SOVEREIGN RATINGS IMPLIED BY COUPLED CDS-BOND MARKET DATA

STEFANO MARMI, ALDO NASSIGH, AND DANIELE REGOLI

Abstract. We propose an approach to sovereign market implied ratings based on informations coming both from Credit Default Swap spreads and bond spreads in a unified way. Operationally speaking, we implement a Support Vector Machine type of selection in the plane CDS-bond. Our numerical results seem to confirm that introducing the bond dimension accounts for implied ratings more accurate and with greater predictive power with respect to the 1-dimensional CDS implied ratings.

Contents
1. Introduction 1
2. Market implied ratings: various existing approaches 2
3. sovereign market implied ratings: a CDS penalty-function analysis 3
4. CDS+bond based analysis 5
5. statistical analysis: SVM vs. KV implied ratings 10
5.1. accuracy 11
5.2. precision & recall 15
5.3. gap-conditioned analysis 17
6. Conclusions 19
Appendix A. data details 24
Credit Default Swaps 24
Bonds 24
Agency ratings 25
References 25

1. Introduction

Ratings provided by credit rating agencies (CRAs) are meant as indicators/opinions about the issuer’s ability to pay back their debt in due time [1, 2, 3]. Their role in the modern financial system could be hardly overestimated. They are of key importance in at least three respects: for investors, who need information on the risks in which they are going to incur if investing in certain products; for issuers of financial products, who need ratings in order to attract investors and inform them on their creditworthiness; finally, credit ratings have become increasingly important in regulatory policies: indeed, financial institutions can use credit ratings from approved agencies in order to assess their capital requirements.

Given that their importance is out of doubt, nevertheless there is plenty of criticism about CRAs and ratings themselves, in particular after their crucial role during the 2008 mortgage crisis and the 2011 euro-zone crisis.

Leaving aside the structural critics, one of the main concerns about agency ratings is their (physiological) delay in capturing the actual health or creditworthiness of the issuer. This is the reason why, aside and complementing the information of agency ratings, the idea of capturing some kind of early warnings...
extracted by the market has become increasingly popular. These warnings are known as market implied ratings or “point in time” ratings.

It is worth to stress that agency ratings and market implied ratings should be considered as complementary to each other: the stability in time of agency ratings and their independence of short term market movements are important features, that investors and emitters look for. At the same time, in particular for the investor side and for portfolio management, the market is a source of high frequency information that should not be neglected and can very well suit the role of giving sort of temperature to have a prompt idea of the ‘health’ of debt emitters.

It is in this spirit that market implied ratings have grown in importance and popularity, and that are now concocted and sold by CRAs themselves besides the ‘standard’ ratings (see, e.g. [4, 5, 6, 7, 8, 9, 10]).

In this paper we focus on market implied ratings for sovereign debt emitters and we propose the idea of using both the market of Credit Default Swaps and that of bonds. Indeed, in [11] qualitative arguments are presented, by which, for sovereign debt emitters and in particular in situations in which various countries share a common currency as in the Euro area, CDS alone may not capture the exact information about the real creditworthiness of a state. As far as we know, ratings derived by market data have up to now been calculated with one single source of information as input, namely ratings implied by CDS or bonds or equity. The original proposal of this work is precisely to present a way of calculating market implied ratings taking into account in one single procedure two different sources of market information. The procedure proposed is nonetheless sufficiently flexible to be generalized to different sectors and different input variables.

The paper is structured as follows: section 2 briefly reviews the literature on market implied ratings; in section 3 we derive CDS-implied ratings for our panel of sovereign countries, using the technique of [4] and [5]; we describe our proposal in section 4, and then calculate the implied ratings for our dataset; finally, section 5 is devoted to the statistical analysis of the performances of the two sets of implied ratings. Concluding remarks follow.

2. Market implied ratings: various existing approaches

Most of the literature on market implied ratings refers to corporate agents and typically deals with Credit Default Swap spreads, bonds and equity data. The first attempt in this direction dates back to 1974, with the so-called Merton model [12] whose idea is to see equity as a call option on assets with debt as strike, use equity and equity volatility market data to invert the relation, thus finding the asset value and asset volatility of the firm: from assets and debts a “distance to default” is calculated, which then can be mapped to a probability of default and to a rating score. This methodology, which is at the core of what are now called the “structural” models for default probabilities, has become very popular during the years, and it is actually implemented by both Moody’s and Fitch as an early warning tool over their long-term ratings (Moody’s KMV and Fitch EIR [9, 8]).

