RATING AGENCIES AND INFORMATION EFFICIENCY: DO MULTIPLE CREDIT RATINGS PAY OFF?

THE CASE OF RMBS RATING MIGRATION

STEFAN MORKOETTER[*] ROMAN STEBLER[**] SIMONE WESTERFELD[***]

January 2015

Abstract

We empirically investigate why issuers solicit and pay for multiple ratings not only at issuance but also during the monitoring phase of a debt instrument. Using a unique record of monthly credit rating migration data from Standard & Poor's, Moody's, and Fitch on all U.S. residential mortgage-backed securities from 1985 to 2012 ever rated (154'600 individual tranches), our results provide empirical evidence that rating agencies put more effort in rating and outlook revisions when tranches have assigned multiple ratings. Further, we demonstrate that in the case of multiple ratings, agencies do a better job in discriminating tranches with respect to default risk. Our results contribute to the literature on information production of credit ratings and extend the perspective to the monitoring period after issuance. We also show that in case of multiple ratings, Moody's on average provides the most conservative credit assessment and that this relative pattern remains consistent over a tranche's lifetime.

Keywords: Multiple Ratings, Information Production, Structured Finance, Rating Agencies, Residential Mortgage-backed Securities, Rating Shopping

JEL Classification: G14, G21, G24

---

*University of St. Gallen (mail: stefan.morkoetter@unisg.ch), **University of St. Gallen (mail: roman.stebler@student.unisg.ch), ***University of Applied Sciences Northwestern Switzerland (mail: simone.westerfeld @fhnw.ch). We are grateful to Thomas Nette from Moody's Analytics as well as to Markus Luehe and Mikhail Putkaradze from Fitch Solutions for supporting our research. We thank Martin Brown, Emilia Garcia-Appendini, Zeno Adams, and the participants of the 2014 Paris Financial Management Conference and the University of St. Gallen Research Seminar for useful comments and suggestions. We further thank Tanja Artiga Gonzalez for helping prepare the data.

1. INTRODUCTION

In the wake of the recent financial crisis, credit rating agencies have been heavily criticized by investors, politicians, and the general public for not putting a red flag on the arising U.S. housing bubble and subsequently the subprime crisis. Yet, only limited research exists on the concrete performance of rating agencies throughout the financial crisis. In this context a rather prominent market segment is U.S. residential mortgage-backed securities (U.S. RMBS): banks mainly used securitization structures to sell off U.S. residential mortgages to worldwide investors and, as we know today, this fuelled the U.S. house price bubble. During the months in the run-up of the financial crisis, these securitization transactions actually became more and more complex, involving several layers of different securitizations (so called CDO Squared). Why were investors still interested in buying these complex financial products? Because rating agencies acted as their agents and provided a credit rating for each of the issued tranches, most of which were even rated by multiple agencies and still attested unabated credit quality in the beginning of 2008. In an efficient market environment, however, one rating agency should suffice to fulfill the monitoring function on behalf of the investor base. Yet, we know little why issuers, and ultimately investors, are willing to pay for more than one credit rating. What is the benefit of additional ratings from an investor's perspective?

Against this background, we investigate whether multiple ratings pay off for investors and analyze how accurately the credit rating agencies monitored credit quality throughout the recent financial crisis. Do more ratings justify the additional costs by incorporating more and also better information? Does the number of outstanding ratings increase the monitoring effort of each individual rating agency? Do we observe specific patterns for specific rating agencies? With our analysis, we shed light on the performance of rating agencies throughout turbulent market times and offer valuable insights for investors and regulators for future investment decisions. To our knowledge, we are the first to assess and benchmark the performance of rating agencies focusing on their monitoring activities after tranche issuance.

For the purpose of this paper we define multiple ratings as tranches, which are rated by more than one rating agency (double or triple ratings) as opposed to single rated tranches. We collected data from Standard & Poor's, Moody's, and Fitch on a total of 154'608 different U.S. residential mortgage-backed securities, which corresponds to a total issuance volume of about 7.51 trillion USD. Since the rating market for U.S. RMBS follows an oligopolistic market structure (dominated by three rating agencies), we were able to obtain the complete data set of all rated U.S. RMBS tranches, the corresponding characteristics, as well as monthly migration data between January 1985 and July 2012 for our empirical analysis.

Our findings confirm that multiple ratings are indeed of avail to investors: First, we find empirical proof that rating agencies demonstrate more effort with regard to their monitoring activities in the case of multiple ratings as compared to single rated tranches. Rating agencies publish more reports and comments and we find that it is 15.02% more likely that a rating agency becomes active as compared to a single rated tranche. For tranches which eventually entered into default, this probability increases substantially to 73.94% in the years between 2002 and 2006. Thus, investors get on average more information from each rating agency compared to in a less competitive situation (single rated tranches). Second, we show that rating agencies not only publish more, but also more accurate information in case of multiple ratings as average default prediction accuracy is significantly higher compared to single-rated tranches. Furthermore, we find that disagreement between rating agencies (as measured by numerical notch difference i.e. rating gap) increases over a tranche's lifetime. On average the predicted rating gap widens by a factor of about ten during the first three years after issuance (or 2.66 rating notches) and is accompanied by a drop in mean ratings of 5.45 notches, reflecting a decreasing average credit quality. With regard to the individual performance of rating agencies, we document that Moody's provides on average the most pessimistic credit assessment at issuance, a pattern that remains over a tranche's lifetime. The rating gap between Fitch and S&P is significantly smaller as compared to the rating gap observed between Moody's and Fitch or Moody's and S&P. Furthermore, we report a rather devastating result with regard to the overall performance of securitization transactions: Out of the 154'608 individual tranches, 49'022 (or 31.71% of the total sample) defaulted at one point in time, peaking at 74.05% for the tranches issued in 2007 (13'030 tranches out of 17'597 issued).

Based on these findings, our paper contributes to the literature on information production of credit ratings, most notably by extending the prevalent, rather static "at issuance"-perspective to the monitoring period following the initial debt placement. We find empirical proof for the information production hypothesis (e.g. Cantor & Packer, 1997) in a dynamic environment. We argue that existing empirical research, restricting its focus only to the point of issuance (e.g. Skreta & Veldkamp, 2009; Bolton et al., 2012) underestimates the motivation for rating shopping, which becomes more pronounced throughout a tranche's lifetime and peaks towards the end of the maturity structure as rating differences increase simultaneously.

Finally, we contribute to existing research on the individual patterns of rating agencies and extend it by displaying the relative performance of rating agencies to each other during the monitoring period (e.g. Livingston, 1999). Our results bring good news for investors: They benefit from multiple ratings due to increasing default accuracy. From a regulatory perspective, we provide empirical evidence that a multiplicity of ratings reduces information asymmetries and lowers overall industry opaqueness. Additional ratings indeed increase market transparency and regulators should therefore support initiatives to foster competition between rating agencies.

The remainder of this paper is organized as follows: Section 2 provides an overview of existing literature on information efficiency of multiple ratings and develops the hypotheses based on an extended framework linking multiple ratings to rating agencies' monitoring activities. Section 3 introduces the data sample and applied methodologies. The empirical results are presented in Section 4. Section 5 discusses potential alternative incentives for issuers to solicit multiple ratings and particularly addresses the concept of rating shopping. Our paper concludes with Section 6.

2. MULTIPLE CREDIT RATINGS AND MONITORING ACTIVITY

Asymmetric information is an important characteristic of the securitization market, where products exhibit complex architecture and information about the underlying credit portfolio is highly opaque. Three key market participants can be identified: issuer (or underwriter), investor and credit rating agency. The issuer structures the transaction via a special purpose vehicle (SPV) in order to sell tranches of different maturity and credit quality to investors. In this process, issuers mandate and pay credit rating agencies to assess the credit quality of each tranche and assign a corresponding credit rating. In reliance on this assessment, investors finally buy the tranches based on their individual risk preferences. As a result of this market structure, the balance of information is typically skewed towards the issuer of a security and the market information function of credit rating agencies constitutes a key factor as due diligence is usually delegated to them by investors (e.g. Diamond, 1984). Thus, the solicitation of rating agencies represents a form of agency costs to mitigate the information asymmetry in the principal-agent relationship between issuer (agent) and investor (principal). Given the assumption that credit ratings from different rating agencies can be considered as substitutes, efficient management of agency costs would imply to assign a single rating agency to assess the credit quality of a securitization tranche or deal to avoid duplication of effort. Reality, on the other hand, shows that about 72.2% of tranches in our sample are rated by more than one rating agency and that the fraction of multiple-rated tranches has increased substantially during the past decades.

There are two distinct aspects that can be associated with the market information function of credit ratings: At the time of issuance, rating agencies perform a *signaling function* to investors and regulators, assessing the credit quality of tranches' underlying portfolio of assets in order to help issuers selling their securitization tranches to investors. This delegation of monitoring by investors is not limited to the initial issuance of a tranche. Credit rating agencies rather maintain an important *monitoring function* and further act as agents on behalf of investors: They regularly evaluate the tranche's credit quality and adjust their ratings if necessary.[2] Such a review process can either result in an upgrade (in case the credit quality has improved), a downgrade (in case the credit quality has deteriorated) or result in no action (in case the credit quality has not changed at all). Rating agencies also publish so-called 'outlook reports', which incorporate a positive or negative outlook but do not trigger any rating events. The outcomes of these monitoring efforts are important information sources for investors, since it allows them to keep a check on their own risk-return balance and to control whether or not the tranches are still in line with the individual investment policies. Rating actions also have a severe impact on the price of tranches with downgrades (or upgrades) typically leading to a decrease (or increase). In short, the performance of rating agencies is not limited to the assessment of credit quality at the point of issuance, but remains important throughout the whole maturity of a securitization tranche.

Several studies from the corporate bond market investigate potential explanations for multiple ratings. In an early publication, Hsueh & Kidwell (1988) analyze why borrowers obtain more than one credit rating. Using a large sample of new-issue general obligation bonds sold between 1976 and 1983, they empirically test the impact of a municipality's decision to acquire a second rating for split- and non-split-rated bonds on new issue borrowing cost. The findings suggest that two credit ratings provide additional information and that split-rated bond issues account for reduced borrowing costs. Cantor & Packer (1995) explore whether the motivation for getting additional ratings is driven by regulatory considerations. However, they only find little evidence in support of their hypothesis. Only in the case of junk

[2] In the following, we will hence refer to the time between issuance of a security and the legal maturity date as the monitoring period.

bonds, the availability of a third opinion enables some borrowers to escape the speculative grade zone into investment grade territory. In a later paper, Cantor & Packer (1997) use issuer-level ratings from the year 1994 in an attempt to understand the motivation for obtaining a third rating. They consider information efficiency, rating shopping, and certification effects, but fail to find evidence that the use of a third rating is motivated by any of these considerations, although they demonstrate that the third rating is systematically more optimistic.

Inspired by the recent financial crisis and the allegations against credit rating institutions, several scholars put forward theoretical models on the rating shopping phenomenon. Rating shopping refers to an issuer's practice of engaging in a dialog with multiple rating agencies but mandating only those which offer the most favorable outcomes. In this process, issuers are in a constant exchange with rating agencies to optimize the transaction structure from their perspective, without revealing this dialog to investors. Skreta & Veldkamp (2009) present a framework where incentives for rating shopping increase as the complexity of the products increases. In one of the most recent publications, Bolton et al. (2012) focus on the conflicts of interest in credit rating agencies by modeling competition among agencies with three different sources of conflicts: (i) understating risk to attract business, (ii) issuers' ability to attract only the most favorable rating, and (iii) the overreliance on ratings by some investor clienteles. Based on their model, the authors conclude that competition can reduce information efficiency, as it facilitates rating shopping and that ratings are more likely to be inflated during booms or when investors are more trusting. Sangiorgi, Sokobin & Spatt (2009) develop a theoretical model of rating shopping and explore biases in ratings conditional upon heterogeneity across issuers in the extent to which different rating authorities agree.

Empirical evidence documenting this effect is, however, rather weak both in corporate finance and securitization markets. By focusing predominantly on ratings at issuance, empirical studies typically find only limited evidence that issuers engage in rating shopping behavior. Jewel & Livingston (1999) investigate whether ratings differ systematically across rating agencies. They examine a very large database with monthly observations of bonds and bond ratings over a five-year time period. The results show the average Fitch rating to be significantly better than Moody's and S&P ratings, but the effect disappears once they restrict their sample to bonds rated by all three rating agencies. However, Fitch ratings serve as a tiebreaker in cases where S&P and Moody's fail to reach consensus. The authors also examine whether rating shopping takes place but cannot find any evidence to support this hypothesis. Bongaerts et al. (2012) explore the economic role of credit rating agencies in the corporate bond market by considering three existing theories about multiple ratings: information production, rating shopping and regulatory certification. However, using differences in rating composition, default prediction and credit spread changes, their evidence only supports regulatory certification. The authors conclude that marginal, additional credit ratings are more likely to occur because of regulatory purposes, and seem to matter primarily for them, but do not seem to provide significant additional information related to credit quality. Bannier & Tyrell (2006) focus on reputation and competition among rating agencies. By proving that under certain conditions, public rating announcements and private information collection may be complements rather than substitutes, they argue that rating agencies may spark off a virtuous circle that increases information precision and raises market efficiency. The study also addresses the difference between solicited and unsolicited ratings and the problem of institutional investors.

