THE EFFECT OF LARGE INVESTORS ON ASSET QUALITY: EVIDENCE FROM SUBPRIME MORTGAGE SECURITIES

Manuel Adelino, W. Scott Frame, Kristopher S. Gerardi*

October 2014

Abstract: This paper examines how Fannie Mae and Freddie Mac (the GSEs), the largest investors in subprime private-label mortgage-backed securities (PLS), affected the composition of security collateral and the distribution of risk within the PLS market. To identify the causal effect of the GSEs, we use the fact that PLS deals in which Fannie Mae and Freddie Mac purchased securities included separate mortgage pools: one specifically created for the GSEs and one or more others directed at other triple-A investors. Studying within-deal variation, we find that the mortgage pools bought by Fannie Mae and Freddie Mac had similar ex-ante risk characteristics, but performed much better ex-post relative to other pools in the same deals. These effects were concentrated in low documentation loans, and for issuers that were highly dependent on Fannie Mae and Freddie Mac that also offer higher yielding securities to the GSEs. Our results extend the importance of large, concentrated claimholders beyond information-sensitive securities, such as equities and bank debt, to information-insensitive arm's-length debt.

* Adelino: Fuqua School of Business, Duke University, manuel.adelino@duke.edu. Frame: Federal Reserve Bank of Atlanta, scott.frame@al.frb.org. Gerardi: Federal Reserve Bank of Atlanta, kristopher.gerardi@atl.frb.org. For helpful comments and discussions we would like to thank Brent Ambrose, Andra Ghent, Robert Hansen, Peter Niculescu, Shane Sherlund, Larry Wall, and Paul Willen, as well as seminar participants at the Kellogg Junior Finance Conference and the University of Juan Carlos III. We thank Neil Desai for outstanding research assistance. Frame also thanks the University of North Carolina at Charlotte and the Federal Reserve Bank of Richmond for financial support. This paper was previously circulated under the title "The Role of Blockholders in Debt Markets: Fannie Mae, Freddie Mac and Subprime MBS."

1. Introduction

Informational frictions and incentive conflicts between issuers of financial claims and the owners of those claims are central topics in financial economics. A large literature has shown that the presence of large, concentrated claimholders can reduce such problems and improve security performance. For example, equity block-holders are able to more efficiently acquire information and monitor issuing managers, and thereby reduce free-rider problems associated with diffuse ownership. Incentive conflicts are reduced as these large investors can directly intervene in firm management or exert influence through the threat of exit.[1] In terms of concentrated debt holders, banks are understood to make information-sensitive loans based on long-term relationships with their borrowers. Here, banks incur screening and monitoring costs to acquire private information about borrowers, which reduces adverse selection and moral hazard problems in a repeated game setting.[2]

However, the effect of concentrated claimholders in information-insensitive arms-length debt markets has largely been ignored by the literature, as investors in those markets are generally thought of as more passive providers of capital. This paper seeks to fill this gap in the literature by studying the role played by two government-sponsored enterprises (GSEs), Fannie Mae and Freddie Mac, in the subprime private-label mortgage-backed securities (PLS) market during the recent housing boom.[3]

The participation of the GSEs in the PLS market provides a unique setting for studying whether investor concentration matters in arms-length debt markets for a few reasons. First, Fannie Mae and Freddie Mac together purchased roughly 30 percent of the total dollar volume of subprime PLS issued during the recent

[1] See Shleifer and Vishny (1986), Burkart, Gromb, and Panunzi (1997), Brav, Jiang, Partnoy and Thomas (2008), and Morse (2013) for evidence of blockholder intervention in the management of firms; and Edmans (2009), McCahery, Sautner, and Starks (2008) and Parrino, Sias, and Starks (2003) on governance through exit.

[2] See Bhattacharaya and Thakor (1993) and Boot (2000) for surveys on the extensive literature on bank screening and monitoring.

[3] In addition to investing in PLS and other residential mortgage-related assets, Fannie Mae and Freddie Mac play a central role in the U.S. mortgage market through their "credit guarantee" businesses, whereby mortgage originators exchange pools of loans for securities that represent an interest in the same pool; and the GSEs agree to ensure timely payment of principal and interest on the securities in exchange for a monthly insurance premium ("guarantee fee"). This paper focuses solely on the GSEs' participation in the PLS market.

U.S. housing boom (2003-2007), making them by far the largest single investors in this asset class.[4] Second, we are able to identify the specific securities purchased by the GSEs by exploiting the fact that, by law, Fannie Mae and Freddie Mac are only allowed to acquire mortgages below the conforming loan limit, a fixed dollar amount set annually by the government. This prompted the creation of subprime PLS securities specifically designed for the GSEs that were backed only by conforming mortgages, as well as other securities in the same deals (issued at the same time, by the same issuer, and with the same credit rating) that could not be purchased by the GSEs and were backed by both conforming and non-conforming mortgages. Third, this institutional arrangement allows us to address selection problems that have long been recognized in this literature related to the endogenous matching of large claimholders with securities issuers.[5] The identification strategy in this paper compares the performance of the mortgages backing securities bought by the GSEs with those issued as part of the same deal aimed at other investors.[6] Finally, because the focus of this study is the mortgage-backed securities market, we can control for detailed collateral characteristics that are generally not available at the same level of detail for other debt markets (e.g., corporate, sovereign). This allows us to identify the potential role of issuer private information in the allocation of risk between mortgage pools sold to the GSEs and pools sold to other investors.

Using within-deal variation, and controlling for all relevant observable underwriting characteristics, we find that loans in GSE pools (i.e. those backing securities purchased by the GSEs) performed significantly better than loans backing other securities. The within-deal setup rules out alternative explanations for the results, including effects due to macro conditions at the time of issuance and unobserved heterogeneity at the level of the issuer, originator, and servicer, since these entities were all shared by loans in the same deal. In addition, the split between GSE and non-GSE loan pools applied exclusively to the triple-A securities in these

[4] Ghent, Hernandez-Murillo and Owyang (2013) analyze a random sample of 100 prospectus supplements and find a similar market share for the GSEs.
[5] Previous studies have used other identification strategies to address this omitted variables problem, most recently the quasi-random inclusion of firms into equities indices (see, for example, Aghion, Van Reenen and Zingales, 2013, Chang, Hong, and Liskovich, 2013, Crane, Michenaud and Weston, 2013, and Mullins, 2013).
[6] Issuers are sometimes referred to as sponsors. The primary responsibility of an issuer is to create the mortgage pool(s) by acquiring loans from originators and then issuing securities backed by the pool cash flows. As discussed below, the issuer and originator are sometimes the same institution or affiliated institutions, but in many cases they are unaffiliated entities.

deals, while lower-rated tranches received cash flows from both pool types and provided credit support for all senior securities (GSE and non-GSE triple-A alike). This removes any concern that differences in performance could be driven by differential levels of credit support or risk retention. We also perform a series of robustness checks to ensure that the results are not driven by some other mechanism unrelated to the issue of concentrated investors. These tests include removing non-conforming jumbo mortgages from the analysis, considering differences in the prepayment risk across mortgage products, and controlling for whether the property was located in a low-and-moderate income area and hence eligible for GSE affordable housing goal credit.

Two potential explanations are consistent with the superior realized performance of mortgages in GSE pools. First, Fannie Mae and Freddie Mac may have had better risk assessment models that led them to demand an unobservably different composition of loans than other investors in the same deals (a demand-side hypothesis). This seems plausible given the GSEs' centrality to the U.S. housing finance system and the breadth and depth of data they collected over time. However, the implicit government guarantee that the GSEs benefited from and the potential moral hazard problem induced by this guarantee would cut against this demand side hypothesis as it would create incentives for the GSEs to seek higher returns by buying riskier loan pools.[7] A second (supply-side) explanation is that issuers may have used private information to choose safer mortgages for GSE pools because of reputational concerns and the importance of the GSEs as major investors in the deals, or because the GSEs might be more likely than other investors to trigger contractual clauses with respect to the quality of the mortgages included in a deal.

Consistent with the hypothesis that issuers used private information, we show that the difference in ex-post performance between loans in GSE and non-GSE pools primarily comes from low documentation loans, where soft information has been found to be especially important (Jiang Nelson Vytlacil, 2014a, 2014b; Keys, Seru, and Vig, 2012; Begley and Purnanandam 2013; Saengchote, 2013). The difference in default rates

---

See Archarya et al. (2010) for a detailed discussion of this issue. As the authors point out, Fannie Mae's own strategic plan document written in 2007 explicitly called for a business model that involved taking on more credit risk by purchasing riskier mortgage products.

[8] The CLL for most single-family homes between 2006 and 2013 was $417,000. Single-family homes in certain high cost areas have higher limits. We refer the reader to the appendix for a list of single-family conforming loan limits during our sample period.

between GSE and non-GSE pools for low documentation mortgages is 3.9 percentage points, whereas the difference for full documentation loans is only 0.7 percentage points. This suggests that issuer sorting of mortgages across pools within a deal was the source of the higher quality mortgages in the GSE pools (i.e., a supply-side effect), as this type of information would have been difficult to transmit and incorporate into the GSEs' internal risk models.

In order to further pin down the supply-side mechanism driving the differences in mortgage performance, we construct a measure of the frequency of interactions between issuers and the GSEs over time. To do this, we use information on the identity of subprime PLS issuers and compute the fraction of previous deals for each issuer that included GSE pools and in which Fannie Mae or Freddie Mac purchased securities. The rationale for this measure is that issuers that often include securities designed for Fannie Mae and Freddie Mac in their deals are more likely to want to maintain good standing with the two GSEs, as they are more likely to be dependent on them for future business. In addition, this may be an indication that the GSEs have a preference for interacting with these issuers. We find that deals arranged by issuers that frequently include GSE pools in their deals show significantly larger differences between GSE and non-GSE pool performance. The (within-deal) difference in overall ex-post loan performance across GSE and non-GSE pools is fully explained by the difference in low documentation loans and by the fraction of previous deals with GSE participation.

We further build on these tests by using the identity of the issuers and separating deals into those where the issuer of the securities and the originator of the underlying mortgages were affiliated (i.e., either the same institution or part of the same vertically integrated entity), and those where they were not. The rationale for this test is that issuers that are affiliated with the mortgage originators are more likely to have private information about loan quality. Consistent with the previous tests, we find that the superior performance of low documentation loans in GSE pools and of loans securitized by issuers with a high proportion of deals with GSE participation is especially strong in deals where the issuer and the originator of the mortgages were affiliated institutions. This provides additional evidence in support of a supply-side interpretation by which issuers included unobservably higher quality loans in GSE pools in order to maintain a good reputation and their relationship.

We further supplement the analysis of loan performance by comparing the yield spreads at origination of GSE and non-GSE securities conditional on pool characteristics at issuance. We find that the yields of GSE pools are significantly higher than those of non-GSE pools, and that this difference is also fully explained by the frequency of previous interactions between issuers and the GSEs. This reinforces the hypothesis that issuers that were more dependent on the GSEs provided them with better deals, both through higher quality underlying mortgages and through higher yields at origination.

Finally, we analyze whether credit models that used only historical loan performance and loan/borrower characteristics at origination could have predicted the differences in performance between mortgages in GSE and non-GSE pools. To do this, we construct loan-level *predicted* probabilities of default using historical data, and compare those within deals across GSE and non-GSE pools. We find that GSE and non-GSE pools had very similar *predicted* default rates and that, if anything, the GSE loans look somewhat riskier than other loans in the same deals. This further confirms that the superior performance of the GSE pools did not reflect an observably better composition of loans at the time of origination and that, instead, private information played a role in the allocation of risky mortgages within deals.

The message from this paper is that the presence of concentrated claimholders matters for security performance even in the case of information-insensitive (highly rated) arm's length debt markets. The results suggest a beneficial treatment of large investors, likely owing to reputational concerns on the part of issuers, or due to large investors' higher likelihood of enforcing certain contractual clauses, as they are more likely to internalize both the benefits and costs of ex-post monitoring. It is important to note, however, that the effects we uncover are conditional on the GSEs' participation in this market, so we cannot speak to the separate question of whether overall subprime mortgage security issuance during this time period would have been better or worse had the GSEs not participated as investors in this market.

This paper builds on a large literature that highlights the important role of asymmetric information in mortgage finance. Several studies have analyzed whether securitization gives rise to moral hazard in terms of originator screening incentives (Ashcraft and Schuermann, 2008; Keys, Mukherjee, Seru, and Vig, 2010; Purnanandam, 2011; Keys, Seru and Vig, 2012; Bubb and Kaufman, 2014), as well as the observable risk characteristics of securitized loans and portfolio loans (Ambrose et al, 2005; Krainer and Laderman, 2009; Elul, 2011; Agarwal, Chang and Yavas, 2012; and Jiang, Nelson, and Vytlacil 2014b). Importantly, investors understand the misaligned incentives associated with securitization, and protect themselves from these frictions through contracting features (e.g., by requiring representations and warranties about loan quality), security design features (e.g., requiring higher subordination levels) or required risk premia (Begley and Purnanandam 2013). Issuers of securities, on the other hand, have incentives to signal higher loan quality through risk retention (e.g., DeMarzo, 2005), vertical integration with originators (Demiroglu and James, 2013), and ultimately through reputation-building (Albertazzi, Eramo, Gambacorta, and Salleo, 2014). This paper investigates a related, but unexplored, role of concentrated security claims as another way of mitigating information frictions in this market.

The paper is organized as follows: Section 2 presents our identification strategy based on PLS deal structure and the core empirical specification. Section 3 describes the data sources and sample statistics. Section 4 presents the results and Section 5 concludes.

2. PLS deal structure and identification strategy

One of the key inputs to this paper is the categorization of PLS securities into those that were specifically structured to allow Fannie Mae and Freddie Mac to invest in them and those that were not. To do this, we use the fact that the GSEs were only allowed to purchase triple-A rated securities backed by mortgages below the conforming loan limit (CLL).[8] In the PLS market this restriction was overcome by using more than one collateral pool in the same deal: one pool would include only "conforming" mortgages and one or more other pools included a mix of conforming and non-conforming loans (or "jumbo" loans). PLS issuers then created triple-A securities intended for the GSEs that only received cash-flows from the conforming loan pool, and other triple-A securities that received cash flows from the other pools for all other investors. Remaining cash-

[8] The CLL for most single-family homes between 2006 and 2013 was $417,000. Single-family homes in certain high cost areas have higher limits. We refer the reader to the appendix for a list of single-family conforming loan limits during our sample period.

flows from all pools in a deal accrued to the junior tranches (double-A rated and below).[9] Figure 1 illustrates how a typical subprime PLS deal with GSE participation was structured.

