
Volume 105, Issue 3, September 2012
ISSN 0304-405X
JOURNAL OF Financial ECONOMICS
Contents
Eugene F. Fama and Kenneth R. French, Size, value, and momentum in international stock returns 457
Harrison Hong and Motohiro Yogo, What does futures market interest tell us about the macroeconomy and asset prices? 473
Nicolae Gârleanu, Leonid Kogan and Stavros Panageas, Displacement risk and asset returns 491
Radu Burlacu, Patrice Fontaine, Sonia Jimenez-Garcès and Mark S. Seasholes, Risk and the cross section of stock returns 511
Michael J. Brennan, Tarun Chordia, Avanidhar Subrahmanyam and Qing Tong, Sell-order liquidity and the cross-section of expected stock returns 523
Nikunj Kapadia and Xiaoling Pu, Limited arbitrage between equity and credit markets 542
Dave Berger and Kuntara Pukthuanthong, Market fragility and international market crashes 565
M. Babajide Wintoki, James S. Linck and Jeffry M. Netter, Endogeneity and the dynamics of internal corporate governance 581
John (Xuefeng) Jiang, Mary Harris Stanford and Yuan Xie, Does it matter who pays for bond ratings? Historical evidence 607
Sebastien Betermier, Thomas Jansson, Christine Parlour and Johan Walden, Hedging labor income risk 622
Sumit Agarwal, Yan Chang and Abdullah Yavas, Adverse selection in mortgage securitization 640
Philip Valta, Competition and the cost of debt 661
Published by ELSEVIER
in collaboration with the
WILLIAM E. SIMON GRADUATE SCHOOL
OF BUSINESS ADMINISTRATION,
UNIVERSITY OF ROCHESTER
SciVerse ScienceDirect

Journal of Financial Economics 105 (2012) 640–660

Contents lists available at SciVerse ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Adverse selection in mortgage securitization ☆

Sumit Agarwal [a,*], Yan Chang [b], Abdullah Yavas [c]

[a] National University of Singapore, 15 Kent Ridge Drive, Singapore
[b] Freddie Mac, 8200 Jones Branch Drive, McLean, VA 22102, USA
[c] School of Business, University of Wisconsin-Madison, 5251 Grainger Hall, 975 University Avenue, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 12 May 2010
Received in revised form
7 September 2011
Accepted 12 September 2011
Available online 22 May 2012

JEL classification:
G01
G21

Keywords:
Mortgage
Securitization
Default
Adverse selection

ABSTRACT

Using several large data sets of mortgage loans originated between 2004 and 2007, we find that in the prime mortgage market, banks generally sold low-default-risk loans into the secondary market while retaining higher-default-risk loans in their portfolios. In contrast, these lenders retained loans with lower prepayment risk relative to loans they sold. Securitization strategy of lenders changed dramatically in 2007 as the crisis set in with most unwilling to retain higher-default-risk loans in return for lower prepayment risk. Contrary to the prime market, the subprime market does not exhibit any clear pattern of adverse selection.

1. Introduction

The U.S. economy recently experienced one of the worst financial and economic crises since the Great Depression. The crisis was triggered by a collapse of the bubble in residential real estate markets. Many commentators cite the remarkable growth of securitization in recent years as a major contributor to the rise of the real estate bubble and the ensuing crisis. Part of the argument is that securitization creates additional layers of agency problems in loan origination, which lead to lax underwriting and thus higher default rates (Rajan et al., 2011).

In this paper, we investigate determinants of lenders' choice to securitize loans, focusing on the quality of loans they sell to investors in the secondary mortgage market relative to ones they retain on their balance-sheets. Lenders typically obtain information—both soft and hard (Petersen, 2004; Agarwal and Hauswald, 2010)—about the borrowers when screening their applications (origination) and may use this information when deciding the quality of loans to sell to the investors (post-origination). The conventional wisdom is that lenders may know more about the credit quality of a borrower than what is reflected in the hard information collected, such as the credit score, income, and debt payments of the borrower. Lenders could have incentives to take advantage of their unobservable private information about borrowers and retain higher-quality loans on their balance-sheets while selling inferior-quality loans. However, market mechanisms, such as lender reputation concerns, due diligence practices in the securitization chain including the originator

☆ We thank Jacqui Barrett for excellent research assistance. We are grateful to Brent Ambrose, Marsha Courchane, Tony Sanders, Peter Zorn, and the participants in finance workshops at the University of Toronto, University of South Carolina, National University of Singapore, as well as the participants at the 2010 Finlawmetrics conference, 2010 FDIC/Federal Reserve Mortgage Symposium, and the 2010 ASSA Meeting, for their helpful comments.

* Corresponding author. Tel.: +65 6516 5316.
E-mail addresses: ushakri@yahoo.com (S. Agarwal), yan_chang@freddiemac.com (Y. Chang), ayavas@bus.wisc.edu (A. Yavas).

0304-405X/$ - see front matter
http://dx.doi.org/10.1016/j.jfineco.2012.05.004

representation and warranties may prevent this from occurring. The ultimate impact of lender ability to securitize on the quality of loans they retain is an empirical question—one which we investigate in this paper.

There are marked differences between securitization in the subprime and prime markets. Prime lenders typically sell mortgage loans in the secondary market to Fannie Mae or Freddie Mac, which are GSEs (Government-Sponsored Enterprises) who in turn sell to investors. In contrast, subprime loans, originated largely by different sets of lenders (Mayer and Pence, 2009), are typically packaged and sold to investors by private issuers such as investment banks. Important differences between the control GSEs or private issuers impose on the securitization chain (e.g., provision of incentives/monitoring) can influence both the origination and post-origination practices of the lenders. For instance, GSEs offer investors guarantees against default risk, while private issuers pass the default risk on to parties that are willing to bear it. As a result, it can be expected that GSEs would impose more stringent underwriting standards regarding default risk for lenders who sell loans to them. Similarly, regulatory capital requirements which are a bigger consideration for prime lenders might also influence lenders' decisions to retain risky mortgages in the prime market. Due to such differences in these markets, we examine the origination and post-origination decisions in these markets separately.[1]

In the empirical analysis, we will look at two margins of risk that the lender faces—prepayment and default risk.[2] Prepayment risk refers to the risk that mortgages may be repaid; prepayments often take place in the form of refinancing due to a decline in the interest rate, which is precisely when prepayment is costly for the investor. Default risk refers to the likelihood that the borrower may stop making payments. Earlier studies of adverse selection in mortgage markets focus mostly on default risk. In this paper, we consider both prepayment risk and default risk, and show that both risks play a critical role in lenders' securitization strategies.

We use a large detailed data set of residential mortgage loans from Lender Processing Services (LPS) Applied Analytics, Inc.[3] to compare default and prepayment risks of loans retained on lenders' balance-sheets with those that are sold to investors between 2004 and 2007. We infer the quality of the loan based on the ex post performance (whether the loan defaults or prepays) of the loan. As a result, we need to account for endogeneity, that is, observed securitization and loan performance are co-determined, with each affecting the outcome of the other. To circumvent this problem, the central identification of the paper imposes a certain structure to the securitization process. This structure follows from the institutional background involving various participants in the securitization of mortgage credit.

Our identification method is best understood as follows. We first map the loan-level observables to default and prepayment probabilities. In particular, we infer the quality on these two dimensions based on loan characteristics for all the loans in a training sample. Importantly, we construct the estimates of loan quality as a function of loan characteristics for all loans in the sample. Doing so ensures that these estimates are constructed regardless of the lender decision to sell these loans or keep them on the balance-sheet (i.e., we examine the distribution of entire set of loans). Second, we take these estimates and perform an out-of-sample forecast of default and prepayment probabilities in a holdout sample. Finally, we correlate the estimated default and prepayment probabilities in out-of-sample loans with the actual observed securitization outcomes for these sets of loans. These three steps allow us to better investigate determinants of lenders' choice to securitize loans.

Our analysis of default outcome shows that in the prime market (loans intended for GSEs), originators chose to sell low-default-risk (not high-default-risk) and high-prepayment-risk loans to the secondary market in the pre-crisis period, 2004–06. This strategy changed in 2007 when widespread market disruption was imminent with most of these lenders unwilling to retain higher-default-risk loans in return for lower prepayment risk. In contrast, we do not find a significant difference in the default and prepayment risks between portfolio and securitized loans for subprime loans. In fact, the only year where a significant difference in default risk for subprime loans is found is 2007, when the real estate bubble started to burst.

We also conducted additional tests to examine the robustness of our inferences. Specifically, we conduct reduced form analysis to ensure that our classification of the "prime-like" loan sample does not drive our findings. In addition, we also conduct our analysis in several samples which will also allow better classification and accounting of differences in incentives of participants across the GSE and non-GSE markets. The inferences from this analysis are the same. There is no clear pattern that emerges across subprime loans. In contrast, in the prime market (loans intended for sale to GSEs), banks generally sold low-default-risk loans into the secondary market while retaining higher-default-risk loans in their portfolios. In addition, we also find support for adverse selection with respect to prepayment risk in the prime market.

Our results illustrate the difference in origination and post-origination behavior of lenders across the prime and subprime sectors. We rationalize these findings by arguing that the differences are likely due to GSEs imposing control on default risk of loans originated by lenders since they offer guarantees only against default risk to investors. This control is missing on the prepayment margin—giving lenders more freedom to adversely select on prepayment risk—since this risk is passed to the investors by GSEs. In contrast, there is no private issuer

[1] A more detailed presentation of the securitization process for the prime and subprime markets can be found in the two figures in Appendix A.

[2] Investors in mortgage loans are concerned with three kinds of risk. Interest risk refers to the fact that a change in interest rates leads to an opposite change in the value of the mortgage. Interest rate risk is independent of the borrower's characteristics, and hence, is not subject to potential adverse selection concerns.

[3] LPS Analytics, Inc. was known as McDash Analytics before this company was acquired by Lender Processing Services, Inc. in 2008.

who coordinates the securitization chain in the subprime market.

Our work is broadly related to the growing literature on the role of securitization in the current crisis. Keys et al. (2010, forthcoming) focus on the impact of securitization on the quality of loan screening and servicing and find that mortgage lenders apply weaker screening standards for loans that they are likely to sell in the non-government-sponsored-enterprise (non-GSE) secondary market. Piskorski, Seru, and Vig (2010) study the moral hazard problem created by securitization from the angle that it inhibits renegotiation of loans for distressed borrowers. Agarwal et al. (2010, 2011) also study the role of securitization on loan modification and find that securitized loans are 30% less likely to be modified than portfolio loans.

We also add to the existing literature on the role of adverse selection in the mortgage market. Following Akerlof's (1970) seminal work (1970), there is a small but growing empirical literature on adverse selection in mortgage financing.[4] The research most closely related to ours is Elul (2009) and Ambrose, Lacour-Little, and Sanders (2005). Using a loan-level data set that covers the most recent cycle, Elul (2009) concludes that securitized prime loans have higher default rates than portfolio loans, and the relative performance of securitized loans worsens over the origination years 2003–07, but securitized subprime loans do not perform worse than portfolio loans. As discussed above, he does not address the competing prepayment risk in his analysis. Ambrose, LaCour-Little, and Sanders (2005) use data from a single lender, and find that loans with lower default risk are more likely to be securitized than retained in the lender's portfolio.

We develop our hypothesis in the next section. In Section 3, we describe the data. Then in Section 4, we present the methodology and discuss the results. We show our results from additional robustness tests in Section 5. Finally, in Section 6, we offer our concluding remarks.

2. Adverse selection and mortgage market sectors

In this section we will lay out our key hypothesis. Our hypothesis develops from the differences in the way mortgage markets operate due to incentives of various participants. We first look at differences at origination and post-origination due to differences in securitization practices in the GSE and private market segments. Next, we use this discussion to motivate and develop our hypothesis and testable predictions.

On the origination side, a key difference between the prime and subprime classes of loans is the underwriting criteria. Subprime borrowers typically have lower credit scores, higher debt-to-income (DTI) ratios, and weaker credit histories that include payment delinquencies, judgments, and bankruptcies. Because of the higher risks of subprime loans, subprime lending is based largely on asset value, rather than borrower characteristics (Cutts and Van Order, 2005). As a result, subprime lending was more common in areas with rising house prices (Mayer and Pence, 2009). In addition to high default risk, subprime borrowers also have high rates of prepayment, typically driven by refinancing, as noted in Sengupta and Emmons (2007), Pennington-Cross (2003), and Courchane, Surette, and Zorn (2004). Subprime borrowers have more to gain from refinancing, since they usually have high initial interest rates on their loans and may qualify for a lower rate once they improve their credit through maintaining payment on the current mortgage. Subprime borrowers usually face a step-up mortgage interest rate that applies at the end of a "teaser" period, and they are more likely to experience financial difficulties. In these cases, cash-out refinancing is a way to cure their financial problems. In contrast, since the prime market is supported by the GSEs, the underwriting standards of lenders reflect the stringent guidelines imposed by GSEs.

