|  |
| --- |
| 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>

DBRS appreciates the opportunity to comment on ESMA’s discussion paper on the validation and review of Credit Rating Agencies methodologies[[1]](#footnote-2) (DP). DBRS understands views from stakeholders will help ESMA provide guidance to credit rating agencies (CRAs) with respect to the identification and consistent application of validation and review measures for demonstrating the discriminatory power, predictive power and historical robustness of methodologies, particularly where there is limited quantitative evidence.

The DP discusses a number of different measures. In DBRS’s view, the various measures discussed in the DP are different versions that test the same thing or convey essentially the same information. Most of these measures have a strong dependence on the default probability structure of the underlying credit portfolio. There are also certain risks in elevating any particular measure to the position of a de facto standard measure. In addition, these tests are less meaningful where there is limited available ratings data including defaults. As such, DBRS thinks that one measure should not be mandated over another. Rather, DBRS believes each CRA should be able to choose a measure that makes sense to them within a standardized framework of measures.

The absence of significant quantitative evidence also impacts what ESMA otherwise considers to be back-testing. DBRS finds it difficult to construe the term ‘back-testing’, as used in Article 8, as consisting of measuring discriminatory power, predictive power and historical robustness of ratings issued using a methodology because this would mean Article 8 provides an exemption from the requirement to carry out each of these items (where quantitative evidence is limited) only to impose the same obligations under the heading of back-testing. Rather, where there is limited available ratings data to support the predictive power, there is likely to be limited evidence of discriminatory power.

An approach where CRAs are permitted some choice in the selection of measures and in their approach to back-testing would still meet ESMA’s intended objective of consistent validation of the discriminatory and predictive power and historical robustness of methodologies by CRAs and between CRAs, especially for those firms that have limited available data.

< ESMA\_COMMENT\_VR\_CRA\_1>

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

<ESMA\_QUESTION\_VR\_CRA\_1>

DBRS understands discriminatory power to be the degree to which credits with different ratings produced using a methodology exhibit different default rates.

<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>

DBRS thinks the Accuracy Ratio, as derived from the CAP curve, is one of several measures which assess a methodology’s discriminatory power. In our experience, instances of low levels of discrimination are flagged by the different measures of discriminatory power such that one measure is not superior or preferable to another. Further, measures such as the Accuracy Ratio and others, have a strong dependence on the default probability structure of the underlying portfolio. There are certain risks in elevating a particular measure, such as the Accuracy Ratio, to the position of a de facto standard measure, including undue comparisons of the magnitude of the measure and possibly the calibration of the composition of underlying rated portfolios in order to maximize the specific measure. As a result, DBRS does not see any reason to mandate the use of the Accuracy Ratio over another measure.

<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>

DBRS does not believe these measures add material incremental information regarding the discriminatory power of methodologies. As discussed in our response to question 2, we have observed that these measures convey essentially the same information as the Accuracy Ratio. We also note that the area under the ROC curve contains the same information as the Accuracy Ratio. In our view, the process of assessing the discriminatory power of a methodology includes the selection of the measure(s) which are most appropriate for the particular portfolio of rated entities and methodology.

<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>

The correct ordering of default rates, i.e. higher ratings exhibiting lower default rates than lower ratings, is a consideration which reflects most closely DBRS’s intention with respect to its ratings. Per response to question 3, the measure selection process may point to applicability (or not) of a certain measure, depending on the particular methodology and data being evaluated. As an example, the Kolmogorov-Smirnov statistic may be inapplicable in the case where a methodology model was derived from the ratings data to be used for measuring the methodology’s discriminatory power.

<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>

DBRS has no suggestions to offer.

<ESMA\_QUESTION\_VR\_CRA\_5>

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

<ESMA\_QUESTION\_VR\_CRA\_6>

DBRS agrees that the predictive power of a methodology must presumably be interpreted as the degree to which the realized default rate relates to an a priori expectation of the default probability. While DBRS uses benchmark default rates for validation purposes, these are not usually targets of our methodologies. Instead, we use them as a device in considering whether credits in different sectors appear to be correctly rank-ordered.

<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>

Our experience with statistical measures regarding the performance of ratings (including predictive power) is that they highlight known problem areas. In other words, in the past, statistical measures have confirmed known poor performance, rather than detected it.

