

Neglected Risks in the Communication of the Mortgage-Backed Securities Offering Process

Harold H. Zhang Feng Zhao Xiaofei Zhao*

November 25, 2019

Abstract

Our study provides direct evidence on risks that are embedded in the written communications of security issuance but neglected by the initial investors in a fragile security market with significant financial innovation. Using a merged sample of institutional investors' initial holdings and the entire issuance written communications of non-agency RMBS deals, we find that mutual funds are more attentive to the risk-related information content of issuance written communications than insurance companies, and the latter are only attentive to deal characteristics. Consequently, insurance companies suffer larger losses and pay higher prices than mutual funds on their RMBS initial holdings.

Keywords: Neglected Risks, Institutional Investors, Written Communications, Uncertain Text

*We thank Dion Bongaerts, Phelim Boyle, Zhonglan Dai, Nathan Dong, Mark Flannery, Mathieu Fournier, Umit Gurun, Kathleen Weiss Hanley, Wei Jiang, Benjamin Keys, Jun Li, Tim Loughran, Wenlan Qian, Duane Seppi, Chester Spatt, Dragon Tang, Boris Vallee, Han Xia, Hongjun Yan, Adela Zhang, Haoxiang Zhu, the seminar participants at the Second Annual Conference on Financial Market Regulation co-hosted by the SEC and University of Maryland, the China International Conference in Finance, the Financial Management Association Meetings, the Conference on Financial Economics and Accounting at the University of Toronto, the Fixed Income and Financial Institutions Conference at the University of South Carolina, China International Forum on Finance and Policy, Erasmus University, and University of Waterloo for helpful comments on various versions of the paper. Harold H. Zhang and Feng Zhao are from the Naveen Jindal School of Management, University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas, 75080, and can be reached at harold.zhang@utdallas.edu, feng.zhao@utdallas.edu. Xiaofei Zhao is from the McDonough School of Business, Georgetown University, Washington, DC 20057, and can be reached at xiaofei.zhao@georgetown.edu.

1. Introduction

Recent studies suggest that the combination of significant financial innovation and investor neglect of risks associated with such innovation could lead to a fragile market for new financial securities (e.g., Gennaioli, Shleifer, and Vishny (2012), Cheng, Raina, and Xiong (2014), Gennaioli, Shleifer, and Vishny (2015), and Bordalo, Gennaioli, and Shleifer (2018)). The subprime residential mortgage-backed securities (RMBS) market is a representative example of significant financial innovation that culminated in the 2008 financial crisis. Anecdotal evidence suggests that investors in the RMBS market who bought these purportedly “safe” securities offering higher yields than traditional safe assets were under the impression that housing prices would continue to rise and thereby neglected the risk of surging defaults. Empirically, however, it is difficult to ascertain the risks neglected by investors because it requires the observation of the information set that investors had access to prior to purchasing the securities. We overcome this difficulty by investigating institutional investors’ initial holdings of RMBS and their relation to the information content of written communications in the RMBS offering process. This is accomplished by taking advantage of a novel merged sample of eMAXX dataset on institutional investors’ initial holdings and the entire issuance written communications of RMBS between the RMBS issuers and the investors. The latter is made publicly available on the Securities Exchange Commission Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system under the regulatory change of the Security Offering Reform (SOR) in December 2005.

Communication of information in the security offering process is critical for investors. Written communications in the form of various filings with the SEC play a particularly important role in assessing the valuation and the risks of securities to be offered. The documents used in the securities offering process thus provide crucial information about the securities characteristics and facilitate investors’ decisions on investing in these securities.¹

¹For instance, Hanley and Hoberg (2010) document how greater informative content in the equity IPO prospectus serves as a proxy for premarket due diligence and results in more accurate offer prices. Related

For mortgage-backed securities, widely known loan characteristics such as loan-to-value ratio, debt-to-income ratio, among others, are easy to quantify and are standard inputs in most models for evaluating mortgage loan defaults. By contrast, risks embedded in textual content are more difficult to ascertain and quantify and thus more likely neglected by investors. The issuance communication provides an ideal laboratory to explore the sources of neglected risks at ground zero of the recent financial crisis. It allows us to answer two key questions: (1) Do investors neglect the risks embedded in written communication content? (2) Which groups of investors tend to neglect such risks and the losses they suffer as a result of neglecting those risks? Research on the nature of neglected risks, who neglected risks and the losses they suffered has been very scarce on the participating institutional investors in RMBS. Our study aims to provide direct evidence on the investors' neglected risks and its effect on their non-agency RMBS investment, thus fill this void in the literature.

Gennaioli, Shleifer and Vishny (2012) describe the central feature of the 2008–2009 financial crisis as “the neglect of potentially huge defaults in the housing bubble and of the sensitivity of AAA-rated securities to these defaults”. The neglect of certain improbable risks is associated with “local thinking” phenomena with significant financial innovation. These risks are often stated in the written communications between the issuers and investors. For example, in the Risk Factor descriptions of the First Franklin Mortgage Loan Trust 2006-FF4 prospectus supplement in Appendix B, there are various statements on potential defaults in the descriptions. The second bullet point states that “*The loans were made to borrowers who are not able, or do not wish, to obtain financing from traditional sources. These mortgage loans may be considered to be of a riskier nature than by traditional sources of financing, so that the holders of the certificates may be deemed to be a greater risk of loss.*” The fourth bullet point states that “*borrowers have increasingly financed their homes with new mortgage loan products, which in many cases have allowed them to purchase homes that they might otherwise not have been able to afford ... There is little historical data with respect to these*

studies on equity IPOs are documented in Hanley and Hoberg (2012), Loughran and McDonald (2013), among others.

new mortgage loan products. Consequently, ... borrower delinquencies and defaults could exceed anticipated levels. In that event, the certificates may not perform as you anticipate.”

Under local thinking, the economic agents put lower probability for the scenarios that non-traditional sources of financing and new mortgage loan products experience defaults beyond the anticipated level. These difficult to quantify contents in written communications may pose a challenge to the MBS investors who are more conversant with processing quantifiable deal characteristics in analyzing defaults. This leads RMBS investors to neglect certain improbable risks hard to quantify and causes them to sustain losses in RMBS investment.

We collect all of the written communications in the issuance process for the non-agency RMBS deals that went through the public offering process immediately after the enactment of the Securities Offering Reform (SOR) in December 2005 and Regulation AB in January 2006. The SOR, for the first time since the Securities Act of 1933, allowed permissible written communications prior to the final prospectus in the issuance process.² This enables us to observe and collect all written communications on RMBS in their offering process. We stop at the end of 2007 because the non-agency RMBS market dropped precipitously as the financial crisis began unfolding.

Based on the nature of written communications, we classify RMBS issuance written communications into two categories: (1) final prospectus supplements; and (2) any other written communications preceding the final prospectus supplements.³ The SOR refers to the latter as a free writing prospectus (FWP). According to the SEC, a free writing prospectus is *“a written communication that constitutes an offer to sell or a solicitation of an offer to buy securities that are or will be the subject of a registration statement.”* Incorporating and

²Regulation AB, the first piece of regulation enacted specifically for asset-backed securities in January 2006, explicitly states that written communication in the ABS offering process is subject to the rules under SOR. As a result of the Securities Offering Reform (SOR), which mandated that all written communications produced prior to the final prospectus be filed with the SEC in the form of free writing prospectuses, our analysis is possible. Prior to the SOR and Regulation AB, issuers generally used term sheets for the public offering of ABS, which are not required to be filed with SEC, this denying us access to any term sheets prior to 2006.

³Most public offerings of ABS are conducted through expedited SEC registration procedures known as shelf offerings or shelf registration. The prospectus itself is general and not specific to an ABS deal. Rather, the final prospectus supplement is created specifically for each individual ABS deal.

analyzing both types of written communications provide a holistic perspective on neglected risks in the RMBS offering process and the effects on RMBS market.

Building upon the recent literature on textual analysis of SEC filings (Loughran and McDonald (2011, 2013, and 2014); Hanley and Hoberg (2010, 2012)), our methodology takes an all-inclusive approach to constructing textual measures for the entire body of written communication for each issuance.⁴ Due to the nature of fixed income securities, investors and issuers focus more on the downside risks (such as default) for RMBS (in contrast to the upside potential in the context of equity IPOs; see Hanley and Hoberg (2010, 2012)). Therefore, for the final prospectus supplements (hereafter referred to as prospectus or supplements for brevity), we focus on the usage of uncertain text from the aggregate list of uncertain, weak modal, negative word lists (Loughran and McDonald (2011)) in these documents. The intuition behind our focus on uncertain text to identify neglected risks is best explained by parsing through the text in written communications as shown in the example in Appendix B. The usage of uncertain words such as “may”, “could”, “loss” or “risk” captures the extent and intensity for the losses in potential scenarios. First, the uncertain words always appear when a separate item is discussed in the Risk Factor section, so the usage of uncertain words reflects the extent of risk factors. Second, more uncertain words are used when a particular risk is articulated or stressed, so the usage of uncertain words also reflects the intensity of risks and severity of potential losses. More important, our formal topic analysis confirms the intuition.

We first apply textual analysis to construct an uncertain text usage measure for the entire prospectus supplement as a whole. To identify the source of variation and aid interpretation of the content of uncertain text usage, we then delve into the contents of four key sections of the prospectus: the Risk Factor, the Description of Certificates, the Mortgage Pool, and the Summary, to construct a separate measure of uncertain text usage for each section. Uncertain word occurrence differs across the sections, consistent with the different purposes

⁴See Loughran and McDonald (2016) for a survey on the recent development of this literature.

served by each of the four sections. Our in-depth topic analysis suggests that uncertain text usage in the Risk Factor section is associated with underwriting standard, geographic and uninsurable risks, among others, while the uncertain text usage in the Mortgage Pool section is associated with stating the approximate average. This contrast further helps us ascertain the effect of risk-relevant information of uncertain text.

For written communications prior to the final prospectus or supplements, i.e., FWPs, their content is non-standard and the uncertain text measure is not applicable. We distinguish between these pre-issuance written communications and classify them into two major types: loan tape and textual documents. The former is an electronic file or a set of numeric files that encompasses individual loan data. The latter consists of textual documents including various updates and disclosures on the deal. Liberti and Petersen (2018) stress the distinctions between soft and hard information arguing that hard information is always numeric and easy to convey, while soft information is always communicated in text and its exact meaning depends on the context. In our setting, the loan tape is a standard and mostly numeric data that can be easily transmitted, thus, represents a form of hard information. Textual FWPs require the context in which these communications are generated and used, thus, are the main sources of soft information. Since MBS issuers are required to file FWPs with the SEC for written communications with any potential investor, local thinking may lead some investors neglecting the updated information in textual FWPs.

Our analysis reveals drastically different attentiveness of the initial holdings of mutual funds from that of insurance companies and pension funds with respect to the contents of the issuance written communications and key deal characteristics.⁵ We find that mutual funds invest less in deals with more textual FWPs and higher uncertain text usage in Risk Factor section of prospectus supplements. Furthermore, their holdings are positively related to the numerical documents measured by the size of loantape FWPs. By contrast, insurance companies are inattentive to issuance written communication contents, textual or numerical.

⁵We combine insurance company and pension fund holdings because they share similar long-term obligations. Hereafter, we simply refer to the combined group as insurance company holdings.

However, they are attentive to salient deal characteristics such as reputation of the issuer, level of borrower documentation, FICO, and loan type (adjustable rate versus fixed rate). This divergence between mutual funds and insurance companies reveals their different attitudes toward improbable risks in non-agency RMBS comprised of considerable financial innovation. It suggests that mutual funds are to some extent more attentive to the risks in the issuance written communications which are neglected by insurance companies.

