The stability and accuracy of credit ratings *

Paulo Viegas de Carvalho
Department of Finance
ISCTE - University Institute of Lisbon
jpvfc@iscte.pt

Paul Anthony Laux
Department of Finance
University of Delaware
laux@udel.edu

João Pedro Pereira
Department of Finance
ISCTE - University Institute of Lisbon
joao.pereira@iscte.pt

First Draft: February 3, 2014
This Draft: September 30, 2014

Abstract

This paper investigates the motivations of Credit Rating Agencies (CRAs) to change their ratings, and the consequences of those changes. We propose a new measure of ratings stability that summarizes the information in a ratings transition matrix into a single scalar number. We find that the intensity with which CRAs change ratings varies through time. In particular, rating changes are more intense during economic bad times. Surprisingly, the periods of stronger rating changes are not associated with higher ratings accuracy. Instead, we find that rating changes are more intense in anticipation of higher corporate bond issuance. These results are consistent with CRAs positioning themselves as useful coordinators between firms and investors to facilitate efficient capital allocation.

JEL classification: G18; G24; G32.
Keywords: Credit ratings; Stability; Accuracy; Coordination.

*We are grateful for the helpful comments of seminar participants at the IFABS 2014 Conference. Financial support from FCT Fundação para a Ciência e Tecnologia under project PTDC/EGE-GES/119274/2010 is gratefully acknowledged.

"Our conversations with investors, issuers and regulators have led us to conclude that many market participants have a strong preference for credit ratings that are not only accurate but also stable. They want ratings to reflect enduring changes in credit risk because rating changes have real consequences—due primarily to ratings based portfolio governance rules and rating triggers—that are costly to reverse. Market participants, moreover, do not want ratings that simply track market-based measures of credit risk. Rather, ratings should reflect independent analytical judgments that provide counterpoint to often volatile market-based assessments."

—*Special Comment*, Moody's Investors Service, September 2006

1 Introduction

The traditional function of Credit Rating Agencies (CRAs) has been independent certification of corporate creditworthiness. Given the quasi-regulatory role that ratings have assumed, ratings are expected to be not only accurate, but also stable (e.g., Altman and Rijken (2006)). Considering that corporate and macroeconomic conditions change over time, this obviously leads to tradeoffs. This article studies CRA behavior with respect to the tradeoff of accuracy and stability. Our evidence suggests that CRAs' deviations from stability may be designed more to promote coordinated actions of issuers and investors, rather than to enhance accuracy.

CRAs claim to pursue stability by implementing rating systems that, according to their own statements, produce estimates of "each issuer's relative fundamental creditworthiness . . . without reference to explicit time horizons", not estimates of absolute probabilities of default (per Cantor and Mann (2006) of Moody's, one of the two dominant CRAs). Similarly, on the website of Standard & Poor's, the other dominant CRA, is the statement (as of September 2014) that ". . . ratings express relative opinions about the creditworthiness of an issuer or credit quality of an individual debt issue, from strongest to weakest, within a universe of credit risk" (Standard and Poor's, 2014). Hence, it seems that changes in macroeconomic variables that affect all firms in the same way should not induce major changes in ratings. In other words, ratings should be more "through-the-cycle" than "point in time." In this case, one would expect to see rating changes with a constant intensity through time, being driven mostly by changes in the characteristics of individual firms and not by changes in business cycle variables.[1]

However, the pattern of rating changes over the last decades suggests that ratings agencies behave in more complex ways, perhaps as a natural response to the incentives of the rating business. In an influential theoretical paper, Bar-Isaac and Shapiro (2013) reason that the quality of CRAs' products (which presumably includes accuracy as a prominent component) will be countercyclical, in that CRAs have incentives to build reputation in bad times (when analyst labor is cheap and mistakes are apt to be noticed) and to milk reputation in good times (when analyst labor is scarce and not many risky firms will default). Bolton, Freixas, and Shapiro (2012) also propose a model where the quality of ratings is more likely to decrease during boom times.[2]

Traditional thinking about CRAs, and even the theories just cited, are premised on the centrality of the "issuer pays" model, in which bond issuers pay a fee to have their bonds rated. This is thought to set up a conflict of interest for the CRAs. Bouvard and Levy (2013) and Frenkel (2014) even argue that CRAs have an incentive to maintain reputations with investors for stringency, as well as with issuers for leniency. In the face of such conflicts of interest, it is not immediately clear how CRAs manage to be economically useful at all.

The answer may come from another stream of theoretical work focusing

[1] Even so, what the rating agencies intend is not entirely clear, as contradictory statements can be found. For example, on the same Standard & Poor's webpage as quoted above we find: "...some risk factors tend to affect all issuers—an example would be growing inflation that affects interest rate levels and the cost of capital." Still, the main sense is that the CRAs mean to register mainly the effects of secular rather than cyclical factors, preferring not to downgrade temporarily only to upgrade later.

[2] Also, Mathis, McAndrews, and Rochet (2009) argue that it may be optimal for CRAs to vary their ratings standards in order to milk their reputation, particularly when rating complex structured products. Mariano (2012) analyzes how competition and reputation concerns affect information revelation by CRAs.

on the fact that CRAs also serve bond investors, who are important indirect customers. A coordination hypothesis, first suggested by Boot, Milbourn, and Schmeits (2006), proposes that CRAs' usefulness derives from coordinating the actions and beliefs of issuers and investors. In their theory, credit ratings may provide a focal point for investors' beliefs about the riskiness of the projects that will be undertaken with bond proceeds, and the threat of credit ratings changes may encourage issuers to undertake projects consistent with the expectations. Importantly, the fact that institutional investors are unwilling or prohibited from holding bonds without investment grade ratings is the "bite" that makes the threat meaningful.

Other variations on the coordination hypothesis have been developed very recently, some of which focus on other channels. For example, ratings changes can influence an issuer's decision to default by providing a focused notice of the likelihood of future funding (see Holden, Natvik, and Vigier (2014) and Manso (2013)). The common feature of these CRA theories is that coordination of issuer and investor expectations is the key reason for CRAs to exist, not prediction of default based on private information or expert processing of public information.

This paper contributes to empirical validation of the theories above by analyzing the causes and consequences of rating changes. We depart from the existing empirical literature by proposing a new measure for the concept of ratings "stability". The main advantage of our measure is to condense in a single (scalar) number all the information for each time period that is sometimes presented in a (two-dimensional) ratings transition matrix. Our measure, denoted Ratings Volatility (*RatVol*), is similar to a standard deviation. It is high when many issuers experience small rating changes or when fewer issuers experience very large rating changes. Hence, it captures more information that other scalar measures of stability typically used by CRAs to characterize their actions, such as the fraction of large rating changes or rating reversals (see, for example, Cantor and Mann (2003)). Being a scalar, our *RatVol* can thus be readily used in time-series tests. Additionally, because the measure aggregates all of a CRA's updating activity, it naturally lends itself to studying CRAs' overall strategies and industrial organization outcomes.

We focus on CRAs' initial ratings and updates to ratings together. Ratings changes, in addition to the initial ratings, seem the obvious laboratory for studying rating stability. Studies that focus on initial ratings only may add to an understanding of CRAs' accuracy, but cannot comment on stability. Using a sample of all domestic long-term credit ratings assigned by Moody's and Standard & Poor's for U.S. non-financial firms between 1994 and 2011, we find the following results.

First, we find that the intensity of rating changes depends on the economic cycle, with CRAs making more intense changes to ratings during bad economic times. These changes during bad times are mostly downgrades; upgrades are insensitive to the cycle. Furthermore, the changes in ratings do not seem to be related to changes in the individual probabilities of default of the rated firms. Hence, despite the CRAs' claims of rating firms through-the-cycle and only relative to each other, our initial results thus suggest that CRAs target a cycle-dependent absolute level of credit risk.

The results with our measure of ratings volatility add to findings in the previous literature that use less-complete measures. For example, Amato and Furfine (2004) find that S&P's ratings of investment grade firms vary according to the state of the business cycle. Nickell, Perraudin, and Varotto (2000) find that Moody's ratings of lowly graded issuers depend on the business cycle.

Second, we provide evidence on theories that propose that ratings quality increases during recessions (Bar-Isaac and Shapiro (2013), Bolton, Freixas, and Shapiro (2012), among others, as discussed above). We measure ratings quality by the Accuracy Ratio (AR). The usage of the accuracy ratio to measure ratings' quality has the advantage of being robust to time-series variation in rating standards and rating inflation (papers documenting variation in rating standards include, for example, Jorion, Shi, and Zhang (2009), Becker and Milbourn (2011), Baghai, Servaes, and Tamayo (2014)). The AR will be high when defaults are concentrated in the firms with the worst rating notations, regardless of the absolute level of the probability of default implied in each rating classification.

We find some evidence of countercyclical CRA accuracy, consistent with the theories. More precisely, we find that the AR is higher towards the end of recessions, when current times are still bad, but signals of better times ahead are already starting to appear.