Based on this institutional feature of the subprime PLS market, we design a simple algorithm to identify mortgage pools that backed securities that were eligible to be purchased by either Fannie Mae or Freddie Mac between 2003 and 2007. We classify a pool as being a "GSE pool" if at least 99 percent of the loans in the pool are below the CLL at the time that securities in the deal are issued to investors.[10] Crucially for our identification strategy, the only reason why an issuer would structure a deal in this way would be to attract the GSEs as investors, as no other investors in the market are constrained by the CLL. In Section 3.2, we show that we are able to closely match the aggregate amount of purchases by the GSEs. The Online Appendix provides additional details about the algorithm, and includes a validation exercise that shows that we likely capture most of the relevant securities purchased by the GSEs. For example, of the 478 subprime PLS securities that were included in the series of lawsuits brought by the Federal Housing Finance Agency (FHFA) in 2011 alleging fraudulent marketing and sales materials for GSE-purchased PLS, 476 are captured by our algorithm as being "GSE" securities, while the other triple-A securities in those same deals are classified as "non-GSE".[11] This gives us confidence that our algorithm is not a significant source of type I error. The evidence using the lawsuit securities and the aggregate amount of purchases also indicates that the GSEs purchased most (if not all) of the pools that were created with these characteristics (i.e., those made up exclusively of conforming loans). In particular, if the GSEs were not the sole buyers of these pools, we would likely miss the total amounts purchased by a substantial amount, which is not the case.

---

[9] Each of the two (or more) groups of triple-A securities in the same deal are protected by subordinate claims on the remaining cash-flows from all other pools if interest and principal from their own pool are insufficient to cover the promised cash-flows. Still, each group of triple-A securities is mostly subject to the performance of the mortgages on which they have priority claims, in particular given that shocks to the performance of different pools in the same deal are likely to be highly correlated.

[10] We also add a condition that less than 75 percent of loans in the pool are second liens (although the results are not sensitive to this additional condition). Since the vast majority of second lien mortgages have outstanding balances well below the conforming loan limit, the CLL tells us very little about whether or not the GSEs purchased securities collateralized by those loan pools.

[11] Details on the lawsuits are available on the FHFA website at fhfa.gov/webfiles/22599/PLSLitigation_final_090211.pdf

Using this simple algorithm we find that the GSEs invested heavily in triple-A subprime PLS from 2003 through 2007, accounting for approximately one-third of overall issuance during this period. There are a couple of potential reasons why Fannie Mae and Freddie Mac were interested in investing so heavily in the subprime PLS market. First, the GSEs benefitted from an implicit federal guarantee of their obligations that resulted in a significant funding advantage and little market discipline (e.g., Greenspan, 2005). Hence, Fannie Mae and Freddie Mac faced strong incentives to grow and acquire eligible mortgage-related assets with yields above their cost of funding. Second, many mortgages funded through PLS deals could be counted against the GSEs' affordable housing goals (in proportion to their investment).[12] Analysis by the U.S. Department of Housing and Urban Development illustrates the goal-richness of subprime PLS acquired by Fannie Mae and Freddie Mac in 2004 and 2005 (Bunce, 2007).

The buyers of triple-A PLS securities with claims on the "non-GSE" pools were much more dispersed. According to Greenlaw, Hatzius, Kashyap, and Shin (2008, p.35), there were seven main groups of non-GSE investors in subprime PLS before 2007: Commercial banks, investment banks, insurance companies, hedge funds, finance companies, mutual funds, and pension funds. For various regulatory reasons, U.S. and foreign commercial banks, investment banks, and insurance companies were the most likely investors (on and off-balance sheet) in non-GSE triple-A subprime securities. For the purposes of this paper, the key assumption is that these investors were much more dispersed than the GSEs.

2.1 Empirical specification

Our primary empirical analysis involves comparing the ex-post default rates of mortgages in GSE pools versus those in non-GSE pools in order to determine whether there are systematic differences in loan quality that were unobservable to PLS investors at the time of contracting. We separate performance differences due to factors observable to investors from those that were unobservable by conditioning on a large set of loan and

---

[12] As Robert Levin, the former chief business officer of Fannie Mae, told the Federal Crisis Inquiry Commission, buying private-label mortgage–backed securities "was a moneymaking activity—it was all positive economics. . . . [T]here was no trade-off [between making money and hitting goals], it was a very broad-brushed effort" that could be characterized as "win-win-win: money, goals, and share", (FCIC, 2011).

borrower characteristics used in most mortgage underwriting models that were readily available to most institutional investors. After documenting differences in ex-post default rates between GSE and non-GSE loan pools likely caused by factors unobservable to PLS investors, we conduct a battery of tests designed to determine whether the performance differences are driven by private information held by the securities issuers.

We supplement the analysis of ex-post performance with an analysis of ex-ante credit risk. Specifically, we construct and compare ex-ante expected default probabilities between loans in GSE and non-GSE pools, conditional on observable information, to determine if there were significant differences in expected performance that could be forecast by investors in real-time. Finally, we analyze PLS security yield spreads at issuance for evidence on preferential treatment of the GSEs relative to other investors on the pricing dimension.

Our main identification strategy relies on comparing the ex-post performance of GSE and non-GSE mortgage pools *in the same deal*, after controlling for the observable risk characteristics of the loans. This allows us to exclude all deal-level unobservable characteristics as drivers of our results, such as issuer, originator, and servicer, since these entities were all shared by loans in the same deal. This is a unique setting, in that we can isolate the impact of a specific investor (or class of investors) on the quality of the underlying assets and on the behavior of issuers. In other corporate finance settings (equities, loans or corporate bonds) different investors either hold the same securities issued at the same point in time, or securities differ either in terms of the timing of issuance or in terms of security characteristics, or usually both.

Specifically, we estimate loan-level regressions of the form

$$LHS_{ijz} = \alpha + \beta_1 X_{ijz} + \beta_2 GSE_{iz} + \eta_j + \varepsilon_{ijz} \tag{1}$$

where z identifies each mortgage in pool i within deal j. LHS_{ijz} is a measure of either mortgage delinquency (described in the next section) or of the yield spreads at the time of security issuance. X_{ijz} is a vector of mortgage-level control variables that includes all relevant observable borrower, loan, and geographic characteristics and GSE_{iz} is an indicator variable that is equal to 1 for mortgages in GSE pools and 0 otherwise. The term η_j represents deal-level fixed effects. Note that by including deal-level fixed effects η_j we are accounting for very fine time fixed effects (effectively one fixed effect for each date that we observe an issuance).

The vector of mortgage-level control variables includes the combined loan-to-value ratio, the logarithm of the original loan balance, the original interest rate, the credit score, the original term, the number of months between origination and issuance (seasoning); and indicator variables for first lien loans, low documentation loans, interest-only loans, balloon loans, negative amortization loans, residence status (owner-occupied, investor/vacation home), loan purpose (cash-out refinance, other refinance, purchase), property type (condominium, multi-family, single-family), and the existence of a prepayment penalty. We also include the county-level unemployment rate and the level of the house price index at the time of issuance, as well as the changes in these series from the time of issuance through the end of the default horizon, as well as a full set of state-level fixed effects. Additional indicator variables are included whenever there are missing observations for any of the controls.

3. Data and summary statistics

3.1 Data sources

The data used in this paper come from two main sources. All loan-level data come from CoreLogic's private label securities database, which covers virtually the entire PLS market.[13] This dataset contains information on the underwriting characteristics of the loans underlying the mortgage-backed securities at origination (as discussed above) and the performance of the loans from the month of origination through 2012. Importantly, this dataset includes an identifier for the mortgage pool that each loan belongs to, which allows us to construct pool-level variables using individual loan data (as opposed to datasets that just include the deal to which a loan

[13] According to CoreLogic's website, the dataset contains information on mortgages that make up over 97 percent of outstanding non-agency MBS pool balances (http://www.corelogic.com/solutions/data-resources-for-capital-markets.aspx#rmbs).

belongs, which does not allow for the distinction between GSE and non-GSE mortgage pools that we use in this paper). In addition, the dataset contains security identifiers (CUSIPs) and deal identifiers.

The CoreLogic raw data sample includes 13,189,213 mortgages that backed subprime PLS issued between 2000 and 2007 (we exclude Alt-A and other PLS issues). We drop loans that are not first or second liens (22,395 loans), loans that were seasoned more than 12 months at the time of issuance (813,901 loans), and loans originated in the U.S. territories of Puerto Rico and the Virgin Islands (126 loans).[14] This leaves us with 12,352,791 loans in 3,987 mortgage pools that collateralized 2,161 subprime deals issued between 2000 and 2007. In most of the empirical analysis we restrict attention to loans backing deals issued between 2003 and 2007 so that the majority of securities are not paid off (or close to being paid off) by the third quarter of 2007 when the financial crisis begins.

Additional information on the attributes of residential mortgage-backed securities was hand-collected from Bloomberg. The data fields include all of the security identifiers (including CUSIP and ticker), the issuer name, the date of issuance, the identification of the loan pool that the security has claims on, the spread over one-month Libor at origination, and the weighted average life as advertised in the prospectus. The dataset we obtain from Bloomberg covers over 90 percent of all subprime PLS issued in the U.S. between 2000 and 2007. We are able to combine the two datasets by merging on individual security CUSIPs. Given that Fannie Mae and Freddie Mac only purchased triple-A PLS, we focus exclusively on the highest rated tranches.

Finally, we obtain monthly unemployment rates at the county-level from the Bureau of Labor Statistics and monthly county-level house price indices from CoreLogic.

3.2 Summary statistics

Table 1 displays the aggregate dollar amount of subprime PLS issued each year over the period 2000—2008 (column 1) obtained from Inside Mortgage Finance's 2011 Mortgage Market Statistical Annual. The table

[14] We eliminate seasoned loans to avoid potential survivorship bias. More than 90 percent of the loans in the CoreLogic database were less than 12 months old at the time of securitization. Other studies that use similar datasets have adopted similar restrictions.

shows the rapid growth in the market that took place during the housing boom and the steep decline that occurred at the onset of the financial crisis. In 2000, a little more than $52 billion subprime PLS was issued. Issuance peaked at $465 billion in 2005, remained roughly constant in 2006, and then dropped precipitously in 2007 to just over $200 billion. Since 2007 the subprime PLS market has virtually disappeared.

Unfortunately, there is no publicly available information on the exact dollar amount of subprime PLS purchased by Fannie Mae and Freddie Mac over the entire sample period. We were able to obtain this information for 2006—2008 from the FHFA's 2011 Annual Report to Congress, which we display in the second column of the table. Prior to 2006, the report does not break out PLS purchases by type of security (subprime, alt-a, prime) for Freddie Mac. In the third column of Table 1, we show estimates of the annual amount of subprime PLS purchased by Fannie Mae and Freddie Mac based on our algorithm. In order to obtain these totals, we sum over all triple-A securities we classify as being eligible for GSE purchase and that are included in Bloomberg. The numbers that we obtain for 2006—2008 are very close to the FHFA figures, which suggests that our algorithm is truly identifying mortgage pools backing subprime PLS purchased by the GSEs. (As mentioned in Section 2.2, we also found that we capture the securities included in the lawsuits brought by the FHFA.) The last column of the table displays the GSEs' combined subprime PLS market shares based on aggregate amounts derived from our algorithm.[15] In 2001 Fannie Mae and Freddie Mac purchased less than four percent of total subprime PLS issued, but in 2004 they bought almost 40 percent. Their market share of purchases fell slightly after 2004, but still remained quite high at around 25 percent through 2007.

Table 2 shows detailed, loan-level summary statistics broken down by mortgages in GSE versus non-GSE pools. The top panel of the table shows information on the distributions of the continuous variables that we include in our mortgage performance regressions. There are some notable differences between GSE and non-GSE mortgage pools. For example, while the FICO distributions are both lower than those in the overall U.S. population, reflecting the fact that many subprime mortgage borrowers have poor credit histories, the

[15] Note that the GSEs' market share of *triple-A* subprime PLS is higher than the percentages listed in the table. Since the triple-A tranches comprised about 80 percent of the typical subprime PLS deal, to obtain a rough estimate of the total amount of triple-A subprime PLS issued over the sample period, we would have to reduce the numbers in column 1 by 20 percent. This suggests that the GSEs' market share of triple-A subprime PLS was about 25 percent higher than the percentages listed in Table 1, so that in 2004 they purchased approximately 50 percent of all triple-A subprime PLS issued.

distribution of loans in GSE pools is significantly lower than the distribution of loans in non-GSE pools (26 points lower on average). Average loan size appears to be similar for both pool types, although this masks important differences across the distributions. The distribution of loan size is much more dispersed in the non-GSE pools, as the top of the distribution is significantly higher due to the presence of non-conforming (jumbo) loans, which are not included in the GSE pools. The bottom of the non-GSE pool loan size distribution is also significantly lower (e.g., tenth percentile is $30,000 compared to almost $60,000 in the GSE pools). The distribution of combined loan-to-value ratios (i.e., loan-to-value ratios including first and second mortgages on the same property) is significantly lower in the GSE pools (the median is about seven percentage points lower and the average is about four points lower), while the distribution of interest rates (at the time of origination) in the GSE pools is also significantly lower than the distribution in non-GSE pools (the median is approximately 60 bps lower and the average is almost 70 bps lower). Another interesting difference between the two pool types can be seen in the distribution of mortgage terms. Virtually all loans in GSE pools have 30-year maturities, while more than 25 percent of the loans in non-GSE pools have maturities less than 20 years. This difference is explained by the fact that non-GSE pools contain many more second lien mortgages than GSE pools.[16] Loans in GSE pools were originated in counties with initially higher unemployment rates, but with slower growth in unemployment rates over the entire sample period. GSE pools are comprised of loans originated in counties in which house prices declined less on average over the crisis period.[17] The last line of the top panel of Table 2 labeled "UAG Zip Code Fraction", shows that, on average, 49.4 percent of the mortgages in non-GSE pools were secured by properties located in low-and-moderate income areas and hence eligible for GSE housing goal credit, whereas 52.4 percent of mortgages in GSE pools qualified towards the underserved areas

[16] Second lien mortgages often have maturities of less than 30 years. In the appendix (Table A.8) we show that our empirical results are not driven by this difference in loan composition by estimating our regression specifications on a sample that excludes second lien mortgages.