Other, non-structural, attempts make a direct use of CDS or bond spreads, which are indeed a proxy of the credit risk of the issuer (see, for instance [13]). One of the first proposals (2003) can be found in [4] and is based on the assumption that “agency ratings are, on average, informative”. In this study, the authors create a penalty function minimizing which results in obtaining spread regions of implied ratings. This approach was recovered by Jaiming Kou and Simone Varotto in 2008 [5] with some modifications and with statistical analysis to measure the goodness of the implied ratings in anticipating the agencies down/upgrades. The same penalty-function approach is implemented by Fitch [7], with the difference (among others) that the dataset is now composed by CDS spreads rather than bond spreads. Moody’s MIR - Market Implied Ratings [9] provides three types of market implied ratings: bond-implied, CDS-implied and equity-implied. While we have already mentioned the latter, bond-implied and CDS-implied MIR ratings are derived by filtering the data excluding extreme values, then calculating the median spread for each agency rating class and taking boundaries between two classes by geometric means of two contiguous median spreads. Standard & Poor’s MDS - Market Derived Signals [6] uses a linear regression approach, with regressors: agency rating class, the industrial sector (for the corporate case), the Document Type (the definition of what is a credit default) and others.
In a recent paper [14] a proposal that uses unsupervised learning methods\(^1\) was presented, namely a clustering algorithm based on an Hidden Markov Model, taking as inputs surviving probability curves bootstrapped by CDS spreads.

As far as we are aware, all the published studies are mainly dedicated to the corporate sector with some minor hints to sovereigns. Here we deal explicitly with the sovereign sector, firstly calculating implied ratings with a standard penalty-function methodology over CDS (which we call, for simplicity, Kou-Varotto (KV) approach), and then proposing and implementing a new methodology for calculating market implied sovereign ratings (but portable to corporates as well). We then use some statistical metrics to compare the performances of the two implied ratings sets.

3. Sovereign market implied ratings: a CDS penalty-function analysis

In this section, we are going to briefly review the ideas of using a penalty-function to calculate implied ratings from CDS spreads and then we present the results of such a calculation on a pool of 36 sovereign countries for a 10-year time interval starting from 2004.

The basic assumption for this kind of non-structural approaches, mentioned in the previous section, is that the agency ratings are considered on average informative, but with “errors”, due e.g. to delay of the agencies in officially updating their opinion on the country’s ability to pay back its debt. Namely, in this kind of approaches, the agency ratings are inputs to the calculation. Thus, the problem one faces is the following: we have, for each trading day, a certain number of CDS spreads, let us say one for each country we are considering\(^2\); attached to each country-spread there is a rating label (given by a chosen agency, or by an average rating among all agencies); if agency ratings exactly captured the credit risk inherent in the CDS spreads on that particular day, then the spreads of each rating class would be perfectly separated along the CDS dimension: typically, this is not the case. We refer to figure 1 for an example: you can see that AAA and AA rated countries (according to S&P’s) do not perfectly separate along the CDS dimension, even if, on average, they do. The penalty-function approach aims to find a clever way to draw a boundary in the spreads in order to give a new label (the implied ratings) on the basis of that boundary, as clarified in figure 1.

The penalty function is derived from \([4]\) and \([5]\), and has the following expression (for AAA/AA boundary)

$$
\text{penalty}(b) = \frac{1}{N_{AAA}} \sum_{j=1}^{N_{AAA}} (s_j^{AAA} - b)^+ + \frac{1}{N_{AA}} \sum_{j=1}^{N_{AA}} (b - s_j^{AA})^+ ,
$$  \(3.1\)

where:

- \(N_{AAA} (N_{AA})\) is the number of AAA (AA) rated countries,
- \(s_j^{AAA} (s_j^{AA})\) is the spread of the \(j\)-th country among the AAA (AA) rated ones,
- \(b\) is the spread of the boundary that we are looking for,
- \((\cdot)^+\) denotes the positive part.

The penalty function increases its value whenever a AA rated country has a spread lower than \(b\) or a AAA rated country trades at spreads higher than \(b\). Minimizing \(3.1\) with respect to \(b\) gives the desired spread of the boundary between AA implied and AAA implied rated spreads. What we are actually doing is to choose the boundary that minimizes the “errors” of the agency ratings; the measure of the error is simply the distance from the boundary itself.

The dataset we shall use comprehends a set of 36 countries\(^3\) covering all the rating classes; 5-year maturity CDS spreads (being the more liquid ones) as given by Markit; long-term ratings according to Standard & Poor’s; a time interval going from January, 1st, 2004 to January, 14th, 2014. In order for statistical analysis to be more reliable, we map the S&P’s rating classes\(^4\) into 5 larger classes, according to the following scheme:

\[\text{AAA} \rightarrow \text{AAA}\]

---

\(^1\)See, e.g., [15].

\(^2\) Actually, for each country there are CDS for various tenors, but let us focus on one single maturity.

\(^3\)See the appendix A for more details on the dataset.

\(^4\) See, e.g., the S&P’s website [www.standardandpoors.com](http://www.standardandpoors.com) for reference to S&P’s ratings classes and their meaning.
Figure 1. snapshot of the 5-year maturity CDS spreads of AAA, AA rated countries (according to S&P’s) in early October 2008, with a sketch of CDS implied rating construction for one day.

