Full	33.81%
Non-full	48.11%
Unknown	18.08%

Balloon type	
Yes	9.46%
No	26.87%
Unknown	63.67%

IO type	
Yes	26.74%
No	56.94%
Unknown	16.32%

Neg Am type	
Yes	11.72%
No	0.00%
Unknown	88.28%

I again limit my sample to 7,747,693 verifiable first-lien loans securitized during 2005-2007. As expected, substantial amount of loans are missing numerical and categorical risk factors. For

example, CLTVs are not reported for 2,103,075 (27.1%) loans. The missing rate hits 88.3% for NEGAMSTATUSIND.

Tables 1.8.1 and 1.8.2 show the marginal effects estimates from logit regressions of mortgage default on the loan characteristics and risk factors. I report specifications for a two-year default horizon as well as a four-year default horizon. The t-statistics appear in parentheses below the coefficient estimates.

The dependent variable is 1 if the mortgage in the pool defaults within two years after origination and 0 otherwise.[34] A mortgage is defined to default if it is seriously delinquent. Following the definition of serious delinquency by Office of the Comptroller of the Currency (OCC)[35], a mortgage is seriously delinquent if the loan becomes 60 or more days past due according to the Mortgage Bankers Association (MBA) definition (that is, 60 days after the missed payment date or 90 days after the last received payment was due) as defined by the BBx performance variable MBADELINQUENCYSTATUS="6" in the BBx date.[36]

Unlike previous research, I allow unreported data to enter into the specification using dummy variables for "unknown" if such a condition exists. In particular, FICOSCOREORIGINATIONCALC and COMBINEDLIENLTVCALC, two most important continuous risk factors, are broken down to several ranges to incorporate missing dummies into the model.

---

[34] Note that I do not account for loans that may have defaulted and been removed from trusts according to substitution clauses common in securitizations in this period. While that is another important element of potential bias in the securitized loan data there is, again, no specific reporting of such repurchases so that I leave imputing such repurchases as a topic for future research.

[35] See http://www.occ.gov/publications/publications-by-type/other-publications-reports/mortgage-metrics-q1-2009/definitions-and-methods-2009-1-quarter.html

[36] For the examples of delinquency calculation, see http://www.securitization.net/pdf/content/ADC_Delinquency_Apr05.pdf

Table 1.8.1. Two-year performance logit regression including undisclosed loans

DV = 1 if the mortgage defaults anytime within 2 years from origination; 0 otherwise			
	Model 1	Model 2	Model 3
FICO unknown (d)	0.0634***	0.0722***	0.0740***
	(80.597)	(90.306)	(92.271)
CLTV unknown (d)	0.1193***	0.1281***	0.1252***
	(244.750)	(259.706)	(245.217)
Doc unknown (d)		0.0271***	0.0356***
		(60.788)	(75.167)
Balloon unknown (d)			0.0002
			(0.563)
IO unknown (d)			-0.0300***
			(-72.826)
FICO < 620 (d)	0.1971***	0.2319***	0.2304***
	(451.224)	(495.373)	(474.560)
620 ≤ FICO < 680 (d)	0.1250***	0.1358***	0.1324***
	(324.887)	(349.463)	(333.444)
80 ≤ CLTV < 100 (d)	0.1097***	0.1139***	0.1121***
	(261.827)	(271.811)	(267.535)
100 ≤ CLTV (d)	0.2047***	0.2126***	0.2063***
	(330.525)	(343.803)	(331.262)
ARM (d)	0.0885***	0.0855***	0.0801***
	(312.315)	(299.839)	(275.739)
Low doc		0.0664***	0.0656***
		(202.337)	(199.993)
Balloon (d)			0.0579***
			(107.073)
IO (d)			0.0146***
			(42.866)
Originated before 2005 (d)	-0.1685***	-0.1655***	-0.1583***
	(-583.641)	(-561.642)	(-501.921)
Originated in 2005 (d)	-0.1948***	-0.1913***	-0.1814***
	(-519.394)	(-510.863)	(-472.894)
Originated in 2006 (d)	-0.0596***	-0.0594***	-0.0590***
	(-159.214)	(-159.106)	(-159.485)
pseudo R-sq	0.1258	0.1314	0.1347
N	7,421,653	7,421,653	7,421,653

Table 1.8.2. Four-year performance logit regression including undisclosed loans

DV = 1 if the mortgage defaults anytime within 4 years from origination; 0 otherwise			
	Model 1	Model 2	Model 3
FICO unknown (d)	0.0654***	0.0874***	0.0859***
	(89.078)	(116.185)	(114.511)
CLTV unknown (d)	0.1196***	0.1380***	0.1236***
	(244.393)	(282.640)	(242.725)
Doc unknown (d)		-0.0049***	0.0136***
		(-9.856)	(25.641)
Balloon unknown (d)			-0.0300***
			(-76.000)
IO unknown (d)			-0.0683***
			(-143.718)
FICO < 620 (d)	0.1539***	0.1984***	0.2004***
	(349.933)	(422.170)	(414.586)
620 ≤ FICO < 680 (d)	0.1252***	0.1414***	0.1386***
	(301.844)	(341.304)	(328.966)
80 ≤ CLTV < 100 (d)	0.1436***	0.1484***	0.1466***
	(325.348)	(337.088)	(334.566)
100 ≤ CLTV (d)	0.2297***	0.2362***	0.2259***
	(386.271)	(400.990)	(379.394)
ARM (d)	0.1120***	0.1098***	0.1029***
	(329.277)	(321.161)	(296.054)
Low doc		0.0963***	0.0958***
		(252.833)	(252.093)
Balloon (d)			0.0428***
			(68.859)
IO (d)			0.0210***
			(53.492)
Originated before 2005 (d)	-0.2439***	-0.2354***	-0.2199***
	(-586.435)	(-545.858)	(-470.735)
Originated in 2005 (d)	-0.1901***	-0.1840***	-0.1680***
	(-397.015)	(-382.701)	(-339.736)
Originated in 2006 (d)	-0.0288***	-0.0288***	-0.0270***
	(-59.034)	(-59.226)	(-55.605)
pseudo R-sq	0.0960	0.1044	0.1089
N	7,522,814	7,522,814	7,522,814

As seen in Table 1.8.1, loans with unknown values are riskier than the safest categories of each attribute, on average, although they are not as risky as the next riskiest cohort. It is particularly interesting to look at the disclosure of FICO scores and CLTV ratios – two of the most widely considered data items for risk analysis of mortgages. Across all of the models, an unreported FICO score is associated with a 6.3% - 8.7% greater chance of default than if it were reported to have been greater than 680, or slightly lower – by itself – than if the loan was associated with a reported FICO between 620 and 680. This suggests the possibility that credit scores are less likely to be reported for the mortgages with intermediate level of FICO scores (approximately with the mean of 680). This may imply the convexity of the curve on the coordinates whose X- and Y-axes represent FICO scores and the corresponding disclosure rate respectively. This is an arguably reasonable disclosure pattern for two reasons: 1) sponsors may be more willing to convey the FICO score if it is high; 2) mortgages for borrowers with low FICO scores may have riskier other loan attributes than those with higher FICO scores, and MBS investors would demand the disclosure of FICO more aggressively if other loan attributes indicate higher risk.

When CLTV is unreported, there is an 11.9% - 13.8% greater chance of default than if it were reported to be lowest range of CLTV lower than 80%, slightly higher than if the CLTV were between 80 and 100 for the two-year default model, and slightly lower than if the CLTV were between 80 and 100 for the four-year default model. Given the association between the omission of CLTV and the intermediate level of ex-post default risk, it is also possible to make an inference that extremely low or high CLTV ratios are more likely to be reported than intermediate level of CLTV.

Similarly, Doc unknown is associated with a 1.4-3.6% greater chance of default than a full doc loan except for Model 2 in Table 1.8.2, still lower than disclosed Low Doc loan (6.6-9.6% greater chance of default). IO unknown is associated with a lower probability of default than a non-IO loan.

Table 1.9: Performance forecast ratios for the two-year and four-year models with and without loans with unknown characteristics using logit specifications – comparison between predicted and actual likelihood of default

2 year default logit regression including unknowns			
	Actually current	Actually default	Total
Predicted current	5,749,493	1,223,910	6,973,403
	82.45%	17.55%	100%
Predicted default	226,187	229,792	455,979
	49.60%	50.40%	100%
Total	5,975,680	1,453,702	7,429,382
	80.43%	19.57%	100%
	% correct prediction		80.48%

2 year default logit regression without unknowns			
	Actually current	Actually default	Total
Predicted current	692,606	233,901	926,507
	74.75%	25.25%	100%
Predicted default	55,134	62,090	117,224
	47.03%	52.97%	100%
Total	747,740	295,991	1,043,731
	71.64%	28.36%	100%
	% correct prediction		72.31%

4 year default logit regression including unknowns			
	Actually current	Actually default	Total
Predicted current	4,565,831	1,662,992	6,228,823
	73.30%	26.70%	100%
Predicted default	544,719	757,774	1,302,493
	41.82%	58.18%	100%
Total	5,110,550	2,420,766	7,531,316
	67.86%	32.14%	100%
	% correct prediction		70.69%

4 year default logit regression without unknowns			
	Actually current	Actually default	Total
Predicted current	430,920	220,727	651,647
	66.13%	33.87%	100%
Predicted default	167,872	235,836	403,708
	41.58%	58.42%	100%
Total	598,792	456,563	1,055,355
	56.74%	43.26%	100%
	% correct prediction		63.18%

The results seem to indicate that the unknown loan characteristics are more associated with higher-quality loans, but not the highest quality loans. If loans with unknown loan characteristics are systematically less likely to fail than others in disclosed risk categories, studies that omit such loans are biased toward a riskier cohort.

If the loans with unknown characteristics add explanatory value to the model, even left specified as unknown, modeling mortgage default without including such loans could have implications for performance forecast accuracy. Running the two-year model without the loans with unknown characteristics has little effect upon adjusted R-squared, but including such loans substantially increases the pseudo R-squared in the four-year model.

Of course, pseudo R-squared is not a good measure for comparison between two models with different number of regressors and different number of observations. Thus, in Table 1.9, I present comparisons of forecast accuracy from my two-year and four-year models with and without the loans with unknown characteristics. It is immediately apparent that including the loans with unknown characteristics increases the forecast accuracy, from 72.31% in the two-year model without loans with unknown characteristics to 80.48% in the two-year model with loans with unknown characteristics. Similarly, forecast accuracy in the four-year model increases from 63.18% to 70.69%.

Forecast accuracies are also measured with Root Mean Squared Errors (RMSE) in OLS regressions. As shown in Table 1.10, RMSEs are also smaller when unknowns are controlled. Model 3 has a higher RMSE difference than Model 1, showing that the more "unknown" variables controlled for, the better.

Table 1.10: Performance forecast ratios for the two-year and four-year models with and without loans with unknown characteristics using OLS specifications – comparison of root mean squared errors

	2 year default OLS		4 year default OLS	
	With unknowns	Without unknowns	With unknowns	Without unknowns
Model 1	38.4%	38.4%	44.0%	44.7%
Model 2	38.3%	38.6%	43.8%	45.0%
Model 3	38.1%	41.9%	43.6%	47.1%

Such results confirm that models built to accommodate loans with unknown characteristics can be expected to perform better than others, most likely because they address – even if incompletely – the sample bias imputed from omitting such loans.

1.6. Conclusion

In this paper, I explore the causes of the financial crisis by examining how loan-level information was disclosed and reported during the pre-crisis period. Using representative loan-level data from trustees, I show that important loan-level attributes are not reported for substantial non-random portion of securitized loans. Thus, the loan-level sample used in many popular studies may not properly represent the population. In such cases, it is becoming apparent in "big data" research that observational biases can be crucial to academic and policy research on the financial crisis.

I also provide empirical evidence that loan-level disclosure is correlated across data fields and that loan performance is worse than the safest loans and better than the riskiest ones when the loan characteristics are not reported. These findings have two important implications. First, if we blindly drop the mortgages with missing values for major risk factors, then the resulting sample may not be random leading to biased estimations. Second, current regulatory regimes do not address improving disclosure practices in ways that can alleviate this data limitations. Therefore, the SEC is probably

right to pursue increased reporting pursuant to Reg AB II, which will most likely increase investors' ability to model accurately loan performance when investing in future mortgage products.

CHAPTER 2. SELECTION BIAS AND THE ESTIMATION OF SPONSOR AFFILIATIONS IN MORTGAGE RISK

2.1 Introduction

The mortgage securitization process consists of multiple transactions conducted by several institutions that are not necessarily affiliated with each other. For example, for transferring mortgages into the trust, there are typically three entities involved in this single step. As a diagram from the prospectus supplement for BSMF 2006-SL1 shows in Figure 2.1, as the sponsor, EMC acquired or aggregated mortgages from originators. As the depositor, Bear Stearns Asset Backed Securities I built up a "bankruptcy firewall" between EMC and the BSMF Trust 2006-SL1 that issued bonds, establishing a two-stage transfer and thus completing "true sale" of assets[37] Since the depositor and the issuer are limited purpose entities created by the sponsor, this paper refers to the three of the entities jointly as the securitizer.

One important characteristic of securitization is that banks often outsource the underwriting function (Gorton and Metrick, 2012). Policymakers were cognizant of the possibility that affiliations among various parties in securitizations may have substantial influences on market outcomes. Hence, "Affiliations and Certain Relationships and Related Transactions" is one of the basic disclosure items in Regulation AB (Reg AB) adopted by Securities Exchange Commission in 2004.[38]

[37] The sale of mortgages from the sponsor to the trust should be "complete and true" for the trust to be designated as a Real Estate Mortgage Investment Conduit (REMIC). Mortgages in the REMIC trust are isolated from the risk of seizure by creditors in the event of sponsor's bankruptcy. Issuance of bonds though REMIC trusts enables sponsoring banks to enjoy the benefits of inexpensive off-balance sheet debt because the cost of securitized debt financing does not contain a bankruptcy risk premium. See Moody's Investors Service (2002) and Gorton and Souleles (2007).

[38] See 17 CFR 229.1119.

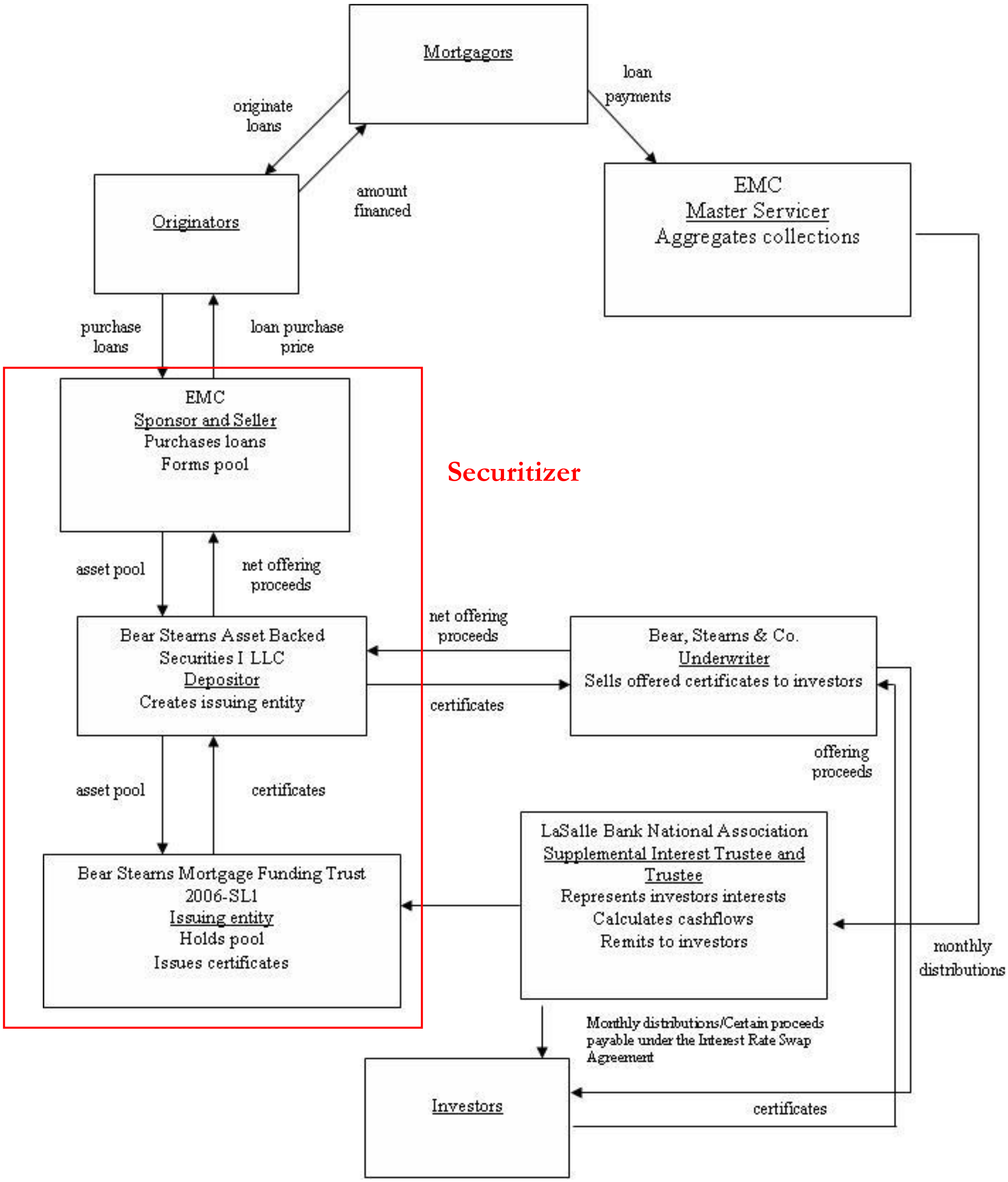

Figure 2.1. Transaction structure of the mortgage securitization of BSMF 2006-SL1

There is also growing academic interest in affiliations. Researchers have examined how the frictions arising from affiliation status contributed to the recent crisis. For example, Demiroglu and James (2012) argue that banks engaged in less screening of applicants for mortgages they planned to sell to unaffiliated securitizers. Titman and Tsyplakov (2010) document that originators cherry-picked less risky mortgages to deposit in their own trusts. The previous analyses of originator-sponsor affiliations rely on the assumption that the identities of originators are correctly reported, and the sample whose affiliation information is available truthfully represents the population.

My paper identifies two sources of bias associated with this assumption. First, the originator information available from prospectus supplements and loan-level datasets is subject to measurement error in that institutions reported as originators may not truly be originators. There are likely multiple institutions such as brokers and loan officers who handle applicants at the front end of origination process[39], banks that temporarily keep mortgages in their portfolio, and institutions specialized in purchase and aggregation of mortgages prior to securitization.

The set of originators reported at the deal level may be erroneous. Reg AB requires securitizers to disclose in the prospectus supplements originators who provide 10% or more of the mortgage pool,[40] however Reg AB provides little guidance with regard to which institution in the origination process should be reported. Hence, sponsors usually report the loan providers they directly transacted with as the originators in their prospectus supplements regardless of whether the loan providers actually closed the mortgages or not. Measurement error for originator identification may be more serious in loan-level data sets voluntarily reported by various institutions across the securitization process. Due to lack of regulation, the reported value for originator name variable

---

[39] In most of the prospectus supplements, the institutions reported as originators are documented to *acquire from third-party originators* as well as originate mortgages in the pool.
[40] See the section concerning originators in p1538 of 33-8518FR.

could be a broker, a correspondent lender, an originating arm of the sponsor, or even the sponsor itself. For example, sponsors like IndyMac and Countrywide consider correspondent lenders as an in-house origination channel while the other sponsors like Wells Fargo treat them as unaffiliated third party originators. The impossibility of identifying originators imply that academics cannot rigorously test the frictions on the part of originators. Institutions identifiable from prospectus supplements and loan-level data sets are loan providers who transfer mortgages to affiliated or unaffiliated sponsors. It is these loan providers that I examine in the analysis of how mortgages are allocated between affiliated and unaffiliated securitizations depending on their credit risk.