[17] Loans in GSE pools were also in areas with slightly higher house-price appreciation over the 12 months preceding the date of issuance. This fact, combined with the observation that loans in GSE pools were in areas with higher unemployment, suggests that Fannie Mae and Freddie Mac may have used the these mortgage- backed securities to assist in meeting their affordable housing goals.

goal.[18] This difference is small and consistent with the affordable housing goals not having been a major driver of loan sorting for GSE and non-GSE pools.[19]

The bottom panel of Table 2 displays averages of the dichotomous variables that we include in our mortgage performance regressions also broken down by GSE and non-GSE pools. The most striking difference between the pool types in the panel is the difference in the share of adjustable-rate mortgages (ARMs). Adjustable-rate loans account for 74 percent of the loans in GSE pools, compared to only 49 percent in non-GSE pools.[20] Below we estimate our regressions using only the sample of ARMs as a robustness test to account for any bias that this difference in composition might introduce in our analysis. The GSE pools are also characterized by lower fractions of low documentation mortgages, interest-only loans, and loans with balloon payments at the time of maturity. A higher fraction of loans in GSE pools contained prepayment penalties. Loans in GSE pools were also less likely to be purchase-money mortgages and more likely to be cash-out refinances.

Finally, the lower part of the panel displays unconditional average default rates for mortgages in both GSE and non-GSE pools. We assume that a mortgage is in default if the borrower is at least two payments behind (60+ days delinquent).[21] Ex-ante default probabilities are calculated from a set of models that use only information before the issuance date of the corresponding deal. We calculate these for horizons of 12 months, 24 months, and 36 months, and describe the details of the empirical models in Section 4.4. Ex-post default rates are calculated directly from the CoreLogic database using information on loan performance from the time

---

[18] Low and moderate income areas are defined as census tracts in which household median income is less than 90 percent of the median income in the metropolitan statistical area (MSA), or the county if the census tract is not located in a MSA. Since the CoreLogic database only includes the identity of the zip code where the property is located we had to use a mapping between zip codes and census tracts to calculate the statistics in Table 2. We provide more details regarding the affordable housing goals and these calculations in Section 4.2 below.

[19] See Ghent, Hernandez-Murillo, and Owyang (2013) for an analysis of the effect of the GSE housing goals on the PLS market.

[20] Virtually all adjustable-rate mortgages in subprime PLS pools were what the industry refers to as "hybrid-ARMs." These loans were typically characterized by a fixed interest rate for 2--3 years at which point the rate would reset to an adjustable-rate that was indexed to a market rate (typically the 6-month LIBOR).

[21] This is conventional in the literature and includes properties that are in the foreclosure process and bank-owned (REO). In the appendix we provide results for our main empirical specifications where we define default to be borrowers that are 90+ days delinquent rather than 60+ days delinquent.

of origination. We report ex-post default rates through the end of 2008, 2010, and 2012 (the end of our Corelogic data sample). The table shows that ex-ante default probabilities associated with GSE pools are similar to those associated with non-GSE pools, and that average ex-post default rates associated with GSE pools are significantly *lower* (by approximately five percentage points for the 2010 and 2012 horizons).

Table 3 displays summary statistics (from Bloomberg) for the triple-A securities that are collateralized by the loans in the CoreLogic database broken down by whether the securities are derived from GSE or non-GSE mortgage pools. For each year, the table shows the number of associated mortgage pools, the average size of the mortgage pools, the spread between the average coupon of the triple-A securities and the one-month LIBOR, and the weighted average expected life of the associated triple-A tranches using the sizes of each individual security as the weights. The last column in the table displays the differences in the summary statistics between the pool types and whether those differences are statistically different from zero. The average pool size is between $19 and $20 million for both GSE and non-GSE pools in each year of our sample. The average spread over one-month LIBOR on GSE triple-A securities was between two and six basis points higher than the average spread on the non-GSE triple-A securities in each year of our sample. Finally, the GSE and non-GSE pools have similar weighted average lives (around 2.4 to 2.5 years), where the life of each security is taken from the prospectus and is based on predicted prepayment behavior on the part of the borrowers.

4. GSE participation and mortgage performance

4.1 Realized performance

The first part of our analysis considers the ex-post performance of mortgages included in GSE pools relative to those in non-GSE pools. We estimate linear probability models (LPM) where the dependent variable is equal to one if a loan is in default (i.e., 60 days or more delinquent) between origination and the end of 2008, 2010 or 2012. We choose to estimate LPMs rather than non-linear discrete choice models (such as logit or probit) for two reasons related to the inclusion of deal-level fixed effects. First, the combination of a large sample (over 10 million loans) and the roughly two thousand deal fixed effects makes it computationally costly to estimate non-linear models. Second, estimating non-linear models with fixed effects can result in inconsistent estimates due to the incidental parameters problem (Neyman and Scott, 1948). Nonetheless, in the Appendix, we re-estimate our main specifications using a logit model to make sure that the results are not driven by the linear functional form assumption.

The main set of results for all three horizons is displayed in Table 4. The first column in each horizon panel corresponds to a regression that only includes issue year effects in the covariate set. The estimated GSE coefficient is positive for all three horizons, and the quantitative magnitude of the coefficients implies that loans backing GSE pools have default rates that are on average 120 to 160 basis points higher than loans backing non-GSE pools. The second column in each panel displays results from regressions that include a full set of controls for loan and borrower characteristics, which includes relevant underwriting variables, as well as economic factors that might impact default rates after origination such as unemployment rates and house price indices. The sign of the GSE coefficient flips, becoming negative and highly statistically significant. Loans in GSE pools default by approximately one percentage point less than loans in non-GSE pools, controlling for observable loan and borrower characteristics.[22] This pattern suggests that the GSEs purchased subprime PLS securities comprised of observably riskier mortgages, but that those mortgages performed better in ways that are unobservable to the econometrician and were likely unobservable to PLS investors at the time of contracting. The third column in each panel adds deal fixed effects to the specification, and thus only uses within-deal variation to estimate the difference in performance between pool types. The addition of deal fixed effects increases the absolute magnitude of the coefficient estimates by almost a factor of two. Mortgages in GSE pools default, on average, 150 to 190 basis points less than loans in non-GSE pools in the same subprime PLS deal. These results show the importance of controlling for deal-level factors like the mortgage servicer, issuer, and originator.[23]

---

[22] The set of borrower and loan control variables is described in detail in Section 2. In order to conserve space, we do not report the associated coefficient estimates in our tables. However, in the Appendix we do report these estimates for two of the regression specifications.

[23] In the appendix, we show results in which linear probability models are substituted with logit models, and models where default is defined as 90+ days delinquent instead of 60+ days delinquent. The results of both modifications are virtually identical to those reported in Table 4. The appendix also presents results for horizons that are based on time from issuance

The observation that mortgages in GSE pools perform much better than those in non-GSE pools in the same deals after controlling for a large set of observable characteristics suggests that the loans backing GSE pools are different in unobservable ways. There are a couple of potential explanations for this pattern. The first is that the GSEs had superior screening technologies (compared to other investors), which are unobservable to us, and this resulted in the loans in GSE pools performing better.[24] For example, it is possible that the GSEs chose to invest in specific mortgage pools based on superior credit risk models that were able to predict loan performance more accurately than other investors. A second possible explanation is that issuers used private information about the quality of the mortgages to give the GSEs higher quality loans within the same deal. Since Fannie Mae and Freddie Mac were such important investors in the market, PLS issuers would have had an incentive to maintain a good reputation with the two institutions in order to ensure a stable source of demand for future business.

4.1.1 Low documentation loans

One natural place to examine whether issuer private information could explain the results is to look specifically at low documentation mortgages. The existing literature has argued that low documentation lending is the segment of the market where banks putting together the deals are most likely to have private information that could lead to systematically differential performance across mortgage pools that is not accounted for by observable loan characteristics (e.g. Keys, Seru and Vig, 2012; Saengchote, 2013; Begley and Purnanandam 2013; Jiang, Nelson and Vytlacil, 2014a, 2014b).

Returning to Table 4, in the fourth column of each panel, we find that the difference in default risk between mortgages in GSE and non-GSE pools is significantly reduced (by 50 percent or more) once we include in the regressions an interaction between the GSE dummy and an indicator variable for low documentation mortgages (note that a low documentation indicator is already included in the model as an

---

rather than a fixed calendar date (24-month and 36-month default horizons from the month of issuance). The point estimates are very similar to those reported in Table 4.

[24] Superior monitoring of borrowers on the part of servicers is unlikely to explain the performance differences since the inclusion of deal fixed effects eliminates variation in mortgage servicing. The same institution virtually always serviced all of the mortgages in a given deal.

underwriting characteristic). This means that differences in performance are relatively small for full documentation loans across GSE and non-GSE pools, but that these differences are significantly amplified in the sample of low documentation loans. Specifically, we find that low documentation loans in GSE pools default by 2.6 to 3.2 percentage points less on average compared to low documentation loans in non-GSE pools. This evidence is consistent with private information on the part of subprime PLS issuers driving the differences in performance between loans in GSE pools and loans in non-GSE pools – suggesting a supply-side driven explanation for the performance differences we observe. If differences in pool performance were driven by better GSE risk assessment, it is not clear why so much of the effect would be concentrated in low documentation loans, as it is unlikely that the GSEs could identify higher quality *low documentation* loans. Figure 2 shows that the results are not driven by one or two quarters alone, but rather that the effects of both the GSE dummy and the low documentation interaction are always below zero, and become especially strong after the first quarter of 2005.

4.1.2 Proportion of deals with GSE participation

In order to further distinguish between the supply (issuer-driven) and demand (GSE-driven) hypotheses, we construct a measure of the frequency of interactions between the GSEs and subprime PLS issuers. Specifically, for each deal issuer in the CoreLogic dataset, we take the number of deals by that issuer that contain a GSE pool up to a given quarter and divide that number by the total number of deals issued by the issuer from the beginning of our sample until that same quarter. We call this variable the "GSE deal fraction", so that a value of one means that all of the issuer's previous deals involved the GSEs, while a value of zero means that none involved the GSEs. Our contention is that this variable measures the closeness between issuer and the GSEs (or the extent of issuer dependency on the GSEs) at a given point in time, such that an issuer with a high GSE deal fraction will have an incentive to supply the GSEs with higher quality loans in order to maintain a good reputation. If the difference in ex-post performance is largely explained by loans in pools arranged by issuers with high GSE deal fractions, then a supply story is the most likely explanation.

In Table 5, we show descriptive statistics of the GSE deal fraction variable for each year in our sample period. Some issuers almost exclusively created deals that included pools aimed at the GSEs (e.g., Fremont had a GSE pool in 100 percent of their deals, as did Wells Fargo, Barclays and Fieldstone in 2004 and 2005). But the mean of this variable across all issuers is about 60 percent, suggesting that other issuers also sold a significant fraction of deals without a pool specifically directed at the GSEs.[25] One notable observation from the table is that there is substantial variation in the GSE deal fraction variable both across issuers as well as over time for a given issuer.

In Table 6 we assess the extent to which prior deals with the GSEs affects the finding that within-deal mortgage performance was better for GSE pools -- particularly for low documentation mortgages. We add an interaction between the GSE deal fraction variable and the GSE dummy to the specification in Table 4. The first three columns in Table 6 display the estimation results for the three different default horizons. The coefficient estimate associated with the interaction term is negative and statistically significant, with a magnitude between 0.026 and 0.033 depending on the horizon. Thus, a loan in a GSE pool arranged by an issuer with a GSE deal fraction of one is approximately 2.6 to 3.3 percentage points less likely to default than a loan in a GSE pool arranged by an issuer with no prior experience with the GSEs (i.e. a GSE deal fraction of zero). The interaction between the GSE dummy and the low documentation dummy is largely unaffected, remaining negative and statistically significant. However, the addition of the new interaction term causes the sign of the GSE dummy coefficient to flip from negative to positive. This suggests that full documentation loans in GSE pools arranged by issuers with no prior experience with the GSEs are *more* risky than similar loans in non-GSE pools within the same deal. Thus, controlling for documentation status and previous experience with the GSEs fully explains the ex-post performance differential between mortgages in GSE pools and non-GSE pools in the same deal.

[25] Identifying the deal issuers was not completely straightforward. We obtained the identity of the issuer of many deals from Bloomberg. However, for some deals, Bloomberg does not have information on the identity of the issuer. In these cases we took the ticker number, and cross-referenced it against the SEC database on company filings. In most cases we were able to obtain the pooling and servicing agreement (PSA) for the corresponding deal, which contains the identity of the deal issuer.

4.1.3 Issuer-originator affiliation

The previous section demonstrated that subprime PLS issuers with a greater proportion of deals that included securities designed for Fannie Mae and Freddie Mac delivered unobservably better-quality low documentation mortgages into GSE pools. This is consistent with those issuers having private information and using it to sort loans into pools. A natural question is how PLS issuers might come by such information, since it is the originator rather than the issuer that directly interacts with borrowers and underwrites mortgages. There are a couple of possible ways in which private information could be transferred from originators to issuers. First, there are direct relationships between many issuers and originators in the subprime PLS market. In some cases the originator and issuer are the same institution, while in others they are part of the same vertically integrated corporation (in which case the originator is typically a subsidiary of the issuer). To the extent that issuers are more likely to obtain private information about loans that are underwritten by affiliated originators (an argument also made by Demiroglu and James, 2012; He, Qian, Strahan, 2012; and Furfine, 2014), we would expect to find stronger results for the sample of loans in which the issuer and originator are affiliated corporations. Second, it is also possible for private information to be transferred between unaffiliated originators and issuers. For example, an issuer may have sufficient experience with a group of originators to be able to identify those that are especially meticulous in their screening of loans (beyond what can be inferred from the set of observable borrower characteristics). Our contention, however, is that it is likely easier to transfer private information when there is a direct affiliation.