Upon establishing the relation between holdings and issuance communications, we examine whether the textual content predicts subsequent loss in the aftermath of financial crisis. As argued in Gennaioli, Shleifer, and Vishny (2012), when investors neglect these risk relevant information in the issuance process, they suffer unexpected losses especially when housing prices decline. The unforeseen and dramatic defaults in the housing market caught investors by surprise, leading to the crash of the mortgage-backed securities widely held by institutional investors. It is a common practice for RMBS investors to use the known characteristics to assess the mortgage loan default risk, such as a borrower's loan-to-value ratio, debt-to-income ratio, among others. The textual content of the issuance communication is often not included in models for mortgage loan default estimation.

Controlling for an extensive list of disclosed deal characteristics (including underwriter and time fixed effects) and macro and local economic conditions that could forecast default, we find that when uncertain word usage in the Risk Factor section (*or the amount of textual FWP content*) increases by one standard deviation, the future deal cumulative net losses increase by 1.17 (*or 1.0*) percentage points. This accounts for 10% (*or 8.5%*) of the average deal loss in our sample. Further, the initial yield spreads for deals with more uncertain textual content are actually lower than that with less uncertain content controlling for everything else. The effect of uncertain textual content on subsequent deal loss is stronger for high-priced deals, consistent with our prior that more neglected risks lead to higher deal prices everything else equal. These results suggest that our textual content measures capture risk relevant information that is neglected by non-agency RMBS investors.

Our analysis on the losses at the RMBS securities level shows that securities initially held by insurance companies have significantly higher losses than those initially held by mutual funds by 2-3%, after controlling for deal and time fixed effects as well as macroeconomic conditions. Furthermore, related to our textual based measures, we show that this loss differential is larger in deals with more textual contents related to neglected risks. Our analysis on RMBS securities by different institutional investors show that the yield spread on RMBS securities held more by insurance companies is lower than that held more by mutual funds, though statistically insignificant. This suggests that insurance companies did not receive commensurate compensation for the losses, lending support to the risks in the offering written communications being neglected.

Our findings upon comprehensive analyses of written communications provide strong evidence that RMBS investors neglected risk-related information in the issuance process. They provide direct empirical evidence on recent theoretical studies regarding how investors neglected risks and the consequence of such neglect for a fragile security market with financial innovation (e.g., Gennaioli, Shleifer, and Vishny (2012), Gennaioli, Shleifer, and Vishny (2015), Bordalo, Gennaioli, and Shleifer (2018)). Our paper also contributes to the call for “*the need to expand the incentive view of the 2007-2008 housing crisis to incorporate the role for beliefs*” (Cheng, Raina, and Xiong (2014)) and provides direct evidence on how institutional investors neglected downside risks during credit expansion (e.g., Baron and Xiong (2017)). Chernenko, Hanson, and Sunderam (2016) show that inexperienced mutual fund managers invest more in the securitized assets linked to subprime mortgages in the credit boom before the 2007-2008 financial crisis. Celerier and Vallee (2017) use European retail structured products and investigate how banks design these products to entice investors to focus on the best-case yield—the headline rate—that is nested in a complex payoff formula. By increasing the headline rates and complexity of certain financial products, banks made a greater profit from selling riskier products, a finding consistent with the view that investors seeking high yields possibly neglected the downside risks on more complex

and riskier products.

In terms of research methodology, our study is related to Ghent, Torous, and Valkanov (2019) which explores deal structure complexity for non-agency RMBS and reports evidence of issuers using complexity to obfuscate higher default risks in the higher rated tranches. While Ghent, Torous, and Valkanov (2019) focus on the decision process on the part of issuers of non-agency RMBS, our study focuses on the decisions made by investors. As stressed in Merrill, Nadauld, and Strahan (2019), "...without demand from final investors, it is difficult to imagine that the supply of private-label securities would have increased so dramatically." While Ghent, Torous, and Valkanov (2019) also find that securities from more complex deals have lower yield spreads, they do not analyze the investor holdings. Another interesting difference in empirical implementation is that some of their complexity measures are constructed from the mortgage pool and certificates sections in the prospectus supplements, our main findings are based on the measures constructed from the Risk Factor section.

Thus, our study offers a novel perspective and provides evidence on the investment decisions of various institutional investors involved in the RMBS securitization chain relative to recent papers documenting the role of incentives of various participants,⁶ particularly that various participants in the securitization supply chain might have engaged in misrepresentation and fraud (e.g., Ben-David (2011), Jiang, Nelson, and Vytlačil (2014a), Piskorski, Seru, and Witkin (2014), Griffin and Maturana (2014), Garmaise (2015)). To the best of our knowledge, our study is one of the first to conduct comprehensive textual analysis of the entire all-inclusive written communications in fixed income security offerings. Our findings on

⁶Notable examples include Mian and Sufi (2009), Barlevy and Fisher (2011), Haughwout, Lee, Tracy, and Van der Klaauw (2011), Keys, Seru, and Vig (2012), Chinco and Mayer (2012), Agarwal, Amromin, Ben-David, Chomsisengphet, and Evanoff (2014) on weakened standards; Ben-David (2011), Jiang, Nelson, and Vytlačil (2014b), Piskorski, Seru, and Witkin (2014), Griffin and Maturana (2014), Garmaise (2015), of misrepresentations and fraud; Keys, Mukherjee, Seru, and Vig (2010), Purnanandam (2011), Nadauld and Sherlund (2013), Rajan, Seru, and Vig (2015), of a market fueled by poor ratings models and the rapid expansion of non-agency securitization markets; Demiroglu and James (2012), Dai, Zhang, and Zhao (2014), of affiliation-related agency issues; Tzioumis and Gee (2013), Agarwal and Ben-David (2014), Agarwal and Ben-David (2014) of loan officer pay structure and the use of other credit derivatives; among others.

neglected risk-related information uncovered in textual communications and uncertain text usage in Risk Factor section for RMBS represent a significant contribution to recent studies in the corporate disclosure setting. For example, Cohen, Malloy, and Nguyen (2016) analyze the price reactions to changes in corporate quarterly and annual SEC filings. They find that the changes in Risk Factor section are most relevant for predicting future abnormal returns and that investors pay more attention to numerical items than textual ones. Similarly, Zhou (2017) finds that higher usage of numbers (relative to words) in corporate conference calls reflects positive information and induces positive price reactions. Hanley and Hoberg (2019) propose a new approach for detecting emerging risks in the financial sector using data from banks' 10-K filings.

2. RMBS Issuance Written Communications

An issuance's written communications for RMBS can be classified into two categories: (1) final prospectus supplements; and (2) pre-issuance written communications prior to the final prospectus supplements (FWPs). According to Securities Offering Reform (SOR) enacted in December 2005, FWPs can be used by any eligible issuer or offering participant after a registration statement has been filed. Pre-issuance written communications in the form of FWPs are therefore used frequently in the offering process of publicly issued asset-backed securities (ABS) including RMBS.⁷ Because ABS issuances rely on shelf-registration, after the FWPs, the last set of written communications is the final prospectus supplement, equiv-

⁷In the mid 1990s, SEC no-action letters have permitted issuers/underwriters to use term sheets for public offering of asset-backed securities. Term sheets are required to be filed with the SEC on FWPs after the SOR. There are three types of term sheets: "structural term sheets," which provide factual information regarding the financial terms of an asset-backed securities transaction, including the securities to be offered and the structure of the offering; "collateral term sheets," which provide data about the assets underlying the offering; and "series term sheets," which combine aspects of a structural term sheet with a collateral term sheet. In the SEC letters, "computational material" is another term used to describe both structural term sheets and collateral term sheets. Decrement tables, which contain maturity projections such as those commonly found in mortgage-backed and complex owner trust transactions, would appear to be series term sheets because they describe proposed structures for the offered securities but are based on data concerning the collateral pool.

alent to the final prospectus in an equity IPO.⁸ In this section, we first describe the sample construction and then conduct a textual analysis of the entire written communications in the RMBS offering process to quantify the contents.

2.1 Data Description

Our data on RMBS issuance written communications come from the SEC EDGAR system. We start with a list of RMBS deals in Bloomberg and link deals to SEC filings by matching deal names. For each deal, we compare its name with the ABS names in EDGAR.⁹ If multiple matches are found, we then read the final prospectus supplements filed with the SEC to pin down the correct match.

Our sample consists of non-agency RMBS deals issued in 2006 and 2007. We focus on this time period because SOR was enacted in December 2005 which clarified the content, permissible use, and potential liability of written communications in ABS public offerings. This allows us to collect these offering documents systematically. The drastic housing market decline resulted in a precipitous drop in the number of securitized mortgage deals post 2007, delineating the endpoint of our sample. Figure 1 shows that the number of deals had already reached a low level starting from August 2007 (below 40), and by December 2007, the number of deals had dropped to below 10.

Figure 1 about here

2.2 Textual Contents of Final Prospectus Supplements

We extract the final prospectus supplements for our sample deals from the SEC filings. To analyze the textual content, we first parse the individual documents following the procedure in Loughran and McDonald (2011, 2013). We add one additional step in the parsing

⁸Most public offerings of ABS are conducted through expedited SEC registration procedures known as shelf offerings or shelf registration. The ABS prospectus itself is general and not specific to an ABS deal. Instead, the final prospectus supplement is created specifically for each individual ABS deal.

⁹SEC has a directory for all types of ABS filers (over 10,000 with SIC code 6189).

procedure to separate the final prospectus supplement from the general shelf-registration statement. This is because the final prospectus generally includes both and only the supplement is prepared for a specific deal. Hence, all the textual measures are based on the supplement. The style and format of the final prospectuses vary across deals and some deals are submitted to SEC in XML format while others are in txt format. We also extract the table of contents from each final prospectus supplement in order to obtain the textual content for individual sections. We exclude deals for which the key sections are unidentifiable due to unorganized formatting.

The final prospectus supplements contain a more standardized format and their contents are predominantly textual in nature. Due the fixed income feature of RMBS, investors are generally concerned with the downside risk on these securities in contrast to the upside potential in equity IPOs. We thus focus on the usage of uncertain words in the final prospectus supplements taking into consideration of the context in which the uncertain words are used. Specifically, we define “uncertain words” based on the aggregate list of *uncertain*, *weak modal*, *negative* word lists (Loughran and McDonald (2011)). We first measure the usage of uncertain words for the entire document of final prospectus supplements. Among the 2600 Loughran and McDonald (2011) aggregate uncertain words, we find that two words: *may* and *approximately* alone account for 24.7% of the uncertain words used in the RMBS final prospectus supplements. The word *may* accounts for 18.6% and the word *approximately* accounts for 6.1%.¹⁰ To investigate the relation between neglected risks and the uncertain word usage, we further explore the uncertain word usage for four key individual sections: the Risk Factor section, the Description of Certificate section, the Mortgage Pool section, and the Summary section.¹¹

Table 1 Panel A reports the summary statistics of the usage of uncertain words in the whole document of the final prospectus supplement (*UncProp*) and its four key sections

¹⁰We carefully exclude the cases in which “May” is used to refer to the calendar month. To do so, we identify and exclude the instances for “20XX May” and “May 20XX”.

¹¹Appendix A provides more details on how we extract these individual sections from the final prospectus supplements.

(*UncRiskFactor* for the Risk Factor section; *UncCertificate* for the Description of Certificates section; *UncMortgagePool* for the Mortgage Pool section; and *UncSummary* for the Summary section), respectively. Panel B reports the correlation coefficients among these measures. On average, a prospectus supplement uses 1267 uncertain words with a standard deviation of 493 uncertain words. The four key sections account for slightly above 60% of the uncertain word usage in the final prospectus supplement. In the four sections, there is an average of 277 uncertain words in the Risk Factor section and 269 in the Description of Certificates section, followed by 124 in the Mortgage Pool section and 102 in the Summary section. Another noteworthy observation is that the correlation between uncertain word usage in the overall final prospectus supplement and individual sections is the highest for the Risk Factor section (0.62) and the lowest for the Summary section (0.12).