AA+,AA,AA- → AA
A+,A,A- → A
BBB+,BBB,BBB- → BBB
BB+,BB,BB-,B+,B,B-,CCC+,CCC,CCC-,CC,SD → BB

the last of which is necessary since we have few data for the single categories of lowest ratings.

The methodology is the following: for each day collect the spreads of the past 30 days (around 22 trading days); divide all the spreads according to the 5 rating classes; minimize the penalty function 3.1 for each couple of nearby ratings (AAA/AA, AA/A, ...) thus finding 4 boundaries per day. The reason for using data of the entire previous month is to prevent the implied ratings to be too volatile and too connected to the daily movements of the market: Kou and Varotto in [5] collect data of two months before the desired date. We believe that an accurate analysis of the best temporal window from which collecting the data to calculate the daily boundary is worth doing, and we shall possibly deal with that in future work.

Figure 2 shows the result of the calculation: you see the 4 barriers during the 10-year period. Of course, the barriers move in time, reflecting the fact that there are some dynamical features common to all the ratings, depending on common underlying scenarios. Indeed, the values of spreads are not influenced by credit risk alone: market-wide liquidity situation, average risk aversion of investors, are just some of the factors impacting on CDS premiums. From the plot, it is easy to see that before the 2008 crisis all the spreads were extremely low and rather stable, then you have a huge increase in all the spreads, with very high correlation among the different classes; there is a peak in early 2009, then a decrease in late 2009 and 2010, and then a new bump, nearly as high as the previous one, during the so-called euro-zone sovereign crisis; then again a decrease in 2013. Notice that, after the triggering of the 2008 crisis, the spread barriers never revert to something comparable to the pre-crisis situation.
The core idea is that a crossing of one of these barriers by the CDS spread related to one country detects precisely a change in the genuine credit risk part of the information encoded in spreads.

Once you have the boundaries, calculating the implied ratings for the countries is straightforward. As an example, we plot the dynamics of the S&P’s and implied ratings for three countries, Italy, Spain and Germany in figures 3, 4, 5 respectively.5

Some comments on the results shown in figures 3, 4, 5: firstly, it is clear that the KV implied ratings anticipate all the downgrades, both for Italy and Spain; secondly, the anticipation horizon is quite long: strong signals of downgrade starts between 1 and 2 years before the actual rating adjustment, sometimes even more; KV implied ratings seem to be a little bit on the pessimistic side, particularly regarding the recent years. These stylized features are common to the majority of the 36 countries analyzed. Moreover, focusing on the German situation, we see that the CDS market was giving signals of downgrade for many days in a row, both for the 2008 crisis and for the more recent euro-zone sovereign distress period, while the CRAs didn’t change their excellent opinion on the creditworthiness of Germany.

4. CDS+bond based analysis

As we mentioned at the end of the previous section 3, KV analysis gives signals of downgrade for Germany during the two main crisis of the last few years. Actually, even if German CDS spreads were raising, resulting in downgrade market signals, we know that precisely in that periods, Germany, in particular German bund, was the “safe haven” of the entire euro-zone. Indeed, as figure 6 shows, in 2011-2013, while the CDS spread was very high, the bond was actually very low. The very same pattern was in place during the 2008 crisis as well. As was pointed out in [11], for sovereign countries the CDS alone may not be able to capture enough information to assess credit risk. In [11] it is also argued that taking into account both CDS and bonds should improve the ability of deducing from the market the rating for countries. And that is precisely what we are going to investigate in what follows.

5The dataset for Italy starts in 2006 and not in 2004 because of a lack of data for the previous period.
Figure 3. Kou-Varotto implied rating together with historical S&P’s rating for Italy

Figure 4. KV implied rating together with historical S&P’s rating for Spain
Figure 5. KV implied rating together with historical S&P’s rating for Germany

Figure 6. plot of bond spread (upper) and CDS spread (lower) for Germany
Adding 5-year maturity bonds means adding one dimension to the input space: a single sovereign state is represented no more by a single number (the CDS spread), but by a couple bond spread + CDS spread. Thus, as time goes on, each country describes a trajectory in the bond-CDS plane. Now the problem amounts to divide, on a daily basis, such a plane in regions, each corresponding to a different implied rating. As in the one dimensional case we needed a prescription (the Kou-Varotto minimization approach) by which separate the points into implied-rating classes, we need a prescription here as well.

One natural choice to classify points in multi dimensional spaces is to use the so called Support Vector Machine approach (by which separate the points into implied-rating classes, we need a prescription here as well. The minimization problem to uniquely fix the linear boundary reads:

\[
\text{maximize } \frac{1}{2} \left( \beta \cdot \beta \right) \quad \text{subject to } y_i (\beta^T x_i + \beta_0) \geq M, \forall i
\]

(4.1)

which is pretty self explanatory.

