|  |
| --- |
| 17 November 2015 | ESMA/2015/1736 |

|  |
| --- |
| Reply form for the Discussion Paper on the  validation and review of Credit Rating Agencies’ methodologies |
|  |

|  |
| --- |
| Date: 17 November 2015 |

Responding to this paper

The European Securities and Markets Authority (ESMA) invites responses to the specific questions listed in the ESMA Discussion Paper on the validation and review of Credit Rating Agencies’ methodologies, published on the ESMA website.

*Instructions*

Please note that, in order to facilitate the analysis of the large number of responses expected, you are requested to use this file to send your response to ESMA so as to allow us to process it properly. Therefore, ESMA will only be able to consider responses which follow the instructions described below:

* use this form and send your responses in Word format (pdf documents will not be considered except for annexes);
* do not remove the tags of type <ESMA\_ QUESTION\_VR\_CRA\_1> - i.e. the response to one question has to be framed by the 2 tags corresponding to the question; and
* if you do not have a response to a question, do not delete it and leave the text “TYPE YOUR TEXT HERE” between the tags.

Responses are most helpful:

* if they respond to the question stated;
* contain a clear rationale, including on any related costs and benefits; and
* describe any alternatives that ESMA should consider

**Naming protocol**

In order to facilitate the handling of stakeholders responses please save your document using the following format:

ESMA\_VR\_CRA\_NAMEOFCOMPANY\_NAMEOFDOCUMENT.

E.g. if the respondent were ESMA, the name of the reply form would be:

ESMA\_VR\_CRA\_ESMA\_REPLYFORM or

ESMA\_VR\_CRA\_ESMA\_ANNEX1

To help you navigate this document more easily, bookmarks are available in “Navigation Pane” for Word 2010 and in “Document Map” for Word 2007.

***Deadline***

Responses must reach us by **19 February 2016.**

All contributions should be submitted online at [www.esma.europa.eu](http://www.esma.europa.eu) under the heading ‘Your input/Consultations’.

***Publication of responses***

All contributions received will be published following the end of the consultation period, unless otherwise requested. **Please clearly indicate by ticking the appropriate checkbox in the website submission form if you do not wish your contribution to be publicly disclosed. A standard confidentiality statement in an email message will not be treated as a request for non-disclosure.** Note also that a confidential response may be requested from us in accordance with ESMA’s rules on access to documents. We may consult you if we receive such a request. Any decision we make is reviewable by ESMA’s Board of Appeal and the European Ombudsman.

***Data protection***

Information on data protection can be found at [www.esma.europa.eu](http://www.esma.europa.eu) under the headings ‘Legal notice’ and ‘Data protection’.

# Introduction

Please make your introductory comments below, if any:

< ESMA\_COMMENT\_VR\_CRA\_1>

The Discussion Paper forces any reader and all interested stakeholders to join into a specific way to analyse and comment the validation and review of CRA's methodologies. It urges respondents to think in the categories of "methodologies with sufficient quantitative evidence" and those with "limited quantitative evidence".

It goes without saying that the border line between those "groups" of methodologies is not clear cut. Since everybody might agree with this challenge of differentiating “sufficient” and “limited” quantitative evidence and that there are no clear rules available to differentiate between "sufficient" and "limited", the more important aspect is missed and overlooked, namely that "sufficient quantitative evidence" is – if ever - the great exception and not at all the rule.

Measures to describe the characteristics of distributions are not the problem, but the conclusions. Assumptions of the statistical models are not met at most CRAs. Since credit ratings are dealing with human beings, it's not like evaluating the quality of weather forecasts. Credit ratings have turned out to be self-fulfilling prophecies in some cases, especially credit ratings of the worldwide leading CRAs. There is a circular logic using default rates to predict default rates. The importance of the interaction of the rating analysts with the issuers is underestimated or not even taken note of when dealing with credit ratings as if they were stochastic variables.

Various scale effects of credit rating services lead to biased results. Other benefits of a rating system than just to "detect" likely defaults are not taken care of. The Alpha and Beta errors can be enormous. Most measures proposed do not take care of the weight of an error (importance of the issuer, number of issues, volume of issues, transparency of the issuer).

The non-deterministic character of a default probability is ignored in models like the CAP curve, since the CAP curve can be misinterpreted not only in the extreme cases that all issuers did or all did not go into default. As can be proven mathematically, different outcomes of the CAP curve for different CRAs could be totally coincidental.