This difference in origination practices is also reflected in the fact that compared with prime originations, subprime originations are more concentrated in a few subprime specialists (Mayer and Pence, 2009). In addition, there is also an observed difference between the characteristics of prime and subprime originators. Studying the list of 210 subprime lenders published by the U.S. Department of Housing and Urban Development (HUD) in 2005, Sengupta and Emmons (2007) note that, relative to prime lenders, subprime lenders have fewer originations and have a higher share of refinance loans as a proportion of total originations.

We now move to explaining post-origination differences across the two markets. Prime lenders typically sell mortgage loans in the secondary market to Fannie Mae or Freddie Mac, which are GSEs who in turn sell to investors. In contrast, subprime loans are typically packaged and sold to investors by private issuers such as investment banks.[5] Important differences between the control GSEs and private issuers impose on the securitization chain can influence both the origination and post-origination practices of the lenders.

First, GSEs and private issuers differ with respect to default risk. GSEs offer investors guarantees against default risk, while private issuers pass the default risk on to parties that are willing to bear it.[6] The fact that default risk is not passed on to investors, and instead it is retained by the GSEs in the prime market, contributes to

[4] An, Deng, Gabriel (2011) find empirical evidence supporting the presence of adverse selection problems in the market for commercial mortgage loans. Downing, Jaffee, and Wallace (2009) show that Freddie Mac sells more lower-credit-quality residential mortgage-backed securities to bankruptcy-remote special purpose securitization vehicles than it retains in its portfolio.

[5] Secondary market institutions often create pools of loans and sell the payment rights of the loans in the pool to investors around the globe. Of the total volume of \$7.6 trillion in pooled mortgages at the end of 2008, about \$5 trillion is securitized or guaranteed by GSEs or government agencies. The remaining \$2.6 trillion is pooled by private mortgage conduits (*source*: www.federalreserve.gov/econresdata/releases/mortoutstand/mortoutstand20090331.htm).

[6] At times, additional protection against credit risk is provided at the security level by third parties through credit enhancement, which typically offers protection against defaults up to a certain fixed amount.

GSEs imposing more control on default risk of loans originated by lenders since they offer guarantees only against default risk to investors. This control should be missing on the prepayment margin—giving lenders more freedom to adversely select on prepayment risk—since this risk is passed to the investors by GSEs.

Second, GSEs have historically purchased only traditional fixed-rate mortgage products. They only began to purchase alternative mortgages, such as hybrid adjustable-rate mortgages (ARMs) and interest-only products, near the height of the cycle. Private label issuers have been purchasing these alternative mortgages on a much larger scale and for a longer time. These differences give more flexibility to private label issuers, enabling them to create securities that better diversify the risks of individual loans. And as a result, private label issuers might be willing to purchase some loans that GSEs would not. These differences indicate that private label issuers may have a higher preference for default risk among the loans they purchase, which could also help explain difference in securitization propensities between the prime and subprime markets over time.

Having summarized the practices in the two markets, we now lay out the economic forces that may drive the decision of lenders to keep some types of loans on their balance sheet while selling others. If lenders keep higher-quality loans on their balance-sheet while selling lower-quality loans to the investors, we will call it adverse selection (Akerlof, 1970). We note that it is also possible that lenders may retain lower-quality loans while selling better quality to investors (for example, to build reputation).

First, the GSEs tend to purchase mortgage loans from originators that have high underwriting standards. In order for an originator to be able to sell a loan to a GSE, the loan needs to satisfy a certain set of criteria. A lender wishing to sell a loan to Fannie Mae needs to enter the loan and borrower data to Fannie Mae's designated automated underwriting software program, the DeskTop Underwriter, to obtain approval. Freddie Mac has similar software called Loan Prospector. It is possible that a subset of loans that fail to meet GSE and private issuer criteria are still acceptable to lenders who may approve and retain some of these loans. As a result, default risk on bank-held loans by lenders who sell to GSEs may be higher relative to loans that are sold to investors via GSEs. This is also predicted by the fact that GSEs would tend to impose more control on default risk of loans originated by lenders. In contrast, since there is no private issuer who coordinates the securitization chain in the subprime market, one might expect a different pattern in this market. Specifically, the lenders might use their private information more freely to adversely select the loans to securitize in the private market. As a result, default risk on bank-held loans might be lower.[7]

Second, the lender decision to sell may vary across markets due to the regulatory capital requirement in the prime market which motivates lenders to securitize loans with lower default probabilities (Ambrose, LaCour-Little, and Sanders, 2005). Specifically, the current risk-based capital rules require banks to have more capital reserves for higher-risk classes of loans. For prime loans, this gives banks incentives to retain riskier mortgage loans with higher expected return and securitize less risky mortgage loans, as long as both groups of loans have the same capital requirements when held on banks' balance-sheets. This would suggest that relative to loans sold to GSEs, those retained by lenders who sell to them might carry more risk. In contrast, for subprime loans, the cost of holding subprime loans on the balance-sheet is already high. As a result, this channel suggests that there might be no differences in the quality of loans sold to investors versus those that are retained by the banks in the subprime market.

Note that we have not been very precise about defining high or low quality based on risk. In the empirical analysis, we will look at two margins of risk that the lender faces—prepayment and default risk. Our analysis will focus on evaluating the difference in quality of loans on the bank balance-sheet relative to loans that are sold on these margins. Based on our discussion, it is clear that the origination and post-origination practices across the prime and subprime markets differ. As a result, we will conduct our analysis separately in these two markets.

3. Data, coverage across data sets, and descriptive statistics

3.1. Types of data sets

Our data are provided by LPS Applied Analytics, Inc., and includes loan-level information collected from residential mortgage servicers. As of July 2008, the data set included loans from nine of the top ten servicers, and represented around two-thirds of the mortgage market in the United States, or more than 39 million active mortgage loans.[8] As the information is collected from mortgage servicers rather than from investors, agency and non-agency mortgage-backed securities as well as portfolio loans are included in the data set.

The LPS data set provides extensive information about the loan, property, and borrower characteristics at the time of origination, as well as dynamically updated loan

[7] One instance of GSEs implementing more stringent underwriting practices follows from the rep and warrant agreement that it has with the lenders. GSEs selectively check loans that go into default, and if they discover that the lender's representation and warrants were violated, they can force the lender to purchase the loan back at par. GSEs also check a random sample of non-defaulted loans, and can force repurchase of all loans with any rep and warrant violations. GSEs keep track of the repurchase record of originators and impose higher fees and capital requirements on originators with high repurchase rates. These measures make it costly for lenders to have securitized loans go into default, and may induce them to be more conservative with loans that they sell into the secondary market than loans that they retain in their portfolios. There are similar rep and warrant requirements in the non-GSE market, but typically, repurchases are less enforced because the rep and warrant descriptions are less specific and more subject to interpretation (Piskorski, Seru, and Vig, 2010).

[8] See www.lpsvcs.com/NewsRoom/Pages/20080722.aspx.

information subsequent to origination. Property-related variables are appraisal amount, geographic location, and property type (single-family residence, condo, or other type of property). Loan characteristics available to us are origination amount, term to maturity, lien position, whether or not the loan is conventional, loan purpose (purchase or refinance), and lender-defined subprime flag, as well as coupon rate on the mortgage. Credit-risk-related variables include debt-to-income ratio, FICO credit score, loan-to-value (LTV) ratio of the borrower at origination, and level of documentation provided.

Beyond the data that are available at origination, dynamically updated variables capture changes made to the loans since origination, as well as their performance at a monthly frequency. Variables of interest are coupon rates (which change for ARMs and have the potential to change for loan modifications), delinquency status (current, 31–60 days delinquent, 61–90 days delinquent, over 91 days delinquent, foreclosure, REO,[9] or paid off), investor type (held in portfolio; private securitization; GNMA, FNMA, and FHLMC[10]; GNMA buyout loans; Local Housing Authority; or Federal Home Loan Bank), and actual principal balance, as well as scheduled principal balance if the borrower pays according to the original terms of the loan. Most critical to this research, the investor type variable tracks securitization decisions regarding the loan made over time, and the delinquency variable provides information on the loan's default and prepayment events. The complication in these data is that they do not identify which loans on the bank's balance-sheet are intended for sale to GSEs or non-GSEs. As discussed earlier, since the incentives of participants in these markets may differ, we will employ several methods to classify these loans appropriately before making comparisons with loans sold to GSEs or in the private market.

We also have access to variables through Home Mortgage Disclosure Act (HMDA) data. The merging of the LPS data set with HMDA data gives us access to additional information on the borrower and the lender. For example, it provides us with socioeconomic and demographic information on the borrower, such as borrower income. We are also able to use the HMDA data to control for lender differences (for example, the number of loans originated by a lender in a given year). In addition, as will become clear, we will also use and discuss Loan Performance (LP) data for some of the empirical analysis. For more details on LPS, see Keys et al. (forthcoming) and Agarwal et al. (2010, 2011), for LP, see Keys et al. (2009), and for HMDA, see Elul (2009). A brief description of the three data sets is provided in Appendix B.

To summarize, relative to LPS data, CoreLogic LoanPerformance Securities data provide data only on securitized subprime (sold to non-GSEs) mortgage data which is underrepresented in LPS. Moreover, Home Mortgage Disclosure Act data cover a wider set of loans (across GSEs and non-GSEs) but only has limited loan-level information relative to LPS. Both LPS and LoanPerformance Securities data contain subprime identification provided by each lender or loan servicer and standardized by the data source. There are two ways to potentially identify subprime loans in the HMDA data. One is based on loans with high yield spreads and another is based on HUD-identified subprime lenders.

Each of the data sets and identification methods has its merits and drawbacks. LPS data have both securitized and portfolio loans, but underrepresent the subprime market. The LoanPerformance Securities data have better coverage of the non-prime market but contain solely the securitized loans (leaving out the portfolio loans completely). The HMDA data's high-cost loan definition designates a loan as subprime if the spread between the loan's coupon and the standard index rate is higher than a certain margin, and therefore, relying on it may lead to undercounting the adjustable-rate loans with low teaser rates but high rates following the initial period. The HMDA HUD lender list definition comprises all loans originated by these lenders in a given year without further distinction at the loan level, and leaves out subprime loans originated by lenders whose main business is in the prime mortgage area.[11]

Note that our main results are derived from the LPS data set. Since these data better represent high-quality ("prime-like") loans, it is not surprising that our main results for GSEs remain robust to several robustness checks. In contrast, we find that the results on subprime loans using LPS—which underrepresents these types of loans—are sensitive to different definitions and data sets.

3.2. Comparison of coverage across data sets

We now discuss the coverage of the LPS data set used in our analysis relative to the other data sets described earlier. First, LPS coverage of the market changes over time; in particular, its coverage increased substantially after 2005. Therefore, one needs to be cautious in making longitudinal inferences from the results that span several years, especially when comparing results from the years up through 2005 and afterward. Second, subprime loans are substantially underrepresented in the data set. Table 1, Panel A compares the origination volume ascribed as prime or subprime in the different data sets and definitions for years 2004–07. It is clear that while LPS reports a similar volume of high-quality ("prime-like") loan originations as LoanPerformance, the subprime loan count from LPS is only about one-third of that reported in LoanPerformance for each of the years 2004–06. The subprime loan count reported by LoanPerformance lies between the two definitions under HMDA

[9] REO stands for real estate owned. See footnote 13 for further details on REO sales.

[10] GNMA refers to Government National Mortgage Association (Ginne Mae); FNMA refers to Federal National Mortgage Association (Fannie Mae); and FHLMC refers to Federal Home Loan Mortgage Corporation (Freddie Mac).

[11] Keys et al. (2010b) compare and contrast LPS and LoanPerformance Securities data sets. Mayer and Pence (2009) give an excellent summary on the details of the three definitions of subprime loans—namely, those identified with the LoanPerformance Securities data set criteria, those based on the HMDA higher-priced loans criterion, and those identified through the HMDA HUD lender list.

Table 1
Descriptive statistics.

Table numbers are from authors' calculation based on individual data sets. Only first liens are included. LoanPerformance prime statistics is based on the LoanPerformance TrueStandings Servicing platform. LoanPerformance subprime statistics is calculated from LoanPerformance Securities data set. High-quality/prime-like loans are defined as conventional, fixed-rate mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000. Additionally prime-like loans have FICO scores above 620 and loan-to-value ratios below 95%.