The second source of bias is the possibility that the distributions of major variables are truncated in a non-random fashion. First, the sample of mortgages whose affiliation status is available from loan-level data sets may fail to represent the population. As shown later in this paper, regardless of the true affiliations, sponsors tend to report themselves rather than unaffiliated small lenders as the loan providers more often for ex-ante riskier mortgages.[41] A lack of consideration for the selective disclosure of loan providers by sponsors may lead to a biased estimation for the relationship between the default risk of a mortgage and its sponsor-loan provider affiliation. Second, a variety of risk factors and loan attributes are typically controlled in the loan performance model. Academics often drop all the mortgages whose characteristics variables are missing, potentially resulting in another sample selection problem.

This paper aims to explore how the measurement error and the incidental truncation of the identity of loan providers impact the study of two empirical problems: how the affiliation between sponsor and loan provider is associated with 1) loan attributes like FICO, LTV, etc. as the proxy of ex-ante

[41] They may do this because they arguably seek to exploit their reputation capital in order to facilitate the sale of bonds backed by observably riskier loans. However, more investigation is necessary to confirm this hypothesis.

default risk; and 2) post-securitization loan performance as the proxy of ex-post default risk. These two problems are important because they provide clues for two hypotheses of whether banks cherry-pick less risky mortgages for their own securitizations exploiting: 1) their private information[42]; and 2) public information[43] for mortgage quality.

Using a unique dataset where the identities of sponsors and loan-providers are hand-collected, and linked to loan attributes in BBx data (BBx)[44], I show that sponsor-loan provider affiliation is seemingly associated with higher ex-ante and ex-post default risk if the affiliation variable is naively constructed based on the identity of institution reported as the originator in the loan-level data. To check the validity of these results, I construct and use the actual affiliation between the sponsor and the true loan provider identified with a simple algorithm through which the number of correctly identified loan providers increases from 973,298 to 3,326,077, or from 20.3% to 69.4% of the population of mortgages privately securitized in 2006. When the missing or erroneous values for loan provider identities are properly back-filled and corrected, I show that loan providers channel ex-ante riskier mortgages to their own trusts significantly less often than to unaffiliated securitizers. This result that loan providers cherry-pick ex-ante less risky mortgages for their affiliated securitizations also holds in control function method, or two-step procedure (Heckman, 1976; Heckman 1979; Lee 1982) where sponsor's selective disclosure of loan providers is explicitly addressed.

[42] As suggested in Stein (2002), institutions may take advantage of "soft information" for mortgage quality that cannot be transferred to the counterparty of mortgage transactions.
[43] The public information includes loan-level data available to investing public as documented in Piskorski, Seru, and Witkin (2013).
[44] BBx is one of the major loan-level data marketed by BlackBox Logic. For details, see http://www.bbxlogic.com/bbx-logic-US-RMBS-non-agency-solutions.php. More information for the coverage and structure of BBx is provided in the following section.

I also show there is a dramatic discrepancy in correlation between ex-post default risk and the sponsor-loan provider affiliations depending on treatments of missing and erroneous values for affiliation variable. If the originator name variable is naively used for the construction of affiliation between loan provider and the sponsor, then affiliation seemingly increases the likelihood of default. In contrast, with the sample where the loan providers are correctly identified and backfilled, I show that mortgages perform significantly better for affiliated loans than for unaffiliated ones. More interestingly, if the missing indicators for each risk factor are introduced into the model, so if the sample is not sliced and diced based on the availability of variables, then loan providers with capacity of securitizations are shown to transfer ex-post riskier mortgages to their own shelves than to unaffiliated sponsors, which is different from Demiroglu and James' (2012) result that sponsoring ability has little effects on mortgage performance.

The remainder of the paper is organized as follows. Section 2 provides detailed information about the structure and coverage of loan-level data provided by BBx. Section 3 illustrates how loan providers are disclosed and reported in the prospectus supplements and BBx. Section 4 introduces a simple algorithm for the calculation of the affiliation between sponsor and true loan provider. Section 5 contrasts the difference in empirical results depending on whether the missing and measurement error problems for the affiliation variable are addressed or not. Section 5 also provide empirical models and results that the affiliation between loan provider and sponsor is associated with lower ex-ante and ex-post default risk. Section 7 concludes.

2.2 BBx data

I examine how mortgage quality varies with the affiliation between loan providers and sponsors using the loan-level data compiled by Blackbox Logic, LLC (BBx). If we are free from concerns about selective omissions for major risk factors, BBx would be enough to be referred as "Big data" for securitized mortgages. In this section, I show that BBx represents sufficiently large portion of

the private-label mortgage market in terms of the number of loan-level variables, geographical coverage, origination amount, and outstanding balance. BBx contains extensive loan-level information for mortgages securitized by private institutions. Borrowers' credit scores and mortgage underwriting characteristics at the time of origination such as loan-to-value (LTV) ratio, loan balance, initial interest rate, insurance information, a variety of indicators related to purpose, occupancy, documentation, maturity, etc. are available from BBx. Also, BBx provides time-varying information on the history of delinquency, payoff status, and details regarding critical events such as modification, prepayment, and loss from liquidation. As of December 2013, BBx includes the information on 7,480 deals, 21,656,677 loans, and 708,373,906 remittance records.

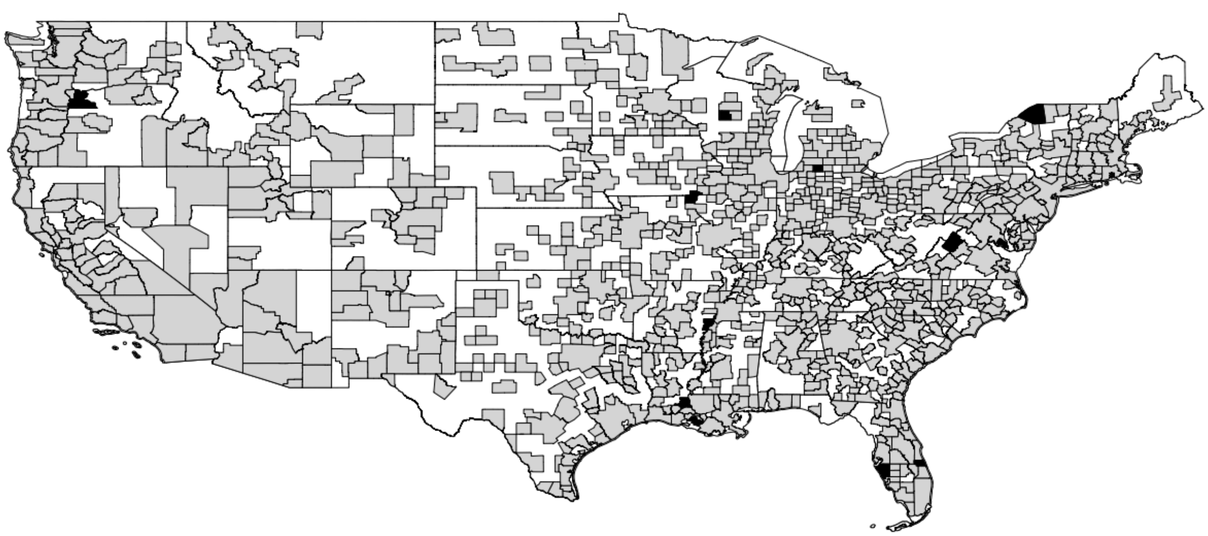

Figure 2.2. Distribution of the properties used as collateral against the mortgages in BBx.

BBx has reasonable coverage of the markets for Jumbo A, Subprime and Alt-A mortgages with senior and junior lien positions as well as prime mortgages. In Figure 2.2, I show how much BBx covers across core based statistical areas (CBSAs) in the United States. To identify the geographical distribution of BBx properties, I merge the location information of properties in BBx and 2003 cartographic boundary files from Census Bureau. Black areas are CBSAs BBx does not cover. Grey

areas represent the CBSAs where BBx properties are located. White areas are not part of a CBSA. As presented in Figure 2.2, virtually every CBSA (920 out of 933) is represented in BBx.

Table 2.1. Comparison of origination amount between BBx and the market

	Total	Non-agency	BBx	% of total population	% of non-agency sector
2005	$3,120	$1,940	$1,146	36.73%	59.07%
2006	$2,980	$1,910	$1,065	35.74%	55.76%

Notes: Figures are in billion.

Table 2.1 contrasts the dollar amount of mortgage originations across the entire mortgage industry, non-agency sector, and BBx in both 2005 and 2006.[45] The 2005 and 2006 sum of origination amount in BBx is respectively 1.146 and 1.065 trillion dollars, which accounts for 59.07% and 55.76% of non-agency sector.

Table 2.2. Comparison of mortgage debt outstanding between non-agency sector and BBx

Measurement date	private conduits population	BBx	BBx coverage
Dec-05	$2,937	$1,340	45.62%
Dec-06	$2,755	$2,061	74.81%
Dec-07	$2,936	$2,393	81.51%
Dec-08	$2,586	$2,066	79.89%
Dec-09	$2,219	$1,722	77.60%
Dec-10	$1,899	$1,420	74.78%
Jun-11	$1,788	$1,304	72.93%

45 The origination figures for total market and non-agency sector are from the 2010 Mortgage Market Statistical Annual.

Table 2.2 compares the mortgage debt outstanding over time between the private securitization industry and BBx. The figures for non-agency sector volume is from the Federal Reserve Board Statistics and Historical data. Outstanding debt amount in BBx is self-calculated. To summarize, 45.62% to 81.51% of outstanding debt amount is covered in BBx from the first quarter of 2005 to the second quarter of 2011.

2.3 Practices of disclosing originators

2.3.1 Deal-level disclosure of originators in prospectus supplements

In response to the need to increase investor protection and facilitate the efficient operation of MBS market, on December 15 2004, the SEC approved Reg AB, which establishes registration, disclosure, and reporting requirements for securitized pool of mortgages. Prospectus disclosures were required to be compliant by Dec 31, 2005, and all shelf registrations were required to conform by Mar 31, 2006.

Specifically, Reg AB requires registrants to disclose the identity of originators who could be important in evaluating risk. According to the disclosure requirements for originators set forth in Item 1110, securitizers need to report the identity of institutions who provide 10% or more of the pool assets and further information[46] for those who provide 20% or more. This "step-ladder threshold for disclosure" is the result of the SEC's effort to balance the public need for information with the risk of over-disclosure.[47] It is notable that there may be multiple institutions within the origination process, and Reg AB provides little guidance about which institution should be reported as the originator. Hence, securitizers usually report the identity of loan providers they directly

[46] Additional items are required to be disclosed for originators larger than 20% including organization forms, underwriting criteria, unfavorable legal proceedings, information of affiliated transaction parties, etc.
[47] See Walworth, Novomisle, and Wetzler, "The Role of Reg AB", New York Law Journal, Jun. 14, 2010.

transact with, and do not trace back the entire mortgage supply chain to identify the original mortgagees.

The most common practice is loan providers are reported as the originator if the outstanding balance of mortgages they provide is larger than 10% of the aggregate principal balance on the cut-off date. Figure 2.3 provides an example of how loan providers are reported by Deutsche Bank in its prospectus supplement for ACE 2006-HE2. Argent, Chapel, and CIT are reported to provide 33.81%, 10.77%, and 10.76% of the pool assets while the loan providers smaller than 10% are categorized to be "various originators" and remain undisclosed.

Originators..........................	Argent Mortgage Company LLC, a Delaware limited liability company, with respect to approximately 33.81% of the mortgage loans, Chapel Mortgage Corp., a New Jersey corporation, with respect to approximately 10.77% of the mortgage loans and CIT Group Inc., a Delaware corporation, with respect to approximately 10.76% of the mortgage loans, in each case, by aggregate principal balance as of the Cut-off Date. The remainder of the mortgage loans were originated by various originators, none of which have originated more than 10% of the mortgage loans. SEE "THE ORIGINATORS" IN THIS PROSPECTUS SUPPLEMENT.

Figure 2.3. Excerpt for disclosure of loan providers in the prospectus supplement for ACE 2006-HE2

Even though Reg AB only requires disclosure of loan providers larger than 10%, some securitizers report the distribution of loan providers in full regardless of the loan amount they provide. Figure 2.4 presents how BancCap Advisors reported loan providers in its prospectus supplement for the deal BASIC 2006-1. Flexpoint Funding, Maribella, and Oak Street are disclosed as loan providers even though their dollar contributions to the composition of the deal are just 3.76%, 8.58%, and 0.46% respectively, which is much smaller than 10%.

ORIGINATORS	NUMBER OF MORTGAGE LOANS	PRINCIPAL BALANCE AS OF THE CUT-OFF DATE ($)	% OF PRINCIPAL BALANCE AS OF THE CUT-OFF DATE	REMAINING TERM TO MATURITY (MONTHS)	MORTGAGE RATES (%)	FICO
<S>	<C>	<C>	<C>	<C>	<C>	<C>
Encore Credit	480	115,877,993.46	54.02	355	7.024	634
Flexpoint Funding	41	8,074,674.24	3.76	358	8.320	602
Funding America	400	71,164,088.18	33.17	357	8.206	622
Maribella Mortgage	98	18,411,072.09	8.58	355	6.986	629
Oak Street Mortgage	7	984,951.42	0.46	360	7.895	611
TOTAL:	1,026	214,512,779.39	100.00	356	7.466	629

Figure 2.4. Excerpt for disclosure of loan providers in the prospectus supplement for BASIC 2006-1

Some securitizers purchase mortgages from a single loan provider. In that case, the name for the trust often includes the name of the single loan provider. For example, in ABFC 2006-OPT2, OPT refers to Option One which provided 100% of mortgages in the deal. Figure 2.5 presents how Option One is reported to be the loan provider in ABFC 2006-OPT2. Notably, the prospectus supplement documents that Option One *acquired* as well as originated the mortgages it provided to ABFC 2006-OPT2, implying that Option One may not be the true originator for some mortgages.

Originator

Option One Mortgage Corporation originated or acquired all of the mortgage loans.

Figure 2.5. Excerpt for disclosure of loan provider in the prospectus supplement for ABFC 2006-OPT2

In some securitizations, sponsors themselves originate loans. As shown in Figure 2.6 for WFMBS 2006-1, Wells Fargo is not only the sponsor but also the loan provider which accounts for 64.42% of the underlying mortgage assets. Loan providers for the remaining 33.58% were not disclosed since, individually, their sizes did not exceed the 10% Reg AB threshold. According to the information regarding origination channel, this 33.58% came from unaffiliated correspondents. Still, Wells Fargo actually originated 41.58% of the pool through its retail loan officers, while 22.84% were acquired from brokers through the wholesale channel, the identities of which are not disclosed.

In this section, I presented how prospectus supplements provide deal-level information about the identity of loan providers who are not necessarily originators who underwrite and create mortgages at the front-end of securitization process. Deal-level patterns of reporting loan providers show why it is not possible to test the hypothesis associated with the opportunism on the part of originators.

Mortgage Loan Data Appearing in Appendix A

The Mortgage Loans were originated by Wells Fargo Bank or its affiliates or purchased from other mortgage lenders. No originator or group of affiliated originators, apart from the Sponsor or its affiliates, has originated or is expected to originate 10% or more (by aggregate unpaid principal balance as of the Cut-Off Date) of the mortgage pool.

ORIGINATORS

Originator	Number	Aggregate Unpaid Principal Balance	Percentage of Total Aggregate Unpaid Principal Balance
Wells Fargo Bank or Affiliate	501	$ 290,192,654.10	64.42%
Other Originators	271	160,255,016.54	35.58
Total	772	$ 450,447,670.64	100.00%

ORIGINATION CHANNELS

Origination Channel	Number	Aggregate Unpaid Principal Balance	Percentage of Total Aggregate Unpaid Principal Balance
Correspondent	271	$ 160,255,016.54	35.58%
Retail	340	187,305,132.03	41.58
Wholesale	161	102,887,522.07	22.84
Total	772	$ 450,447,670.64	100.00%

Figure 2.6. Excerpt for disclosure of loan provider in the prospectus supplement for WFMBS 2006-1

The distribution of loan providers presented in the prospectus supplements is a deal-level information, however it is still useful even for loan-level analysis. I explore the disclosure practices and limitations of loan-level information about loan providers in the next section.

2.3.2 Loan-level disclosure of originators in trustee data

A mortgage's originator is defined to be disclosed at the loan level if the mortgage's originator name is not missing in the loan-level data. Most of the loan-level information has been collected from trustees by data vendors such as BBx from BlackBox Logic, LLC., ABSnet Loan of Lewtan[48], and

48 Lewtan is the subsidiary of Standard & Poor's.

LoanPerformance from First American CoreLogic. These trustee datasets consist of two tiers of loan-level information: “loan tapes” and “remittance data” (Mason, Imerman, and Lee, 2014).

Reg AB does not require any disclosure of originators at the loan level. Therefore, the variable of originator name (ORIGINATORNAME in BBx) frequently indicates loan providers. Furthermore, the originator has been disclosed for a substantially small amount of mortgages. Figure 2.7 shows what portion of loan providers have been disclosed for mortgages depending on whether their trustee is Wells Fargo or not.[49] For the mortgages whose loan level information is obtained from Wells Fargo, the disclosure of loan providers dramatically increased from 2.1% in July 2006 to 62.2% in August 2006 while the disclosure rates remained relatively low at around 10 to 20% for the other trustees. This implies that the loan-level disclosure of loan providers may substantially vary across different trustees.

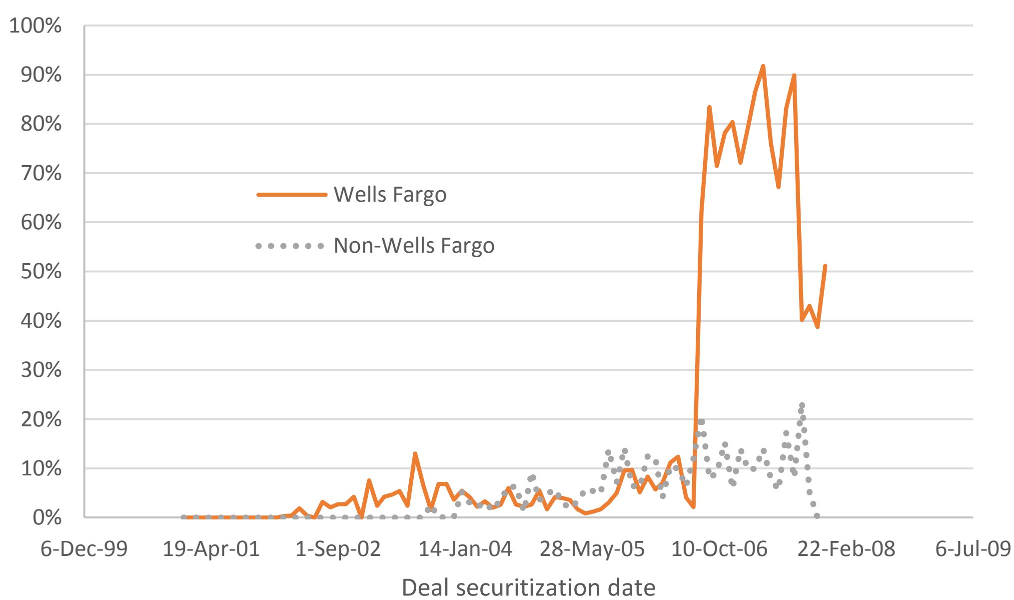

Figure 2.7. The disclosure rate for loan provider identity in BBx by trustees

[49] The “other trustees” includes Deutsche Bank, JP Morgan Chase, Bank of New York, LaSalle Bank, US Bank, Washington Mutual, Citi, Countrywide, Residential Funding Corporation, and IndyMac.

Due to the absence of regulation and voluntary nature of loan-level disclosure, sponsors can choose to report themselves or unaffiliated loan providers as the originator, or even choose not to disclose loan-level origination information. Table 2.3 shows that Taylor, Bean, and Whitaker (TBW) chose three loan-level disclosure options for the deals it securitized in 2006. In the first half of 2006, it rarely reported loan provider identities for securitizations TBW 2006-1, 2, and 3 as opposed to TBW 2006-4, 5, and 6 securitized during the second half of 2006 where originator name is missing for none of the mortgages. In BBx, TBW reported itself as the single originator for TBW 2006-4 and 6, however it fully disclosed all 1,114 brokers who provided loans into TBW 2006-5. This contrasts with information provided in the prospectus supplements for TBW securitizations, where TBW is always reported to be the single originator.