Overall, recent literature provides only limited empirical evidence to explain the existence of multiple credit ratings. A potential explanation for the lack of compelling proof might be seen in the strong focus on the signaling function of credit ratings by prevalent literature. Relevant literature so far falls short of including the entirety of rating agencies' monitoring activities and is limited to a rather static "at issuance" perspective. Little is known of how ratings migrate after a particular tranche was issued and why issuers solicit and pay for multiple ratings not only at issuance but also during the monitoring phase of an asset's term. The main aim of this paper is therefore to extend the existing literature on multiple ratings by explaining its existence against the background of monitoring activities following the issuance of securitization tranches. In the following, we will focus our attention on the *information production hypothesis* (e.g. Cantor & Packer, 1997) as the prevailing theory in the established literature on the existence of multiple ratings and apply it to a dynamic framework.

In contrast to Bongaerts (2012), our first hypothesis is motivated not from the perspective of issuers, but from the investor's point of view. According to the information production hypothesis, more ratings reduce uncertainty about the underlying credit quality of a security. As investors are adverse to uncertainty, issuers may apply for additional ratings due to the demand for increased information production. Investors are interested in additional information, since it allows them to better assess the credit quality of the underlying debt instruments. Rating agencies may also apply different models or specialize in evaluating particular drivers of default and might thereby develop comparative advantages to justify their existence. Thus, the advantageous effect of rating agencies' different perspectives is expected to provide additional information on the uncertainty associated with credit quality and default probabilities. An additional rating in agreement with the existing rating would reduce credit quality uncertainty, whereas a difference in credit ratings might indicate a higher level of uncertainty, e.g. due to increased opaqueness of underlying assets. But does this argumentation really pay off for investors? Do additional ratings really lead to more and better information for investors?

In a first step, we argue that competition between rating agencies is more intense during the monitoring period in case of multiple ratings: Since their activities are directly benchmarked to their peers', rating agencies are induced to show more effort with regard to their monitoring obligations than observed for single-rated tranches. Consequently, we hypothesize that multiple ratings lead to more activity at the level of each individual rating agency. More activity or effort in turn can be interpreted as more information production:

Proposition 1: *Monitoring effort is higher for multiple-rated tranches compared to single-rated tranches.*

So far, we were only concerned about the amount of information being produced. However, more information does not necessarily correspond to better information. Investors delegate their monitoring activities to credit rating agencies in order to get a most accurate understanding of the underlying credit quality. Credit quality in turn is measured by default probabilities. Thus, we argue in the following that empirical evidence for the information production hypothesis should also be related to the level of accuracy achieved by the rating agencies. The ultimate measure to benchmark accuracy of rating agencies is of course how good they are in predicting the default of debt instruments, not only at issuance but also throughout the whole lifetime of a debt instrument. Thus, to assess default prediction accuracy of rating agencies it is important to include the complete monitoring period and continuously control for credit rating agencies' ability to assessing credit risk. It is not sufficient to only focus on the ratings at issuance, since they are normally based on one-year default probability and do not take the whole lifetime of a credit-linked instrument into account. In addition, rating processes in structured finance diverge significantly from those in the corporate bond market. Unique features are the limited accessibility of rating tools used by the agencies, the different methodologies used and the close cooperation between agency and issuer during the negotiation phase.[3] If multiple ratings come along with comparative advantages due to more disciplined behavior of rating agencies and heterogeneity of the individual rating processes or models, we argue that multiple ratings should indeed lead to higher accuracy of predicting the default of debt instruments:

Proposition 2: *Rating classification accuracy is higher for multiple-rated tranches compared to single-rated tranches.*

Moreover, we would expect this pattern to be consistent over the entire monitoring period and different tranche characteristics.

3. DATA

Our analysis is based on a dataset combining credit rating information from three different sources. The joint data comprises the complete daily long-term credit rating migration of residential mortgage-backed securities rated by Standard & Poor's, Moody's Investors Service, and Fitch Ratings. Standard & Poor's ratings are obtained through the S&P Credit Ratings database on Wharton Research Data Services (WRDS). The global database uses the combined information of S&P RatingsXpress and Compustat fundamental and market data, providing the credit rating migration of 205'670 RMBS tranches issued between 1977 and mid-2012. Moody's credit ratings are taken from the Structured Finance Default Risk Services database (SF-DRS) which covers the credit histories and material impairment of all Moody's-rated structured finance products issued since 1982. As of July 2012, the database includes historical changes in credit ratings of 94'216 RMBS tranches, segmented into 10'704 deals. Fitch credit ratings are provided by the Fitch Solutions Integrated Data Services (IDS) and include the global record of historical credit ratings since 1985 on both issuer- and tranche-level for the Fitch-rated structured finance universe offering detailed sub-level debt classification of each rated tranche. The available record up to mid-2012 comprises 79'305 RMBS tranches from 1'515 different originators.

Besides daily rating migration, numerous deal- and tranche-level characteristics are available for each data set. All three records commonly feature a number of tranche-level se-

[3] The latter has been heavily criticized in the recent past by politicians and regulatory authorities. We do not intend to discuss independency issues of rating agencies in this context; however, we proceed with the assumption that the relationship and exchange between rating agencies and issuers is very close and thereby impacts information efficiency.

curity identifiers[4] along with tranche name, original amount, asset type, domicile of assets, debt currency, issue launch date, legal maturity date, sub-industry classification, and information on credit enhancement or specialized financial structuring. Rating migration data include daily long- and short-term credit ratings and rating changes from the respective rating agencies, as well as watch list and rating outlook indications where applicable. Moreover, S&P and Fitch provide additional information on issuer- and entity-level for each tranche alongside with interest rates paid on debt obligations. Moody's record on the other hand includes debt classification according to tranche seniority, impairment calculations for tranches in default, and several deal-level characteristics such as unique deal identification, deal name, and original sale amount.

3.1 SAMPLE CONSTRUCTION

In a first step, we eliminate tranches with missing data, which cannot be complemented by information provided by either one of our data sources. In particular, we exclude tranches without a unique security identifier code such as CUSIP, ISIN, or CINS. In order to rule out currency- and country-specific effects, we limit our analysis to the U.S. market by clearing the data of all transactions which are not denominated in U.S. dollar, and we also drop deals whose majority of assets is not domiciled in the United States. We then match all the tranches of the records from S&P, Moody's, and Fitch based on available security identifier codes. This enables us to identify single-rated and multiple-rated (i.e. double- and triple-rated) tranches. We refer to the terms single-, double-, and triple-rated with respect to coverage by S&P, Moody's, and Fitch as these three rating agencies cumulatively account for about 91% of outstanding credit ratings of securitized assets in 2012[5]. However, we implicitly accept the possibility that some tranches might be rated by additional agencies which are not covered by our sample (NRSRO or non-NRSRO certified).

[Table 1 about here]

Concerning the matching of individual rating scales used by Moody's and Fitch, we refer to Table I, which outlines the mapping code of the individual alphanumerical rating classes on a numerical reference scale based on underlying one-year default probabilities (as re-

[4] Identifiers include CUSIP (Committee on Uniform Security Identification Procedures), ISIN (International Securities Identification Number), CINS (CUSIP International Numbering System), GVKEY (Compustat ID), and CIK (Central Index Key), among others.
[5] Annual Report on Nationally Recognized Statistical Rating Organizations, U.S. Securities and Exchange Commission (SEC), December 2012.

ported by S&P, Moody's and Fitch). This approach is commonly used in finance literature to be able to compare different rating scales (see e.g. Cantor & Packer, 1997; Jewell & Livingston, 1999). The matching on the lower end of the rating scale deserves some further explanation. Following the methodologies of S&P (2013), Moody's (2013), and Fitch (2013) the event of a default is defined as either (i) a missed or delayed disbursement of a contractually-obligated interest or principal payment (excluding missed payments cured within a contractually allowed grace period), as defined in credit agreements and indentures; or (ii) a situation where the issuer has entered into bankruptcy filings, administration, receivership, liquidation or other formal winding-up procedure, or such a situation is believed to be inevitable based on the rating agencies' opinion. This definition corresponds to a credit rating of C on Moody's global long-term rating scale and a rating of D on S&P and Fitch's international credit rating scale, respectively. However, a closer examination of near-to-default tranches rated by multiple agencies indicates that downgrades to C on behalf of Moody's (where C is the lowest rating category) rather corresponds to S&P and Fitch downgrades to C (their second lowest rating category) than to their actual default rating D. We thus account for differences in the practice of assigning ratings to indicate default by considering a tranche to be in default if it has been flagged by a rating below Ca on the Moody's rating scale or an equivalent CC on the S&P and Fitch rating scales. From the individual numerical ratings we calculate the notch difference which enables us to identify divergence in the credit quality assessment of multiple-rated tranches, which, given the longitudinal nature of the sample, can be analyzed on a time-continuous basis.

3.2 DESCRIPTIVE SAMPLE STATISTICS

A first overview on the scope of our sample is given in Figure 1. The chart reports the number of RMBS tranches for which an outstanding rating from S&P, Moody's, and/or Fitch is available, segmented according to year. It impressively illustrates the rapid growth of the RMBS market in the U.S., especially during the post-millennial period, before the outbreak of the subprime mortgage crisis led to a sudden collapse of the market in late 2007. Multiple ratings increasingly gained popularity although their growth has been stemmed in the most recent years of the sample. Whereas roughly one out of two tranches was rated by more than one rating agency in 1992, the share of single ratings has diminished to about 27% in the following decade and only constitutes a mere 24% of outstanding tranches in 2012.

[Figure 1 about here]

Monthly cross-sectional mean rating levels are shown in Figures 2 and 3. Not surprisingly, the picture is dominated by the collapse of the RMBS market in late 2007. Average credit ratings, which have remained constantly on high levels for the past fifteen years, lost substantial ground and fell over twelve rating notches after finally beginning to level off in 2010. Differences in agency specific rating records are generally small, yet S&P seems to be slightly more optimistic about the credit quality of its mandated tranches across the sample period. On the other hand, mean tranche ratings differ considerably with respect to the number of assigned ratings as shown in Figure 3. Single-rated tranches experienced a continuous deterioration of mean rating levels and appear to be relatively more conservative compared to multiple ratings, which remained particularly optimistic throughout the expansion of the structured finance market. In addition, multiple-rated tranches have been more severely downgraded in the aftermath of the subprime crisis. While single ratings have lost on average about 5 notches between 2008 and 2010, mean multiple ratings deteriorated by roughly twice that much. In fact, the pattern points towards the assumption that the majority of multiple ratings were systematically overoptimistic, and that the subprime crisis forced rating agencies to correct exaggerated expectations. In a recent study, Efing & Hau (2013) support this view by demonstrating that credit ratings were biased towards issuer clients that provide the agencies with more rating business. Overall, the considerable difference in rating levels with respect to the single/multiple rating dichotomy demand for a more detailed investigation and will be further discussed in the empirical analysis.

[Figures 2 and 3 about here]

A more nuanced picture of the final sample is given in Table 2 which reports selected tranche characteristics for a number of subsamples, particularly for each rating agency and for single-, double, and triple-rated tranches. The combined record comprises a total of 154’608 tranches, of which 42’668 (27.6%) are single-rated, 91’118 (58.9%) carry double ratings, and 20’822 (13.5%) have ratings assigned from all three rating agencies. As in the corporate bond market, S&P constitutes the largest share of outstanding credit ratings, providing credit statements for about 81.2% of tranches in the sample. Moody's is solicited for 57.5% of tranches while Fitch is still contracted with almost one out of two tranches. In general, tranche maturities do not appear to vary substantially across subsamples, although single-rated tranches tend to be of shorter maturity, compared to other subsamples. With an average of $72 million, issuance volumes are particularly high for senior securities while subordinate tranches on average only amount to $13 million. Additionally, issuance volumes seems to be positively correlated with the number of assigned ratings, suggesting that the size of a tranche might play a role in the issuer's decision, whether or not to solicit multiple ratings. This appears reasonable, as economies of scale allow allocating the costs for additional ratings to a wider asset base. Both maturity and volume are important factors for debt instruments and will be taken into account in the empirical analysis.

[Table 2 about here]

An interesting peculiarity of multiple ratings is that they allow for a direct comparison between different rating agencies as they refer to the same debt instrument. Due to differences in length of the credit quality assessment process, minor time lags might occasionally arise in our sample between the initial publications of ratings for multiple-rated tranches. In these cases we adjust for this effect by calculating the respective means over the first date on which the ratings of all involved rating agencies are publicly available. However, the resulting variations are negligible, as time lags typically do not exceed the period of three months. Among double ratings, rating differences between S&P and Fitch are close to zero whereas Moody's ratings at issuance appear to be significantly more conservative.[6] The ratings of triple-rated tranches seem to confirm this pattern. Moreover, the presence of a third rating agency coincides with an even stronger diverging opinion of Moody's, while the assessment of S&P and Fitch remains roughly consistent. The relatively low market share of Moody's for single-rated tranches might further indicate that investors know about this conservative attitude and thus refrain from soliciting Moody's as a sole provider of credit opinions. Since Moody's tends to be the most conservative rating agency, this underrepresentation in terms of single-rated tranches is not surprising. We will address this notion in more detail in the empirical analysis. With respect to defaults, the statistics show a strong relationship between the number of assigned ratings and the rate of tranches which received a rating of C or below during the sample period. A possible approach to explain this observation may be grounded on the thought that issuers rely on additional ratings particularly for low-quality assets in an attempt to con-

[6] Deviations in ratings issued by Moody's might at least partially be explained by different methodologies in determining the overall creditworthiness of an instrument. Whereas S&P and Fitch ratings are based on probability of default (PD), Moody's credit models are based on expected loss (EL), reflecting both the likelihood of default and expected financial losses in the event of a default (loss given default). As indicated by Peretyatkin & Perraudin (2002), ratings based on EL may therefore ceteris paribus be more favorable to large senior tranches than a PD approach, and less favorable towards more junior tranches that tend to be of smaller size.

vince potential investors. This argumentation is supported by the fact that the proportion of multiple-rated tranches increased towards the run-up of the financial crisis, which was accompanied by a decreasing quality of the assets being securitized.