Columns 4 through 9 of Table 6 replicate the first three columns in the same table, but separate deals based on whether the issuer and originator were affiliated at the time that the deal was issued (either the same institution or part of the same vertically integrated corporation). Approximately two-thirds of the observations in the CoreLogic database (64.2 percent of our sample) contain information on the identity of the originator of the mortgages included in the deal; and for this subset we are able to determine whether the originator and issuer are affiliated. The columns with issuer-affiliated originators include 396 deals where all loans are by affiliated originators, and the "unaffiliated" column has 695 deals where no loans are made by issuer-affiliated originators. This leaves out 85 deals that had a mix of both affiliated and unaffiliated issuers.[26] We find that within issuer-originator affiliated deals (columns 4 through 6), low documentation mortgages in GSE pools perform significantly better compared to those in non-GSE pools; and that the fraction of previous deals made with the GSEs is strongly correlated with performance. For unaffiliated deals, we also obtain negative point estimates for both low documentation and GSE deal fraction variable interactions, but the results are statistically and economically weaker (especially for the 2010 and 2012 horizons). This is consistent with better transmission and use of private information within the same organization, and suggests that unaffiliated issuers were less likely to have private information that they could use to form mortgage pools.

4.2 Robustness Tests

There are two potential sources of differences between the GSE and non-GSE pools driven by institutional features of the GSEs. The first is directly related to the way we identify the GSE pools in the first place, namely that they do not contain jumbo loans. To the extent that jumbo loans might behave systematically worse due to unobserved characteristics, this could explain some of the results above. The second potential issue concerns the GSEs' percent-of-business affordable housing goals set by regulation. It could be the case that the GSEs demanded more loans in census tracts that qualified for the affordable housing goals, and that these census tracts are different along dimensions unobservable to us, which could explain the performance differences that we are finding.

Table 7 excludes jumbo loans from the estimation altogether. The first thing to note is that only a very small fraction of mortgages in the non-GSE pools are jumbos (about 680,000, or 6.5 percent of the whole sample). When we exclude them from the regressions, we obtain essentially the same estimates as in Table 4 and Table 6, both for the GSE "main effect" and for the interactions with low documentation and the GSE deal fraction variables.

[26] The 85 mixed deals are dropped due to our lack of confidence in the identity of the originator and/or our ability to identify a relationship between the issuer and originator (the raw data on originator identities in the CoreLogic database is somewhat messy, so we were forced to expend significant effort in cleaning and standardizing the names in order to integrate the information into our empirical analysis).

Table 8 repeats the regressions in Table 4 and Table 6 accounting for whether a loan was located in a low-and-moderate income area and hence eligible for GSE affordable housing goal credit. In order to determine underserved area goal (UAG) eligibility, we match tract-level UAG data obtained from the FHFA to the zip code associated with each mortgage using a population-weighted bridge provided by the Missouri Census Data Center. We use 1990 tract definitions for tract data up to 2002, and 2000 tract definitions for the later years.[27] We include a zip-code level variable that varies between 0 if the zip code does not contain a census tract that is eligible for the UAG, and 1 if the entire population in that the zip code is in UAG-eligible census tracts. This variable takes intermediate values when only a fraction of the zip code's population is in such tracts. We find that the results are virtually unchanged when we include these additional controls.[28]

As an additional robustness test, we also consider whether our results are driven by potential differences in exposure to pre-payment risk across GSE and non-GSE pools. In particular, it could be that the GSEs looked for a different pre-payment profile from their PLS securities than other investors, and that those differences in pre-payment profiles simply happened to be correlated with better credit risk ex post. This would be consistent with the lower proportion of fixed-rate mortgages in GSE pools relative to non-GSE pools that can be inferred from the second panel of Table 2. We should note that all previous regressions control for these observable characteristics (i.e., they control for mortgage type), but here we test whether unobservable characteristics of a specific subset of loans (fixed-rate mortgages in particular) that have different exposure to pre-payment risk could explain the results. Table 9 shows that the results from Table 4 and Table 6 are essentially unchanged when we look only at adjustable-rate loans. By restricting the sample to only ARMs we

[27] A census tract in a metropolitan area is classified as an underserved area if the tract's median income is no greater than 90 percent of median income for the metropolitan area; or if minority households comprise at least 30 percent of the tract's population and tract median income is no greater than 120 percent of area median income. A similar definition, based on counties, is employed for nonmetropolitan areas.

[28] In fact, we include a series of dichotomous variables rather than a single continuous variable. In the appendix, we describe how they are constructed and show the coefficient estimates associated with the underserved areas goal variables. We find that loans originated in zip codes with a higher fraction of the population in goal-eligible census tracts are characterized by higher default rates, ceteris paribus.

are able to keep pre-payment profiles largely fixed, which indicates that results are not driven by unobserved credit factors that could be correlated with pre-payment risk.[29]

4.3 Comparison of ex-ante risk characteristics

In this section we compare the riskiness of the mortgages underlying GSE and non–GSE pools based on only the borrower and loan characteristics available at the time that the deals were issued. A comparison of ex-ante default risk allows us to verify whether differences in loan performance between GSE and non-GSE pools were predictable by investors using only information at the time of contracting. Additionally, we are also able to specify more flexible models than those in Tables 4 and 6, as well as explicitly control for differences in expected prepayment behavior.

We construct ex-ante default probabilities for each mortgage in our sample in the spirit of the model in Ashcraft, Goldsmith-Pinkham and Vickery (2010). For each loan in the sample, we determine the quarter in which the corresponding deal was issued. We then take all loans in deals issued between 24 months and 12 months *before* that quarter, and track those mortgages over the subsequent 12 months, creating indicator variables that take values of one if the mortgage is 60 days delinquent, 90 days delinquent, in foreclosure or in REO (or any other liquidation status following foreclosure) at any point during the 12 month period, respectively.[30] We perform the same exercise for horizons of 24 and 36 months.[31] We then estimate three different discrete choice models: a linear probability model, a logistic regression, and a multinomial logistic regression that specifically accounts for the fact that mortgages can prepay as well as default. In addition to these three models, we estimate a variant of a competing risks hazard model, which is a standard model used

---

[29] An alternative method to control for variation in prepayment rates in the regression would be to jointly model prepayment and default, which is often done in the mortgage valuation literature using non-linear models such as multinomial logit or hazard models with competing risks. We do this below as part of an analysis of ex-ante default probabilities. However, this is simply not feasible in our analysis of ex-post performance due to the large number of deal fixed effects that are included in the regressions.

[30] For the 12-month predicted default variable we take loans that were originated between 24 months and 12 months before that quarter, so that we have a full 12 months of history for each loan.

[31] For the 24-month horizon we take all loans that collateralized deals issued between 36 months and 24 months before the quarter of interest and track those mortgages over the subsequent 24 months, while for the 36-month horizon we take all loans in deals issued between 48 and 36 months before the quarter of interest and track those loans over the subsequent 36 months.

in the empirical mortgage default literature.[32] The hazard model allows for the inclusion of covariates that vary over time at a monthly frequency, and thus, could potentially allow us to incorporate forecasts of variables like house prices, unemployment rates, and interest rates into the calculation of ex-ante default probabilities.[33]

The regressions are estimated each quarter over the period 2003—2007 and include most of the variables in Table 2 as well as state fixed effects.[34] We take the estimated coefficients from these loan-level credit risk models and apply them to the characteristics of the loans in deals issued in the current quarter to create the 12-month, 24-month, and 36-month loan-level default probabilities. This means that ex-ante default probabilities are created using only information available at the time in which the deals are issued.[35]

Table 10 shows the results from loan-level regressions of the 12-month, 24-month and 36-month ex-ante (predicted) default probabilities on a dummy variable that is equal to one for GSE pools identified by our algorithm. As discussed above, we use four alternative models to compute the default probabilities at the loan level. Panels A, B, C, and D correspond to the linear probability model, logit model, multinomial logit model, and competing risks hazard model, respectively, as the underlying models for constructing ex-ante default probabilities.

---

[32] The competing risks are default and prepayment. We assume a multinomial logit specification for the hazards.

[33] The results on ex-ante probabilities derived from the hazard model in Table 11 assume that the time-varying variables do not change over the horizon (i.e. flat house prices and unemployment rates). The estimated differences in ex-ante probabilities are not sensitive to this assumption however.

[34] Specifically, the variables we use in the model are the combined loan-to-value ratio, the logarithm of the original loan balance, the original interest rate, the credit score, the original size of the loan, the original term, the number of months between origination and issuance (seasoning), indicator variables for low documentation, interest-only loans, first lien loan, negative amortization, residence status (owner-occupied, investor/vacation home), loan purpose (cash-out refinance, other refinance, purchase), property type (condominium, multi-family, single-family), and the existence of a prepayment penalty. We also include the level of the unemployment rate and the level of the house price index at the time of issuance, as well as the 12-month trailing house price appreciation at the county level and the 12-month trailing unemployment rate at the county level. Additional indicator variables are included whenever there are missing observations in any of the controls. We also used more disaggregated fixed effects at the county level, but this change had a trivial impact on the predicted default probabilities. Finally, we also tried estimating separate models for jumbo loans, second lien loans, and adjustable-rate loans. These variations, which are displayed in the appendix, also had little impact on the results.

[35] Most mortgages are seasoned only one or two months so that the month of origination closely corresponds to the month in which the deal is issued. In addition we eliminate mortgages seasoned more than 12 months, so the two months are never more than one year apart. We have estimated the credit risk models relative to the month of origination rather than issuance. The results are not significantly affected by this change.

The results in Table 10 show that loans in GSE pools either do not look riskier or the difference in predicted default probabilities is, in fact, positive, meaning that the GSE pools look somewhat riskier using information available at origination. The linear probability model shows statistically insignificant and economically very small differences in the ex-ante riskiness of loans in GSE pools versus those in pools directed at other investors, whereas the logit and multinomial logit models suggest that there is a higher ex-ante default probability of 60-100 basis points at a two or three-year horizon (although there is no difference at the 12-month horizon). We should point out that a very large share of the variation in ex-ante mortgage risk can be explained by the deal fixed effects included in these regressions (R-squared goes from 1 percent to over 30 percent). These results confirm that there were no observable differences between GSE and non-GSE pools, and that the differences that emerge ex post are likely the result of *unobservable* differences between pools.

4.4 Analysis of MBS prices

Finally, we turn to the pricing of subprime MBS and we consider whether GSE and non-GSE pools were priced differently at the time of issuance. We focus in particular on the yield spreads of triple-A securities with claims on GSE and non-GSE pools, and limit the sample to deals where all the triple-A securities were either floating rate tranches or inverse floaters. Focusing on floating rate tranches has two main advantages: First, because the yield on those tranches is always quoted as a spread over one-month Libor in our sample, we can cleanly aggregate the yield for multiple tranches in the same deal and construct a pool-level spread. Second, because these tranches have a very short duration, we can ignore interest rate risk and the negative convexity issue that arises with fixed-rate mortgage-backed securities. Including only the floating rate mortgage pools drops 177 observations, leaving 3,290 unique pools for this set of regressions.

Table 11 reports the results of regressions of the at-issuance yield spreads of the securities in our sample (relative to Libor) on the GSE dummy variable, a control for the weighted average of the expected life of each security and pool-level loan characteristics. We find that the yield spreads on GSE-purchased securities are, on average, three to six basis points higher than those purchased by non-GSE investors. This holds both with and without deal fixed effects, and also when we control for observable characteristics of the mortgages in each pool, which indicates that Fannie Mae and Freddie Mac were able to obtain higher yielding securities relative to other investors in the same deals buying similarly triple-A rated securities. Given that these pools performed significantly better in terms of realized default, the GSEs seem to have gotten a better deal based on realized returns than other investors in the same deals.

In columns 3 and 6 we consider whether the spread difference between GSE and non-GSE pools co-varies with the "GSE deal fraction" variable. Indeed, we find that the whole spread difference can be explained by the interaction of the GSE dummy with this variable, suggesting that Fannie Mae and Freddie Mac received particularly good deals from issuers that frequently included GSE pools in their deals.

5. Conclusion

Fannie Mae and Freddie Mac have long played a central role in the U.S. housing finance system as both securitizers and investors in conforming prime mortgages. Moreover, during the recent U.S. housing boom, the GSEs were also the two largest investors in subprime PLS – a fact that has largely escaped academic attention.

In this paper we use a unique feature of the structure of subprime PLS deals to show that, conditional on observable risk characteristics, the loan pools that were eligible to be bought by Fannie Mae and Freddie Mac performed significantly better during the crisis relative to mortgage pools backing securities sold to other investors in the same deals. We document that this difference is concentrated in low documentation loans, which suggests that issuers were using private information to sort mortgages into GSE and non-GSE pools. Deals sold by issuers that frequently structured deals for Fannie Mae and Freddie Mac also exhibit larger differences in performance between GSE and non-GSE pools, especially when the issuer of the deals and the originator of the mortgages are affiliated institutions. This further supports the view that this was largely a supply-driven response by the issuers to the presence of these large investors in the subprime PLS market.

The subprime mortgage securities purchased by the GSEs that are the focus of this analysis were largely thought of as information-insensitive at the time of issuance, as they were rated triple-A. Our results suggest that the presence of large concentrated investors positively effects security performance in cases of information-insensitive arm's-length debt where investors are generally thought of as more passive providers of capital. This is an important finding as it implies that large and concentrated claimholders can influence security design and performance well beyond corporate equity and bank debt, where their role has been well established by the literature.

References

Acharya, V. V., Richardson, M., Van Nieuwerburgh, S., and White, L. J., 2011, "Guaranteed to Fail: Fannie Mae, Freddie Mac, and the Debacle of Mortgage Finance." Princeton University Press.

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

Aghion, P., Van Reenen, J., Zingales, L., 2013, "Innovation and institutional ownership." American Economic Review, 103(1), 277-304.

Ambrose, B. W., and T. G. Thibodeau, 2004, "Have the GSE Affordable Housing Goals Increased the Supply of Mortgage Credit?" *Regional Science and Urban Economics* 34, 263-273.

Ambrose, B. W., LaCour-Little, M., and A.B. Sanders, 2005, "Does regulatory capital arbitrage, reputation, or asymmetric information drive securitization?" *Journal of Financial Services Research*, *28*(1-3): 113–133.

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

Ashcraft, A., Goldsmith-Pinkham, P., and J. Vickery, 2010, "MBS Ratings and the Mortgage Credit Boom." Federal Reserve Bank of New York Staff Report 449.

Begley, T., and A. Purnanandam, 2013, "Design of Financial Securities: Empirical Evidence from Private-label RMBS Deals." Working Paper.