The discussion paper does not explicitly mention other dimensions of credit rating than just Probability of Defaults or Loss Given Default. It draws mainly on research of the last three decades, while the more ordinal than cardinal character of credit ratings dominated most of the history of CRAs over an entire century. Rating symbols of rating scales should be regarded in the context of definitions and as a means of communication. Credit rating is not a specific theory, method or model, but much more a formal system of signs governed by rules of combination to communicate meaning.

The idea of "rating evidence" is not properly taken care of in the discussion paper. The approach of "rating evidence" was developed well before the financial crises, since it became apparent that credit ratings for structured finance were provided on a much lower level of evidence than credit ratings for corporates or banks.

Given the current state of credit rating sciences, there is no way to rate the rating agencies with one single approach. Justice cannot be provided for a whole CRA, but only for individual cases of credit ratings.

Following the constitutional principle of equality the validation approach with respect to CRA's methodologies should not preclude certain business models of CRAs. All presented statistical methods are gaining significance with larger numbers of credit ratings. Therefore, they penalize CRAs that concentrate on a small number of issuers, issues or entities to be rated. The financial markets in the EU are highly fragmented and have many different segments. In many market niches, there are only very few participants. In empirical social research, surveys with fewer than 1,000 or even 10,000 respondents are often referred to as inadequate. In the AAA, AA or A rating categories defaults are detected in the per thousand range. For some market segments, it would therefore never be possible to begin a proof. It would not be possible to make use of special expertise of experienced analysts in a CRA focussing exclusively on such market segments. To conclude with an analogy, analysts are not more successful in detecting rotten fruit when apples and pears are mixed with a lot of other fruit in just one large fruit basket, than if they each specialize on identifying rotten apples in a small apple basket and rotten pears in a small pear basket.

< ESMA\_COMMENT\_VR\_CRA\_1>

1. Do you agree with ESMA’s view regarding the discriminatory power of methodologies?

<ESMA\_QUESTION\_VR\_CRA\_1>

No. Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. More probability density is found as one gets closer to the expected (mean) value in a normal distribution. Since ESMA is of the view that the discriminatory power of a methodology relates to ability to rank order the rated entities in accordance to their future status (defaulted or not defaulted) at some predefined time horizon, all inferences and conclusions can safely extend from descriptive statistics only when there is an extremely large number of observations.

In the rating industry, any number beyond 100 is already "extremely large". There is e.g. for obvious reasons no CRA who could come up with much more than 200 sovereign ratings, since there are simply not so many sovereigns active in the capital markets. To give another example, in most banking systems of the world there are only a few banks working under the same regulatory and economic environment. It cannot be proven that a CRA capable of rating a few banks in one country is equally capable in another country. Vice versa, a CRA with a large number of bank ratings around the globe does not necessarily be in a better position to understand bank risks under the many different regimes than a CRA with a few ratings only because the statistical base is bigger.

It is not possible to raise standards in the industry if all CRAs simply use the same set of statistical measures in demonstrating the discriminatory power of their methodologies.

<ESMA\_QUESTION\_VR\_CRA\_1>

1. Do you agree that the Accuracy Ratio, as derived from the CAP curve, is the minimum statistical measure that a CRA should use as part of its validation processes for demonstrating the discriminatory power of its methodologies?

<ESMA\_QUESTION\_VR\_CRA\_2>

No. The CAP curve deals with credit ratings as if they were prophecies, the prophecy being "this issuer will default" or "that issuer will not default". Most CRAs do not intend to make prophecies, but to provide probabilities of default. The Accuracy Ratio measures only the quality of prophecies.

E.g., if a CRA makes perfect prophecies, putting all issuers into the lowest category which actually default later on, the CRA would be considered as "best". In reality, there is no economic gene code with which to predict which being is "good" and which is "bad". In most cases default is the result of a stochastic process.

It is the rare exception, that a CRA could make a prophecy. Such qualities are available only for a very limited number of models (e.g. OK-Score).

<ESMA\_QUESTION\_VR\_CRA\_2>

1. Do you agree that complementary measures such as the Kolmogorov-Smirnov statistic and the ROC curve (along with a confusion matrix) add further information to the discriminatory power of methodologies? If not, please explain why.

<ESMA\_QUESTION\_VR\_CRA\_3>

Yes, it adds further information, but it should be noted that a Kolmogorov distribution is the distribution of a random variable. Therefore it should carefully be considered whether the Kolmogorov-Smirnov statistic is applied to exactly the same “variables” at different CRAs.