4. EMPIRICAL RESULTS

To measure the effects of additional ratings on agency effort and the coherence of multiple ratings over time, we rely on fixed-effects multiple linear regression analysis, while differences in classification accuracy of credit ratings are quantified using receiver operating characteristic. Both applications are widely accepted measures in financial literature and also commonly used in the context of multiple ratings (e.g. Jewell & Livingston, 1999; Guettler & Kraemer, 2008: Bongaerts et al., 2012). We address the hypotheses formulated in Section 2 by individually defining the identification strategy and methodology for each of the two research question, followed by a discussion of the empirical results.

4.1 DO MULTIPLE RATINGS FOSTER MONITORING EFFORT?

In order to determine the degree of rating and revision effort on behalf of rating agencies for single- and multiple-rated tranches, we use rating activity on tranche level as a proxy for agency effort. We quantify rating activity as the number of reviews in credit ratings (upgrade, downgrade, confirmed) and rating outlook (positive, negative, stable) for each tranche over a specific period of time. Frequently, a periodic credit assessment does not lead to a change in rating or outlook designation but the current estimates are confirmed by a more recent rating date. Accordingly, we also include rating and outlook confirmations as they, alike actual rating changes, provide evidence for revision effort on behalf of rating agencies. Table 3 provides a first overview on rating activity and reports mean number of rating actions along with standard deviation, minimum and maximum values for single and multiple ratings, rating agencies and different years. It is not surprising that the number of rating actions by S&P, Moody's, and Fitch about tripled from 2007 to 2008, when yearly rating activity typically assumed maximum values, shortly after the crisis began to reveal its entirety. In general, rating effort appears to be higher for multiple-rated tranches compared to single ratings as the corresponding test statistics are significant at the 0.01 confidence level, for the most part. Moreover, the effect appears to emerge even more clearly during the crisis period, when especially S&P and Moody's, but also Fitch reinforce their activities, particularly among multiple-rated tranches.

[Table 3 about here]

In the following we perform a multivariate regression analysis in order to assess the impact of multiple ratings on overall monitoring activity in the presence of covariates. Accordingly, we set up a linear regression equation and specify the covariates to be included in the model.

$$E_{i,t} = \alpha_{Issuer} + \alpha_{Year} + \alpha_{Vintage} + \beta_1 Mult_i + \beta_2 S\&P_{i,t} + \beta_3 Moody's_i + \beta_4 Fitch_i \quad (1)$$
$$+ \beta_5 Default_i + \beta_6 Rating_{i,t} + \beta_7 TTM_{i,T-t} + \beta_8 Size_i + \beta_9 Collateral_i + \varepsilon_i$$

The dependent variable $E_{i,t}$ is *Revision Effort* and indicates the intensity of monitoring activity. It is defined as the total number of credit rating (and rating outlook) reviews performed for each tranche i in year t. We employ issuer fixed-effects to control for unobserved heterogeneity at the issuer level. As related research observes considerable differences in delinquency rates of residential mortgages with respect to loan vintage (e.g. Demyanyk & Van Hemert, 2011) we also include dummy variables for each calendar year α_{Year} and tranche vintage year $\alpha_{Vintage}$ to control for time-varying heterogeneity in credit quality and more precisely identify the causal effect of multiple ratings on monitoring effort. The main explanatory variable *Multiple Ratings* ($Mult_i$) is dichotomous and coded one if tranche i is rated by more than one rating agency and zero otherwise. $S\&P_{i,t}$, $Moody's_{i,t}$, and $Fitch_{i,t}$ are dichotomous variables coded 1 if tranche i is rated by the respective rating agency at time t and 0 otherwise. We account for the notion that tranches on the path to default might be under close scrutiny of rating agencies and hence are likely to be subject to higher monitoring effort compared to tranches whose credit quality is still unabated. *Ex-post Performance* ($Default_i$) controls for this effect by indicating whether a tranche experiences default at some point in time (1), or not (0). On a related note, rating activity tends to be higher among tranches at the lower end of the rating scale. We therefore also include the (mean) numerical *Tranche Rating* ($Rating_{i,t}$) at the end of year t to control for individual tranche credit quality. In addition, the frequency of rating and outlook revision typically decreases as tranches approach final maturity (T). Thus, we let *Time To Maturity* ($TTM_{i,T-t}$) denote remaining tranche lifetime in months defined as $T - t$. *Size* refers to the natural logarithm of the tranche's original amount, denominated in U.S. Dollar and captures differences in revision effort related to tranche size. Finally, *Collateral*$_i$ is a zero-one variable and captures differences in the opaqueness of the underlying pool of assets by distinguishing assets backed by prime-rate borrowers (1) from those with underlying mortgages of inferior credit quality such as subprime or Alt-A papers[7] (0). There is no formal definition of prime or subprime borrowers but according to industry standards, subprime borrowers were historically defined as having a FICO score[8] below 640, although this has varied over time and circumstances (Lo, 2012). We do not include a variable related to tranche subordination since this effect is to a large extent already captured by the numerical tranche rating (*Rating*$_{i,t}$) and inclusion of such does not significantly alter the goodness-of-fit statistics or coefficients of our explanatory variables. The error term is assumed to be normally distributed with $\varepsilon_i \sim N(0, \sigma^2)$. Yet, we relax the assumption of independent and identical distribution and account for potential clusters on tranche level by cluster-robust standard errors, which are a clustered version of Huber-White sandwich estimators.

[Table 4 about here]

The results are presented in Table 4. In Column (1), we run a first regression on rating effort, considering only rating upgrades, downgrades, and confirmations. Column (2) presents full sample results of monitoring activity including both rating and outlook review effort. As the dependent variable captures the number of tranche reviews in a given year, we can interpret the coefficient as follows: for a one unit change in the predictor variable, the response variable is expected to change by the respective regression coefficient, given the other predictor variables are held constant. Thus, the estimated coefficient for *Multiple Ratings*$_i$ predicts an increase in the number of rating reviews per year of about 15.02% if a tranche is rated by more than one rating agency, compared to single ratings. However, the effect in the total sample reduces to 1.66% if rating outlook reviews are taken into account. In Column 3 we run the regression over a subsample, including only tranches which eventually defaulted at some point in time, and further segment the sample into a five-year pre-crisis period (Column 4) and a crisis period of equal length (Column 5). In line with previous results, the subsample regressions display significantly higher relative rating activities for multiple-rated tranches and the effect even appears to be substantially larger for tranches on the path to default. In

[7] Alternative A-paper is a type of U.S. mortgage that is considered to be riskier than A-paper, or "prime", and less risky than "subprime".
[8] Originally created by the Fair Isaac Corporation, the FICO score is a type of credit score that uses mathematical models and takes into account various factors on payment history, current level of indebtedness, types of credit used, length of credit history, and new credit, to quantify the overall credit risk of an applicant. A person's FICO score will range between 300 and 850. Scores above 650 indicate good credit history, while individuals with scores below 620 may often find it difficult to obtain financing at a favorable rate.

particular, this is the case during the pre-crisis period when estimated annual monitoring effort is about 73.94% higher for multiple-rated tranches, and almost five times higher compared to the crisis period. When comparing relative levels of monitoring effort of each rating agency, the results point towards a highly consistent pattern: relative rating revision activity is typically highest for Fitch, while Moody's appears to be the least active player in terms of monitoring effort, although only slightly surpassed by S&P.

We perform several robustness checks in order to evaluate the sensitivity of our results towards different model specifications. First, we run a number of additional regressions on the same regression equation as our original model, but with slightly altered subsample constraints, e.g. we allow the length of the observation period to vary for both, the pre-crisis and crisis period. In general, the analysis confirms the robustness of our results towards different model specifications. Variations with respect to our main variables of interest are negligible and obtained coefficients are in line with those of the original models in Table 4.

[Table 5 about here]

From the baseline model in Column 1 (Table 4) we then calculate a series of margin post-estimations to get a more nuanced view on the effect of multiple ratings on monitoring effort. Margins are frequently used in Epidemiology and Biostatistics, but recently also in Economics[9] as an informative means for summarizing how the value of a response variable is related to changes in a particular covariate or combination of such, holding the remaining variables at their means. The margins of predicted number of rating reviews per annum (p.a.) for different subsamples and selected time periods are presented in Table 5 and graphical illustrations, alongside with 95% confidence intervals are provided in Figures 3.1 and 3.2. Several peculiarities deserve to be highlighted in this context. In general, monitoring effort on behalf of rating agencies varies considerably with respect to a tranche's ex-post performance, vintage year, and, to a lesser extent, also calendar year. As expected, the number of rating reviews depends on whether a tranche defaulted at some point in time, or not. It is reasonable to conclude that tranches on the path to default are under closer scrutiny and hence have their ratings revised more frequently. When comparing different levels of revision effort among single- and multiple-rated tranches in the subsamples of defaults and non-defaults, we can observe that effort is considerably higher for multiple ratings across relevant calendar years. For

[9] For a detailed discussion of margins and marginal effects, in particular the distinction between average marginal effects (AME) and marginal effects at the mean (MEM) see e.g. Long (1997), Long & Freese (2005), Bartus (2005), or Cameron & Trivedi (2010).

example in 2008, non-defaulting tranches received on average 0.28 reviews p.a. if they were rated by one agency, whereas multiple-rated tranches have been reviewed 0.81 times. This is equivalent to the interpretation that on average 28.37% of single-rated tranches have been reviewed once p.a. in contrast to 81.17% of multiple-rated tranches. The same pattern holds for defaulted tranches during the same period, where monitoring effort was 0.85 reviews p.a. (single ratings) and 1.30 reviews p.a. (multiple ratings), respectively. Figure 3.1 confirms the robustness of these results by showing that confidence intervals for all factor combinations do not overlap at any point in time. The analysis also reveals that monitoring effort is highly sensitive to tranche vintage, i.e. the year a tranche is issued. Between 2007 and 2011, tranches issued after 2005 received about three times more rating reviews compared to tranches issued before 2000. Subsample comparisons for each vintage group again indicate that multiple-rated tranches are under closer due diligence and subject to roughly twice as many rating reviews p.a. compared to single-rated tranches. Figure 3.2 reinforces the systematic pattern and displays the statistical significance of these results. Overall monitoring effort appears to be highest for multiple-rated tranches of more recent vintages while single-rated tranches of older vintages receive the least attention in terms of monitoring activity.

[Figures 3.1 and 3.2 about here]

With respect to Proposition 1, the above results allow us to reject the null hypothesis that rating effort does not vary with respect to the number of assigned ratings, and not reject the alternative hypothesis for all sample periods and subsample analyses. The findings provide strong empirical evidence that rating agencies at least partially condition the level of rating and outlook revision effort on competition, i.e. the availability of peer ratings. We further observe lower revision activity in the run-up, and higher activity during and after the subprime crisis. A possible interpretation of this result might be seen in the competitive pressure on rating agencies during tightening market conditions, when credit ratings are typically under very close scrutiny by regulators and investors. In such an environment, reputational concerns might be a crucial factor in rating agencies' effort to correctly determine an instrument's inherent credit risk. The threat of below-average performance, together with the fact that multiple-rated tranches allow for direct benchmarking with peers, creates a strong incentive for rating agencies to increase monitoring effort on this particular set of tranches even further. Our findings are in line with Efing & Hau (2013) in a sense that relatively lower revision rates in the pre-crisis period support the view that issuer-friendly ratings, which presumably received little attention as long as the market thrived, were particularly prevalent in this period.

4.2 DO MULTIPLE RATINGS INCREASE RATING ACCURACY?

In the previous section, we focused on rating agencies' degree of effort in monitoring credit quality of single- and multiple-rated tranches and different market environments, and observed increased revision effort during and in the aftermath of the subprime mortgage crisis. As a consequence, we argue that higher levels of revision effort on behalf of rating agencies should ultimately be reflected in their overall rating performance. More precisely, one should be able to observe an improvement in rating agencies' ability to correctly discriminate between different levels of credit risk.

The classification accuracy of credit ratings can be described by the receiver operating characteristic (ROC) curve, which is a plot of the true positive rate (TPR), i.e. the proportion of actually defaulted tranches which are ex-ante correctly identified as such, versus the false positive rate (FPR), i.e. the proportion of non-defaults which are falsely identified as defaults. Equivalently, the ROC curve can be represented as the cumulative distribution function of the case marker observations (defaults), standardized with respect to the control distribution (non-defaults). We account for covariates that affect the distribution of the marker among controls. Following Janes & Pepe (2008, 2009) we employ a covariate-adjusted measure of classification accuracy called the covariate-adjusted ROC curve, or the *A*ROC. We then compare the summary measures of the *A*ROC curves for single- and multiple-rated tranches to identify differences in the predictive power of credit ratings for the two subsamples. As the number of tranche defaults is particularly low in the pre-crisis period, we restrict our analysis to the crisis and post-crisis sample (2007 to 2011) where we have sufficient coverage of events to obtain statistically meaningful results.