Panel A: Origination volume reported in LPS, LoanPerformance, and HMDA data (in millions)

Origination year	LPS		LoanPerformance		HMDA definition high cost loans		HMDA definition HUD lender list	
	High-quality/ "prime-like"	Low-quality/ "subprime-like"	Prime	Subprime	Prime	Subprime	Prime	Subprime
2004	6.7	0.6	8.3	1.9	11.3	1.6	10.8	2.1
2005	6.7	0.7	7.9	2.3	9.6	3.0	10.5	2.1
2006	5.8	0.6	6.7	1.8	7.6	2.8	9.1	1.4
2007	4.9	0.2	7.2	0.3	6.8	1.5	7.9	0.4

Panel B: Descriptive statistics of LPS, LoanPerformance, and HMDA data

Variable	Origination year	LPS				LPS-HMDA (High-cost loan)				LoanPerformance			
		High-quality/ "prime-like"		Low-quality/ "subprime-like"		High-quality/ "prime-like"		Low-quality/ "subprime-like"		Prime		Low-quality/ "subprime-like"	
		Full doc	Low/no doc	Full doc	Low/no doc	Full doc	Low/no doc	Full doc	Low/no doc	Full doc	Low/no doc	Full doc	Low/no doc
Low/no doc share	2004	72%	28%	97%	3%	74%	26%	95%	5%	77%	23%	66%	34%
	2005	68%	32%	84%	16%	68%	32%	81%	19%	72%	28%	63%	37%
	2006	66%	34%	85%	15%	67%	33%	84%	16%	66%	34%	62%	38%
	2007	70%	30%	87%	13%	71%	29%	86%	14%	69%	31%	66%	34%
Average FICO	2004	712	705	610	630	713	704	612	634	718	721	610	634
	2005	716	709	606	641	717	710	605	645	719	722	611	638
	2006	709	708	605	628	711	711	606	633	714	721	607	635
	2007	707	700	597	611	707	702	595	608	711	723	606	630
Average coupon rate	2004	5.46	5.06	7.38	7.26	5.49	5.24	7.34	7.25	5.24	5.12	7.25	7.36
	2005	5.73	5.21	7.52	7.76	5.78	5.38	7.40	7.75	5.49	4.82	7.46	7.61
	2006	6.45	6.14	8.32	8.75	6.49	6.27	8.32	8.75	6.28	5.81	8.36	8.59
	2007	6.48	6.38	8.42	8.40	6.47	6.42	8.29	8.35	6.39	6.33	8.54	8.70
Average LTV	2004	72.9	71.5	80.1	80.9	73.4	71.9	80.4	81.1	72.1	72.2	81.2	79.7
	2005	72.1	72.0	79.9	80.0	72.3	72.3	80.4	80.1	72.1	72.7	80.8	79.4
	2006	73.7	73.2	79.4	79.6	74.3	73.4	79.4	79.3	74.1	73.5	80.8	79.7
	2007	76.0	75.5	79.9	78.6	76.9	75.2	79.6	78.2	76.7	73.8	80.8	78.6

Table 1 (*continued*)

Panel C: Descriptive statistics for high-quality/"prime-like" full-documentation loans

	Kept in portfolio				Sold to GSEs				Sold to private label			
	2004 n=13,122	2005 n=14,606	2006 n=45,345	2007 n=61,496	2004 n=304,631	2005 n=224,020	2006 n=342,928	2007 n=278,619	2004 n=77,633	2005 n=59,402	2006 n=46,063	2007 n=25,191
FICO	729 (51.3)	729 (51.6)	731 (53.2)	733 (53.2)	732 (50.1)	727 (52.4)	727 (53.2)	725 (54.1)	732 (47.2)	730 (50.0)	722 (51.1)	728 (52.0)
Income (in $1,000s)	106.11 (199.1)	95.03 (158.3)	94.24 (151.7)	107.94 (158.4)	87.04 (102.2)	84.54 (87.9)	91.38 (93.4)	91.24 (97.6)	121.82 (153.2)	113.71 (127.3)	124.75 (157.9)	133.10 (171.9)
LTV ratio	70.46 (18.2)	70.39 (17.7)	70.48 (18.0)	72.18 (18.3)	66.24 (18.6)	66.93 (18.5)	68.65 (17.7)	71.15 (17.2)	69.99 (15.69)	70.99 (15.83)	71.00 (16.17)	70.86 (16.64)
Origination amount ($)	207,201 (143,148)	206,549 (131,545)	196,418 (119,778)	209,205 (136,443)	152,595 (78,191)	168,066 (87,509)	177,577 (94,383)	183,244 (97,384)	218,077 (156,621)	230,901 (161,735)	254,013 (176,720)	272,087 (173,578)
Conform	0.62 (0.48)	0.72 (0.45)	0.77 (0.42)	0.76 (0.43)	1.00 (0)	1.00 (0)	1.00 (0)	1.00 (0)	0.69 (0.46)	0.70 (0.46)	0.64 (0.48)	0.65 (0.48)
Jumbo	0.18 (0.39)	0.12 (0.33)	0.05 (0.22)	0.08 (0.28)	0.01 (0.07)	0.01 (0.11)	0.00 (0.05)	0.00 (0.05)	0.24 (0.43)	0.23 (0.42)	0.25 (0.44)	0.25 (0.44)
Coupon rate	5.62 (0.49)	5.68 (0.48)	6.25 (0.48)	6.36 (0.59)	5.67 (0.52)	5.83 (0.42)	6.47 (0.42)	6.43 (0.46)	5.88 (0.51)	6.00 (0.48)	6.60 (0.78)	6.44 (0.68)

Panel C reports the descriptive statistics for single-family prime full documentation loans originated between January 2004 and June 2007. Means are reported. Standard deviations are in parentheses. Only conventional, fixed-rate mortgages are included. Second liens and loans above $650,000 are excluded. High Quality/Prime-like loans are defined as conventional, fixed-rate mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000. Additionally prime-like loans have FICO scores above 620 and loan-to-value ratios below 95%.

Panel D: Descriptive statistics for high-quality/"prime-like" low/no-documentation loans

	Kept in portfolio				Sold to GSEs				Sold to private label			
	2004 n=1,732	2005 n=1,206	2006 n=2,860	2007 n=4,303	2004 n=55,943	2005 n=83,459	2006 n=138,976	2007 n=111,910	2004 n=18,610	2005 n=15,996	2006 n=17,162	2007 n=9,482
FICO	721 (48.05)	731 (48.16)	738 (44.95)	726 (47.62)	724 (48.63)	721 (51.33)	723 (51.38)	717 (53.38)	725 (43.12)	722 (46.96)	715 (47.65)	719 (50.35)
Income (in $1,000s)	105.81 (120.08)	102.47 (91.90)	113.15 (103.89)	141.23 (137.72)	87.53 (73.81)	83.52 (71.82)	91.46 (83.15)	91.03 (93.20)	137.85 (135.27)	118.03 (109.25)	134.81 (121.67)	147.76 (155.41)
LTV ratio	68.67 (16.86)	71.19 (17.75)	69.06 (17.98)	69.03 (17.89)	66.65 (18.15)	69.44 (17.11)	70.33 (16.42)	72.11 (16.18)	69.65 (15.35)	71.25 (14.48)	71.37 (15.44)	71.36 (15.35)
Origination amount ($)	207,911 (131,318)	212,638 (140,827)	219,451 (131,454)	281,805 (174,702)	172,009 (82,668)	166,174 (88,279)	181,161 (96,361)	181,206 (98,473)	242,890 (167,711)	245,501 (165,059)	273,118 (172,918)	309,165 (184,068)
Conform	0.75 (0.43)	0.79 (0.41)	0.87 (0.34)	0.67 (0.47)	1.00 (0)	1.00 (0)	1.00 (0)	1.00 (0)	0.60 (0.49)	0.63 (0.48)	0.62 (0.49)	0.53 (0.50)
Jumbo	0.13 (0.34)	0.14 (0.34)	0.08 (0.27)	0.26 (0.44)	0.00 (0.07)	0.01 (0.10)	0.00 (0.04)	0.00 (0.06)	0.31 (0.46)	0.27 (0.45)	0.27 (0.44)	0.36 (0.48)
Coupon rate	5.80 (0.63)	5.92 (0.57)	6.59 (0.65)	6.90 (0.89)	5.70 (0.48)	5.89 (0.47)	6.49 (0.45)	6.47 (0.49)	6.07 (0.56)	6.08 (0.43)	6.78 (0.72)	6.58 (0.69)

Panel D reports the descriptive statistics for single-family prime low/no-documentation loans originated between January 2004 and June 2007. Means are reported. Standard deviations are in parentheses. Only conventional, fixed-rate mortgages are included. Second liens and loans above $650,000 are excluded. High-quality/prime-like loans are defined as conventional, fixed-rate mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000. Additionally prime-like loans have FICO scores above 620 and loan-to-value ratios below 95%.

Panel E: Descriptive statistics for "subprime-like" loans

	Kept in Portfolio				Sold to GSEs				Sold to Private Label			
	2004 $n=153$	2005 $n=1,258$	2006 $n=238$	2007 $n=764$	2004 $n=0$	2005 $n=0$	2006 $n=0$	2007 $n=0$	2004 $n=442$	2005 $n=4,950$	2006 $n=10,371$	2007 $n=1,281$
FICO	606 (55.4)	615 (53.2)	626 (60.3)	612 (58.3)					630 (60.4)	620 (55.2)	625 (56.2)	617 (54.4)
Income (in $1,000s)	77.70 (63.0)	81.65 (176.8)	87.67 (89.9)	84.28 (150.2)					109.97 (257.2)	78.03 (123.2)	84.74 (137.2)	80.14 (95.3)
LTV ratio	74.75 (11.1)	76.55 (11.3)	76.76 (13.9)	73.60 (14.5)					78.29 (12.2)	78.21 (13.5)	76.46 (14.1)	75.44 (15.1)
Origination amount ($)	192,710 (122,160)	159,944 (105,193)	206,834 (139,727)	200,121 (133,052)					125,613 (96,280)	144,992 (109,235)	179,996 (123,375)	191,505 (131,215)
Conform	0.07 (0.26)	0.14 (0.34)	0.16 (0.37)	0.11 (0.32)					0.22 (0.41)	0.13 (0.33)	0.14 (0.35)	0.09 (0.29)
Jumbo	0.14 (0.35)	0.06 (0.24)	0.11 (0.31)	0.09 (0.29)					0.05 (0.22)	0.06 (0.23)	0.06 (0.25)	0.08 (0.28)
Low/no documentation	0.14 (0.35)	0.21 (0.41)	0.21 (0.41)	0.09 (0.29)					0.34 (0.47)	0.25 (0.43)	0.18 (0.39)	0.15 (0.36)
Coupon rate	7.33 (1.19)	7.71 (1.12)	8.46 (1.58)	8.47 (1.53)					8.10 (1.18)	8.03 (1.21)	8.25 (1.40)	8.28 (1.39)

Panel E reports the descriptive statistics for single-family subprime loans originated between January 2004 and June 2007. Means are reported. Standard deviations are in parentheses. None of the subprime loans in our data set were sold to the GSEs. Subprime-like loans are as defined as mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000 and reported as subprime by the LPS database (FICO < 620 and having Grade "B" or Grade "C").

(which we go into further detail about later). This issue of LPS undercounting the subprime sector is also documented in Elul (2009) and Keys et al. (forthcoming). Third, compared with the LoanPerformance Securities data set, the LPS data set also underreports loans with low or no documentation in the subprime sector. As seen in Panel B of Table 1, for each year shown, the low/no documentation share in the subprime sector reported in the LoanPerformance data is more than double that reported in LPS. The subsample from LPS that merges with HMDA successfully further reduces the sample size, usually by 40–60%, depending on the segment. To address the concern that the LPS sample may not be representative of the full market, we also compare in Panel B of Table 1 the key credit characteristics of average FICO score, loan-to-value ratio, and mortgage coupon rate at origination for each sector of prime and subprime loans and the documentation type across data sets. It shows that merging with HMDA does not change the average characteristics significantly for each sector in LPS. Compared with LoanPerformance, the LPS average prime loan has a slightly lower credit score, but otherwise similar features. Larger differences exist between LPS and LoanPerformance average coupon rates at origination—which, given the similarities in FICO and LTV statistics, could be due to the different loan product mixes in each data set, other credit characteristics, or pricing differences. In Section 5, we deal with the data issue by employing various robustness tests, including performing analysis on a different subprime data set and using different subprime definitions, such as the HUD subprime lender list.

3.3. Sample construction

We focus on conventional, fixed-rate mortgages for single-family residences and condos originated between January 2004 and December 2007. Second mortgages, home equity lines of credit (HELOCs), and loans above $650,000 are excluded. We choose to examine only fixed-rate mortgages to reduce model specification errors and enhance comparability across different market segments and time periods. Although we allow both prime and subprime loans to enter the data set, we impose additional restrictions on the prime loans. We confine the analysis to high-quality loans with FICO scores above 620 and loan-to-value ratios below 95%. Ideally, we would have liked to have loans that were intended for GSEs and compared the quality of these loans relative to those on the lender balance-sheets to reduce the issues of looking at loans across different markets with differences in incentives for participants at origination and post-origination stages. These criteria, though not sufficient, enable us to get closer to this ideal scenario (we are left with both GSE and non-GSE securitized loans). Our other sample, "subprime-like" loans, were purchased only by the non-GSEs, and therefore, no additional constraints are required for this group of loans. We acknowledge that while trying to make the subprime loan sample homogeneous, we lose a significant number of observations in the structural analysis. However, it is reassuring that in the reduced-form analysis when we include a much larger sample, the results of structural analysis are confirmed.

