

Rating Alignment, Rating Shopping and Reputation of Credit Rating Agencies: Evidence from the Subprime Crisis

Annalisa Croce

Politecnico di Milano, Department of Management, Economics and Industrial Engineering. Via Lambruschini 4/b, 20156 Milano, Italy. Phone: +390223992743; Fax:+390223992710; E-mail address: annalisa.croce@polimi.it.

Stefano Lugo

Politecnico di Milano, Department of Management, Economics and Industrial Engineering. Via Lambruschini 4/b, 20156 Milano, Italy. Phone: +390223993965; Fax:+390223992710; E-mail address: stefano.lugo@mail.polimi.it.

Robert Faff

University of Queensland, UQ Business School, 4072, St. Lucia, Australia. Phone:61733468055; Fax:+61733468166; E-mail address: r.faff@business.uq.edu.au

Abstract

This paper compares conflict of interest incentives and reputational concerns of credit rating agencies (CRAs) in the context of the subprime crisis. We argue that, during up-market periods, ratings levels are affected by both a strong tendency for alignment across CRAs and ratings “shopping” by issuers, while, during periods of economic slowdown, these distortions disappear since CRAs are then more concerned about their long-run reputation. We test our hypotheses by analyzing the gap between Moody’s and S&P’s ratings on US residential, subprime mortgage-backed securities before and after the 2007 crisis. Overall, our results show a clear reduction in ratings alignment. Moreover, we find strong evidence that harsher downgrades came from S&P, which had higher ratings before the crisis, and that the gap reduction is strongly correlated with the rating gap before the crisis. We interpret this as evidence that CRAs try to “reverse the gap”, to reduce the impact on their (relative) reputation. Finally, we find that harsher downgrades tend to occur for securities not jointly rated and that the relation between downgrades and initial rating is significantly different across the two agencies, this being consistent with the rating shopping hypothesis.

EFM classification codes: 540, 790

Keywords: credit rating agencies, subprime, conflicts of interest, rating alignment, rating shopping

1. Introduction

Using the subprime crisis as our experimental setting, we study “collusive” behavior amongst credit rating agencies (CRAs) and “ratings shopping” in the context of a large sample of residential mortgage-backed securities (RMBS). We argue that when a crisis occurs, the ratings alignment on jointly-rated securities strongly decreases, since each CRA is more concerned about its own reputation, and thus is more likely to give a truly independent evaluation. Also, we expect harsher downgrades to come from the CRA which was giving higher ratings before the crisis occurred, this behavior also being consistent with a refocus on reputational capital by the agency. In other words, we expect alignment to decrease for cases in which ratings agreement previously existed and to increase when a “split rating” was present.¹ Finally, we test for ratings shopping by studying whether the crisis induced harsher downgrades on not jointly rated (NJR) securities rather than on their jointly rated (JR) counterparts, and whether the relationship between the initial evaluation and downgrade is different across CRAs.

The literature on ratings shopping suggests issuers could choose a particular agency knowing that its evaluation system would be more preferable than its competitors in producing a more favorable rating for a specific security. Baker & Mansi (2002) conduct a survey among rated issuers, finding that only 2.7% of them claim that the rating agencies business model induces agencies to assign high ratings in order to satisfy issuers. Nevertheless, CRAs themselves admit ratings shopping exists – especially for structured finance products, for which there is no marking-to-market and unsolicited ratings are not common. In a series of studies,² Moody’s reports hypothetical evaluations of securities it hadn’t rated, finding that average differences would have been higher than those verified on jointly rated securities. Thus, Moody’s rating opinions and the opinions of other agencies often diverge substantially in situations when Moody’s is not asked to

¹ Cantor et al. (2007b) suggests that split ratings mainly come from different rating scales. However, if that was the only major source of misalignment, we should then expect only average ratings level, and not relative valuation, to be affected by a downturn. We verify this is not the case during the subprime crisis.

² See, for example, Moody’s 2002a and Moody’s 2002b.

rate a particular security. This is highly suggestive of rating shopping phenomenon: when asking only one agency for a rating, the issuer targets the agency that it believes would assign the higher evaluation. At the same time, this shows that rating alignments on securities evaluated by more agencies is higher than it would have been on securities rated by only one agency.

The crisis on subprime RMBS, starting in 2007, is an appealing context to study the aforementioned phenomena, since it is a widely acknowledged and clear structural break jointly affecting thousands of similar issues, allowing for a comparative analysis of ratings before and after such a major trigger event. The observed growth of the issuance of these securities between 2002 and 2006 was exponential for several reasons: low interest rates, less tight credit access conditions, predatory lending and new forms of mortgages (such as adjustable rate mortgages with a two to three year period of no principal repayment) burst into the real estate and mortgage market, especially for nonprime borrowers. Subprime mortgages grew from \$421 billion in 2002 to \$640 billion in 2006, representing approximately 22% of the mortgage pool (14% in 2002) according to Moody's (2008b). At the same time, the crisis on high-yield corporate bonds seen between 2001 and 2003 drove a linked crisis for Collateralized Debt Obligations (CDOs) built upon these securities: CDO issuers replaced troubled company debt with residential mortgages as their main source of collateral for their issues. Accordingly, the dramatic rise of RMBS, ABS and CDOs drove a strong increase in their importance in CRA business. For example, Moody's revenues from structured finance ratings grew from \$US50 million to nearly \$US850 million over the period 1995 and 2006.³

The rapid growth in these products increased the relevance of conflicts of interest as well as ratings complexity. In an analysis conducted by the SEC on CRA's behavior in evaluating RMBS and CDO, they find an inability of analysts to deal with the increased amount and complexity of these securities. They also verified a high concentration in the market of subprime RMBS underwriting: regarding 642 analyzed deals there were only 22 different arrangers, and half of them

³ J. Eisinger: "Overrated". Portfolio.com, August 13 2007.

accounted for nearly 80% of both deal volume and dollar value; in the CDO market, the degree of concentration is even higher (SEC, 2008). While the first signs of a slowdown in real estate and associated securities surfaced by the end of 2006, the crisis really began showing its true teeth in June 2007. Specifically, on the 21st of June that year, Merrill Lynch made a selling audit for its \$850 million share in two Bear Sterns funds mainly invested in structured finance products. Despite several liquidity injections into the two funds by Bear Sterns, JP Morgan and others, no buyer was found, forcing a severe write down. On July 3, 2007 Moody's and S&P, which hadn't shown strong rating revision activity up to that time, started a series of frequent and strong downgrades on these securities, which are the object of our analysis.