However, our results also suggest that business cycle theories of CRAs that focus on reputation with issuers are likely incomplete. The reason is that we do not find any significant relation between the volatility and the accuracy of ratings. This is a surprising result. In an issuer-pay rating system, ratings volatility has an implicit cost for the rating agency: the current issuers that are downgraded are less likely to do business with the same CRA in the future. If the CRA is not getting a clear increase in accuracy, which would presumably improve its reputation, then why is it changing ratings more intensely in the first place?

Our final set of results suggest the answer. We find that ratings volatility is higher just before higher activity in the primary bond market. Consistent with the coordination hypothesis as in (Boot, Milbourn, and Schmeits, 2006), the results suggest that CRAs update ratings when it is necessary to "remind" investors that CRAs monitor and penalize firms that deviate from the initial risk profile. The advertising value of the reminder would be greatest when aggregate bond issuance is set to increase, because upcoming paying issuers will want a highly credible rating (i.e., one that helps coordinate with investors). Moreover, to the extent that the existing issuers who are re-graded and the new issuer group do not overlap, the new issuers are not harmed at all by the CRAs' actions. While CRAs must obviously care about accuracy, they also seem to be timing their rating changes in order to attract the business of future bond issuers.

The remainder of our article is structured as follows. Section 2 discusses the sample, analysis variables, and descriptive statistics. Section 3 provides an analysis of ratings volatility, and Section 4 provides an analysis of ratings quality, including the tradeoff between accuracy and stability. Section 5 studies the relation of ratings revisions to future aggregate bond issuance, and the last section concludes.

2 Data and definition of variables

2.1 Credit ratings sample

We start by collecting all credit rating announcements by Moody's and Standard & Poor's directly from the agencies' own databases (Moody's Default and Recovery Database and Standard & Poor's Capital IQ Database). We select all U.S. nonfinancial firms rated by these agencies between 1991 and 2011. For each firm, we select the domestic long-term issuer credit rating. The S&P/Moody's rating notation is converted to the following numerical scale: AAA/Aaa = 1, AA+/Aa1 = 2, ..., D/C = 22. On this scale, a higher number thus corresponds to lower credit quality.

Credit rating agencies (CRAs) revise their rating decisions periodically, usually within 1 or 2 years of the last assessment. However, there is typically no commitment to revise the rating within any given time frame, which means that the time between announcements varies from firm to firm. In our sample, Moody's(S&P) revises the rating for a given issuer after 1.1(1.6) years on average and 83%(73%) of the ratings are revised within 2 years. Hence, we assume that a given rating remains valid for the period of 2 years after it is announced or until a new rating is issued, whichever happens first. We only include a firm in the sample after we observe the first rating announcement and drop it from the sample two years after the last rating announcement.

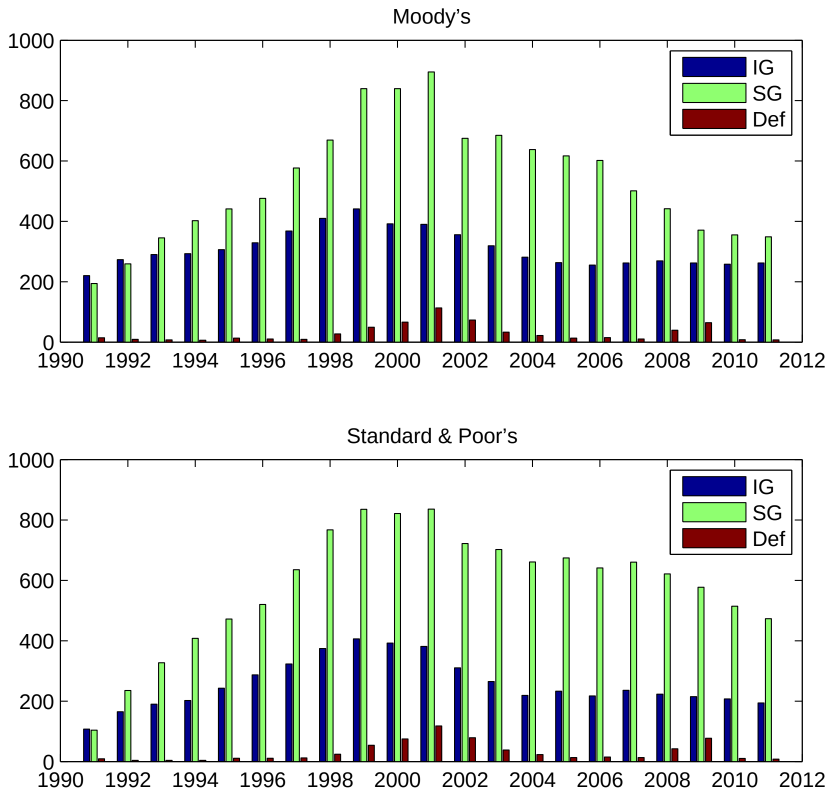
We complement the CRA's ratings databases by gathering information on credit defaults from Bloomberg (CACT: Capital Change; Bankruptcy Filing), CRSP (delisting code 574), COMPUSTAT (inactivation code 02), and from the UCLA - LoPucki Bankruptcy Research Database. We classify a firm as in default as soon as a default is registered in any of these databases.

Figure 1 shows the evolution through time of the number of issuers rated by each CRA. The total number of rated issuers increases up to 2001 and decreases after that. The small number of defaults in the first years of the sample are insufficient to compute accuracy ratios, as described below. Therefore, we exclude the years up to 1994 from the sample.

We conduct our empirical time-series analysis below at a quarterly fre-

Figure 1: Number of issuers rated by each agency

The figure shows, for each year and for each credit rating agency, the total number of firms rated as Investment Grade (IG) or Speculative Grade (SG), and also the number of firms that defaulted (Def).



quency. This frequency is high enough to detect most rating reversals and, at the same time, it is low enough to ensure sufficient rating changes from period to period. Additionally, it is a reasonable frequency to analyze business cycle effects. Hence, our final sample is from 1994Q1 to 2011Q4.

2.2 Volatility of ratings

2.2.1 Definition

We propose a new measure for the stability of credit ratings or, more precisely, for its inverse — the volatility of ratings. It is based on the same information required to compute a standard ratings transition matrix. The main goal of our measure is to condense all the information about rating changes into a single scalar.

Let t measure time in quarters and $t = 1, 2, \ldots, T$ denote the end of each quarter in the sample. Let

$$w_t(s, f) := \frac{n_t(s, f)}{\sum_{s=1}^{K} \sum_{f=1}^{K} n_t(s, f)} \tag{1}$$

where $n_t(s, f)$ is the number of firms that ended the last quarter (time $t-1$) with rating s and ended the current quarter (time t) with rating f. Note that by using the actual *number* of firms (rather than just the percentages that are usually shown in transition matrices) we give more weight to the transition paths with more observations. K denotes the number of rating classes (in our case, $K = 22$, with 22 representing default).

We define the volatility of ratings as

$$RatVol_t := \sqrt{\sum_{s=1}^{K} \sum_{f=1}^{K} w_t(s, f) \times (f - s)^2} \tag{2}$$

This measure is very similar to a standard-deviation. $RatVol_t$ will be high when there are some large rating changes during the quarter, but it can also be high when there are *many* small rating changes. While CRAs already monitor large rating changes (as discussed further below), we argue that a ratings' system is also unstable if it has frequent widespread rating changes, albeit small. $RatVol_t$ uses all the information in the ratings transition matrix and thus captures more fully the concept of ratings' instability.

Table 1 provides descriptive statistics on $RatVol$. The time-series mean is similar for both agencies: approximately 0.7 for Moody's and 0.62 for S&P. The minimum value across both agencies is 0.28 and the maximum is 1.16, which indicates substantial variation through time.

Table 1: Descriptive statistics

This table shows summary statistics for the time series of CRA's choices and outcomes. The sample is 1994Q1–2011Q4.