Over time, the LPS data set has grown dramatically with the addition of new reporting servicers. The addition of these servicers to the data set means that both seasoned loans and new originations are included, but only information available *after* servicers sign on with LPS Applied Analytics, Inc., is in the LPS data set. This could potentially left-censor the data because earlier loans that have defaulted or prepaid prior to the servicer beginning reporting will not be included, while loans that have remained current will. To reduce the extent of left-censoring in the data, we eliminate from our sample loans that entered LPS more than 4 months after origination, consistent with Piskorski et al. (2010).[12]

Another concern is to distinguish between the loans intended for securitization and those that the lenders were able to securitize. The originator may be forced to keep a loan on their books if it goes to delinquency before it can be securitized, although the loan may have been intended for securitization at origination. To correct for this, we calculate the average time it takes between a loan's origination and securitization for the prime and subprime sectors separately, and remove from our sample any loans that were delinquent within this period. The average time to securitize is 4–5 months for subprime loans and 5–6 months for prime loans, consistent with those reported in Keys et al. (forthcoming).[13]

For the "prime-like" loans, we categorize the loans as being held in portfolio, sold to the GSEs, or sold to private issuers. As noted before, this is less than ideal, but in our robustness tests we are able to confirm that this contamination does not adversely affect our inferences. GSE loans include those with Ginnie Mae, Fannie Mae, and Freddie Mac, as well as the Ginnie Mae buyout loans. We define "investor type" as the most common type of investor within the 12-month period after origination. In our sample, 78.3% of the loans with an observable investor type are classified as GSE loans, 7.4% as held in portfolio, and 14.3% as privately securitized.

We define a loan in default if it is over 61 days delinquent, is foreclosed, or has experienced an REO sale.[14] A loan is considered prepaid if it has been paid in full in a month when the scheduled principal balance amount is greater than $500 and the prepayment is not preceded by delinquency events.[15] We also create a dummy variable to indicate whether a loan conformed to GSE standards at the time of origination.

Since we do not have an official GSE credit standard, we follow the definition proposed by Ambrose et al. (2005). We label a loan as conforming if it was held by one of the agencies above at some point during the 12 months after origination. If the loan was not held by a GSE, we label it as conforming if the FICO score was higher than 660, the origination amount was below the conforming limit for that geographic area and time, and the loan has private mortgage insurance if the LTV ratio is above 80. Overall, 92.2% of the loans in our sample are identified as conforming according to this definition.

3.4. *Descriptive statistics*

We now discuss summary statistics of our sample. Specifically, panels C and D of Table 1 provide the descriptive statistics for the high-quality ("prime-like") loan sample, broken down by documentation type first and then by origination year and investor type. The average FICO score for loans sold to the GSEs shows a slight decline over the years, while the LTV ratio for these loans increases every year. There is also a substantial increase in the share of loans that are low or no documentation from 2004 through 2007. The credit quality of the loans that are privately securitized show a similar pattern as those sold to the GSEs.

Note that the year 2007 shows a divergence in the trends between the two channels: The average FICO score for the privately securitized loans improves over the previous year, whereas it is the opposite for the loans sold to the GSEs. In conjunction with the observed drop in the share of privately securitized loans in the prime market, this signals credit tightening at the private label channel. The quality of prime loans held in portfolio appears to improve slightly from 2004 through 2006, with signs of deterioration in 2007. For full-documentation prime loans (Table 1, Panel C), the average FICO score increases slightly every year. The proportion of prime loans considered to be of conforming quality increases over time but then drops in 2007. In a similar fashion, the LTV ratio hovers at around 70% for every year except for a higher level in 2007. For low/no documentation prime loans (Table 1, Panel D), both average FICO score and conforming loan share decline in 2007 as well. Lenders choose to keep a much smaller portion of loans that are low or no documentation, and the share relative to its portfolio declines from 2004 through 2006, yet increases in 2007. Lenders retain a steady share of about 3–4% of the prime loans that they originate in 2004 and 2005. In 2006, as the housing bubble reaches its peak and house prices start to decline in the latter half of the year, about 8% of prime loans originated were held in portfolio. In 2007, when the downturn in real estate becomes widespread, the rate of prime loans being held in portfolio increases drastically to 13.4%, indicating the increasing difficulty of securitization.

Table 1, Panel E reports the summary statistics for the "subprime-like" loans. Only five to 90 subprime loans in our data set were purchased by the GSEs in any year, so their statistics are not reported. The great majority of them were privately securitized. As expected, the average

[12] Alternatively, we find that using a 12-month cutoff point, as in Elul (2009), does not materially change our results.

[13] Elul (2009) excludes from his sample any loans delinquent within 3 months of origination in one version of his analysis.

[14] According to the Office of Thrift Supervision (OTS), a loan is in delinquency if a monthly payment is not received by the loan's due date. This is a slightly less strict definition of delinquency than the Mortgage Bankers Association's definition. An REO sale follows an unsuccessful foreclosure when a buyer for the property cannot be found and the mortgage lender repossesses the property to sell separately.

[15] The minimum principal balance of $500 is used to differentiate a prepayment from a scheduled final month's payment of a loan.

FICO score for the subprime loans is about 100 points fewer than the average FICO score for the prime loans, and the average LTV ratio is about 10 percentage points higher for the subprime loans. Compared with the subprime loans held in portfolio, the subprime loans that are privately securitized tend to have a slightly higher average FICO score for every year in the sample and the origination unpaid principal balance (UPB) tends to be lower. For every other variable of interest, subprime loans that are kept in portfolio seem comparable to the subprime loans that are privately securitized. The year 2006 saw the largest portion of subprime fixed-rate loans securitized. In 2005, lenders kept roughly 20% of the subprime loans originated in portfolio, but in 2006, this proportion drastically dropped to only 2%. Similar to the prime market in 2007, the subprime market in 2007 shows a significant increase in the share of loans that lenders keep in their portfolios, up to a high of 37%, which reflects the fact that securitization channels have tightened.

4. Methodology and results

In this section, we empirically examine the determinants of lenders' choice to securitize loans, focusing on the quality of loans they sell to investors in the secondary mortgage market relative to ones they retain on their balance-sheets.

There are two main challenges for the empirical tests to investigate these issues. The first is the issue of endogeneity—that is, observed securitization and loan performance may be jointly determined—with each affecting the outcome of the other. As a result, simply regressing one outcome on the other is likely to constitute an inadequate analysis and may well lead to biased conclusions. To circumvent this problem, the central identification of the paper is based on imposing certain structure to the securitization process (as in Ambrose et al., 2005). This structure follows from the institutional background involving various participants in the securitization of mortgage credit. We present the structural approach as the main result, but we also employ reduced-form regressions for robustness.

The other challenge emerges due to market segmentation. As discussed earlier, we will employ several strategies to ensure that our inferences are not affected due to not being able to infer which of the bank-held loans in LPS is directly comparable to loans securitized to GSEs or to private issuers (which differ dramatically for reasons highlighted earlier).

4.1. Structural approach outline

To determine the relationship between a lender's securitization choice and expected loan performance, we adopt a structural approach consisting of three steps. First, we map the loan-level observables to default and prepayment probabilities. The idea here is that we model the loan quality, regardless of whether it is sold or held on the balance-sheet, based on observables. In other words, we infer the quality based on loan characteristics for all the loans in the distribution. This allows us to construct the estimates of loan quality—independent of the lender's selling decision—since both loans sold and on balance sheet are considered together. Second, we take these estimates and perform an out-of-sample forecast of default and prepayment probabilities. Finally, we correlate the estimated default and prepayment probabilities in out-of-sample loans with the actual observed securitization outcomes for these loans.

To conduct our analysis, we first segment the data into "prime-like" and "subprime-like" sectors by origination year to ensure alignment between securitized loans and those kept in portfolio.[16] For "prime-like" and "subprime-like" loans and for each year of origination, we divide the sample population into a random 75% estimation sample and a 25% holdout sample. First, based on the 75% estimation sample, we construct a hazard model, using the observed default and prepayment outcomes in the next 24 months. In the second step, we apply the coefficients obtained from the first step to the holdout sample consisting of the other 25% of the population, and calculate its expected default and prepayment probabilities. In the last step, we regress the observed securitization outcome on the loans in the holdout sample on their expected default and prepayment probabilities obtained from the previous two steps, as well as on other variables, controlling for the pricing of the loan and the market environment at the time of origination. Based on these results, we can assess the relationship between a loan's expected performance and the lender's securitization choice.

To account for variations in lenders' formation of expectations regarding loan performance, we use a "rational expectations" approach, where we assume that the lender has perfect foresight regarding the contribution of loan characteristics to the outcome probabilities, with the expectations for loans in the holdout sample being formed in the same way as those in the estimation sample. This way, we use parameters estimated from loans originated in the same year to apply to the holdout sample. For example, to form expected prepayment and default probabilities in the next 24 months for loans originated in 2006, we use parameters estimated from a different sample, also originated in 2006, observed through 2008.[17]

[16] We perform propensity score tests for each year to verify that the "prime-like" and "subprime-like" segmentation of the portfolio loans is as expected. The score model is based on each year's securitized GSE and non-GSE loans. The analysis reveals some contamination in the "prime-like" loans since some of the loans in this sample end up with non-GSEs. We discuss this issue in the robustness section and conclude this contamination does not alter our inferences.

[17] We also performed an alternative approach to modeling lenders' expectations, the "adaptive expectations" approach, where lenders are assumed to form their expectations based on their experiences up to the time of loan origination. In other words, they draw conclusions for the 2006 loans by learning from the performance of loans originated before 2006. For this case, we lag the estimation sample by 2 years compared with the holdout sample, so that parameters estimated will be from a 24-month period before origination of the new loan. We use parameters estimated from the 2004 and 2005 full populations to the holdout sample for 2006 and 2007, respectively. The results from this approach

4.2. Default and prepayment estimation

We model the loans' default and prepayment probabilities in a competing risk hazard framework. At each point in time, the borrower may decide to terminate the loan by refinancing or moving and prepaying the balance owed, or the borrower may decide to default on the loan (giving away the house to the lender). If neither of these events occurs at that point, the loan survives for another period, and the observation is considered censored. We implement the hazard model as a multinomial logit model with a quadratic baseline function for the timing of an event.

We control for borrower and mortgage characteristics in the default and prepayment estimation. To ensure consistency between the estimation and forecast, the explanatory variables are taken as of the time of origination. These variables are borrower credit score (FICO), borrower income (Income), origination loan-to-value ratio (LTV ratio), whether the loan was considered conforming by GSE standards (Conform), whether the loan amount was above conforming loan limits (Jumbo), and loan underwriting documentation level (Low documentation). We also include the time since origination and its squared term to control for loan age effects. As discussed before, we include all the loans (bank-held and securitized) to get the estimates for the entire distribution of loans.

The estimation coefficients are reported in panels A and B of Table 2 for "prime-like" and "subprime-like" loans. Borrower credit score (FICO) is negatively correlated with default probability for both the prime and subprime sectors and across all origination years 2004–07, indicating that borrowers with better credit standing are less likely to default. We also find a reduced probability to prepayment for borrowers with higher FICO scores, except for subprime loans originated in 2005 and 2007, where the effect is insignificant, and for prime loans originated in 2007. The 2007 cohorts were originated at the start of the downturn, the positive correlation between FICO and prepayment probability signals the credit tightening subsequent to the downturn, and only those with good credit standing can qualify for refinance following the downturn.

Higher-income borrowers tend to have reduced probability of default for loans originated in 2004–06, and an increased probability of default for loans originated in 2007. This may be due to the fact that particular geographic areas with the most severe residential loan delinquencies—namely, California, Florida, Nevada, and Arizona—were also areas where housing prices had risen the most and thus required higher income to qualify for mortgages. Higher income is also associated with higher prepayment propensity among prime borrowers for most of the years, although among subprime borrowers the income effect on prepayment is insignificant.

(*footnote continued*)
are largely consistent with the main approach and are available upon request.

Higher loan-to-value (LTV) ratio measured as of origination is found to contribute positively to borrower default probabilities for both prime and subprime loans in all years. A high LTV increases the probability that the house will go underwater, adding to the borrower's incentive to default. A high LTV also restricts the borrower's ability to refinance in cases of financial distress, as we estimate that higher LTV loans are less likely to prepay in nearly all cases.