Table 2.3. Loan-level disclosure by Taylor, Bean, and Whitaker in BBx

Deal name	# of originators disclosed	# of all loans	Disclosed institution	Securitization date
TBW 2006-1	0	1196	N/A	30-Mar-06
TBW 2006-2	7	2763	TBW only	27-Apr-06
TBW 2006-3	0	3149	N/A	30-Jun-06
TBW 2006-4	1738	1738	TBW only	30-Aug-06
TBW 2006-5	3001	3001	TBW and 1,114 brokers	26-Oct-06
TBW 2006-6	2565	2565	TBW only	21-Dec-06

Other sponsors have a variety of reporting patterns for ORIGINATORNAME in BBx. Figure 2.8 presents the number of different values for ORIGINATORNAME within a deal. It is immediately obvious that majority of sponsors do not report ORIGINATORNAME at all in BBx. ORIGINATORNAME is missing for every mortgage in the 774 deals. The group of sponsors who make no reports for ORIGINATORNAME includes Countrywide, Residential Funding Corporation, IndyMac, WaMu, and Wells Fargo.

For the 80 deals securitized by Lehman Brothers, Greenwich, Bank of America, JP Morgan Chase, and UBS, sponsors report themselves or other loan providers as the single originator. For the 66 deals securitized by EMC, Deutsch Bank, Morgan Stanley, UBS, and Goldman Sachs, sponsors report more than 30 institutions as originators due to the fact that they usually operate conduits through which they securitize mortgages from numerous brokers and correspondents.

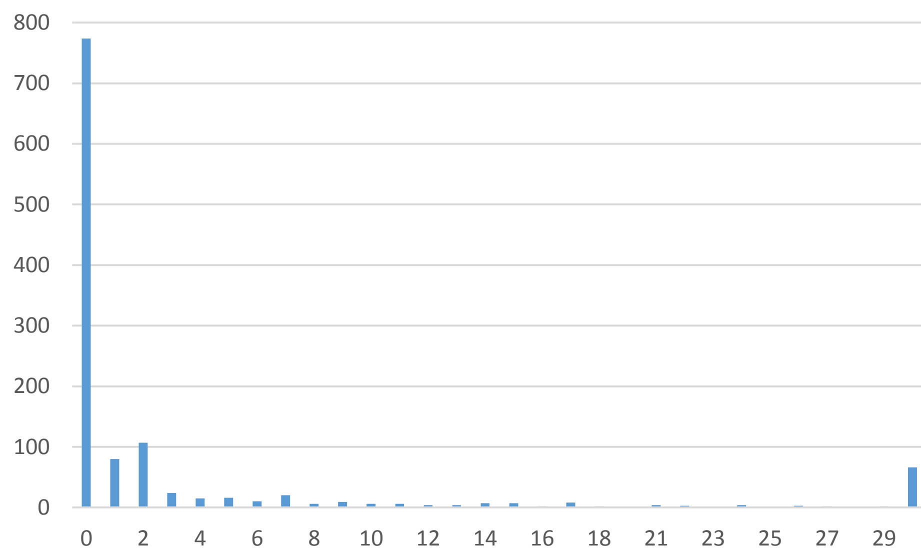

Figure 2.8. Distribution of the number of disclosed loan providers within a deal securitized in 2006

The loan-level disclosure practices suggest that the variable of originator name in the loan-level data alone may not be helpful for calculating sponsor affiliations. In the next section, I introduce a simple algorithm which actively employs information about loan providers not only from loan-level data but also from deal-level prospectus supplements.

2.4 Calculation of true affiliation between loan provider and sponsor

As presented in the previous section for disclosure practices, the institutions reported to be originators are often loan providers who sold mortgages acting as a middleman between original mortgagees and sponsors, therefore they are not necessarily originators. Main purpose of this paper is to examine how these loan providers allocate their mortgages between affiliated and unaffiliated

securitizations. However, it is possible to calculate the affiliation between loan providers and sponsors only when both of them are identified. Sponsor is easily identifiable from prospectus supplements[50], while identity of loan providers is missing for the majority of mortgages in BBx. This may make the sample of mortgages whose ORIGINATORNAME are reported not representative of the population. Therefore, in the examination of how affiliation choices are made by loan providers, it is important to consider the possibility that the exclusion of observations with no ORIGINATORNAME from the sample leads to bias.

One way to deal with potential bias is to reconstruct the population by backfilling and correcting ORIGINATORNAME in the loan level data. In order to recover the identity of loan providers, I develop a simple algorithm. First, identification is straightforward for the deals whose sponsors purchase mortgages from a single loan provider. In this case, the single loan provider's name is available from prospectus supplements or even from their trust name. Among 1,195 deals securitized in 2006, at least for 553 deals, sponsors acquired 2,486,352 (51.86%) mortgages from a single loan provider. For example, for ABFC 2006-OPT1 securitized by Bank of America, the single loan provider is Option One. The five largest sponsors in this category are Countrywide, Lehman Brothers, IndyMac, Long Beach, and Greenwich, and the top five single loan providers are Countrywide, New Century, Option One, First Franklin, and Long Beach.

Second, if the value for ORIGINATORNAME in BBx is identical to one of the institutions reported as the originators in prospectus supplements for the deals with multiple loan providers, then I treat the institution listed in ORIGINATORNAME to be the loan provider. Notably, I do not directly use ORIGINATORNAME to identify loan providers because it may indicate the other

[50] Prospectus supplements are available from SEC Edgar, Bloomberg terminal or data vendors such as ABSnet.

institution who sold mortgages to the loan provider. According to the prospectus supplement for ABFC 2006-HE1, Accredited Lenders, Ameriquest, New Century, Option One, and WMC provided 98% of the loans in the pool. BBx reports 6,791 (88%) loans were from those five loan providers, implying that the other institutions in earlier part of mortgage supply chain were misreported as one of those five loan providers for the other 10%.

Table 2.4. The distribution of originators and primary servicers

	% Origination in Pro Supps	% Servicers in BBx
Bank of America	3.16%	2.25%
GMAC	4.08%	12.55%
National City	13.25%	3.63%
PHH	0.11%	0.08%
RFC	3.84%	3.19%
SunTrust	2.48%	1.89%
Washington Mutual	16.86%	2.33%
Wells Fargo	56.22%	58.04%
PNC		5.53%
JP Morgan Chase		10.52%

Third, if the value for SERVICERNAMECALC, the variable for the identity of primary servicer, in BBx is identical to one of the originators reported in prospectus supplements, then I treat the institution in SERVICERNAMECALC to be the loan provider. This method could be justified by the industry practices that loan providers often maintain their servicing rights even after they sell mortgages to the sponsor. For the deal, ACE 2006-SD3 securitized by Deutsche Bank, ORIGINAORNAME in BBx is 100% missing, however IndyMac and Washington Mutual are reported to provide 21% of the pool in the prospectus supplement, and they service 7% of the pool

according to BBx. Moreover, the distributions of originators in the prospectus supplements and SERVICERNAMECALC in BBx are sometimes relatively close to one another.

Table 2.4 presents the distribution of originators reported in the prospectus supplements and SERVICERANMECALC in BBx for BAFC 2006-1. It is immediately obvious that the two distributions are substantially similar. WaMu has small share of servicing relative to the amount of loans it provided because Chase took over servicing rights after it acquired WaMu in 2008.

Table 2.5. List of sponsors with no capacity for origination

Non-originating sponsors	N loans	% loans
EMC	342846	23.28
Goldman Sachs	251760	17.09
Morgan Stanley	204418	13.88
Deutsche Bank	198378	13.47
Greenwich	146944	9.98
Nomura	73637	5
Carrington	49113	3.33
Sutton Funding	35270	2.39
IXIS	32323	2.19
Barclays	30631	2.08
C-BASS	28583	1.94
Delta Funding	18927	1.29
Societe Generale	16048	1.09
Luminent	11949	0.81
NewCastle	11272	0.77
Saxon	6247	0.42
HomeBanc	3908	0.27
GSC Capital	2776	0.19
Ocwen	2689	0.18
CSE Mortgage	2177	0.15
RWT Holdings	1800	0.12
BancCap Advisors	1026	0.07
Total	1472722	100

Fourth, I can always treat loan providers and sponsors as unaffiliated so long as the sponsor does not have any in-house origination platform. For the sample of mortgages securitized in 2006, sponsors with no capacity for origination securitized 1,472,722 (30.7%) loans.

Table 2.5 presents the list of non-sponsoring originators and their market shares. EMC, Goldman Sachs, Morgan Stanley, Deutsche Bank and Greenwich are the five largest sponsors who exclusively acquired mortgages on flow or bulk basis from unaffiliated loan providers.

Fifth, if the institution reported as the originator in BBx or in prospectus supplements is just the mortgage acquisition channel of the sponsors, then loan provider is treated to be missing even though ORIGINATORNAME is not missing. ORIGINATORNAME indicates conduit or purchase programs for 44,664 loans. We can observe this type of loans in the shelves of Bear Stearns, Morgan Stanley, Goldman Sachs, and UBS. Table 2.6 shows how major conduit sponsors report ORIGINATORNAME variables in BBx.

Table 2.6. Disclosure for conduits

Sponsor	Value for ORIGINATORNAME	Bloomberg ID for trusts
EMC Mortgage Corp	EMC, EMC FLOW, EMC RESIDENTIAL, EMC MORTGAGE CORP	BSABS 2006-HE3, HE4, HE10 , SD4, BSMF 2006-AR3, 4, 5, 8, BALTA 2006-8, EMCM 2006-A, SACO 2006-3, and 10
Morgan Stanley Mortgage Capital	ASSURANCE PARTNERS BANK (CONDUIT), CENDANT CONDUIT, FNBA CONDUIT, GREENPOINT CONDUIT	MSM 2006-13AX, 15AX, 16AX, MSM 2006-12XS, 15XS, and 17XS
Goldman Sachs Mortgage Company	GS CONDUIT	GSAA 2006-14, 15, 16, 17, 19, 20, and GSR 2006-8F
UBS Real Estate Securities	Conduit, UBS conduit	MALT 2006-1, MASL 2006-1

Among 4.8 million mortgages securitized in 2006, ORIGINATORNAME is populated only for 973,298 (20.3%) loans. However, through this simple algorithm, I can confidently identify 3,326,077 (69.4%) loan providers. Furthermore, I can calculate the affiliation status for 3,904,931 (81.5%) loans.

2.5 Empirical models and results

The paper examines how loan providers choose between affiliated and unaffiliated securitizations. The loan provider's choice of affiliations can be associated with either ex-ante or ex-post default risk. For the relation between sponsor-loan provider affiliation and the ex-ante default risk, I examine how various ex-ante risk factors such as FICO, LTV, etc. affect the likelihood of affiliation between the loan provider and the sponsor. To answer the question of how the affiliation is related to ex-post risk, I estimate the effect of affiliation on the post-securitization mortgage performance.

I start the empirical analysis of the relation between sponsor-loan provider affiliation and default risk with the presentation of summary statistics in Table 2.7. Column 1 shows sample means for major attributes for the entire group of 4,793,923 mortgages securitized in 2006. These approximately 4.8 million mortgages are broken down to two groups based on whether ORIGINATORNAME is missing or not in columns 2 and 3. 973,298 mortgages in column 2 are divided again to two groups in columns 4 and 5 depending on whether the institution in ORIGINATORNAME is affiliated with the sponsor. I identified true loan providers for 3,904,931 mortgages using a simple algorithm introduced in section 2.4, which are categorized based on actual affiliation between loan provider and sponsor, and accordingly grouped in columns in 6 and 7 in the table.

Table 2.7 shows interesting variation in risk factors and loan attributes. Average FICO scores are higher, and simple LTVs and interest rates are lower when ORIGINATORNAME is reported. The risk factors substantially vary depending on the measure of affiliation as well.

Table 2.7. Averages for major numeric loan attributes by disclosed affiliation

	(1) Total	ORIGINATORNAME reported		Reported to be affiliated		Affiliated confirmed	
		(2) Yes	(3) No	(4) Yes	(5) No	(6) Yes	(7) No
FICO	668.26	662.44	670.26	661.47	662.66	680.53	653.75
Simple LTV	68.22%	72.36%	67.17%	70.58%	72.76%	64.91%	71.43%
Combined-lien LTV	84.96%	83.94%	85.30%	85.35%	83.61%	83.79%	84.42%
Interest rate	8.00%	8.18%	7.96%	8.50%	8.11%	7.66%	8.39%
Term	338.2	338.6	338.1	343.3	337.6	345.6	335.2
Property appraisal value	349127.7	318068.4	356871.8	315225.3	318792.9	384620.8	312779.6
Loan amount	221701.2	222027.0	221618.1	229673.6	220304.4	238735.1	205542.1
2nd lien	21.6%	25.0%	20.8%	25.7%	24.8%	22.0%	22.2%
Low doc	36.1%	44.7%	40.4%	45.6%	44.5%	45.9%	34.3%
Balloon	21.3%	25.8%	20.1%	24.7%	26.1%	17.9%	26.3%
IO	20.6%	26.0%	19.2%	26.0%	26.0%	16.0%	20.5%
Investment loan	9.2%	9.1%	9.2%	10.1%	8.8%	7.9%	7.9%
Cash-out refinancing	31.3%	29.6%	31.7%	32.4%	29.0%	40.4%	29.2%
N loans	4,793,923	973,298	3,820,625	178,939	794,359	1,558,176	2,346,755

When the institution in ORIGINATORNAME is affiliated with the sponsor, mortgages have lower FICO scores, and higher combined-lien LTV and coupon rate. In contrast, FICO scores are higher and simple LTVs, combined-lien LTVs, and coupon rates are lower when the loan provider is affiliated with the sponsor. This suggests that inferences on the relation between sponsor affiliations and ex-ante risk factors may be dramatically different depending on the measure of affiliations.

2.5.1. Affiliation and ex-ante default risk

The first empirical question of the paper is how the loan provider's affiliation choices are affected by a variety of loan characteristics. This relation is important because it provides some insight into the question of whether loan providers cherry-pick ex-ante less risky mortgages for their own securitizations. The variation in average risk factors across different measures of affiliation suggests that it may not be desirable to naively use the variable of originator name that is mostly missing in the loan-level data sets. Therefore, I explore how and why the empirical relation between the affiliation and ex-ante risk factors varies depending on whether the missing problem is addressed or not.

2.5.1.1 Estimation of affiliation determination without consideration of selective disclosure

I begin the empirical work with the affiliation calculated based on ORIGINATORNAME in BBx without backfilling missing values nor correcting incorrect values for ORIGINATORNAME. This affiliation variable equals 1 if the institution in ORIGINAOTRNAME transfers the mortgage to its affiliated sponsor. I call this affiliation "reported affiliation." If we naively use the reported affiliation, this means we do not effectively address the limitations generally inherent in loan-level data sets which lack substantial amount of information in a non-random fashion.

Table 2.8 shows the results for an OLS regression of reported affiliation on various mortgage risk characteristics. The sample is restricted to 726,067 first lien mortgages where ORIGINATORNAME in BBx is populated. I focus on first-lien mortgages to correctly measure the

effect of combined-lien loan-to-value (CLTV) ratio. The dependent variable equals one if the institution in ORIGINATORNAME is affiliated with the sponsor and zero otherwise.

Table 2.8. One-shot OLS for reported affiliation without consideration of selective disclosure

DV=1 if the loan provider is reported to be affiliated with the sponsor; 0 otherwise				
	Model 1	Model 2	Model 3	Model 4
FICO < 620 (d)	0.0069***	0.0148***	0.0102***	0.0082***
	(6.188)	(12.723)	(8.500)	(6.904)
620 ≤ FICO < 680 (d)	-0.0002	0.0049***	0.0037***	-0.0022**
	(-0.224)	(4.395)	(3.293)	(-1.964)
FICO unknown (d)	0.0424***	0.0824***	0.0577***	0.1114***
	(10.806)	(20.671)	(14.504)	(27.454)
80 ≤ CLTV < 100 (d)	0.0171***	0.0141***	0.0186***	0.0118***
	(14.752)	(12.228)	(16.148)	(10.321)
100 ≤ CLTV (d)	0.0316***	0.0294***	0.0185***	0.0062***
	(20.557)	(19.149)	(12.076)	(4.019)
CLTV unknown (d)	-0.0238***	-0.0075***	-0.0276***	-0.0378***
	(-18.005)	(-5.092)	(-18.174)	(-24.723)
ARM (d)		0.0467***	0.0533***	0.0408***
		(47.067)	(52.407)	(37.717)
Low doc (d)		0.0153***	0.0192***	0.0062***
		(14.622)	(18.475)	(5.862)
Doc unknown (d)		-0.0370***	0.0001	0.0051***
		(-25.185)	(0.075)	(3.329)
Balloon (d)			0.0145***	0.0073***
			(8.802)	(4.418)
Balloon unknown (d)			0.0304***	0.0356***
			(19.535)	(23.007)
IO (d)			-0.0064***	-0.0076***
			(-5.706)	(-6.720)
IO unknown (d)			-0.1414***	-0.1395***
			(-87.702)	(-86.913)
Originated before 2004 (d)				-0.1359***
				(-75.295)
Originated in 2004 (d)				-0.1530***
				(-79.098)
Originated in 2005 (d)				-0.0912***
				(-75.687)
Constant	0.1713***	0.1275***	0.1152***	0.1549***
	(175.574)	(95.969)	(55.670)	(72.192)
Adj. R-sq	0.002	0.007	0.022	0.032
N Obs	726067	726067	726067	726067

I present four different specifications which vary with the number of controls ranging from only two controls (FICO credit scores and CLTV ratio) in model 1 to seven controls including various loan types associated with interest rate, level of documentation, balloon loan features, and interest-only (IO) features.

I use dummy variables for continuous numeric attributes like FICO and CLTV to flexibly address potential non-linearity. Notably, the base case is always set as the least risky category. Hence, the base cases are mortgages with FICO higher than 680, LTV at origination less than 80, fixed rate, full documentation, no balloon, and no IO. I present t-statistics in parenthesis below the coefficient estimates. I use White-Huber sandwich estimator for standard errors to address potential heteroskedasticity.

Under the one-shot model with no consideration of the potential for selective disclosure of loan provider identity by sponsors, ex-ante risk generally seems to increase the likelihood of affiliation between the loan provider and the sponsor. For example, loan providers channel mortgages with FICO scores less than 620 to their affiliated securitizations more often (0.7% - 1.5%) than they do those with FICO scores higher than 680. Likewise, the affiliation is similarly more likely for the mortgages with high CLTV, adjustable rate mortgages (ARM), low-doc and balloon mortgages. This positive correlation between ex-ante risk characteristics and affiliation, nevertheless, is based on how sponsors *report* their affiliations with loan providers, which may not provide an accurate picture on how loan providers make choices between affiliated and unaffiliated securitizations.

Although this model for reported disclosure does not lead to unbiased estimation for the relation between ex-ante risk factors and the affiliation choices made by loan providers, it provides some insight into the potential source of selection bias. Table 2.7 shows that sponsors report themselves as originators more often for the mortgages with ex-ante riskier underwriting characteristics. One

potential reason for this selective disclosure is that the sponsor can more easily issue and sell bonds backed by suspicious-looking mortgages when they carry the brand names of sponsors of large banks. This reasoning is consistent with the complaint lodged against E*Trade that they allegedly reported many of its loans to be originated in-house rather than truthfully disclosing unaffiliated and notorious loan providers. As sponsors continue to exploit their reputational capital, loan characteristics become more and more risky for the mortgages whose sponsors and originators are reported to be affiliated. Thus, the correlation between affiliation and risk characteristics could be seemingly and spuriously positive in the selected sample of mortgages whose originators are reported.

2.5.1.2. Estimation of affiliation determination based on sample reconstruction

In the analysis of how loan providers choose to sell among different securitizations, there are two distinct sources for bias. First, given the absence of regulation for loan-level disclosure of originator identity, sponsors provide information on originator name to BBx with the names of virtually any institutions in the origination process, leading to serious measurement errors. Second, since it is possible to calculate affiliations only for the mortgages whose loan provider is disclosed, the analysis may be vulnerable to selection bias if sponsors do not randomly disclose the identity of loan providers. To address these problems, I reconstruct my sample by backfilling and correcting the identity of loan providers based on the algorithm that was presented in section 2.4. The measure of affiliation based on the algorithm in section 2.4 is called "actual affiliation."