<ESMA\_QUESTION\_VR\_CRA\_3>

1. Are there additional quantitative measures that CRAs should use and which would add further insight into the discriminatory power of methodologies? If yes, please explain the measures and your rationale.

<ESMA\_QUESTION\_VR\_CRA\_4>

Based on its research during the 1980ies and 1990ies, RATING EVIDENCE GmbH was founded in February 2004 to provide an alternative measure which results in a grading of individual credit ratings which are expressed by AAA, AA, A, BBB etc. in a range from 0 % to 100 % evidence.

In the extreme case that credit ratings are knowingly derived from a wheel of fortune, so that there could not be any relationship between any default rate and any credit rating provided this way, the evidence is 0 %. No matter, whether the wheel of fortune had generated a BB or AA+, the evidence would be 0%.

The other extreme is a 100 % evidence, when there is no doubt about the credit rating, e.g. in the case of a rating in the category D for an issuer who is already filing for bankruptcy. 100 % evidence might also be seen under certain conditions for a AAA rating of a central bank on its debt denominated in its own currency, since the central bank would always be able to "print" its own money. Many hyper-inflations are proof of the fact that central banks have not been stopped doing this.

<ESMA\_QUESTION\_VR\_CRA\_4>

1. Are there qualitative measures that are appropriate for demonstrating the discriminatory power of methodologies? If yes, please explain the measures and your rationale.

<ESMA\_QUESTION\_VR\_CRA\_5>

Yes. A lot can be learned from forensic sciences.

<ESMA\_QUESTION\_VR\_CRA\_5>

1. Do you agree with ESMA’s view regarding the predictive power of methodologies?

<ESMA\_QUESTION\_VR\_CRA\_6>

Yes, CRAs whose ratings refer to a creditworthiness measure other than default probabilities should develop and employ relevant statistical tests for comparing the expected behaviour of the ratings to the observed results.

In the first 80 years of their existence, CRAs never published any comprehensive statistic relating defaults to rating categories. Thanks to the advent of computer technologies, only in the late 1980ies they began to make use of methodologies described by ESMA. Most of the methodologies proposed by ESMA treat credit ratings as if they were cardinal, while in reality their many uses in banking, in business and on the capital markets are based on their ordinal characteristic.

<ESMA\_QUESTION\_VR\_CRA\_6>

1. Do you agree that statistical measures of predictive power increase the quality of validation of CRAs methodologies and should be performed by the CRAs?

<ESMA\_QUESTION\_VR\_CRA\_7>

Yes, if the theoretical conditions and requirements for applying the measures are met.

<ESMA\_QUESTION\_VR\_CRA\_7>

1. Do you agree that the binomial and the chi-square tests are the minimum statistical measures that a CRA (when its ratings refer to default probabilities) should use as part of its validation processes for demonstrating the predictive power of its methodologies?

<ESMA\_QUESTION\_VR\_CRA\_8>

No. Even when it is the intention of the CRA to let its ratings refer to default probabilities, the CRA might meet the intentions of the regulator. Even without using chi-square tests rating activities can be conducted in accordance with the principles of integrity, transparency, responsibility and good governance, in order to ensure that resulting credit ratings used in the Community are independent, objective and of adequate quality.

<ESMA\_QUESTION\_VR\_CRA\_8>

1. Do you agree that complementary measures such as the Brier score and the Vasicek one-factor model test add further information to the predictive power of methodologies (when the CRAs’ ratings refer to default probabilities)? If not, please explain why.

<ESMA\_QUESTION\_VR\_CRA\_9>

Yes, if conditions to apply these scores and models are met. E.g. predictions of the CRA must assign probabilities to a set of mutually exclusive discrete outcomes.

<ESMA\_QUESTION\_VR\_CRA\_9>

1. Are there additional measures that CRAs should use and which would add further insight into the predictive power of methodologies when the CRAs’ ratings refer to default probabilities? If yes, please explain the measures and your rationale.

<ESMA\_QUESTION\_VR\_CRA\_10>

Yes. CRAs should be encouraged to make use of all kinds of summary statistics, describing location, spread, shape and dependence, when applicable, e.g. arithmetic mean, median, mode, and interquartile mean, standard deviation, variance, range, interquartile range, absolute deviation, distance standard deviation, coefficient of variation, skewness, kurtosis, L-moments, Pearson product-moment correlation coefficient, Spearman's rank correlation coefficient etc.