Loans with reduced documentation require less paperwork in the underwriting process, and are generally issued to borrowers with variable or unverifiable income, such as borrowers who are self-employed or citizens of another country. LaCour-Little and Yang (2010) find that such loans issued during the most recent housing boom have a higher propensity of default. Our results confirm that low- or no-documentation loans have a higher default probability; a stronger effect among such loans is observed in the subprime sector. These loans are also more likely to be prepaid.

4.3. Cumulative default and prepayment probability

We use the estimated coefficients to calculate the expected cumulative 12-month and 24-month prepayment and default probabilities for each loan in the holdout sample. The general form of the calculation of probability of outcome for loan i for each of the three outcomes at each point in time is

$$P_i(outcome=j)=\left(\frac{\exp(\alpha_j+x'_{1,2,\dots,n}\beta_j)}{1+\sum_{j=1}^{2}\exp(\alpha_j+x'_{1,2,\dots,n}\beta_j)}\right),$$

where $j=1,2$ (default, prepay). And to ensure that probabilities sum up to 1, we use the following:

$$P_i(outcome=survive)=\left(\frac{1}{1+\sum_{j=1}^{2}\exp(\alpha_j+x'_{1,2,\dots,n}\beta_j)}\right)$$

where α is the constant, X is the value of the independent variables, and β is the vector of coefficient estimates.

Panels A and B of Table 3 present the cumulative expected probabilities calculated assuming lenders have rational expectations. In the "prime-like" sample, compared with loans held in portfolio, loans sold to GSEs consistently have lower expected default probabilities. Loans sold to GSEs have the highest prepayment probabilities in all years, while loans held in portfolio and loans privately securitized have lower prepayment probabilities. In the "subprime-like" sample, loans that are securitized to non-GSEs have the lowest default probabilities of the three types in 2004 and 2005, but then the highest default probabilities in 2006 and 2007. Also, in this sample, securitized loans have lower expected default probabilities in all years of our sample.

4.4. Decision to securitize

In the last step, we use the observed securitization choice from the holdout sample to model the adverse selection. Our base model includes the expected default probability and expected prepayment probability calculated in the previous steps. In addition, we control for the

Table 2
Competing risks model of mortgage outcome.

This table states the results from a competing risks model of the outcome to prepay, default, or remain current on a given mortgage as estimated by a multinomial logit model. The dependent variable is whether a loan experienced default, prepayment, or remained current within 24 months of origination. The independent variables are information available to lenders at the time of underwriting and include the borrower's FICO score (FICO), the borrower's income (Income), the loan-to-value ratio for the mortgage (LTV ratio), whether or not the loan conforms to GSE standards (Conform), whether or not the loan amount exceeds GSE limits (Jumbo), and whether the loan application was low- or no-documentation (Low documentation).

Panel A: Competing risks model of mortgage outcome for high-quality/"prime-like" loans

	2004		2005		2006		2007	
Default outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	0.408	0.262	1.066**	0.001	−0.082	0.654	−0.696**	0.000
FICO	−0.018**	0.000	−0.018**	0.000	−0.017**	0.000	−0.015**	0.000
Income	−0.008**	0.000	−0.006**	0.000	−0.001**	0.000	0.000	0.879
LTV ratio	0.034**	0.000	0.032**	0.000	0.032**	0.000	0.030**	0.000
Conform	−0.208**	0.003	−0.079	0.270	−0.209**	0.000	−0.001	0.969
Time (in months)	0.342**	0.000	0.216**	0.000	0.288**	0.000	0.324**	0.000
Time^2	−0.009**	0.000	−0.005**	0.000	−0.006**	0.000	−0.007**	0.000
Jumbo	−0.174	0.242	0.054	0.634	0.315**	0.000	0.496**	0.000
Low documentation	0.262**	0.000	0.288**	0.000	0.135**	0.000	0.124**	0.000
Prepay outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	−4.930**	0.000	−3.088**	0.000	−5.315**	0.000	−8.211**	0.000
FICO	−0.005**	0.000	−0.004**	0.000	−0.001**	0.000	0.004**	0.000
Income	0.000**	0.000	0.000**	0.000	0.000**	0.000	0.000**	0.000
LTV ratio	0.004**	0.000	−0.001**	0.000	−0.005**	0.000	−0.009**	0.000
Conform	−0.018	0.487	−0.076**	0.008	0.282**	0.000	0.422**	0.000
Time (in months)	0.380**	0.000	0.158**	0.000	0.148**	0.000	0.109**	0.000
Time^2	−0.011**	0.000	−0.004**	0.000	−0.004**	0.000	−0.002**	0.000
Jumbo	−0.023	0.448	−0.290**	0.000	0.083**	0.010	0.066	0.056
Low documentation	0.273**	0.000	0.094**	0.000	0.088**	0.000	0.042**	0.000
	Observations	8,056,598	Observations	6,744,001	Observations	9,855,260	Observations	8,015,286
	Pseudo-R^2	0.042	Pseudo-R^2	0.021	Pseudo-R^2	0.025	Pseudo R^2	0.045

Panel A results are estimated from a 75% sample of prime loans taken for each year of data. High-quality/prime-like loans are defined as conventional, fixed-rate mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000. Additionally, prime-like loans have FICO scores above 620 and loan-to-value ratios below 95%. ** Significant at 1% level, * significant at 5% level.

Panel B: Competing risks model of mortgage outcome for "subprime-like" loans

	2004		2005		2006		2007	
Default outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	−5.285*	0.013	−5.576**	0.000	−4.311**	0.000	−2.876**	0.000
FICO	−0.010**	0.002	−0.008**	0.000	−0.010**	0.000	−0.011**	0.000
Income	0.000	0.783	0.000	0.924	0.000	0.143	0.000	0.472
LTV ratio	0.049**	0.001	0.035**	0.000	0.040**	0.000	0.034**	0.000
Conform	−0.871	0.132	0.129	0.463	0.139	0.203	0.193	0.280
Time (in months)	0.273*	0.013	0.325**	0.000	0.307**	0.000	0.252**	0.000
Time^2	−0.007	0.064	−0.009**	0.000	−0.009**	0.000	−0.006**	0.000
Jumbo	−1.118	0.276	0.603**	0.000	0.447**	0.000	0.272	0.157
Low documentation	0.517	0.114	0.458**	0.000	0.793**	0.000	0.532**	0.000
Prepay outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	−3.162*	0.013	−5.293**	0.000	−2.075**	0.000	−4.729**	0.000
FICO	−0.004*	0.038	0.000	0.566	−0.004**	0.000	−0.002	0.212
Income	0.000	0.934	0.000	0.101	0.000	0.418	0.000	0.443
LTV ratio	−0.009	0.194	−0.011**	0.000	−0.020**	0.000	−0.014**	0.003
Conform	−0.155	0.605	−0.237*	0.041	0.008	0.941	−0.059	0.823
Time (in months)	0.260**	0.000	0.299**	0.000	0.240**	0.000	0.277**	0.000
Time^2	−0.007**	0.005	−0.010**	0.000	−0.009**	0.000	−0.009**	0.000
Jumbo	0.096	0.792	−0.242	0.102	−0.171	0.258	−0.696	0.086
Low documentation	0.287	0.162	0.268**	0.000	0.358**	0.000	0.197	0.385
	Observations	10,901	Observations	96,724	Observations	164,975	Observations	34,380
	Pseudo-R^2	0.050	Pseudo-R^2	0.033	Pseudo-R^2	0.049	Pseudo-R^2	0.055

Panel B results are estimation from a 75% sample of subprime loans taken for each year of data. Subprime-like loans are as defined as mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000 and reported as subprime by the LPS database (FICO < 620 and having Grade "B" or Grade "C"). ** Significant at 1% level, * significant at 5% level.

mortgage's yield spread, whether the loan is above the conforming loan limit, the market credit spread premium (log_credit_spd), the shape of the yield curve (log_yield_curve), and interest rate volatility (log_sigma_int).

Besides termination risks, we need to take into consideration the yield on individual loans in determining lenders' decisions to securitize. A lender may choose to keep a loan in its portfolio if the loan is overpriced relative to its risk. We measure a loan's pricing by the difference between a loan's coupon rate at origination and the contemporaneous yield on the 10-year Treasury bond. The steepness of the yield curve is measured by the ratio of the 10-year Treasury bond rate to the one-year Treasury note rate. The credit spread premium is defined as the difference between the AAA bond index and the Baa bond index. We proxy for interest rate volatility by the standard deviation in the one-year Treasury bond rate 15 months before the origination of a loan. This is estimated in a multinomial logit equation that takes the general form as follows:

$$\log\left(\frac{\pi^j}{\pi^r}\right) = \alpha_j + x'_{1,2,\ldots,n}\beta_j$$

In our estimation, π^j ($j=1,2$) represents the probability of the two outcomes of interest—that is, sold to GSEs and securitized through private issuers—and π^r is the residual probability of being kept in portfolio. The model reduces to a logit model for the subprime sector, where there are only two outcomes—that is, sold to private label securities or kept in portfolio. For the prime-like loan sample, this equation is run with lender fixed effects, with the loans of the same holding company given the same identifier.[18] To test for a small lender effect, we combine all lenders that issued fewer than 20 loans in that particular year and assign them one identifier. Under the rational expectations assumption, the cumulative default and prepayment probabilities are estimated using parameters from observed performance on loans originated in the same year. The results are summarized in panels A and B of Table 4.

Our results show that, for the prime-like sample, during the pre-crisis years lenders are less likely to sell a loan to a GSE if the loan is expected to have a higher default probability. The coefficients on the expected cumulative default probability (cumdefaultprob_24) are negative and significant for each year from 2004 through 2006. The negative relationship between default probability and probability of securitization weakens through the years examined. The effect is stronger for the private label securities than the GSEs. The year 2007 is a turning point. As we know now, during the late 2006 through 2007 period, the house price downturn started, credit standards were tightened, subprime lending and securitization came to a halt, and private label securitization dwindled in volume relative to GSE securitization. Over this period, we find a strong reversal effect in lenders' strategies for the loans that they originate such that loans securitized with GSEs have higher expected default rates than those kept in portfolio. In return for retaining higher-default-risk loans, lenders were more likely to securitize loans with higher expected prepayment probabilities to the investors through GSEs. This is seen in the positive and significant coefficients on the expected cumulative prepayment probability (cumprepayprob_24) variable. Note that we are reluctant to interpret the coefficients on the private securitized market for the "prime-like" loans due to the issue of contamination discussed earlier.

In the subprime sector, lenders are less likely to securitize higher-default-risk loans in 2007, but are indifferent to the default probabilities before that. Expectations about prepayment probabilities are not a significant factor in the securitization of subprime loans under either expectations model in any of the years.

We note here that the behavior of lenders in the pre-crisis and during-crisis periods are quite different. They can, however, be rationalized relatively clearly. During the boom period, the benefits of the expected continued business relationship with the GSEs outweighed gain in deviating from their usual strategy. In times when the market is in distress and the survival of the lender is questionable, the benefits from a gain in the current period by cutting losses outweigh future concerns regarding reputational risk, and the lenders may deviate from the behavior during the boom times. In addition, after the contraction of the private securitization market, most of the take-up in the mortgage market was by the GSEs. As a result, the incentives of participants and their behavior underwent a dramatic change—which also manifests itself in these results. The significance of 2007 being different from the previous years is also observed in Keys et al. (forthcoming) and Elul (2009).

Our results regarding default risk and securitization are consistent with Ambrose et al. (2005) despite differences in the data sources. Instead of being based on one lender as in Ambrose et al. (2005), our results are based on a data set involving more than 4,500 lenders. Furthermore, our results show that the results are applicable to a different time period from the one studied by Ambrose et al. (2005). Our data source is similar to that of Elul (2009)—we too use the LPS Analytics data set—but our findings are contrary to his findings. We find that the default rates of securitized loans are lower than portfolio loans for each of the years 2004–06, and in addition, we do find evidence that lenders are more likely to securitize loans that have higher expected default rates with the GSEs in 2007. The difference in our results is likely due to differences in methodologies. Unlike Elul (2009), we construct a model of lenders' expectations of default and prepayment probabilities for the loans they originate, and use these expectations to estimate the lenders' choice of whether to securitize a loan. We also control for the spread that the loan enjoyed over the ten-year Treasury rate and the pricing of risk in the market at the time, as the pricing of the loan's risk and the riskiness of the loan are likely to impact the lender's securitization decision. A third difference is that we construct the test for each prime/subprime segment and origination year separately

[18] Loans from originating lenders with the same holding company (e.g., Wells Fargo California and Wells Fargo Washington) are grouped together.

Table 3
Predicted cumulative default and prepayment probabilities.