Table 2.9 shows the impact of ex-ante risk characteristics upon actual affiliation in OLS regression setting. The sample contains 2.1 million first lien mortgages whose loan provider is identifiable though the algorithm. The dependent variable is defined to be one if the loan provider I identified is affiliated with the sponsor and zero otherwise. I use an identical set of controls employed in the naïve approach in section 2.6.1.

Table 2.9. One-shot OLS for actual affiliation

DV=1 if the loan provider is actually affiliated with the sponsor; 0 otherwise				
	Model 1	Model 2	Model 3	Model 4
FICO < 620 (d)	-0.0804***	-0.0527***	-0.0353***	-0.0357***
	(-104.319)	(-65.271)	(-43.017)	(-44.795)
620 ≤ FICO < 680 (d)	-0.0456***	-0.0302***	-0.0211***	-0.0264***
	(-57.157)	(-37.803)	(-27.013)	(-34.509)
FICO unknown (d)	0.3221***	0.3223***	0.3061***	0.3514***
	(207.442)	(213.330)	(210.986)	(248.425)
80 ≤ CLTV < 100 (d)	-0.0677***	-0.0537***	-0.0443***	-0.0490***
	(-75.412)	(-60.742)	(-52.063)	(-59.548)
100 ≤ CLTV (d)	-0.0261***	-0.0128***	-0.0154***	-0.0242***
	(-22.940)	(-11.436)	(-14.155)	(-22.988)
CLTV unknown (d)	-0.2953***	-0.2375***	-0.2589***	-0.2311***
	(-353.874)	(-275.079)	(-295.997)	(-265.367)
ARM (d)		-0.1310***	-0.1127***	-0.1151***
		(-188.722)	(-161.859)	(-167.163)
Low doc (d)		0.0090***	0.0051***	-0.0080***
		(12.132)	(6.977)	(-11.335)
Doc unknown (d)		-0.1584***	-0.1305***	-0.1283***
		(-191.258)	(-148.022)	(-148.468)
Balloon (d)			-0.2546***	-0.2572***
			(-279.260)	(-283.650)
Balloon unknown (d)			-0.1099***	-0.0748***
			(-142.308)	(-98.852)
IO (d)			-0.0467***	-0.0435***
			(-63.021)	(-59.764)
IO unknown (d)			-0.2683***	-0.2301***
			(-270.692)	(-230.837)
Originated before 2004 (d)				-0.2974***
				(-168.937)
Originated in 2004 (d)				-0.3178***
				(-126.358)
Originated in 2005 (d)				-0.2077***
				(-308.713)
Constant	0.4934***	0.5706***	0.6989***	0.7340***
	(619.523)	(552.434)	(557.639)	(599.963)
Adj. R-sq	0.094	0.124	0.164	0.204
N Obs	2,190,202	2,188,736	2,188,736	2,188,736

All coefficients are negative and significant, suggesting that loan providers partition their mortgages into two groups depending on the ex-ante risk. Also, ex-ante less risky mortgages were transferred to loan providers' own securitizations while ex-ante riskier mortgages are sold to unaffiliated sponsors. The riskier ranges of FICO and CLTV reduce the likelihood of sponsor-loan provider affiliation respectively by up to 8% and 6.7%. ARMs are deposited into affiliated trusts less often by 13.1% than FRMs. Exotic payments structures such as balloon and IO decreases affiliation by 26% and 4.7%. Additionally, the estimates of vintage fixed effects in model 4 show that loan providers make affiliation choices more often for recently closed mortgages. Notably, lending standards have monotonically deteriorated in the years leading up to the crisis (Demyanyk and Van Hermert, 2011).

The effects of some ex-ante risk characteristics attenuate when other attributes are added into the model. However, they are still economically and statistically significant even in model 4. Interestingly, the adjusted R-squared is substantially larger when I use actual affiliation than when I use reported affiliation. For example, the adjusted R-squared for full model increases from 0.032 in Table 2.8 to 0.204 in Table 2.9.

2.5.1.3 Estimation of affiliation determination based on control function approach

Calculation of the affiliation between loan provider and sponsor depends crucially on whether the identity of loan provider is available or not. This means there could be another layer of choice associated with sponsors' disclosure of loan provider as well as the choice of affiliation made by loan providers. Therefore, I divide the problem into two parts:

1) $D^* = BX + \Gamma Z + \varepsilon$

2) $A = \Pi X + \mu$

Equation 1 is the selection equation which models mortgages' selection into the group whose loan providers are disclosed by sponsors. The dependent variable D^* in equation 1 is sponsor's utility-

maximizing propensity of disclosing loan providers, or latent index of disclosure whose observable counterpart $D = 1$ if $D^* > 0$ and $D = 0$ otherwise. X is a vector of ex-ante default risk factors including the dummies for FICO, LTV, ARM, lien, and documentation. Z is a vector of variables excluded from equation 2. Z contains the number of loans in the deal, the indicator of whether the sponsor acquires mortgages from a single loan provider, and the quarterly deal vintages. These three variables are expected to affect the disclosure but do not otherwise directly affect the affiliation. The cost of tracking loan providers will be lower for the pools with a smaller number of mortgages. If the sponsor acquires a group of loans from a single provider, then there will be no need for additional disclosure at the loan level since the identity of the single loan provider should be available from the prospectus supplements or from the trust name. Disclosure choices are expected to vary with deal securitization dates as shown in Figure 6 where the disclosure rate has increased during the years leading up to the crisis.

Equation 2 is the equation of interest which illustrates how loan providers make affiliation choices. The dependent variable A equals one if the identity of a loan provider reported in ORIGINATORNAME is affiliated with the sponsor, and zero otherwise. Importantly, A is "incidentally truncated" (Greene, 1990) in that it is observed only when the sponsor decides to disclose the loan provider. Sponsors may exploit their own reputational capital by reporting themselves as the originators for observably riskier mortgages as shown in section 2.5.1, which leads to affiliated mortgages having ex-ante riskier characteristics. In other words, equation 2 cannot be identified on its own because the non-zero correlation between μ and ε.

The system of equations 1 and 2 are often called Type II Tobit model (Amemiya, 1985). Wooldridge (2010) discuss the estimation of this model based on the following assumptions:

a) (X, Z, D) are always observed, while A is observed only when $D = 1$,

b) (ε, μ) is independent of (X, Z) with zero mean,

c) $\varepsilon \sim N(0,1)$, and

d) μ is a linear projection of ε

Following the two-step procedure developed by Heckman (1979) and Lee (1982), I use a function that controls for selection bias, which is also called "Inverse Mills ratio" or "control function". This control function represents the component of μ associated with selective disclosure in equation 1 which arguably causes the omitted variable problem in equation 2, and thus is the source of selection bias (Heckman, 1979). I estimate the control function based on equation 1 using all the mortgages securitized in 2006, which is embedded into equation 2.[51]

Table 2.10 presents the marginal effects estimates for equations 1 and 2. I use Heckprob procedure in Stata because the dependent variable in the outcome equation is binary. The correlation between two residual terms from selection and outcome equations is estimated to be 0.297 with standard errors of 0.005. This statistical significance of correlation estimate means there does exist the sample selection problem. The positive value of correlation estimate implies the existence of some unobserved factor which moves disclosure and affiliation in the same direction.

Column 1 corresponds to probit regression of sponsors' disclosure choices upon ex-ante risk characteristics in equation 1. The selection equation results in Column 1 provides evidence that sponsors do disclose the identity of loan providers more often for ex-ante riskier mortgages, consistent with the notion that sponsors are pushed to provide information about loan provider identity for mortgages with worse underwriting characteristics.

[51] This procedure is often called "Heckit" after Heckman (1976).

Table 2.10. Two-stage control function estimation

	(1) Disclosure	(2) Affiliation
FICO < 620 (d)	0.0185***	-0.0005
	(12.662)	(-0.301)
620 ≤ FICO < 680 (d)	0.0064***	-0.0059***
	(5.125)	(-4.067)
FICO unknown (d)	-0.1848***	-0.0708***
	(-155.157)	(-30.048)
80 ≤ LTV < 100 (d)	0.0276***	-0.0037**
	(22.818)	(-2.575)
100 ≤ LTV (d)	0.0750***	-0.0191***
	(31.787)	(-8.444)
LTV unknown (d)	0.0982***	0.0019
	(21.661)	(0.438)
ARM (d)	0.0601***	0.0080***
	(49.879)	(5.207)
2nd lien (d)	0.0790***	-0.0001
	(41.491)	(-0.052)
Lien unknown (d)	-0.2098***	-0.0728***
	(-242.518)	(-27.331)
Low doc (d)	-0.0100***	-0.0048***
	(-7.734)	(-3.044)
Doc unknown (d)	0.0263***	-0.0143***
	(16.064)	(-8.088)
Number of loans in the pool	-0.0000***	
	(-39.448)	
Provided by a single LP (d)	-0.0830***	
	(-74.790)	
Securitized in 2006 Q1 (d)	-0.1956***	
	(-197.737)	
Securitized in 2006 Q2 (d)	-0.2014***	
	(-197.516)	
Securitized in 2006 Q3 (d)	-0.0450***	
	(-38.902)	
N Obs	478,032	104,087

According to the results in section 6.1., sponsors may take advantage of the brand value of their originating arms to securitize those suspicious-looking loans. Disclosure is more likely for the mortgages with FICO less than 620 by 1.85% compared to those with FICO higher than 680. Increase in LTV ratio from the lowest group under 80 to the highest group above 100 is associated with more frequent disclosure by 7.5%. Disclosure rate is also higher for mortgages with adjustable rate and second lien than for their less risky counterpart by 6 to 8%. The relation between the number of loans in the pool and the disclosure incidence is not economically significant, however the disclosure rate is lower for the single loan-provider deals than for the others by 8.3%. The timing of securitization and disclosure also significantly affects sponsors' disclosure choices. Disclosure frequency decreases by 4% to 20% when the pool of mortgages were securitized in the quarters earlier than the 4Q 2006.

Column 2 in Table 2.10 shows the marginal effects estimates of ex-ante risk characteristics upon affiliation choices made by the loan providers. Affiliation is more likely for the mortgages with ex-ante riskier characteristics, which contrasts with the results from the naïve approach without consideration of selection bias in section 2.5.1.1, and consistent with the results based on sample reconstruction in section 2.5.1.2. The likelihood of reported affiliation is the lowest for the intermediate level of FICO scores, and the change in LTV from the least risky to the most risky range reduces the reported affiliation by 1.9%.

2.5.2. Affiliation and ex-post default risk

I now examine how the post-securitization loan performance is associated with whether or not the loan provider is affiliated with the sponsor. This empirical relation is important to examine whether loan providers channel ex-post less risky mortgages for their own securitizations.

2.5.2.1. Estimation of affiliation effects on loan performance without consideration of selective disclosure

Researchers may naively use ORIGINATORNAME variable in BBx to identify loan providers and their affiliation with the sponsors. Alternatively, I calculate affiliation using the algorithm presented in section 2.4. To investigate the possibility of bias arising from naively using loan-level disclosure of loan providers, or ORIGINATORNAME in BBx, I illustrate how the correlation between the post-securitization loan performance and the sponsor-loan provider affiliation varies depending on different methods of identifying affiliations. Figure 2.9 shows what happens if the naïve measure of affiliation is used.

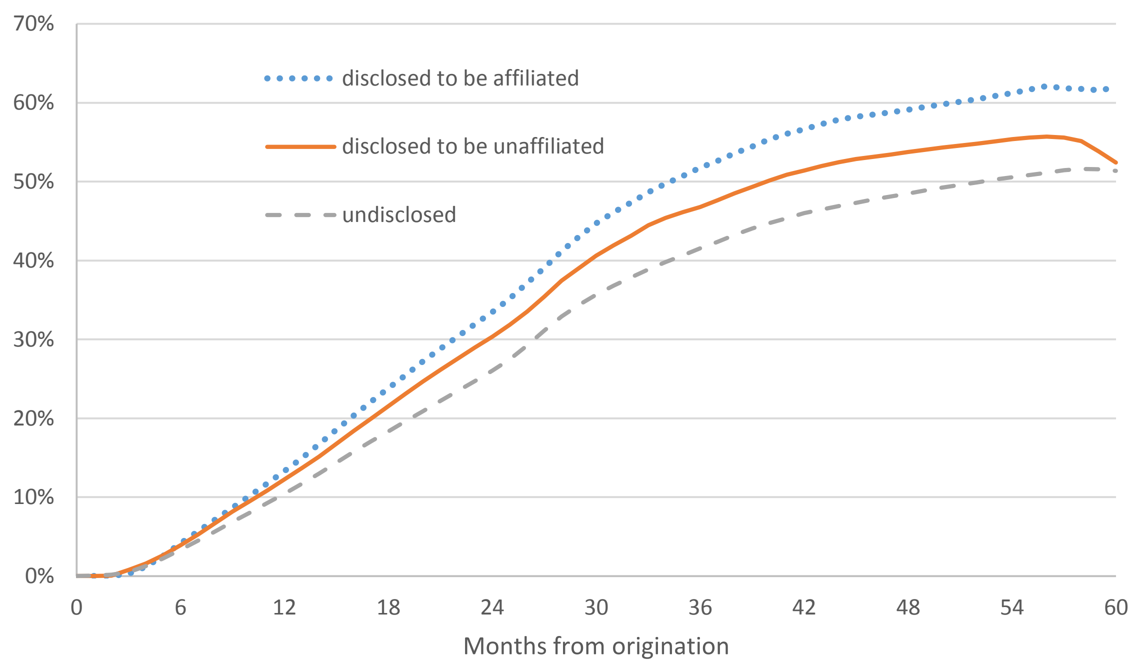

Figure 2.9. Cumulative serious delinquency rate over the life of mortgage by disclosed affiliation

I define a mortgage's originator and sponsor as being "disclosed to be affiliated" if the originator identity is not missing in the loan-level BBx and the originator is reported to be the sponsor's subsidiary, parent, or the sponsor itself. A mortgage is considered to be seriously delinquent if it falls into 60 days in arrears or worse status. Mortgages become seriously delinquent the most often when

their originators are reported to be affiliated with the sponsor. Given the naïve metric of affiliation, mortgages perform the best when their originator identity information is missing in the trustee data. If I rely on the reported affiliation, the loan-provider and sponsor affiliation seems to increase the likelihood of mortgage distress.

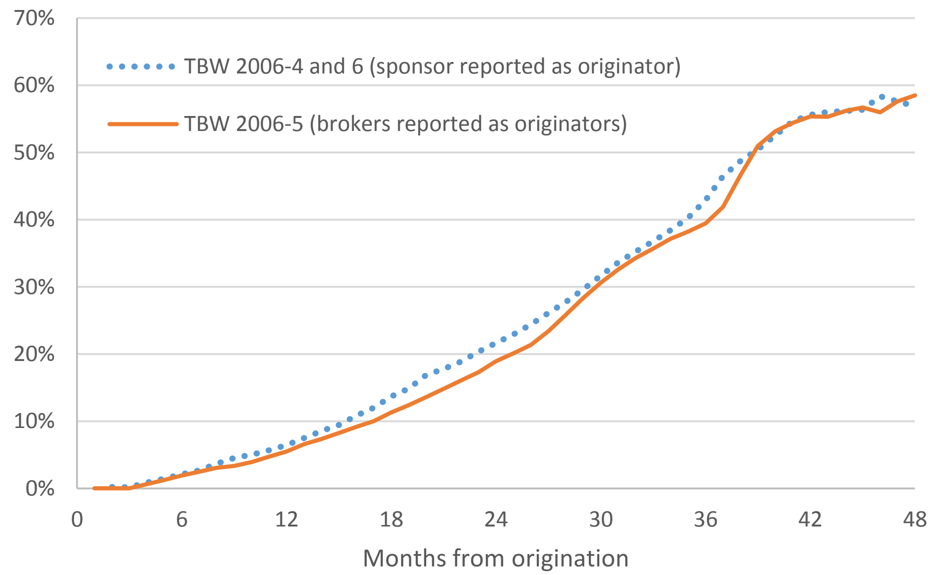

Figure 2.10. Cumulative serious delinquency rate over the life of mortgage for TBW securitizations

For a robustness check, I use the sample of mortgages securitized by TBW in 2006 that shows the most dramatic differences in reporting ORIGINATORNAME in BBx across different deals. I include mortgages in TBW 2006-4 and 6 for which TBW reported itself as the originator in BBx, and those in TBW 2006-5 for which TBW reported one of 1,104 brokers and correspondents as the originator. Hence, the loan provider and the sponsor are measured to be affiliated for TBW 2006-4 and 6 while it appears to be zero affiliation for the mortgages in TBW 2006-5. Figure 2.10 confirms the positive correlation between affiliation and the likelihood of defaults by showing that the mortgage performance is significantly worse at 5% level from 8th to 28th month from origination

when originators and sponsors are reported to be affiliated than when they are reported to have no affiliation.

2.5.2.2. Estimation of affiliation effects on loan performance based on sample reconstruction

The positive correlation between affiliation and mortgage failure, however, is not consistent with the findings of Demiroglu and James (2012) and Titman and Tsyplakov (2010) that affiliation decreases defaults. Thus, I conduct the same examination using actual affiliation based on the true identity of loan providers derived from the algorithm in section 2.4. Figure 2.11 presents the expected pattern seen in previous studies where the mortgages whose loan providers and sponsors are affiliated perform better than unaffiliated loans.

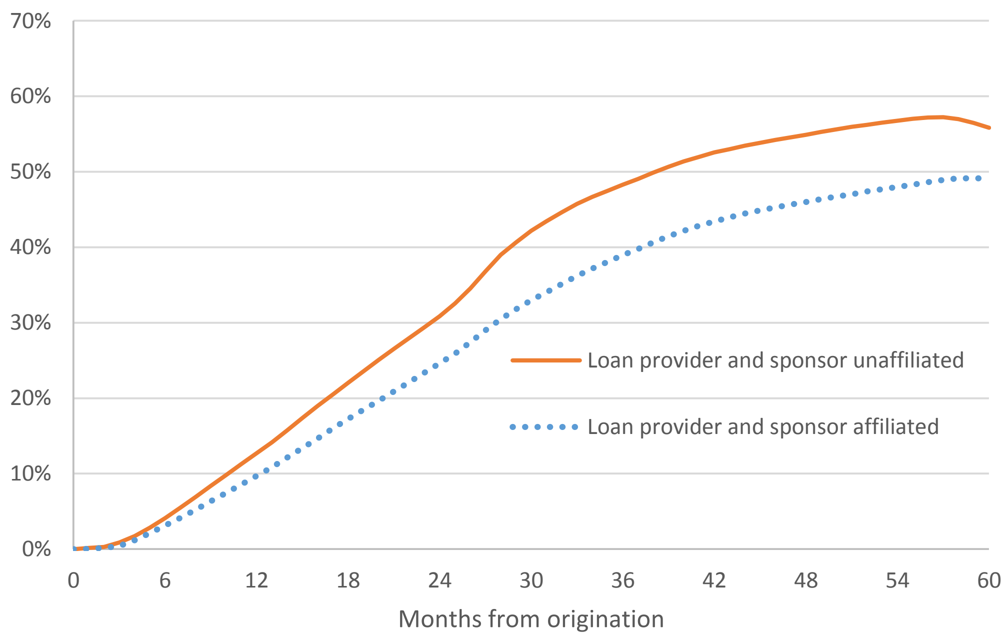

Figure 2.11. Cumulative serious delinquency rate over the life of mortgage by actual affiliation

The estimated relation between post-securitization mortgage performance and sponsor-loan provider affiliation varies depending on the measure of affiliation. This suggests that if I rely only on the loan provider identity available from loan-level trustee data for the construction of affiliation status, I may obtain biased results. There could be two reasons for the positive correlation between

reported affiliation and the likelihood of defaults. First, MBS investors may demand more information regarding originator identity for riskier loans, leading to better performance of undisclosed loans than disclosed loans. Second, sponsors may exploit their reputational capital to form and sell MBS backed by the riskiest loans. According to a related lawsuit brought against E*Trade, the company allegedly concealed from investors that it "purchased loans from troubled subprime lenders such as National City, GMAC, Countrywide, Opteum, Inc. and Fremont General – who had become notorious for poor underwriting standards, illegal practices, government investigations, delinquencies, and the mortgage crisis."