Panel C presents the frequency of the top 20 uncertain words within each section. Uncertain word occurrence differs across the sections, consistent with the different purpose served by each of the four sections. For example, in the Risk Factor section, the usage of “MAY” (used 77888 times or 25%) dominates, followed by “LOSS” (used 32569 times or 10%), “RISK” (used 17515 times or 6%), and “COULD” (used 12266 times or 4%).¹² These words are used primarily to signal the contingencies of various potential risks without having to provide much numeric detail in the Risk Factor section. Appendix B provides an actual example of uncertain word usage in the Risk Factor section of the First Franklin Mortgage Loan Trust 2006-FF4.¹³ In sharp contrast, the top 4 words used in the Mortgage Pool section are “APPROXIMATELY” (used 36950 times or 27%), “MAY” (used 24568 times or 18%), “DELINQUENCY” (used 8470 times or 6%), and “CLOSING” (6377 times or 5%). Among these 4 words, “APPROXIMATELY” is commonly used for numerical approximation/rounding when discussing numeric information while “CLOSING” and “DELINQUENCY” are often

¹²The percentage usage is calculated based on the total number of uncertain words in each individual section and in this case based on Risk Factor section.

¹³In the Online Appendix, we also apply topic analysis to the content of risk factor sections and find that in general the risks are clustered in various categories such as underwriting standard risk, geographic and uninsurable risk, credit score risk, etc.

used when discussing the closing date or the percentage of loans in delinquency. Therefore, the uncertain words in the Mortgage Pool section are much less likely to be used for the purposes of describing material risk factors. The meaning behind the usage of uncertain words in the Description of Certificates section is somewhat in-between the motivations outlined for the Risk Factor section and the Mortgage Pool section. As the main purpose of this section is to describe the certificates themselves, it covers the description of cash flow distributions including how “UNPAID” interests are distributed among the certificates or reference to defaulted swap “TERMINATION” payment, as well as outlining some contingent scenarios through the use of “MAY”. Lastly, the Summary section is a mixture of various sections and serves to summarize.

Table 1 about here

2.3 Written Communication Contents in FWPs

Given that FWPs are a new form of SEC filing, in Appendix C, we present a list of FWPs on one specific mortgage deal—First Franlin Mortgage Loan Trust 2006-FF4. FWPs provide data such as basic statistics for collateral assets, structural and collateral term sheets, detailed loan level data, and corresponding updates. The document format of FWPs is typically non-standard. We thus apply a different approach to examine the content of FWPs. Specifically, we analyze the textual content of FWPs and classify these documents as loan data FWPs or textual FWPs. A loan data FWP covers loan-level information, which is commonly referred to as loantape. A loantape FWP provides information on an individual loan such as its FICO score, LTV ratio, specific loan terms, and the like. The primary characteristic of a loantape FWP is its quantitative nature. To provide a measure for the quantitative nature of loantape FWPs, we break down its content into alphabet and numeric tokens.¹⁴

¹⁴A token is a structure representing a lexeme that explicitly indicates its categorization for the purpose of parsing. To obtain tokens from a FWP document, we first parse the document to strip off various html

Panel A of Table 2 shows the content characteristics of loantape FWPs. It is clear that loantape FWPs are dominated by numeric tokens. The average number of numeric tokens in a loantape FWP is over 820 thousand whereas the average number of alphabet tokens is about 380 thousand. Moreover, those alphabet tokens found in loantape FWPs are likely to index quantitative information since some dummy variables may be presented in alphabet form. For instance, for the documentation of an individual loan, the value could be “Full” or “Low.” For the occupancy of the property for each loan, the value could be “Owner Occupied” or “Investor”.¹⁵ The quantitative nature of loantape FWPs can also be seen from a low average alphabet to numeric token ratio of 0.5.

In contrast, a textual FWP is dominated by alphabet-tokens and has a descriptive nature. Panel B of Table 2 shows that the average number of numeric tokens for textual FWPs is 23 thousand and the number of alphabet tokens is around 301 thousand. The average alphabet to number ratio is 51 for textual FWPs, in sharp contrast to an average alphabet to number ratio of 0.5 for loantape FWPs. Moreover, examining the interquartile difference (Q3/Q1) for the alphabet-number ratio, we observe that the ratio of the 75th percentile to the 25th percentile is close to 20 for textual FWPs. The interquartile difference measured by Q3 to Q1 ratio is only about 2 for loantape FWPs. This reflects another important difference between these two types of FWPs.¹⁶

We create two measures of FWP content for each deal, one for loantape FWP and another for textual FWP. For loantape FWP, we sum up the number of alphabet tokens and the number of numeric tokens and use the natural logarithm of this sum to measure the content of loantape FWP. We label this “Log(Loantape FWP)” in our tables. Similarly, we label the measure of textual FWP content as “Log(Textual FWP)”. Table 2 Panel C reports the summary statistics of the loantape and textual FWPs. The Log(Loantape FWP) has

tags, then break the remaining document into tokens by any white space. We define a token as a numeric token if the first character is a number (0 to 9). Similarly, we define a token as an alphabet token if the first character is a letter (a to z or A to Z).

¹⁵Certainly different deals may use different words to represent these values. The example here is based on the deal in Appendix C.

¹⁶Loantape FWPs are also much more standardized in format while textual FWPs are much less structured.

an average of 4.6 with a standard deviation of 6.6, while for textual FWPs, the average is higher at 7.9 with a slightly lower standard deviation at 6.3.

Table 2 about here

3. Institutional Investors and Written Communication Content

We now investigate institutional investors' initial holdings of RMBS upon issuance with respect to the information in the content of RMBS issuance written communications. This analysis is crucial for at least two reasons. First, it advances research on understanding institutional investors' decisions in RMBS investment leading up to the financial crisis. Second, analyzing how different institutional investors' initial holdings corroborate the issuance written communications provides direct evidence on the information that these investors neglected and paid attention to. While this analysis is crucial, it is nonetheless greatly limited by the availability of data on the participating institutional investors that initially invested in RMBS.

To fill this critical void in the existing literature, we obtain institutional investors' initial holdings of RMBS from Thomson Reuter's eMAXX solution. eMAXX data mainly cover the holdings of mutual funds, insurance companies and pension funds. We rely on eMAXX data to construct the ownership of these RMBS for mutual funds and combined insurance company and pension fund holdings and evaluate the attentiveness of the holdings of these two groups of institutional investors to the issuance written communication content and to other commonly used deal characteristics. To accomplish this objective, we first identify the RMBS deals that are present in both issuance written communications sample and in the eMAXX data. We then calculate the percentage of the initial holdings by mutual funds, reported within one year of the issuance dates (to account for the potential lag in reporting), relative to the issuance amount and denote it as mutual fund initial holdings (%). Similarly

we calculate the percentage of combined insurance companies' and pension funds' initial holdings and denote it as insurance company initial holdings (%). To make the holdings comparable between mutual funds and insurance companies, we only keep securities with an initial credit rating of AAA in the analysis.

In addition, we collect information on deal characteristics which are widely used to assess default risks in existing literature. These include the original collateral balance, the number of tranches, the percentage of loans with low documentation (including loans with no documentation), the average FICO score, the loan-to-value (LTV) ratio, the percentage of adjustable rate mortgages, negative amortization, purchase loans, single family property, owner-occupied house, and second lien. We also collect regional and macroeconomic variables, the credit spread, and 10-year treasury yield to use in our analyses. The details of the variable construction can be found in Appendix D and the summary statistics are presented in the Online Appendix to save space.

Table 3 presents the summary statistics on mutual fund and insurance company initial holdings of RMBS in our combined sample. On average, mutual funds hold about 5.3% of RMBS initial issuance. Insurance companies and pension funds combined on average hold more of the RMBS with an average of 11.6% of initial issuance. Both holdings exhibit noticeable cross-sectional variations with a standard deviation of 10.6% for mutual fund holdings and 13.3% for insurance company holdings. In the meantime, mutual fund initial holdings can be as high as 20.4% and the insurance company initial holdings can reach almost 40% at the top 95 percentile, respectively. These statistics indicate that mutual fund and insurance company initial holdings play an important role in RMBS issuance.

Table 3 about here

3.1 Mutual Fund Holdings and Written Communication Content

To examine whether investors' investment decision in RMBS takes into account the written communication contents in its offering process, we conduct regression analysis on the deter-

minants of investors' initial holdings in RMBS. For our analysis on the mutual fund initial holdings, we choose the dependent variable to be the percentage of all the initial holdings by mutual funds relative to the issuance amount. The explanatory variables of our interest are the RMBS offering written communication contents. We control for all widely used deal characteristics that could influence investors' investment decision in RMBS.

Table 4 about here

Table 4 reports the results of analyzing whether mutual fund initial holdings of RMBS are related to orthogonalized version of FWP content and the uncertain word usage of the four key sections of the final prospectus supplements. To generate the orthogonalized version of our variables from written communications, we regress each of the variables on textual FWPs, loantape FWPs, uncertain word usage in the entire prospectus supplements and each of the four main sections on regressors including deal characteristics, and time and lead underwriter fixed effects. We keep the residuals from these regressions and use these orthogonalized variables in our analysis. Our results show that mutual funds tend to hold less of the RMBS deals that contain more textual FWP but hold more deals with greater loantape FWP. Our estimated coefficients indicate that for a one percent increase in the textual FWPs the mutual fund initial holdings would decrease by 0.75 percent, while a one percent increase in Loantape FWPs is associated with an increase of 0.68 percent of mutual fund initial holding.

Mutual funds also tend to hold fewer deals with more uncertain words in the final prospectus supplement. This effect is however concentrated in the Risk Factor Section of the prospectus supplement. Our estimated coefficient suggests that a one standard deviation increase in the uncertain word usage in the Risk Factor Section is associated with a 2.05% increase in mutual fund initial holdings. Overall, our results indicate that mutual funds were more cautious of mortgage deals with greater textual FWP content and higher uncertain word usage in the Risk Factor section, and consequently invested less in such RMBS deals. On the other hand, mutual fund holdings are not sensitive to other deal characteristics except for negative

amortization. We focus on securities with initial credit rating of AAA. These securities can be created from collaterals of varying characteristics through credit enhancements such as over-collateralization and subordination. Our results show that mutual fund managers select AAA rated securities across the spectrum of collateral characteristics such as FICO, LTV, documentation level, among others, in order to achieve their investment objectives.

3.2 Insurance Company Holdings and Written Communication Content

For comparison, we conduct the same analysis for insurance company initial holdings. The dependent variable is similarly defined as the percentage of all the initial holdings by insurance companies and pension funds relative to the issuance amount. The explanatory variables of interest and the control variables are the same as those for the mutual fund initial holding analysis. Table 5 presents the regression analysis results on whether insurance company RMBS initial holdings are related to the orthogonalized version of the FWP content and the uncertain word usage in the four key sections of the final prospectus supplements.

Table 5 about here

In contrast to the findings on mutual fund initial holdings, insurance company initial holdings do not seem to be significantly related to either uncertain word usage in final prospectus supplements or FWP content. Interestingly, insurance companies appear to pay more attention to the headline numbers in deal characteristics. For example, insurance companies hold less of the deals with low FICO scores and high percentage of loans with low documentation. They also hold more deals that have highly-reputable underwriters.¹⁷ Furthermore, insurance companies hold more securities from smaller deals. Since larger deals usually offer better diversification in the mortgage pool, uncertainty in collateral quality can more adversely affect the senior tranches in smaller deals.

¹⁷Griffin, Lowery, and Saretto (2014) show that more reputable underwriters do not produce better securities in this market.