Overall, our results show a clear reduction in ratings alignment. Moreover, we find strong evidence that harsher downgrades came from S&P, which assigned higher ratings before the crisis, and that the gap reduction is strongly correlated with the rating gap before the crisis. We interpret this as evidence that CRAs were keen to "reverse the gap", to reduce the impact on their (relative) reputation. Finally, we find that harsher downgrades tend to occur for securities not jointly rated and that the relation between downgrades and the initial rating is significantly different across the two firms, this being consistent with the rating shopping hypothesis.

Our paper proceeds as follows. Section 2 provides a brief literature review culminating in our core empirical predictions. In Section 3 we present the dataset, while in Section 4 the empirical analysis and results are shown. Section 5 closes with some final remarks and conclusions.

2. Background, Literature Review and Empirical Predictions

The relevance and importance of studies examining the behavior of credit rating agencies has grown considerably over recent years, and particularly so, since CRAs emergence as a key player in the evolution and regulation of financial markets (see, for example, SEC, 2003). Market trust in CRA evaluations has driven coupon spreads and debt value to be strongly influenced by

ratings (see, for example, Ederington et al., 1987 and Kliger and Sarig, 2000). Moreover, many financial regulatory institutions have used their ratings as official benchmarks for the credit risk of bonds and structured obligations in defining restrictions on investors' behavior. For example, the US SEC use their evaluations to define in which classes of asset, certain types of agents are allowed to invest/trade, and the Basel II agreement makes bank capital requirements depend on their assets credit risk, which are also estimated using agencies ratings. Therefore, it is quite plausible that the CRAs ratings deliberations influences economic and financial cycles by driving market rate spreads and liquidity.

However, conflicts of interest exist in CRA behavior. Because it is the debt issuer who pays the fees to obtain a rating, rating agencies could be tempted to increase their revenues by assigning higher ratings. Higher ratings would have the effect of lowering interest rates requested by the market and, as a consequence, raise the total issuing volume, which in turn would induce a higher overall demand for ratings. Besides this global effect, assignment of an "inflated" rating would also likely attract more issuers regarding other competitors. This is especially true for complex and illiquid securities such as RMBS and Asset Backed Securities (ABS): fees paid by issuers are particularly high (between 13 and 16 bps), and the absence of a marking-to-market process makes such information produced by rating agencies even more critical.

Collectively, these conflict of interest considerations can produce a ratings shopping effect: issuers ask for ratings assessments from different agencies, making public only favorable ratings; CRAs themselves can indicate what kind of enhancement and/or structuring of the security is needed to achieve the desired rating under their evaluation models. Notably, Skreta and Veldkamp (2009) document the rating shopping phenomenon by showing that when issuers can choose from several ratings suppliers, product complexity can lead to rating inflation, even though rating agencies produce unbiased ratings.

Critics of rating agencies behavior claim that the absence of competition does not help to remove the effects of the CRA conflict of interest (see, for example, Bolton et al., 2009). The rating

market looks, in fact, much more like an oligopoly than a perfectly competitive market. In the US, the SEC designates ten Nationally Recognized Statistical Rating Organizations (NRSROs);⁴ however, among these chartered agencies, the so-called “big three” (Moody’s, Standard and Poor’s and Fitch) dominate the market. For example, in 2008 they controlled the 98 percent of the US market for debt ratings, according to the SEC.⁵ This situation, as modeled theoretically by Stolper (2009), gives these three agencies an incentive to align their evaluations. Specifically, if they assign aligned ratings, in the scenario where actual losses exceed the expected losses predicted by CRAs, it would be difficult for a regulator to distinguish between an exogenous, unpredictable common shock effect and an inflated ratings effect. Conversely, deviating from general alignment exposes a single CRA to either a reputational risk (where they produce higher ratings than peers) or to a loss of potential profits during a positive cycle (where they produce lower ratings than peers).

The regulatory solution proposed by Stolper (i.e. rewarding CRAs not colluding by reducing the number of chartered competitors in the future) would be ineffective in practice, since the market is already dominated by just three out of ten NRSROs (with Moody’s as the most prominent player and Fitch as a clear third). In other words, ratings alignment gives CRAs a type of insurance against the detrimental effects on their reputations of negative shocks on the affordability of their ratings, without compromising the positive short-term effect of giving high ratings. Moreover, aligned evaluations on jointly-rated securities could reduce the uncertainty on effective risk perceived by the market, thus reducing requested spreads. Cantor et al. (1997) and Bongaerts et al. (2009), find both Moody’s and S&P ratings produce useful information for pricing jointly rated corporate bonds. To the best of our knowledge, no similar study exists on RMBS and ABS spreads, which is notable since it is likely that the relationship is even stronger, given the illiquidity and complexity of these securities. Strong correlation between ratings and market spreads also makes CRAs less influenced

⁴ The ten NRSROs, as of 18th July 2010, were A.M. Best Company, Inc, DBRS Ltd., Egan-Jones Rating Company, Fitch, Inc., Japan Credit Rating Agency Ltd., LACE Financial Corp., Moody’s Investors Service, Inc., Rating and Investment Information, Inc., Realpoint LLC, Standard & Poor’s Ratings Services. <http://www.sec.gov/answers/nrsro.htm>

⁵ D. Evans and C. Salas: “Flawed Credit Ratings Reap Profits as Regulators Fail Investors”. Bloomberg.com, April 29, 2009.

to promptly change their evaluations as new information becomes available, since this could endogenously trigger a deterioration of issuers' affordability. Loffler (2007) argues that this is a rational behavior, since overly volatile ratings wouldn't furnish greatly valuable information on the long-term risk of the market.

The typical CRA response to this kind of criticism is that their reputation is their most important asset:⁶ if ratings were not affordable, markets would not rely on them, thus reducing their signaling power and making no issuer willing to pay for them. Indeed, Covitz and Harrison (2003) find some support for this view, showing that the timing of downgrades does not relate to issuer interest and importance as a customer for the CRA. However, this argument understates three nontrivial considerations: the impact of economic cycles on both (a) the monitoring of CRA behavior and (b) CRAs rating reliability; and (c) the potential gain coming from inflated ratings on complex structured securities when fees are significant.