Actually, there a scale invariance in the problem: to see this, notice that one can replace the constraints \(\| \beta \| = 1\), \(y_i (\beta^T x_i + \beta_0) \geq M\) with

\[
y_i (\beta^T x_i + \beta_0) \geq M \| \beta \|
\]

(where \(\beta_0\) has been redefined appropriately); now notice that this constraint is invariant to rescaling of the vector \((\beta_0, \beta)\). Then, we can arbitrarily fix the norm \(\| \beta \|\). A smart choice is

\[
\| \beta \| = \frac{1}{M}.
\]

Then, the problem (4.1) is equivalent to the following

\[
\min_{\beta, \beta_0} \frac{\| \beta \|^2}{2} \quad \text{subject to: } y_i (\beta^T x_i + \beta_0) \geq 1, \forall i
\]

(4.2)

In most cases, and for sure in our situation, data are not linearly separable. SVM prescribes a procedure to deal with non-linearly separable data, as in the example of the right panel of figure 7: we still try to maximize the margin, but we allow some points to be inside the margin or even on the “wrong” side of the boundary line. These “errors” are weighted with the parameter \(\xi_i\) that is the distance by which they enter inside the margin. The minimization problem to uniquely fix the linear boundary reads:

\[
\text{optimization } \left\{ \begin{array}{l}
\min_{\beta, \beta_0} \left( \frac{\| \beta \|^2}{2} + C \sum_{i=1}^{n} \xi_i \right); \quad C \rightarrow \text{cost parameter} \\
\text{subject to: } y_i (\beta^T x_i + \beta_0) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1 \ldots n
\end{array} \right.
\]

(4.3)
Figure 7. Support Vector Machine separating plane: linearly separable (left) and non-linearly separable (right) cases.

where $C$ is an exogenous parameter that measures how much we are willing to make “errors” at all. The minimization problem is then a trade off between the maximization of the margin and the minimization of the cost we pay when making “errors”.

Actually, the real utilization of SVM is to use that boundary line to classify new data of which we know $x$ but we don’t know $y$. That’s precisely the philosophy behind all the Machine Learning algorithms: to train the machine with some known data and to use it to classify unknown ones. We have instead to face a slightly different problem: we want to separate points on a plane of which we know the classification (agency ratings) but we also know that this classification contains “errors”, and we need a good way to perform that separation.

Of course, SVM is not confined to work on a plane: SVM can deal with an arbitrary (finite, but in some cases also infinite) number of dimensions. Moreover, SVM can deal with non-linear boundaries as well, via so called kernel methods. In the present work we confine ourselves to linear boundaries and to 2-dimensional analysis, but these are two features that we are going to investigate in future research: adding to CDS spreads and bonds some extra information, and analyzing the possibility of implementing some curved decision boundary.

The dataset we work with is the following: same 36 countries, ratings and CDS spreads as for the previous section 3; 5-year maturity bond interest rate spreads over zero-interest curves, as calculated by UniCredit risk management staff. The time interval is again January, 1st, 2004 → January, 14th, 2014.

The methodology for calculation is as follows:

- for each trading day, we collect CDS spreads and bond spreads for the 36 countries for 30 days backward in time, resulting in approximately 750 input data per day,
- with these data we calculate the boundary lines for that day, with the SVM procedure,
- having the plane for that day divided into implied rating regions, it is easy to infer the implied rating for each CDS-bond couple for that day.

All the calculations were done with R open software [21] and in particular with package e1071 [22] and caret [23] for SVM analysis and cross-validation tuning respectively.

Actually, there are a couple of technicalities we must dwell on a while: one regarding the cost parameter $C$, and the other referring to the multi-class nature of our problem. The cost parameter $C$ must be fixed a priori of the actual boundary calculation. Remember that fixing $C$ means fixing the price we are willing to pay for “misplaced” points on the plane. In the literature on SVM, one can find several approaches regarding the parameter tuning of the support vector approach, among which we choose the cross-validation. Namely, the routine does the following: she divides in 10 parts all the training data for that day ($\sim$ 750 points), then uses 9/10 of the data as training set on which she calculates the boundaries, and the other 1/10 as a test set. This procedure is repeated for each 1/10 of the data set, and the overall procedure is repeated for 10 times independently of one another. Then, the accuracy (i.e. how many
points of the test set are classified correctly by the trained boundaries), averaged over all these trials, is calculated. This is done for every value of the cost parameter along the sequence \((2^{-3}, 2^{-2}, \ldots, 0.2, 4)\). Finally, the cost parameter that maximizes the accuracy is chosen. This procedure is run through every ten training days, keeping the latest optimal cost for the other 9 days. The overall scheme results in a further (this time statistical) optimization of the separation boundary.

Now with the multi-class issue: the SVM procedure we described above, refers to binary classification problems, namely, problems where data belongs to one among 2 classes. Here, instead, we are dealing with a 5-class classification. The routine we used adopts the so-called one-versus-one approach, which runs like this: she calculates 10 linear boundaries, 1 for each couple of rating classes; then, each point on the plane in assigned an implied rating by a voting scheme: the point is classified according to all the 10 couples, and the rating class that obtains more “votes” is the one assigned to that point. This procedure determines, daily, the implied rating regions of the CDS-bond plane\(^6\).