The difference between mutual funds and insurance companies in their attentiveness to the written communication contents in the offering process provides insight into the reasons on why institutional investors neglect risks in non-agency RMBS investment. Previous studies have shown insurance companies can behave differently from other institutional investors due to regulatory constraints on their risk-taking behavior. For instance, Merrill, Nadauld, and Strahan (2019) show that insurance companies load up on structured securities leading up to the financial crisis due to regulatory capital requirement. Becker and Ivashina (2015) find that insurance companies reach for yields in corporate bond market. Our findings suggest that insurance companies neglect the risks that they deemed improbable and loaded up the AAA rated non-agency RMBS from deals with higher level of loss-relevant information which are conveyed in the offering written communications of these securities.

4. Written Communications and Neglected Risks

In this section, we demonstrate that the offering written communication contents on unattended risks are strongly related to future losses in the aftermath of 2008 financial crisis. Gennaioli, Shleifer, and Vishny (2012) argue that the unforeseen and dramatic defaults in the housing market caught investors by surprise, leading to the crash of the mortgage-backed securities widely held by institutional investors. We expect that this effect is stronger for deals with more communication contents related to neglected risks. Therefore, this analysis establish the direct link between neglected risks and deal losses by showing the risk related textual information in issuance written communications predicting the cross sectional variation in deal performance. We measure deal cumulative net loss rate as the sum of all losses of principal suffered prior to December 2010, divided by the total original balance of all mortgages.

4.1 Written Communication contents and Deal Performance

Our goal is to extract information from written communications that are not spanned by deal characteristics, macroeconomic conditions, underwriter fixed effects, and deal pricing. If the unspanned components from written communications can predict deal losses, these information could be neglected by the investors. Our intuition is that if investors were able to fully account for the risk-related information in issuance written communications into pricing, then if we control for the variables related to deal pricing and other control variables, the written communications should no longer have predictive power for deal losses. On the other hand, if investors neglected certain risk-related information in the issuance written communications, then the risk-related written communication contents would not be fully priced and they should have an incremental effect on deal losses even after controlling for deal pricing variables such as deal initial yield spread and credit enhancement.

As discussed in Section 3, we regress each of the variables on textual FWPs, loantape FWPs, uncertain word usage in the entire prospectus supplements and each of the main sections, on deal characteristics, and time and lead underwriter fixed effects, and keep the residuals from these regressions and refer them as the orthogonalized versions of the original variables. We have also performed the orthogonalization of the variables for written communications by incorporating deal pricing variables as additional regressors, and find the two orthogonalized versions are highly correlated with correlation coefficient above 0.99. This provides an initial indication that the information from written communications may not have affected pricing significantly.

We next regress deal losses on the orthogonalized variables from issuance written communications using the following specification:

$$\begin{aligned} \text{Cumulative net loss} = & \alpha + \beta_0 \times \text{Uncertain Word Usage in Final Prospectus Supplement} \\ & + \beta_1 \times \text{FWP content} + \beta_2 \times \text{Yield spread} \end{aligned}$$

$$+\beta_3 \times \text{Credit enhancement} + \text{Deal charact.} + \text{Fixed effects.}$$

Table 6 reports the main findings. Our estimation results show that, controlling for mortgage deal pricing, characteristics, regional and macroeconomic conditions, orthogonalized textual FWPs have a significantly positive effect on the cumulative net loss. For example, the estimated coefficient in column (1) of the table suggests that when the aggregate content of textual FWPs increases by one standard deviation (the standard deviation of textual content variables are normalized to one unit), the cumulative net loss is higher by 1.00%.¹⁸ In the meantime, the loantape FWP has a negative and yet insignificant relation to a deal's cumulative net loss, suggesting that the amount of disclosure in loantape FWPs provide limited additional information on cumulative net loss beyond the reported deal characteristics. Considering that textual FWPs are the more likely the source of soft information that is difficult to interpret and quantify, the stark contrast between the loantape and textual FWPs suggests that risk-related information in the pre-issuance written communication contained in textual FWPs is more likely to be neglected by investors.

Table 6 about here

Turning to the textual contents of final prospectus supplements, we find that the uncertain word usage orthogonalized to other deal characteristics is positively related to a deal's cumulative net loss. This effect is above and beyond that associated with the pre-issuance FWP contents. Upon analyzing further the uncertain word usage in the individual sections of the supplements, we find that uncertain word usage in the Risk Factor section has the strongest predictive power for deal cumulative net losses, consistent with our expectation that risk-related information in the issuance written communications is neglected. For example, in column (2) of this table, the coefficient associated with UncRiskFactor is 1.18 (t -stat=4.84). Economically, this suggests that a one standard deviation increase in

¹⁸In the Online Appendix, we show that the results of including FWP content but excluding the final prospectus supplement content in the analysis are very similar.

UncRiskFactor is associated with 1.18 percentage points higher deal cumulative net losses, representing close to 10% of the average deal losses.

When we put UncRiskFactor, UncCertificate, UncMortgagePool, and UncSummary together in the same regression (column (6) of the table), the coefficient on UncRiskFactor remains the strongest at 1.03 (t -stat of 4.90), whereas the coefficients for the other three sections are either insignificant or substantially attenuated.¹⁹ We also note the sharp contrast between UncRiskFactor and UncMortgagePool. In the latter the uncertain words are more often used to indicate the estimated averages of mortgage pool statistics, thus less related to deal risks. This finding clearly demonstrates that risk-related textual contents can be very informative about future deal losses. However, it is precisely this type of textual information that is also prone to investor neglect.

The associations between other control variables and the cumulative net losses are generally consistent with our intuition. For example, we find that the number of tranches increases the cumulative net losses. To the extent that the number of mortgage deal tranches serves as a proxy for product complexity of securitization, this finding suggests that the complexity of structured products with significant financial innovation is adversely associated with their performance, consistent with recent empirical evidence in the RMBS market (Ghent, Torous, and Valkanov (2019)) and the retail structured financial product market (Celerier and Vallee (2017)). High FICO score is associated with a lower cumulative net loss while high percentage of low documentation and LTV increase the cumulative net loss. The former reflects higher borrower credit worthiness and thus better loan quality. The latter indicates higher risk because of higher borrower leverage.

A high percentage of loans with second lien also increases cumulative net losses. This reflects both the higher risks of these loans and possible under-reporting of second lien, as documented in some recent studies (see Griffin and Maturana, 2013). On the other

¹⁹Note that in our analysis we measure deal performance using the cumulative net loss as of the end of 2010. Our results are robust to using different cutoff dates. For example, if we use the cumulative net loss as of the end of 2012, we obtain very similar results. For brevity we do not tabulate these results, although they are available upon request.

hand, high underwriter reputation had no significant effect on deal cumulative net losses, possibly due to controlling for the lead underwriter fixed effect on deal performance. In all regression specifications, house price change has a negative effect on the cumulative net loss. This suggests that housing price decline contributed to the large losses of RMBS. It is noteworthy that our measure of housing price index is constructed based on the geographical composition for each deal. This helps to alleviate the concern of possible correlation between communication contents related and subsequent housing price changes.

Overall, our findings suggest that both the pre-issuance textual communications and uncertain word usage in the final prospectus supplements, particularly in the Risk Factor section, are significantly related to the cumulative net losses of these securities. This is consistent with the prediction from Gennaioli, Shleifer, and Vishny (2012) that investors neglect improbable risks and believe the securities involving financial innovations to be safe, and subsequently suffer substantial losses.

4.2 Written Communication Contents and Deal Pricing

To lend further support to the neglected information from the offering written communications to be related to unattended risks, we investigate whether or the extent such information is reflected in the initial deal pricing. The actual prices that investors paid are difficult to obtain. Therefore, we follow existing studies (e.g., He, Qian, and Strahan (2012), Piskorski, Seru, and Witkin (2014)) that approximate the initial deal pricing using the deal initial yield spread. We measure the initial yield spread as the average yield of all securities issued by the trust weighted by the face value of the securities minus the yield on the 10-year Treasury bond.

To facilitate our analyses below, we construct two textual risk indices from the orthogonalized variables for the written communications. Specifically, using the estimates for these variables in predicting deal losses, we extract the loss-related components that are not spanned in other control variables. The first textual risk index is based on column (1) in Table 6 that

includes orthogonalized uncertain word usage in the final prospectus supplement, textual and loantape FWP, i.e., Textual Risk Index 1 = $0.89 \cdot \text{UncProp} + 1.00 \cdot \text{Log}(\text{Textual FWP}) - 0.31 \cdot \text{Log}(\text{Loantape FWP})$. The second textual risk index is based on column (6) Table 6, i.e., Textual Risk Index 2 = $1.03 \cdot \text{UncRiskFactor} + 0.52 \cdot \text{UncCertificate} + 0.91 \cdot \text{Log}(\text{Textual FWP}) - 0.49 \cdot \text{Log}(\text{Loantape FWP})$. These indices are highly correlated with the deal losses by construction. A higher value from the indices represents higher losses due to the unspanned information from written communications. The correlation coefficient between the two indices is 0.77.

In our first analysis, we compare the initial yield spreads between deals with high versus low index values assuming that the deals are otherwise comparable. The natural choice of methodology for this analysis is to match observations with the set of covariates including deal characteristics, regional and macroeconomic variables, deal lead underwriters and issuance year, and then comparing the deal pricing for the treatment (high index value) and control (low index value) samples. Matching method can alleviate the concern that any potential difference in pricing may reflect the differences in the characteristics of mortgage deals correlated with different levels of written communications. In other words, the uncovered relation between deal pricing and written communications may be due to misspecification if mortgage deals with higher yields have disproportionately higher level of written communications.

Table 7 about here

Table 7 reports the results of matching analysis on deal initial yield spreads. We control for deal, regional and macroeconomic variables, such as the original collateral balance, high reputation, number of tranches, FICO, low documentation, LTV, adjustable rate mortgage, negative amortization, purchase loans, single family owner occupied, second lien, housing price change, credit spread and 10 year treasury. We also require match exactly on deal lead underwriters and issuance year. Using the nearest-neighbor matching with replacement to implement one-to-one matching of treatment and control samples, we compute the z -statistic

using the analytical estimator of the asymptotic variance of matching estimators proposed by Abadie and Imbens (2006, 2008) to make statistical inference on the significance of the difference in deal initial yield spreads. The Abadie-Imbens robust z -statistics are reported in the parentheses below the estimates of the difference.

We use different cutoff values of our constructed indices to separate high versus low values in our comparison. The reported sample size includes observations satisfying the overlapping requirement from both the treated and control samples. Our results show that the deals with higher index values do not have higher yield spreads, but rather their yield spreads are often lower, with statistical significance for the top 10% deals in the issuance written communication contents (about 20 bps lower in yields), than comparable deals with lower index values. This result suggests that the loss relevant information in deal issuance communication contents are not reflected in the deal initial yields. Ghent, Torous, and Valkanov (2019) find that RMBS from more complex deals do not have higher yield spread even though they suffer higher losses. Our findings reinforce the narrative that investors in non-agency RMBS do not take into account certain loss-relevant information into pricing.

The effect of issuance written communication contents on deal losses can vary across deals depending upon the extent of neglected risks. Intuitively, deals with more severe neglected risks on issuance written communication contents see a larger price upward bias everything else the same. For deals with the same characteristics other than initial yields, higher initial deal pricing may reflect greater extent of neglected risks pertaining to the issuance written communication contents. In our second analysis, we compare the effect of the issuance communication contents on deal performance between high-priced and low-priced deals. Because neglect of loss-relevant information by investors makes overpricing more likely, these deals are likely overpriced controlling for everything else. As a result, we expect the interaction term between the issuance written communication contents and deal pricing to have a positive effect on deal losses. In other words, if comparing two deals with the same level of information contents, we expect the information contents to

have a stronger predicative power for future deal losses for the higher priced deals. We perform similar regression analysis on deal losses including the interaction term between initial yield spread and the indices constructed from the issuance written communication contents. Table 8 reports the estimation results. The estimated coefficient for the interaction term is significantly negative in both specifications. Since high-priced deals have lower yield spreads, we indeed find that neglected information has a stronger predicative power for defaults in high-priced deals controlling for everything else.