<ESMA\_QUESTION\_VR\_CRA\_10>

1. Are there qualitative measures that are appropriate for demonstrating the predictive power of methodologies when the CRAs’ ratings refer to default probabilities? If yes, please explain the measures and your rationale.

<ESMA\_QUESTION\_VR\_CRA\_11>

The validation approach of CRAs can be divided into two parts. It might be obvious that forensic methods can be used for the first part, which refers to the validation processes and governance developed by CRAs, such as the gathering of necessary information from reliable sources and the assessment of data quality (e.g. accuracy, completeness, timeliness and appropriateness of the data used for validation purposes), the documentation of the relevant policies and procedures, etc. It might be less obvious that similar methods are applicable to the second part, which the discussion paper focuses on.

<ESMA\_QUESTION\_VR\_CRA\_11>

1. Do you agree that CRAs using methodologies related to creditworthiness measures other than default probabilities should use statistical measures to demonstrate the predictive power of their methodologies? If yes, please state the potential creditworthiness measures that methodologies could relate to and the corresponding statistical measures as well as any appropriate qualitative measures.

<ESMA\_QUESTION\_VR\_CRA\_12>

Yes, the statistical measure of rating evidence from 0 % to 100 % can be applied in any case. Since the concept of rating evidence is looking at individual ratings and scaling evidence on a cardinal scale, even ordinal credit ratings can be compared in this third dimension of rated evidence by common statistical measures. E.g. different kinds of averages of rating evidence can be calculated, resulting in comparisons like "credit ratings of CRA 1 have an evidence of 56 % on average, while CRA 2's credit ratings have an evidence of 78 % on average".

<ESMA\_QUESTION\_VR\_CRA\_12>

1. If ESMA establishes that there is a need for further guidance to the industry, should this guidance also cover the demonstration of predictive power of methodologies related to creditworthiness measures other than default probabilities?

<ESMA\_QUESTION\_VR\_CRA\_13>

No. When the "Handbuch Rating" was published in 1996, it was the very first collection of articles on credit ratings in German language. It was the first time that CRAs Moody's, S&P's and Fitch Ratings and experts joined in such an endeavour. Scientists in many other European countries failed to look at the subject at all. Most of them became aware of the subject only in the aftermath the global financial crises 2008, so that their perceptions are biased. The very first doctoral thesis in German language on "credit rating by international rating agencies" was published only in 1991. This is just a reminder of the comparatively short history of any science of credit ratings. In addition, it can easily be proven that Western literature fails to make use of the knowledge and expertise of Chinese CRAs, suggesting a different approach to the global credit rating system. Therefore further guidance should only be provided after a leap in the scientific knowledge of the subject.

<ESMA\_QUESTION\_VR\_CRA\_13>

1. Do you agree with ESMA’s view regarding the historical robustness of methodologies?

<ESMA\_QUESTION\_VR\_CRA\_14>

The scope of this question is too broad to be answered with a simple "yes" or "no". In any case, yes, historical robustness of a methodology can be demonstrated by assessing other dimensions that do not relate to its discriminatory or predictive power.

<ESMA\_QUESTION\_VR\_CRA\_14>

1. Do you agree that stability statistical measures and the transition (migration) matrices are the minimum measures that a CRA should use as part of its validation processes for demonstrating the historical robustness of its methodologies?

<ESMA\_QUESTION\_VR\_CRA\_15>

No. If a minimum measure is required, this measure would be endorsed to be reliable enough to deliver meaningful results. Currently there is no such measure which could count on a scientific consensus. <ESMA\_QUESTION\_VR\_CRA\_15>

1. Do you agree that complementary measures such as distribution analysis, the univariate analysis of rating determinants and benchmarking add further information to the historical robustness of methodologies? If not, please explain why.

<ESMA\_QUESTION\_VR\_CRA\_16>

Yes, they add further information, if premises and conditions are met.

<ESMA\_QUESTION\_VR\_CRA\_16>

1. Are there additional measures (qualitative or quantitative) that CRAs should use and which would add further insight into the historical robustness of methodologies? If yes, please explain the measures and your rationale.

<ESMA\_QUESTION\_VR\_CRA\_17>

Yes. The above mentioned concept of "rating evidence" generates additional information if historical values (i.e. 0 % to 100 % for each rating) are analyzed.