Beginning with the first point, the attention paid by financial markets to CRA behavior is closely linked to economic cycles, strongly increasing during downturns while being much less effective during expansionary phases.⁷ Accordingly, Nickell et al. (2000) using Moody's data from 1970 to 1997 find that business cycles are the main factor in explaining rating transition probabilities; Bangia et al. (2002), using S&P data from 1981 to 1998, also provide evidence supporting the importance of separating the economy into two states (i.e. expansion and contraction) when analyzing the determinants of ratings downgrades: they show that default probabilities increase significantly during contractions. Overall, during a contraction phase CRAs virtually compete with each other in being the first to signal increased credit risk, seemingly to compensate for their benign behavior during expansion phases. Such a perception is well illustrated

⁶ See, for instance, SEC 2003.

⁷ For example, the CRAs failure to predict the crisis at firms such as Enron, WorldCom or Parmalat focused public attention on the affordability of CRA evaluations – however, this hasn't changed the market structure or the signaling power of ratings. For a review of research evidence on criticisms of rating agencies, see Frost (2007).

by the recent case of the Greek Government lamenting the harsh downgrades received on its bonds in 2011.⁸

Second, as already stated, illiquid complex securities lack the marking-to-market check for ratings reliability, thus making a misevaluation clearly visible only in the case of a sudden increase in default rates driven by an economic downturn. When this happens, as in Stolper's model, excessive ratings alignment makes it difficult to distinguish between misevaluation effects and exogenous unpredictable shock effects, thus preserving the rating sector *status quo*.

Third, the costs associated with potential loss in reputation could be overwhelmed by the gain derived from giving less severe ratings. Relevant to this concern, the Bolton et al. (2009) and Mathis et al (2009) models of CRA behavior show, respectively, that in the presence of naïve investors and when a large fraction of the CRA's income stems from rating complex products, it can be rational for agencies to inflate their evaluations.

Accordingly, based on the arguments of conflicts of interest, ratings shopping and reputation effects, our empirical predictions are as follows:

P1: Compared to the pre-crisis period, credit ratings assigned by the two key agencies become less aligned after the advent of the crisis

Underlying this prediction, we assume that during expansion periods, ratings levels are affected by a strong tendency for alignment across CRAs, while, during periods of economic slowdown, this distortion would disappear since CRAs are then more concerned about their long-run reputation.

P2: On average, the agency assigning higher credit ratings in the pre-crisis period invokes harsher downgrades after the advent of the crisis

With this prediction we aim to assess whether reputational incentives drive CRA behavior during a crisis period. We assume that when a crisis occurs, each CRA is more concerned about its own

⁸ R. Milne and D. Oakley: "Hard to Credit". *Financial Times*, 28 March 2011.

reputation. Thus, we expect harsher downgrades to come from the CRA which was giving higher ratings before the crisis, thereby reflecting a refocus on reputational issues by the agency.

P3: On average, there is a reversal of the credit ratings gap between the two key agencies after the advent of the crisis

According to this prediction, the “reversing gap” phenomenon is assumed to be in evidence regardless of which agency gave the higher rating initially.

P4: On average, harsher downgrades occur after the advent of the crisis for securities not jointly rated compared to their jointly rated counterparts

Not jointly rated (NJR) securities are predicted to receive harsher downgrades for two reasons. First, if rating shopping exists, these tranches would likely have been more overrated than their jointly rated (JR) counterparts before the crisis; and second, the absence of the competitor evaluation increases the potential negative impact on reputation, as the responsibility for a misevaluation is born by only one CRA.

Similarly, consistent with the rating shopping hypothesis, we investigate the possibility that the relationship between the downgrade intensity and the pre-crisis rating differs: (a) between jointly-rated and not-jointly rated securities and (b) across agencies for not-jointly rated securities. Specifically, our final empirical prediction is the following:

P5: On average, the relationship between the pre-crisis rating and the observed downgrade for the two agencies differs more in the case of not-jointly-rated securities than for their jointly-rated counterparts

Both empirical and theoretical studies⁹ document how, on average, Moody’s (by using Expected Losses as the main indicator), tend to give more severe ratings to subordinate, typically speculative grade (SG) tranches, while S&P (based on Probability of Default) can assign slightly worse ratings (relative to Moody’s) to senior, investment grade (IG) tranches. Accordingly, we test

⁹ See, for example, Moody’s (2007), Fender and Kiff (2004) and Peretyatkin and Perraudin (2002).

whether the relationship between initial rating and subsequent downgrade is different between the two CRAs. Moreover, while in jointly rated tranches this difference in evaluation might be smoothed by rating alignment, we expect it to be more evident in the single-rated tranches. In other words, if issuers, trying to secure inflated evaluations, target different agencies for different expected levels of rating (i.e. investment versus speculative grades), this should be reflected in different levels of downgrade for a given initial rating among the two agencies in the NJR sample. However, for jointly-rated securities the general alignment predicted by *PI* smoothes this difference.

3. Data

Since our goal is to analyze CRAs behavior with respect to rating downgrades on non-prime RMBS following the 2007 structural break, our sample is created based on a search in Bloomberg of all US ABS Home Equity Loan (HEL) tranches¹⁰ issued between 2005 and 2007 that experienced a downgrade or have been placed on a watchlist by Moody's and/or S&P between June 1st 2007 (just before the Bear Sterns hedge funds default) and January 31st 2008. According to Moody's classification, the HEL category includes "securities backed by subprime (B&C) mortgage loans, home improvement loans, high loan-to-value (high LTV) loans, home equity lines of credit (HELOCs), and closed-end second-lien loans, as well as net interest margin (NIM) securitizations";¹¹ aside from Alt-A mortgages, all non-prime RMBS should then be considered in the category. We then integrate the dataset by searching across Moody's and S&P websites for the ratings of tranches which experienced a downgrade by only one of the two CRAs during the analyzed period. In total, we identify 4,142 tranches, 330 (8%) of which were rated by only one of the two CRAs (70 by Moody's and 260 by S&P) and 3,812 were jointly rated (JR).