2.5.2.3. Estimation of loan providers' sponsoring ability on loan performance

Underperformance of unaffiliated mortgages can be interpreted in two ways. First, loan providers as the originators may screen loan applicants more diligently for the mortgages they use to feed their own securitizations. Second, loan providers may cherry-pick less risky mortgages for their own securitization. According to Demiroglu and James (D & J, 2012), these two competing hypotheses are testable by examining the relation between loan providers' sponsoring ability and loan performance. Loan providers with no capacity of securitization are not incentivized to cherry-pick good mortgages because they do not have their own shelves to feed. Hence, those loan providers always randomly pass through mortgages to unaffiliated sponsors. If the mortgages from loan providers with their own shelves are as safe as the mortgages from loan providers with no shelves, it is possible to argue that the former do not sell "lemons" to unaffiliated sponsors.

D & J argue lax screening for unaffiliated mortgages because they find loan performance is not affected by loan providers' ability to securitize. However, they dropped all mortgages whose risk factors are missing from their sample. The inclusion of those dropped mortgages makes a substantial difference in the estimated effects of loan providers' sponsoring ability.

Table 2.11. The effects of sponsoring ability on loan performance

	(1) Demiroglu and James	(2) This paper
Loan provider affiliated with sponsor (d)	0.780***	0.864***
	(-13.08)	(-3.14)
Loan provider cannot sponsor (d)	0.977	2.083***
	(-0.75)	(9.85)
Missing indicators included	N	Y
Risk factors controlled	Y	Y
House price index controlled	Y	Y
Loan vintage controlled	Y	Y
N	373,871	206,679[52]

Table 2.11 presents the estimates of odds ratio from the logit regression of loan performance. The effects of affiliation and sponsoring ability estimated by D & J are presented in column (1). In column (2), I replicate D & J's model with the sample of mortgages in BBx. Missing indicators for a variety of risk factors and loan attributes are included in model (2) while D & J dropped all the mortgages with missing attributes in model (1). The effects of affiliation are qualitatively similar between (1) and (2). However, as shown in (2), mortgages from the loan providers with no shelves such as brokers default significantly more often than those from the loan providers who can securitize.[53] Significantly positive effects of sponsoring ability on mortgage performance suggests that loan providers may not randomly pass through mortgages to unaffiliated sponsors.

[52] Although I retain all the mortgages whose variables are missing, the size of my sample for model (2) is still smaller than D & J's. This is because my sample consists only of 7 series (BOAA, BALTA, CWALT, DBALT, JPALT, NCAMT, and WMALT) presented as examples in D & J's table 1. This positive association between loan provider's sponsoring ability and the likelihood of default is robust for the entire sample of mortgages securitized in 2006.

[53] This is not surprising given Jiang, Nelson, and Vytlacil's (2010) results that brokered mortgages are associated with "borrower information falsification."

2.6 Conclusion

This paper sheds light on the potential bias in the analysis of how loan providers divide their mortgages between their own securitizations and unaffiliated sponsors. Through detailed examination of current practices of disclosing loan providers in prospectus supplements and in a large loan-level dataset, I show that naïve use of reported loan provider identity could lead to substantial measurement errors and selection bias. Without proper consideration of these data and econometric issues, the loan providers spuriously seem to in-house securitize ex-ante risker mortgages selling less risky loans to unaffiliated securitizers. I suggest two methods to effectively address these issues. First, the loan provider identity in loan-level dataset can be recalculated based on several pieces of information available from prospectus supplements. After the missing or incorrect identities are recovered, it is possible to identify true affiliation between loan providers and sponsors. Second, the problem can be divided into two parts: 1) the selective disclosure of loan providers by the sponsors; and 2) the affiliation choices made by the loan providers. These two approaches show that loan providers cherry-pick ex-ante less risky mortgages for their own secuirtizations.

CHAPTER 3. WHY DO SPONSORS CARE ABOUT PERFORMANCE OF MORTGAGES THEY SECURITIZE: EVIDENCE FROM THE SPONSOR-SERVICER EFFECTS UPON MILD DELINQUENCIES

3.1 Introduction

In private mortgage securitizations, the sponsor is a non-agency financial institution that sells a pool of mortgages it originated or purchased to the issuer of securities backed by those underlying mortgages.[54] The sponsor is a pivotal player in the securitization process in that its affiliations with other institutions have attracted attention from regulators and investors. Regulation AB (Reg AB), adopted by the U.S. Securities and Exchange Commission (SEC) in 2004, requires disclosure of whether and how the sponsor is affiliated with its material transaction parties including servicers and originators of at least 10% of the mortgage pool.[55]

Among others, Demiroglu and James (2012) and Titman and Tsyplakov (2010) show a significantly positive correlation between the sponsor-originator affiliation and the post-securitization performance of mortgages.[56] The argument that the credit quality of mortgages varies with sponsor affiliations is based on the assumption that sponsors have "skin in the game," or that sponsors' profits are affected by the performance of securitized mortgages. At first glance, this is puzzling because it is investors in MBS, not the originator nor the sponsor, who are the ultimate bearers of risk.

There are two possible reasons why sponsors care about the quality of mortgages they already sold off in the form of MBS. The first explanation relates to the capital structure of trusts, or the entity

[54] See the overview provided on p. 1508 of 33-8518FR.
[55] See "Affiliations and Certain Relationships and Related Transactions," p.1550 of 33-8518FR.
[56] Demiroglu and James (2012) argue the affiliation between the sponsor and the originator encourages more stringent screening of loan applicants. Titman and Tsyplakov (2010) argue that originators cherry-pick less risky mortgages for their own securitizations.

that issues MBS. DeMarzo (2005) developed a theoretical model that showed "tranching" of an asset pool into senior, mezzanine, and residual tranches may lead to optimal securitization. Dermiroglu and James (2012) document that sponsors are exposed to default risk even after MBS issuance and sale because they often hold the bonds backed by the residual tranche. Indeed, the initial owner of the residual certificates, or securities backed by residual tranches, is stated to be the sponsor in prospectus supplements for some deals such as SAMI, SACO, and BALTA sponsored by Bear Stearns, ACE, and DBALT sponsored by Deutsche bank, to name a few. However, residual certificates are also sold to third-party investors as documented in Figure 3.1 for the prospectus supplement of ARSI 2006-M1 sponsored by Ameriquest.

> The Class CE, Class P and Residual Certificates, which are being issued simultaneously with the Offered Certificates, are not offered by this prospectus supplement. Such certificates may be delivered to the Seller as partial consideration for the Mortgage Loans or alternatively, the Depositor may sell all or a portion of such certificates to one or more third-party investors in one or more private transactions.

Figure 3.1. Excerpt from "The Certificates" in the prospectus supplement for ARSI 2006-M1

The second source of "skin in the game" for the sponsors is their reputation.[57] Given the information asymmetry that exists between sponsors and MBS investors about the credit quality of collateralized mortgages (Ashcraft and Schuermann, 2008), it is necessary for sponsors to maintain their reputations in order to sell bonds at a fair market price in the long term.[58] Poor performance of

[57] Titman and Tsyplakov (2010) argue that distressed originators do not carefully underwrite mortgages to earn short-term revenue in exchange for their reputation.

[58] This is consistent with the theories developed by Klein and Leffler (1981), Shapiro (1983), and Allen (1984) that sellers whose product quality is unknown to buyers seek for long-term benefits from favorable reputation rather than exploiting their reputation to sell low-quality goods at higher price.

mortgages may erode subordinate tranches damaging sponsors' reputations and future profits. Thus, it is natural for sponsors to be concerned about performance of mortgages even after securitizations.

There are multiple strategies for sponsors to isolate MBS investors from loss and maintain their reputations. For example, sponsors may provide implicit recourse to investors (Calomiris and Mason, 2004)[59]. If the sponsor is also servicing the mortgages it securitized, then it may provide better quality of servicing as shown in this paper.

I empirically address the question of why sponsors care about the performance of mortgages they have already sold in the form of MBS. To my best knowledge, this is the first paper that explores the sources of sponsors' "skin in the game." Using a large, unique dataset constructed with BBx, ABSnet, prospectus supplements for securitized RMBS deals, and electronic government archives, I show that a substantial portion of mortgages move back and forth between being current and 60 days in arrears, without ever reaching foreclosure and liquidation. In this paper, mortgages are defined to be in mild delinquency (MDQ) if a mortgage repeatedly has the status of 30 to 60 days in arrears, but never reaches a more serious state of delinquency. MDQ is arguably a better measure for servicing quality than other performance measures such as serious delinquency, default, or foreclosure which could be determined by a variety of factors other than servicing quality.[60] I show that mortgages experience MDQ less often when a sponsor "takes care" of mortgages as the primary servicer than when mortgages are serviced by institutions unaffiliated with the sponsor. More importantly, this affiliation effect upon MDQ becomes stronger after the most junior tranche has dried up. These results imply that sponsors provide better servicing than external servicers do,

[59] According to Calomiris and Mason (2004), in the context of credit card securitization, regulatory capital arbitrage through implicit recourse may be socially beneficial for the purposes of reputation, signaling, and efficient risk allocation.
[60] I discuss MDQ in detail in section 3.2.

particularly after their subordinate tranches and reputation begin to suffer damage, which strongly supports the reputation approach to sponsors' "skin in the game."

I structure the rest of the paper as follows. The following section discusses MDQ in greater detail. Section 3.3 explains how my dataset is compiled from sources including BBx, ABSnet, and electronic government archives. Section 3.4 presents the estimates for the relation between sponsor-servicer affiliation and MDQ, and the evolution of these affiliation effects around the date when the residual tranche has dried up based on hazard and linear approaches. Finally, section 3.5 concludes.

3.2 Mild delinquencies as the measure for servicing quality

Default management is the most important task of servicers. MBS investors can directly benefit from proper servicing regardless of how severe the delinquency has become (Mason, 2009). The degree to which servicing quality affects returns to investors, however, could vary with different mortgage failure types. Serious delinquency is expected to have lower correlation with servicing quality than MDQ because it is driven more by variation in the mortgagor's ability to repay than the servicer's due diligence. Moody's (2003) documents the importance of "distinction between the quality of the mortgage loans and the quality of servicer."[61]

Alternatively, foreclosure and modification are noisier measures of servicing quality because they are influenced by many factors including regional variation in foreclosure procedures and regulations[62]; credit quality of the mortgages; servicer or lender's willingness and ability to renegotiate; government subsidies for modification, etc. Moreover, I cannot always interpret fewer foreclosure and more modifications to be directly connected with high quality servicing because loss mitigation is often

[61] See Residential Mortgage Servicer Quality ("SQ") Ratings in EMEA: Moody's Methodology (2003).
[62] See Ghent and Kudlyak (2011). See http://www.realtytrac.com/real-estate-guides/foreclosure-laws/ for the variation in foreclosure laws and procedures.

possible with prompt foreclosures, and modification can also be used as a means of predatory servicing (Mason, 2007).

If a mortgagor is one or two months behind the payment schedule, three interpretations are possible: 1) mistakes on the part of borrower; 2) a prelude to default; and 3) the result of poor or predatory servicing. To correctly identify MDQ as a result of poor or predatory servicing, I exclude one-time 30 days in delinquency (borrower's innocent mistakes) and 30 to 60 days delinquencies followed by 90+ days in arrears (a prelude to default).

I argue one cause of MDQ is poor servicing. It is notable that mortgagors often repeatedly fall into MDQ within a relatively short time frame. They also stay in MDQ for an extended period of time without rolling into serious delinquency. This means payments are continuously made, but the borrower remains just one or two months behind. If servicers adequately manage MDQ, then mortgagors should be able to quickly catch up with their payment schedule. As documented in Moody's (2003), however, "in order to manage arrears effectively, a servicer must establish contact with the borrower and determine the cause of the arrears," which is costly to do. Servicers are compensated with a fixed fee as a percentage of outstanding mortgage pool, hence the amount of their servicing effort is not expected to vary with the occurrence of MDQ if all other conditions are held constant. However, if the servicer is also the sponsor who has a vested interest in mortgage performance, then the servicer has the incentive to expend its resources for telephone calls, mailings, and incurrence of legal and administrative costs to attempts to cure the MDQ.

Another reason for MDQ could be predatory servicing. MDQ may be an opportunity for servicers to earn late fee (as high as 5% of monthly payments) [63] in addition to servicing income which

[63] See Herkenhoff and Ohanian (2012).

typically accounts for 12.5 to 50 basis points of remaining principal balance[64]. Servicers are required to make payments on behalf of delinquent borrowers out of their own pockets to MBS holders. Hence, if the servicers can cheaply fund the interest expenses accruing on the monthly payments they make for delinquent borrowers, then servicers may push healthy borrowers into delinquencies.

There is a body of testimonial, anecdotal and indirect evidence for fraudulent and predatory servicing practices. Thompson testified "For many subprime servicers, late fees alone constitute a significant fraction of their total income and profit…Servicers thus have an incentive to push homeowners into late payments and keep them there" in front of the U.S. Senate Committee.[65] Servicers may take advantage of mortgagors' inability to prove the exact date of payment, charging late fees even when mortgagors are not behind the payment schedules (U.S. General Accounting Office [GAO] 1989). For example, FTB Mortgage Services misapplied the payments from borrowers, collecting unwarranted late fees and ignoring contact from borrowers' attorneys.[66] Some servicers claimed that borrowers missed payments even before the end of the allowed grace period (Brennan 1998). Several servicers sent inaccurate monthly payment demands to charge late fees (Isaac 2001). Additional fees for default management are another plausible motivation for servicers to cause MDQ. According to Renuart (2003), MDQ is a good excuse for servicers to charge fees for property inspections and appraisals regardless of whether or not those services are actually necessary. Servicers can usually keep all the fees generated from MDQ. Revenue from extra fees is typically large enough to cover the operating costs of servicers (Cornwell 2004).[67] The potential

[64] See Mason (2007).

[65] Diane E. Thompson, a counsel of National Consumer Law Center, testified before the U.S. Senate Committee on Banking, Housing, and Urban Affairs on "Preserving Homeownership: Progress Needed to Prevent Foreclosures" on July 19, 2009. Her other testimonial on "Problems in Mortgage Servicing From Modification to Foreclosure" on November 16, 2010 also illustrates how servicers can profit from late fees.

[66] Ronemus vs. FTB Mortgage Services, 201 B.R. 458 (1996)

[67] For example, Ocwen Financial Corporation earned $46 million as late fees, accounting for about 18% of Ocwen's servicing income in 2008. Countrywide charged and earned $285 million as late fees in 2006. Although

effects of predatory servicing may be non-negligible because mortgage terms are typically so complicated that borrowers may not be able to figure out whether fees are legitimately charged (Medine 2000).

I hypothesize that servicing quality increases when the sponsor itself handles mortgages as the primary servicer. For the purpose of empirical tests, mortgage performance can be a measure which represents the quality of servicing.[68] I classify mortgage performance into three categories. First, serious delinquency refers to 90 or more days in arrears followed by foreclosure. Second, foreclosure and modification are the actions taken by servicers for distressed mortgages. Third measure is MDQ, or 30 to 60 days in arrears which is the relevant measure for this paper.

3.3 Data

This paper employs a unique dataset constructed with three different sources. BBx provides information on individual mortgages. Deal-level information is available from ABSnet. I also obtained macro-economic variables from electronic archives of government agencies.

3.3.1 BBx

The main data source is BBx compiled by Blackbox Logic, LLC.[69] BBx provides detailed information on mortgages securitized by private institutions other than government sponsored entities (GSEs) such as Fannie Mae and Freddie Mac. BBx consists of three files: two loan-level datasets including loan tapes (CHARs) and remittance data (PERIODICs), and one deal-level data

---

a single inspection brings only $15, large servicers such as Wells Fargo with 7.7 million mortgages can earn up to $115 million through inspections (Thompson 2010).

68 A servicer can handle mortgages it securitized as the sponsor. The same institution may also handle mortgages for unaffiliated sponsors as the external servicer. The quality of servicing is expected to be different between these internally and externally serviced mortgages. Hence, the "Servicer quality ratings" issued by Moody's cannot be used in my analysis because it cannot capture the variation in servicing quality within the institution. See "Moody's Rating Symbols & Definitions" at https://www.moodys.com/sites/products/AboutMoodysRatingsAttachments/MoodysRatingsSymbolsand%20Definitions.pdf

69 See Mason, Imerman, and Lee (2014), and Mason and Lee (2014) for detailed information about BBx data.

(DEALs). I use CHARs to retrieve the names of primary servicers to construct my key indicator variable of whether a mortgage's servicer is affiliated with the sponsor who securitizes it.

I hypothesize that the servicing quality increases when the sponsor and the servicer are affiliated, which reduces MDQ. However, MDQ is also associated with negative credit events on the part of mortgagors. Hence, it is important to estimate the effects of servicer-sponsor affiliation controlling for various factors that capture the credit quality of mortgages. CHARS also provides underwriting characteristics such as a mortgagor's FICO score, loan-to-value ratio, interest rate type, lien type, credit category, and origination date.

The first risk measure is the Fair Isaac Corporation (FICO) credit score designed to rank individuals based on their financial history. FICO in BBx ranges from 350 to 800 in proportion to the probability of timely repayment. Loan-to-value ratio (LTV) is the ratio of original loan amount and the appraisal value of the property. LTV is an important risk factor which represents how much equity borrowers hold against the mortgage debt amount, and thus how likely lenders are to absorb losses in the event of foreclosure. There may exist more than one loan or lien against the same property. Second lien mortgages are riskier because the second lien lenders will only get paid after the first mortgages are satisfied in event of foreclosure.

Interest rate type is also informative in estimating default risk. Borrowers with adjustable rate mortgages (ARMs) are exposed to higher risk in that they need to make significantly more monthly payments over the life of mortgages than the fixed rate mortgages (FRMs). Exotic features such as interest only (IO) and negative amortization are more conspicuous among ARMs.[70] Issuers of MBS

[70] The portions of interest only and negative amortization mortgages are respectively 28.1% and 18.2% of ARMs while they account for only 11.6% and 2.7% among FRMs.

designate credit categories for the collateralized mortgages as prime, alt-a, and subprime, which are also added into the model to control for negative credit events which are not driven by servicers.

To construct MDQ, I need to measure the evolution of mortgage performance over time. Specifically, I use a data field in PERIODICS called MBADELINQUENCYSTATUS, or the delinquency status defined by Mortgage Bankers Association (MBA). MBADELINQUENCY STATUS is a set of codes which illustrates the location of the loan in its performance curve. Mortgagors are defined to be in MDQ if they stay in 30 to 60 days in arrears after they miss their payments for the first time, however if they never fall into worse status.[71]

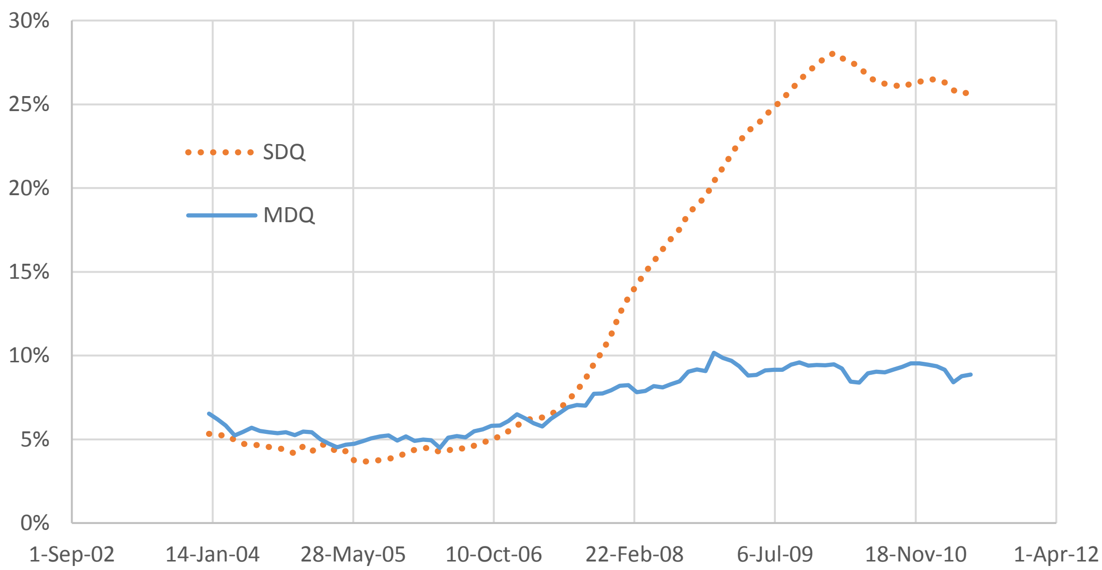

Figure 3.2. The portion of the MDQ across time

I am interested in the situation where a mortgagor repeatedly fails to catch up with their payment schedules due to poor or predatory servicing, not because of negative credit events against

[71] The values of MBADELINQUENCYSTATUS include "C", "3", "6", and "9", respectively indicating current payment, 30, 60, and 90+ days in arrears. In terms of BBx, a mortgage is defined to be in MDQ if its MBADELINQUENCYSTATUS changes between "3" and "6" back and forth after it had "3" for the first time, and never falling into "9" or worse. According to MBA's convention, a borrower is 30-day delinquent if she doesn't repay until one day before the next due date.

borrowers. Hence, first failure of payment and serious delinquency followed by foreclosure are not counted as MDQ.