Panel A: Predicted cumulative default and prepayment probabilities for loans under rational expectations

	2004			2005			2006			2007		
	Held in portfolio $n=3{,}691$	Sold to GSEs $n=90{,}265$	Sold as private label $n=24{,}260$	Held in portfolio $n=3{,}960$	Sold to GSEs $n=76{,}890$	Sold as private label $n=18{,}845$	Held in Portfolio $n=12{,}150$	Sold to GSEs $n=120735$	Sold as private label $n=15{,}734$	Held in portfolio $n=16{,}529$	Sold to GSEs $n=97{,}881$	Sold as private label $n=8{,}591$
Mean cumulative default probabilities												
Month 12	0.29%	0.21%	0.19%	0.37%	0.37%	0.33%	0.49%	0.46%	0.57%	1.10%	1.20%	1.24%
Month 24	1.13%	0.82%	0.74%	1.28%	1.29%	1.15%	2.75%	2.59%	3.23%	6.20%	6.75%	7.01%
Mean cumulative prepayment probabilities												
Month 12	3.94%	3.87%	3.92%	5.35%	5.59%	5.25%	6.45%	6.99%	6.47%	7.09%	7.37%	6.74%
Month 24	14.84%	14.61%	14.78%	14.14%	14.86%	13.97%	17.65%	19.13%	17.71%	21.77%	22.63%	20.69%

Panel A reports the predicted cumulative default and prepayment probabilities for prime loans both 12 and 24 months after origination. The probabilities are calculated for each loan in the holdout sample using the coefficients estimated from the same year estimation sample as reported in Table 2 Panel A. High-quality/prime-like is defined as conventional, fixed-rate mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000. Additionally prime-like loans are defined as loans with FICO scores above 620 and loan-to-value ratios below 95%.

Panel B: Predicted cumulative default and prepayment probabilities for "subprime-like" loans under rational expectations

	2004			2005			2006			2007		
	Held in portfolio $n=34$	Sold to GSEs $n=0$	Sold as private label $n=89$	Held in portfolio $n=313$	Sold to GSEs $n=0$	Sold as private label $n=1{,}194$	Held in portfolio $n=66$	Sold to GSEs $n=0$	Sold as private label $n=2{,}578$	Held in portfolio $n=184$	Sold to GSEs $n=0$	Sold as private label $n=277$
Mean cumulative default probabilities												
Month 12	3.26%		4.17%	4.48%		4.74%	7.02%		7.10%	8.50%		8.34%
Month 24	11.49%		14.63%	15.11%		15.96%	24.19%		24.48%	34.19%		33.57%
Mean cumulative prepayment probabilities												
Month 12	9.89%		8.78%	9.86%		9.86%	8.38%		8.11%	4.37%		4.48%
Month 24	33.34%		29.58%	26.89%		26.89%	18.54%		17.94%	11.36%		11.64%

Panel B reports the predicted cumulative default and prepayment probabilities for subprime loans both 12 and 24 months after origination. The probabilities are calculated for each loan in the holdout sample using the coefficients estimated from the same year estimation sample as reported in Table 2 Panel B. Subprime-like loans are as defined as mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000 and reported as subprime by the LPS database (FICO < 620 and having Grade "B" or Grade "C").

Table 4
Probability of securitization.

This table reports the coefficients of a multinomial logit model, which estimates the probability that a loan will be sold to the GSEs or privately securitized. This regression was estimated using the holdout sample that was created from the same origination years as the estimation samples. The independent variables are the yield spread at origination (yield_spread), whether or not the loan is above GSE loan limits (jumbo), the difference between the AAA bond index and the Baa bond index at the time of origination (credit_spread), the ratio between the 10-year Treasury rate and the 1-year Treasury rate (yield_curve), the interest rate volatility over the 24 months before origination (sigma_int), the cumulative 24-month prepayment and default probabilities (cumprepayprob_24 and cumdefaultprob_24), as well as the dummy variables to indicate whether the loans had a high yield spread (high_spd) or a low yield spread (low_spd). Lender fixed effects are also included but their coefficients are not reported here. **Significant at 1% level, *significant at 5% level.

Panel A: Probability of securitization for high-quality/"prime-like" loans under rational expectations

	2004		2005		2006		2007	
GSE outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	5.109**	0.000	7.711**	0.000	10.472**	0.000	2.360**	0.000
Yield_spread	0.074	0.065	0.979**	0.000	0.935**	0.000	0.171**	0.000
Jumbo	−1.988**	0.000	−1.385**	0.000	−1.465**	0.000	−2.279**	0.000
Credit_spread	−1.553**	0.000	−0.718	0.151	−5.661**	0.000	−1.146**	0.000
Yield_curve	0.363**	0.000	−0.533	0.191	−10.024**	0.000	3.690**	0.000
Sigma_int	0.219	0.530	−1.864**	0.000	9.478**	0.000	−0.836**	0.000
Cumprepayprob_24	4.171**	0.000	14.936**	0.000	24.749**	0.000	4.560**	0.000
Cumdefaultprob_24	−40.733**	0.000	−18.612**	0.000	−5.448**	0.000	3.081**	0.000
Low documentation	−0.077	0.080	0.482**	0.000	0.314**	0.000	0.457**	0.000
Private label outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	7.158**	0.000	11.397**	0.000	−1.711	0.253	5.975**	0.000
Yield_spread	1.148**	0.000	2.051**	0.000	1.859**	0.000	0.630**	0.000
Jumbo	0.224**	0.000	0.660**	0.000	1.050**	0.000	0.434**	0.000
Credit_spread	−4.169**	0.000	−4.108**	0.000	−3.181**	0.000	1.283**	0.000
Yield_curve	−0.147	0.083	−1.888**	0.000	2.088	0.090	−4.133**	0.000
Sigma_int	−2.683**	0.000	−5.292**	0.000	7.098**	0.000	0.557**	0.004
Cumprepayprob_24	6.047**	0.000	4.929**	0.000	5.794**	0.000	−0.391	0.252
Cumdefaultprob_24	−55.826**	0.000	−39.671**	0.000	−6.427**	0.000	−2.147**	0.000
Low documentation	0.299**	0.000	0.520**	0.000	0.549**	0.000	0.443**	0.000
	Observations	118,216	Observations	99,695	Observations	148,619	Observations	123,001

Panel A reports the coefficients of a multinomial logit model, which estimates the probability that a prime loan will be sold to the GSEs or privately securitized. High-quality/prime-like loans are defined as conventional, fixed-rate mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000. Additionally, prime-like loans have FICO scores above 620 and loan-to-value ratios below 95%.

Panel B: Probability of securitization for "subprime-like" loans under rational expectations

	2004		2005		2006		2007	
Private label outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	30.442	0.233	14.695**	0.000	−38.112**	0.003	271.500**	0.001
Yield_spread	1.630**	0.000	0.233**	0.001	−0.359**	0.001	0.587*	0.022
Jumbo	0.611	0.206	0.105	0.507	−0.371	0.091	−0.567	0.267
Credit_spread	−67.460*	0.035	−9.149**	0.000	21.075**	0.000	−56.973	0.173
Yield_curve	7.482	0.540	−6.422**	0.000	37.143**	0.001	−226.200**	0.000
Sigma_int	10.127	0.759	11.161**	0.000	−10.434*	0.020	−4.604	0.599
Cumprepayprob_24	−14.453**	0.001	−0.954	0.470	0.271	0.836	−11.539	0.168
Ccumdefaultprob_24	1.599*	0.657	−1.226	0.252	1.867	0.067	−4.783**	0.002
Low documentation	0.850*	0.012	0.223*	0.018	−0.135	0.406	0.194	0.581
	Observations	123	Observations	1,507	Observations	2,644	Observations	461

Panel B reports the coefficients of a multinomial logit model which estimates the probability that a subprime loan will be privately securitized. Subprime-like loans are defined as mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000 and reported as subprime by the LPS database (FICO < 620 and having Grade "B" or Grade "C").

to allow both the coefficients on the covariates and the baseline control group to vary by year and across market segments, whereas Elul (2009) pools all the loans originated in different years together in his analysis and uses loans originated in 2003 as the control group. Testing for the effect of securitization by comparing the performance of loans securitized versus those kept in portfolio calls for close alignment between the two comparison groups in

Table 5
Lender size and securitization choice

This table reports the coefficients of a multinomial logit model, which estimates the probability that a prime loan will be bought by the GSEs or privately securitized. This regression was estimated using the holdout sample that was created from the same origination years as the estimation samples. The independent variables are the yield spread at origination (yield_spread), whether or not the loan is above GSE loan limits (jumbo), the difference between the AAA bond index and the Baa bond index at the time of origination (credit_spread), the ratio between the 10-year Treasury rate and the 1-year Treasury rate (yield_curve), the interest volatility over the 24 months before origination (sigma_int), the cumulative 24-month prepayment and default probabilities (cumprepayprob_24 and cumdefaultprob_24) as well as the dummy variables to indicate whether the loans had a high yield spread (high_spd) or a low yield spread (low_spd). Lender fixed effects are also included but their coefficients are not reported here. High quality/prime-like loans are defined as conventional, fixed-rate mortgages for single-family residences and condos originated between January 2004 and December 2007 below $650,000. Additionally prime-like loans have FICO scores above 620 and loan-to-value ratios below 95%. **Significant at 1% level, *significant at 5% level.

Panel A: Probability of securitization for high-quality/"prime-like" loans under rational expectations—small lenders

	2005		2005		2006		2007	
GSE outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	5.175*	0.012	9.456*	0.031	−1.067	0.924	10.292	0.208
Yield_spread	0.691	0.120	0.220	0.605	−0.745*	0.035	−0.704**	0.004
Jumbo	−1.565**	0.000	−1.011**	0.000	−1.360**	0.000	−1.694**	0.000
Credit_spread	0.340	0.897	−3.950	0.157	2.685	0.586	0.607	0.904
Yield_curve	−1.378*	0.020	−2.531	0.237	−0.756	0.936	−8.613	0.098
Sigma_int	−4.380	0.102	−3.526	0.233	−2.058	0.567	−1.800	0.419
Cumprepayprob_24	1.862	0.589	13.703**	0.003	23.790**	0.000	5.305	0.200
Cumdefaultprob_24	−20.771	0.091	−21.605*	0.041	14.605**	0.007	12.710**	0.003
High_spd	−0.327	0.061	−0.081	0.602	−0.023	0.892	0.250	0.141
Low_spd	−0.072	0.701	0.210	0.173	−0.150	0.339	−0.020	0.881
Private label outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	8.247**	0.000	13.379**	0.003	−3.808	0.751	15.888	0.088
Yield_spread	2.312**	0.000	1.309**	0.003	−0.765*	0.037	−0.729*	0.022
Jumbo	0.872**	0.000	0.880**	0.000	1.236**	0.000	1.252**	0.000
Credit_spread	−3.784	0.172	−6.507*	0.024	5.607	0.287	−1.180	0.833
Yield_curve	−1.422*	0.023	−3.019	0.174	5.482	0.587	−10.577	0.079
Sigma_int	−4.388	0.121	−6.901*	0.024	−1.500	0.698	−1.186	0.640
Cumprepayprob_24	−11.036**	0.002	5.687	0.240	−10.200	0.129	−7.697	0.110
Cumdefaultprob_24	2.581	0.841	−8.415	0.444	15.468**	0.006	5.977	0.181
High_spd	−0.378*	0.038	−0.072	0.659	0.482**	0.007	0.771**	0.000
Low_spd	0.011	0.957	0.160	0.327	−0.337	0.051	−0.561**	0.001
	Observations	4,752	Observations	6,989	Observations	6,441	Observations	5,903

Panel A reports the coefficients of a multinomial logit model which estimates the probability that a prime loan originated by a small lender (defined as those originating 20 or fewer loans in a given year in our sample) will be bought by the GSEs or privately securitized.