Another minor technical remark is about the asymmetry of data: namely, for each day we have (sometimes big) differences in the numbers of points for each class (namely, we have a lot of AAA countries but few BB, see appendix A). Given that accuracy is the metric for cross-validation, unbalanced data may result in unbalanced model selection. The solution to this problem is to assign a cost parameter \(C\) different for each class, weighted on the basis of the size of the training dataset of each class. See, e.g., [16] for some details.

Unfortunately, showing the 2-dimensional boundaries evolving in time is neither illustrative nor easy as in the one-dimensional case. We are going to show just a couple of daily snapshots, in order to give an idea of how the SVM works. Some general comments regarding the two figures 8, 9:

- interestingly, you see that while the high ratings (AAA, AA) are separated from the others along the SE-NW direction, the low ones tend to be separated along the orthogonal direction: this means that while for high rated countries CDS and bond spreads tend to be negatively correlated, the opposite holds for low rated countries,
- a comparison of the two plots clarifies the very idea of implied ratings and of SVM as well: in fig. 9 you see that Romania and Turkey, despite being rated BB by Standard & Poor’s, have actually quite low bond spreads, and not very high CDS spreads as well (comparable to BBB values); SVM decides however to rate them BB, minimizing the errors; in fig. 8 on the other hand, while Romania is still reclassified BB, Turkey is too mixed with AA guys, and classifying it BB would have cost way too much: thus the calculation puts Turkey in the AA region. Indeed, the market on those days whose trading Turkey debt issues and CDS at the level of most AA countries.

As for the KV case, let us show the time evolution of implied ratings for Italy, Spain and Germany. Figures 10, 11, 12 shows the results: historical S&P’s ratings, SVM implied ratings and, for comparison, KV implied ratings. Regarding the issues we addressed for KV implied ratings, we see here that SVM classification has at least two differences/improvements:

- it is more timely, meaning that the downgrading signals usually start nearer to the actual rating change,
- it is more “optimistic”, meaning that tends to give higher implied ratings, in particular in the recent period.

Regarding the German situation, we clearly see that adding the bund has the (expected) result of getting rid of the downgrading signals both for the 2008 and for the euro-zone crisis.

5. Statistical analysis: SVM vs. KV implied ratings

In the present section, we are going to present the results of some statistical tests done to measure the goodness of our implied ratings, in comparison with the KV CDS-based approach.

\(^6\)Another popular scheme in dealing with \(k\)-class classification is the one-versus-all, in which only \(k\) SVM boundaries are calculated, and the classes are assigned according to the distance of a point from the \(k\) boundaries, on the basis of the fact that the more you are far from a boundary the more likely your classification is right.
SVM classification plot

![SVM classification plot](image)

**Figure 8.** (part of the) bond-CDS plane on 2012-09-15: different colored regions denote different implied ratings as calculated by the SVM procedure; the colored circles are the data used to train the machine for this day. Colors of circles correspond to agency (S&P’s) ratings according to: black-AAA, red-AA, green-A, blue-BBB, cyan-BB. Some countries are indicated explicitly.

The usual tests to check the goodness of a classification are the ROC (Receiver Operator Characteristic) curve and the Accuracy Ratio, with varying time horizon to test the predictive power as well. Both these tools measure the ability of rating assignments essentially by considering the percentage of defaulters that were assigned low ratings. Unfortunately, these are completely useless in our settings (and, in general, for sovereigns), since the number of defaults of sovereign states, even if we considered longer period of time and larger pool of states, is ridiculous in the perspective of having a statistical test. Thus, we are obliged to abandon the idea of having a test of the *absolute* goodness of our implied ratings, and we check their performance with respect to S&P’s.

5.1. **accuracy.** In order to check the accuracy of SVM implied ratings with respect to S&P’s, and to compare it to that of KV implied ratings, we do the following: we take the implied ratings on each day in the interval considered (2004, January 1st to 2014, January 14th) and we calculate the accuracy of the SVM (and KV) implied ratings versus the agency rating of 1 year later. Then we plot the results of accuracy day by day in figure 13. Recall that accuracy in a classification problem is simply given by

\[
\text{accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{entire population}};
\]

namely, the fraction of right guesses made by the implied ratings.

In figures 14-16 other accuracy graphs are displayed, namely the daily accuracy of prediction for 3-month, 6-month and 1 month horizon. It is pretty clear that in all the situations considered the SVM implied ratings behave better than KV. Just to be clear, we can calculate the cumulate accuracy for the

---

7“Right” with respect to the (time-shifted) agency ratings.
Figure 9. (part of the) bond-CDS plane on 2013-09-12: different colored regions denote different implied ratings as calculated by the SVM procedure; the colored circles are the data used to train the machine for this day. Colors of circles correspond to agency (S&P’s) ratings according to: black-AAA, red-AA, green-A, blue-BBB, cyan-BB. Some countries are indicated explicitly.