Figure 3.2 exhibits the portions of two mortgage groups in mild delinquency (MDQ) and serious delinquency (SDQ) across time from Jan 2004 to Dec 2010. A mortgage is serious delinquent if the borrower fell into 90 days in arrears or worse status. Both groups accounted for around 5% of the population with the MDQ cohort slightly larger than the SDQ one during the years leading up to the crisis. However, the portion of SDQ group dramatically increased up to 28% as the crisis deepened.

3.3.2 ABSnet

I use ABSnet for deal-level information which is not available in BBx. The key variable in the analysis is the affiliation between the primary servicer and the sponsor. I obtain the identity of sponsors from prospectus supplements (forms 424B2, 424B3, and 424B5) of MBS trusts which are downloadable in the section of legal reports in ABSnet.[72] Notably, the sponsor may take over the servicing rights for some of the mortgages it acquires from multiple loan providers, or the sponsor sometimes retains its servicing rights for the mortgages it closed as the originator. Therefore, there may be both internal and external servicers concurrently within a deal.

Another key variable in the analysis of how the affiliation effects varies depending on whether a sponsor retains financial stakes in its mortgage trust is the date when the most junior tranche dries up. ABSnet provides the time-varying capital structure from which I can observe the dynamics of outstanding balance for each tranche. I hand-collected the information of outstanding balance for

[72] Alternatively, prospectus supplements are also available from SEC Edgar and Bloomberg terminal.

junior tranches (X to R) over time and calculated when the balances of subordinate tranche reached zero.

3.3.3 Macro variables

It is also important to control for changes in macroeconomic conditions in order to quantify the effects of sponsor-servicer affiliation upon the quality of servicing, as measured by the likelihood of MDQ. To control for time-varying influences of macroeconomic conditions upon mortgage performance, I use two variables, unemployment rates and housing price indices. The monthly unemployment rate at the county level comes from the U.S. Bureau of Labor Statistics. The home price indices (HPI) are obtained from Federal Housing Finance Agency (formerly Office of Federal Housing Enterprise Oversight HPI). This is quarterly price index at state level for single-family, conventional mortgage transactions which conform to the guide line set by government sponsored entities. I decide to choose FHFA-HPI instead of S&P Case-Shiller HPI because I am interested in looking at a broad geographic area, and do not wish to limit myself to single family houses in twenty metropolitan areas.

3.4 Empirical models and results

3.4.1 Summary statistics

This paper empirically examines how the affiliation between sponsor and servicer affects the likelihood of mild delinquencies. Additionally, I am interested in how the affiliation effects vary depending on the amount of financial stakes the sponsor retains in the mortgage trust. I begin the empirical analysis with the examination of statistical characteristics for my sample.

Table 3.1 presents the number of observations for which variables are not missing and the means for the whole sample of mortgages securitized in 2006, the group of mortgages serviced by external servicers (institutions unaffiliated with the sponsor), and the mortgages internal servicers (the sponsor or its affiliate) take care of.

Table 3.1. Descriptive statistics for mortgage characteristics by affiliations

	All		Serviced by external servicers		Serviced by internal servicers	
	N	Mean	N	Mean	N	Mean
% internal servicers	4632938	53.86%				
FICO at origination	3750912	668.26	1670560	651.45	1957578	681.19
LTV at origination	4681745	68.22	2079222	71.61	2441934	66.58
Initial interest rate	4791286	7.92	2137497	8.24	2492806	7.66
Original balance	4791108	221701.2	2136121	206589.1	2494002	238165.7
Appraisal value of the property	4009429	349127.7	1699667	304480.4	2192650	384070
% ARM	4725490	55.86%	2112143	59.77%	2452375	53.36%
% full documentation	3218813	38.54%	1409139	42.15%	1731898	35.88%
% first lien	3894538	73.40%	1743671	73.75%	2032823	73.33%
% owner occupied	4793915	77.47%	2137631	76.61%	2495299	79.83%
% single family	4735009	71.62%	2093687	74.37%	2481405	69.58%
% purchase loans	4766786	45.08%	2116468	46.59%	2489749	44.46%
% subprime	4612530	34.03%	2066675	42.31%	2422097	28.05%

Among the approximately 4.6 million mortgages sold into private securitizations in 2006, internal servicers account for over half (53.86%) of the sample. On average, observable underwriting characteristics suggest that mortgages handled by internal servicers are less risky. Internally serviced mortgages have higher FICO credit scores and lower loan-to-value ratios by 4.6% and 7.6% respectively. The initial interest rate is 14% higher for externally serviced mortgages than internally serviced ones. Average loan amount and the appraisal value of the property at loan origination are higher for internal mortgages by 15.3% and 26%. External parties service ARMs and subprime mortgages 12% and 51% more than internal servicers.

Table 3.2. Top 10 servicers by affiliation

Top 20 External Servicers	Percent	Top 20 Internal Servicers	Percent
Wells Fargo	20.88	Countrywide Home Loans	26.44
Ocwen	9.87	Residential Funding Corp	13.63
Litton Loan Servicing	7.32	Aurora Loan Services	14.17
Bank of America	5.66	EMC Mortgage	8.69
American Home Mortgage	4.29	Washington Mutual	7.39
Countrywide Home Loans	3.94	IndyMac	5.3
Select Portfolio Servicing	3.85	Bank of America	4.87
Wilshire Credit Corp	3.58	Wells Fargo	3.98
Homeq Servicing Corp	3.19	JP Morgan Chase	3.16
GMAC	2.93	Ameriquest Mortgage	2.59
National City Bank	2.87	Option One	2.32
Lasalle Bank	2.08	Carrington Mortgage Services	1.41
Option One	1.93	Chase Manhattan Bank	1.25
Chase Manhattan Bank	1.9	American Home Mortgage	1.15
Saxon Mortgage Services	1.87	Long Beach Bank	0.76
JP Morgan Chase	1.37	First Horizon	0.68
Home Loan Services	1.06	Bayview Loan Servicing	0.51
New Century	0.91	New Century	0.47
Doral Financial Corp	0.9	Fremont	0.39
PHH Mortgage	0.77	Equity One	0.38

Table 3.2 presents the twenty largest external and internal primary servicers and their respective market shares. The sample is restricted to the mortgages securitized in 2006 for which sponsors are identified from prospectus supplements in ABSnet. Wells Fargo, Ocwen, Litton, Bank of America, and American Home Mortgage account for 48.02% of mortgages whose servicers and sponsors are not affiliated. Notably, unlike Ocwen and Litton whose businesses are primarily focused on servicing, other large external servicers are major originators and sponsors in the non-agency securitization industry. Countrywide, Residential Funding Corporation, Aurora, EMC and Washington Mutual handle 70.32% of internally serviced mortgages.

3.4.2 The relation between affiliation and MDQ - Hazard analysis

Defaults and prepayments typically occur only once over the life of a mortgage. However, MDQs are repeatable events because borrowers can fall behind and catch up with their payment schedules

many times over the life of the loan. I argue that the frequency and duration of MDQ are correlated with the quality of servicing in that mortgagors who frequently experience 30 to 60 days delinquencies may do so in past because of poor servicing quality. I posit that these delinquencies perhaps could have been avoided if primary servicers were affiliated with the sponsors and thus more diligent in servicing.

In the context of mild delinquencies, hazard is defined as the risk that the borrower falls behind her payment schedule due to poor servicing in a particular remittance period. Figure 3.3 illustrates how MDQ hazards are constructed based on an example of mortgage remittance reports.

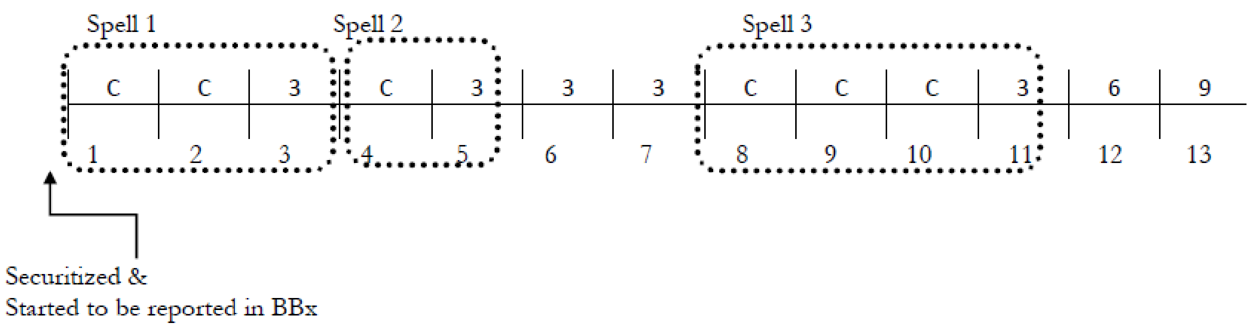

Figure 3.3. Construction of MDQ hazards

The numbers underneath the timeline indicate the number of months since the loan has been securitized. A spell is the total number of remittance reports observed prior to a borrower falling 30 days in arrears.[73] I contrast the length of spells 1 and 2 with spell 3 because 30-day delinquencies in the former are temporary and cured later while the 30-day delinquency in the latter is followed by more serious delinquencies and thus censored from the right. In the example from Figure 3.3, there

[73] The length of spell corresponds to the duration in a typical hazard model.

are three spells, whose durations are respectively 3, 2, and 4 months. Empirical hazard is the number of spells which end during the interval divided by the effective sample size at the beginning of the interval. I estimate a hazard function using the following formula:

$$h(t_{im}) = \frac{d_i}{b_i(n_i - \frac{w_i}{2} - \frac{d_i}{2})} \tag{1}$$

I define one interval as three months, or one quarter, in this context. The midpoint for the i-th interval is denoted as t_{im}. The number of MDQ events is detnoed as d_i. The width of the interval is b_i. The number still at risk at the beginning of the interval is n_i. The number of cases withdrawn within the interval, or censored due to transfer out of BBx or serious delinquency is w_i.

I begin empirical tests with estimation of the Kaplan-Meier empirical hazard to examine the univariate relationship between sponsor-servicer affiliation and the likelihood of MDQ. Figure 3.4 presents the estimates of empirical hazard of MDQ for the mortgages securitized in 2006 during the time when they stay in the trust. In the context of MDQ, an empirical hazard refers to the ratio of the number of mortgages that experience MDQ to the number of mortgages that have not yet experienced MDQ. The horizontal axis denotes the number of months since the most recent MDQ. Intervals for hazard estimates are given in increments of three months. The vertical axis presents the probability of MDQ at time t conditional that the mortgagors were current on their payment at time $t-1$. The solid line corresponds to the mortgages serviced by external servicers, and the dotted line is associated with internally serviced mortgages. The sample period covers from Jan 2006 to June 2011.

The two empirical hazards are significantly different from each other at 1% level. The hazard curves are generally downward sloping with earlier quarters containing higher hazards. This can be interpreted as once a mortgagor falls into MDQ, she repeatedly experiences subsequent MDQs within a relatively short time frame. Another noticeable pattern is that the hazards are higher in the second and the eighth quarters since securitization than surrounding periods, regardless of servicers' affiliation type.

MDQ is a function of both credit quality of mortgagors and poor or predatory servicing. Mortgagors may miss their monthly payments because they experience negative credit events, or a downturn in housing markets may lead some borrowers to strategically default; thus MDQ is related to factors beyond servicer quality. Therefore, it is important to control for MDQ factors which are not related to servicing frictions.

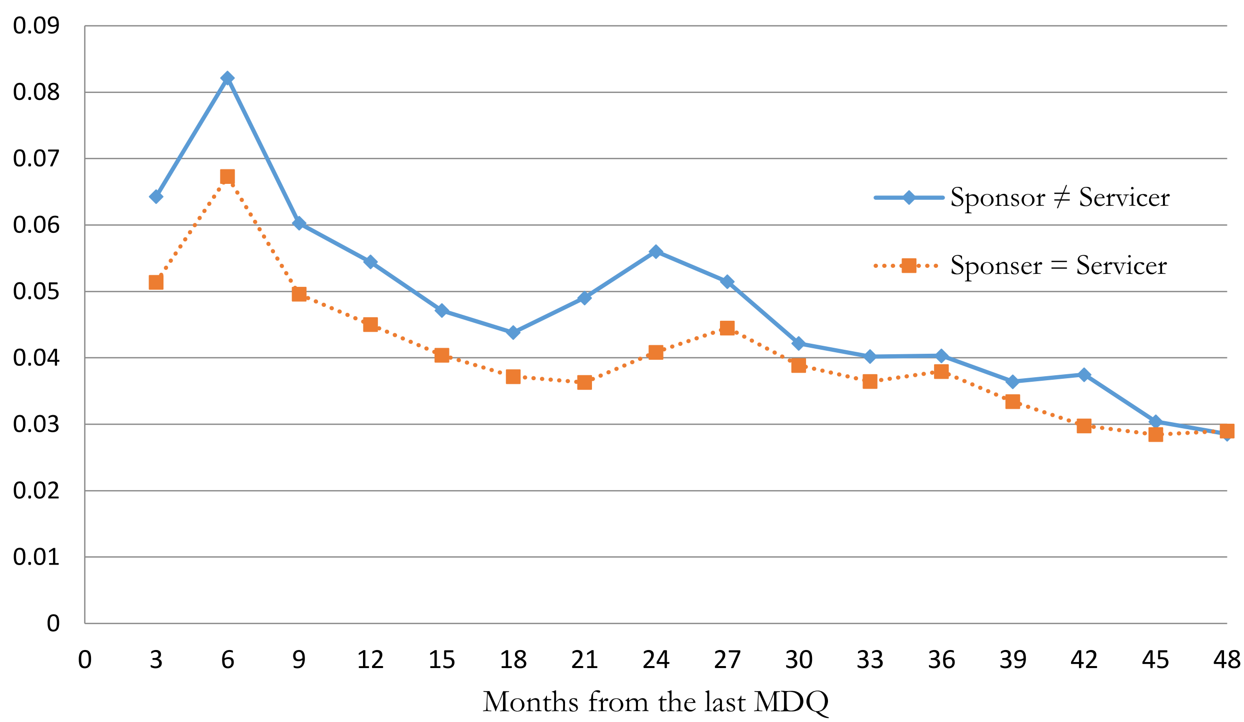

Figure 3.4. Kaplan-Meier Empirical Hazard of Mild Delinquencies

I next examine whether the affiliation between the servicer and the sponsor decreases the hazard of 30 to 60 days in arrears even after controlling for mortgage characteristics such as FICO credit scores, LTV ratios, interest rate type, lien type, credit category, property location, and mortgage vintages. I estimate Cox proportional hazard regressions of MDQs upon affiliation using the following specification:

$$h_i(t) = \lambda_0(t)\exp[\alpha \times internal_i + \beta' X_i + \gamma' State_i + \theta' Vintage_i] \qquad (2)$$

where t is the length of a spell. $h_i(t)$ is the function for the hazard of mild delinquency. The baseline hazard is denoted as $\lambda_0(t)$. The key independent variable $internal$ equals one if the servicer and the sponsor are affiliated for a particular mortgage. A vector of characteristics for mortgage i is denoted as X_i. $State$ is a vector of dummy variables for property location at the state level.[74] $Vintage$ is a vector of dummy variables which indicate when the mortgages were originated. Notably, equation 2 is constructed as one-stage model because in this setting I am not concerned with potential sample selection issues. The names for primary servicers and sponsors are available from BBx loan tapes and prospectus supplements for virtually every loan, which eases the concerns about selection bias.[75]

[74] Due to computational burden, I use state dummies to control for regional differences in Cox proportional hazard model.
[75] In contrast, the studies associated with loan providers (or originators) are subject to selection bias because disclosure of loan providers is a choice of securitizers based on ex-ante risk characteristics. See Mason and Lee (2014) for details.

Table 3.3 presents the hazard ratio estimates for the effects of affiliation between the servicer and the sponsor on the hazard of MDQ based on a Cox-proportional hazard model as specified in equation (2). I use about 5 million observations in the longitudinal format.

Table 3.3. Cox hazard regression with loan attributes

Dependent variable: the hazard of mild delinquency				
	Model 1	Model 2	Model 3	Model 4
Sponser = Servicer (d)	0.954***	0.98***	0.934***	0.965***
	(0.0009)	(0.0011)	(0.0011)	(0.0011)
Log of FICO at origination	0.045***	0.086***	0.043***	0.088***
	(0.0048)	(0.0088)	(0.0048)	(0.0087)
Log of LTV at origination	0.996***	1.068***	0.995***	1.082***
	(0.0008)	(0.0012)	(0.0008)	(0.0012)
ARM (d)		1.318***		1.296***
		(0.0012)		(0.0012)
2nd lien (d)		1.337***		1.367***
		(0.0018)		(0.0019)
Alt-A (d)		1.148***		1.157***
		(0.0017)		(0.0017)
Subprime (d)		1.179***		1.213***
		(0.0025)		(0.0025)
Originated before 2000 (d)			0.948***	0.998
			(0.0107)	(0.0111)
Originated in 2000 (d)			0.9***	0.973***
			(0.0151)	(0.0153)
Originated in 2001 (d)			0.906***	0.945***
			(0.0132)	(0.0136)
Originated in 2002 (d)			0.908***	0.985*
			(0.0105)	(0.0109)
Originated in 2003 (d)			0.903***	0.98***
			(0.0083)	(0.0083)
Originated in 2004 (d)			0.901***	0.934***
			(0.0069)	(0.0069)
Originated in 2005 (d)			0.855***	0.853***
			(0.0013)	(0.0013)
Property location FE	N	N	Y	Y
Log-likelihood value	-68183085	-67417634	-68017106	-67254046
N	5,047,718	4,996,261	5,038,012	4,986,562

The sample period covers the remittance reports for the mortgages in the trusts securitized in 2006. The dependent variable is the hazard of MDQ. I present four specifications to show results, with and without consideration of mortgage characteristics beyond FICO and LTV, property locations and mortgage vintage fixed effects.

I present the simplest specification as model 1 containing only variables for the affiliation between the servicer and the sponsor, and borrowers' FICO and LTV. I add more controls for mortgage types related to interest rate, lien, and credit ratings in models 2 and 4. Loan vintage dummies are included in models 3 and 4. Depending on specification, the servicer-sponsor affiliation significantly decreases the hazard of MDQ by 2 to 6.6%.[76] This is consistent with the first hypothesis that primary servicers provide superior servicing when they are also sponsors of the securitizations. This negative relation between servicer-sponsor affiliation and the hazard of MDQ is robust even when the sample is restricted to the group of mortgages serviced only by sponsors with servicing platform.

MDQ is also significantly associated with ex-ante risk factors such as FICO, interest rate variability, credit category, and lien types. MDQ is more likely for the mortgages with low FICO scores, adjustable interest rates, second liens and non-prime category whose base servicing costs are expected to be higher than their less risky counterparts.