Table 8 about here

5. Security Losses, Pricing, and Investors' Holdings

We document above that, compared to mutual funds, insurance companies and pension funds neglected risk-relevant information in issuance written communications in the offering process of RMBS. In this section, we analyze whether the initial security holdings of insurance companies and pension funds are associated with higher defaults than those of mutual funds. To this end, we construct measures at the security level. Our security level loss rate measure is the cumulative principal losses as a percentage of the initial principal for each security. We aggregate the initial holdings of insurance companies (including pension funds) for each security and denote it as `Insure_holding`. The mutual fund security-level holding is constructed in the same manner and denoted as `Mutual_holding`. For ease of comparison, we standardize both `Insure_holding` and `Mutual_holding` so that they have mean of 0 and standard deviation of 1. Consistent with previous analysis, we only include the securities that have initial credit ratings of AAA.

We regress the security level loss rate on `Insure_holding` and `Mutual_holding` and the results are presented in Table 9. Since these are security-level analyses, we include deal fixed effect to control for all variables that are the same for all securities in the same deal. The inclusion of deal fixed effect mitigates the influence of deal characteristics, both observed

and unobserved, and allows us to identify the relation between the security loss rate and the institutional investor's initial holdings across different investor types. It is clear from the estimation result that more insurance company holdings are associated with higher future losses at the security level, whereas the mutual fund holdings are insignificantly associated with future losses. For example, one standard deviation increase in insurance company (mutual fund) holdings is associated with 4.1% (1.7%) higher losses (column (1)). The difference is significant both statistically and economically.

More importantly, when we conduct subsample analysis on deals with higher (lower) level of neglected risk-related information in the issuance written communications, we find that the difference between insurance company and mutual fund holdings is more pronounced in the subsample of higher level of neglected risks. Using our previously defined Textual Risk Indices for the level of neglected risks, we find that the difference in future losses between insurance company and mutual fund holdings is 3.3% (column 2) and 4.0% (column 4) for the deals with higher index values using Textual Risk Index 1 and 2, respectively, both estimates are statistically significant. On the other hand, the corresponding difference is 1.2% and 0.7%, respectively, for the deals with lower textual risk index values and are statistical insignificant. Our results indicate that in contrast to mutual funds, insurance companies neglect loss-relevant information in the RMBS issuance written communications.

Table 9 about here

Finally, we analyze the relation between security pricing and insurance company and mutual fund holdings to answer the question whether insurance companies are compensated for holding securities from deals of higher level of neglected risk-related information in the issuance written communications. We measure security level pricing by the yield spread between the securities that have initial credit ratings of AAA and Treasury yields of the corresponding maturity for fixed-rate bonds and LIBOR index for floating-rate bonds. We regress the security level yield spread on `Insure_holding` and `Mutual_holding` and the results

are presented in Table 10. Since these are security-level analyses, we include deal fixed effect to control for all variables that are the same for all securities in the same deal.

We find that the securities with higher initial holdings by insurance companies and pension funds relative to mutual funds actually have lower yield spreads, yet statistically insignificant. This is the case across deals with high or low level of neglected risk-related information in the issuance written communications. Therefore, we do not find evidence that insurance companies get compensated for holding securities from deals that have more loss relevant communication contents. An alternative explanation for insurance companies holding more securities from deals with a higher level of neglected risk-related information communication contents is that they reach for yield within a rating category because they are constrained in their risk-taking by ratings-based regulation. Our findings do not support such an alternative explanation.

Table 10 about here

6. Concluding Remarks

This paper is the first to analyze the entire body of issuance written communications in the offering process of RMBS and the first to document direct empirical evidence of the neglected risks embedded in issuance written communications by RMBS investors. We find that the RMBS issuance written communications are closely related to the underlying asset quality with respect to future deal losses. Using textual analysis on the pre-issuance written communications and the final prospectus supplements, we find that textual FWPs and uncertain word usage are strongly positively related to future deal losses. Our analyses show that the compensation to investors for bearing such risks are insufficient for the losses incurred, indicating that investors neglected a significant portion of the risk-related information in issuance written communications.

Our study also fills a critical void in the existing literature by providing direct empiri-

cal evidence of institutional investors' initial holdings of RMBS and the differing extent to which risk-related information was neglected by different types of investors with respect to issuance written communications. In particular, we find that mutual funds are more attentive to issuance written communication contents. They hold more securitized mortgage deals whose pre-issuance written communications contained more numerical information and relatively less textual information as well as less uncertain word usage in the final prospectus supplements, especially in the Risk Factor section. In contrast, insurance companies and pension funds are unattentive to the issuance written communication contents; rather, their RMBS initial holdings are very closely related to salient headline deal characteristics such as the FICO score, low documentation, underwriter reputation, among others. Analysis at the RMBS security level shows that the initial holdings of insurance companies and pension funds suffered higher losses than average RMBS while the initial holdings of mutual funds performed similarly to average RMBS. Our findings demonstrate that there is substantial risk-related information in issuance written communications that is neglected by investors. This neglected risk-related information is associated with significant investment losses, leading to the fragility of a security market with financial innovations such as the RMBS market preceding the recent financial crisis.

References

- Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D Evanoff, 2014, Predatory lending and the subprime crisis, *Journal of Financial Economics* 113, 29–52.
- Agarwal, Sumit, and Itzhak Ben-David, 2014, Do loan officers' incentives lead to lax lending standards?, *Working Paper*.
- Barlevy, Gadi, and Jonas DM Fisher, 2011, Mortgage choices and housing speculation, Federal Reserve Bank of Chicago Working Paper.
- Baron, Matthew, and Wei Xiong, 2017, Credit expansion and neglected crash risk*, *The Quarterly Journal of Economics* 132, 713–764.
- Becker, Bo, and Victoria Ivashina, 2015, Reaching for yield in the bond market, *The Journal of Finance* 70, 1863–1902.
- Ben-David, Itzhak, 2011, Financial constraints and inflated home prices during the real estate boom, *American Economic Journal: Applied Economics* pp. 55–87.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2018, Diagnostic expectations and credit cycles, *The Journal of Finance* 73, 199–227.
- Celerier, Claire, and Boris Vallee, 2017, Catering to investors through security design: Headline rate and complexity, *The Quarterly Journal of Economics* 132, 1469–1508.
- Cheng, Ing-Haw, Sahil Raina, and Wei Xiong, 2014, Wall street and the housing bubble, *American Economic Review* 104, 2797–2829.
- Chernenko, Sergey, Samuel G. Hanson, and Adi Sunderam, 2016, Who neglects risk? investor experience and the credit boom, *Journal of Financial Economics* 122, 248 – 269.

- Chinco, Alex, and Chris Mayer, 2012, Distant speculators and asset bubbles in the housing market, *Columbia Business School mimeo*, April.
- Cohen, Lauren, Christopher J. Malloy, and Quoc H. Nguyen, 2016, Lazy prices, Working Paper.
- Dai, Zhonglan, Harold H. Zhang, and Feng Zhao, 2014, Tug-of-war: Incentive alignment in securitization and loan performance, *Working Paper*.
- Demiroglu, Cem, and Christopher James, 2012, How important is having skin in the game? originator-sponsor affiliation and losses on mortgage-backed securities, *Review of Financial Studies* 25, 3217–3258.
- Garmaise, Mark J, 2015, Borrower misreporting and loan performance, *The Journal of Finance* 70, 449–484.
- Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2012, Neglected risks, financial innovation, and financial fragility, *Journal of Financial Economics* 104, 452 – 468 Market Institutions, Financial Market Risks and Financial Crisis.
- , 2015, Neglected Risks: The Psychology of Financial Crises, *American Economic Review* 105, 310–314.
- Ghent, Andra C, Walter N Torous, and Rossen I Valkanov, 2019, Complexity in Structured Finance, *The Review of Economic Studies* 86, 694–722.
- Griffin, John, Richard Lowery, and Alessio Saretto, 2014, Complex securities and underwriter reputation: Do reputable underwriters produce better securities?, *Review of Financial Studies* 27, 2872–2925.
- Griffin, John M, and Gonzalo Maturana, 2014, Who facilitated misreporting in securitized loans?, *Review of Financial Studies*, forthcoming.

- Hanley, Kathleen Weiss, and Gerard Hoberg, 2010, The information content of ipo prospectuses, *Review of Financial Studies* 23, 2821–2864.
- , 2012, Litigation risk, strategic disclosure and the underpricing of initial public offerings, *Journal of Financial Economics* 103, 235–254.
- , 2019, Dynamic interpretation of emerging risks in the financial sector, *Review of Financial Studies* p. forthcoming.
- Haughwout, Andrew, Donghoon Lee, Joseph S Tracy, and Wilbert Van der Klaauw, 2011, Real estate investors, the leverage cycle, and the housing market crisis, *FRB of New York Staff Report*.
- He, JIE (JACK), JUN (QJ) Qian, and PHILIP E. Strahan, 2012, Are all ratings created equal? the impact of issuer size on the pricing of mortgage-backed securities, *The Journal of Finance* 67, 2097–2137.
- Jiang, Wei, Ashlyn Nelson, and Edward Vytlacil, 2014a, Liar’s loan? – effects of origination channel and information falsification on mortgage loan delinquency, *Review of Economics and Statistics* 91, 1–18.
- Jiang, Wei, Ashlyn Aiko Nelson, and Edward Vytlacil, 2014b, Securitization and loan performance: Ex ante and ex post relations in the mortgage market, *Review of Financial Studies* 27, 454–483.
- Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2010, Did securitization lead to lax screening? evidence from subprime loans, *The Quarterly Journal of Economics* 125, 307–362.
- Keys, Benjamin J., Amit Seru, and Vikrant Vig, 2012, Lender screening and the role of securitization: Evidence from prime and subprime mortgage markets, *Review of Financial Studies* 25, 2071–2108.

- Loughran, Tim, and Bill McDonald, 2013, IPO first-day returns, offer price revisions, volatility, and Form S-1 language, *Journal of Financial Economics* 109, 307 – 326.
- , 2016, Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* 54, 1187–1230.
- Merrill, Craig B., Taylor D. Nadauld, and Philip E. Strahan, 2019, Final demand for structured finance securities, *Management Science* 65, 390–412.
- Mian, Atif, and Amir Sufi, 2009, The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis, *The Quarterly Journal of Economics* 124, 1449–1496.
- Nadauld, Taylor D., and Shane M. Sherlund, 2013, The impact of securitization on the expansion of subprime credit, *Journal of Financial Economics* 107, 454–476.
- Piskorski, Tomasz, Amit Seru, and James Witkin, 2014, Asset quality misrepresentation by financial intermediaries: Evidence from RMBS market, *Journal of Finance*, *forthcoming*.
- Purnanandam, Amiyatosh, 2011, Originate-to-distribute model and the subprime mortgage crisis, *Review of Financial Studies* 24, 1881–1915.
- Rajan, Uday, Amit Seru, and Vikrant Vig, 2015, The failure of models that predict failure: Distance, incentives, and defaults, *Journal of Financial Economics* 115, 237–260.
- Tzioumis, Konstantinos, and Matthew Gee, 2013, Nonlinear incentives and mortgage officers decisions, *Journal of Financial Economics* 107, 436–453.
- Zhou, Dexin, 2017, Good news in numbers, Working Paper.

Appendix A: Extracting the content of prospectus supplement

We rely on the “Table of Contents” in the prospectus supplement to locate the starting and ending positions of each section. The four sections that we focus on are as follows:

- Summary
- Risk Factors
- Description of Certificates
- The Mortgage Pool

To arrive at the above common section names, we first extract all the section names and cluster them by frequency, then we go through all the names manually to consolidate the section names. Therefore, in an individual prospectus supplement, the section name may take a slightly different form from the one presented above. For example, “The Mortgage Pool” could also appear in the following names: “Description of the mortgage pool”, “Description of the mortgage pools”, “The mortgage pools”, “Mortgage pool characteristics”, and “Information about the mortgage pool”. The “Description of Certificates” could also appear as: “Description of the certificates”, “The offered certificates”, “Offered certificates”, “Definitive certificates”, “General description of certificates”, “Description of the securities”, and “Description of securities”.