[Insert Table 1 about here]

¹⁰ Indexed in Bloomberg as ABS/CMO HOME EQ category, USA market.

¹¹ See Moody's (2008a).

Table 1 displays the ratings distribution and average differences in ratings of JR securities as at March 2007 (Panel A); June 2007 (Panel B) and February 2008 (Panel C). As shown in the table, more than 90% of tranches in our sample were rated as investment grade (i.e. ratings of Aaa, Aa, A or Baa) in June 2007. This high proportion occurs partly because the bulk of all rated HEL were investment grade before the crisis (nearly 75% of them were Aaa, Aa or A) and partly because (by construction) our dataset does not include securities which defaulted without previously experiencing a rating review during the analyzed period. In stark contrast, by February 2008, approximately 43% and 20% of JR securities were rated “junk” (Caa-CCC or below) by S&P and Moody’s, respectively.

Similar to average rating levels, relative evaluations between Moody’s and S&P have also changed between June 2007 and February 2008. Prior to the crisis, 2,432 JR tranches (64%) were rated the same, 1,139 (30%) had a higher rating granted by S&P, while only 241 (6%) had a higher rating ranked by Moody’s. Notably, all tranches rated triple A by Moody’s were equally rated by S&P. Qualitatively, the relative rating distribution and average differences in June 2007 were similar to those reported by Moody’s on all JR ABS for March 2007.¹² In February 2008, only 15% of tranches had the same rating, and the bulk had a higher rating by Moody’s. It is clear from this anecdotal picture that the alignment of ratings dramatically declined after the crisis (consistent with empirical prediction *P1*), and that S&P became the more severe agency, while Moody’s ratings, even though generally more aligned, were predominantly lower before the crisis (consistent with empirical prediction *P2*).¹³

To measure the extent of split rating on JR securities, we follow a standard approach in studies on CRAs: we match Moody’s and S&P alphanumerical scales with a master scale based on 21 notches (Aaa/AAA=21, Aa1/AA+ = 20.....C/D=1). Since studies on default prediction accuracy (see, for example, Hamilton and Cantor, 2004) show that the inclusion of securities on watchlist for

¹² See Moody’s (2007).

¹³ A test of the rating difference between S&P and Moody’s is significant at the 1% level.

possible future downgrades brings significant information on the level of distress, we consider watchlist inclusion as a 1 notch downgrade (i.e. we subtract 1 from the numerical value previously assigned). Based on our sample, we see that there has not only been a strong increase in the number of tranches rated higher by Moody's, but also the average gap, when S&P is more severe, has strongly increased and is bigger than the average gap when the Moody's rating is lower, for all classes of ratings. For example, JR securities rated Baa by Moody's in February 2008 have a lower evaluation by S&P in 75% of the cases. Moreover, the average notch gap of 4.56 at the Baa level implies an average S&P rating of BB-/B+ on the same securities; and considering only those tranches rated higher by Moody's, the average S&P rating drops to B- (given the average gap of 6.91). In contrast, prior to the crisis, for Moody's Baa rated securities only 9.71% of tranches were rated higher by Moody's while 40.38% of them were rated higher by S&P. However, the average S&P rating was basically very similar to Moody's (there was less than a 1-notch average gap).

[Insert Table 2 about here]

Table 2 reports the average relative rating transitions. It is clear from this table how the preceding patterns have occurred: when the S&P rating was considerably higher than Moody's (more than 2 notches) in June 2007, about half the time this gap was compensated or reversed in February 2008. In particular, almost one out of three tranches had a substantially reversed gap (i.e. the Moody's rating ended more than 2 notches higher). In cases where the rating was equal in June 2007, 63% of such tranches changed to be more highly rated by Moody's in February 2008, and for half of these cases the gap exceeded 2 notches. The preceding observations are, again, consistent with empirical prediction *P2*.

4. Empirical evidence

4.1 Rating gaps and reputational effect on jointly rated securities

To begin, we assess whether and how reputation influenced ratings during the crisis. To this end, we analyze the gap variation before and after the crisis started using the variable Delta, defined as follows:

$$\Delta_{t,i} = S\&P_{t,i} - \text{Moody}'s_{t,i} \quad (1)$$

where t is the given time period (either June 2007 or February 2008), i is a jointly rated tranche and S&P and Moody's are the respective ratings assigned by each agency, measured according to the main numerical scale as described in Section 3. Delta thus takes integer values defined over the interval -20 to 20: negative (positive) values indicate S&P (Moody's) giving the lower rating.

In the absence of a reputational effect, Delta in February 2008 should be equal to Delta in June 2007, implying that crisis did not cause any change in the (relative) behavior of CRAs. In contrast, in the presence of reputational effects as predicted by $P2$, the correlation between the gap variations before and after the crisis started should be negative: the agency giving the higher rating before the crisis should apply the harsher downgrade, aiming to reverse the gap. However, we expect Delta in February 2008 to be nonlinearly correlated with Delta in June 2007: the higher the difference in rating before the crisis, the higher the expected "gap reversing" after the crisis took place. Accordingly, we specify a cubic relation between Delta in February 2008 and in June 2007, given by equation [2]:

$$\Delta_{Post_i} = \alpha + \beta_1 \Delta_{Pre_i} + \beta_2 \Delta_{Pre_i}^2 + \beta_3 \Delta_{Pre_i}^3 + \varepsilon_i \quad (2)$$

where Δ_{Post} and Δ_{Pre} are values of Delta in February 2008 and June 2007, respectively.

The intercept term, α , represents the average gap between S&P and Moody's ratings after the crisis on securities that were similarly rated before the crisis took place. According to *P1* we expect this coefficient to be significantly different from 0. More specifically, a negative value of this coefficient would confirm a post-crisis higher downgrade from S&P than from Moody's. The signs of the coefficients β_1 , β_2 and β_3 should be consistent with predictions *P2* and *P3*, which state that Δ_{Post} should increase (decrease) when Δ_{Pre} decreases (increases) for Δ_{Pre} sufficiently distant from zero, while allowing for a positive/neutral relationship when Δ_{Pre} is close to zero.