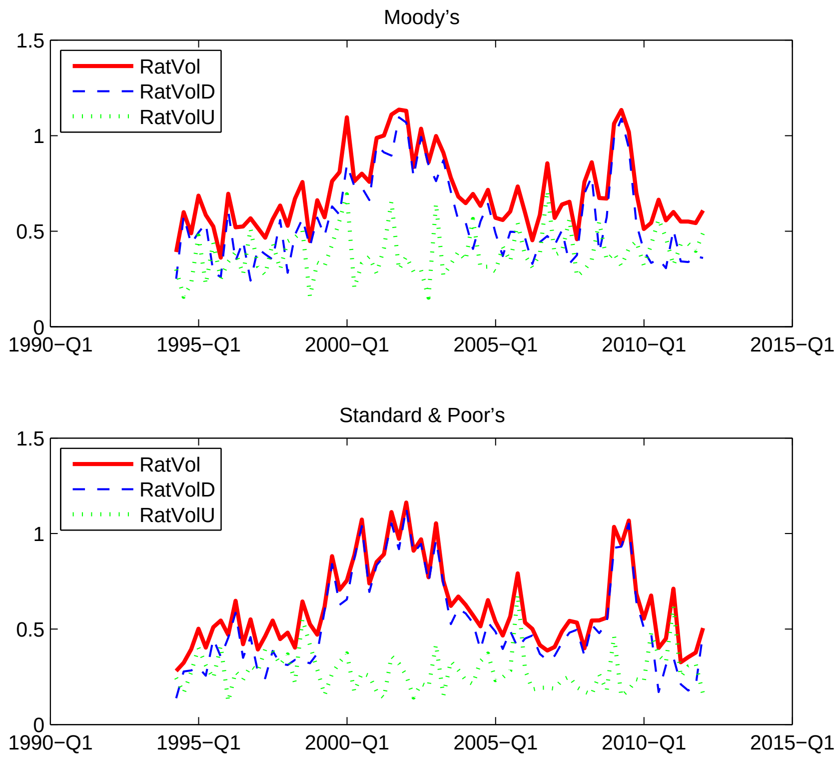
	Mean	Std Dev	Min	Median	Max
Panel A: Moody's					
Average rating level	11.7048	0.6259	10.6352	11.6027	12.6904
Volatility of ratings (RatVol)	0.6969	0.1946	0.3612	0.6578	1.1362
Vol. from upgrades (RatVolU)	0.3828	0.1213	0.1515	0.3672	0.7113
Vol. from downgrades (RatVolD)	0.5590	0.2239	0.2415	0.5002	1.0962
Accuracy ratio of ratings (AR)	0.7149	0.1573	-0.0020	0.7369	0.9922
Rate of default	0.0231	0.0187	0.0019	0.0138	0.0622
Average of PD	0.0437	0.0338	0.0100	0.0250	0.1384
Volatility of PD (PDVol)	0.0300	0.0309	0.0004	0.0142	0.1110
Accuracy ratio of PD	0.7392	0.1944	-0.0351	0.7882	0.9975
Panel B: Standard and Poor's					
Average rating level	11.8627	0.4692	11.0151	12.0364	12.6497
Volatility of ratings (RatVol)	0.6177	0.2176	0.2803	0.5450	1.1626
Vol. from upgrades (RatVolU)	0.2800	0.1103	0.1200	0.2591	0.6809
Vol. from downgrades (RatVolD)	0.5264	0.2484	0.1377	0.4774	1.1339
Accuracy ratio of ratings (AR)	0.7283	0.1369	0.2454	0.7384	0.9897
Rate of default	0.0292	0.0250	0.0040	0.0170	0.0865
Average of PD	0.0597	0.0486	0.0104	0.0376	0.1877
Volatility of PD (PDVol)	0.0317	0.0377	0.0002	0.0128	0.1409
Accuracy ratio of PD	0.7004	0.2024	-0.3158	0.7242	1.0000

We emphasize that our study covers the entire universe of credit rating changes. We begin with the granular data of every change for every issuer, but then summarize the data into one time series for each CRA because we want to study CRAs' behavior and outcomes, not individual bond ratings updates. The point, then, is not to understand the cross-section of a CRA's updates, but the nature of its actions overall, and how CRA's compare.

Figure 2 better illustrates the time evolution of $RatVol$. The periods of 2000–2002 and 2008–2009 clearly stand out as periods of strong instability in ratings.

Figure 2: Volatility of ratings

The figure shows the volatility of the ratings issued by each credit rating agency. $RatVol$ is the total volatility of ratings defined in (2), $RatVolD$ is the volatility due to downgrades defined in (5), and $RatVolU$ is the volatility due to upgrades defined in (4).



2.2.2 Decomposition into downgrades and upgrades

$RatVol$ indexes the effects of both rating upgrades and downgrades. To identify the separate contribution of these two channels, note that the total squared volatility can be decomposed into:

$$\begin{aligned} RatVol_t^2 &= \sum_{s=1}^{K}\sum_{f=1}^{K} w_t(s,f) \times (f-s)^2 \left(I_{\{f<s\}} + I_{\{f>s\}}\right) \\ &= RatVolU_t^2 + RatVolD_t^2 \end{aligned} \tag{3}$$

where the volatilities due to upgrades ($RatVolU_t$) and downgrades ($RatVolD_t$) are

$$RatVolU_t := \sqrt{\sum_{s=1}^{K}\sum_{f=1}^{K} w_t(s,f) \times (f-s)^2 I_{\{f<s\}}} \tag{4}$$

$$RatVolD_t := \sqrt{\sum_{s=1}^{K}\sum_{f=1}^{K} w_t(s,f) \times (f-s)^2 I_{\{f>s\}}} \tag{5}$$

and the indicator function $I_{\{f<s\}}$ equals 1 when the number associated with the final rating (f) is lower than the initial rating (s), i.e., when there is an upgrade.

Figure 2 also shows these series. As expected, most of the ratings' volatility during the crisis periods of 2000–2002 and 2008–2009 is due to downgrades. In good times, both upgrades and downgrades contribute with similar proportions to total volatility. These results are consistent with Nickell, Perraudin, and Varotto (2000) and Amato and Furfine (2004).

2.2.3 Comparison with alternative measures

CRAs compute two alternative measures of ratings' stability (e.g., Cantor and Mann (2003)): *Large Rating Changes*, defined as the ratio of the number of rating changes of three or more notches to the total number of rating announcements; and *Rating Reversals*, defined as the ratio of the number of rating reversals to the total number of rating announcements, where a reversal is an upgrade that was preceded by a downgrade or vice-versa.

To compare our measure of volatility, *RatVol*, with those alternatives, table 2 shows the correlations between them. First, we note that the correlation between *RatVol* and *Large Rating Changes* is high and significant: 0.72 for Moody's and 0.89 for S&P. These correlations are almost equal to the correlations with the part of volatility that is due to downgrades (*RatVolD*), suggesting that CRAs make large changes mostly only when it is necessary to downgrade firms. These high correlations are to be expected due to the square function in (5) which exacerbates any large change. Second, we observe that the correlation between our *RatVol* and *Reversals* is very small (0.21 for Moody's) or even insignificant (for S&P).

Hence, we conclude that our measure of volatility contains information that is different from traditional alternative measures of stability. While *RatVol* is strongly influenced by large rating changes, it provides a full account of all rating transitions, including both small and large rating changes.

Table 2: Correlation between the volatility of ratings and traditional measures of stability

This table shows correlation coefficients between our measure of volatility, *RatVol*, and two other traditional measures of ratings' stability: Large Rating Changes (L.R.C.) and Rating Reversals (Revers.). The values above(below) the main diagonal are for Moody's(S&P). Stars denote statistical significance at the 10% (*), 5% (**), and 1% (***) levels. The sample is 1994Q1–2011Q4.

	RatVol	*RatVolD*	*RatVolU*	L.R.C.	Revers.
RatVol	$\rightleftharpoons$	0.92***	0.25**	0.72***	0.21*
RatVolD	0.95***	$\rightleftharpoons$	−0.12	0.74***	0.13
RatVolU	0.09	−0.21*	$\rightleftharpoons$	−0.02	0.19
L.R.C.	0.89***	0.86***	0.04	$\rightleftharpoons$	0.27**
Revers.	−0.14	−0.16	0.07	−0.08	$\rightleftharpoons$

2.3 Accuracy of ratings

The accuracy ratio (AR) is a well-established indicator to evaluate the performance of a rating system (e.g., Cantor and Mann (2003), Altman and

Rijken (2006)). The AR measures how well a particular rating system discriminates risk, that is, it measures the correlation between ratings and defaults. The AR can be interpreted like a correlation coefficient: if all defaults are concentrated in the worst rating categories (C, CC, ...), the AR is +1; if defaults are unrelated to ratings, the AR is 0; if all defaults are concentrated in the best rating categories (AAA, AA, ...), the AR is -1.[3]

Following Cantor and Mann (2006) we compute the AR over a one-year horizon. Let $r = 1, 2, \ldots, 21$ index the rating category, ordered from highest credit risk (excluding default) to lowest credit risk. We compute the accuracy ratio at time t as:

$$AR_t = \frac{\sum_{i=\underline{r}}^{\bar{r}} [n(i) - n(i-1)] \, [d(i) - n(i) + d(i-1) - n(i-1)]}{1 - D/N} \tag{6}$$

where

$\underline{r}$	=	minimum r in the sample at t.
$\bar{r}$	=	maximum r in the sample at t.
$n(r)$	=	$\sum_{i=\underline{r}}^{r} N_i/N$, for $r \geq 1$, and $n(0) = 0$
$d(r)$	=	$\sum_{i=\underline{r}}^{r} D_i/D$, for $r \geq 1$, and $d(0) = 0$
N_r	=	number of issuers with rating r at t.
D_r	=	number of issuers with rating r at t that defaulted over the following year.
N	=	total number of issuers at t.
D	=	total number of defaults over the following year.