As a first step, we fit a logistic regression model which is estimated using maximum likelihood, and specify how the covariates act on the distribution of the marker, i.e. the rating classification.

$$D_{i,t} = \alpha_{Vintage} + \beta_0 + \beta_1 Rating_{i,t} + \beta_2 Effort_{i,t} + \beta_3 Mult_i + \beta_4 TTM_{i,T-t} + \beta_5 Size_i + \beta_6 Collateral_i + \varepsilon_i \qquad (2)$$

The binary outcome variable $D_{i,t}$ assumes 1 if a tranche i has experienced a default event during the observation period t and 0 otherwise. It can thus also be interpreted as the default probability with respect to t. The marker, or classification variable *Tranche Rating* ($Rating_{i,t}$) is the numerical tranche rating at the beginning of period t and ranges from 1 (best rating) to 21 (worst rating/default). A number of covariates account for effects that potentially affect the distribution of the marker among controls. *Monitoring Effort* ($Effort_{i,t}$) captures tranche differences in the number of rating reviews and is defined as the number of rating reviews that occur between inception at t_0 and the beginning of observation period at time t, divided by elapsed tranche lifetime (in months). In this respect, we create a dynamic measure of agencies' monitoring effort where the effect of each additional rating review becomes less pronounced over time. *Multiple Ratings* (*Mult*) is binary and indicates whether a tranche is rated by multiple rating agencies. As we expect ratings of tranches with longer time to maturity to be more prone to uncertainty, we include *Time to Maturity* (*TTM*), defined as remaining tranche lifetime in months ($T - t$) to control for this effect. The remaining covariates remain the same as in equation (1) where they are described in more detail: *Size* refers to the natural logarithm of tranche original issuance amount, denominated in U.S. Dollar, and *Collateral* (*Coll*) is a dichotomous variable and coded 1 if a tranche's assets are backed by prime-rate borrowers, and 0 otherwise. The distribution of the random error, ε, is estimated empirically by using the residuals from the linear model. The stratified measure of rating performance is then defined as

$$A\mathrm{ROC}(f) = P(1 - PV_{DZ} \leq f) \tag{3}$$

where PV stands for percentile value, and $PV_{DZ} = F_Z(Y_{DZ})$ represents the case observation with the covariate value $Z(Y_{DZ})$ standardized with respect to the control population with the same value of Z. Accordingly, we estimate F_Z, the distribution of the marker in controls as a function of Z. That is, for each case subject i we calculate the PV:

$$\widehat{PV}_{DZ_i} = \hat{F}\{(Y - \widehat{\beta_0} - \widehat{\beta_1}Mult_i - \widehat{\beta_2}Size_i - \widehat{\beta_3}Coll_i)/\hat{\sigma}\} \tag{4}$$

where $\widehat{\beta_0}$, $\widehat{\beta_1}$, $\widehat{\beta_2}$, $\widehat{\beta_3}$, and $\hat{\sigma}$ represent estimates from the logit model. The cumulative distribution function of the estimated case percentile values is estimated empirically. We then calculate the area under the AROC curve, $A\mathrm{AUC} = \int_0^1 A\mathrm{ROC}(f)df$, which is given by

$$\widehat{A\mathrm{AUC}} = \sum_{i=1}^{n_D} \widehat{PV}_{DZ_i}/n_D \tag{5}$$

The AAUC estimate is the sample average of the case standardized marker values, where the sum is over the n_D case observations and can be interpreted as the probability that, for a random case and control marker observation with the same covariate value, the case observation is higher than the control. Accordingly, the angle bisecting line, also called reference line, represents a random model where the marker provides no discriminatory power in ex ante distinguishing between case and control observation. We calculate the classification accuracy for each rating agency and different prediction periods, using credit ratings at the beginning of the respective period as classification variables. We let the binary outcome variable assume one if a tranche experiences a default in a given period and zero otherwise. The variable thus captures a rating agency's capability to predict default up to the duration of the prediction period.

[Figures 4.1 to 4.3 about here]

Figures 4.1, 4.2 and 4.3 provide a graphical illustration of the estimated AROC curves for a one-year prediction period starting in the beginning of 2008. One can see that classification accuracy varies with respect to the number of assigned rating agencies. More precisely, the precision of rating schemes appears to increase with additional rating agencies joining the monitoring process. The corresponding AAUCs for each rating agency and different prediction periods are reported in Table 6. ΔAAUC captures the difference in classification accuracy for single- and multiple-rated tranches and results from subtracting the AUC of single-rated tranches from the AUC of multiple-rated tranches. A Wald statistic is obtained by dividing the observed difference by its standard error and compared to the standard normal distribution to obtain a p-value. Generally, the findings indicate significantly higher classification accuracy for multiple-rated tranches, particularly among S&P and Fitch ratings where ΔAAUC displays strong statistical significance in the crisis and post-crisis period. The results in the Moody's panel are less conclusive. The predictive power of Moody's single ratings dominates those of multiples in 2009 and 2010, but is consistent with S&P and Fitch in earlier periods. In addition, S&P dominates its peers in terms of classification accuracy among multiple-rated tranches in the outbreak of the crisis in 2007 and 2008, but is outperformed by Fitch in the following years. Moody's ability to discriminate credit quality is typically below peer level for multiples, yet displays strong performance among single-rated tranches.

[Table 6 about here]

Overall, the analysis of receiver operating characteristics suggests a predominantly positive impact of multiple ratings on rating agencies' ability to correctly classify tranches with respect to credit risk. Based on Proposition 2, we can thus reject the null hypothesis that classification accuracy does not vary with respect to number of assigned ratings, and not reject the alternative hypothesis for Fitch and S&P. In addition, the results of our *A*ROC analysis emphasize a steady decrease in the explanatory power of credit ratings towards the end of the sample period. Relating these findings to those of the previous section provides a strong argument that the observed increase in rating and revision effort on behalf of rating agencies since the market turmoil in late 2007 ultimately might have manifested in higher classification accuracy of credit quality among multiple-rated tranches. The economic implications of these results are significant. Total defaults among single-rated tranches amounted to 26.44% in our sample. In dollar terms, this corresponds to roughly 110 USDbn estimated potential losses to investors in this segment. More disciplined behavior on behalf of rating agencies among single-rated tranches could indeed have contributed to a more timely and accurate prediction of these losses up to several years ahead.

5. DISAGREEMENT AMONG RATING AGENCIES

As outlined in Section 2, related literature brought forth a number of additional incentives which serve as potential motivation for issuers to solicit multiple ratings. Most notable in this context are incentives related to overcome specific regulatory certification hurdles, and the *rating shopping hypothesis* (e.g. Bongaerts et al., 2012). Empirical evidence around multiple ratings in connection with regulatory certification is typically weak (e.g. Cantor & Packer, 1995/1997). However, Bongaerts et al. (2012) conclude that marginal, additional credit ratings are more likely to occur because of, and seem to matter primarily for, regulatory purposes. As a result of the disruptive events during the recent financial crisis, the regulatory framework around structured finance assets has been, and still is, subject to a major overhaul. While entering the depths of the regulatory guidelines issued by relevant supervisory authorities is beyond the scope of this paper, reflections around the concept of rating shopping deserve some further attention.

Since the abandonment of the subscription approach by Moody's, Fitch, and finally S&P in the early 1970s, the rating market for securitization transactions is traditionally characterized by solicitation, meaning that the issuer selects and pays involved rating agencies for evaluating a security's credit quality. Such an issuer-pays model creates a range of conflicts of interests between market participants and might trigger a behavior that is referred to in literature as rating shopping (e.g. Jewell & Livingston, 1999). Under the rating shopping hypothesis, issuers 'shop' for additional ratings in the hope of improving their rating or meeting regulatory certification standards. According to this theory, rating shopping can emerge when rating agencies do not perfectly agree or there is increased uncertainty about an instrument's credit quality. In this case, issuers, who have additional, private information about the tranches' credit quality, can seek to maximize their average rating by soliciting multiple bids. In exchange for a small break-up fee, issuers can keep an already solicited credit rating confidential as they own the publication rights for solicited ratings (Mählmann, 2008).

In contrast to the information production hypothesis, the incentive to obtain multiple ratings is now motivated solely from the issuer's perspective: Issuers benefit from a very good credit rating, because investors' return expectations are a function of the underlying credit quality. The higher the credit risk, the higher the interest rates demanded by potential investors. Thus, issuers have, by definition, a very strong incentive to get c. p. the best possible rating for the debt instruments they intend to sell to investors. We further assume that investors assess the yield they demand on the basis of the average credit ratings in case of more than one credit rating. Accordingly, issuers have an interest in additional (and better) ratings, because it will lead to lower refinancing costs. On the other hand, investment policies, particularly of large institutional investors, may constrain rating shopping as they quite often demand more than one rating. In such cases, the issuer is required to publish a second rating in addition to the (most favorable) rating which would reflect his rational choice.

In general, we see two main arguments which motivate issuers to engage in rating shopping. On the one hand, split ratings at the time of issuance might incentivize issuers to solicit rating agencies which provide a more optimistic credit assessment than their competitors. This well-documented effect has been the primary focus of recent literature related to the rating shopping phenomenon. In addition, we argue that rating shopping is also driven by issuer's expectations about relative future rating migration. We therefore extend the existing theory to a more dynamic environment taking into account the monitoring period after issuance.

Two potential scenarios may exist: (i) Rating agencies are indeed trading in favorable ratings at issuance for being solicited by the issuer (e.g. AAA instead of AA). Throughout the post-issuance period, rating agencies are expected to adjust their credit assessment to reflect the fair credit quality. The average mean rating would come down and in case of split ratings, we would also expect that the rating gap would become smaller and even vanish overtime completely. The cost of rating shopping is borne by the investors in the form of lower yields at issuance. (ii) Structural differences between rating processes exist that cause the rating gap to remain stable or even increase over a tranche's lifetime, meaning that the full (dis)advantage of rating shopping becomes evident over time. These structural differences can, for example, result from different estimation processes for recovery rates or applied mathematical concepts (e.g. note that S&P and Fitch use an approach based on probability of default, whereas Moody's use an expected loss concept). In this case, investors have to bear the (increasing) cost of rating shopping throughout the whole maturity structure. As existing literature only focuses on rating shopping at tranche issuance (e.g. Skreta & Veldkamp, 2009), it may actually underestimate the real costs of rating shopping that is borne by investors. In this section, we do not focus on actual refinancing costs but rather focus on how potential incentives for rating shopping evolve across the maturity structure, during the period after tranche issuance.

In the following, we aim to shed light on the motives for rating shopping on behalf of issuer clients, by focusing on the monitoring function of multiple credit ratings. Multiple-rated tranches provide a unique opportunity to directly compare credit assessment and subsequent monitoring of different rating agencies on the same asset. We argue that, if systematic dispersions of credit ratings are persistent in the long run as described in scenario (ii), the related literature actually underestimates issuer's incentives for rating shopping. In this case, such long-run considerations might even outweigh marginal differences in ratings at the time of issuance. In order to investigate systematic differences in credit ratings of multiple-rated tranches, we rearrange our sample according to a tranche-term perspective. More precisely, we define a discrete variable *Tranche Age* ($Age_{i,t}$) indicating the current age of tranche i, which is defined as the months between issue launch date and time of observation in month t. We then assemble the individual end-of-month observations according to a term perspective, where 0 represents the month of issuance.

[Figures 5.1 to 5.3 about here]

Figures 5.1, 5.2, and 5.3 show the resulting rating term curves for multiple-rated tranches and pairwise combinations of rating agencies. Two observations deserve particular attention: First, numerical notch differences of multiple ratings appear to increase with respect to tranche age for all pairings of rating agencies, but particularly for S&P/Moody's-rated tranches. In general, ratings seem to diverge stronger for agency combinations involving S&P, while Fitch ratings tend to accord stronger with Moody's, but also display systematic dispersion over time. Second, the direct comparison of multiple ratings suggests a more conservative credit assessment on behalf of Moody's as opposed to S&P and Fitch. Consistent with the analysis of multiple ratings at issuance, it confirms the persistence of Moody's relative conservativeness in the long run and might serve as a potential explanation why only few issuers solicit Moody's on a standalone basis. The systematic and time-persistent differences in rating levels provide a strong motivation for issuer clients to engage in rating shopping activities. Issuers might be tempted to exploit the inconsistent credit assessments of different rating agencies in order to maximize the rating of their securities. The distribution of solicited single ratings in Table 2 supports this view. For example, Moody's share of single-rated tranches of 5.57% is much lower compared to single ratings in the S&P (21.86%) and Fitch (14.09%) portfolios, although Moody's market share in structured finance ratings is substantially higher than Fitch's. Indeed, this might be explained by Moody's relative conservatism in the credit rating process. As we can see from Figures 5.1 and 5.3, whereas rating differences at inception are almost negligible, S&P and Fitch ratings start to become relatively less pessimistic with tranche age. The same observation holds also for Moody's peers, i.e. the relative conservativeness of each rating agency is consistent with its respective market share of single-rated tranches. Hence, these figures suggests that issuers, to some extent, know about the relative differences in ratings from S&P, Moody's, and Fitch, and make use of this information in the solicitation process. That is, they tend to refrain from soliciting a conservative rating agency if that rating will end up being the only one which is publicly available. Instead, they frequently publish conservative ratings in combination with credit opinions from a more optimistic rating agency in order to avoid the adverse effects of inferior ratings.