Bhattacharaya, S., and A.V. Thakor, "Contemporary Banking Theory." *Journal of Financial Intermediation*, 3, 2-50.

Bhutta, N., 2012, "GSE activity and mortgage supply in lower-income and minority neighborhoods: the effect of the affordable housing goals." *The Journal of Real Estate Finance and Economics*, 45(1), 238-261.

Boot, A., 2000, "Relationship Banking: What Do We Know?" *Journal of Financial Intermediation*, 9(1): 7–25.

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

Brav, A., Jiang, W., Partnoy, F., and R. Thomas, 2008. "Hedge fund activism, corporate governance, and firm performance," *Journal of Finance*, 63(4): 1729–1775.

Bubb, R., and A. Kaufman, 2014. "Securitization and moral hazard: Evidence from a lender cutoff rule," *Journal of Monetary Economics*, forthcoming.

Bunce, H., 2007, "The GSEs Funding of Affordable Loans: A 2004-05 Update." U.S. Department of Housing and Urban Development, Offce of Policy Development and Research Working Paper #HF-018.

Burkart, M., Gromb, D., and F. Panunzi, 1997, "Large shareholders, monitoring, and the value of the firm," *Quarterly Journal of Economics,* 112(3): 693–728.

Chang, Y-C., Hong, H. G., Liskovich, I., 2013, "Regression Discontinuity and the Price Effects of Stock Market Indexing." *Review of Financial Studies*, forthcoming.

Crane, A. D., Michenaud, S., Weston, J., 2013, "The Effect of Institutional Ownership on Payout Policy: A Regression Discontinuity Design Approach." SSRN Working Paper.

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

DeMarzo, P., 2005, "The Pooling and Traching of Securities: A Model of Informed Intermediation," *Review of Financial Studies*, *18*(1): 1–35.

Edmans, A., 2009, "Blockholder Trading, Market Efficiency, and Managerial Myopia," *Journal of Finance*, 64(6): 2481–2513.

Elul, R. 2011, "Securitization and mortgage default," Available at SSRN 1786317.

Eramo, G., Gambacorta, L., and C. Salleo, 2014. "Asymmetric Information in Securitization: An Empirical Assessment," *Journal of Monetary Economics*, forthcoming.

Federal Crisis Inquiry Commission, 2011, "The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States." Government Printing Office.

Furfine, C. H., 2014. "Complexity and Loan Performance: Evidence from the Securitization of Commercial Mortgages," Review of Corporate Finance Studies, forthcoming.

Ghent, A., Hernandez-Murillo, R., and Owyang, M., 2013, "Did Affordable Housing Legislation Contribute to the Subprime Securities Boom?" Working Paper.

Greenlaw, D., Hatzius, J., Kashyap, A., and Shin, H. S., 2008, "Leveraged Losses: Lessons from the Mortgage Market Meltdown," *Proceedings of the U.S. Monetary Policy Forum 2008*.

Greenspan, A., 2005, *Regulatory Reform of the Government-Sponsored Enterprises*, Testimony before the Committee on Banking, Housing and Urban Affairs, U.S. Senate (April 6).

He, J. J., Qian, J. Q., and Strahan, P. E., 2012, "Are All Ratings Created Equal? The Impact of Issuer Size on the Pricing of Mortgage-Backed Securities." *The Journal of Finance*, 67(6), 2097-2137.

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

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

Krainer, J., and E. Laderman, 2009, "Mortgage loan securitization and relative loan performance," *Journal of Financial Services Research*, 1-28.

Jiang, W., Nelson, A. A., and E. Vytlacil, 2014a, "Liar's loan? Effects of origination channel and information falsification on mortgage delinquency." *Review of Economics and Statistics*, forthcoming.

Jiang, W., Nelson, A. A., and E. Vytlacil, 2014b, "Securitization and Loan Performance: Ex Ante and Ex Post Relations in the Mortgage Market," *Review of Financial Studies*, 27(2): 454–483.

McCahery, J., Sautner, Z., and L. Starks, 2010, "Behind the scenes: The corporate governance preferences of institutional investors," Tilburg Law School Research Paper No. 010/2010.

Morse, A., 2013, "Influence in Delegated Management: Active Investors in Private Equity Funds", Working Paper.

Mullins, W., 2013, "The Governance Impact of Index Funds: Evidence from Regression Discontinuity," MIT Working Paper.

Neyman, J., Scott, E.L., 1948, "Consistent estimation from partially consistent observations," *Econometrica* 16: 1–32.

Passmore, W., 2005, "The GSE Implicit Subsidy and the Value of Government Ambiguity." Federal Reserve Board Staff Working Paper 2005–05.

Parrino, R., Sias, R., and L. Starks, 2003, "Voting with their feet: Institutional ownership changes around forced CEO turnover," *Journal of Financial Economics* 68(1): 3–46.

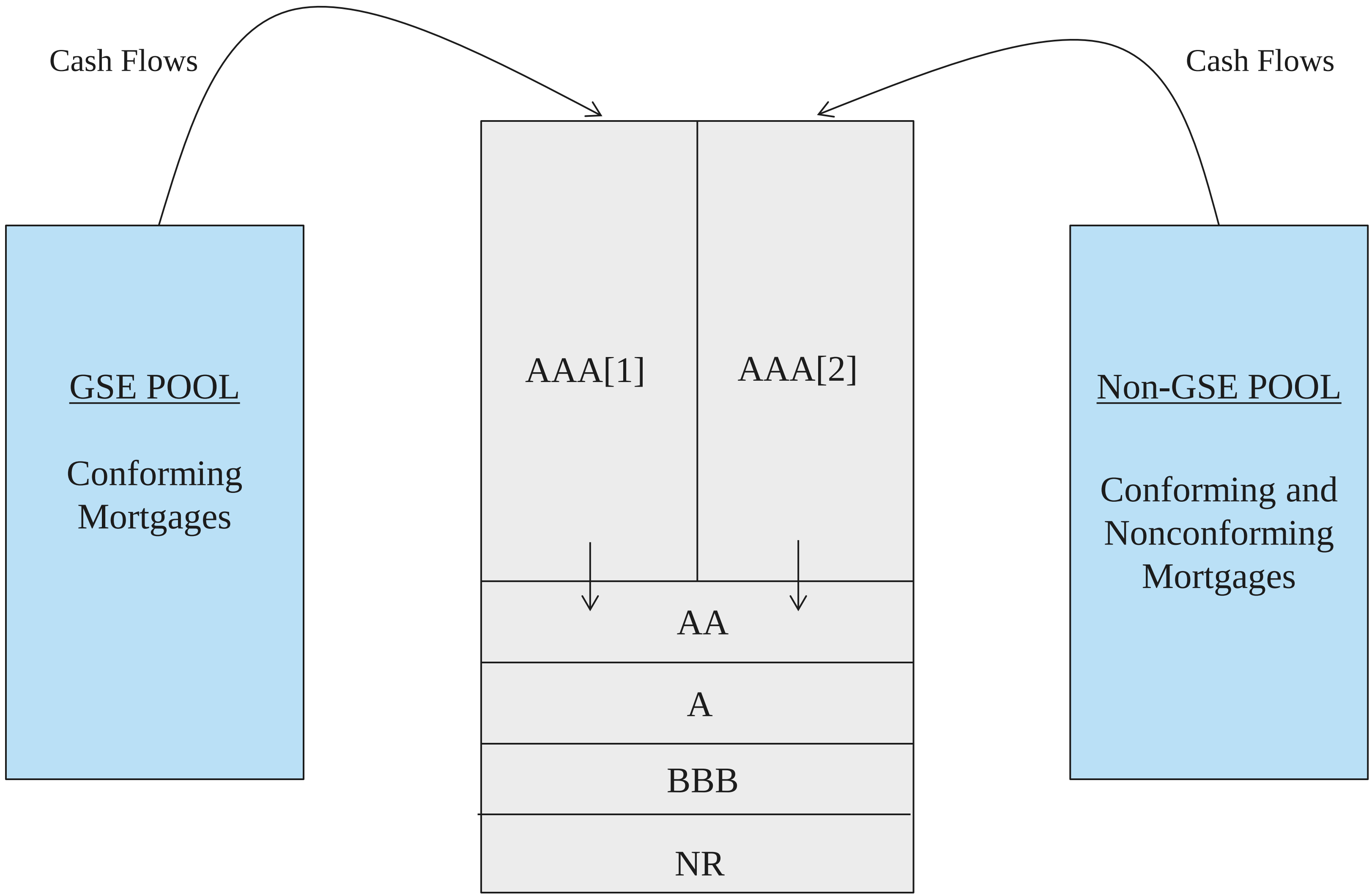
Purnanandam, A, 2011, "Originate-to-distribute model and the subprime mortgage crisis," *Review of Financial Studies*, *24*(6): 1881–1915.

Roberts, M. R., and A. Sufi, 2009, "Renegotiation of financial contracts: Evidence from private credit agreements," *Journal of Financial Economics,* 93(2): 159–184.

Saengchote, K., 2013, "Soft Information in the Subprime Mortgage Market," 26th Australasian Finance and Banking Conference. Available at SSRN: http://ssrn.com/abstract=2169592.

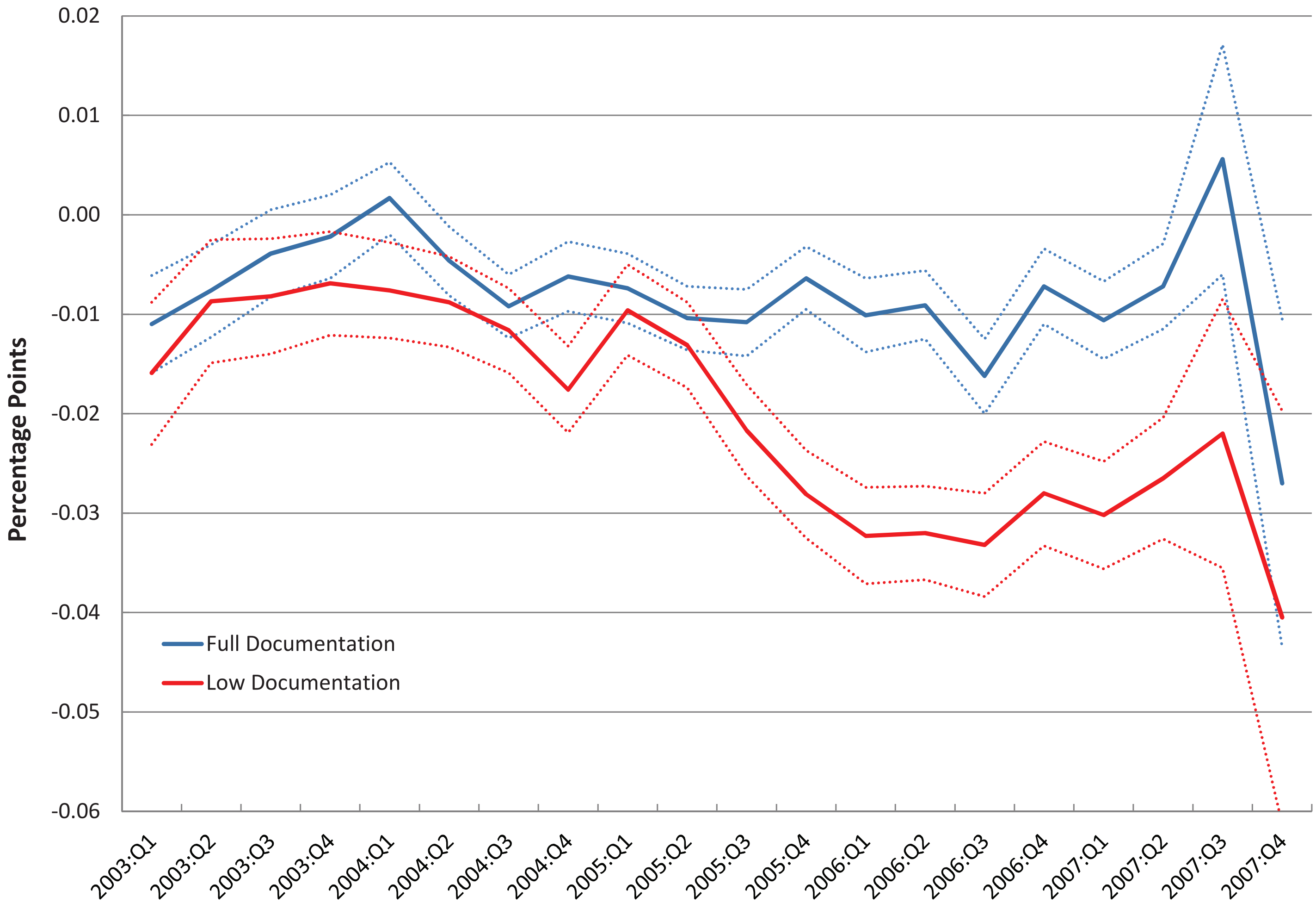
Shleifer, A., and R. Vishny, 1986, "Large shareholders and corporate control," *Journal of Political Economy*, 94(3): 461–488.

Figure 1: Typical Subprime PLS Deal Structure with GSE Participation



This figure displays the structure of a typical subprime PLS deal purchased by the GSEs. These deals involved more than one mortgage pool: one consisting of only conforming loans ("GSE pool") and at least one other pool made up of both conforming and non-conforming (jumbo) loans ("Non-GSE Pool"). The lower rated securities derived their cash flows from all pools, while the triple-A securities purchased by the GSEs derived their cash flows exclusively from the conforming pool and the triple-A securities purchased by other investors derived their cash flows from the other pools.

Figure 2: Evolution of Ex-Post Performance Difference between GSE and Non-GSE Mortgage Pools over Sample Period



Notes: This figure plots the coefficient estimates of the GSE dummy variable and the interaction of the GSE dummy and the low documentation dummy from a series of linear probability models estimated separately for each quarter of issuance over the sample period 2003–2007. The dependent variable is the ex-post default rate measured through 2008, where default is defined as a loan being 60 or 90 days delinquent, in foreclosure or REO. The model specification is identical to column 4 in Table 5, and is estimated separately for each quarter (of issuance) in the sample. The solid lines correspond to the point estimates from each linear probability model, while the 95 percent confidence intervals are given by the dashed lines.