3.4.3 The relation between affiliation and MDQ - Linear analysis

The hazard scheme is flexible enough to capture repetitive nature of MDQ and the length of time up to the occurrence of events. However, hazard analysis has one drawback. Mortgagors may

---

[76] The hazard ratio means the variation in the rate of MDQ when the independent variable of interest increases by one unit with all other variables held constant. Hence, from the hazard ratio of 0.965 for sponsor-servicer affiliation dummy in model 4, I say the rate of MDQ decreases by 3.5% (= 100% − 96.5%) as the servicer for a mortgage changes from external to internal one.

continuously stay in 30 to 60 days in arrears for an extended period without coming back to be current, and the subsequent MDQs within a spell is not considered in the construction of the previously presented hazard. For example, in Figure 3.3, the mortgagor falls into 30 days in arrears again in the fifth month and remains in that state for the next three months. Spell 2, however, does not include MDQs in the sixth and seventh months. In order to overcome this problem, I use the mild delinquency (MDQ) ratio[77] as an alternative measure of the propensity for MDQs. The MDQ ratio is defined as the number of months for which borrowers face 30 to 60 days in arrears divided by the number of months for which mortgages do not fall into serious delinquencies. Using this measure, the value of MDQ ratio for the mortgage in Figure 3.3 is 4/11 or 0.364.

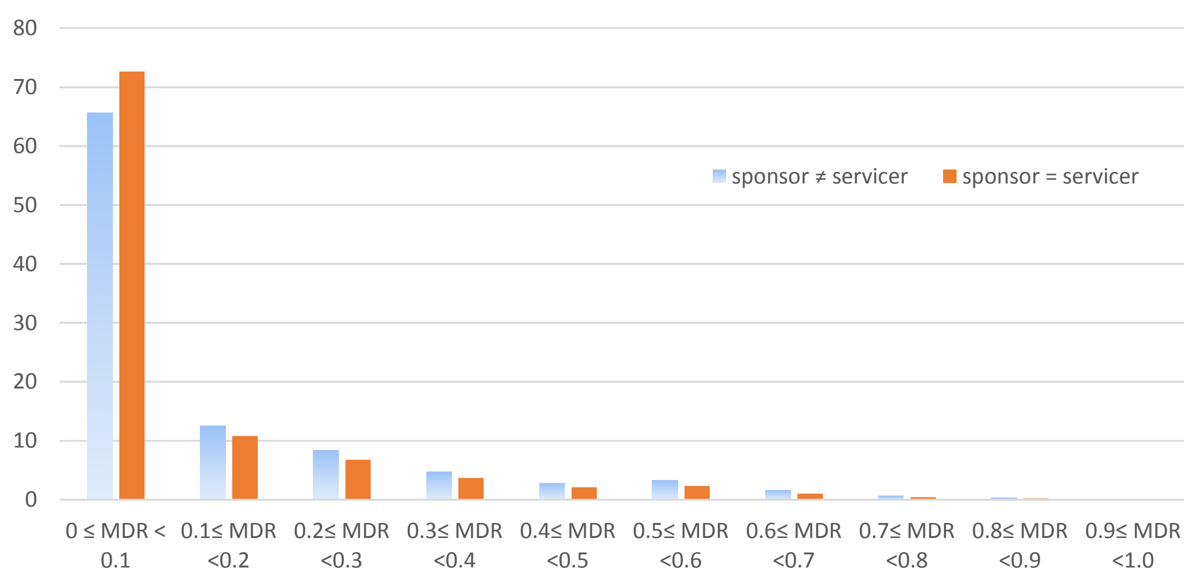

Figure 3.5. The distribution of mild delinquency ratio (MDR)

I present the histograms of MDQ ratio by affiliation types in Figure 3.5. Regardless of sponsor-servicer affiliation, the majority of mortgages have a MDQ ratio less than 10%. Fewer and fewer

[77] I thank Charles Calomiris for his suggestions related to the MDQ ratio.

mortgages have higher MDQ ratios for both groups. The size of the cohort with MDQ ratio less than 10% is larger for mortgages whose sponsors and servicers are affiliated than for the others. In contrast, the size of cohorts with a MDQ ratio higher than 10% are always smaller for internally serviced mortgages than for externally serviced ones. This is consistent with my expectation that mortgages will experience longer periods of MDQ when they are externally serviced because servicers are relatively more negligent and predatory for the mortgages they did not securitize.

I examine whether this negative relation between the MDQ ratio and servicer-sponsor affiliation holds when other ex-ante risk characteristics are controlled, using the following model:

$$MDQ\ ratio_i = \alpha \times internal_i + \beta' X_i + \gamma' CBSA_i + \theta' Vintage_i + \varepsilon_i \quad (3)$$

where the dependent variable is the MDQ ratio for mortgage i. The MDQ ratio is modeled as a linear function of the affiliation between the servicer and the sponsor ($internal_i$), ex-ante risk characteristics or loan attributes (X_i), property location ($CBSA_i$)[78], and loan vintage ($Vintage_i$).

I present OLS regression estimates for the relation between MDQ ratio and sponsor-servicer affiliations in Table 3.4. Of the approximately 4 million mortgages securitized in 2006, I use 500,000 random sample in this analysis. I present four specifications with varying levels of controls. Models 2 and 4 have additional controls for loan types. Fixed effects associated with loan vintages and property locations are additionally controlled in models in 3 and 4. OLS estimates are consistent with the Cox regression results.

The proportion of loan life spent in MDQ significantly decreases when the sponsor plays the role of the primary servicer. This negative correlation between MDQ ratio and sponsor-servicer affiliation is

[78] Core based statistical area (CBSA) is the union of metropolitan and micropolitan statistical areas.

robust across different models. The MDQ ratio increases when mortgages have higher ex-ante risk characteristics, including low FICO, high LTV, adjustable interest rate, and 2nd lien status.

Table 3.4. OLS regression of MDQ ratio on loan attributes[79]

	Dependent variable = the ratio of MDQ			
	Model 1	Model 2	Model 3	Model 4
Sponsor = Servicer (d)	-0.0136***	-0.0128***	-0.0123***	-0.0084***
	(-33.090)	(-30.907)	(-29.442)	(-19.691)
FICO < 620 (d)	0.0858***	0.0854***	0.0844***	0.0066***
	(148.699)	(145.513)	(144.774)	(4.434)
620 ≤ FICO < 680 (d)	0.0304***	0.0299***	0.0301***	0.0177***
	(57.951)	(56.657)	(57.275)	(24.662)
FICO unknown (d)	0.0437***	0.0438***	0.0432***	-0.0072***
	(76.419)	(61.185)	(75.040)	(-6.118)
LTV < 80 (d)	-0.0164***	-0.0154***	-0.0156***	-0.0120***
	(-19.462)	(-18.110)	(-18.446)	(-13.659)
80 ≤ LTV < 100 (d)	-0.0036***	-0.0032***	-0.0026***	-0.0020**
	(-4.258)	(-3.585)	(-3.076)	(-2.192)
LTV unknown (d)	0.0414***	0.0706***	0.0554***	0.0904***
	(27.372)	(41.235)	(34.585)	(50.814)
ARM (d)		0.0115***		0.0125***
		(24.726)		(26.202)
2nd lien (d)		0.0100***		0.0079***
		(16.523)		(12.668)
Lien unknown (d)		0.0051***		-0.0118***
		(6.802)		(-13.725)
Vintage fixed effects	N	N	Y	Y
Location fixed effects	N	N	Y	Y
adj. R-sq	0.061	0.066	0.066	0.078
N	476582	471164	476582	460605

3.4.4 The variation in affiliation effects around the dry-up date

In sections 3.4.2 and 3.4.3., I showed that MDQ is more likely for mortgages whose sponsors outsource servicing from unaffiliated institutions. This implies that servicing quality is better when

[79] Dummy variables are denoted as (d).

sponsors themselves service mortgages they deposit in their own trust. Hence, it is natural to posit that sponsors care about the performance of mortgages in their deals.

In this section, I examine why sponsors are incentivized to take care of mortgages whose bonds are already sold off to investors. There are two possible reasons. First, sponsors may retain the residuals (Stanton, 2005; Demiroglu and James, 2012). Second, sponsors may care more about their reputation that is a function of the returns on the bonds they issued. Both arguments are consistent with the notion that sponsors' profits could be affected by the performance of underlying mortgages. However, the former predicts that the sponsor-servicer affiliation effects should disappear as soon as a sufficient number of mortgages default and the value for the first loss position held by the sponsor becomes zero. In contrast, under the latter argument, the affiliation effects are expected to increase after the most junior tranche dries up, which prompts sponsors to better service the mortgages in their trusts.

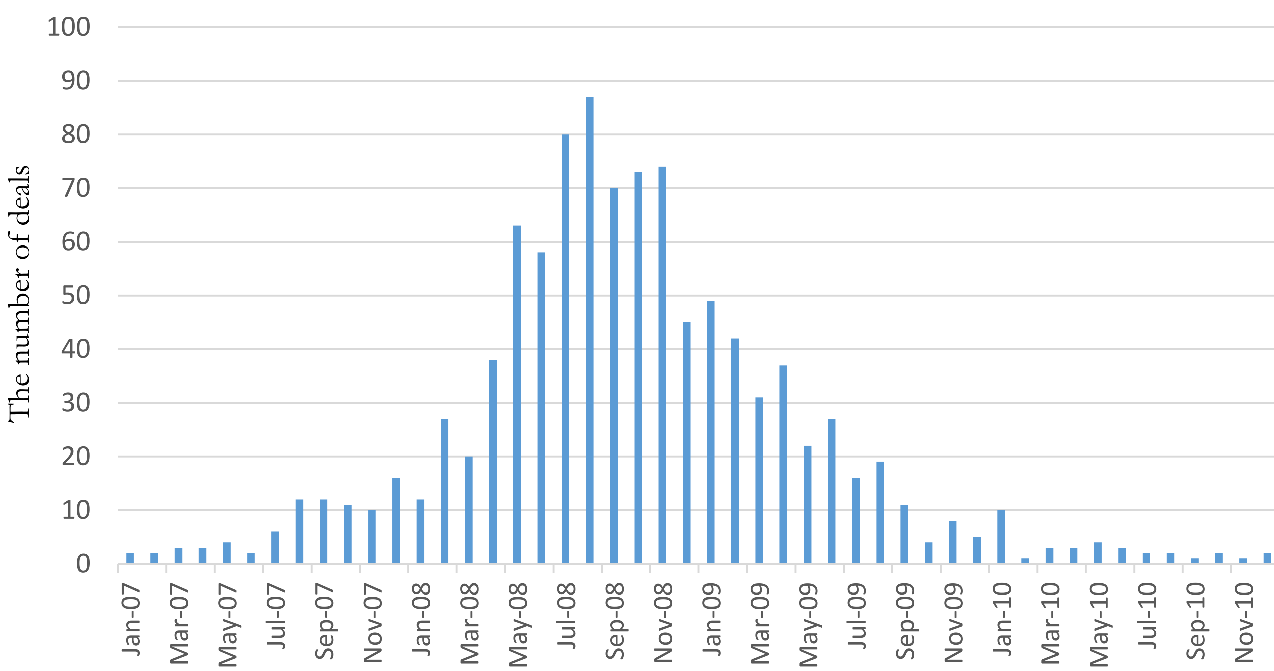

Figure 3.6. Distribution of dry-up date

The dry-up is a key event in the examination of sponsors' incentives to take care of mortgages in their trusts. Figure 3.6 shows the distribution of dry-up dates. The vertical axis shows the number of deals whose equity tranche has dried up. Across all securitizations, the loss from defaulted loans eroded the bottom floor of the securitization tower and reached the mezzanine tranches the most often in the third quarter of 2008. In particular, there are 87 deals whose bottom tranches were wiped out in August 2008.

Figure 3.7 shows the portion of borrowers who fall into MDQ around the time when the most junior tranche dries up. The solid and dotted lines indicate the MDQ rates respectively for the mortgages whose servicers are unaffiliated and affiliated with the sponsors. The horizontal axis is associated with the number of months around the date when the value for equity tranche drops to zero. The difference in MDQ rate between externally and internally serviced mortgages is amplified after the dry-up date. MDQ rate is only slightly higher for external mortgages than for internal ones (0.53%) 24 months before the dry-up date, however the difference in MDQ rate increases to 3.7% approximately six months after the dry-up date, which is more consistent with the reputation hypothesis.

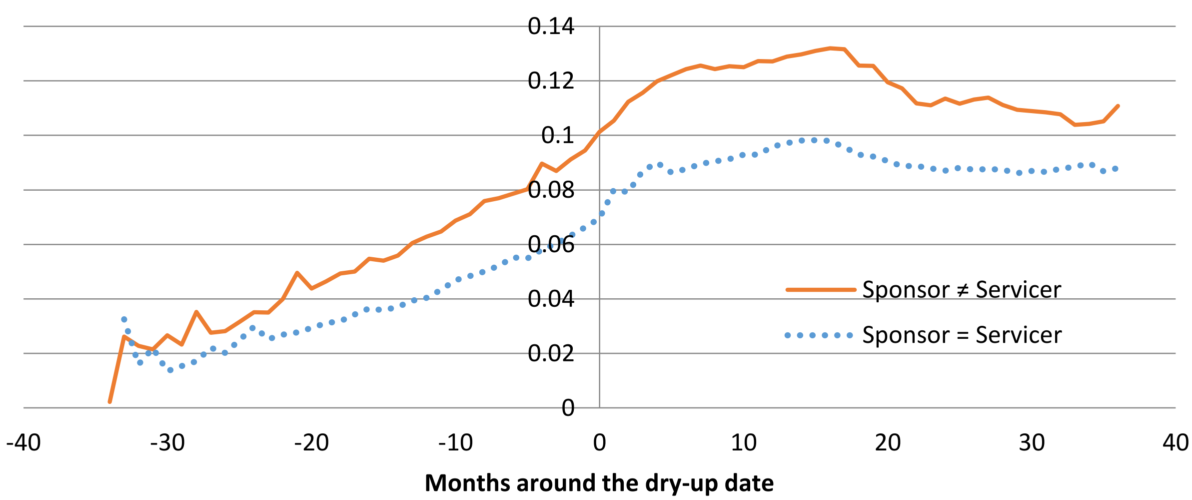

Figure 3.7. MDQ rate by servicer-sponsor affiliation

It could be argued that the dynamics of affiliation effects around the dry-up date may be driven by the difference in ex-ante risk characteristics for each mortgage or macro environment mortgagors face. Therefore, I examine how the effects of affiliation vary across time since the dry-up date, controlling for loan attributes and macro variables.

$$1(mortgage\ is\ in\ MDQ\ |t) = \alpha \times internal_i + \beta' X_i + \theta' M_{it} + \varepsilon_{it} \quad (4)$$

In the regression equation (4), the dependent variable is whether the mortgage experiences MDQ in tth month from the dry-up ($t = 0$). The vector of underwriting characteristics for mortgage i is denoted as X_i. A vector of macro variables is M_{it} including percentage change in unemployment rate and FHFA house price index for the property location between origination and tth month from $t = 0$.

In addition to estimating a regression for the dry-up date, I repeat the estimation for 12 regressions for each month between six months before and after $t = 0$, which is presented in Table 3.5. Similar to previous models, the sample is restricted to the mortgages securitized in 2006. The regressions are separately run for each month around the dry-up date when the most junior tranche was exhausted for a deal. T=0 is the tranche dry-up date. $T = -t$ means t months before the dry-up date while the number of months after the dry-up date is given by $T = t$. Loan vintage and servicer fixed effects are controlled for all 13 regressions. Standard errors are clustered by ZIP codes. As suggested in my first essay, mortgage underwriting characteristics may be selectively reported in the loan-level dataset. Hence, to minimize the number of mortgages dropped from the sample, I keep all mortgages who are missing some attributes by classifying those in the “unknown” categories.

Table 3.5. Loan-level MDQ regressions: The effect of servicer-sponsor affiliation and the value of the most junior tranche

Dependent variable=1 if the mortgage is mildly delinquent (30 or 60 days in arrears but not followed by 90+days in arrears or by foreclosure); 0 otherwise.													
	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
Servicer = sponsor (d)	-0.0104***	-0.0105***	-0.0128***	-0.0157***	-0.0178***	-0.0200***	-0.0229***	-0.0245***	-0.0253***	-0.0271***	-0.0082***	-0.0276***	-0.0275***
	(-14.070)	(-13.631)	(-15.775)	(-18.867)	(-20.174)	(-22.257)	(-24.734)	(-25.225)	(-25.714)	(-26.835)	(-7.200)	(-26.918)	(-26.241)
FICO < 620 (d)	0.1479***	0.1497***	0.1565***	0.1581***	0.1605***	0.1644***	0.1694***	0.1744***	0.1759***	0.1753***	0.1720***	0.1766***	0.1761***
	(142.379)	(140.488)	(141.431)	(139.261)	(137.318)	(136.501)	(139.552)	(139.653)	(137.123)	(135.561)	(131.815)	(134.380)	(130.946)
620 ≤ FICO < 680 (d)	0.0465***	0.0485***	0.0522***	0.0539***	0.0562***	0.0587***	0.0609***	0.0640***	0.0666***	0.0673***	0.0687***	0.0701***	0.0710***
	(79.085)	(80.634)	(84.628)	(84.306)	(85.242)	(83.755)	(86.163)	(87.215)	(88.134)	(86.687)	(87.607)	(87.915)	(87.308)
FICO unknown (d)	0.0521***	0.0512***	0.0584***	0.0547***	0.0571***	0.0587***	0.0605***	0.0582***	0.0626***	0.0602***	0.0618***	0.0634***	0.0629***
	(58.795)	(56.645)	(62.581)	(57.334)	(58.149)	(56.416)	(57.965)	(53.341)	(55.926)	(51.649)	(53.359)	(53.183)	(52.775)
80 ≤ LTV < 100 (d)	0.0092***	0.0095***	0.0117***	0.0124***	0.0125***	0.0133***	0.0145***	0.0138***	0.0145***	0.0160***	0.0176***	0.0169***	0.0171***
	(17.847)	(18.091)	(21.363)	(21.986)	(21.955)	(22.869)	(24.605)	(22.541)	(23.052)	(24.973)	(26.393)	(25.301)	(25.137)
LTV ≥ 100 (d)	0.0064***	0.0037***	0.0112***	0.0132***	0.0114***	0.0044***	0.0118***	0.0044***	0.0079***	0.0081***	0.0092***	0.0095***	0.0111***
	(5.877)	(3.297)	(9.680)	(11.261)	(9.281)	(3.536)	(9.329)	(3.411)	(5.960)	(5.879)	(6.428)	(6.771)	(7.816)
LTV unknown (d)	-0.0080	-0.0031	0.0011	-0.0023	-0.0056	-0.0176**	0.0018	0.0022	-0.0023	0.0117	0.0153	0.0224*	0.0206*
	(-1.195)	(-0.453)	(0.155)	(-0.310)	(-0.710)	(-2.196)	(0.202)	(0.239)	(-0.229)	(1.083)	(1.379)	(1.940)	(1.684)
Second lien (d)	-0.0129***	-0.0101***	-0.0253***	-0.0248***	-0.0230***	-0.0058***	-0.0243***	-0.0116***	-0.0216***	-0.0224***	-0.0224***	-0.0236***	-0.0251***
	(-14.561)	(-11.076)	(-28.702)	(-27.673)	(-24.649)	(-5.770)	(-24.790)	(-10.844)	(-20.110)	(-20.165)	(-19.216)	(-20.257)	(-21.872)
Lien unknown (d)	0.0051***	0.0028***	0.0081***	0.0068***	0.0090***	0.0147***	0.0097***	0.0076***	0.0069***	0.0144***	0.0037***	0.0088***	0.0114***
	(5.874)	(3.147)	(8.740)	(7.118)	(9.199)	(14.228)	(9.391)	(7.118)	(6.403)	(12.693)	(3.166)	(7.542)	(9.410)
ARM (d)	0.0134***	0.0149***	0.0163***	0.0164***	0.0184***	0.0176***	0.0234***	0.0240***	0.0265***	0.0271***	0.0253***	0.0267***	0.0264***
	(25.768)	(27.965)	(29.555)	(29.410)	(31.658)	(29.952)	(38.619)	(37.831)	(40.499)	(39.872)	(36.440)	(38.035)	(37.217)
Low doc (d)	0.0054***	0.0046***	0.0072***	0.0064***	0.0062***	0.0096***	0.0062***	0.0064***	0.0063***	0.0059***	0.0068***	0.0087***	0.0093***
	(8.938)	(7.272)	(11.204)	(9.625)	(9.066)	(13.558)	(8.721)	(8.703)	(8.527)	(7.573)	(8.504)	(10.821)	(11.243)
Doc unknown (d)	0.0262***	0.0286***	0.0212***	0.0186***	0.0193***	0.0151***	0.0182***	0.0307***	0.0201***	0.0231***	0.0184***	0.0235***	0.0220***
	(29.035)	(31.046)	(23.555)	(20.167)	(20.088)	(15.158)	(18.049)	(28.271)	(18.531)	(20.481)	(16.302)	(20.291)	(18.853)
%Δ Unemployment rate	0.0209***	0.0222***	0.0196***	0.0217***	0.0192***	0.0144***	0.0127***	0.0128***	0.0120***	0.0063***	0.0077***	0.0041***	0.0028***
	(23.504)	(25.415)	(22.827)	(25.365)	(23.281)	(18.385)	(16.495)	(16.600)	(15.458)	(8.233)	(10.017)	(5.391)	(3.610)
%Δ House price index	-0.0047	-0.0081***	-0.0086***	-0.0023	-0.0036	-0.0087***	-0.0174***	-0.0187***	-0.0164***	-0.0275***	-0.0268***	-0.0331***	-0.0338***
	(-1.615)	(-2.843)	(-2.960)	(-0.800)	(-1.246)	(-2.933)	(-5.894)	(-6.169)	(-5.465)	(-8.986)	(-8.612)	(-10.789)	(-10.821)
Loan vintage FE (d)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Servicer FE (d)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
adj. R-sq	0.054	0.052	0.058	0.056	0.056	0.057	0.059	0.059	0.061	0.061	0.060	0.060	0.060
N	1156187	1129125	1099378	1077894	1056651	1032551	1011611	992668	973902	956850	941237	926035	912308

In general, the affiliation between servicer and sponsor decreases the likelihood of MDQ. The negative effects of affiliation nearly monotonically increase from -1.04% to -2.76% during the 13 months around the dry-up date. If sponsors' interest in loan performance was determined only by their first loss positions, the affiliation effects should decline or disappear after the value of residual tranche became zero. In contrast, my result shows that as residual value declines and goes to zero, the effect of affiliation between the sponsor and the servicer becomes more negative. These results are consistent with the hypothesis that sponsors may be incentivized to actively manage the performance of collateralized mortgages through affiliated servicers due to their desire to maintain reputational capital.