In a next step, we confirm these findings by regressing rating notch difference against tranche age in a multivariate setup. The dependent variable $G_{i,t}$ denotes the *Rating Gap*, i.e.

the absolute value of the numerical notch difference between two ratings of tranche i in month t. We take account of temporal variation in the dependent variable related to the year of tranche issuance include year-fixed effects $\alpha_{Vintage}$. The main explanatory variable *Tranche Age* ($Age_{i,t}$) indicates the age of tranche i at time t (in months). We include an additional binary variable *Triple Rating* ($Triple_i$) to distinguish between double- (0) and triple-rated tranches (1). We do not directly consider numerical credit ratings, as given the setup, each observation involves multiple ratings. But we account for the overall level of seniority by means of a zero-one variable *Seniority* ($Senior_i$), coded one for senior tranches, and zero for AAA-subordination. Further, we employ the same control variables for remaining time to maturity, tranche size, and collateral quality as in previous models. Thus, we obtain the following regression equation:

$$G_{i,t} = \alpha_{Vintage} + \beta_1 Age_{i,t} + \beta_2 Triple_i + \beta_3 Senior_i + \beta_4 TTM_i + \beta_5 Size_i + \beta_6 Collateral_i + \varepsilon_i \qquad (6)$$

As we model a count variable, i.e. a variable that take more than two values and all of the values are integers, we employ a fixed-effects negative binominal regression model which is estimated using maximum likelihood. We also run the regression in a multiple linear setup. Both models yield very similar results and particularly the variation with respect to our main variables of interest is negligible. We therefore adhere to the negative binominal model, as the standard for modelling count variables as it is more robust with respect to assumptions on the underlying distribution of the dependent variable, which does not display the properties of a normal distribution. Cluster-robust standard errors account for potential clusters on tranche level.

Table 7 reports the results of the fixed-effects negative binominal regression for pairwise combinations of multiple-rated tranches. We run regressions for each combination of rating agencies for the total period (1985-2012) and a reduced pre-crisis sample (1985-2006). Coefficients for tranche age are positive and highly significant for all regressions, and appear to be higher for the total period. Although the effect is also clearly present in the pre-crisis sample, the most recent years seem to have strongly reinforced the divergence of multiple credit ratings. Interestingly, the zero-one variable for triple ratings is also highly significant, suggesting that the rating gap for any agency pairings is larger when a third rating agency is involved. From the perspective of this analysis, a third rating thus appears to increase the level of uncertainty among the first and second agency about a tranche's credit quality, rather than decrease it. Moreover, gap levels are highly affected by tranche seniority. Predictor variables for subordination are negative and highly significant and indicate that rating gaps are considerably larger for mezzanine and subordinate tranches. This might be due to the fact that uncertainty in assessing credit quality is generally higher among junior tranches of a deal, as they are the first to absorb the losses should home owners be unable to pay back their mortgages. This increased level of uncertainty might additionally aggravate the diverging tendency between different ratings of multiple-rated tranches.

[Table 6.1 about here]

As a further robustness check, we calculate a series of marginal effects based on equation (6) to investigate how changes in the response variable are related to changes in a particular covariate. At this point, it is important to distinguish between the average marginal effect (AME) and the marginal effect at the mean (MEM). The former refers to the computation of each observation's marginal effect with respect to an explanatory factor, averaged over the estimation sample. In contrast, MEM measures change in the response while holding all other variables at their means. Current practice tends to favor the use of AME for several reasons.[10] In accordance with these concerns, we follow the established methodology of computing average marginal effects and synonymously refer to AME when we discuss marginal effects in the remainder of this paper.

The computation of marginal effects is different for discrete (i.e. categorical) and continuous variables and, in the context of linear statistical models, also varies in terms of interpretation. With binary independent variables, marginal effects measure discrete change, i.e. the average change in the expected value of the response variable, in our case $G_{i,t}$, if one independent variable changes from 0 to 1, holding all other variables constant. That is, for a categorical variable Z_k the AME is

$$AME_k = \frac{1}{n}\sum_{x=1}^{n}[F(\beta Z^x | Z_k^x = 1) - F(\beta Z^x | Z_k^x = 0)] \tag{7}$$

where βZ^x denotes the value of the linear combination of parameters and variables for the xth observation and $F(\cdot)$ is the cumulative distribution function that maps the values of βZ^x to

[10] For example, MEM are not good approximations of AME, computed as means of marginal effects evaluated at each observations, if some of the parameter estimates are large. But issues also arise in terms of interpretability. Notably, it is generally viewed to be problematic to evaluate marginal effects at means of dummy variables since means of dummies refer to nonexistent observations (e.g. Bartus, 2005).

the [0, 1] interval. On the other hand, marginal effects for continuous variables measure the instantaneous rate of change and provide an approximation to the amount of change in $G_{i,t}$ that will be produced by a 1-unit change in Z_k.[11] In this case, researchers typically estimate the effect of an infinitely small change. Let $f(\cdot)$ be the derivative of $F(\cdot)$ with respect to βZ. The AME of the continuous variable Z_k is then given by

$$AME_k = \beta_k \frac{1}{n} \sum_{x=1}^{n} f(\beta Z^x) \tag{8}$$

The AMEs for selected values of tranche age in combination with binary variables for triple ratings and seniority for the total sample are depicted in Table 6.2. The instantaneous rate of change in the rating gap shows a positive, nonlinear correlation with tranche age. In other words, not only do rating gaps become larger, but the rate of change also increases with tranche age. For example, the expected monthly change in the rating gap for S&P/Fitch-rated tranches increases from 0.016 notches at a tranche age of 0.5 years to 0.235 notches three years after issuance. Change rates for S&P/Moody's and Moody's/Fitch combinations increase slightly more moderately during the same period from 0.025 to 0.196 notches, and from 0.019 to 0.139 notches, respectively. The post estimation of marginal effects also confirms the effects of a third rating, and subordination on the rating gap. We observe systematically higher AMEs for the subsamples of triple-rated tranches for combinations including a rating from Fitch, but also the difference in rating gaps between S&P and Moody’s is moderately significant with respect to triple ratings. For mezzanine and subordinate tranches, the sensitivities of age on the rating gap are even higher.

[Table 8 about here]

We further compute AMEs for selected vintages. Age effects on the predicted rating gap three years after issuance are on average about three to four times larger for tranches which have been issued after 2002. In fact, tranches of the 2007 vintage series were equally prone to age-depending changes in rating gaps as the average subordinate tranche in the sam-

[11] A potential issue may arise with continuous variables in the sense that there is no guarantee that a bigger increase in Z_k would produce an increase in the response variable equal to the increase in Z_k times the instantaneous rate of change. This is because the relationship between Z_k and the response variable is nonlinear. Yet, when Z_k is measured in small units, the effect of an increase in Z_k by unity may match up well with the marginal effect for Z_k. However, in response to the fact that the presentation of a single marginal effect for each covariate may or may not be informative in assessing the effect of changes on the response variable, Long (1997) and others suggest to examine adjusted predictions across a range of discrete values for one or more covariates (continuous or discrete). That is, we can look at the effects of discrete changes in categorical and continuous variables simultaneously, in order to get a more nuanced picture of the impact of covariates on the response variable.

ple. Overall, the marginal effects on predicted rating gap are less pronounced when computed over the pre-crisis subsample. Rating gaps tend to increase linearly with tranche age but the effects lack statistical significance. However, there is a significant difference in rating gaps with respect to triple ratings, subordination, and, to a limited extent, also tranche vintage.

In summary, the time series analysis of multiple ratings provides suggestive empirical evidence that structural differences in rating methodologies are reflected by diverging credit ratings of mortgage-backed assets, which in turn constitute a strong incentive for issuers to engage in rating shopping. In addition, the findings suggest that the effects of distinct tranche characteristics on the rating gap of multiple-rated tranches, although veritably existing before the subprime mortgage crisis, have been amplified to a large extent by it.

6. CONCLUSION

This study explores potential reasons for the existence of multiple credit ratings in the securitization market. Based on the complete rating migration for U.S. RMBS transactions, and thus avoiding a potential selection bias, we analyze potential incentives to obtain multiple ratings, induced by credit agencies' monitoring behavior following the issuance of securitized assets. In the solicited market for mortgage-backed securities, investors heavily depend on rating agencies to better understand the complex transaction structures. The recent financial crisis has highlighted the role of RMBS transactions as well as the importance of rating agencies quite precisely in this context. We supplement existing research on multiple ratings by focusing on incremental information produced by additional credit ratings. So far, existing research was limited to an at-issuance perspective, neglecting the monitoring period after the debt issuance. Our results hold good news for investors: In a competitive situation of multiple ratings outstanding, we observe that the rating effort of each individual rating agency is increasing, leading to more information being produced. In a subsequent step, we document that multiple ratings not only lead to more but also to better information: Default prediction accuracy increases with the number of outstanding ratings. Thus, we empirically support the information production hypothesis and extend it to a dynamic framework. The economic implication of these results is that more disciplined behavior on behalf of rating agencies among single-rated tranches could have contributed to a more timely and accurate prediction of about 110 USDbn of potential losses to investors.

Manifested by an increasing rating gap, we further find that disagreement among rating agencies widens over a tranche’s lifetime. Direct comparisons of multiple ratings suggest a more conservative credit assessment on behalf of Moody's as opposed to S&P and Fitch throughout the whole monitoring period. Consistent with the analysis of multiple ratings at issuance, this confirms the persistence of Moody's relative conservatism in the long run and might serve as a potential explanation of why only a few issuers solicit Moody's on a standalone basis: Moody's market share of all single-rated tranches adds up to only 11.62%, far lower compared to the agency's total market share in structured assets of 32.76%.[12] These quantitative and qualitative differences in methods and models applied, as well as individual assessment, appear to be structurally dominant over a tranche's lifetime. We thus conclude that rating shopping is not only motivated by split ratings at issuance, but also by an issuer's expectations about relative future rating migration, a notion not considered in the prevalent literature on multiple ratings. However, we do not find empirical evidence that rating agencies are trading favorable ratings in exchange for being mandated by an issuer and subsequently revising their overly optimistic assessments in the monitoring period.

[12] Annual Report on Nationally Recognized Statistical Rating Organizations, U.S. Securities and Exchange Commission (SEC), December 2012, p. 6.

REFERENCES

Bannier, C. & Tyrell, M. (2006). Modelling the Role of Credit Rating Agencies – Do They Spark Off a Virtuous Cycle? Finance and Accounting Working Paper No 160, Goethe-Universität Frankfurt/M.

Bartus, T. (2005). Estimation of marginal effects using magreff. *The Stata Journal*, 5 (3), pp. 309-329.

Bolton, P., Freixas, X. & Shapiro, J. (2012). The Credit Ratings Game. *The Journal of Finance*, 67 (1), pp. 85-111.

Bongaerts, D., Cremers, K.J.M. & Goetzmann, W.N. (2012). Tiebreaker: Certification and Multiple Credit Ratings. *The Journal of Finance*, 67 (1), pp. 113-152.

Cameron, A.C. & Trivedi, P.K. (2010). Microeconometrics using Stata, Revised Edition. StataCorp LP.

Cantor, R. & Packer, F. (1995). The Credit Rating Industry. *Journal of Fixed Income*, 5 (3), pp. 10-34.

Cantor, R. & Packer, F. (1997). Differences of opinion and selection bias in the credit rating industry. *Journal of Banking & Finance*, 26 (10), pp. 1395-1417.

Diamond, D.W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51 (3), pp. 393-414.

Demyanyk, Y. & Van Hemert, O. (2011). Understanding the Subprime Mortgage Crisis. *Review of Financial Studies*, 24 (6), pp. 1848-1880.

Efing, M. & Hau, H. (2013). Structured Debt Ratings: Evidence on Conflicts of Interest. *Journal of Financial Economics*, forthcoming.

Guettler, A. & Kraemer, W. (2008). On Comparing the Accuracy of Default Predictions in the Rating Industry. *Empirical Economics*, 34 (2), pp. 343-356.

Hsueh, L.P. & Kidwell, D.S. (1988). Bond Ratings: Are Two Better Than One? *Financial Management*, 17 (1), pp. 46-53.

Janes, H., & M.S. Pepe (2008). Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: An old concept in a new setting. *American Journal of Epidemiology*, 168 (1), pp. 89-97.

Janes, H. & Pepe, M.S. (2009). Adjusting for covariate effects on classification accuracy using the covariate-adjusted ROC curve. *Biometrika*, 96 (2), pp. 371-382.

Jewell, J. & Livingston, M. (1999). A Comparison of Bond Ratings from Moody's S&P and Fitch IBCA. *Financial Markets, Institutions & Instruments*, 8 (4), pp 1-45.

Lo, A.W. (2012). Reading about the Financial Crisis: A Twenty-One-Book-Review. *Journal of Economic Literature*, 50 (1), pp. 151-178.

Long, J.S. (1997). Regression Models for Categorical and Limited Dependent Variables. In: Advanced Quantitative Techniques in the Social Sciences Number 7. Sage Publications.