Panel B: Probability of securitization for high-quality/"prime-like" loans under rational expectations—large lenders

	2004		2005		2006		2007	
GSE outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	2.751**	0.000	7.413**	0.000	16.475**	0.000	1.413	0.221
Yield_spread	−0.290**	0.001	0.868**	0.000	1.394**	0.000	0.994**	0.000
Jumbo	−2.001**	0.000	−1.072**	0.000	−1.700**	0.000	−2.435**	0.000
Credit_spread	−0.925*	0.038	−2.360**	0.000	−7.936**	0.000	−1.224	0.058
Yield_curve	0.382**	0.000	−1.235**	0.000	−13.466**	0.000	4.750**	0.000
Sigma_int	0.609	0.153	−1.858**	0.000	12.640**	0.000	−3.623**	0.000
Cumprepayprob_24	13.530**	0.000	14.965**	0.000	32.530**	0.000	6.294**	0.000
Cumdefaultprob_24	−59.838**	0.000	−46.127**	0.000	−4.940**	0.000	−0.975**	0.007
High_spd	0.169**	0.000	−0.093**	0.000	−0.270**	0.000	−0.093**	0.000
Low_spd	−0.195**	0.000	−0.097**	0.000	−0.087**	0.000	−0.091**	0.000
Private label outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Constant	5.281**	0.000	10.122**	0.000	3.394*	0.043	21.918**	0.000
Yield_spread	1.324**	0.000	1.933**	0.000	1.884**	0.000	1.001**	0.000
Jumbo	0.309**	0.000	0.743**	0.000	0.971**	0.000	0.389**	0.000
Credit_spread	−5.414**	0.000	−4.703**	0.000	−4.358**	0.000	−5.527**	0.000
Yield_curve	0.126	0.200	−1.484**	0.000	1.578	0.253	−12.703**	0.000
Sigma_int	−1.684**	0.000	−3.855**	0.000	9.853**	0.000	−5.819**	0.000
Cumprepayprob_24	11.339**	0.000	3.595**	0.000	3.883**	0.000	1.590**	0.007

Table 5 (*continued*)

Panel B: Probability of securitization for high-quality/"prime-like" loans under rational expectations—large lenders

GSE outcome	2004 Coefficient	2004 $P>\|z\|$	2005 Coefficient	2005 $P>\|z\|$	2006 Coefficient	2006 $P>\|z\|$	2007 Coefficient	2007 $P>\|z\|$
Cumdefaultprob_24	−77.474**	0.000	−34.518**	0.000	−5.485**	0.000	−2.473**	0.000
High_spd	0.122**	0.001	−0.106**	0.000	0.031	0.229	0.214**	0.000
Low_spd	−0.061	0.080	−0.140**	0.000	−0.079**	0.002	−0.251**	0.000
	Observations	84,955	Observations	124,847	Observations	115,739	Observations	71,206

Panel B reports the coefficients of a multinomial logit model, which estimates the probability that a prime loan originated by a large lender (defined as those originating more than 20 loans in a given year in our sample) will be bought by the GSEs or privately securitized.

every relevant aspect other than securitization. Accounting for vastly different incentives for participants selling loans to GSE or non-GSE markets seems first-order. In addition, mismatched groups with inadequate controls will likely lead to poor identification of the securitization indication variable, with effects driven by differences inherent to the underlying populations rather than securitization.

5. Robustness and extensions

As discussed before, the conclusions regarding each market segment need to be drawn carefully, since the results may be sensitive to methodology or data choices. In this section we perform two broad tests. First, we are worried that contamination in our classification of the "prime-like" loan sample, as discussed earlier, may be affecting our results. To alleviate this concern, we will sort our main sample into loans originated by large and small lenders. We will then compare the results for these groups of lenders with those predicted by the economic arguments we laid earlier to gauge the robustness of our tests. Second, we are aware that our structural approach imposes a strong assumption on the nature of the securitization process, namely, that the decision for securitization follows a sequence as discussed before. Therefore, we use the reduced-form approach which relaxes the assumption regarding the securitization process. Moreover, we will also conduct this analysis in several samples which will also allow better classification and accounting of differences in incentives of participants across the GSE and non-GSE markets.

5.1. Large vs. small lenders

To examine the adverse selection behavior of different types of lenders further, we segment the data by lender size, proxied by their origination volume in a given year. Those originating 20 or fewer loans are grouped as small lenders, and the analysis is repeated on each group's securitization choice. Lender fixed effects are included in the large lender group. We do not include lender fixed effects in the small lender group because each lender contributes 20 loans or fewer to the sample and lender-specific variations are not considered significant enough to bias the estimation results. The results under rational expectations are shown in panels A and B of Table 5.[19]

In Panel A of Table 5, the coefficients on the default probability variable are positive and significant for securitization with GSEs in years 2006 and 2007 and for securitization with private labels in 2006. This is contrary to the results for the large lender group, presented in Panel B of Table 5, which are similar to the results obtained from the overall sample reported in Table 4, Panel A. In 2006, loans sold by small lenders to GSEs have both higher default risk and higher prepayment risk than the loans held in portfolio. The loans sold to private labels have higher default risk but not significantly different prepayment risk from loans held in portfolio. In 2007, prepayment probability does not play a significant role in small lenders' securitization decisions. We attribute the difference to two factors. One is that smaller lenders are more likely to be small-town lenders who have private information about their borrowers to adversely select loans for securitization. The other is that small lenders are less likely to retain higher-default-risk loans even if the loans have lower prepayment probability because of their less diversified borrower base and their smaller cost of gaining a damaged reputation in the secondary market.

In 2006, the loans small lenders sold into the secondary market had both higher default risk and higher prepayment risk. This is in contrast with loans sold by large lenders, which had higher prepayment risk but lower default risk in each year. We interpret this difference between the loans originated by small and large lenders as evidence of asymmetric information in the market. While large banks handle more loan originations and often evaluate them at their regional underwriting centers, small banks are more likely to be small-town or neighborhood banks. Hence, as we mentioned before, small banks are more likely to know the borrower personally and to have private (soft) information about the borrower. The asymmetric information that small lenders enjoy enables them to identify the "lemons," that is, loans that are riskier on both default and prepayment fronts, and sell them into the secondary market.

[19] The results under adaptive expectations are qualitatively similar.

Table 6
GSE, jumbo, HUD classified subprime and portfolio loans.
This table states the results from a competing risks model of the outcome to prepay, default, or remain current on a given mortgage as estimated by a multinomial logit model. The dependent variable is whether a loan experienced default, prepayment, or remained current within 24 months of origination. The independent variables are information available to lenders at the time of underwriting and include the borrower's FICO score (FICO), the borrower's income (Income), the loan-to-value ratio for the mortgage (LTV ratio), whether the loan is securitized, and whether the loan application was low- or no-documentation (Low documentation). **Significant at 1% level, *significant at 5% level.

Panel A: Probability of default and prepayment for GSE securitized loans versus *GSE-like portfolio loans*

	2004		2005		2006		2007	
Default outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Securitized	−0.195*	0.050	0.09	0.343	−0.226**	0.000	−0.102**	0.000
Prepay outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Securitized	0.004	0.882	−0.017	0.522	0.038**	0.005	0.400**	0.000
	Observations	9,460,382	Observations	8,129,794	Observations	13,794,348	Observations	12,234,342

Panel B: Probability of default and prepayment for jumbo (non-GSE) securitized loans versus *portfolio loans*

	2004		2005		2006		2007	
Default outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Securitized	−0.148	0.649	−0.265	0.249	−0.006	0.958	−0.188**	0.002
Prepay outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Securitized	−0.249**	0.000	−0.275**	0.000	−0.119*	0.030	0.155**	0.000
	Observations	667,666	Observations	556,898	Observations	465,019	Observations	408,852

Panel C: Probability of default and prepayment for subprime loans (defined as per the HUD lender list)

	2004		2005		2006		2007	
Default outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Securitized	−0.421	0.156	−0.008	0.934	0.366	0.467	−0.39	0.237
Prepay outcome	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$	Coefficient	$P>\|z\|$
Securitized	−0.187	0.104	−0.117	0.063	−0.247	0.124	0.242	0.266
	Observations	184,334	Observations	340,252	Observations	219,582	Observations	103,899

5.2. *Reduced form analysis of comparing GSE, jumbo, subprime, and portfolio loans*

As a robustness exercise to our structural analysis, we conduct a reduced-form exercise where we study the behavior of GSE loans relative to bank-held loans and also separately compare private label securitized loans to the portfolio loans. Based on the discussion in Section 2, it is clearly important to separately study the role of adverse selection in the GSE and private label securitization markets due to the different institutional structures of these two markets. However, in our LPS data set, it is not possible to identify which segment of the market the loans on the balance-sheet were intended for. For instance, there may be some loans on a bank's portfolio that might be intended for sale to GSEs but remain on the lender's books for some reason. These bank-held loans might be loans that are ex ante of different quality than loans that were originated to be privately securitized loans. Though our "prime-like"/high-quality sample tried to circumvent this problem, we had noted that several of the high-quality loans as per our classification were sold to private label as well (suggesting that our method of classification was not perfect). We now provide reduced-form tests to conduct robustness tests that address this issue.

We start by estimating the reduced-form regressions on a sample of GSE securitized and bank-held loans that are directly comparable. To help us to make comparisons, we classify portfolio loans as GSE-like or non-GSE-like, following the propensity score matching procedure of Agarwal et al. (2011) and Keys et al. (forthcoming). In particular, we run a probit regression on a sample of all securitized loans (private label and GSE), in which the dependent variable is whether a loan is a GSE loan. The explanatory variables are FICO and LTV at origination, as well as indicators for year of origination, for whether a mortgage has adjustable interest rates, for non-owner occupancy, and for not fully documented loans (low or no documentation). Then, we predict the GSE dummy for each portfolio loan. We classify loans with a propensity score of 0.5 or more as GSE-like and the rest as non-GSE-like. We find that a vast majority of the portfolio loans are GSE-like.[20] This is likely because our sample is high-quality (and only Fixed Rate Mortgages) and most of these loans were intended for sale to GSEs. We take these

[20] Note that this also reiterates that in our structural analysis, when we took the high-quality/"prime-like" sample, while there was some contamination in that some loans going to private market also got included, this contamination was not that large.

bank-held loans and estimate the competing risk hazard regression model for this sample.

The results are presented in Table 6, Panel A which shows the probability of default and prepayment for GSE securitized loans as compared to GSE-like portfolio loans. We find that across the years, securitization is associated with a lower probability of default (except in 2007) and higher probability of prepayment. This is consistent with our results in the structural analysis and also consistent with the arguments outlined in Section 2.

Next, we do the reduced-form analysis for loans that are primarily intended for subprime loans. This will allow us to make reduced form comparisons for non-GSE loans similar to the comparison in Panel A of Table 6 for GSE loans. To do so, we will study the jumbo loan and the HUD-classified loans, respectively. Panels B and C of Table 6 repeat the analysis for two sets of loans where securitization is done by private issuers (non-GSEs). In Panel B of Table 6 we show the probability of default and prepayment for jumbo loans that are securitized through non-GSEs and compare them to portfolio loans. We find that across the years, there is some evidence that securitization caused a higher probability of prepayment for jumbo loans though the evidence on defaults is not conclusive (except for 2007).

In Panel C of Table 6, we define loans privately securitized in an alternative fashion. As discussed earlier, the definition of subprime itself is not standard and uniform across different data sets (or sometimes even within a single data set). We take advantage of the LPS–HMDA merged data set and use the HUD lender list as the means of identifying subprime loans, and estimate a model of default and prepayment on this group of loans. As can be observed, we find no clear pattern of default or prepayment for subprime loans. This is also consistent with our results from the structural analysis.

Overall, in the reduced-form analysis for subprime loans, there is no clear pattern that emerges. In contrast, in the prime market (loans intended for sale to GSEs), banks generally sold low-default-risk loans into the secondary market while retaining higher-default-risk loans in their portfolios. In addition, we also find support for adverse selection with respect to prepayment risk in the prime market. These results are consistent with our results from the structural analysis and give us comfort on the robustness of our inferences.

6. Conclusion

Are loans sold into the secondary mortgage market of different quality than loans that lenders retain on their balance-sheet? Our analysis of a large data set of mortgage loans originated between 2004 and 2007 reveals strong evidence that the answer differs dramatically depending on the segment of the mortgage market. Banks sold low-default-risk loans into the secondary market and retained higher-default-risk loans in their balance-sheet for loans intended to be sold to GSEs. In addition, we find support for adverse selection with respect to prepayment risk: securitized loans sold to GSEs entail a higher prepayment risk than loans on lenders' balance-sheets. In contrast, we do not find any conclusive pattern for loans sold in the private subprime market.

We suggest several reasons for these differences. Origination and post-origination practices in the prime and subprime market differ significantly due to GSEs imposing control on default risk of loans originated by lenders since they offer guarantees only against default risk to investors. This control is missing on the prepayment margin—giving lenders more freedom to adversely select on prepayment risk—since this risk is passed to the investors by GSEs. In contrast, there is no private issuer who coordinates the securitization chain in the subprime market. These differences are also accentuated due to capital requirement arbitrage for prime lenders.

Our results suggest that in return for selling loans with lower default risk, lenders retained loans with lower prepayment risk in the prime market. This would have been a profitable strategy in the early years of our sample period—when prepayment risk driven by high refinancing activity was a bigger concern for lenders than default risk in the prime market. Interestingly, as the bursting of the bubble approached and default concerns in the market started growing, these same lenders became less willing to retain higher default risk in return for lower prepayment risk.

Our findings in the subprime market also suggest that the results in this market may be sensitive to various assumptions and definition differences. As a result, we clarify and caution researchers against drawing conclusions in this sector without being fully aware of alternative data sets and the underlying assumptions (e.g., we warn against using LPS data coverage for low-documentation subprime loans, especially pre-2005).