Appendix B: Sample Risk Factor descriptions (First Franklin Mortgage Loan Trust 2006-FF4)

The following sentences are extracted from the Risk Factor section of First Franklin Mortgage Loan Trust 2006-FF4 (with “may” and “could” highlighted by us).

- Less stringent underwriting standards and the resultant potential for delinquencies on the mortgage **could** lead to losses on your certificates.
- The mortgage loans were made, in part, to borrowers who, for one reason or another, are not able, or do not wish, to obtain financing from traditional sources. These mortgage loans **may** be considered to be of a riskier nature than mortgage loans made by traditional sources of financing, so that the holders of the certificates **may** be deemed to be at greater risk of loss than if the mortgage loans were made to other types of borrowers.
- The underwriting standards used in the origination of the mortgage loans held by the trust are generally less stringent than those of Fannie Mae or Freddie Mac with respect to a borrower’s credit history and in certain other respects. Mortgage loan borrowers **may** have an impaired or unsubstantiated credit history. As a result of this less stringent approach to underwriting, the mortgage loans purchased by the trust **may** experience higher rates of delinquencies, defaults and foreclosures than mortgage loans underwritten in a manner which is more similar to the Fannie Mae and Freddie Mac guidelines.
- Increased use of new mortgage loan products by borrowers **may** resulted in a decline in real estate values generally. In recent years, borrowers have increasingly financed their homes with new mortgage loan products, which in many cases have allowed them to purchase homes that they might otherwise not have been able to afford. Many of these new products feature low monthly payments during the initial years of the loan that can increase (in some cases, significantly) over the loan term. There is little historical data with respect to these new mortgage loan products. Consequently, as borrowers face potentially higher monthly payments for the remaining terms of their loans, it is possible that, combined with other economic conditions such as increasing interest rates and deterioration of home values, borrower delinquencies and defaults **could** exceed anticipated levels. In that event, the certificates, and your investment in the certificates, **may** not perform as you anticipate.

Appendix C: Background on ABS offering process and example FWPs

While the offering process for asset-backed securities (including MBS) shares some features with the offering process for corporate securities (such as equity and other fixed income securities), they also differ from each other in important ways. As noted by the SEC in its July 2004 release on the rule change on communication to allow more information to reach investors, “ABS investors are generally interested in the characteristics and quality of the underlying assets, the standards for their servicing, the timing and receipt of cash flows from those assets and the structure for distribution of those cash flows. As a general matter, there is essentially no business or management (and therefore no management’s discussion and analysis of financial performance and condition) of the issuing entity, which is designed to be a solely passive entity.” As a result, information regarding characteristics and quality of the assets is critical for investors in assessing how an asset pool and tranches of the asset pool as ABS securities will perform.

Because the primary risk profiles are delineated when the asset pool is formed, information critical to the risk assessment of asset-backed securities is contained in the written communications in the issuance process of these securities. Below we provide some example FWPs from First Franklin Mortgage Loan Trust 2006-FF4.

- FWP size: 25 KB (20060222)
Link: <http://www.sec.gov/Archives/edgar/data/807641/000091412106000390/0000914121-06-000390.txt>
Content: Basic summary statistics for collateral analysis
- FWP size: 372 KB (20060302)
Link: <http://www.sec.gov/Archives/edgar/data/1353977/000091412106000515/0000914121-06-000515.txt>
Content: Structural and collateral term sheet (including some selected mortgage pool data)
- FWP size: 491 KB (20060302)
Link: <http://www.sec.gov/Archives/edgar/data/1353977/000091412106000517/0000914121-06-000517.txt>
Content: More summary statistics on the underlying loans
- FWP size: 296 KB (20060303)
Link: <http://www.sec.gov/Archives/edgar/data/1353977/000091412106000531/0000914121-06-000531.txt>

Content: Updated structural and collateral term sheet (with modified deal structure such as amount and characteristics of certain class of certificates)

- FWP size: 7 MB (20060308)

Link: <http://www.sec.gov/Archives/edgar/data/1353977/000091412106000573/0000914121-06-000573.txt>

Content: Detailed loan level data

- FWP size: 9 MB (20060309)

Link: <http://www.sec.gov/Archives/edgar/data/1353977/000091412106000585/0000914121-06-000585.txt>

Content: Updated loan level data

- FWP size: 363 KB (20060323)

Link: <http://www.sec.gov/Archives/edgar/data/1353977/000091412106000795/0000914121-06-000795.txt>

Content: Summary statistics of loan characteristics by loan types

Appendix D: Variable definitions

- Cumulative net loss: Historical percentages of cumulative loss on the underlying loans comprising the entire collateral that backs the deal
- No. of FWPs: Number of free writing prospectuses prior to the deal issuance date
- Multiple FWPs: Equals 1 if no. of FWPs is greater than 1; 0 otherwise
- Log(Textual FWP): The logarithm of 1 plus the sum of the number of alphabet and numeric tokens for all the textual FWPs within a deal
- Log(Loantape FWP): The logarithm of 1 plus the sum of the number of alphabet and numeric tokens for all the loantape FWPs within a deal
- UncProp: Number of uncertain words in a deal's final prospectus supplement
- UncRiskFactor: Number of uncertain words in a deal's Risk Factor section of the final prospectus supplement
- UncCertificate: Number of uncertain words in a deal's Description of Certificate section of the final prospectus supplement
- UncMortgagePool: Number of uncertain words in a deal's Mortgage Pool section of the final prospectus supplement
- UncSummary: Number of uncertain words in a deal's Summary section of the final prospectus supplement
- Original collateral balance: The original balance of the underlying loans comprising the entire collateral
- High reputation: Equals 1 if the deal has an underwriter IPO reputation score greater than or equal to 8 (from Professor Jay Ritter's website); 0 otherwise
- No. of tranches: Number of securities in a deal
- FICO: Weighted average original credit score of the underlying loans
- Low documentation: percent of underlying loans with limited, as distinguished from full, documentation
- LTV: Original loan to value percentage of the loan

- Adjustable rate mortgage: The percent of the adjustable rate mortgage loans
- Negative amortization: Equals 1 if the deal consists of mortgages with negative amortization features; 0 otherwise
- Purchase loans: The percent of the Loan Purpose (the reason for the loan) for Purchase
- Single family: percent of Single Family Mortgaged Properties, the type of properties against which the loans were written
- Owner occupied: percent of the Occupancy (the purpose of the property) for Owner Occupied
- Second lien: percentage of the loans comprising the collateral that are second lien

Table 1: Summary statistics on final prospectus supplement content

This table presents the summary statistics on the uncertain word usage in final prospectus supplement. The total number of uncertain words (in 100) in the prospectus supplement is denoted as *UncProp*. The total number of uncertain words (in 100) in four key sections of the final prospectus supplement are denoted as *UncRiskFactor* for the Risk Factor section, *UncCertificate* for the Description of Certificates section, *UncMortgagePool* for the Mortgage Pool section, and *UncSummary* for the Summary section. For a deal to be included in this sample, we require that these 4 sections be non-missing. Panel A presents the summary statistics. The statistics reported include N (number of observations), Mean, Std. Dev. (standard deviation), the k^{th} percentile (P_k for $k = 5, 25, 50, 75, 95$) of each variable. Panel B reports the correlation matrix. Panel C presents the occurrence frequency of uncertain words in four key sections of the final prospectus supplement. The column *freq* represents the frequency of the top 20 words in each section and the column *pct* represents the ratio of a word’s frequency to the total number of uncertain words in each section.

Panel A: summary statistics	N	Mean	Std. Dev.	P5	P25	P50	P75	P95
<i>UncProp</i>	1419	12.67	4.93	6.69	9.39	12.10	15.45	20.05
<i>UncRiskFactor</i>	1419	2.77	1.11	1.27	2.02	2.66	3.56	4.70
<i>UncCertificate</i>	1419	2.69	1.17	1.03	1.85	2.59	3.31	4.85
<i>UncMortgagePool</i>	1419	1.24	0.77	0.31	0.74	1.09	1.59	2.69
<i>UncSummary</i>	1419	1.02	0.42	0.46	0.75	0.94	1.22	1.73

Panel B: correlation	<i>UncProp</i>	<i>UncRiskFactor</i>	<i>UncCertificate</i>	<i>UncMortgagePool</i>
<i>UncRiskFactor</i>	0.62			
<i>UncCertificate</i>	0.31	0.28		
<i>UncMortgagePool</i>	0.23	0.18	0.34	
<i>UncSummary</i>	0.12	0.17	0.16	0.06

Panel C: frequency of uncertain words

Risk Factor			Certificate			Mortgage Pool			Summary		
UncWord	freq	pct									
MAY	77888	25%	LOSS	40561	14%	APPROXIMATELY	36950	27%	MAY	16595	15%
LOSS	32569	10%	MAY	32388	11%	MAY	24568	18%	LOSS	14295	13%
RISK	17515	6%	TERMINATION	29372	10%	DELINQUENCY	8470	6%	APPROXIMATELY	11489	10%
COULD	12266	4%	UNPAID	18640	6%	CLOSING	6377	5%	TERMINATION	8117	7%
APPROXIMATELY	12053	4%	SHORTFALL	18246	6%	RISK	4101	3%	CLOSING	7625	7%
DEFAULT	11942	4%	APPROXIMATELY	12199	4%	REFINANCE	2962	2%	SHORTFALL	5718	5%
SHORTFALL	10236	3%	RISK	10657	4%	DEFAULT	2919	2%	UNPAID	5059	4%
DELINQUENCY	9912	3%	LIQUIDATION	9633	3%	UNPAID	2625	2%	RISK	4138	4%
ADVERSELY	7598	2%	UNSCHEDULED	8479	3%	PENALTY	2237	2%	DELINQUENCY	2490	2%
TERMINATION	5965	2%	CLOSING	8376	3%	LOSS	2026	1%	VARIABLE	2270	2%
BANKRUPTCY	3441	1%	DELINQUENCY	5831	2%	BANKRUPTCY	1968	1%	LIKELIHOOD	1847	2%
FORECLOSURE	3288	1%	DEFAULT	5426	2%	LATE	1677	1%	BREACH	1779	2%
DECLINE	3165	1%	BANKRUPTCY	4854	2%	NEGATIVE	1590	1%	COULD	1640	1%
POSSIBLE	2797	1%	FAILURE	3533	1%	BREACH	1434	1%	SOMETIMES	1618	1%
INSUFFICIENT	2502	1%	LIQUIDATED	3521	1%	FORECLOSURE	1337	1%	AGAINST	1597	1%
DELAYS	2501	1%	DEPLETION	3142	1%	VARY	1283	1%	ADVERSELY	1401	1%
LIQUIDATION	2400	1%	FORECLOSURE	2992	1%	VARIANCE	1137	1%	LIQUIDATION	1172	1%
LIKELIHOOD	2341	1%	DEFAULTED	2905	1%	PROBABILITY	1075	1%	POSSIBLE	1072	1%
UNPAID	2170	1%	UNREIMBURSED	2646	1%	BELIEVES	937	1%	NEGATIVE	1024	1%
SLOWER	2004	1%	TERMINATE	2365	1%	HAZARD	935	1%	DEFECT	885	1%

Table 2: Characteristics of FWP contents

This table presents the content characteristics of the two main types of FWPs: loantape FWP (Panel A) and textual FWP (Panel B). A token is a structure representing a lexeme that explicitly indicates its categorization for the purpose of parsing. To obtain tokens from a FWP document, we first parse the document to strip off various html tags, then we break the remaining document into tokens by any white space. We define a token as a numeric token if the first character is a number (0 to 9). Similarly, we define a token as an alphabet token if the first character is a letter (a to z or A to Z). We classify an FWP as a loantape FWP if it is dominated by numeric tokens and as a textual FWP if it is dominated by alphabet tokens. Panel C reports the summary statistics on FWP contents. The statistics reported include N (number of observations), Mean, Std. Dev. (standard deviation), the k^{th} percentile (Pk for $k = 50, 75$) of each variable.