Long, J.S. & Freese, J. (2005). Regression Models for Categorical Outcomes using Stata, Second Edition. College Station, TX: Stata Press.

Lugo, S., Croce, A. & Faff, R. (2014). Herding Behavior and Rating Convergence among Credit Rating Agencies: Evidence from the Subprime Crisis. Review of Finance Advanced Access, pp. 1-29.

Mählmann, T. (2008). Rating agencies and the role of rating publication rights. *Journal of Banking and Finance*, 32 (11), pp. 2414-2422.

Morkoetter, S. & Westerfeld, S. (2009). Do Multiple CDO Ratings Impact Credit Spreads? *The Journal of Fixed Income*, 19 (1), pp. 55-72.

Sangiorgi, F., Sokobin, J. & Spatt, C. (2009). Credit-Rating Shopping, Selection and the Equilibrium Structure of Ratings. NBER Working Paper No. 14712.

Skreta, V. & Veldkamp, L. (2009). Ratings shopping and asset complexity: A theory of ratings inflation. *Journal of Monetary Economics*, 56 (5), pp 678-695.

Table 1: Numerical Rating Mapping Code

This table shows the numerical mapping code which is applied to the individual global long-term credit rating scales based on underlying one-year default probabilities reported by S&P, Moody's and Fitch.

Rating Code	S&P Long-term Rating Class	Moody's Long-term Rating Class	Fitch Long-term Rating Class
1	AAA	Aaa	AAA
2	AA+	Aa1	AA+
3	AA	Aa2	AA
4	AA-	Aa3	AA-
5	A+	A1	A+
6	A	A2	A
7	A-	A3	A-
8	BBB+	Baa1	BBB+
9	BBB	Baa2	BBB
10	BBB-	Baa3	BBB-
11	BB+	Ba1	BB+
12	BB	Ba2	BB
13	BB-	Ba3	BB-
14	B+	B1	B+
15	B	B2	B
16	B-	B3	B-
17	CCC+	Caa1	CCC+
18	CCC	Caa2	CCC
19	CCC-	Caa3	CCC-
20	CC	Ca	CC
21	C	C	C
21	D		D / DD / DDD

Table 2: Summary Statistics

This table shows summary statistics and selected tranche characteristics of all available U.S. Dollar-denominated RMBS tranches issued between January 1985 and July 2012. The sample only includes securities whose underlying assets are predominantly domiciled in the United States. S&P ratings are obtained from the S&P Credit Ratings database in WRDS, Moody's ratings are from the Structured Finance Default Risk Services database (SF-DRS), and Fitch ratings are provided by Fitch Solutions Integrated Data Services (IDS). Test statistics of equality of mean rating codes at issuance are obtained by a two-sample t–test.

Subsample	# of rated Tranches	in % of Total Sample	Mean Maturity (in years)	Mean Volume (in USDm)	Mean Rating Code at Issuance			# of Tranches in Default	in % of Sub-sample
					S&P	Moody's	Fitch		
Total Sample	154'608	100.00%	28.83	48.60	3.15	2.98	3.04	49'022	31.71%
Rating Agency									
S&P	125'498	81.17%	28.96	52.10	3.15	-	-	41'949	33.43%
Moody's	88'955	57.54%	29.05	59.20	-	2.98	-	31'855	35.81%
Fitch	72'917	47.16%	28.97	46.70	-	-	3.04	22'836	31.32%
Subordination									
Senior (AAA)	97'599	63.13%	28.34	72.10	1.00	1.05	1.01	14'942	15.31%
Subordinate	57'009	36.87%	29.68	13.30	7.06	6.56	7.36	34'080	59.78%
Collateral									
First Mortgage	85'457	55.27%	29.08	41.40	2.38	2.32	2.42	22'991	26.90%
Subprime Mortgage	35'246	22.80%	28.80	67.20	3.99	4.17	4.19	15'469	43.89%
Other	33'905	21.93%	28.23	45.20	3.93	2.70	3.56	10'562	31.15%
Single Rating									
Total	42'668	27.60%	28.08	27.70	-	-	-	11'281	26.44%
S&P	27'439	17.75%	27.89	27.70	4.98	-	-	7'171	26.13%
Moody's	4'957	3.21%	27.85	39.80	-	3.68	-	580	11.70%
Fitch	10'272	6.64%	28.87	20.90	-	-	7.14	3'530	34.37%
Double Rating									
Total	91'118	58.93%	29.01	52.70	2.45	2.73	1.87	27'864	30.58%
S&P / Moody's[a]	49'295	31.88%	29.26	62.20	2.79	2.92	-	18'435	37.40%
S&P / Fitch[b]	27'942	18.07%	29.05	41.60	1.83	-	1.84	6'466	23.14%
Moody's / Fitch[c]	13'881	8.98%	28.04	39.00	-	2.05	1.92	2'963	21.35%
Triple Rating									
S&P / Moody's / Fitch[d]	20'822	13.47%	29.52	69.50	3.36	3.57	3.35	9'877	47.44%

H_0: Mean Pairwise Notch Difference = 0 a, c, d) $Pr(|T| > |t|) < 0.001$ b) $Pr(|T| > |t|) < 0.1$

Table 3: Rating and Outlook Revision Effort

This table reports mean number of rating actions per tranche within a given year. Rating actions include rating revisions (upgrade, downgrade, confirmed) and outlook revisions (positive, negative, stable). Revision effort is further segmented according to rating agency and number of assigned ratings. Statistical significance levels for differences in means are reported as results of a two-sample t–test. ***, **, and * denote significance at the 1%, 5%, and 10% confidence level.

Year	Single-rated Tranches					Multiple-rated Tranches					Diff
	N	Mean (M_S)	SD	Min	Max	N	Mean (M_M)	SD	Min	Max	($M_M - M_S$)
Panel A: S&P											
2000	4'473	0.180	0.404	0	3	16'408	0.248	0.593	0	8	0.068***
2001	4'372	0.045	0.227	0	3	19'017	0.053	0.244	0	4	0.008*
2002	4'807	0.037	0.235	0	10	23'757	0.092	0.458	0	7	0.055***
2003	6'112	0.069	0.272	0	4	31'614	0.111	0.333	0	6	0.041***
2004	7'745	0.062	0.264	0	5	36'473	0.112	0.460	0	10	0.049***
2005	9'823	0.050	0.234	0	8	46'685	0.089	0.289	0	3	0.038***
2006	11'491	0.047	0.262	0	4	62'159	0.105	0.353	0	10	0.058***
2007	13'109	0.369	0.984	0	9	74'103	0.298	0.859	0	11	-0.071***
2008	13'357	0.895	1.625	0	10	72'791	1.099	1.931	0	10	0.204***
2009	14'132	1.249	1.484	0	9	70'830	1.695	1.778	0	8	0.446***
2010	15'684	0.618	1.065	0	6	68'714	0.756	1.255	0	6	0.138***
2011	15'348	0.475	0.866	0	8	67'424	0.673	1.296	0	8	0.198***
Panel B: Moody's											
2000	1'502	0.029	0.169	0	1	12'587	0.034	0.194	0	3	0.005
2001	1'802	0.001	0.024	0	1	15'111	0.043	0.203	0	1	0.043***
2002	2'186	0.012	0.108	0	1	18'702	0.133	0.356	0	4	0.121***
2003	2'122	0.057	0.270	0	2	23'787	0.178	0.405	0	2	0.121***
2004	1'642	0.112	0.444	0	2	28'750	0.253	0.477	0	3	0.141***
2005	1'405	0.157	0.584	0	4	40'390	0.279	0.478	0	6	0.122***
2006	1'335	0.072	0.330	0	3	56'161	0.162	0.400	0	4	0.090***
2007	1'399	0.084	0.320	0	2	65'850	0.261	0.579	0	6	0.178***
2008	1'358	0.588	0.967	0	5	64'844	0.819	1.098	0	8	0.231***
2009	1'333	0.638	0.674	0	3	63'189	0.895	0.893	0	5	0.257***
2010	1'305	0.270	0.516	0	2	61'295	0.839	0.835	0	4	0.569***
2011	1'179	0.417	0.556	0	2	60'611	0.220	0.444	0	4	-0.198***
Panel C: Fitch											
2000	4'101	0.049	0.334	0	6	16'808	0.036	0.372	0	6	-0.013**
2001	4'813	0.145	0.434	0	5	19'580	0.032	0.209	0	4	-0.112***
2002	5'041	0.548	0.584	0	5	22'768	0.101	0.496	0	7	-0.447***
2003	4'992	0.726	0.595	0	5	28'165	0.345	0.576	0	6	-0.381***
2004	3'178	0.467	0.633	0	5	26'968	0.338	0.547	0	5	-0.129***
2005	2'428	0.320	0.678	0	7	29'532	0.324	0.548	0	4	0.004
2006	3'237	0.384	0.734	0	10	36'963	0.499	0.589	0	5	0.114***
2007	4'155	0.703	0.983	0	9	44'229	0.847	0.973	0	9	0.144***
2008	4'173	2.744	2.326	0	10	43'958	2.121	1.978	0	11	-0.623***
2009	5'201	1.853	1.040	0	10	43'436	2.080	1.474	0	10	0.227***
2010	5'305	1.802	1.112	0	6	42'506	1.216	0.852	0	6	-0.586***
2011	5'197	3.129	1.134	0	8	41'686	2.598	2.039	0	10	-0.532***

Table 4: Monitoring Effort Regression Results

This table provides the results of the fixed-effects multiple linear regression for rating revision effort of single- and multiple-rated tranches. The dependent variable is *Rating Effort* and refers to the number of credit rating reviews (upgrade, downgrade, confirmed). In addition to rating revisions, *Total Effort* also includes rating outlook revisions (positive, negative, stable). (D) indicates that the regression is performed over a subsample including only tranches which eventually defaulted at some point in time. We employ fixed effects at issuer-, year-, and tranche vintage-level to control for unobserved heterogeneity. *S&P*, *Moody's*, and *Fitch* are dichotomous variables coded 1 if a tranche is rated by the respective rating agency at time *t* and 0 otherwise. All other control variables are defined in the main text. Cluster-robust standard errors are a clustered version of Huber-White sandwich estimators and account for potential clusters on tranche level.

Dependent: Review Effort	(1) Rating Effort	(2) Total Effort	(3) Rating Effort (D)	(4) 2002-2006 (D)	(5) 2007-2011 (D)
Multiple Ratings$_i$	0.1502***	0.0166***	0.3902***	0.7394***	0.1568***
	(0.00360)	(0.00549)	(0.00634)	(0.00863)	(0.01000)
S&P$_{i,t}$	0.1133***	0.3211***	0.2924***	-0.4774***	0.6299***
	(0.00323)	(0.00504)	(0.00658)	(0.01107)	(0.00945)
Moody's$_{i,t}$	0.0967***	0.2828***	0.1763***	-0.5779***	0.5440***
	(0.00286)	(0.00442)	(0.00566)	(0.00832)	(0.00981)
Fitch$_{i,t}$	0.7553***	1.0878***	1.1711***	0.0933***	1.6400***
	(0.00371)	(0.00568)	(0.00634)	(0.01032)	(0.00862)
Controls	Yes	Yes	Yes	Yes	Yes
Issuer-fixed effects	Yes	Yes	Yes	Yes	Yes
Year-fixed effects	Yes	Yes	Yes	Yes	Yes
Vintage-fixed effects	Yes	Yes	Yes	Yes	Yes
Method	MLR	MLR	MLR	MLR	MLR
McFadden's adj. R^2	0.480	0.426	0.491	0.519	0.451
Observations	981'702	981'702	333'503	60'699	224'993

Cluster-robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Predicted Number of Tranche Rating Reviews per Year

This table reports the margins of predicted number of rating reviews per annum (p.a.) calculated from the baseline model (1) in Table 4 for selected subsamples and sample periods. The Defaulted Tranches subsample includes only tranches which eventually experienced default at some point in time. Vintage subsamples are based on the year of tranche issuance. All additional model covariates are described in the main text. Standard errors are obtained by Delta method.