Table 1 shows descriptive statistics for the accuracy ratio. The time-series average of AR is very similar for both CRAs, namely, 71% for Moody's and 73% for S&P (these values are consistent with the 82.6% for Moody's reported in Cantor and Mann (2003)). Nevertheless, the distributions are different across the two CRAs: while the AR of S&P never drops below 25%, the AR of Moody's reaches a minimum of -0.2%. On the other hand, both agencies are able to reach impressive accuracies of almost 100% in some

[3]The Accuracy Ratio contains exactly the same information as the Area Under the Receiver Operating Characteristic (AUROC) curve. As shown in Engelmann, Hayden, and Tasche (2003), $AR = 2 \times AUROC - 1$.

periods.

Interestingly, S&P is able to achieve a slightly higher accuracy than Moody's (73% versus 71%) with a *lower* average volatility of ratings ($RatVol =$ 62% for S&P versus $RatVol =$ 70% for Moody's). In other words, S&P's rating system is both more accurate and more stable than Moody's.

2.4 Business cycle variables

We use the following standard macroeconomic variables to characterize the business cycle:

GDP_t	:=	Real GDP growth over quarter t.
$YieldSlope_t$	:=	Yield curve slope (10-year minus 2-year Treasury Bond yields) at t.
$CreditSprd_t$	:=	Credit spread (BBB - AAA yields) at t.
VIX_t	:=	CBOE volatility index at t.
ADS_t	:=	Aruoba-Diebold-Scotti Business Conditions Index at t.

The yield curve slope is commonly used to characterize the business cycle, with an inverted yield curve usually signalling an incoming recession. A high credit spread typically means current bad times. The VIX proxies for overall economic uncertainty. The Aruoba-Diebold-Scotti Business Conditions Index is published by the Federal Reserve Bank of Philadelphia. It is designed to track real business conditions at high frequency. It has an average of zero and progressively bigger positive values indicate progressively better-than-average conditions.

Given that our paper focuses on credit risk, we also use the rate of default to better characterize the current economic conditions. We compute the rate of default as the fraction of firms that defaulted over the past year:

$RateDef_t$	:=	Number of firms that defaulted over the year (from $t-4$ to t) divided by the number of rated firms at the beginning of the year ($t-4$).

Note that this rate of default is computed for each CRA.

2.5 Estimation of credit risk with public information

To benchmark the accuracy and stability of ratings, we compute an easily available alternative measure of credit risk. Specifically, we follow Shumway (2001) and Campbell, Hilscher, and Szilagyi (2008) and estimate the probability of default for each firm through a logit model using only publicly available information:

$$PD_{i,t} = \frac{1}{1 + \exp\left[-(X_{i,t}\beta + \varepsilon_{i,t})\right]} \tag{7}$$

where $PD_{i,t}$ is the probability, estimated at time t, that firm i defaults over the following year.

To select the best predictors ($X_{i,t}$), we start with all the variables of Shumway (2001) and Campbell, Hilscher, and Szilagyi (2008), and then identify the best ones for our sample through a top-down variable selection procedure, taking into account the statistical and economic significance of each coefficient, as well as the variance inflation factors.[4] Table 3 shows our best model specification and the estimation results.

Table 1 shows descriptive statistics for the probabilities of default estimated with the model specified in table 3. On average, the firms rated by Moody's have a PD of 4.37%, while the firms rated by S&P have 5.97%. The table also shows the Accuracy Ratios for the logit model, computed in a similar way as described for ratings. The AR of the estimated PD for Moody's clients is 74% and for S&P's clients is 70%. These values are very close to the AR of the ratings issued by the CRA's, suggesting that our simple model of assessing credit risk using only publicly available information is as powerful as agency ratings.

Similarly to the volatility of ratings, we also compute the volatility of the probabilities of default over quarter t as

$$PDVol_t := \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{PD_{i,t}} - \widehat{PD_{i,t-1}})^2} \tag{8}$$

[4] All market and accounting variables for each firm are from CRSP and COMPUSTAT.

Table 3: Logit model for credit default prediction

This table reports the estimates of a logistic regression of credit default on the following variables: log of total assets (Size), total debt divided by market value of assets (TDLM), annual standard deviation of the firm's daily stock return (Sigma), net income divided by the market value of assets (NIATM), total liabilities divided by total assets (LTAT), cash and short term investments divided by the market value of assets (CHATM), and the market-to-book ratio (MB). Heteroskedasticity-consistent standard errors are obtained using Hubber-White estimators.

	Estimates	Robust S.E.	z-value	p-value
Intercept	-12.34	0.24	-52.41	< 0.00
Size	0.54	0.02	26.97	< 0.00
TDLM	1.78	0.24	7.49	< 0.00
Sigma	46.03	1.74	26.42	< 0.00
NIATM	-5.17	0.27	-19.04	< 0.00
LTAT	3.41	0.15	22.02	< 0.00
CHATM	-1.99	0.41	-4.84	< 0.00
MB	-0.36	0.05	-6.93	< 0.00
Pseudo R^2		0.433		
Wald χ^2		4019.4 (p-value < 0.000)		
Observations		109767		

where $\widehat{PD_{i,t}}$ is the probability of default for firm i estimated through (7) and N is the number of firms rated by a particular CRA. A high value of $PDVol_t$ means that the credit risk of the firms in the sample (as measured by public information) changed a lot over the current quarter. We use this as a control for ratings volatility.

3 Anaylsis of the volatility of ratings

This section documents how rating changes depend on the business cycle. Both simple univariate conditional means comparisons and more formal regression tests show that ratings are *not* "through-the-cycle".

3.1 Univariate analysis of business cycle effects

We start with a preliminary univariate analysis of how rating changes depend on the state of the business cycle as described by a single macroeconomic variable. We cut the sample across the median value of a given business cycle variable, and then compare the means of $RatVol$ for the quarters where the business cycle variable is above or below its median.

Table 4 shows the results. We find that on average the volatility of ratings is significantly higher in periods when GDP growth is lower than usual, when the yield curve slope is higher than usual, when the credit spread is higher than usual, or when VIX is higher than usual. The ADS index shows a similar picture: when its value is lower, $RatVol$ is higher. Hence, the results indicate that both Moody's and S&P change their ratings with more intensity during bad or uncertain economic periods.

3.2 Multivariate regression analysis

We now proceed to more formal tests of whether ratings change according to the business cycle. We run the following regression for each CRA:

$$\begin{aligned} RatVol_t = \alpha + \beta_1 GDP_t + \beta_2 YieldSlope_t + \beta_3 CreditSprd_t + \beta_4 VIX_t \\ + \gamma_1 RateDef_t + \gamma_2 PDVol_t + \varepsilon_t \quad (9) \end{aligned}$$

Table 4: Business cycle effects on the volatility of ratings: univariate tests

This table compares two conditional means of the volatility of ratings. The sample is split in two according to the macroeconomic variable identified in each row. The first(second) column shows the mean of *RatVol* for the subsample where the macroeconomic variable is below(above) its median. The last two columns show the t-statistic and corresponding p-value for the test that the two conditional means are equal. The sample is 1994Q1–2011Q4.

	Below median	Above median	t-stat	p-value
Panel A: Moody's RatVol				
Split on GDP	0.769	0.632	3.138	0.002
Split on YieldSlope	0.650	0.747	-2.169	0.033
Split on CreditSprd	0.594	0.804	-5.389	0.000
Split on VIX	0.628	0.766	-3.173	0.002
Split on ADS	0.769	0.628	3.267	0.002
Panel B: Standard and Poor's RatVol				
Split on GDP	0.692	0.561	2.695	0.009
Split on YieldSlope	0.583	0.666	-1.673	0.099
Split on CreditSprd	0.505	0.745	-5.702	0.000
Split on VIX	0.536	0.708	-3.691	0.000
Split on ADS	0.709	0.540	3.599	0.001

In addition to typical macroeconomic business cycle variables, the model includes two control variables specific to each CRA: $RateDef_t$, as CRAs are likely to be particularly sensitive to changes in the intensity of defaults; $PDVol_t$, to control for changes in the underlying characteristics of rated firms that might induce CRAs to update their ratings. As an alternative description of the business cycle and to account for possible collinearity between the business cycle variables, we also estimate a model where the four business cycle variables (GDP_t, $YieldSlope_t$, $CreditSprd_t$, and VIX_t) are replaced by the single ADS_t index.

The results in table 5 show that both Moody's and S&P make more changes to their ratings in bad times. When GDP growth is lower, the credit spread is higher, or the rate of default is higher, CRAs increase their $RatVol$. This result is robust to measuring the state of the business cycle with the ADS index, that is, when the ADS is lower (bad times), the volatility of ratings is higher. The explanatory power of the model is similar for both CRAs, with adjusted R^2 above 60%.