Panel A: Content characteristics of Loantape FWP

N=794 (Loantape FWP)	Mean	St. Dev.	Q1	Median	Q3
No. of alphabet tokens	379,006	376,664	128,205	269,461	507,007
No. of numeric tokens	820,367	670,928	366,878	642,897	1,078,382
Alphabet-number ratio	0.50	0.38	0.28	0.41	0.61

Panel B: Content characteristics of Textual FWP

N=2055 (Textual FWP)	Mean	St. Dev.	Q1	Median	Q3
No. of alphabet tokens	300,890	302,020	56,752	169,712	566,700
No. of numeric tokens	23,099	23,382	3,962	14,811	36,777
Alphabet-number ratio	51	681	3	12	58

Panel C: Summary statistics for FWP contents

	N	Mean	Std. Dev.	P50	P75
Log(Loantape FWP)	1743	4.58	6.58	0.00	13.37
Log(Textual FWP)	1743	7.89	6.33	12.00	13.39

Table 3: Summary statistics on institutional investors' initial holdings of RMBS

We obtain institutional investors' initial holdings of RMBS from Thomson Reuter's eMAXX solution. The eMAXX data mainly cover the holdings of insurance companies (pension funds) and mutual funds. We calculate the percentage of all the initial holdings by mutual funds (insurance companies and pension funds), reported within one year of the issuance dates, relative to the issuance amount and denote it as mutual fund (insurance company) initial holdings (%). All the securities included in the analysis are required to have an initial rating of AAA. This table presents the summary statistics on mutual fund holdings and insurance company holdings. The statistics reported include N (number of observations), Mean, Std. Dev. (standard deviation), the k^{th} percentile (Pk for $k = 5, 25, 50, 75, 95$) of each variable.

	N	Mean	St. Dev.	P5	P25	P50	P75	P95
Mutual fund holdings (%)	924	5.25	10.59	0.00	0.19	2.03	5.78	20.39
Insurance company holdings (%)	924	11.55	13.28	0.00	2.33	6.78	16.65	39.74

Table 4: Mutual fund initial holdings and issuance communication

This table reports the results of analyzing whether mutual fund initial holdings of RMBS are related to FWP and prospectus supplement content. The dependent variable is the percentage of all the initial holdings by mutual funds, reported within one year of the issuance dates, relative to the issuance amount. All the securities included in the analysis are required to have an initial rating of AAA. All the other variables are defined as in previous tables. The t-statistics based on standard errors clustered at lead-underwriter-level are reported in the parentheses below each coefficient estimate. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(Textual FWP)	-0.82** (-2.12)	-0.82** (-2.12)	-0.70* (-1.95)	-0.85** (-2.10)	-0.85** (-2.12)	-0.79** (-2.28)	-0.75* (-1.93)
Log(Loantape FWP)	0.57* (1.80)	0.57* (1.80)	0.68** (2.16)	0.61* (1.82)	0.56* (1.81)	0.57* (1.79)	0.68* (2.06)
UncProp		-0.04 (-0.20)					
UncRiskFactor			-0.80** (-2.49)				-0.74* (-1.77)
UncCertificate				-0.32 (-0.78)			-0.14 (-0.36)
UncMortgagePool					-0.34 (-1.44)		-0.20 (-0.87)
UncSummary						-0.22 (-0.56)	0.04 (0.10)
Original collateral balance	-0.09 (-0.23)	-0.09 (-0.23)	-0.13 (-0.32)	-0.11 (-0.27)	-0.11 (-0.26)	-0.09 (-0.21)	-0.14 (-0.36)
High reputation	2.07 (1.02)	2.07 (1.01)	2.13 (1.05)	2.11 (1.04)	2.13 (1.05)	2.12 (1.07)	2.17 (1.07)
No. of tranches	0.34 (1.28)	0.34 (1.28)	0.37 (1.34)	0.35 (1.28)	0.34 (1.25)	0.33 (1.20)	0.37 (1.29)
Low documentation	0.50 (1.02)	0.50 (1.02)	0.44 (0.89)	0.51 (1.01)	0.52 (1.05)	0.51 (1.03)	0.46 (0.89)
FICO	0.34 (0.68)	0.34 (0.68)	0.40 (0.76)	0.38 (0.78)	0.31 (0.65)	0.33 (0.66)	0.39 (0.80)
LTV	-0.56 (-0.84)	-0.56 (-0.83)	-0.49 (-0.70)	-0.55 (-0.81)	-0.56 (-0.86)	-0.56 (-0.84)	-0.49 (-0.72)
Adjustable rate mortgage	0.58 (0.93)	0.58 (0.93)	0.54 (0.89)	0.58 (0.94)	0.59 (0.96)	0.57 (0.92)	0.56 (0.92)
Negative amortization	7.84*** (3.35)	7.84*** (3.35)	8.01*** (3.44)	7.83*** (3.26)	7.87*** (3.29)	7.83*** (3.23)	8.01*** (3.35)

Continued on Next Page...

Table 4 – Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Purchase loans	-0.31 (-0.56)	-0.32 (-0.56)	-0.32 (-0.57)	-0.34 (-0.59)	-0.31 (-0.54)	-0.31 (-0.55)	-0.33 (-0.57)
Single family	0.42 (1.02)	0.42 (1.02)	0.43 (1.06)	0.43 (1.07)	0.43 (1.06)	0.43 (1.10)	0.44 (1.12)
Owner occupied	-0.73 (-1.04)	-0.73 (-1.03)	-0.74 (-1.05)	-0.71 (-0.99)	-0.70 (-0.98)	-0.73 (-1.03)	-0.71 (-0.98)
Second lien	-0.24 (-0.45)	-0.24 (-0.44)	-0.17 (-0.28)	-0.19 (-0.36)	-0.24 (-0.44)	-0.21 (-0.40)	-0.15 (-0.27)
House prices change	-0.15 (-0.26)	-0.15 (-0.27)	-0.19 (-0.32)	-0.13 (-0.23)	-0.14 (-0.24)	-0.16 (-0.29)	-0.17 (-0.28)
Credit spread	-0.62 (-1.21)	-0.62 (-1.21)	-0.60 (-1.22)	-0.66 (-1.39)	-0.64 (-1.25)	-0.62 (-1.21)	-0.63 (-1.38)
10 Year Treasury	0.37 (0.81)	0.37 (0.81)	0.36 (0.79)	0.35 (0.79)	0.36 (0.79)	0.35 (0.78)	0.35 (0.78)
Adj. R^2	0.120	0.119	0.124	0.120	0.120	0.119	0.121
N	758	758	758	758	758	758	758
Lead underwriter FE	Yes						
Year FE	Yes						

Table 5: Insurance company initial holdings and issuance communication

This table reports the results of analyzing whether mutual fund initial holdings of RMBS are related to FWP and prospectus supplement content. The dependent variable is the percentage of all the initial holdings by insurance companies (including pension funds), reported within one year of the issuance dates, relative to the issuance amount. All the securities included in the analysis are required to have an initial rating of AAA. All the other variables are defined as in previous tables. The t-statistics based on standard errors clustered at lead-underwriter-level are reported in the parentheses below each coefficient estimate. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(Textual FWP)	0.04 (0.10)	0.09 (0.22)	0.09 (0.22)	0.04 (0.10)	-0.02 (-0.06)	0.04 (0.09)	0.00 (0.01)
Log(Loantape FWP)	0.19 (0.37)	0.18 (0.36)	0.23 (0.43)	0.20 (0.37)	0.17 (0.34)	0.19 (0.37)	0.19 (0.37)
UncProp		-0.73 (-1.51)					
UncRiskFactor			-0.34 (-0.67)				-0.29 (-0.48)
UncCertificate				-0.02 (-0.06)			0.12 (0.28)
UncMortgagePool					-0.69 (-1.51)		-0.68 (-1.41)
UncSummary						0.03 (0.10)	0.17 (0.52)
Original collateral balance	-1.54*** (-3.56)	-1.54*** (-3.53)	-1.55*** (-3.63)	-1.54*** (-3.56)	-1.57*** (-3.65)	-1.54*** (-3.57)	-1.57*** (-3.72)
High reputation	9.09*** (3.16)	9.13*** (3.10)	9.11*** (3.13)	9.09*** (3.17)	9.21*** (3.16)	9.08*** (3.15)	9.18*** (3.10)
No. of tranches	0.34 (1.13)	0.34 (1.13)	0.35 (1.20)	0.34 (1.13)	0.33 (1.09)	0.34 (1.13)	0.34 (1.19)
Low documentation	-1.11** (-2.44)	-1.18** (-2.49)	-1.14** (-2.53)	-1.11** (-2.45)	-1.08** (-2.39)	-1.11** (-2.42)	-1.11** (-2.49)
FICO	5.66*** (3.12)	5.71*** (3.16)	5.76*** (3.13)	5.66*** (3.12)	5.64*** (3.04)	5.67*** (3.10)	5.76*** (3.04)
LTV	-1.28* (-1.76)	-1.29* (-1.76)	-1.27* (-1.76)	-1.28* (-1.76)	-1.30* (-1.78)	-1.28* (-1.77)	-1.30* (-1.81)
Adjustable rate mortgage	-1.17 (-1.67)	-1.18 (-1.68)	-1.19 (-1.67)	-1.17 (-1.66)	-1.16 (-1.63)	-1.17 (-1.67)	-1.16 (-1.62)
Negative amortization	2.06** (2.15)	2.14** (2.38)	2.14** (2.30)	2.06** (2.15)	2.11** (2.16)	2.06** (2.15)	2.21** (2.30)

Continued on Next Page...

Table 5 – Continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Purchase loans	-0.21 (-0.45)	-0.24 (-0.50)	-0.21 (-0.45)	-0.21 (-0.46)	-0.20 (-0.44)	-0.21 (-0.45)	-0.19 (-0.44)
Single family	-0.10 (-0.20)	-0.10 (-0.21)	-0.10 (-0.21)	-0.10 (-0.20)	-0.08 (-0.18)	-0.10 (-0.21)	-0.10 (-0.22)
Owner occupied	0.86 (1.28)	0.87 (1.33)	0.86 (1.29)	0.86 (1.30)	0.92 (1.40)	0.86 (1.28)	0.91 (1.41)
Second lien	-0.32 (-0.83)	-0.28 (-0.76)	-0.29 (-0.81)	-0.31 (-0.78)	-0.31 (-0.76)	-0.32 (-0.87)	-0.32 (-0.80)
House prices change	2.59*** (3.89)	2.57*** (3.88)	2.57*** (3.90)	2.59*** (3.88)	2.62*** (3.98)	2.59*** (3.89)	2.60*** (3.92)
Credit spread	1.54 (1.62)	1.58 (1.68)	1.55 (1.67)	1.54 (1.65)	1.51 (1.59)	1.54 (1.61)	1.53 (1.67)
10 Year Treasury	1.22 (1.59)	1.22 (1.59)	1.22 (1.58)	1.22 (1.58)	1.21 (1.60)	1.22 (1.63)	1.24 (1.61)
Adj. R^2	0.307	0.308	0.306	0.306	0.308	0.306	0.306
N	758	758	758	758	758	758	758
Lead underwriter FE	Yes						
Year FE	Yes						