In particular, if we examine the marginal effects as shown in equation [3], it appears that a negative sign for β_3 implies decreasing marginal effects when Δ_{Pre} moves "far" from 0 and eventually becoming negative when the magnitude of Δ_{Pre} is sufficiently large. When the marginal effect is less than unity an increase in the absolute gap before the crisis results in a less than proportional increase in the expected absolute gap after the crisis; when the marginal effect becomes negative, an increase in the absolute gap in June 2007 *decreases* the expected absolute gap in February 2008 or even results in an expected gap with the opposite sign. In other words, if β_3 is negative, the higher is the initial rating difference, the greater is the gap reduction (or even reversion) once the crisis occurs, consistent with prediction *P3*. Thus, we expect β_3 to be negative.

$$\frac{\delta \Delta_{Post}_i}{\delta \Delta_{Pre}_i} = \beta_1 + 2\beta_2 \Delta_{Pre}_i + 3\beta_3 \Delta_{Pre}_i^2 \quad (3)$$

Moreover, we expect the marginal effect when Δ_{Pre} is equal to 0 (*i.e.* β_1) to be close to unity, as ABS issues characterized by aligned evaluations before the crisis shouldn't constitute a matter of concern for the CRAs relative reputation.

Figure 1 provides an illustrative characterization of how the cubic specification captures the predicted relation. Specifically, when Δ_{Pre} falls in the intermediate range of about (-5 to 5), the curve lays under the 45° straight line (assuming a negative value of the α coefficient in

accordance with the descriptive statistics that show S&P downgrades are generally harsher), but no strong reversal effect is present. However, once *Delta_Pre* breaches this range in either direction, if β_3 is negative the cubic relation comes into play sending the slope into negative territory, consistent with reputation effects dominating such that there is a reversal of the ratings gap.

[Insert Figure 1 about here]

The estimation results for equation [2] are reported in Table 3. As shown in the table, we find the estimated coefficients to be significant and with expected signs such that the relationship between *Delta_Pre* and *Delta_Post* assume the shape illustrated in Figure 1 within the *Delta_Pre* domain (-20, +20). First, the estimated intercept term is negative, indicating that S&P tended to downgrade their credit assessments more severely than Moody's, thus confirming *P1*: after the crisis began the previously less severe CRA becomes, on average, the more severe. Second, as illustrated in Figure 1, the results for the quadratic and cubic terms indicate that, according to prediction *P2*, the more disparate are the ratings between S&P and Moody's prior to the crisis, the greater is the move in ratings such that the gap is reduced or even reversed. This effect is in evidence regardless of which agency gave a higher rating initially. This finding is thus consistent with empirical prediction *P3*.¹⁴ Moreover, looking at the marginal effects, we verify that they become significantly negative within the domain of *Delta_Pre*, regardless of which agency was assigning the higher rating in June 2007. However, it must be noted that the marginal effects become significantly negative (only at the 10% confidence level) for lower rating gaps when Moody's was the more severe agency, a result consistent with S&P generally being more keen to reverse contrary evaluations, as they proved to be higher on average before the crisis erupted. Finally, we cannot reject the hypothesis that the marginal effect is not statistically different from 1 when *Delta_Pre* is equal to 0 (p-value equal to 0.627), supporting the prediction presented above.

¹⁴ Our model specification is robust to checks for the model degree using the Likelihood ratio test: the explanatory power of a second-degree model (i.e. including only the constant, *Delta_Pre* and *Delta_Pre*²) is significantly lower than the explanatory power of our cubic model.

[Insert Table 3 about here]

4.2 Rating shopping and not jointly rated securities

To test prediction *P4*, we measure, separately for the two CRAs, the difference in average downgrade (measured in the numerical scale) between JR and NJR tranches, excluding JR tranches downgraded only by the other CRA. We verify that, in our sample, the average downgrade for NJR tranches is significantly higher than that for JR tranches for Moody's: the difference, equal to 1.9 notches is significant at the 99% confidence level. In contrast, for S&P, the difference is -0.1 notches and is not significant at standard confidence levels. This finding thus supports empirical prediction *P4* for Moody's but not for S&P. This differential finding is not surprising once we recognize the fact that S&P made significantly harsher downgrades on JR tranches than Moody's, this reducing the pressure to be even more severe on NJR tranches. In support of this interpretation, we find that Moody's average downgrade on NJR tranches (6.57 notches) is not statistically different from S&P (6.32), while its average rating reduction on downgraded JR tranches (4.70) is significantly lower compared to S&P (6.42), at 99% confidence level.

Finally, aiming to test prediction *P5* we create two datasets: one including all tranches downgraded by Moody's (both JR and NJR) and a second based on all tranches downgraded by S&P (both JR and NJR). Accordingly, we estimate, separately for the S&P and Moody's samples, the following model:

$$DownGrade_i = \alpha + \beta_1 NJ_i + \beta_2 Rating_Pre_i + \beta_3 NJ_i Rating_Pre_i + \varepsilon_i \quad (4)$$

where $DownGrade_i$ is the difference between the ratings in February 2008 and June 2007 given by Moody's (S&P) on tranche i , NJ_i is a dummy variable taking a value of 1 if the tranche is rated by only one of the CRAs (i.e. by Moody's or by S&P, but not by both) and 0 otherwise and $Rating_Pre_i$ is the rating assigned in June 2007 by Moody's (S&P). Again, ratings and downgrades are measured using the numerical scale presented in Section 3.

The expected downgrade for the JR tranche i is thus given by the following expression:

$$DownGrade_{JR_i} = \alpha + \beta_2 Rating_{Pre_i} \quad (5)$$

while the expected downgrade for the NJR tranche j is measured by:

$$DownGrade_{NJR_j} = \alpha' + \beta'_2 Rating_{Pre_j} \quad (6)$$

where α' equals to $(\alpha + \beta_1)$ and β'_2 equals to $(\beta_2 + \beta_3)$.

Table 4 shows results for the estimation of equation [4]. The first column refers to Moody's sample, while the second reports estimates on the S&P sample. In the last two rows, a Wald test on the linear combinations of coefficients, denoted α' and β'_2 , are reported.

[Insert Table 4 about here]

As expected, for jointly rated tranches, the two agencies show a similar relation between initial rating and downgrade, given by the estimated β_2 : -0.61 versus -0.69. In particular, a negative and significant estimated coefficient indicates that the worse the initial rating on the jointly rated tranches, the harsher is the downgrade and the magnitude of this effect is very similar between the two agencies.