Table 1: Subprime Private-label Securities Issuance (PLS): 2000–2008

Year	Subprime PLS Issuance ($ billions)	GSE Subprime PLS Purchases: Public Data ($ billions)	GSE Subprime PLS Purchases: Proprietary Data ($ billions)	GSE Subprime PLS Purchases: Market Share (%)
2000	52.5	.	.	.
2001	87.1	.	3.4	3.8
2002	122.7	.	14.6	11.9
2003	195.0	.	67.7	34.7
2004	362.6	.	141.0	38.9
2005	465.0	.	134.4	28.9
2006	448.6	110.4	106.0	23.6
2007	201.6	59.6	50.1	24.9
2008	2.3	0.7	.	.

Notes: Subprime PLS Issuance is obtained from the 2011 Mortgage Market Statistical Annual (volume II, page 31). Publicly available data on PLS purchased by the GSEs is obtained from the 2012 Federal Housing Finance Agency's (FHFA) Annual Report to Congress. The FHFA report only breaks out PLS purchases into subprime and Alt-A for Freddie Mac beginning in 2006. Proprietary data on PLS purchased by the GSEs is obtained from CoreLogic's Asset-Backed Securities database. The GSEs' market share of PLS purchases (column 5) is obtained by dividing GSE PLS purchases (column 4) by total subprime PLS issuance (column 2).

Table 2: Loan-level Summary Statistics: Corelogic Subprime PLS Issued 2003–2007

Continuous Variables	Non-GSE (N = 6,324,311)						GSE (N = 4,140,711)					
	Mean	10 perc.	25th perc.	Median	75th perc.	90th perc.	Mean	10 perc.	25th perc.	Median	75th perc.	90th perc.
FICO (Points)	642	558	600	641	684	729	616	538	574	615	653	692
Balance ($)	159,224	30,000	54,900	109,180	214,000	375,000	156,907	59,600	90,980	140,000	209,600	282,000
CLTV (P.Points)	88.8	70.0	80.0	91.8	100.0	100.0	84.4	65.0	78.8	85.0	95.0	100.0
Orig. Rate (P. Points)	8.64	6.44	7.20	8.33	9.90	11.45	7.94	6.25	6.88	7.75	8.75	9.95
Term (months)	314	180	240	360	360	360	350	360	360	360	360	360
Unemployment (P. Points)	5.09	3.40	4.10	4.90	5.80	6.90	5.39	3.60	4.30	5.20	6.20	7.20
Trailing 12-month unemployment change	-6.5%	-18.6%	-13.2%	-7.4%	-1.1%	7.0%	-5.3%	-17.6%	-12.4%	-6.4%	0.0%	8.6%
Unemployment change through 2012	54.7%	7.1%	25.7%	48.1%	79.2%	108.5%	47.1%	2.8%	20.0%	41.7%	66.7%	96.2%
Trailing 12-month HPA	12.1%	1.2%	4.4%	9.8%	18.9%	26.8%	12.3%	1.9%	4.8%	9.9%	18.7%	26.8%
HPA through 2012	-17.5%	-42.8%	-32.2%	-18.8%	-3.8%	8.4%	-13.8%	-39.5%	-28.1%	-14.6%	-0.4%	11.1%
UAG Zip Code Fraction	49.4%	0.0%	8.2%	49.9%	88.3%	100%	52.4%	0.0%	12.0%	55.7%	93.4%	100%

Indicator Variables	Non-GSE Mean	GSE Mean
Low Documentation (share)	0.412	0.347
Non-Owner Occupied (share)	0.083	0.084
Purchase Loan (share)	0.508	0.356
Cash-Out Refinance (share)	0.422	0.563
Interest-Only (share)	0.137	0.096
Balloon (share)	0.225	0.094
ARM (share)	0.489	0.744
Prepay Penalty (share)	0.616	0.719
12-m. Pred. Loan Default (P. Points)	0.119	0.123
24-m. Pred. Loan Default (P. Points)	0.150	0.171
36-m. Pred. Loan Default (P. Points)	0.163	0.195
Default Rate through 2008:Q4	0.338	0.315
Default Rate through 2010:Q4	0.421	0.376
Default Rate through 2012:Q4	0.444	0.393

Notes: This table shows summary statistics for the loans underlying the GSE and non-GSE triple-A subprime PLS securities in the Corelogic database issued between 2003 and 2007. GSE refers to mortgage pools made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit.The variables coincide with the covariate set included in the ex-ante and ex-post default regressions below. The mean, median, 10th, 25th, 75th, and 90th percentiles of the respective distributions of the continuous variables are displayed, while the mean of the dichotonomous variables is shown. FICO score is the credit score of the borrowers on the loan application; balance is the size of the loan at origination; CLTV is the size of all loans on the property relative to the price of the house (transaction price or appraisal amount, depending on whether it is a purchase or a refinance); original rate is the interest rate on the loan at origination; term is the original term of the loan; unemployment is measured at the county-level (from the BLS); HPA is the county-level house price appreciation (from CoreLogic); low-documentation is a 1 if the loan was either low documentation or no-documentation; non-owner occupant is a 1 if the property is for investment purposes or is a vacation/second home. The bottom of the table shows the 12-month, 24-month, and 36-month predicted default rates for each loan at the time of issuance using all information in the data for the previous two vintages of subprime PLS and defining default as being 60 or 90 days delinquent, in foreclosure or REO. In addition the realized default rates as of 2008:Q4, 2010:Q4, and 2012:Q4 are also displayed.

Table 3: Pool-level Summary Statistics: Corelogic Subprime PLS Issued 2003–2007

Year		Non-GSE	GSE	Difference
2003	# Pools	312	172	140
	Pool Size ($ millions)	19.34	19.83	0.49***
	Spread (bps)	36.96	38.83	1.87
	Average Life (years)	2.91	2.85	-0.07
2004	# Pools	419	297	122
	Pool Size ($ millions)	19.59	20.11	0.52***
	Spread (bps)	30.16	33.17	3.01***
	Average Life (years)	2.66	2.76	0.10*
2005	# Pools	511	316	195
	Pool Size ($ millions)	19.92	20.02	0.11**
	Spread (bps)	20.02	25.88	5.86***
	Average Life (years)	2.31	2.51	0.19***
2006	# Pools	537	314	223
	Pool Size ($ millions)	20.05	19.72	-0.32***
	Spread (bps)	13.46	16.44	2.98***
	Average Life (years)	2.15	2.30	0.15***
2007	# Pools	241	171	70
	Pool Size ($ millions)	19.90	19.58	-0.32***
	Spread (bps)	23.47	25.27	1.80
	Average Life (years)	2.20	2.18	-0.02
All	# Pools	2,020	1,270	750
	Pool Size ($ millions)	19.79	19.88	0.09***
	Spread (bps)	23.41	26.92	3.51***
	Average Life (years)	2.42	2.51	0.09***

Notes: This table shows summary statistics for triple-A subprime PLS issued between 2003 and 2007 broken down by whether the security was collateralized by GSE or non-GSE mortgage pools. GSE refers to mortgage pools made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. The spread refers to the difference between the average coupon of triple-A subprime securities (weighted by the size of the tranche in each pool) and the one-month LIBOR. The average life refers to the average expected life for the tranches as advertised in the prospectus (where the average for the pools was weighted by the size of each tranche). Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 4: Effect of GSE participation on Ex-Post Default Rates

	Horizon through 2008:Q4				Horizon through 2010:Q4				Horizon through 2012:Q4			
GSE (d)	0.016***	-0.011***	-0.019***	-0.007***	0.014**	-0.008***	-0.016***	-0.006**	0.012**	-0.008***	-0.015***	-0.006
	(3.04)	(5.97)	(10.69)	(4.17)	(2.49)	(4.03)	(8.77)	(2.53)	(2.30)	(3.74)	(7.99)	(2.55)
Low Doc			0.057***	0.070***			0.060***	0.072***			0.058***	0.069***
			(8.22)	(10.04)			(8.94)	(10.77)			(9.58)	(11.60)
GSE*Low Doc				-0.032***				-0.029***				-0.026***
				(9.19)				(8.76)				(8.73)
Deal F.E. ?	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Covariates ?	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Issue Year F.E. ?	Y	Y	.	.	Y	Y	.	.	Y	Y	.	.
# Loans	10,465,022	10,465,022	10,464,165	10,464,165	10,465,022	10,465,022	10,464,165	10,464,165	10,465,022	10,465,022	10,464,165	10,464,165
# Deals	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809
Adjusted R^2	0.04	0.14	0.16	0.16	0.09	0.19	0.20	0.20	0.11	0.20	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 2010Q4 and 2012Q4. Default is defined as a loan being 60 or 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 3. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 5: List of Sponsors with Highest Values of “GSE Deal Fraction”

Sponsor	Average value of GSE Deal Fraction (%)						# Deals
	All Years	2003	2004	2005	2006	2007	(2003 - 2007)
Fremont	100	100	100	100	100	.	28
Fieldstone	98.3	.	100	100	94.1	91.7	13
Wells Fargo	94.4	.	100	100	87.4	63.6	11
Barclays	91.8	.	100	100	88.8	84.0	36
Washington Mutual	83.7	84.7	78.2	82.5	86.1	87.9	43
UBS	82.5	100	97.3	89.4	68.7	61.3	42
Morgan Stanley	80.4	75.3	79.5	83.5	81.8	78.6	111
National City	77.8	73.1	77.3	78.2	79.4	.	65
Goldman Sachs	77.3	100	91.3	78.0	70.9	69.0	65
Deutsche Bank	75.7	64.4	81.7	78.9	74.3	73.2	74
All Sponsors	59.7	43.1	59.8	62.4	61.4	58.1	1,751

Notes: This table displays the ten subprime PLS sponsors with the highest values of the “GSE Deal Fraction” measure used in the analysis. “GSE Deal Fraction” is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Only sponsors involved in at least ten deals over the sample period (2003–2007) are included in the table.

Table 6: Ex-Post Default Rates, "GSE Deal Fraction," and Issuer/Originator Affiliation

	All Deals			Affiliated Deals			Unaffiliated Deals		
Horizon	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4
GSE (d)	0.013^{***}	0.012^{***}	0.011^{***}	0.026^{**}	0.025^{***}	0.024^{***}	-0.005	-0.014^{**}	-0.015^{**}
	(3.59)	(4.03)	(4.21)	(2.43)	(2.91)	(3.01)	(0.78)	(2.24)	(2.45)
GSE*Low Doc	-0.032^{***}	-0.029^{***}	-0.026^{***}	-0.034^{***}	-0.025^{***}	-0.022^{***}	-0.032^{***}	-0.009	-0.008
	(8.99)	(8.76)	(8.72)	(6.29)	(6.26)	(6.19)	(7.08)	(0.80)	(0.80)
GSE*"GSE Deal Fraction"	-0.033^{***}	-0.027^{***}	-0.026^{***}	-0.052^{***}	-0.051^{***}	-0.050^{***}	-0.005	0.009	0.010
	(5.73)	(4.56)	(4.43)	(3.21)	(3.58)	(3.70)	(0.58)	(1.00)	(1.14)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	10,156,202	10,156,202	10,156,202	2,668,773	2,668,773	2,668,773	3,374,320	3,374,320	3,374,320
# Deals	1,724	1,724	1,724	396	396	396	695	695	695
Adjusted R^2	0.16	0.20	0.21	0.15	0.19	0.21	0.15	0.19	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 2010Q4 and 2012Q4. The first three columns display results for the sample of all deals, while columns 4-6 display results where the originator of all loans in a deal is affiliated with the issuer, while the last three columns display results for the sample of deals in which the originator of all loans in a deal is not affiliated with the issuer. Default is defined as a loan being 60 or 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 3. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 7: Robustness Check: Conforming Mortgages Only

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.022***	-0.010***	0.011***	-0.019***	-0.007***	0.011***	-0.018***	-0.007***	0.011***
	(10.67)	(5.69)	(2.87)	(8.81)	(3.07)	(3.48)	(8.04)	(2.94)	(3.66)
GSE*Low Doc		-0.034***	-0.033***		-0.032***	-0.033***		-0.030***	-0.030***
		(9.28)	(9.07)		(9.40)	(9.40)		(9.58)	(9.56)
GSE*"GSE Deal Fraction"			-0.034***			-0.029***			-0.028***
			(5.85)			(4.64)			(4.44)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	9,783,310	9,783,310	9,495,412	9,783,310	9,783,310	9,495,412	9,783,310	9,783,310	9,495,412
# Deals	1,809	1,809	1,724	1,809	1,809	1,724	1,809	1,809	1,724
Adjusted R^2	0.15	0.15	0.15	0.20	0.20	0.20	0.21	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 2010Q4 and 2012Q4. Default is defined as a loan being 60 or 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Loans above the conforming loan limit (jumbo mortgages) are excluded from the sample.

Table 8: Robustness Check: Controlling for Affordable Housing Goal Effects

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.019***	-0.008***	0.013***	-0.016***	-0.006***	0.011***	-0.015***	-0.006***	0.011***
	(11.13)	(4.73)	(3.29)	(9.54)	(2.87)	(3.67)	(8.61)	(2.83)	(3.67)
GSE*Low Doc		-0.032***	-0.032***			-0.029***		-0.026***	-0.026***
		(9.04)	(8.90)			(8.53)		(8.43)	(8.44)
GSE*"GSE Deal Fraction"			-0.032***			-0.027***			-0.026***
			(5.53)			(4.46)			(4.21)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	9,823,431	9,823,431	9,554,443	9,823,431	9,823,431	9,554,443	9,823,431	9,823,431	9,554,443
# Deals	1,809	1,809	1,691	1,809	1,809	1,691	1,809	1,809	1,691
Adjusted R^2	0.16	0.16	0.16	0.20	0.20	0.20	0.21	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 2010Q4 and 2012Q4. In addition to the full list of those controls discussed in section 4 of the text, we include a series of indicator variables that measure the fraction of the population in a zip code that resides in census tracts which meet the qualifications for the underserved area affordable housing goal (UAG). Default is defined as a loan being 60 or 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 9: Robustness Check: Adjustable-Rate Mortgages Only

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.015***	-0.005**	0.019***	-0.015***	-0.007**	0.015***	-0.015***	-0.007***	0.013***
	(7.23)	(2.04)	(5.19)	(6.76)	(2.62)	(4.72)	(6.61)	(2.90)	(4.39)
GSE*Low Doc		-0.028***	-0.028***		-0.023***	-0.023***		-0.021***	-0.021***
		(8.92)	(8.54)		(7.87)	(7.69)		(7.51)	(7.32)
GSE*"GSE Deal Fraction"			-0.037***			-0.033***			-0.032***
			(6.06)			(5.59)			(5.43)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	6,161,367	6,161,367	5,971,766	6,161,367	6,161,367	5,971,766	6,161,367	6,161,367	5,971,766
# Deals	1,634	1,634	1,557	1,634	1,634	1,557	1,634	1,634	1,557
Adjusted R^2	0.14	0.15	0.14	0.2	0.2	0.19	0.21	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 2010Q4 and 2012Q4. Default is defined as a loan being 60 or 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Fixed-rate mortgages are excluded from the sample.