<table>
<thead>
<tr>
<th>SVM</th>
<th>AAA’</th>
<th>AA’</th>
<th>A’</th>
<th>BBB’</th>
<th>BB’</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>27562</td>
<td>3634</td>
<td>223</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>AA</td>
<td>3828</td>
<td>9423</td>
<td>1531</td>
<td>496</td>
<td>53</td>
</tr>
<tr>
<td>A</td>
<td>365</td>
<td>2640</td>
<td>8805</td>
<td>1506</td>
<td>175</td>
</tr>
<tr>
<td>BBB</td>
<td>8</td>
<td>184</td>
<td>1871</td>
<td>8411</td>
<td>961</td>
</tr>
<tr>
<td>BB</td>
<td>61</td>
<td>331</td>
<td>47</td>
<td>1252</td>
<td>5283</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KV</th>
<th>AAA’</th>
<th>AA’</th>
<th>A’</th>
<th>BBB’</th>
<th>BB’</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>24604</td>
<td>6051</td>
<td>768</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AA</td>
<td>3598</td>
<td>8270</td>
<td>2883</td>
<td>398</td>
<td>182</td>
</tr>
<tr>
<td>A</td>
<td>136</td>
<td>3260</td>
<td>8009</td>
<td>1297</td>
<td>789</td>
</tr>
<tr>
<td>BBB</td>
<td>0</td>
<td>125</td>
<td>2219</td>
<td>6265</td>
<td>2826</td>
</tr>
<tr>
<td>BB</td>
<td>0</td>
<td>13</td>
<td>808</td>
<td>1346</td>
<td>4807</td>
</tr>
</tbody>
</table>

Table I. Cumulate contingency table for SVM (left) and KV (right) implied ratings for 1-month horizon (primed ratings stand for implied).

<table>
<thead>
<tr>
<th>SVM</th>
<th>AAA’</th>
<th>AA’</th>
<th>A’</th>
<th>BBB’</th>
<th>BB’</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>25141</td>
<td>2490</td>
<td>61</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AA</td>
<td>3434</td>
<td>8550</td>
<td>1280</td>
<td>218</td>
<td>9</td>
</tr>
<tr>
<td>A</td>
<td>365</td>
<td>2641</td>
<td>7729</td>
<td>1151</td>
<td>2</td>
</tr>
<tr>
<td>BBB</td>
<td>8</td>
<td>301</td>
<td>1845</td>
<td>7182</td>
<td>1249</td>
</tr>
<tr>
<td>BB</td>
<td>61</td>
<td>351</td>
<td>266</td>
<td>1787</td>
<td>3654</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KV</th>
<th>AAA’</th>
<th>AA’</th>
<th>A’</th>
<th>BBB’</th>
<th>BB’</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>22634</td>
<td>4586</td>
<td>472</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AA</td>
<td>2941</td>
<td>7625</td>
<td>2698</td>
<td>171</td>
<td>56</td>
</tr>
<tr>
<td>A</td>
<td>136</td>
<td>3359</td>
<td>7473</td>
<td>841</td>
<td>79</td>
</tr>
<tr>
<td>BBB</td>
<td>0</td>
<td>119</td>
<td>1947</td>
<td>5952</td>
<td>2567</td>
</tr>
<tr>
<td>BB</td>
<td>0</td>
<td>14</td>
<td>905</td>
<td>1267</td>
<td>3933</td>
</tr>
</tbody>
</table>

Table II. Cumulate contingency table for SVM (left) and KV (right) implied ratings for 1-year horizon (primed ratings stand for implied).

entire period for the different horizons:

svm 1 month: $\sim 0.756$  svm 3-month: $\sim 0.768$  svm 6-month: $\sim 0.763$  svm 1yr: $\sim 0.757$
Figure 10. SVM implied ratings for Italy, in comparison with KV implied ratings and together with the historical S&P’s.

Figure 11. SVM implied ratings for Spain, in comparison with KV implied ratings and together with the historical S&P’s.
Figure 12. SVM implied ratings for Germany, in comparison with KV implied ratings and together with the historical S&P’s.

Figure 13. Plot of daily accuracy of prediction for 1 year horizon, both for SVM and KV implied ratings (irs).
The best performance is given by SVM implied ratings over a 1 month predicting horizon. You can also see the contingency tables for 1-year and 1-month horizon in tables I and II.

This analysis reflects two facts:

- SVM implied ratings are, on average, fairly better in matching future agency ratings with respect to KV ones.
- SVM implied ratings (for the very same way in which they are concocted) tend to be pretty stable around the agency ratings, resulting in an increasing accuracy the shorter the horizon.