The real difference in the behavior of the two CRAs becomes clear when we look at the NJR tranches (i.e. coefficient β'_2), confirming prediction $P5$: for Moody's the relation between initial rating and downgrade is negative, similar to that shown in JR securities for both CRAs. Thus, Moody's downgrade behavior is even higher for solely-rated securities that, before the crisis, received a low rating. In contrast, S&P shows an opposite trend in NJR securities: downgrades increase with the initial rating, meaning that higher downgrades are applied to securities that, before the crisis, received a better evaluation. Specifically, the estimated coefficient on initial rating interacting with the NJR dummy (β_3) is significantly positive and more than fully counteracts the negative β_2 coefficient: the sum of the two estimated coefficients, reported in the last row, is

significantly positive. This finding of a differential downgrade behavior for not-jointly rated securities, is consistent with empirical prediction *P5*. Results, based on the estimated coefficients, are graphically presented in Figure 2.

[Insert Figure 2 about here]

These results show that, for NJR securities, both agencies made harsher downgrades on rating grades theoretically unfavoured by their own evaluation methods. In one sense this is surprising, as we would have expected lower downgrades on tranches that should have been already penalized by the evaluation method used before the crisis. However, it might be the case that adjustments on their models parameters after the crisis exacerbate the tendency to penalize more the investment or the speculative grades tranches for S&P and Moody's, respectively. Of course, it could also be that each CRAs ratings on NJR tranches were inflated particularly on grades theoretically penalized by their models, a possibility that could be investigated in future research. Interestingly, the different relationship between initial rating and downgrades, along with the higher average downgrades, supports the view that NJR tranches were particularly subject to a ratings shopping phenomena, driving inflated evaluations before the crisis and, at least for Moody's, representing a bigger concern for CRA reputation once the crisis started.

5. Conclusions

In this paper we analyze the CRA behavior during the subprime crisis. We find evidence that ratings alignment has strongly reduced, and that S&P, which gives on average higher ratings than Moody's just before the crisis, was more severe in rating downgrades eight months later. We interpret this alignment reduction as evidence of more independent evaluations. When a big misalignment was present before the crisis, the less severe agency generally reduced or reversed this gap. Both higher downgrades by the previously less severe agency and the gap reversion on split ratings are evidence of a major focus on reputation once the crisis began. We also find that

Moody's made harsher downgrades on tranches that were not jointly rated; this finding supports the idea of inflated evaluations when the benchmark of alternative peer ratings is missing, as well as a stronger reputational concern linked to solely rated securities during a downturn. Finally, with regard to the solely rated securities, we find that the relationship between the evaluation before the crisis and the subsequent downgrade is different for the two agencies, giving further support to the idea that ratings shopping might exist and thus induces ratings inflation.

Our findings give further support to concerns about the effect of incentives distortion and market concentration on CRA behavior, already highlighted in previous financial scandals (such as Enron and Parmalat frauds). While distortions such as the alleged rating shopping phenomenon are not easily identifiable without a structural break, future research should further investigate the relationship between leading-following on downgrades and split ratings. Ideally, ratings alignment, especially on complex, structured and illiquid products, should be monitored over a longer time horizon, as it could revert over time to the high (and arguably excessive) levels observed prior to the onset of the subprime crisis.

References

- Baker, H., and Mansi, S. (2002). Assessing credit agencies by bond issuers and institutional investors. *Journal of Business Finance and Accounting* 29 , 1367-1398.
- Bangia, A., Diebold, F. X. and Schuermann, T. (2002). Ratings migration and the business cycle, with applications to credit portfolio stress testing. *Journal of Banking and Finance* 26 , 445-474.
- Bolton, P., Freixas, X. and Shapiro, J. (2009). The credit ratings game. NBER Working Paper No. 14712.
- Bongaerts, D., Cremers, M. and Goetzmann, W. N. (2009). Multiple ratings and credit spreads. NBER Working Paper No. 15331.
- Cantor, R., Packer, F. and Cole, K. (1997). Split ratings and the pricing of credit risk. Federal Reserve Bank of New York Research Paper No. 9711.
- Covitz, D. and Harrison, P. (2003). Testing conflicts of interest at Bond Ratings Agencies with market anticipation: evidence that reputation incentive dominate. Board of Governors and the Federal Reserve System, Finance and Economics Discussion Series, 1-36.
- Ederington, L.H., Yawitz, J.B. and Roberts, B.E. (1987). The informational content of bond ratings. *Journal of Financial Research* 10, 211-226.
- Fender, I. and Kiff, J. (2004). CDO Rating Methodology: Some Thoughts on Model Risk and its Implications. *Journal of Credit Risk*, Vol. 1, No. 3
- Frost, C. (2007). Credit Rating Agencies in Capital Markets: A Review of Research Evidence on Selected Criticisms of the Agencies. *Journal of Accounting, Auditing and Finance*, 469-492.
- Guttler, A. and Wahrenburg, M. (2007). The adjustment of credit ratings in advance of defaults. *Journal of Banking and Finance* 31, 751-767.
- Hamilton, D. and Cantor, R. (2004). Rating transitions and default conditional on watchlist, outlook and rating history. Moody's investor service special comment, February.
- Kliger, D. and Sarig, O. (2000). The Information Value of Bond Ratings. *Journal of Finance*, 55(6), 2879-2902.
- Loffler, G. (2007). Profits first or clients first? Some lessons from Moody's Stock Price. Working Paper.
- Mathis, J., McAndrews, J. and Rochet, J. (2008). Rating the raters: Are reputation concerns powerful enough to discipline rating agencies? *Journal of Monetary Economics* 56, 657-674.
- Moody's Investor Service (2002a). Moody's Studies Ratings of Non-Moody's Rated CDOs and Confirm Rating Estimate Approach. March 22, 2002.
- Moody's Investor Service (2002b). Moody's study of Ratings of Non-Moody'-rated RMBS. April 18, 2002.
- Moody's Investor Service (2007). Comparing Ratings on Jointly-Rated U.S. Structured Finance Securities: 2007 Update. March 2007.
- Moody's Investor Service (2008a). Structured Finance Rating Transitions: 1983-2007. Moody's Credit Policy special comment, February 2008.

Moody's Investor Service (2008b). A short guide to Subprime. March 25, 2008.