Table 10: Ex-Ante Default Probabilities for Loans in GSE and Non-GSE Pools

Panel A: OLS (Linear Probability Model)

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0053*	-0.0064***	0.0200***	-0.0025	0.0251***	-0.0029
	(1.93)	(3.79)	(4.47)	(1.07)	(5.40)	(1.21)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,465,022	10,465,022	10,465,022	10,465,022	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
R^2	0.00	0.37	0.01	0.31	0.01	0.30

Panel B: Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0035	-0.0021	0.0222***	0.0063***	0.0322***	0.0071***
	(1.20)	(1.25)	(5.55)	(3.30)	(8.72)	(3.44)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,438,305	10,438,305	10,330,159	10,330,159	8,779,903	8,779,903
# Deals	1,809	1,809	1,807	1,807	1,556	1,556
Pseudo R^2	0.00	0.33	0.01	0.24	0.02	0.19

Panel C: Multinomial Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0042	-0.0016	0.0214***	0.0077***	0.0315***	0.0103***
	(1.48)	(1.01)	(5.03)	(3.92)	(9.20)	(5.15)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
Pseudo R^2	0.00	0.32	0.01	0.24	0.01	0.19

Panel D: Competing Risks Duration Model

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0019	-0.0019	0.0116***	0.007***	0.0084*	0.0106***
	(0.67)	(1.15)	(3.06)	(3.91)	(1.92)	(5.66)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,344,000	10,344,000	10,344,000	10,344,000	10,344,000	10,344,000
# Deals	1,804	1,804	1,804	1,804	1,804	1,804
Pseudo R^2	0.00	0.29	0.01	0.30	0.01	0.30

Notes: This table shows loan-level, OLS regressions where the dependent variables are the 12-month, 24-month, and 36-month predicted default rates at the time the loan is originated using all information in the data for the previous two years for the 12-month rate and three years for the 24-month and 36-month predicted rates. Default is defined as a loan being 60 or 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. In Panel A we compute ex-ante default rates using OLS regressions, while Panel B uses ex-ante default rates using logistic regressions, and Panel C uses ex-ante default rates using multinomial logistic regressions. In Panel D we compute ex-ante default probabilities using a competing risks duration model, where we assume a logistic form for the hazards. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 11: Yield Spreads for GSE and Non-GSE Pools

	(1)	(2)	(3)	(4)	(5)	(6)
GSE (d)	2.71***	3.55***	-2.41	4.68***	6.36***	2.75
	(4.59)	(3.99)	(1.28)	(7.20)	(3.88)	(1.26)
Average Life	5.41***	6.99***	6.99***	4.52***	5.49***	5.50***
	(6.53)	(6.66)	(7.66)	(4.73)	(4.78)	
GSE * "GSE Deal Fraction"			9.61***			6.57**
			(3.29)			(2.03)
Pool Characteristics?				Y	Y	Y
Issue Quarter FE?	Y			Y		
Issue FE?		Y	Y		Y	Y
# Pools	3,290	3,290	3,290	3,290	3,290	3,290
Adjusted R^2	0.56	0.79	0.79	0.62	0.84	0.84

Notes: This table shows pool-level, OLS regressions where the dependent variable is the pool-level average spread (in percentage points) over the one-month LIBOR rate. The average spread is calculated by weighting the spread on individual tranches included in each pool by their original dollar amount. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Regressions with "Pool Characteristics" (columns 4,5, and 6) have pool-level controls for the loan characteristics in the pool. A full list of those controls is given in the text in Section 3. Pool-level average life is the average weighted expected life for the tranches in each pool as advertised in the prospectus where the average is weighted by the size of each tranche. The sample includes only triple-A floating rate tranches that are part of deals where all the triple-A tranches are either floating rate or inverse floaters. Standard errors are heteroskedasticity-robust and clustered at the quarter of origination level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Appendix (NOT FOR PUBLICATION)

A.1 Algorithm for Identifying GSE Pools

This section describes our algorithm for identifying pools of mortgages backing subprime PLS deals involving Fannie Mae and Freddie Mac.

As described in the text, we use a unique feature of the subprime PLS market to indirectly identify triple-A subprime PLS purchased by the GSEs. PLS deals involving Fannie Mae and Freddie Mac were split into conforming and nonconforming mortgage pools. This split was necessary to facilitate the GSEs' purchases of PLS since, by law, Fannie Mae and Freddie Mac are only allowed to securitize and/or invest in mortgages below the conforming loan limit. The GSEs would thus purchase triple-A securities that were backed by loans from the pool(s) of exclusively conforming mortgages, while other investors would purchase securities from the pool(s) that contained both conforming and non-conforming loans.

Based on this institutional feature of the PLS market, we design an algorithm to identify mortgage pools that backed securities purchased by either Fannie Mae and Freddie Mac between 2003 and 2007.[1] The algorithm is quite simple, with the following two conditions required for a pool to be categorized as a "GSE pool":

1. At least 99% of loans in the pool must be below the conforming loan limit at the time that securities in the deal are issued to investors.

2. Less than 75% of loans in the pool are second liens.

[1] Unfortunately this algorithm does not allow us to distinguish between mortgage pools backing securities purchased by Fannie Mae versus pools backing securities purchased by Freddie Mac.

Table A.1: Conforming Loan Limits: 2000–2007

Year	Conforming Loan Limit (Single Family Property)
2000	$ 252,700
2001	$ 275,000
2002	$ 300,700
2003	$ 322,700
2004	$ 333,700
2005	$ 359,650
2006	$ 417,000
2007	$ 417,000
2008	$ 417,000

The first restriction is the most important. It is based on whether the loan lies above or below the conforming loan limit at the time that the deal is issued, rather than the time that the mortgage is originated.[2] Table A.1 shows the conforming loan limits for the period 2000–2007, which applied to all states within the continental U.S.

We allow up to 1 percent of the loan pool to be composed of non-conforming mortgages to take into account potential measurement error in the data. Specifically we are concerned with potential error stemming from two variables. First, there may be cases in which the outstanding balance reported in the CoreLogic database is incorrect. Second, there may be cases in which the variable that indicates whether a property is single-family or 2-4 family is incorrectly reported. The conforming loan limits for 2-4 family properties were significantly higher than those for single-family properties. Thus, if an observation is incorrectly categorized as pertaining to a single-family home instead of a 2-4 family property, then we would likely misclassify the observation as a non-conforming mortgage.[3]

[2]This distinction is potentially important because there are often seasoned loans in the mortgage pools so that the loan amount at origination can be higher than the outstanding balance at the time that the deal is issued.

[3]Note that our algorithm only considers single-family mortgages due to the fact that 2-4 family properties were subject to different conforming loan limits depending on the exact number of units, and CoreLogic does not distinguish between properties by the number of units (i.e. it groups 2, 3, and 4 family units

We impose the restriction on the proportion of second lien mortgages because the vast majority of them have outstanding balances below the conforming loan limit.[4] Hence, the conforming loan limit tells us very little about whether or not the GSEs purchased securities collateralized by those loan pools.[5]

A.2 Additional Validation of Algorithm

Table 1 in the paper compares annual, aggregate GSE purchases of subprime PLS as calculated using our algorithm with those listed in the 2011 FHFA Annual Report to Congress. In the table we were only able to compare numbers for 2006 and 2007 because the FHFA report does not break out Freddie Mac's purchases by the type of security (subprime versus Alt-A), whereas it does for Fannie Mae going back to 2003. Another source of information about the aggregate purchases of subprime securities by the GSEs is the Federal Crisis Inquiry Commission (FCIC) Report. In order to infer the annual purchases in 2003–2005 by Freddie Mac, we use a Figure in the FCIC report entitled "Buyers of Non-GSE Mortgage-Backed Securities" (see page 124 of the report). Of course, we cannot obtain precise numbers for Freddie Mac from the figure, but we are able to obtain approximate numbers. Table A.2 displays Fannie Mae's numbers from the FHFA Annual Report, as well as Freddie Mac's inferred numbers from the FCIC figure (as an interval, to allow for potential measurement error). We add these numbers to arrive at a total annual figure for GSE subprime PLS purchases from 2003–2007. In the last column of the table we show the annual purchases derived from our algorithm, which closely tracks the purchases derived from the public sources.

together).

[4]In our sample of pools backing securities issued between 2003 and 2007 there are 245 pools for which the share of second lien loans is greater than 75%.

[5]We also imposed the restriction that a GSE pool can only be associated with a deal that contained at least two mortgage pools. We did this because of our focus on deal-level fixed effects, however there is nothing that prohibited an issuer from structuring a deal with only a single conforming loan pool. There were only a handful of these deals in the CoreLogic database, so the restriction has no effect on the algorithm.

Table A.2: GSE Subprime PLS Annual Purchases: 2003–2007

	FHFA Report to Congress Total	Fannie Mae	FCIC Report Freddie Mac	FHFA + FCIC Reports Total	Algorithm Total
2003	.	25.8	[44-48]	[69.8-73.8]	67.7
2004	.	67.0	[70-74]	[137-141]	141.0
2005	.	24.4	[112-116]	[136.4-140.4]	134.4
2006	110.4	35.6	[72-76]	[107.6-111.6]	106.0
2007	59.6	16.0	[37-41]	[53-57]	50.1

Although a complete list of PLS securities purchased by Fannie Mae and Freddie Mac is not publicly available, one source of information for validating our algorithm at the security-level is a disclosure by the Federal Housing Finance Agency (FHFA) announcing lawsuits against PLS issuers in September of 2011.[6]

In the lawsuits, the FHFA focuses on 718 securities that were purchased by the GSEs, and includes the associated tickers (e.g. "ABFC 2006-HE1 A1"). We attempt to match the 718 securities to subprime PLS tickers obtained from Bloomberg. We are able to identify 478 out of the 718 securities as being subprime, while another 226 have different collateral characteristics.[7] The face values of the 478 subprime securities included in the lawsuit were \$37.3 billion in 2005, \$80.7 billion in 2006, and \$38.3 billion in 2007, vastly less than the total amount of subprime PLS purchased by the GSEs during those years.

We can use the 478 securities to partially validate our algorithm, as we are certain that these securities were purchased by the GSEs. When we match the securities to our GSE indicator variable we find that 476 out of the 478 subprime securities included in the lawsuit (99.6%)

[6] In this lawsuit, the FHFA sued 17 PLS issuers because it concluded that "some portion of the losses that Fannie Mae and Freddie Mac incurred on private-label mortgage-backed securities (PLS) are attributable to misrepresentations and other improper actions by the firms and individuals named in these filings." (FHFA, September 6, 2011, "Federal Housing Finance Agency Statement on Recent Lawsuits Filed").

[7] This leaves out 14 securities that we are not able to match to Bloomberg using the tickers provided in the lawsuit documents.

are classified as GSE, which translates into a type I error rate of 0.4%. We cannot use this test to evaluate the type II error in our algorithm, given that there are many securities purchased by Fannie Mae and Freddie Mac that are not part of the lawsuits.

A.3 Robustness Tables

This section contains tables of the robustness check results referred to in the main text. Tables A.3 and A.4 contain results that correspond to Tables 4, 5, and 7 in the text in which default is defined as 90+ days delinquent rather than 60+ days delinquent. The results are quite similar and suggest that they are not sensitive to the definition of default.

Tables A.5 and A.6 contain ex-ante predicted default probability results (corresponding to Table 4 in the text) in which predicted default probabilities are calculated using slightly different statistical models. Table A.5 uses separate models for first and second lien mortgages and separate models for conforming and non-conforming (jumbo) loans, while Table A.6 uses separate models for adjustable-rate and fixed-rate mortgages. The results are quite similar to those in Table 4, in which a single model was used to calculate all predicted default probabilities.

In Table A.7 we re-estimate our ex-post default rate regressions (Tables 5 and 7 in the text) using logistic models rather than linear probability models. The drawback of using logit models in the presence of fixed effects is the well-known incidental parameters problem. With small numbers of observations within groups, the incidental parameters problem can result in significant bias of the estimates of the slope parameters. Since most of the deals contain thousands of loans in multiple mortgage pools, this is likely not an important issue in our context. In fact, the results displayed in Table A.7 are virtually identical to the results in Tables 5 and 7.[8]

[8] We also considered the conditional logit estimator developed by Chamberlain (1980), which eliminates

In Table A.8 we re-estimate our ex-post default rate regressions (Tables 5 and 7 in the text) on a sample that excludes all second lien mortgages. In the summary statistics displayed in Table 2 in the text, we saw a significant difference in the distribution of loan maturities across GSE and non-GSE pools, which is driven by the fact that GSE pools contain virtually no second lien mortgages. Thus, we want to be sure that the estimation results are not being driven by differences in the prevalence of second lien mortgages across GSE and non-GSE pools. Table A.8 shows that eliminating all second lien mortgages does not affect the results.

Finally, we re-estimate our ex-post default rate regressions in Table A.9 over two different horizons (24 months and 36 months) measured relative to the month of security issuance rather than a specific point in (calendar) time. These horizons are consistent with the methodology used in estimating the ex-ante default probabilities. The estimates in Table A.9 are qualitatively and quantitatively similar to those in Tables 5 and 7 in the text.

---

the fixed effects from the likelihood function, and thus is not susceptible to the incidental parameters problem. However, with large numbers of observations within groups (in our case loans within deals), the estimator becomes difficult to implement computationally.