Subsample	Year (selected)	Single-rated Tranches		Multiple-rated Tranches	
		Reviews p.a.	Std. Err.	Reviews p.a.	Std. Err.
Non-defaulted Tranches	2000	0.0955***	5.98E-03	0.6235***	3.02E-03
	2002	0.1426***	6.03E-03	0.6706***	2.69E-03
	2004	0.1896***	6.19E-03	0.7176***	2.62E-03
	2006	0.2367***	6.46E-03	0.7647***	2.81E-03
	2008	0.2837***	6.82E-03	0.8117***	3.22E-03
	2010	0.3308***	7.27E-03	0.8588***	3.77E-03
Defaulted Tranches	2000	0.6644***	6.39E-03	1.1130***	6.69E-03
	2002	0.7114***	5.98E-03	1.1600***	6.27E-03
	2004	0.7585***	5.66E-03	1.2071***	5.93E-03
	2006	0.8055***	5.45E-03	1.2541***	5.70E-03
	2008	0.8526***	5.37E-03	1.3012***	5.60E-03
	2010	0.8996***	5.43E-03	1.3482***	5.61E-03
Vintages < 2000	2007	0.2145***	1.01E-02	0.5717***	8.38E-03
	2008	0.2380***	1.03E-02	0.5953***	8.57E-03
	2009	0.2615***	1.05E-02	0.6188***	8.77E-03
	2010	0.2850***	1.07E-02	0.6423***	8.99E-03
	2011	0.3086***	1.09E-02	0.6658***	9.22E-03
Vintages 2000-2005	2007	0.4861***	5.03E-03	0.9620***	4.54E-03
	2008	0.5096***	5.05E-03	0.9855***	4.56E-03
	2009	0.5332***	5.11E-03	1.0090***	4.62E-03
	2010	0.5567***	5.21E-03	1.0325***	4.72E-03
	2011	0.5802***	5.33E-03	1.0561***	4.86E-03
Vintages > 2005	2007	0.8183***	1.66E-02	1.3369***	1.04E-02
	2008	0.8418***	1.66E-02	1.3605***	1.04E-02
	2009	0.8653***	1.66E-02	1.3840***	1.04E-02
	2010	0.8889***	1.66E-02	1.4075***	1.04E-02
	2011	0.9124***	1.66E-02	1.4310***	1.04E-02

Delta-method standard errors, *** p<0.01, ** p<0.05, * p<0.1

Table 6: Receiver Operating Characteristics (ROC) for Rating Classification Accuracy

This table provides the covariate-adjusted receiver operating characteristic (*A*ROC) estimation for different combinations of subsamples. In particular, the reported coefficients refer to the effect of multiple ratings on the area under the *A*ROC curve (*A*AUC) for each rating agency. The logistic model is fit using maximum likelihood estimation and the binary reference variable is *Default*, assuming 1 if a tranche is in default at the end of a given prediction period and 0 otherwise. The classification variable is *Tranche Rating*, denoting the numerical credit rating of the respective rating agency at the beginning of a given year. We employ fixed effects at issuer- and tranche vintage-level to control for unobserved heterogeneity. All other control variables are defined in the main text. Robust standard errors based on Huber-White sandwich estimators are reported in brackets. Statistical significance of Δ*A*AUC are obtained by a chi-squared test. ***, **, and * denote significance at the 1%, 5%, and 10% confidence level.

Rating Date	Prediction Period	# of Rated Tranches	# of Defaults	Single-rated Tranches				Multiple-rated Tranches				Δ*A*AUC
				Tranches	Defaults	$A\text{AUC}_{i,t}$	Std. Err.	Tranches	Defaults	$A\text{AUC}_{i,t}$	Std. Err.	
Panel A: S&P												
01/2007	1 Year	60'785	0.51%	9'042	0.66%	0.887	0.0133	51'743	0.49%	0.974	0.0029	0.087***
	2 Years	57'723	3.87%	8'745	6.00%	0.904	0.0053	48'978	3.49%	0.967	0.0015	0.063***
01/2008	1 Year	69'972	3.55%	10'507	5.44%	0.916	0.0044	59'465	3.21%	0.956	0.0020	0.040***
	2 Years	67'393	16.78%	10'460	24.46%	0.816	0.0044	56'933	15.37%	0.905	0.0016	0.089***
01/2009	1 Year	68'069	16.64%	11'122	23.16%	0.843	0.0038	56'947	15.37%	0.924	0.0013	0.081***
	2 Years	66'163	9.81%	10'946	11.41%	0.688	0.0076	55'217	9.50%	0.778	0.0028	0.089***
01/2010	1 Year	67'215	9.68%	11'967	10.55%	0.739	0.0065	55'248	9.49%	0.786	0.0024	0.047***
	2 Years	66'036	7.67%	11'846	6.50%	0.684	0.0084	54'190	7.93%	0.709	0.0030	0.026**
Panel B: Moody's												
01/2007	1 Year	46'300	1.68%	1'143	0.00%	-	-	45'157	1.73%	0.973	0.0016	-
	2 Years	44'619	10.82%	763	7.34%	0.857	0.0198	43'856	10.89%	0.945	0.0013	0.088***
01/2008	1 Year	53'541	12.28%	903	7.97%	0.862	0.0191	52'638	12.35%	0.936	0.0013	0.074***
	2 Years	51'729	16.68%	1'012	34.19%	0.903	0.0096	50'717	16.33%	0.807	0.0024	-0.095***
01/2009	1 Year	51'900	16.66%	1'030	34.17%	0.909	0.0093	50'870	16.30%	0.854	0.0020	-0.055***
	2 Years	49'374	10.09%	597	5.53%	0.879	0.0410	48'777	10.14%	0.747	0.0030	-0.132***
01/2010	1 Year	49'375	10.09%	597	5.53%	0.873	0.0306	48'778	10.14%	0.747	0.0030	-0.126***
	2 Years	48'920	4.36%	806	11.54%	0.877	0.0189	48'114	4.24%	0.886	0.0048	0.009
Panel C: Fitch												
01/2007	1 Year	30'751	2.03%	2'928	6.28%	0.952	0.0055	27'823	1.59%	0.953	0.0027	0.001
	2 Years	30'800	9.18%	2'970	30.37%	0.890	0.0061	27'830	6.92%	0.940	0.0020	0.050***
01/2008	1 Year	36'575	9.98%	3'803	33.45%	0.883	0.0057	32'772	7.26%	0.940	0.0019	0.057***
	2 Years	35'741	20.79%	3'878	36.41%	0.801	0.0073	31'863	18.89%	0.812	0.0025	0.012
01/2009	1 Year	35'906	20.74%	3'888	36.32%	0.795	0.0076	32'018	18.85%	0.876	0.0020	0.081***
	2 Years	35'087	12.69%	3'865	8.87%	0.825	0.0098	31'222	13.16%	0.725	0.0043	-0.100***
01/2010	1 Year	35'087	12.69%	3'865	8.87%	0.749	0.0132	31'222	13.16%	0.800	0.0032	0.051***
	2 Years	34'159	7.85%	3'692	3.17%	0.749	0.0215	30'467	8.42%	0.790	0.0045	0.041*

Table 7: Rating Gap of Multiple-rated Tranches

This table provides the results of the fixed-effects negative binominal regression for rating gaps of multiple-rated tranches. The dependent variable is the rating gap and refers to the absolute value of the numerical rating notch difference between two rating agencies at each point in time. *Tranche Age* is a continuous variable denominated in months and indicates the time since the tranche has been issued. *Triple Rating* is binary and assumes 1 if a tranche has an additional third rating, and 0 otherwise. *Seniority* is binary and distinguishes senior tranches (1), which are typically rated AAA at inception, from mezzanine and subordinate tranches (0). The remaining control variables are defined in the main text. Cluster-robust standard errors are a clustered version of Huber-White sandwich estimators and account for potential clusters on tranche level.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent: Rating Notch Gap	S&P - Moody's (Total)	S&P - Moody's (Pre-Crisis)	S&P - Fitch (Total)	S&P - Fitch (Pre-Crisis)	Moody's - Fitch (Total)	Moody's - Fitch (Pre-Crisis)
Tranche Age$_{i,t}$	0.067***	0.015***	0.093***	0.032***	0.066***	0.015***
	(0.0004)	(0.0007)	(0.0009)	(0.0013)	(0.0006)	(0.0009)
Triple Rating$_i$	0.033***	0.124***	0.118***	0.059	0.172***	0.547***
	(0.0117)	(0.0258)	(0.0201)	(0.0587)	(0.0191)	(0.0380)
Seniority$_i$	-1.170***	-2.478***	-1.556***	-4.265***	-1.393***	-2.304***
	(0.0139)	(0.0357)	(0.0261)	(0.1317)	(0.0244)	(0.0375)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Vintage-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Method	GLM	GLM	GLM	GLM	GLM	GLM
McFadden's adj. R^2	0.144	0.190	0.185	0.249	0.149	0.229
Observations	2'139'858	1'138'932	1'423'584	905'175	1'026'622	615'267

Cluster-robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 8: Average Marginal Effects of Tranche Age on pairwise Rating Gap

This table reports the average marginal effects (AME) calculated from the negative binominal regression model at selected values of the explanatory variables and averaging over the total sample (1985-2012). Additional model covariates are defined in the main text and include tranche size, remaining time to maturity, and a zero-one variable for collateral quality. The (-) indicates that the respective variables are held at their individual values for each observation.

Tranche Age (Months)	Triple Rating (1=Yes, 0=No)	Seniority (1=Senior, 0= Subrd.)	Vintage (Years)	Rating Gap S&P/Moody's		Rating Gap S&P/Fitch		Rating Gap Moody's/Fitch	
				AME	Std. Err.	AME	Std. Err.	AME	Std. Err.
6	-	-	-	0.025***	0.0001	0.016***	0.0006	0.019***	0.0001
12	-	-	-	0.038***	0.0002	0.028***	0.0010	0.028***	0.0002
18	-	-	-	0.057***	0.0004	0.048***	0.0018	0.042***	0.0004
24	-	-	-	0.086***	0.0008	0.081***	0.0033	0.062***	0.0008
30	-	-	-	0.130***	0.0015	0.138***	0.0061	0.093***	0.0015
36	-	-	-	0.196***	0.0028	0.235***	0.0111	0.139***	0.0027
6	1	-	-	0.026***	0.0003	0.026***	0.0012	0.023***	0.0002
12	1	-	-	0.039***	0.0004	0.044***	0.0022	0.035***	0.0003
18	1	-	-	0.059***	0.0007	0.075***	0.0038	0.052***	0.0006
24	1	-	-	0.090***	0.0011	0.127***	0.0068	0.078***	0.0011
30	1	-	-	0.135***	0.0019	0.216***	0.0122	0.117***	0.0020
36	1	-	-	0.204***	0.0033	0.368***	0.0219	0.174***	0.0035
6	0	-	-	0.024***	0.0002	0.008***	0.0002	0.011***	0.0002
12	0	-	-	0.037***	0.0003	0.014***	0.0003	0.016***	0.0003
18	0	-	-	0.056***	0.0005	0.025***	0.0004	0.025***	0.0005
24	0	-	-	0.084***	0.0009	0.042***	0.0008	0.037***	0.0007
30	0	-	-	0.127***	0.0016	0.071***	0.0016	0.055***	0.0012
36	0	-	-	0.192***	0.0029	0.121***	0.0030	0.082***	0.0020
6	-	1	-	0.013***	0.0001	0.009***	0.0003	0.011***	0.0001
12	-	1	-	0.020***	0.0002	0.015***	0.0004	0.017***	0.0002
18	-	1	-	0.030***	0.0003	0.025***	0.0008	0.025***	0.0003
24	-	1	-	0.045***	0.0005	0.042***	0.0014	0.038***	0.0006
30	-	1	-	0.068***	0.0009	0.072***	0.0026	0.057***	0.0010
36	-	1	-	0.102***	0.0016	0.122***	0.0048	0.085***	0.0018
6	-	0	-	0.041***	0.0003	0.034***	0.0014	0.030***	0.0003
12	-	0	-	0.062***	0.0004	0.057***	0.0025	0.045***	0.0005
18	-	0	-	0.093***	0.0008	0.097***	0.0044	0.067***	0.0009
24	-	0	-	0.140***	0.0015	0.166***	0.0080	0.100***	0.0016
30	-	0	-	0.212***	0.0027	0.282***	0.0144	0.150***	0.0028
36	-	0	-	0.320***	0.0048	0.480***	0.0260	0.224***	0.0048
36	-	-	<2002	0.047***	0.0010	0.040***	0.0009	0.066***	0.0017
36	-	-	2002	0.132***	0.0020	0.088***	0.0047	0.132***	0.0028
36	-	-	2003	0.148***	0.0022	0.142***	0.0037	0.147***	0.0030
36	-	-	2004	0.183***	0.0026	0.257***	0.0108	0.154***	0.0031
36	-	-	2005	0.207***	0.0029	0.316***	0.0107	0.160***	0.0032
36	-	-	2006	0.268***	0.0040	0.401***	0.0220	0.174***	0.0034
36	-	-	2007	0.301***	0.0047	0.489***	0.0426	0.199***	0.0043

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Figure 1: Outstanding U.S. RMBS Tranches