NERA (2003). Credit ratings for structured products: A Review of Analytical Methodologies, Credit Assessment Accuracy, and Issuer Selectivity among the Credit Rating Agencies. 6 November 2003

Nickell, P., Perraudin, W. and Varotto, S. (2000). Stability of rating transitions. *Journal of Banking and Finance* 24, 203-227.

Peretyatkin, V. and, Perraudin, W. (2002). EL and DP Approaches to Rating CDOs and the Scope for "Ratings Shopping". *Credit Ratings Methodologies, Rationale and Default Risk*, London, Risk Books.

Ramakrishnan, R. and Thakor, A. (1984). Information Reliability and a Theory of Financial Intermediation. *Review of Economic Studies*, 51 (3), 415-432.

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

Stolper, A. (2009). Regulation of credit rating agencies. *Journal of Banking and Finance* 33, 1266–1273.

The Securities and Exchange Commission (2003). Report on the Role and Function of Credit Rating Agencies in the Operation of the Securities Markets. January 2003.

The Securities and Exchange Commission (2008). Summary Report of Issues Identified in the Commission Staff's Examinations of Select Credit Rating Agencies. July 2008.

Table 1. Ratings distribution and average differences in ratings of jointly rated securities

This table reports the distribution among different ratings of all ABS jointly rated by Moody's and S&P in March 2007, as reported by Moody's, as well as for our dataset of HEL in June 2007 and February 2008. Matchable data on all ABS rating distribution by S&P were not available. Average GAP is measured averaging all rating differences (including tranches rated equally) between S&P and Moody's rating for each Moody's rating class in notches. AGAP if Moody's is higher (lower), is the average GAP only on tranches with a lower (higher) rating by S&P.

| Rating | Moody's frequency | S&P frequency | Average GAP | AGAP if Moody's higher | AGAP if Moody's lower | Moody's lower | Same | Moody's higher |
|-------------------------------------|-------------------|---------------|-------------|------------------------|-----------------------|---------------|---------|----------------|
| <i>Panel A: All ABS, March 2007</i> | | | | | | | | |
| Aaa | 36.66% | . | 0.03 | 2.23 | na | na | 98.60% | 1.40% |
| Aa | 17.06% | . | -0.20 | 1.78 | -1.20 | 26.70% | 66.30% | 7.00% |
| A | 19.52% | . | -0.54 | 1.51 | -1.76 | 36.20% | 57.20% | 6.60% |
| Baa | 20.56% | . | -0.55 | 1.38 | -1.71 | 36.90% | 57.10% | 6.00% |
| Ba | 5.24% | . | -0.88 | 1.94 | -1.89 | 50.30% | 46.10% | 3.60% |
| B | 0.52% | . | -1.29 | 2.19 | -3.57 | 53.50% | 18.60% | 27.90% |
| Caa-Below | 0.43% | . | -2.75 | 1.24 | -4.28 | 71.00% | 5.60% | 23.40% |
| All Ratings | 100% | . | -0.30 | 0.0875 | -1.70 | 22.40% | 72.90% | 4.70% |
| <i>Panel B: HEL, June 2007</i> | | | | | | | | |
| Aaa | 19.96% | 20.17% | 0.00 | . | na | na | 100.00% | 0.00% |
| Aa | 18.68% | 22.90% | -0.17 | 1.21 | -1.10 | 21.77% | 72.33% | 5.90% |
| A | 23.92% | 24.13% | -0.52 | 1.15 | -1.61 | 39.14% | 51.32% | 9.54% |
| Baa | 27.54% | 26.26% | -0.55 | 1.38 | -1.68 | 40.38% | 49.90% | 9.71% |
| Ba | 9.63% | 6.37% | -0.99 | 0.13 | -2.03 | 52.59% | 44.69% | 2.72% |
| B | 0.13% | 0.08% | -6.6 | . | -6.6 | 100% | 0.00% | 0.00% |
| Caa-Below | 0.13% | 0.08% | -6.6 | . | -6.6 | 100% | 0.00% | 0.00% |
| All Ratings | 100% | 100% | -0.42 | 1.32 | -1.68 | 30.10% | 63.80% | 6.10% |
| <i>Panel C: HEL, February 2008</i> | | | | | | | | |
| Aaa | 12.96% | 7.58% | 1.10 | 1.10 | na | na | 0.00% | 100.00% |
| Aa | 21.77% | 29.33% | 1.38 | 4.00 | -1.09 | 19.04% | 41.33% | 39.64% |
| A | 11.04% | 2.91% | 4.89 | 9.60 | -1.62 | 31.58% | 12.15% | 56.28% |
| Baa | 7.71% | 6.22% | 4.56 | 6.91 | -2.86 | 21.90% | 3.10% | 75.00% |
| Ba | 12.54% | 5.88% | 3.46 | 5.42 | -3.80 | 19.67% | 2.72% | 77.62% |
| B | 13.54% | 5.12% | 1.56 | 3.22 | -5.29 | 17.69% | 4.76% | 77.55% |
| Caa-Below | 20.44% | 42.97% | -1.17 | 1.71 | -3.11 | 53.79% | 17.07% | 29.14% |
| All Ratings | 100% | 100% | 1.99 | 4.58 | -2.70 | 26.02% | 15.19% | 58.79% |

Source: Panel A data from Moody's (2007); Panels B and C are from authors analysis.