To further analyze the response of CRA to business conditions, we decompose the total ratings' volatility into upgrades and downgrades (as defined in equation 3), and run separate regressions for $RatVolU$ and $RatVolD$. The results in table 5 show that the component that is most sensitive to the business cycle is $RatVolD$ (downgrades), that is, both Moody's and S&P downgrade more during bad times (periods with lower GDP growth, higher credit spread, higher default rate, or lower ADS index) and downgrade less during good times. On the contrary, upgrades seem to be completely unrelated to the business cycle (the adjusted R^2 is even negative).

Surprisingly, we find that $PDVol_t$ is statistically insignificant in every regression. Once we control for changes in the overall business conditions (which affect all firms in the same direction), further changes in the characteristics of individual firms do not seem to induce changes in ratings. This suggests that CRAs may be updating their ratings mostly according to the cycle and not so much in response to idiosyncratic changes in individual firms. Hence, our results are *not* consistent with the CRA's claim that rating notations are just classifications of firms *relative* to each other (see, for

Table 5: Regression of ratings volatility

This table shows OLS estimates of the regression in (9), with Newey-West standard-errors with 4 lags. The variables are as defined in the text. Values in parenthesis are t-ratios. Stars denote statistical significance at the 10% (*), 5% (**), and 1% (***) levels. The sample is 1994Q1–2011Q4.

	Moody's						**Standard & Poor's**					
	$RatVol_t$		$RatVolD_t$		$RatVolU_t$		$RatVol_t$		$RatVolD_t$		$RatVolU_t$	
Constant	0.51***	0.54***	0.32***	0.38***	0.41***	0.38***	0.40***	0.44***	0.30***	0.32***	0.27***	0.30***
	(7.34)	(32.38)	(4.79)	(16.06)	(5.76)	(17.62)	(5.49)	(18.93)	(4.15)	(12.28)	(6.67)	(14.37)
GDP_t	−2.03**		−2.12**		−0.15		−1.00		−1.53*		0.60	
	(−2.15)		(−2.44)		(−0.19)		(−1.13)		(−1.70)		(1.38)	
$YieldSlope_t$	−0.05**		−0.05*		−0.01		−0.08***		−0.11***		0.02	
	(−2.30)		(−1.94)		(−0.78)		(−3.55)		(−4.19)		(0.84)	
$CreditSprd_t$	9.86**		11.02**		2.47		11.38**		13.85***		−0.28	
	(2.09)		(2.06)		(0.76)		(2.34)		(2.70)		(−0.06)	
VIX_t	0.00		0.00		0.00		0.00		0.00		0.00	
	(0.65)		(1.23)		(−1.36)		(0.67)		(0.19)		(0.37)	
ADS_t		−0.11***		−0.13***		−0.01		−0.08***		−0.10***		0.01
		(−6.06)		(−5.71)		(−0.54)		(−3.76)		(−3.89)		(0.33)
$RateDef_t$	6.67***	6.32***	7.84***	7.62***	0.08	−0.26	7.05***	6.00***	8.54***	7.04***	−0.91*	−0.53
	(6.00)	(7.06)	(6.57)	(8.00)	(0.11)	(−0.41)	(6.25)	(4.97)	(7.17)	(5.41)	(−1.96)	(−1.21)
$PDVol_t$	0.09	−0.40	−0.27	−0.67	0.51	0.20	−0.24	−0.34	−0.36	−0.47	−0.03	−0.06
	(0.13)	(−0.72)	(−0.34)	(−0.92)	(0.96)	(0.39)	(−0.44)	(−0.58)	(−0.72)	(−0.78)	(−0.07)	(−0.15)
R-squared	0.60	0.63	0.62	0.63	−0.06	−0.04	0.64	0.61	0.70	0.65	−0.04	−0.02

example, Cantor and Mann (2006) and Standard and Poor's (2014)). Instead, they suggest that CRAs may be actually targeting an absolute level of credit risk.

In summary, the results in tables 4 and 5 consistently show that both Moody's and S&P change their ratings more intensely in bad times. In other words, these results show that ratings are *not* "through-the-cycle".

4 Analysis of the quality of ratings

Given that the intensity of rating changes varies through time as documented above, we now proceed to investigate how CRAs manage their ratings quality, as measured by the accuracy ratio (AR). While we find moderate evidence that the accuracy ratio depends on the business cycle, we fail to find support for the hypothesis that stronger ratings changes lead directly to higher accuracy.

4.1 Univariate analysis of business cycle effects

Similarly to section 3, we start with a preliminary univariate analysis of how the accuracy ratio depends on the state of the business cycle as described by a single macroeconomic variable. We cut the sample across the median value of a given business cycle variable, and then compare the means of AR for the quarters where the business cycle variable is above or below its median.

The results in table 6 show that the accuracy ratio of Moody's is higher when either the yield curve slope, the credit spread, or the VIX is higher than usual, which are typically indicators of bad economic times. Standard & Poor's also has slightly higher average accuracy ratios during bad times according to these indicators, but the differences between good and bad times are not statistically significant. Hence, our results are moderately supportive of the idea that the quality of ratings is higher in recessions.

Table 6: Business cycle effects on the accuracy ratio of ratings: univariate tests

This table compares two conditional means of the accuracy ratio of ratings. The sample is split in two according to the macroeconomic variable identified in each row. The first(second) column shows the mean of the accuracy ratio for the subsample where the macroeconomic variable is below(above) its median. The last two columns show the t-statistic and corresponding p-value for the test that the two conditional means are equal. The sample is 1994Q1–2011Q4.

	Below median	Above median	t-stat	p-value
Panel A: Moody's Accuracy Ratio				
Split on GDP	0.707	0.715	-0.206	0.838
Split on YieldSlope	0.663	0.760	-2.762	0.007
Split on CreditSprd	0.665	0.758	-2.652	0.010
Split on VIX	0.674	0.747	-2.017	0.048
Split on ADS	0.717	0.705	0.343	0.733
Panel B: Standard and Poor's Accuracy Ratio				
Split on GDP	0.707	0.750	-1.337	0.186
Split on YieldSlope	0.710	0.750	-1.255	0.214
Split on CreditSprd	0.721	0.738	-0.518	0.606
Split on VIX	0.720	0.739	-0.571	0.570
Split on ADS	0.740	0.720	0.619	0.538

4.2 Multivariate regression analysis

We now proceed to a more formal analysis of the determinants of ratings' accuracy. In addition to business cycle variables, we also test the effect of our measure of ratings' volatility. Since CRAs claim to tradeoff accuracy against stability (e.g., Cantor and Mann (2006)), one would expect that higher *RatVol* would be accompanied by a higher Accuracy Ratio. Hence, we run the following regression for each CRA:

$$
\begin{aligned}
AR_t = \alpha + \beta_1 GDP_t + \beta_2 YieldSlope_t + \beta_3 CreditSprd_t + \beta_4 VIX_t \\
+ \gamma_1 RatVol_t + \gamma_2 RateDef_t + \gamma_3 AR_{t-1} + \varepsilon_t \quad (10)
\end{aligned}
$$

The model also includes the CRA-specific rate of default, *RateDef*, to help describe the overall economic conditions, and the first lag of the accuracy ratio to account for persistence in this series.

To obtain a robust characterization of the relation between the business cycle and the accuracy ratio, we estimate this regression with two different econometric methods. First, we do simple OLS regressions (with Newey-West standard errors) for each CRA taken in isolation.

Then, we also estimate the model as a seemingly unrelated regression (SUR) system for both CRAs together, imposing the restriction that coefficients on the business cycle variables are the same for both CRAs. The restriction is appropriate under the assumption that both CRAs models are driven by the same economic forces with respect to the business cycle. In estimating the SUR system, standard errors must be calculated using a method that is both robust to the time-series correlations in the data, and that also has good properties in small samples. To take account of these issues, we estimate standard errors using the bootstrap approach, where bootstrap observations are drawn in clustered fashion, a year at a time, to preserve time series correlations in the data up to 4 quarterly lags.

Table 7 shows the results. In the single-equation model for S&P (panel A), we find that the coefficient on the yield curve slope is positive, while all other coefficients are insignificant. For Moody's there is also a positive relation between accuracy and the yield curve slope, but now we find a contaminating effect of a positive coefficient on GDP growth, which would suggest more accuracy in good times. The restricted dual-equation models (panel B) help to sort out the potential collinearity of the business cycle variables. The SUR estimates show a strongly statistically significant positive coefficient on the yield curve slope, that is, the accuracy ratios of both agencies increase when the interest rate curve is upward slopping. However, the coefficients on all the other business cycle variables are insignificant.