<ESMA\_QUESTION\_VR\_CRA\_17>

1. Do you agree with ESMA’s view regarding the validation of methodologies with limited quantitative evidence?

<ESMA\_QUESTION\_VR\_CRA\_18>

Again, this question is too broad in its scope to be answered by a simple "yes" or "no". In fact, the question might generate biased results on ESMAs discussion paper, since stakeholders are forced to express their views dichotomously.

<ESMA\_QUESTION\_VR\_CRA\_18>

1. Do you agree that CRAs should, as a first step, investigate data enhancement in validating methodologies with limited quantitative evidence?

<ESMA\_QUESTION\_VR\_CRA\_19>

The meaningfulness of data enhancement solutions are strongly overrated. There is very limited meaning in expanding data samples with the use of third party data. The way a CRA generates data is part of its unique franchise. Combining (if meaningful) asset classes or sub-asset classes with similar risk characteristics in order to perform joint validation assessments does not make sense in certain cases. A textbook example is the hypothetical credit rating of the Pope or the Vatican City State. There is only one Pope in the world with very unique features which are not easily comparable to religious leaders in other religious states. Even worse is the creation of hypothetical transactions to expand the available data. <ESMA\_QUESTION\_VR\_CRA\_19>

1. Do you agree that CRAs should, as a second step, investigate measures that may enable them to perform statistical tests to demonstrate the discriminatory power of their methodologies?

<ESMA\_QUESTION\_VR\_CRA\_20>

Yes, ESMA should help to investigate measures and enable CRAs.

<ESMA\_QUESTION\_VR\_CRA\_20>

1. Do you agree that historical robustness measures should be performed when validating methodologies with limited quantitative evidence?

<ESMA\_QUESTION\_VR\_CRA\_21>

No, since there is no single set of historical robustness measures available which is applicable for the validation of all kinds of methodologies.

<ESMA\_QUESTION\_VR\_CRA\_21>

1. Do you agree that the transition (migration) matrices and benchmarking are the minimum measures that a CRA should use as part of its validation processes for methodologies with limited quantitative evidence?

<ESMA\_QUESTION\_VR\_CRA\_22>

No, since any minimum measure would make stakeholders believe that this measure is a reliable source to evaluate whatever quality of a CRA. On the contrary, investors, issuers and the public in general should always be reminded that opinions on the future of an issuer are just opinions and that there still has not been developed a generally accepted science of credit ratings which would make the predictive power of credit ratings as reliable as the landing of man on the moon.

<ESMA\_QUESTION\_VR\_CRA\_22>

1. Do you agree that complementary historical robustness measures add further information to the validation processes for methodologies with limited quantitative evidence? If not, please explain why.

<ESMA\_QUESTION\_VR\_CRA\_23>

Yes, they add further information, but only if conditions for their application are met.

<ESMA\_QUESTION\_VR\_CRA\_23>

1. Are there additional measures that CRAs should use when validating methodologies with limited quantitative evidence? If yes, please explain the measures and your rationale.

<ESMA\_QUESTION\_VR\_CRA\_24>

Yes. The key to validating methodologies with limited quantitative evidence is to look at the evidence of individual ratings, not only the evidence for a group of ratings or the whole universe of ratings published.

E.g., the OK-Score Model has been warning for business failure and fraud at Abengoa since 2011. Abengoa’s has officially announced its default on the 26th of November 2015. The OK-Score Model had already started warning for business failure on the 7th of February 2011. The non-detection error of the OK-Score Model is reportedly 2%, and false warning error 2%.

Credit ratings for FlowTex, WorldCom, Enron, Parmalat and some other issuers are examples in which the error of the CRAs involved could have been detected.

<ESMA\_QUESTION\_VR\_CRA\_24>

1. Do you agree that thresholds should be set for the quantitative validation techniques?

<ESMA\_QUESTION\_VR\_CRA\_25>

No, since there should be no obligation to make use of quantitative validation techniques as long as there is no agreement among scientists and other experts about the meaningfulness of the validation techniques.

<ESMA\_QUESTION\_VR\_CRA\_25>

1. Do you agree that the Internal Review Function should decide on these values?

<ESMA\_QUESTION\_VR\_CRA\_26>

No. There are human beings working at the Internal Review Function who are not less fallible than human beings outside this organisational unit.

<ESMA\_QUESTION\_VR\_CRA\_26>

1. **Do you agree that predefined actions should be defined by the CRAs when the thresholds are met?**

<ESMA\_QUESTION\_VR\_CRA\_27>

No, for the reasons provided above.

<ESMA\_QUESTION\_VR\_CRA\_27>