As we shall see in section 5.3, this second point does not actually mean that the shorter term is the better in predicting upgrades and downgrades of agency ratings. Namely, accuracy captures the information of prediction in every situation, also when there is actually no upgrade or downgrade in the future. Some kind of analysis which focuses mainly on predicting the agency ratings movements is necessary for a thorough description of the goodness of implied ratings.

5.2. precision & recall. Another standard test we are going to show, is the precision & recall for each class. For sake of clarity, we just give the cumulative precision and average recall over the 10-year interval for each class, and we limit ourselves to the 1-month and 1-year horizons of prediction. The results are summarized in tables III and IV.

Let us remind what precision and recall are:

\[
\text{precision of class } k = \frac{\# \text{ true positives}}{\# \text{ positives}} = \text{probability of being in class } k \text{ given that the test was positive}
\]

\[
\text{recall of class } k = \frac{\# \text{ true positives}}{\# \text{ members of class } k} = \text{probability of positive test given that the individual is in class } k
\]
Figure 15. Plot of daily accuracy of prediction for a 6-month horizon for SVM and KV implied ratings.

Figure 16. Plot of daily accuracy of prediction for a 1-month horizon for SVM and KV implied ratings.
The higher precision the fewer type I errors, the higher recall the fewer type II errors. Namely, with high class $k$ precision you can trust the test in signals of being in class $k$, with high class $k$ recall you can be quite sure that if a guy is in class $k$ the test is going to get it right.

We can add as a final indicator the popular $F_1$-score, which is simply the geometric mean of precision and recall, namely

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}},$$

and is a standard way of summarizing together precision and recall. The cumulate results for 1-month horizon are in table V, while in figure 17 we plot the daily values of $F_1$ for all the 5 classes. In table VI and figure 18 you can see the same indicator but for a 1-year predicting horizon.

As it is apparent from table III, both precision and recall (and thus $F_1$ as well) are quite high for all the rating classes, and the SVM implied classification performs systematically better, with the (very slight) exception of the precision of AAA.

5.3. gap-conditioned analysis. In this section we want to address the capacity of sovereign implied ratings to predict agency rating adjustments. Namely, as was mentioned at the end of section 5.1, the metrics seen so far do not discriminate predictions of stability from predictions of down/upgrades. While predicting stable ratings is of course important, a focus on rating adjustments is needed to understand the goodness of implied ratings. For the statistical test analysis of the present section, we take inspiration by Moody’s report [9], where the authors do a similar thing for corporates (with their bond-implied ratings).
Figure 17. F1-score for 1-month horizon for the 5 rating classes.

Figure 18. F1-score for 1-year horizon for the 5 rating classes.
In figure 19 you see a graph displaying predictive power of SVM implied ratings. The methodology is the following: we take the 36 countries at the beginning of each month from 2004, January 1st to 6 months before our end date, which is 2014, January 14; we calculate, for each country in each of these days, the rating gap, namely:

\[
\text{rating gap} = \text{iRTG} - \text{RTG}
\]

were we have mapped rating to integers according to

AAA → 1, AA → 2, ..., BB → 5 ;

thus, a positive rating gap means that, on that particular day, the SVM procedure assigns a rating worse than the agency does. Then we go and look to the situation 6 months later and we count how many countries have been downgraded/upgraded/untouched by S&P’s, conditioned to their rating gap (of 6 months before). We aggregate the data for all the countries and the entire time interval and plot them in a bar graph. Each bar corresponds to a rating gap and counts all the countries that, being in that rating gap class, are subject to an (agency) upgrade (green), a downgrade (blue), or are stable (red). Notice that the bars show only fraction values, namely the count is normalized with respect to the number of events for each class.

If the implied ratings are to lead the agency ratings (i.e. to predict them) than we would expect the following: a very big red bar in the 0-gap position; as we move rightward, an increasing blue region (a worse implied rating 'implies' an increasing fraction of future downgrades); as we move leftward, an increasing green region (a better implied ratings 'implies' an increasing fraction of future upgrades). As we can clearly see, while for gap ≥ 0 we can indeed conclude that implied ratings lead the agency ratings, we cannot say the same for negative rating gaps.

The same calculation is done for a 1 year horizon: the SVM performance is here even better, which means that, when focusing on rating adjustments, the short term horizon is not the optimal choice for prediction.

Table VII and plot 20 show the same results, but with absolute numbers. You can see that the vast majority of events is in the region: gap = 0 - change = stable. Actually, notice that the bars corresponding to gap -4, -3, 3 have far too few events to be statistically meaningful: it is better to consider only the bars with gaps form -2 to 2. Moreover, it is apparent that the upgrade events are actually just a tiny fraction of the overall number, both in the 6-month and 1 year scenario, thus resulting in a statistically poor ensemble for making any prevision regarding the upgrades.