Table 2. Relative average rating transitions

This table shows the frequency distribution of relative evaluations on jointly rated tranches in June 2007 and February 2008. Moody's higher (S&P higher) indicates Moody's (S&P) rating is higher than its peer's by no more than 2 notches. Moody's higher > 2 (S&P higher >2) indicates that the rating spread exceeds 2 notches. Percentages are expressed relatively to each rating gap category in June 2007.

| | | February 2008 | | | | | |
|-----------|-------------------|---------------|----------------|-------------------|------------|---------------|---------|
| | | Equal | Moody's higher | Moody's higher >2 | S&P higher | S&P higher >2 | Total |
| June 2007 | Equal | 17.64% | 31.74% | 31.46% | 12.42% | 6.74% | 100.00% |
| | Moody's higher | 4.04% | 18.39% | 63.68% | 8.52% | 5.38% | 100.00% |
| | Moody's higher >2 | 22.22% | 16.67% | 50.00% | 5.56% | 5.56% | 100.00% |
| | S&P higher | 12.69% | 12.39% | 32.63% | 23.87% | 18.43% | 100.00% |
| | S&P higher >2 | 7.53% | 11.64% | 30.82% | 13.01% | 36.99% | 100.00% |

Table 3. Regression model estimation of credit rating gap post versus pre crisis

This table reports the outcome of estimating a cubic model of Delta, where Delta is the measure of the gap in ratings between S&P and Moody's (as defined in Section 3). The dependent variable, Delta_Post, is the Delta value in February 2008. Delta_Pre is the Delta value in June 2007. Standard errors robust to heteroskedasticity are shown in parentheses. The likelihood ratio test is performed considering a second degree model nested in our cubic model. Marginal effects are reported for the minimum positive and maximum negative values exhibiting a negative marginal effect significant at the 10% level (or lower). Standard errors are shown in parentheses. ***, ** and * indicate, respectively, significance levels of 1%, 5% and 10%.

| | Estimated Coefficient | |
|-------------------------|-----------------------|-----|
| Delta_Pre | 0.9590 | *** |
| | (0.08) | |
| Delta_Pre ² | 0.0078 | |
| | (0.02) | |
| Delta_Pre ³ | -0.0070 | *** |
| | (0.00) | |
| Constant | -2.3640 | *** |
| | (0.08) | |
| R ² | 0.04 | |
| N | 3,812 | |
| Likelihood ratio test | 13.65 | *** |
| <i>Marginal effects</i> | | |
| Delta_Pre= -11 | -1.7453 | * |
| | (1.01) | |
| Delta_Pre= 9 | -0.6661 | ** |
| | (0.34) | |

Table 4. Regression model estimation of Moody's and S&P downgrades on jointly rated (JR) and not jointly rated (NJR) tranches

This table reports the result of OLS regressions for HEL tranches of JR and NJR for Moody's and S&P. The dependent variable is the downgrade effected by each CRA between June 2007 and February 2008. Downgrade is measured as the change in the numerical rating scale. Rating_Pre is the CRA rating in June 2007. NJ is a dummy variable taking the value of 1 when the tranche is rated only by Moody's or by S&Ps (but not by both CRAs) and 0 otherwise. Standard errors are shown in parentheses. ***, ** and * indicate, respectively, significance levels of 1%, 5% and 10%.

| | Symbol | Moody's | | S&P | | |
|--------------------|---------------------|---------------------|---------------------|----------------------|--------------------|-----|
| NJ | β_1 | 6.7524 (2.0604) | *** | -15.0225 (1.2486) | *** | |
| Rating_Pre | β_2 | -0.6139 (0.0181) | *** | -0.6858 (0.0207) | *** | |
| NJ*Rating_Pre | β_3 | -0.3317 (0.1339) | ** | 0.9709 (0.0944) | *** | |
| Const. | α | 14.1107 (0.2844) | *** | 17.7144 (0.3487) | *** | |
| Obs. | | 2,840 | | 3,721 | | |
| Adj R ² | | 0.3002 | | 0.2285 | | |
| | $\alpha + \beta_1$ | α' | 20.8632 (2.0407) | *** | 2.6918 (1.1989) | ** |
| | $\beta_2 + \beta_3$ | β'_2 | -0.9456 (0.1326) | *** | 0.2851 (0.0921) | *** |

Figure 1. Expected gap between Moody's and S&P credit rating evaluations

This figure shows the expected gap between Moody's and S&P ratings evaluation in February 2008 as a function of the evaluation gap in June 2007, based on the cubic model developed in the main text. Both the June and the February gap are measured using the variable Delta, as defined in Section 3. Delta pre is the gap in June 2007, while Delta post is the gap in February 2008. The horizontal and vertical axes represent perfectly aligned evaluation after and before the crisis respectively, while the 45° dotted line represents an unchanged gap before and after the beginning of the crisis.

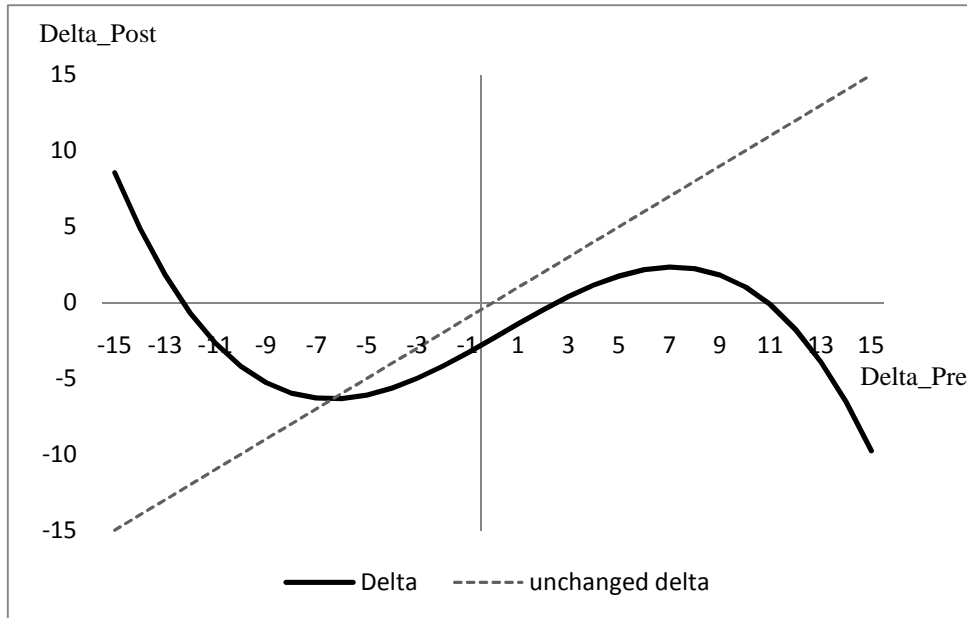


Figure 2. Expected downgrades on jointly rated (JR) and not jointly rated (NJR) tranches

This figure shows the expected downgrade relative to the June 2007 rating – expressed in the main numerical scale (as explained in Section 3) – for both agencies and both JR and NJR tranches. The average NJR line is the average of S&P NJR and Moody's NJR curves.