To the extent that the yield curve slope is higher in recessions (typically due to very low short-term rates), our results support the idea that accuracy increases in bad times. However, our results are only moderately supportive of the notion that accuracy increases in bad times since the estimates on the other business cycle variables are inconclusive. We suspect that an important reason is that our specific business cycle indicator measures are all fairly strongly correlated with each other, making it difficult to identify separate effects. At the same time, they vary in terms of whether the focus on the real economy (i.e., GDP) or the financial sector (VIX and $CreditSprd$), and about whether they measure current conditions (GDP), prospects for the near-term future (VIX) or the somewhat longer-term future ($YieldSlope$). Therefore, we check further using an omnibus summary measure of the business cycle that should resolve multicolinearity issues, and middle-ground set of two measures to allow for richer dynamics.

Specifically, to further investigate the effect of the business cycle on ratings' quality, we estimate (10) with GDP, $YieldSlope$, $CreditSprd$, and VIX replaced by two alternative business cycle descriptors as follows.

First, we use the Aruoba-Diebold-Scotti Business Conditions Index, ADS. Table 7 shows that the coefficients on ADS are negative for both agencies, which suggests less accuracy in good times, but the estimates are not statistically significant. Interestingly, in this specification the rate of default for Moody's shows up with a strongly significant positive coefficient, indicating that the accuracy of Moody's is higher in periods with higher defaults. It is possible that this measure over-summarizes, and that some offsetting effects are lumped together.

Table 7: Regression of Accuracy Ratio

This table shows estimates of the regression in (10) under two different methods. Panel A contains single-equation OLS estimates with Newey-West standard errors. Panel B shows SUR estimates of two-equation systems in which both Moody's and S&P are required to have the same coefficient estimates on business cycle variables, and standard errors are computed with a cluster bootstrap. The variables are as defined in the text. Values in parenthesis are t-ratios. The sample period is 1994Q1–2011Q4.

	Panel A: Single equation models						Panel B: Restricted dual equation models					
	Moody's	*S&P's*	*Moody's*	*S&P's*	*Moody's*	*S&P's*	*Moody's*	*S&P's*	*Moody's*	*S&P's*	*Moody's*	*S&P's*
Constant	0.4952	0.3377	0.4162	0.3519	0.6535	0.3897	0.4501	0.3477	0.4099	0.3632	0.5424	0.4402
	(5.698)	(5.165)	(5.871)	(8.305)	(6.178)	(8.164)	(4.994)	(4.423)	(3.502)	(4.669)	(5.483)	(6.438)
GDP_t	0.0071	0.0028					0.0041	0.0041				
	(2.483)	(0.657)					(1.188)	(1.188)				
$YieldSlope_t$	0.0705	0.0196					0.0428	0.0428				
	(2.155)	(1.746)					(2.721)	(2.721)				
$CreditSprd_t$	-0.0160	0.0074					-0.0042	-0.0042				
	(-0.572)	(0.341)					(-0.154)	(-0.154)				
VIX_t	0.0044	0.0001					0.0019	0.0019				
	(1.794)	(0.038)					(1.002)	(1.002)				
ADS_t			-0.0029	-0.0058					-0.0046	-0.0046		
			(-0.348)	(-0.604)					(-0.381)	(-0.381)		
$PC1_t$					0.0294	0.0085					0.0177	0.0177
					(2.472)	(1.714)					(1.819)	(1.819)
$PC2_t$					0.0628	0.0168					0.0362	0.0362
					(2.362)	(1.450)					(3.505)	(3.505)
$RatVol_t$	-0.2147	0.0207	-0.1426	0.0128	-0.2451	0.0093	-0.1722	0.0121	-0.1334	0.0058	-0.1900	-0.0095
	(-4.356)	(0.236)	(-4.383)	(0.126)	(-4.496)	(0.113)	(-2.597)	(0.156)	(-1.887)	(0.064)	(-3.344)	(-0.129)
$RateDef_t$	0.4340	-0.1095	1.8412	0.4142	0.6205	-0.0035	0.9067	-0.5875	1.7856	0.4166	1.0378	-0.4346
	(0.671)	(-0.126)	(4.236)	(0.571)	(1.379)	(-0.004)	(1.215)	(-0.703)	(2.911)	(0.611)	(1.742)	(-0.548)
AR_{t-1}	0.2978	0.4653	0.4949	0.4929	0.3042	0.4629	0.3883	0.4248	0.4960	0.4837	0.3920	0.4276
	(2.572)	(9.265)	(5.712)	(13.114)	(2.818)	(7.946)	(3.266)	(5.818)	(3.317)	(7.660)	(3.150)	(5.941)
Observations	70	70	70	70	70	70	70	70	70	70	70	70
R-squared	0.407	0.293	0.293	0.280	0.401	0.291	0.387	0.275	0.293	0.279	0.382	0.275

Next, we form two principal component summaries of the business cycle variables to parsimoniously capture their variations without excessive multi-collinearity. Specifically, we use the full time series of the change in GDP, yield curve slope, credit spread and VIX as the basis for extracting these principal components. For analysis purposes, we retain the first two principal components to use as summaries of business cycle variation. We have examined the correlations of these two principal components measures with the raw business cycle variables. The first principal component, $PC1$ in our notation, is correlated most strongly with the credit spread and the VIX, with other correlations being less than half as important as those two. Therefore, we see $PC1$ as a measure of the fear of the downside of the business cycle. As to the second principal component, $PC2$, the largest correlations are to the change in GDP and the yield curve slope. Considering that an upward-sloping yield curve is known to portend future growth in the economy and to be steepest before such periods, we see $PC2$ as being an index of the current and future upside to the business cycle. Very loosely speaking, we have extracted a "fear-oriented" $PC1$ and a "hope-oriented" $PC2$.

Table 7 also provides regression estimates of accuracy ratios using these principal components regressors. In both the single equation models and the restricted dual equation models, it appears that $PC1$ and $PC2$ effectively register some business-cycle variations in accuracy. Focusing on the restricted dual equation models, we have that the fear-oriented $PC1$ is associated with more accuracy at a modestly-high ($< 10\%$) level of statistical significance in a two-tail test. Especially given that perhaps a one-tail test is more appropriate, this suggests that the CRAs are more accurate when the economy is pricing the fear of broadly bad economic outcomes into bonds and stocks. The hope-oriented $PC2$ is also associated with a positive effect on accuracy ratios, and is very strongly statistically significant. The interpretation is that accuracy is higher when better times are starting to appear, i.e., GDP is turning up and the bond market signals growth for the future. Overall, we take these results as supportive of the theoretical prediction that CRAs quality (measured as accuracy in this study) is stronger when current economic conditions are not strong, but when reputation needs to be built for future improved conditions.

The model in (10) also tests the relation between ratings' accuracy and stability. If CRAs were willing to induce less stability during some periods in order to increase accuracy, we would expect to find a positive coefficient on our measure of ratings' volatility, *RatVol*. However, the results in table 7 fail to support this tradeoff.

For S&P, the coefficient on *RatVol* is always insignificant. For Moody's, we even find a surprisingly negative impact of *RatVol* on AR, that is, after controlling for business cycle effects, stronger rating changes by Moody's seem to be associated with *less* accuracy. However, this result is not robust to different specifications of the regression model. In particular, if we exclude all business cycle variables and regress AR on *RatVol* only, or even on several lags of *RatVol*, we find no significant relation between AR and *RatVol* (nor for Moody's, nor for S&P).[5]

Hence, our results do not support the notion that CRAs may be willing to induce more ratings' instability to increase accuracy during some periods. We investigate this relation further in the following section.

4.3 Further analysis of the relation between accuracy and stability

4.3.1 Alternative measures of stability

Our results above do not support the hypothesis that CRAs might be trading off stability against accuracy through time. To confirm that this lack of relationship is not due to our particular measure of ratings' stability, we also analyze other measures typically used by CRAs.

Figure 3 compares the relation between the Accuracy Ratio and three alternative measures of instability. The first panel plots the AR against our measure of stability, *RatVol*. If there was a tradeoff, we should observe an upward slopping plot of AR versus *RatVol*, i.e., as *RatVol* increased (stability decreased), accuracy should also increase (similarly to figure 1 in

[5]Results not shown, but available upon request.

Cantor and Mann (2006)). The figure shows instead a mostly flat relation — higher ratings' volatility (lower stability) is *not* compensated by higher accuracy.

The second plot in figure 3 shows *AR* against *Large Rating Changes* (number of rating changes of three or more notches over the total number of rating announcements) and the third plot against *Rating Reversals* (number of rating reversals over the total number of rating announcements). CRAs routinely state that they aim to avoid both large rating changes and rating reversals (e.g., Cantor and Mann (2003)). In both cases, we find a mostly flat relation, that is, the accuracy ratio does not seem to depend on Large Rating Changes nor on Rating Reversals. Hence, the lack of a tradeoff between accuracy and stability is not due to our particular measure of ratings' volatility.