Table A.3: Predicted Default Probabilities for Loans in GSE and Non-GSE Pools: Alternative Default Definition

Panel A: OLS (Linear Probability Model)

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	-0.0009	-0.0075	0.0085	-0.0067	0.0103	-0.0079
	(0.48)	(5.96)	(2.47)	(3.32)	(2.71)	(3.77)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,465,022	10,465,022	10,465,022	10,465,022	9,168,963	9,168,963
# Deals	1,808	1,808	1,808	1,808	1,570	1,570
R^2	0.00	0.40	0.00	0.33	0.00	0.30

Panel B: Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	-0.0033	-0.0050	0.0108	0.0024	0.0188	0.0039
	(1.39)	(2.88)	(3.35)	(1.55)	(7.14)	(2.36)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,438,305	10,438,305	10,330,159	10,330,159	8,779,903	8,779,903
# Deals	1,808	1,808	1,806	1,806	1,555	1,555
Pseudo R^2	0.00	0.34	0.00	0.24	0.01	0.17

Panel C: Multinomial Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	-0.0025	-0.0046	0.0106	0.0035	0.0186	0.0065
	(1.11)	(2.72)	(3.14)	(2.21)	(7.85)	(3.99)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	9,168,963	9,168,963
# Deals	1,808	1,808	1,808	1,808	1,570	1,570
Pseudo R^2	0.00	0.34	0.00	0.24	0.01	0.17

Notes: This table shows loan-level, OLS regressions where the dependent variables are the 12-month, 24-month, and 36-month predicted default rates at the time the loan is originated using all information in the data for the previous two years for the 12-month rate and three years for the 24-month and 36-month predicted rates. Default is defined as a loan being 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Panel A computes predicted default rates using OLS regressions, while Panel B computes predicted default rates using logistic regressions, and Panel C computes predicted default rates using multinomial logistic regressions. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics.

Table A.4: Ex-Post Default Rates for Loans in GSE and Non-GSE Pools: Alternative Default Definition

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.019	-0.007	0.012	-0.016	-0.005	0.009	-0.015	-0.006	0.008
	(10.43)	(4.66)	(3.45)	(9.62)	(2.68)	(3.18)	(8.75)	(2.65)	(3.21)
GSE*Low Doc		-0.033	-0.033		-0.029	-0.030		-0.027	-0.027
		(9.88)	(9.66)		(8.91)	(9.07)		(8.76)	(8.95)
GSE*"GSE Deal Fraction"			-0.031			-0.023			-0.021
			(5.41)			(3.84)			(3.66)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	10,464,165	10,464,165	10,156,202	10,464,165	10,464,165	10,156,202	10,464,165	10,464,165	10,156,202
# Deals	1,809	1,809	1,724	1,809	1,809	1,724	1,809	1,809	1,724
Adjusted R^2	0.15	0.15	0.15	0.21	0.21	0.21	0.22	0.22	0.22

Notes: This table shows loan-level, OLS regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 20010Q4 and 2012Q4. Default is defined as a loan being 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.5: Ex-Ante Default Probabilities for Loans in Conforming and Non-Conforming Pools: Alternative Model for Generating Predicted Default Rates

Panel A: OLS (Linear Probability Model)

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0006	-0.0034	0.0241	0.0096	0.0232	0.0104
	0.11	-1.07	6.25	3.31	8.72	2.55
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,465,022	10,465,022	10,465,022	10,465,022	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
R^2	0.00	0.35	0.01	0.22	0.01	0.19

Panel B: Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0044	-0.0023	0.0212	0.0079	0.0213	0.0090
	(0.92)	(0.82)	(6.08)	(4.38)	(5.71)	(3.90)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,438,305	10,438,305	10,330,159	10,330,159	8,779,903	8,779,903
# Deals	1,809	1,809	1,807	1,807	1,556	1,556
Pseudo R^2	0.00	0.32	0.01	0.18	0.01	0.19

Panel C: Multinomial Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0035	-0.0026	0.0211	0.0092	0.0189	0.0093
	(0.74)	(0.91)	(6.04)	(5.59)	(5.39)	(3.53)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
Pseudo R^2	0.00	0.33	0.01	0.18	0.01	0.16

Notes: This table shows loan-level, OLS regressions where the dependent variables are the 12-month, 24-month, and 36-month predicted default rates at the time the loan is originated using all information in the data for the previous two years for the 12-month rate and three years for the 24-month and 36-month predicted rates. Predicted default rates are calculated by estimating separate regressions for first and second mortgages and separate regressions for conforming and non-conforming (jumbo) loans. Default is defined as a loan being 60 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Panel A computes predicted default rates using OLS regressions, while Panel B computes predicted default rates using logistic regressions, and Panel C computes predicted default rates using multinomial logistic regressions. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics.

Table A.6: Ex-Ante Default Probabilities for Loans in Conforming and Non-Conforming Pools: Alternative Model for Generating Predicted Default Rates

Panel A: OLS (Linear Probability Model)

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0017	-0.0038	0.0159	-0.0052	0.0045	-0.0092
	(0.55)	(1.75)	(3.82)	(2.53)	(0.84)	(3.85)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,465,022	10,465,022	10,465,022	10,465,022	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
R^2	0.00	0.32	0.00	0.23	0.00	0.26

Panel B: Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0038	-0.0018	0.0214	0.0064	0.0197	0.0057
	(0.89)	(0.84)	(4.91)	(2.77)	(4.86)	(2.42)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,438,305	10,438,305	10,330,159	10,330,159	8,779,903	8,779,903
# Deals	1,809	1,809	1,807	1,807	1,556	1,556
Pseudo R^2	0.00	0.33	0.01	0.17	0.01	0.17

Panel C: Multinomial Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0038	-0.0023	0.0219	0.0054	0.0182	0.0046
	(0.90)	(1.04)	(6.56)	(2.28)	(4.79)	(1.84)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
Pseudo R^2	0.00	0.34	0.01	0.14	0.00	0.17

Notes: This table shows loan-level, OLS regressions where the dependent variables are the 12-month, 24-month, and 36-month predicted default rates at the time the loan is originated using all information in the data for the previous two years for the 12-month rate and three years for the 24-month and 36-month predicted rates. Predicted default rates are calculated by estimating separate regressions for adjustable-rate and fixed-rate mortgages. Default is defined as a loan being 60 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Panel A computes predicted default rates using OLS regressions, while Panel B computes predicted default rates using logistic regressions, and Panel C computes predicted default rates using multinomial logistic regressions. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics.

Table A.7: Ex-Post Default Rates for Loans in GSE and Non-GSE Pools: Logit Model

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.015	-0.004	0.016	-0.014	-0.005	0.013	-0.014	-0.005	0.012
	(10.70)	(2.56)	(4.56)	(9.09)	(2.53)	(4.15)	(8.91)	(2.87)	(3.81)
GSE*Low Doc		-0.030	-0.030		-0.027	-0.027		-0.025	-0.025
		(9.08)	(8.95)		(7.75)	(7.84)		(7.48)	(7.60)
GSE*"GSE Deal Fraction"			-0.032			-0.027			-0.026
			(5.82)			(4.55)			(4.32)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	10,018,355	10,018,355	9,742,002	10,018,355	10,018,355	9,742,002	10,018,355	10,018,355	9,742,002

Notes: This table shows average partial effects from loan-level, logistic regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 20010Q4 and 2012Q4. Default is defined as a loan being 60 days or more delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the average partial effects, the second row shows z-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table A.8: Ex-Post Default Rates for Loans in GSE and Non-GSE Pools: First Lien Mortgages Only

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.014	-0.006	0.013	-0.013	-0.006	0.011	-0.013	-0.006	0.010
	(7.28)	(2.91)	(3.46)	(6.37)	(3.11)	(3.80)	(5.96)	(3.15)	(3.88)
GSE*Low Doc		-0.026	-0.026		-0.020	-0.020		-0.018	-0.018
		(9.48)	(8.99)		(8.11)	(7.89)		(7.47)	(7.30)
GSE*"GSE Deal Fraction"			-0.030			-0.027			-0.026
			(4.83)			(4.53)			(4.50)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue Year F.E. ?	.	.	.	.	.	.	.	.	.
# Loans	7,743,382	7,743,382	7,530,399	7,743,382	7,743,382	7,530,399	7,743,382	7,743,382	7,530,399
# Deals	1,632	1,632	1,561	1,632	1,632	1,561	1,632	1,632	1,561
Adjusted R^2	0.14	0.14	0.14	0.19	0.19	0.18	0.19	0.19	0.19

Notes: This table shows loan-level, OLS regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 20010Q4 and 2012Q4. Default is defined as a loan being 60 or 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Second lien mortgages are excluded from the sample.

Table A.9: Ex-Post Default Rates for Loans in GSE and Non-GSE Pools: Alternative Horizons

Horizon	24 Months			36 Months		
GSE (d)	-0.016	-0.006	0.008	-0.016	-0.005	0.012
	(6.89)	(3.26)	(1.77)	(8.50)	(2.42)	(3.27)
GSE*Low Doc		-0.027	-0.028		-0.031	-0.031
		(6.78)	(7.08)		(9.39)	(9.41)
GSE*"GSE Deal Fraction"			-0.021			-0.027
			(2.83)			(4.46)
Deal F.E. ?	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y
# Loans	10,464,165	10,464,165	10,156,202	10,464,165	10,464,165	10,156,202
# Deals	1,809	1,809	1,724	1,809	1,809	1,724
Adjusted R^2	0.16	0.16	0.16	0.20	0.20	0.20

Notes: This table shows loan-level, OLS regressions where the dependent variable is the actual default rate of loans backing subprime PLS issued between 2003 and 2007 calculated over two different horizons relative to the month of security issuance: 24 months and 36 months. Default is defined as a loan being at least 60 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

A.4 Additional Analysis

In the first column of Table A.10 we display output from our main specification (Table 5) for (virtually) all covariates, in order to provide the reader with an idea of the quantitative magnitudes of the estimates associated with the control variables in our regressions. In the second column, we display a similar set of estimates from the regression specification that includes controls for the fraction of the zip code (in which the loan was originated) that lies in census tracts that are eligible for the underserved area affordable housing goal (UAG).[9] We specify the variable has a set of indicators that correspond to each decile. For example, the first variable is an indicator for whether the UAG zip code fraction is between 0 and 0.1, the second is an indicator for whether the UAG fraction is between 0.1 and 0.2, etc. We omit the indicator that corresponds to the highest UAG values (between 0.9 and 1). The estimation results show that loans originated in zip codes with higher UAG fractions are more likely to default, *ceteris paribus*.

Table A.10: Ex-post Default Rate Linear Probability Model Coefficient Estimates

	Horizon through 2008	
	Baseline Specification	Include UAG Controls
GSE (d)	-0.007	-0.008
	(4.17)	(4.73)
Low Doc (d)	0.070	0.068
	(10.04)	(10.16)
GSE*Low Doc	-0.032	-0.032
	(9.19)	(9.04)
Owner Occupant (d)	-0.047	-0.044

[9] Both regressions also include a full set of state fixed-effects, deal fixed-effects, and dummy variables that control for missing covariate values.

	(11.72)	(11.29)
Prepay Penalty (d)	0.047	0.046
	(9.47)	(9.49)
1-unit Single Family Prop. (d)	-0.004	-0.002
	(3.03)	(1.83)
Condominium (d)	-0.024	-0.019
	(10.71)	(9.00)
Balloon (d)	0.049	0.048
	(12.12)	(11.88)
# Months Seasoned	0.000	0.000
	(0.00)	(0.04)
ARM (d)	-0.003	0.031
	(0.08)	(0.88)
Interest-Only (d)	0.046	0.046
	(10.31)	(10.46)
Negatively Amortizing (d)	0.046	0.045
	(1.73)	(1.70)
First Lien (d)	0.037	0.024
	(2.54)	(1.56)
Purchase Loan (d)	0.012	0.012
	(3.23)	(3.41)
Refinance Cash-Out (d)	-0.017	-0.018
	(13.21)	(13.65)
LTV	0.002	0.002
	(8.53)	(8.28)
$70 \leq$ LTV < 80 (d)	0.023	0.024
	(5.04)	(5.11)

$80 < \text{LTV} < 90$ (d)	0.047	0.048
	(5.20)	(5.30)
$90 \leq \text{LTV} < 100$ (d)	0.074	0.075
	(6.37)	(6.49)
$\text{LTV} \geq 100$ (d)	0.130	0.131
	(8.24)	(8.32)
$\text{LTV} = 80$ (d)	0.026	0.025
	(6.78)	(6.62)
FICO	-0.001	-0.001
	(25.93)	(26.03)
$\text{FICO} < 580$	0.025	0.026
	(7.05)	(7.39)
$580 < \text{FICO} \leq 620$	0.022	0.023
	(5.90)	(6.12)
$620 < \text{FICO} \leq 660$	0.004	0.004
	(1.24)	(1.54)
$660 < \text{FICO} \leq 700$	-0.010	-0.009
	(4.32)	(4.12)
Interest Rate	0.030	0.029
	(17.04)	(17.29)
Log (Loan Balance)	0.020	0.027
	(2.27)	(2.90)
Term	0.000	0.000
	(8.73)	(8.55)
Jumbo (d)	0.023	0.025
	(4.85)	(5.37)
Unemp. Level at Origination	0.004	0.003

	(6.17)	(4.97)
Price Index Level at Origination	0.001	0.001
	(7.90)	(7.41)
Δ Unemp. through 2008	0.013	0.019
	(2.76)	(3.67)
HPA through 2008	-0.190	-0.191
	(4.30)	(4.38)
$0 \leq$ UAG Fraction < 0.10	.	-0.040
	.	(11.57)
$0.10 \leq$ UAG Fraction < 0.20	.	-0.038
	.	(11.31)
$0.20 \leq$ UAG Fraction < 0.30	.	-0.030
	.	(8.62)
$0.30 \leq$ UAG Fraction < 0.40	.	-0.029
	.	(10.60)
$0.40 \leq$ UAG Fraction < 0.50	.	-0.023
	.	(7.59)
$0.50 \leq$ UAG Fraction < 0.60	.	-0.026
	.	(9.04)
$0.60 \leq$ UAG Fraction < 0.70	.	-0.017
	.	(11.77)
$0.70 \leq$ UAG Fraction < 0.80	.	-0.014
	.	(7.04)
$0.80 \leq$ UAG Fraction < 0.90	.	-0.008
	.	(7.06)
Deal F.E. ?	Y	Y
State F.E. ?	Y	Y

# Loans	10,464,165	9,823,431
# Deals	1,809	1,809
Adjusted R^2	0.16	0.16