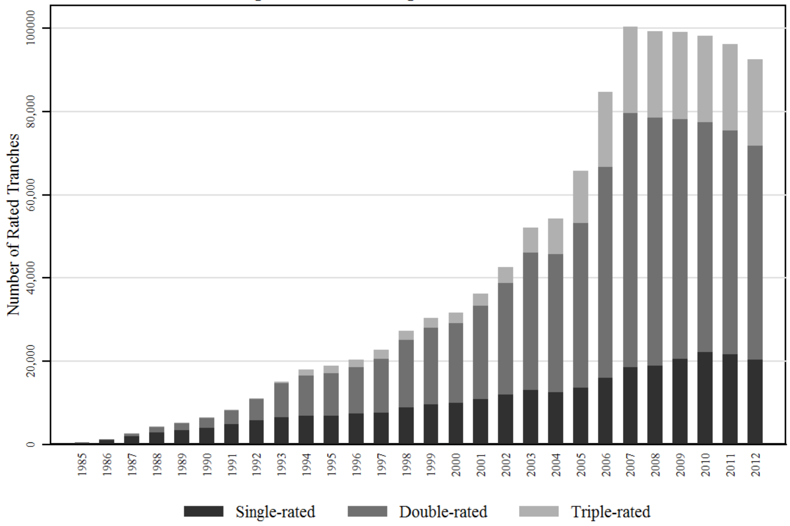

Figure 2.1: Mean Tranche Rating by Rating Agency

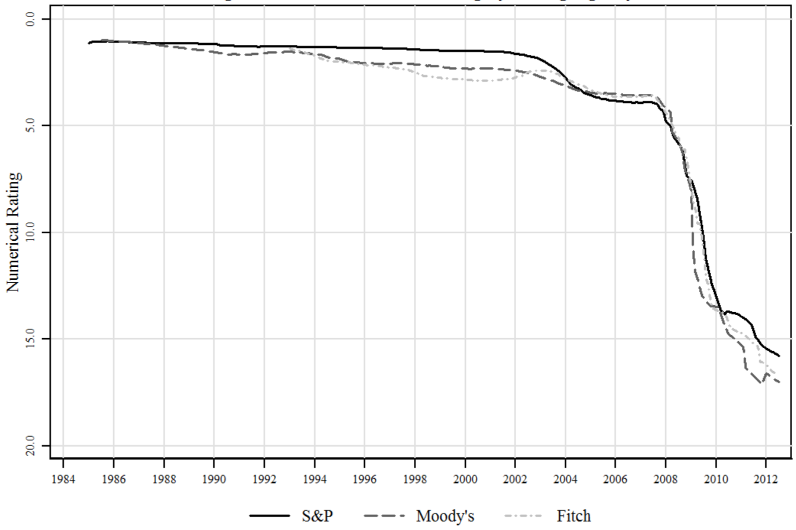

Figure 2.2: Mean Tranche Rating by # of Ratings

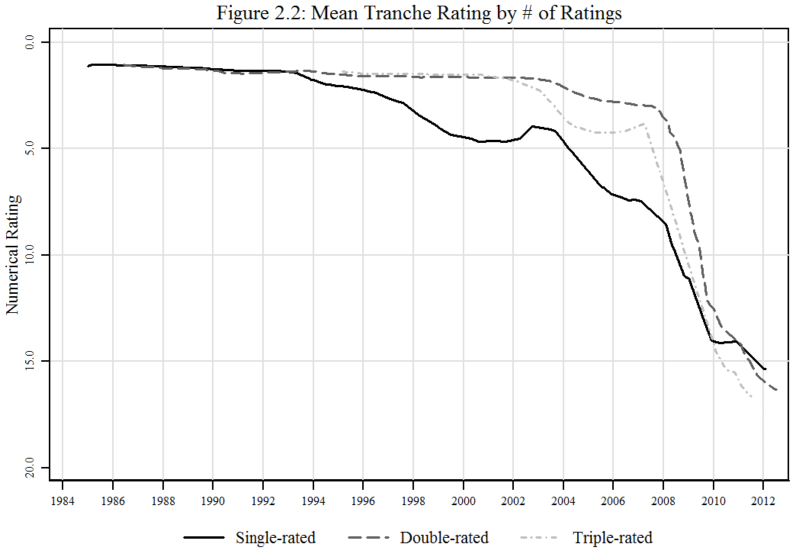

Figure 3.1: Predicted Number of Tranche Rating Reviews per Year

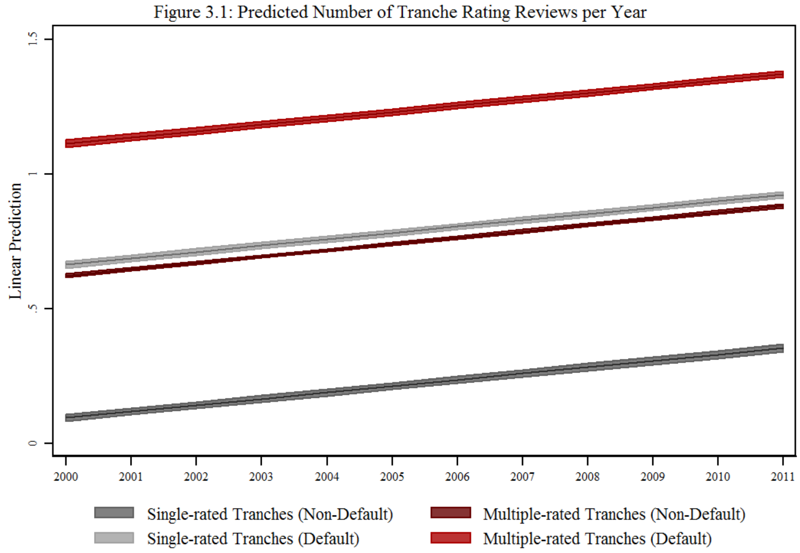

Figure 3.2: Predicted Number of Tranche Rating Reviews per Year

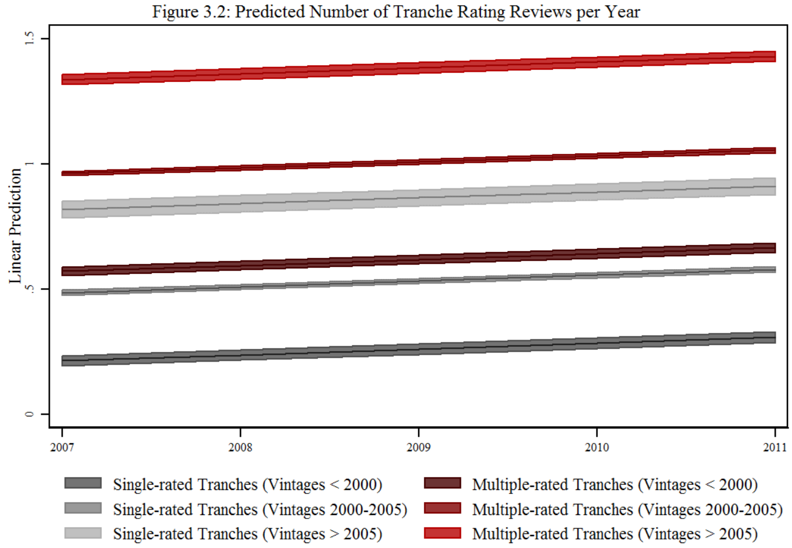

Figure 4.1: S&P ROC Curve for 1-Year Rating Accuracy (2008)

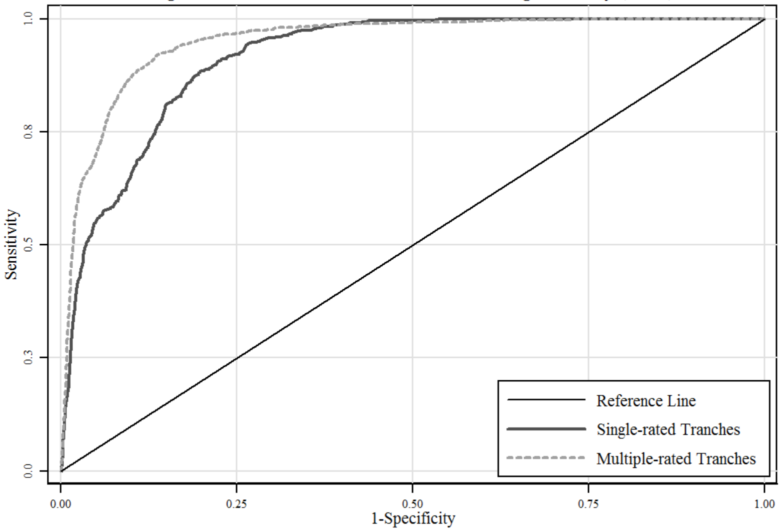

Figure 4.2: Moody's ROC Curve for 1-Year Rating Accuracy (2008)

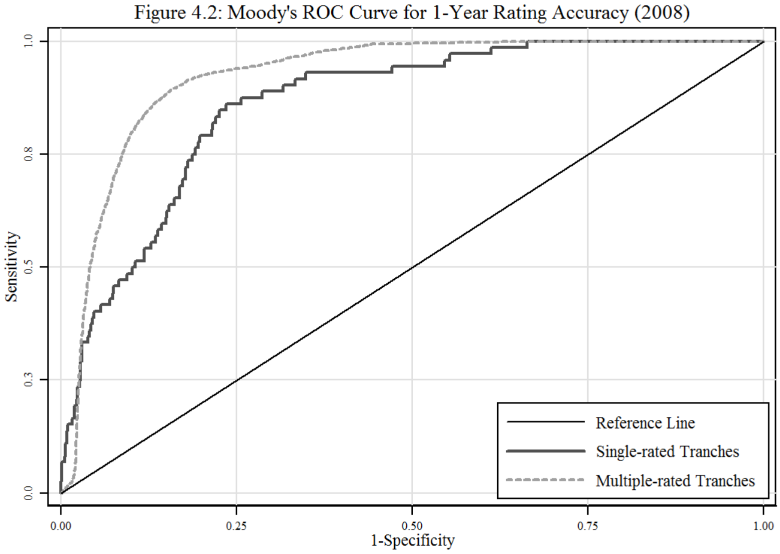

Figure 4.3: Fitch ROC Curve for 1-Year Rating Accuracy (2008)

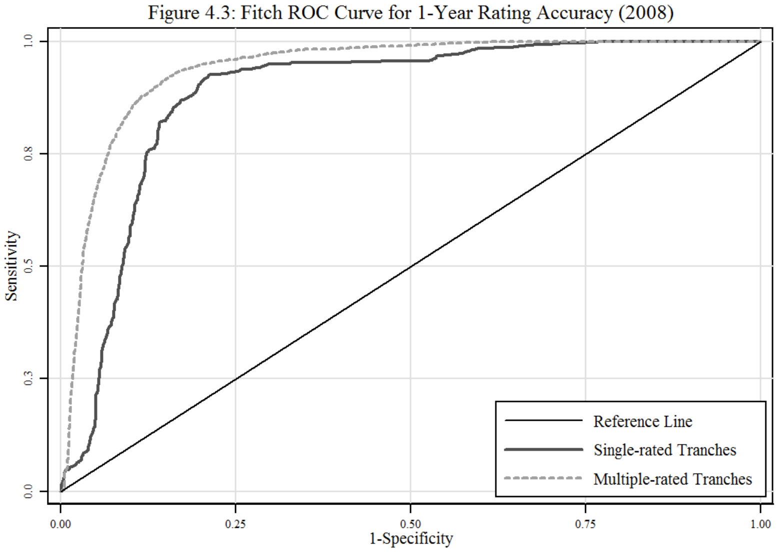

Figure 5.1: Term Curve of Senior Tranches Rated by S&P and Moody's

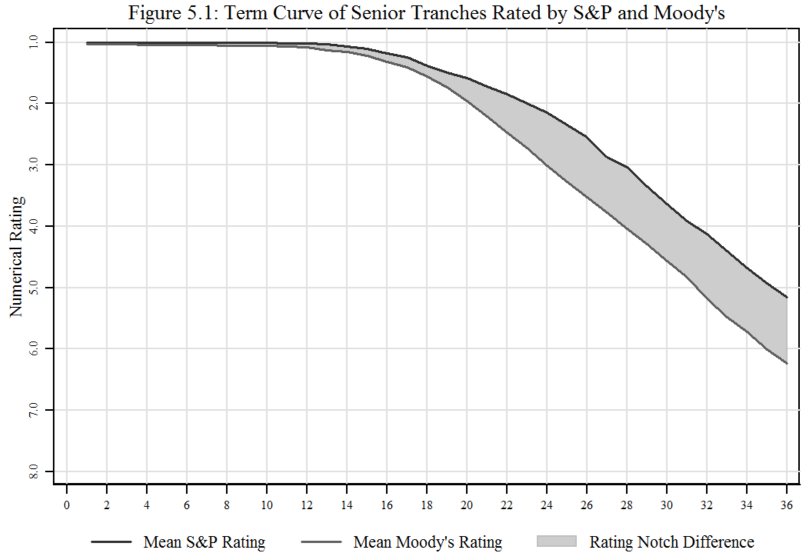

Figure 5.2: Term Curve of Senior Tranches Rated by S&P and Fitch

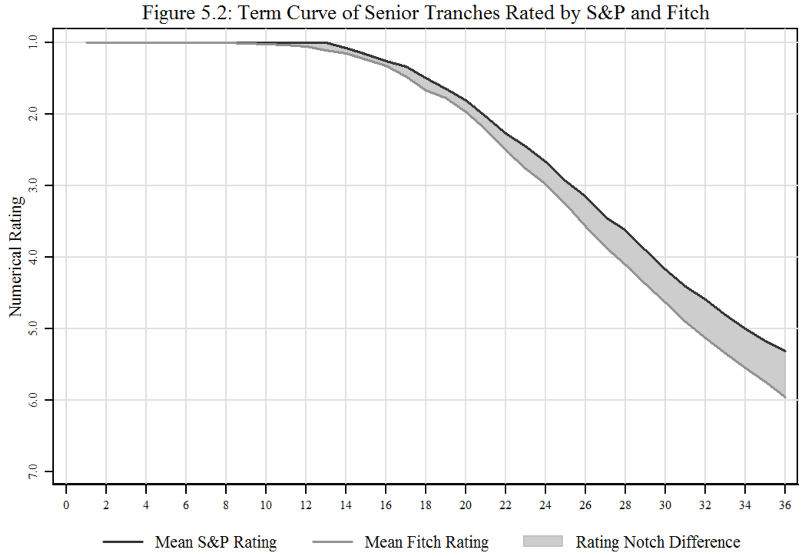

Figure 5.3: Term Curve of Senior Tranches Rated by Moody's and Fitch