It should be noted that observing that some securitized loans are higher quality than bank-held loans does not necessarily mean securitization did not play a role in the rising default probabilities that triggered the recent financial crisis. Rather, securitization could have led to a greater supply of funds for mortgage lending, which in turn might have contributed to house price growth (Mian and Sufi, 2009) and the deterioration in underwriting standards (Greenspan, 2010).

Appendix A. The securitization process

See Figs. A1 and A2.

Appendix B. Data terms

LPS: Lender Processing Services Applied Analytics, Inc. data set (formerly McDash): provides loan level data set including both prime and subprime loans.

LP: CoreLogic LoanPerformance data set. Provides a variety of data products. TrueStandings Servicing offers aggregate-level prime and subprime servicer-reported characteristics and performance data. TrueStandings Securities offers loan level non-prime asset backed securities.

HMDA: Passed in 1975, the Home Mortgage Disclosure Act (HMDA) requests lenders of certain size to report loan application and origination information. An annual collection of this information is available for public use. This

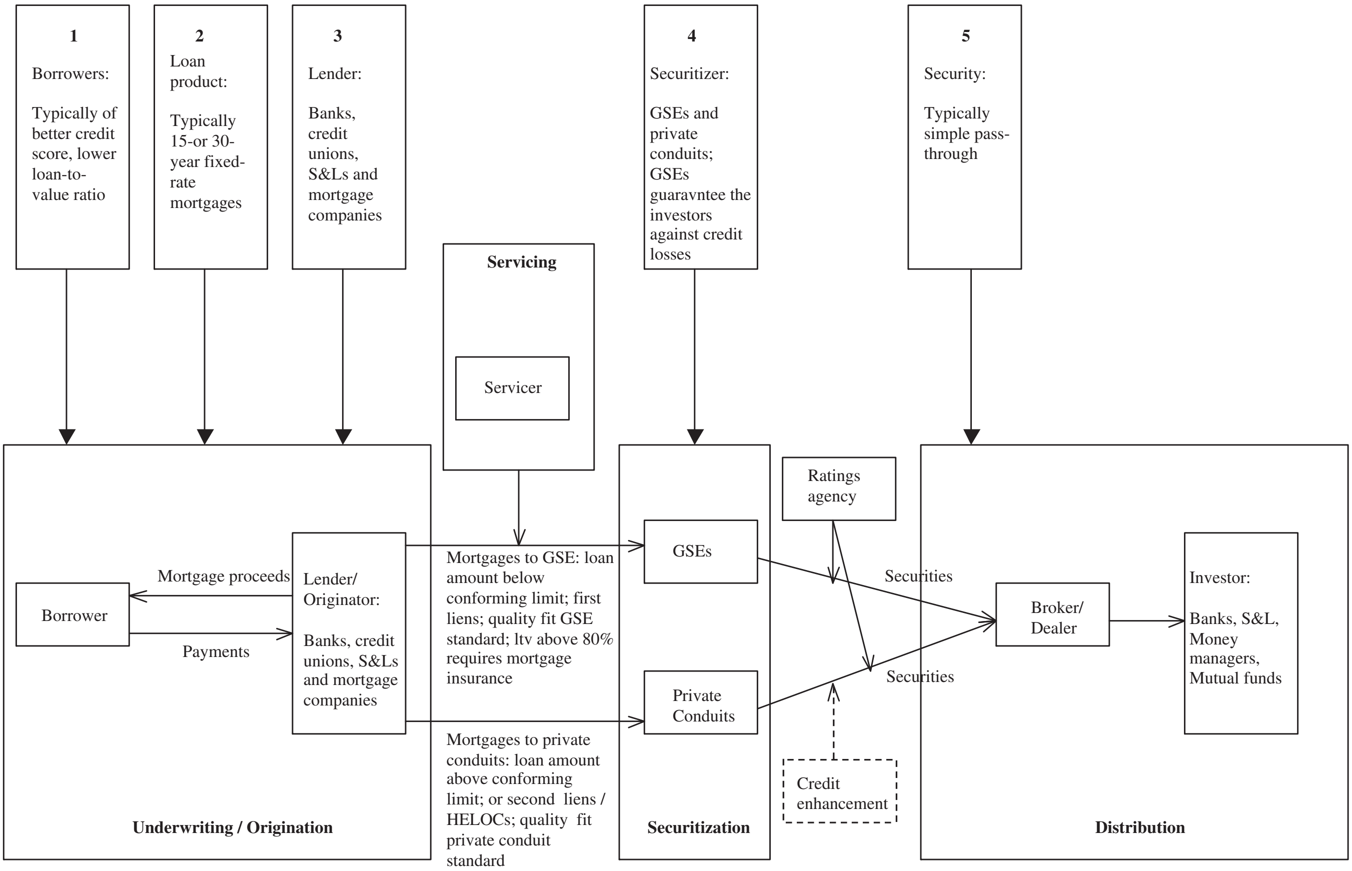

Fig. A1. Securitization process: The prime market.

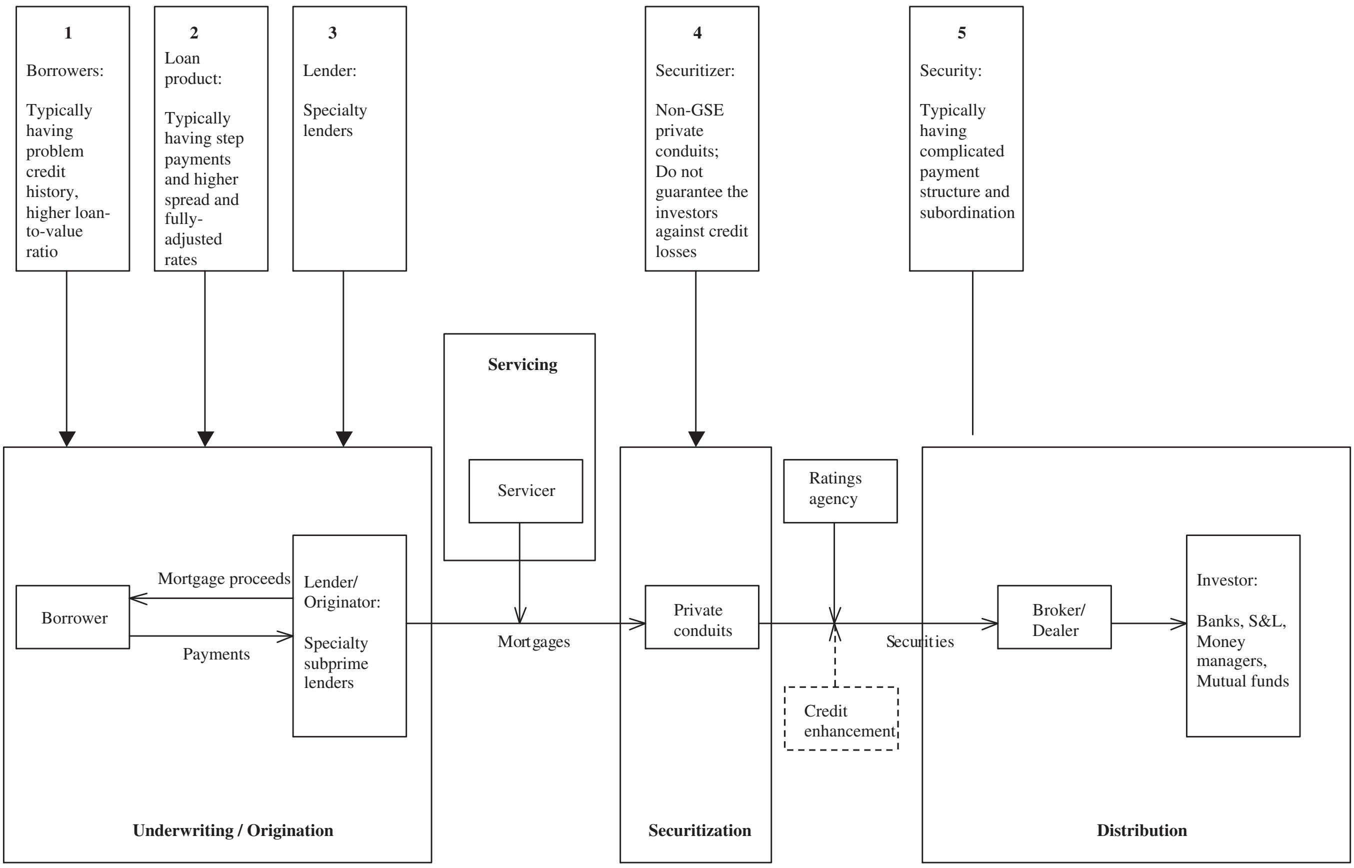

Fig. A2. Securitization process: The subprime market.

covers both prime and subprime lenders, but does not provide loan performance information. However, algorithms can be developed to merge with other data sets (such as LPS) to get the performance information.

Sample 'prime-like' loans: conventional loans with FICO above 620 and loan-to-value below 95%.

Sample 'subprime-like' loans: Reported as subprime by the LPS database, with FICO < 620 and credit grade of "B" or "C". Credit grade depends on such factors as the credit score of the borrower, loan-to-value ratio and the debt ratio (ratio of the total monthly debt to the monthly gross income of the borrower). If a loan has high ranking on all these factors, the loan is assigned Grade "A" and qualifies for a lower interest rate. If the loan has lower ranking on one or more of these factors, the quality of the loan is downgraded to Grade "B", "C", or "D" and the loan is charged a higher interest rate.

References

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., Evanoff, D.D., 2011. Role of securitization in mortgage renegotiation. Journal of Financial Economics 102, 554–578.

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S., Evanoff, D.D., 2010. Market-based loss mitigation practices for troubled mortgages following the financial crisis. Unpublished working paper. Federal Reserve Bank of Chicago, Ohio State University, and U.S. Office of the Comptroller of the Currency.

Agarwal, S., Hauswald, R., 2010. Distance and private information in lending. Review of Financial Studies 23, 2757–2788.

Akerlof, G.A., 1970. The market for "lemons": quality uncertainty and the market mechanism. Quarterly Journal of Economics 84, 488–500.

Ambrose, B., LaCour-Little, M., Sanders, A., 2005. Does regulatory capital arbitrage, reputation, or asymmetric information drive securitization? Journal of Financial Services Research 28, 113–133.

An, X., Deng, D., Gabriel, S.A., 2011. Asymmetric information, adverse selection, and the pricing of CMBS. Journal of Financial Economics 100, 304–325.

Courchane, M., Surette, B.J., Zorn, P., 2004. Subprime borrowers: mortgage transitions and outcomes. Journal of Real Estate Finance and Economics 29, 365–392.

Cutts, A.C., Van Order, R., 2005. On the economics of subprime lending. Journal of Real Estate Finance and Economics 30, 167–196.

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

Elul, R., 2009. Securitization and mortgage default: reputation vs. adverse selection. Unpublished working paper. Federal Reserve Bank of Philadelphia.

Greenspan, A., 2010. The crisis. Brookings Papers on Economic Activity Spring, 201–246.

Keys, B., Mukherjee, T., Seru, A., Vig, V., 2009. Financial regulation and securitization: evidence from subprime loans. Journal of Monetary Economics 56, 700–720.

Keys, B., Mukherjee, T., Seru, A., Vig, V., 2010. Did securitization lead to lax screening? Evidence from subprime loans. Quarterly Journal of Economic 125, 307–362.

Keys, B., Seru, A., Vig, V., 620 FICO, take II: securitization and screening in the subprime mortgage market. Review of Financial Studies, forthcoming.

LaCour-Little, M.,Yang, J., 2010. Taking the lie out of liar loans: the effect of reduced documentation on the performance and pricing of Alt-A and subprime mortgages. Unpublished working paper. California State University at Fullerton.

Mayer, C., Pence, K., 2009. Subprime mortgages: what, where, and to whom? In: Glaeser, E., Quigley, J. (Eds.), Housing Markets and the Economy: Risk, Regulation, and Policy: Essays in Honor of Karl E. Case, Lincoln Land Institute, Cambridge, MA, pp. 149–198.

Mian, A., Sufi, A., 2009. The consequences of mortgage credit expansion: evidence from the U.S. mortgage default crisis. Quarterly Journal of Economics 124, 1449–1496.

Pennington-Cross, A., 2003. Credit history and the performance of prime and nonprime mortgages. The Journal of Real Estate Finance and Economics 27, 279–302.

Petersen, M., 2004. Information—soft and hard. Unpublished working paper. Northwestern University.

Piskorski, T., Seru, A., Vig, V., 2010. Securitization and distressed loan renegotiation: evidence from the subprime mortgage crisis. Journal of Financial Economics 97, 369–397.

Rajan, U., Seru, A., Vig, V., 2011. The failure of models that predict failure: distance, incentives and defaults. Unpublished working paper. University of Chicago.

Sengupta, R., Emmons, W.R., 2007. What is subprime lending? Economic Synopses, St. Louis FED, Number 13.