4.3.2 The incremental effect of rating changes

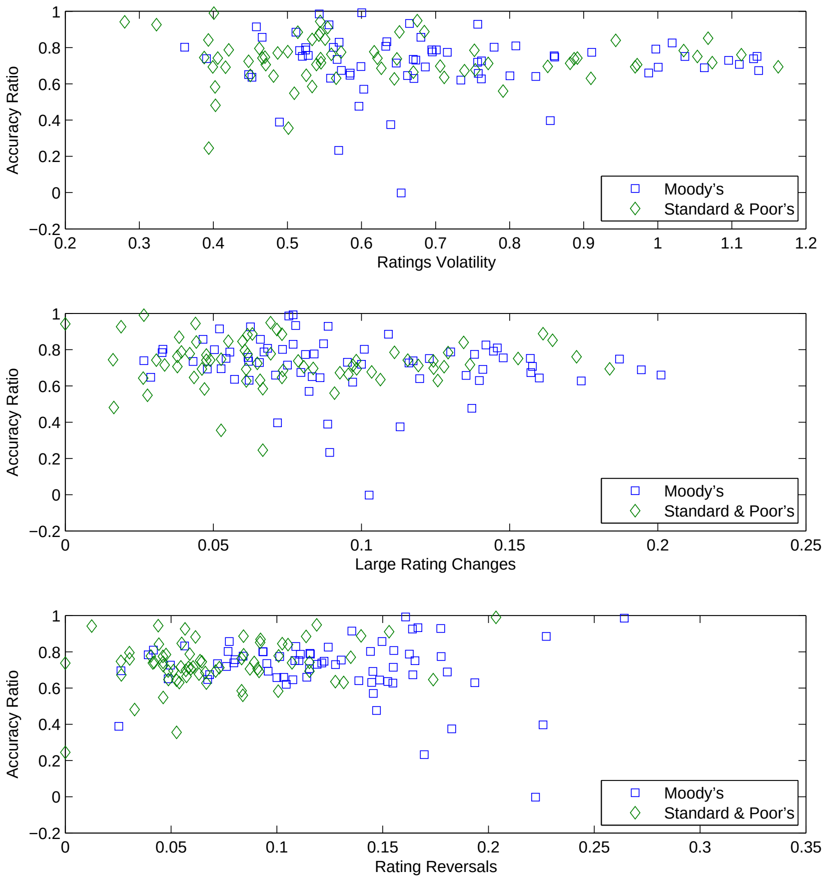
The previous sections fail to support the idea that periods of stronger rating changes lead to higher accuracy. One natural question is then whether rating changes have any information content at all, that is, whether rating changes at least help to *maintain* accuracy.

To investigate this question, we formulate a null hypothesis that rating changes are just random changes, without any predictive power for future defaults. Under this hypothesis, the AR of the latest ratings issued at t would be just as high as the AR if ratings had not changed from $t-1$. To be more precise, denote the AR defined in (6) as $AR_{t,(t,t+4)}$ to stress that this is the AR of ratings issued at time t for defaults occurring in $(t, t+4)$. We want to compare it with $AR_{t-1,(t,t+4)}$, that is, with the ability of old ratings issued at $t-1$ to predict defaults over the same $(t, t+4)$ period.

The sample means for Moody's are the following: the accuracy of current ratings is $E[AR_{t,(t,t+4)}] = 0.71$, while the accuracy of old ratings is $E[AR_{t-1,(t,t+4)}] = 0.64$, with a t-test rejecting that these means are equal (p-value = 0.01). For S&P, the accuracies are $E[AR_{t,(t,t+4)}] = 0.74$ and $E[AR_{t-1,(t,t+4)}] = 0.66$, with a t-test again rejecting that these means are

Figure 3: Accuracy versus stability

This figure shows, for each CRA, the relation between the accuracy ratio (defined in (6)) and three alternative measures of stability. The top panel uses our *RatVol* (defined in 2); the middle panel uses *Large Rating Changes* (number of rating changes of three or more notches over the total number of rating announcements); and the bottom panel uses *Rating Reversals* (number of rating reversals over the total number of rating announcements). Each point in the figure represents a given quarter. The sample is 1994Q1–2011Q4.



equal (p-value = 0.005). As we use older ratings, the average accuracy ratio deteriorates even further.

These results thus reject the hypothesis that rating changes have no information content regarding future defaults. While the estimates for regression (10) in the previous section indicate that CRAs are not actively managing the accuracy of their ratings through time, the current results support the common-sense assumption that most rating changes do at least help to maintain accuracy.

Nevertheless, the strong variability through time of *RatVol*, as documented in section 2.2, suggests that CRAs may be doing more than just updating their ratings in a smooth piece-meal fashion just to maintain accuracy. The following section investigates other potential motivations for CRAs to change the intensity of their rating updates through time.

5 The relation between the volatility of ratings and bond issuance

Given the surprising results in the previous sections that rating changes seem to just follow the business cycle and not lead to higher accuracy, we now further investigate other potential reasons for rating changes. Recall that most of the volatility of ratings comes from downgrades. Since firms do not like to be downgraded, ratings volatility represents an implicit cost to CRAs in the form of "unhappy customers". Hence, there must be some reason why CRAs are willing to incur this cost.

If improvements in their information content do not explain why CRAs update ratings, perhaps other theories of CRA functioning can help us understand. In particular, Boot, Milbourn, and Schmeits (2006) show that "coordination" might be the underlying purpose for CRAs to exist. They argue that credit ratings provide focal points for both bond investors and issuers in terms of anchoring issuers' credit risk levels. Bond investors want comfort that the issuers' projects will, in fact, embody the anticipated risks. Because ratings agencies can and do revise their ratings (and put firms on credit watch, which is the mechanism that Boot, Milbourn, and Schmeits (2006) focus on most), investors will learn if firms behave differently. And because institutional investors often have credit ratings as a key criteria for their willingness to hold an issue, the implied threat has "bite." CRAs then do more than provide information, they provide a specific mechanism that fosters the successful functioning of the bond market more broadly.

In view of these theories, we ask whether coordination might be an important explanation for CRAs to update ratings, and might lie behind the observed ratings volatility. We reason that CRAs who want to sell ratings to new issuers in a competitive environment must be able to show that they are effective at the core job of CRAs: coordination, under this reasoning. Updating ratings, thereby bringing "bite" to the coordinating threat, is a visible sign to issuers and investors that the agency is active and effective. Under the coordination reasoning, issuers are more willing to pay CRAs that provide the bite because investors are more willing to lend to firms that are rated by them. Though the bond issuers pay for the ratings, bond investors are also a strong customer component.

This leads us to a specific hypothesis. We expect that CRAs will update ratings more aggressively shortly prior to periods of strong bond issuance. It is most worth the cost and strain of doing so when new customers will soon be available. Ratings updates, under this hypothesis, are a marketing tool. Formally, of course, we investigate if the null of this hypothesis can be rejected.

Tests of this hypothesis are informative from an additional point of view. Recall that we have found evidence that CRA ratings accuracy increases when measures of impending economy-wide improvements are strong. Bond issuance surges during economic upturns. Therefore, if evidence favors the hypothesis above, we have uncovered suggestive evidence as to why CRA actions might follow a business cycle pattern: following the business cycle pattern helps a CRA be ready to take advantage of a strong bond market.

Note that we are not proposing that ratings volatility causes bond issuance. Bond issuance, at an aggregate level, is somewhat predictable based on economic conditions and has a degree of persistence. Given that, what we are proposing is that CRAs will become diligent and active when they anticipate an increased level of bond issuance, and thus an increased selling potential.

To test this hypothesis, we collect data on total bond issuance from all nonfinancial U.S. corporations from Mergent FISD. The data series is available at the monthly frequency in nominal values, so we first use the CPI to convert it to real dollars, and then aggregate to quarterly values by adding the corresponding monthly values. Let $BndIss_t$ denote the total bond issuance during quarter t in constant 2000 billion dollars. We run the following regression for each CRA:

$$BndIss_t = \alpha + \beta_1 BndIss_{t-1} + \beta_2 MARatVol_{t,\ldots,t-3} + \beta_3 MARatVol_{t-4,\ldots,t-7} + \gamma_1 Macro_{t-1} + \gamma_2 Macro_{t-2} + \varepsilon_t \quad (11)$$

where $MARatVol_{t,\ldots,t-3}$ is a moving average of $RatVol$ over the current year (quarters t to $t-3$) and $MARatVol_{t-4,\ldots,t-7}$ is an average over the previous year. We use these moving averages to obtain a less noisy series than the original series with quarterly observations. $Macro_{t-1}$ is a vector with lagged macroeconomic control variables that may influence the decision to issue bonds: economic growth (GDP), interest rates ($YieldSlope$ and $CreditSpread$), and overall uncertainty (proxied by VIX). We include a lag of bond issuance to control for persistence in this series.

Table 8 shows the results. Specifically, the left side of the table reports on bond issuance predictions using lags of Moody's ratings volatility, while the right side of the table uses lags of Standard & Poor's ratings volatility. Coefficient estimates show that lagged bond issuance is not statistically significant, which helps alleviate concerns that our results are biased by excessive persistence in the dependent variable. The lagged business cycle regressors, except for GDP, are often not highly statistically significant, which is probably a consequence of their multicollinearity, as discussed earlier. Nonetheless, as a group they function effectively as business cycle control variables.