<table>
<thead>
<tr>
<th></th>
<th>down</th>
<th>stable</th>
<th>up</th>
<th></th>
<th>down</th>
<th>stable</th>
<th>up</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>-4</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>-3</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>-3</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>-2</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td>382</td>
<td>19</td>
<td>-1</td>
<td>6</td>
<td>331</td>
<td>33</td>
</tr>
<tr>
<td>0</td>
<td>63</td>
<td>2500</td>
<td>22</td>
<td>0</td>
<td>115</td>
<td>2276</td>
<td>38</td>
</tr>
<tr>
<td>1</td>
<td>44</td>
<td>291</td>
<td>0</td>
<td>1</td>
<td>84</td>
<td>227</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>21</td>
<td>0</td>
<td>2</td>
<td>23</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table VII. 6-month and 1 year SVM gap-conditioned contingency tables

In figure 21 it is displayed the same bar graph, obtained with the very same methodology described above, with rating gap calculated with Kou-Varoteo implied ratings, while VIII and 22 are the corresponding tables and absolute counting bar graphs, respectively. A comparison shows clearly that, as far as downgrades are concerned, SVM implied ratings perform slightly better than KV.

6. Conclusions

Credit ratings have a crucial role in many aspects of modern financial markets, even more so after the 2008 mortgage crisis (when CRA's have been strongly criticized for their 'opinions' on mortgage-backed securities) and in particular euro-zone crisis. Given that rating agencies are quite slow in updating
Figure 19. 6-month and 1 year predictive power: on x axis the gap between SVM implied rating and the corresponding historical one (positive gap means implied rating is worse than agency rating); the bars show the fraction of stable/upward/downward rating change after 6 months-1 year.
Figure 20. 6-month and 1 year absolute counting for rating change conditioned to SVM rating gap on the entire time interval.
Figure 21. 6-month and 1 year predictive power: on x axis the gap between KV implied rating and the corresponding historical one (positive gap means implied rating is worse than agency rating); the bars show the fraction of stable/upward/downward rating change after 6 months-1 year.
Figure 22. 6-month and 1 year absolute counting for rating change conditioned to KV rating gap on the entire time interval.
In the following, some detailed information about the dataset used for statistical analysis is presented. We acknowledge UniCredit for providing all the data used in our calculations, in particular we thank Dr. Andrea Basile and the entire Group of Risk Methodologies and Architecture.

The 36 countries considered are Australia, Austria, Belgium, Bulgaria, Canada, Croatia, Czech Republic, Cyprus, Denmark, Finland, France, Germany, Greece, Hong Kong, Hungary, Ireland, Italy, Japan, Malaysia, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russia, Singapore, Slovenia, Slovak Republic, Spain, Sweden, Swiss, Turkey, Venezuela, UK, US.

**Credit Default Swaps.** CDS data are from Markit Corporation. They are updated to 14-01-2014, with quoted maturities in \{0.5, 1, 2, 3, 4, 5, 7, 10, 15, 20, 30\} years, among which we used the most liquid one, namely 5-year. The Debt Tier considered is Foreign Currency Debt (Domestic Debt CDS quotations are much more rare), and currencies for considered debt are dollars for all countries except for US who has euros denominated contracts.

**Bonds.** Bond data are constructed starting from bond prices by Bloomberg Corporation. The spread derivation goes through the construction of two curves: the zero curve, or risk-free curve, one for each currency area (namely, we have a zero curve for each currency by which the bond contracts are denominated), and the zero-coupon curve based on treasuries (of course, one curve for each country considered). The actual data used are the spreads between the treasury (zero-coupon) yield and the risk-free yield (of the corresponding currency). The zero curves are built essentially starting from quotations of basic derivatives on currency exchange rates. On theoretical grounds put forward in [24], the zero-coupon curves for treasuries are bootstrapped via the the standard risk-free procedures, even if treasuries are in fact defaultable. As for CDS, the most liquid maturity (5-year) only has been used in calculation.
Agency ratings. Agency ratings are by Standard & Poor’s. We use the Long Term Foreign Currency rating assessment, remapped into 5 classes as described in section 3. S&P’s is the agency typically more timely in updating/reviewing its rating assessments.

On average along the period analyzed, roughly 39% of the countries considered are rated AAA by S&P’s7, 26% AA, 17% A, 14% BBB and 10% BB.

As a last remark, notice that we have not complete data for all the countries on the entire period considered (01-01-2004 → 14-01-2014). This refers almost completely to the bond dataset, where countries like Swiss, Italy, Singapore, New Zealand, Bulgaria, Portugal, Malaysia, Venezuela and Romania, have yield curves starting after 2004 (typically around 2007). This, of course, does not influence at all the analysis of this paper, it only amounts to have a shorter time period of analysis for some of the countries.

References


7Taking into account the remapping.


Scuola Normale Superiore and C.N.R.S. UMI 3483 Laboratorio Fibonacci, Pisa, Italy
E-mail address: stefano.marmi@sns.it

UniCredit Group, Milano and Università Commerciale Luigi Bocconi – Milan, Italy
E-mail address: aldo.nassigh@unicredit.eu

Scuola Normale Superiore, Pisa, Italy
E-mail address: daniele.regoli@sns.it