Table 6: Cumulative net loss and issuance communication content

This table reports the results of analyzing how FWP content and the final prospectus supplement content affect the future performance of the underlying loans in terms of cumulative net loss (the dependent variable), controlling for the variables that are related to initial pricing including the initial yield spread and the subordination. The main independent variables of interest are the FWP content: Log(Textual FWP) and Log(Loantape FWP), and the final prospectus supplement content: UncProp, UncRiskFactor, UncCertificate, UncMortgagePool, and UncSummary. The other deal characteristic variables are defined in previous tables and in the Appendix B. All continuous variables are standardized to have unit variance. The t-statistics based on standard errors clustered at lead-underwriter-level are reported in the parentheses below each coefficient estimate. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
UncProp	0.89** (2.27)					
UncRiskFactor		1.18*** (4.84)				1.03*** (4.90)
UncCertificate			0.74*** (3.83)			0.52** (2.23)
UncMortgagePool				0.10 (0.64)		-0.14 (-0.82)
UncSummary					0.53** (2.59)	0.17 (0.92)
Log(Textual FWP)	1.00*** (5.48)	0.88*** (4.34)	1.14*** (4.94)	1.09*** (4.76)	1.02*** (4.41)	0.91*** (4.07)
Log(Loantape FWP)	-0.31 (-1.51)	-0.40** (-2.16)	-0.39 (-1.65)	-0.25 (-1.10)	-0.26 (-1.05)	-0.49** (-2.34)
Initial yield spread	1.55*** (4.67)	1.43*** (4.56)	1.63*** (4.63)	1.60*** (4.82)	1.65*** (4.77)	1.49*** (4.35)
Subordination	0.58 (0.99)	0.39 (0.73)	0.65 (1.06)	0.55 (0.94)	0.63 (1.12)	0.50 (0.92)
Original collateral balance	-0.05 (-0.28)	-0.05 (-0.29)	-0.04 (-0.22)	-0.06 (-0.29)	-0.04 (-0.22)	-0.04 (-0.20)
High reputation	0.90 (0.76)	0.88 (0.81)	0.92 (0.74)	0.88 (0.74)	0.90 (0.72)	0.92 (0.80)
No. of tranches	0.74** (2.17)	0.74* (2.05)	0.74** (2.27)	0.73** (2.16)	0.73** (2.31)	0.75** (2.18)
Low documentation	1.55*** (4.94)	1.56*** (5.01)	1.54*** (4.92)	1.54*** (4.84)	1.55*** (4.91)	1.56*** (5.08)

Continued on Next Page...

Table 6 – Continued

	(1)	(2)	(3)	(4)	(5)	(6)
FICO	-4.18*** (-5.77)	-4.38*** (-6.18)	-4.10*** (-5.25)	-4.17*** (-5.57)	-4.09*** (-5.48)	-4.28*** (-5.90)
LTV	0.91*** (3.56)	0.94*** (3.45)	0.89*** (3.18)	0.92*** (3.25)	0.88*** (3.06)	0.89*** (3.33)
Adjustable rate mortgage	0.90*** (3.18)	0.91*** (3.38)	0.89*** (3.20)	0.90*** (3.24)	0.90*** (3.45)	0.91*** (3.43)
Negative amortization	3.10*** (3.02)	2.89*** (2.89)	3.18*** (3.06)	3.17*** (2.90)	3.22*** (3.09)	2.95*** (3.09)
Purchase loans	2.18*** (7.44)	2.18*** (8.28)	2.17*** (6.75)	2.17*** (7.02)	2.16*** (7.15)	2.18*** (7.95)
Single family	-0.21 (-1.31)	-0.20 (-1.35)	-0.21 (-1.33)	-0.20 (-1.24)	-0.23 (-1.29)	-0.21 (-1.41)
Owner occupied	-0.69*** (-4.30)	-0.69*** (-4.77)	-0.70*** (-3.97)	-0.69*** (-3.79)	-0.68*** (-3.62)	-0.69*** (-4.63)
Second lien	1.15** (2.46)	1.10** (2.33)	1.18** (2.53)	1.21** (2.52)	1.18** (2.44)	1.09** (2.31)
House prices change	-2.29*** (-7.05)	-2.28*** (-7.03)	-2.32*** (-7.17)	-2.27*** (-6.84)	-2.28*** (-6.49)	-2.32*** (-7.12)
Credit spread	0.02 (0.11)	0.03 (0.10)	0.03 (0.13)	0.04 (0.19)	0.04 (0.16)	0.02 (0.07)
10 Year Treasury	0.10 (0.69)	0.10 (0.61)	0.11 (0.86)	0.10 (0.73)	0.12 (0.84)	0.11 (0.75)
Adj. R^2	0.716	0.722	0.712	0.704	0.708	0.726
N	1271	1271	1271	1271	1271	1271
Lead underwriter FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 7: Difference in initial yield spreads between deals with different issuance communication content using matching approach

The table applies a matching sample analysis to examine the difference in the initial yield spreads for deals with different level of textual content. The deal, regional and macroeconomic matching covariates include original collateral balance, high reputation, No. of tranches, FICO, low documentation, LTV, adjustable rate mortgage, negative amortization, purchase loans, single family owner occupied, second lien, housing price change, credit spread and 10 year treasury. All models match exactly on deal lead underwriters and issuance year. We use nearest-neighbor matching with replacement to implement one-to-one matching of Treatment and Control samples and the significance of the difference in cumulative net loss is based on a z-statistic which is computed using the analytical estimator of the asymptotic variance of matching estimators proposed by Abadie and Imbens (2006, 2008). The A.I. robust z-statistics are reported in the parenthesis below the estimates of the difference. The reported number of observations in each model is the combined samples before matching. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

Textual Risk Index 1 - UncProp, Log(Textual), Log(Loantape)			
	High(top 50%)	High(top 25%)	High(top 10%)
Yield (High - Low)	0.05	-0.05	-0.21**
z	(1.46)	(-1.00)	(-2.28)
N	1238	1229	1202

Textual Risk Index 2 - UncRiskFactor, UncCertificate, Log(Textual), Log(Loantape)			
	High(top 50%)	High(top 25%)	High(top 10%)
Yield (High - Low)	-0.01	-0.07	-0.19**
z	(-0.35)	(-1.19)	(-2.11)
N	1257	1239	1082

Table 8: Cumulative net loss and the interaction of issuance communication content and deal pricing

This table reports the results of analyzing how FWP content and the final prospectus supplement content affect the future performance of the underlying loans in terms of cumulative net loss (the dependent variable), controlling for the variables that are related to initial pricing including the initial yield spread and the subordination. The main independent variables of interest are the FWP content: Log(Textual FWP) and Log(Loantape FWP), and the final prospectus supplement content: UncProp, UncRiskFactor, UncCertificate, UncMortgagePool, and UncSummary. The other deal characteristic variables are defined in previous tables and in the Appendix B. All continuous variables except the interaction terms are standardized to have unit variance. The t-statistics based on standard errors clustered at lead-underwriter-level are reported in the parentheses below each coefficient estimate. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

	(1)	(2)
Initial Yield Spread	1.83*** (4.40)	1.93*** (4.97)
Initial Yield Spread×Textual Risk Index 1	-0.35* (-1.73)	
Initial Yield Spread×Textual Risk Index 2		-0.55** (-2.17)
UncProp	0.91** (2.44)	
UncRiskFactor		1.02*** (4.51)
UncCertificate		0.56** (2.45)
UncMortgagePool		-0.15 (-0.82)
UncSummary		0.12 (0.69)
Log(Textual FWP)	0.93*** (5.04)	0.84*** (3.66)
Log(Loantape FWP)	-0.32 (-1.61)	-0.51** (-2.55)
Deal and Macroeconomic Controls	Yes	Yes
Time Fixed Effects	Yes	Yes
Lead Underwriter Fixed Effects	Yes	Yes
Adjusted R ²	0.717	0.728
Observations	1271	1271

Table 9: Security-level losses and institutional investors' initial security holdings

This table reports the results of analyzing whether mutual fund and insurance company initial holdings of RMBS are related to underlying loan losses at the security level (not the deal level). The dependent variable is the cumulative principal losses as the percentage of the initial underlying loan amount at the security level. The independent variable `Insure_holding` is the percentage of the initial security holdings by all insurance companies (including pension funds), reported within one year of the issuance dates, relative to the issuance amount. It is at the security level. Similarly, we calculate the mutual fund security-level holdings and denote it as `Mutual_holding`. For easier comparison, we standardize both `Insure_holding` and `Mutual_holding` so that they have means of 0 and standard deviations of 1. All the securities included in the analyses are required to have initial ratings of AAA. All the other variables are defined as in previous tables. The t-statistics based on standard errors clustered at deal-level are reported in the parentheses below each coefficient estimate. The results for testing the difference between the coefficients of `Insure_holding` and `Mutual_holding` are reported in the row labelled “`Insure_holding – Mutual_holding`”. Note that deal fixed effects are included in the analyses. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

	(1)	(2)	(3)	(4)	(5)
	Whole Sample	Textual Risk Index 1 top 50%	Textual Risk Index 1 bottom 50%	Textual Risk Index 2 top 50%	Textual Risk Index 2 bottom 50%
<code>Insure_holding</code>	0.0411*** (4.87)	0.0411*** (3.74)	0.0405*** (3.32)	0.0497*** (4.45)	0.0334*** (2.77)
<code>Mutual_holding</code>	0.0170 (1.61)	0.0078 (0.95)	0.0286 (1.44)	0.0071 (0.88)	0.0290 (1.55)
Testing the difference between coefficients of <code>Insure_holding</code> and <code>Mutual_holding</code> :					
<code>Insure_holding – Mutual_holding</code>	0.0241** (1.96)	0.0332** (2.54)	0.0119 (0.56)	0.0427*** (3.30)	0.0043 (0.21)
Deal FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Macro variables	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0719	0.0997	0.0501	0.0974	0.0585
N	1688	844	844	844	844

Table 10: Security-level yield spread and institutional investors' initial security holdings

This table reports the results of analyzing whether mutual fund and insurance company initial holdings of RMBS are related to yield spread at the security level (not the deal level). The dependent variable is the yield spread between the securities that have initial credit ratings of AAA and Treasury yields of the corresponding maturity for fixed-rate bonds and LIBOR index for floating-rate bonds. The independent variable `Insure_holding` is the percentage of the initial security holdings by all insurance companies (including pension funds), reported within one year of the issuance dates, relative to the issuance amount. It is at the security level. Similarly, we calculate the mutual fund security-level holdings and denote it as `Mutual_holding`. For easier comparison, we standardize both `Insure_holding` and `Mutual_holding` so that they have means of 0 and standard deviations of 1. All the securities included in the analyses are required to have initial ratings of AAA. All the other variables are defined as in previous tables. The t-statistics based on standard errors clustered at deal-level are reported in the parentheses below each coefficient estimate. The results for testing the difference between the coefficients of `Insure_holding` and `Mutual_holding` are reported in the row labelled “`Insure_holding – Mutual_holding`”. Note that deal fixed effects are included in the analyses. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

	(1)	(2)	(3)	(4)	(5)
	Whole Sample	Textual Risk Index 1 top 50%	Textual Risk Index 1 bottom 50%	Textual Risk Index 2 top 50%	Textual Risk Index 2 bottom 50%
<code>Insure_holding</code>	-0.0168 (-1.10)	-0.0273 (-1.32)	-0.0075 (-0.35)	-0.0112 (-0.73)	-0.021 (-0.87)
<code>Mutual_holding</code>	0.0057 (0.37)	0.0011 (0.03)	0.011 (0.54)	0.0003 (0.01)	0.0123 (0.67)
Testing the difference between coefficients of <code>Insure_holding</code> and <code>Mutual_holding</code> :					
<code>Insure_holding – Mutual_holding</code>	-0.0225 (-1.16)	-0.0282 (-1.22)	-0.0185 (-0.60)	-0.0115 (-0.54)	-0.0341 (-1.08)
Deal FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Macro variables	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.8625	0.8856	0.8367	0.8959	0.8943
N	754	377	377	377	377

Figure 1: Number of deals in each issuance month

This figure plots the total number of deals in each issuance month in our sample period (2006 and 2007).