Current and lagged ratings volatility is a significant and positive correlate of future bond issuance, both in the case of Moody's and Standard & Poor's. Both CRAs update their ratings more strongly prior to higher bond issuance, consistent with the notion that they are preparing for such conditions. More precisely, the results show a persistent long-term relation, in the sense that both higher average ratings volatility over the current year (quarters $t, \ldots, t-3$) and higher average ratings volatility over the prior year (quarters $t-4, \ldots, t-7$) are associated with higher current bond issuance. Thus, the evidence suggests that coordination with the bond market is displayed in the data, in the sense of strong ratings update activity before strong issuance periods.

In addition to the results shown in 8, we have tested this hypothesis in other ways. In particular, we have examined whether residuals from a fitted value equation for ratings volatility (using our previous regression formats as the basis for computing the fitted value) are predictive of bond issuance. We find that they are.

6 Conclusion

Credit rating agencies (CRAs) balance competing goals. Investors and regulators demand accuracy, but also value stability and a long-run view in ratings. Issuers, who pay directly for the ratings, prefer the highest possible ratings in the first instance, but also want ratings with strong credibility so that they support bond holdings by institutional investors. The dominant CRAs, Moody's and Standard & Poors, have been roundly criticized in recent years for bad handling of these tradeoffs. In the extreme, they have been viewed as pandering to issuers in a less-than-competitive marketplace.

Our paper provides evidence on the forces that drive CRAs balancing of accuracy versus stability. At the core of our empirical work is a new measure of ratings instability that summarizes into a single number the information typically presented in two-dimensional transition matrices.

Using this new measure, we find that both Moody's and Standard & Poor's vary the intensity of rating changes through time. In particular, both CRAs have more volatile ratings during bad economic times, which is

Table 8: Regression of bond issuance

This table shows estimates of the regression in (11), with the variables as defined in the text. Values in parenthesis are t-ratios. The sample is 1994Q1–2011Q4.

	Moody's		Standard and Poor's	
$BndIss_{t-1}$	0.0995	0.0163	0.0828	-0.0047
	(0.744)	(0.118)	(0.643)	(-0.035)
$MARatVol_{t,\dots,t-3}$	189.0176	126.3909	182.7027	111.9321
	(3.404)	(2.056)	(3.885)	(1.969)
$MARatVol_{t-4,\dots,t-7}$		135.9018		102.7895
		(2.150)		(2.065)
VIX_{t-1}	-0.2689	-0.0529	0.1823	-0.1883
	(-0.229)	(-0.044)	(0.158)	(-0.157)
GDP_{t-1}	9.4933	8.3897	6.9816	7.3873
	(3.249)	(2.733)	(2.330)	(2.380)
$YieldSlope_{t-1}$	-4.4575	-10.7285	-7.8169	-5.0497
	(-0.196)	(-0.467)	(-0.350)	(-0.225)
$CreditSpread_{t-1}$	29.8162	23.9481	23.0453	24.0529
	(1.526)	(1.209)	(1.194)	(1.228)
VIX_{t-2}	0.3259	0.6651	0.6285	0.5782
	(0.283)	(0.540)	(0.559)	(0.488)
GDP_{t-2}	-3.6954	-4.6168	-5.7050	-5.1865
	(-1.189)	(-1.453)	(-1.888)	(-1.678)
$YieldSlope_{t-2}$	-16.3491	-25.3805	-10.1624	-27.6636
	(-0.726)	(-1.110)	(-0.459)	(-1.162)
$CreditSpread_{t-2}$	-15.5949	-11.8797	-24.5901	-9.4838
	(-0.788)	(-0.592)	(-1.247)	(-0.448)
Constant	-72.8971	-101.1747	-23.2201	-49.6817
	(-1.954)	(-2.496)	(-0.729)	(-1.399)
Observations	69	65	69	65
R-squared	0.525	0.529	0.548	0.544

inconsistent with the CRAs' claim that ratings are only relative rankings of firms and thus mostly independent of the business cycle. Instead, CRAs do seem to target implicit absolute levels of credit risk.

Surprisingly, we do not find evidence that higher ratings volatility leads directly to higher ratings accuracy. In other words, rating changes do not seem to be driven by information motives. Instead, our results show that CRAs increase the volatility of ratings in advance of more issuance in the primary bond market.

Our evidence suggests that coordination of expectations and actions between issuers and investors is an important determinant of CRAs' ratings updating process. Though a coordination role has been suggested in theory, there is little prior evidence of its importance. In this role, ratings put a clear label on the levels of risk in issuers' projects. The threat of downgrades encourages issuers to keep risk within the indicated range, especially because institutional investors use ratings as a central criteria in sorting for investable bonds. Though any one issuer may consider volatility in its own ratings—and in particular downgrades—to be undesirable, issuers overall benefit from credible ratings to support this coordination role. Credible ratings require a level of volatility as business cycle conditions change. Our specific finding that CRAs update more intensely before periods of strong issuance activity is consistent with them putting more "bite" into their ratings overall just before periods when they will need to sell ratings to new issuers who will want them to be strongly credible with investors.

One possible interpretation of this evidence is to regard it as an advertising story. Given that ratings volatility has an implicit cost in the form of "unhappy" old customers that are downgraded, why are CRAs willing to incur the cost? It may be the price of advertising that they will be credible coordinators between future bond issuers and investors. Our results are consistent with CRAs that do not possess a significant information advantage, but that instead try to position themselves as useful coordinators between investors with varied risk appetites and issuers with projects of varied credit qualities.

References

Altman, E. I., and H. A. Rijken, 2006, "A point-in-time perspective on through-the-cycle ratings," *Financial Analysts Journal*, pp. 54–70.

Amato, J. D., and C. H. Furfine, 2004, "Are credit ratings procyclical?," *Journal of Banking & Finance*, 28(11), 2641–2677.

Baghai, R. P., H. Servaes, and A. Tamayo, 2014, "Have rating agencies become more conservative? Implications for capital structure and debt pricing," *The Journal of Finance*, Forthcoming.

Bar-Isaac, H., and J. Shapiro, 2013, "Ratings quality over the business cycle," *Journal of Financial Economics*, 108(1), 62–78.

Becker, B., and T. Milbourn, 2011, "How did increased competition affect credit ratings?," *Journal of Financial Economics*, 101(3), 493–514.

Bolton, P., X. Freixas, and J. Shapiro, 2012, "The credit ratings game," *The Journal of Finance*, 67(1), 85–111.

Boot, A. W., T. T. Milbourn, and A. Schmeits, 2006, "Credit ratings as coordination mechanisms," *Review of Financial Studies*, 19(1), 81–118.

Bouvard, M., and R. Levy, 2013, "Two-Sided Reputation in Certification Markets," *Available at SSRN: http://ssrn.com/abstract=2178119.*

Campbell, J., J. Hilscher, and J. Szilagyi, 2008, "In Search of Distress Risk," *Journal of Finance*, 63, 2899–2939.

Cantor, R., and C. Mann, 2003, "Measuring the performance of corporate bond ratings. Special comment.," Moody's Investors Service.

——— , 2006, "Analyzing the tradeoff between ratings accuracy and stability. Special comment.," Moody's Investors Service.

Engelmann, B., E. Hayden, and D. Tasche, 2003, "Testing rating accuracy," *Risk*, pp. 82–86.

Frenkel, S., 2014, "Repeated interaction and rating inflation: A model of double reputation," *American Economic Journal: Microeconomics*, Forthcoming.

Holden, S., G. J. Natvik, and A. Vigier, 2014, "An Equilibrium Theory of Credit Rating," Working paper, University of Oslo.

Jorion, P., C. Shi, and S. Zhang, 2009, "Tightening credit standards: the role of accounting quality," *Review of Accounting Studies*, 14(1), 123–160.

Manso, G., 2013, "Feedback effects of credit ratings," *Journal of Financial Economics*, 109(2), 535–548.

Mariano, B., 2012, "Market power and reputational concerns in the ratings industry," *Journal of banking & Finance*, 36(6), 1616–1626.

Mathis, J., J. McAndrews, and J.-C. Rochet, 2009, "Rating the raters: are reputation concerns powerful enough to discipline rating agencies?," *Journal of Monetary Economics*, 56(5), 657–674.

Nickell, P., W. Perraudin, and S. Varotto, 2000, "Stability of rating transitions," *Journal of Banking & Finance*, 24(1), 203–227.

Shumway, T., 2001, "Forecasting bankruptcy more accurately: A simple hazard model," *The Journal of Business*, 74(1), 101–124.

Standard and Poor's, 2014, "About Credit Ratings," http://www.standardandpoors.com/aboutcreditratings/.