MDQ may occur for a variety of reasons other than credit quality of mortgages. Hence, ex-ante risk factors may exert mixed influences on the likelihood of MDQ. Mortgages experience MDQ more often when their borrowers have lower FICO scores, and when the mortgages were closed with adjustable interest rate, low documentation, and when the property locations were exposed to adverse economic conditions with rising unemployment and declining house prices.

However, MDQ is the most likely for mortgages with an intermediate level of LTV and senior lien status. This presumably implies the possibility that mortgagors with capacity to continuously pay late fees could be often induced to make payment mistakes.

My results are highly robust across different specifications and different set of samples including the group of mortgages serviced only by sponsors with servicing platforms. The results consistently show that the negative relation between the sponsor-servicer affiliations and the probability of MDQ is amplified as the junior tranches dry up whether servicer effects are controlled, and whether junior-lien mortgages are excluded to control for the effects of home equity.

3.5 Conclusion

There may be substantial differences between traditional mortgage lending models where the financial intermediary functions are vertically integrated and the securitized banking model where transaction parties are not necessarily affiliated. In particular, sponsor affiliations have been shown to significantly affect securitization outcomes by a large body of literature (Demiroglu and James, 2012; Titman and Tsyplakov. 2010). Most of the previous studies about sponsor affiliations assumed that sponsors have "skin in the game," however to date none have examined why sponsors should be concerned about the performance of mortgages they already sold off into securitizations. In this paper, I provide the evidence on the source of sponsors' "skin in the game." Using large loan-level datasets, I show that servicing quality increases when the mortgage is serviced by the sponsor. More importantly, the relationship between sponsor-servicer affiliation and the likelihood of MDQ is stronger after the most junior tranche has dried up. Based on these two sets of results, I conclude that sponsors do have "skin in the game" stemming from financial incentives to maintain their reputational capital.

REFERENCES

Allen, F. (1984). Reputation and product quality. *Rand Journal of economics*,*15*(3), 311-327.

Amemiya, T. (1985). *Advanced econometrics*. Harvard university press.

Agarwal, S., Chang, Y., & Yavas, A. (2012). Adverse selection in mortgage securitization. *Journal of Financial Economics*, *105*(3), 640-660.

Ashcraft, A. B., & Schuermann, T. (2008). *Understanding the securitization of subprime mortgage credit* (No. 318). Now Publishers Inc.

Bubb, R., & Kaufman, A. (2009). *Securitization and moral hazard: Evidence from a lender cutoff rule* (No. 09, 5). Public policy Discussion Papers, Federal Reserve Bank of Boston.

Brennan, William J. Jr. (1998). Statement of William J. Brennan Jr., Director, Atlanta Legal Aid Society, Inc. Testimony before the U.S. Senate Special Committee on Aging. March16, http://aging.senate.gov/minority/public/index.cfm/files/serve?File_id=41c4af52-0ef6-6075-61ac-9de40ae8bfec

Calomiris, C. W., & Mason, J. R. (2004). Credit card securitization and regulatory arbitrage. *Journal of Financial Services Research*, *26*(1), 5-27.

Ciochetti, B. A., Deng, Y., Lee, G., Shilling, J. D., & Yao, R. (2003). A proportional hazards model of commercial mortgage default with originator bias.*The Journal of Real Estate Finance and Economics*, *27*(1), 5-23.

Cornwell, T. (2004). Countrywide: fees cover cost of servicing. *National Mortgage News* 28(34):6.

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

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

Demyanyk, Y., & Van Hemert, O. (2011). Understanding the subprime mortgage crisis. *Review of Financial Studies*, *24*(6), 1848-1880.

Eggert, K. (2004). Limiting abuse and opportunism by mortgage servicers. Housing Policy Debate, 15(3).

Final Rule: Asset-Backed Securities; Release No. 33-8518 (https://www.sec.gov/rules/final/33-8518.htm)

Ghent, A. C., & Kudlyak, M. (2011). Recourse and residential mortgage default: evidence from US states. *Review of Financial Studies*, *24*(9), 3139-3186.

Gorton, G., & Metrick, A. (2012). Securitized banking and the run on repo. *Journal of Financial Economics*, *104*(3), 425-451.

Gorton, G. B., & Souleles, N. S. (2007). Special purpose vehicles and securitization. In *The risks of financial institutions* (pp. 549-602). University of Chicago Press.

Greene, W. H. (1990). Econometric Analysis. Macmillan Publishing Company. New York, NY

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica: journal of the econometric society*, 679-694.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4* (pp. 475-492). NBER.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.

Herkenhoff, K. F., & Ohanian, L. (2012). Foreclosure delay and US unemployment. *Federal Reserve Bank of St. Louis Working Paper Series*, (2012-017).

Isaac, Ronald G. (2001). Prepared statement of the Fedral Trade Commission before the California State Assembly Committee on Banking and Finance. Feburary 21.
http://aging.senate.gov/minority/public/index.cfm/files/serve?File_id=41c4af52-0ef6-6075-61ac-9de40ae8bfec

Jiang, W., Nelson, A. A., & Vytlacil, E. (2010). Liar's loan? Effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics*, (00).

Keys, B. J., Mukherjee, T., Seru, A., & Vig, V. (2010a). Did securitization lead to lax screening? Evidence from subprime loans. *The Quarterly Journal of Economics*, *125*(1), 307-362.

Keys, B. J., Mukherjee, T., Seru, A., & Vig, V. (2010b). 620 FICO, Take II: Securitization and screening in the subprime mortgage market. Working paper.

Keys, B. J., Seru, A., & Vig, V. (2012). Lender screening and the role of securitization: evidence from prime and subprime mortgage markets. *Review of Financial Studies*, *25*(7), 2071-2108.

Klein, B., & Leffler, K. B. (1981). The role of market forces in assuring contractual performance. *The Journal of Political Economy*, *89*(4), 615.

Larry Freudenberg v E*Trade Financial Corporation et al, U.S. District Court for the Southern District of New York, No. 07-8538

Lee, L. F. (1982). Some approaches to the correction of selectivity bias. *The Review of Economic Studies*, *49*(3), 355-372.

Maddala, G. S. (1986). *Limited-dependent and qualitative variables in econometrics* (No. 3). Cambridge university press.

Mason, J. R. (2007). Mortgage Loan Modification: Promises and Pitfalls.*Available at SSRN 1027470*.

Mason, J. R. (2009). Subprime servicer reporting can do more for modification than government subsidies. *Financial Economics Network Professional & Practitioner Paper Series*.

Mason, J. R., Imerman, M., B., & Lee, H., (2014). Self-reporting under SEC Reg AB and transparency in securitizations: evidence from loan-level disclosure of risk factors in RMBS deals.

Mason, Joseph R. and Lee, Hong. "Selection Bias and the Estimation of Sponsor Affiliations in Mortgage Risk," Working Paper, 2014.

Mayer, C., Piskorski, T., & Tchistyi, A. (2013). The inefficiency of refinancing: Why prepayment penalties are good for risky borrowers. *Journal of Financial Economics*, *107*(3), 694-714.

Mian, A., & Sufi, A. (2009). The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *The Quarterly Journal of Economics*, *124*(4), 1449-1496.

Moody's Investors Service (2002). Bullet proof structures revisited: Bankruptcies and a market hang-over teste securitizations' mettle. Special Report. August.

Moody's Rating Symbols & Definitions (2009)

Piskorski, T., Seru, A., & Witkin, J. (2013). *Asset quality misrepresentation by financial intermediaries: Evidence from RMBS market* (No. w18843). National Bureau of Economic Research.

Piskorski, T., Seru, A., & Vig, V. (2010). Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis. *Journal of Financial Economics*, *97*(3), 369-397.

Prospectus supplement for ABFC 2006-OPT2

Prospectus supplements for ACE 2006-ASP2

Prospectus supplement for ACE 2006-HE2

Prospectus supplement for ACE 2006-SD3

Prospectus supplements for AMSI 2006-R1

Prospectus supplement for ARSI 2006-M1

Prospectus supplement for BASIC 2006-1

Prospectus supplements for BAFC 2006-1

Prospectus supplements for BSABS 2006-IM1

Prospectus supplements for BSARM 2006-4

Prospectus supplements for CMLTI 2006-HE1

Prospectus supplements for CMLTI 2006-WFH4

Prospectus supplements for FFML 2006-FF17

Prospectus supplements for RALI 2006-QH1

Prospectus supplement for TBW 2006-1

Prospectus supplement for TBW 2006-2

Prospectus supplement for TBW 2006-3

Prospectus supplement for TBW 2006-4

Prospectus supplement for TBW 2006-5

Prospectus supplement for TBW 2006-6

Prospectus supplement for WFMBS 2006-1

Proposed Rule: Asset-Backed Securities; Release No. 33-9117 (https://www.sec.gov/rules/proposed/2010/33-9117.pdf)

Renuart, E. (2003). Toward one competitive and fair mortgage market: suggested reforms in a tale of three markets point in the right direction. *Texas Law Review*, 82, 421-38.

Residential Mortgage Servicer Quality ("SQ") Ratings in EMEA: Moody's Methodology (2003)

Ronemus vs. FTB Mortgage Services, 201 B.R. 458 (1996)

Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The quarterly journal of economics*, *98*(4), 659-679.

Stein, J. 2002. Information Production and Capital Allocation: Decentralized versus Hierarchical Firms. *Journal of Finance* 57(5), 1891-1921.

Thompson, D. E. (2010). Written testimony on problems in mortgage servicing from modification to foreclosure

_____. (2009). Written testimony on preserving homeownership: progress needed to prevent foreclosures

Titman, S., & Tsyplakov, S. (2010). Originator performance, CMBS structures, and the risk of commercial mortgages. *Review of Financial Studies*, *23*(9), 3558-3594.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

APPENDIX. MISSING RATE BY DATA CATEGORY AND CLEANSING

Data category	Raw item	missing rate	Cleansed item	missing rate
ARM Adjustment Factor	ArmRoundCd	67.05%		
	ArmRoundDesc	75.04%		
	ArmRoundFactor	64.47%		
	NoArmLookBkDays	63.21%		
ARM Conversion	ArmConvertCd	81.96%	ArmConvertStatusInd	49.57%
ARM Index	IndexCd	45.29%		
	IndexShortName	61.99%		
	OrigIdxValue	38.16%		
ARM Payment	FirstPaymtAdjDt	70.70%	FirstPaymtAdjDtCalc	65.21%
	PayAdjFreq	17.83%		
	PeriodicPayCapPct	32.64%		
ARM Rate Adjustment	IntRtAdjFreq	17.44%	IntRtAdjFreqCalc	17.44%
ARM Rate Cap/Floor	ArmRtLifeCap	56.84%	ArmRtLifeCapCalc	56.89%
	LifeMaxIntRtCeiling	15.67%		
	LifeRtFloor	17.82%	LifeRtFloorCalc	17.91%
	PeriodicRtCap	17.95%		
	PeriodicRtFloor	95.40%		
ARM Rate Initial Period	FirstPerRtCap	78.14%		
	FirstRtAdjDt	64.95%	FirstRtAdjDtCalc	61.08%
	InitialFixedRtPer	99.94%	InitialFixedRtPerCalc	64.14%
Credit Documentation	DocCd	73.33%		
	DocCdDesc	75.82%		
	DocType	40.23%	DocTypeSummary	40.33%
	DocTypeDesc	38.82%		
	NoRatioID	97.95%		
Credit Equity	CombinedLienLTV	24.62%	CombinedLienLTVCalc	47.26%
	LienStatus	27.04%	LienType	26.02%

	OrigLTVRatio	8.30%	OrigLTVRatioCalc	6.35%
	PledgedAssetAmt	92.85%		
	PledgedAssetMrtgInd	59.51%	PledgedAssetMortgageStatusInd	59.51%
Credit FICO	FicoRawScore	58.87%		
	FicoScoreOrigination	22.70%	FicoScoreOriginationCalc	29.30%
Credit MI	LenderPaidMIFlag	90.08%	LenderPaidMIStatusInd	90.08%
			MIStatusInd	0.00%
	PMICovPct	50.75%	PMICovPctCalc	50.75%
	PMIIndicator	37.71%		
	PMIInsurerCd	38.41%		
	PMIInsurerName	81.12%		
	PMIPercentage	91.70%		
Credit Rating	CreditGrade	84.76%	CreditCatLoan	10.65%
Duration	FundingDtTm	89.92%		
	MaturityDt	7.50%	MaturityDtCalc	3.41%
	OrigDtNoteDt	22.44%	OrigDtNoteDtCalc	1.37%
	OriginalTerm	5.28%	OriginalTermCalc	3.38%
	AmortizationTerm	67.43%	AmortizationTermCalc	8.34%
Loan Balance	IssuanceBal	45.14%	IssuanceBalCalc	14.19%
	OriginalBal	0.75%	OriginalBalCalc	0.56%
	SaleBalance	39.16%		
	SchedLnBalClosing	89.69%		
Loan Feature I/O	IntOnlyEndDt	74.56%		
	IntOnlyOrigTerm	44.22%	IntOnlyOrigTermCalc	43.95%
	IOFlag	44.05%	IOStatusInd	43.39%
Loan Feature NegAm	ArmNegAmortCap	55.77%	ArmNegAmortCapCalc	62.92%
	HELOCDrawPeriodYrs	99.91%		
	NegAmortCd	75.42%	NegAmStatusInd	94.79%
	NegAmPctg	88.67%		
Loan Feature Teaser	ArmTeaserPeriod	70.72%		

	PayTeaserind	70.90%		
	PayTeaserPeriod	70.90%		
Loan Origination	ChannelCd	94.99%		
	ChannelDesc	95.00%		
	OriginatorCd	99.87%		
	OriginatorName	84.60%		
	Seller	99.87%		
Loan Type	ARMInd	60.80%	IntRtTypeSummary	3.98%
	LoanType	26.22%		
	LoanTypeDesc	26.25%		
	NoteDesc	71.55%		
	NoteType	64.24%		
	ProductDesc	65.70%		
	ProductTypeCd	58.00%		
	ProgramCd	93.12%		
	ProgramName	93.88%		
Loan Type Balloon	BalloonInd	63.03%	BalloonStatusInd	61.73%
Loan Type Hybrid			HybridARMInd	87.33%
Loan Type Option ARM			OptionARMInd	94.79%
Loan Type Heloc			HelocInd	98.35%
Occupancy	LeaseholdID	97.66%		
	OccStatusCd	9.60%		
	OccStatusDesc	9.07%	OccType	10.42%
Payment	FirstPaymtDt	8.82%	FirstPaymtDtCalc	4.77%
	FirstPrinPaymtDt	90.10%		
	IssuePI	14.01%		
	SchedPIAtIssuance	89.26%		
			OrigPI	97.49%
Prepayment Penalty	PayoffPnltyType	94.85%		
	PrepayPenaltyAmt	39.12%		

	PrepayPenaltyEndDt	81.66%		
	PrepayPenaltyFlag	73.15%		
	PrepayPenaltyInd	31.97%	PrepayPenaltyStatusInd	63.24%
	PrepayPenaltyWaived	70.94%		
Property Location	PropertyCity	13.97%	PropertyCityCalc	0.00%
	PropertyCounty	84.62%	PropertyCountyCalc	0.00%
	PropertyCountyCd	87.34%		
	PropertyStAddress	96.69%		
	PropertyState	4.82%	PropertyStateCalc	0.00%
	PropertyZipCd	3.30%	PropertyZipCdCalc	0.00%
Property Type	NoUnits	76.95%	NoUnitsCalc	76.96%
	PropertyTypeCd	8.59%		
	PropertyTypeDesc	8.40%	PropType	8.67%
Property Value	ApprslTypeDesc	96.21%		
	CurAppraisalValue	97.54%		
	OrigAppraisalValue	10.35%	OrigAppraisalValueCalc	18.50%
Purpose	PurposeCd	12.07%		
	PurposeDesc	12.17%	PurposeType	13.82%
	RelocationInd	95.20%		
Rate	FixedRetYldRt	58.91%		
	InitialIntRt	66.17%	InitialIntRtCalc	0.22%
	IntRtAtIssuance	85.86%	IssuanceRtCalc	94.16%
	Margin	15.59%	MarginCalc	15.59%
	NoteRateAdjForLPMI	90.02%		
	OrigIntRt	0.47%	OrigIntRtCalc	5.49%
Servicing	EscrowBal	90.59%		
	MasterServFee	41.94%		
	ServicerCd	73.63%		
	ServicerName	11.74%	ServicerNameCalc	12.38%
	ServicerNo	69.48%		

	ServicingFee	27.05%		
Loan Modifications			NoMods	1.38%
			ModLatestDt	90.58%
Key Events			FirstActivityDt	0.00%
			LatestActivityDt	0.00%
			POActivityDt	27.02%

VITA

Hong Lee was born in Seoul, South Korea. He earned Bachelor of Arts degree in English language and literature in 2004 and Master of Arts in economics in 2006 from Korea University. After he came to the United States, he earned Master of Arts in economics as an intermediate degree from the doctoral program at Brown University in 2007. In 2008, Mr. Lee transferred to the PhD program in finance at Louisiana State University. He is scheduled to receive the degree of Doctor of Philosophy in finance from LSU in August 2014.

His research interests are primarily in the areas of banking and real estate economics. In particular, he examines and reveals the complexity of securitization process and the data structure for private-label mortgages that previous studies failed to consider. His papers have been selected for presentations at the doctoral consortiums and annual meetings of Asian, European, and U.S. Financial Management Association (FMA), Southern Finance Association, and the Bank of Canada.

Mr. Lee recently accepted the offer of a position for tenure-track assistant professor at Department of Finance and Financial Services at Wright State University.