<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>

DBRS agrees that the binomial test is the most obvious starting point for establishing whether anticipated default probabilities and actual default rates are consistent, and the Hosmer-Lemeshow test seems like a logical extension. We note that the binomial test relies on an assumption of independent behaviour of individual credit ratings, which is at odds with our experience. While this does not preclude measuring the probability of the hypothesis that the true default probability of credits is the a priori probability under the assumption of such independence, it means the results cannot be taken at face value.

<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>

The Brier score appears closely related to the Hosmer-Lemeshow test, and so appears unlikely to add material information regarding the predictive power of a methodology. The Vasicek test addresses the point we made in our response to question 8 regarding credits not being independent, which is helpful. Unfortunately, it also requires a (Gaussian) correlation parameter. As this correlation is not known, one has to make an assumption, so that the validity of the hypothesis testing using the Vasicek one-factor model test is dependent on the validity of that correlation assumption. Again, the results of such a test cannot be taken at face value.

<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>

DBRS has no suggestions to offer.

<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>

DBRS has no suggestions to offer.

<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>

DBRS is of the opinion that all CRAs should be subject to similar requirements, regardless of the measure of creditworthiness to which they make reference, subject to data availability. However, similar requirements does not preclude being able to choose a statistical measure within a standardized framework.

<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>

TYPE YOUR TEXT HERE

<ESMA\_QUESTION\_VR\_CRA\_13>

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

<ESMA\_QUESTION\_VR\_CRA\_14>

DBRS’s interpretation of Article 7(2)(a) is that the requirement to describe the “historical robustness” of credit ratings is an inextricable component of the validation of the methodologies’ predictive and discriminatory power. Specifically, given that most commonly accepted validation techniques for measuring predictive and discriminatory power utilize historical observations of quantities such as default frequency and timing, loss at default etc., establishing measures of historical robustness is a necessary condition in order to properly measure predictive and discriminatory power. Hence, at least certain measures of historical robustness will be incorporated in the context of assessing predictive and discriminatory power.

At the same time, any measure of historical robustness, including ratings migration, ratings distribution, as well as rated entities migration (the changes over time in the attributes describing the universe of rated entities), and others, may carry weight in the process of back testing and benchmarking the methodologies’ predictive and discriminatory power. Excluding any such measure a priori may degrade the evaluation of applicability of the back testing technique, for example. Hence, no measure of historical robustness should be excluded a priori from consideration when attempting to measure predictive and discriminatory power.

Combining the above considerations, DBRS concludes that the intent of Article 7 with respect to the historical robustness of credit ratings, is to underline the importance of measuring their predictive and discriminatory power in the context of historical (and observed) outcomes. Therefore, we view measures of historical robustness as tools assisting in the evaluation of the predictive and discriminatory power of rating methodologies, rather than as measures comprising a separate dimension of evaluating rating methodologies’ performance.

<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>

Stability measures: We understand stability measures to be a commonly used tool in the context of credit scoring models. The vast majority of our methodologies are in effect the transposition of our analysts’ expertise rather than scoring algorithms trained on a specific dataset. It is therefore not obvious to which dataset one would compare the current population of credits covered by a methodology (i.e. we do not have an ‘original’ population). It is also not immediately obvious how stability of characteristics of credits or of the distribution of ratings within a sector over time would assist in validating a methodology. For instance, since the 2008 financial crisis, banks have changed with higher regulatory capital requirements and changed business mixes; this change in the characteristics of the credits does not in itself help to confirm or deny the validity of a banking methodology or of its performance.

Migration matrices: we find migration matrices useful pieces of information. If all negative migration rates were to be excessive, this would result in an excessive default rate. Also, unusual large rating changes can be worth examining to see whether they are due to an event that was not considered within the methodology and should be introduced. We note that these considerations are qualitative and lead to further examination rather than being conclusive evidence.

<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>

As noted above, we are uncertain how the distribution of ratings informs the validity of a methodology or its robustness beyond reflecting rating migrations otherwise visible in a migration matrix.

We believe univariate analysis, where relevant, is a consideration that is likely to confirm the rigour with which input variables were chosen. It does not, in our understanding, say a lot about actual realised performance of the ratings since it relates individual inputs to actual realisations for credits and ratings do not feature. We note that paragraph 2(a) of Article 7 refers to the “historical robustness (…) of credit ratings issued using the relevant methodology”.

Benchmarking to other CRA ratings will establish whether two CRAs tend to agree or disagree in their current credit assessments. It isn’t clear to us how this would shed light on the historical robustness of a methodology, i.e. the robustness of the methodology in the past.

<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>

DBRS has no suggestions.

<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>

We find it difficult to construe the term ‘back-testing’, as used in Article 8, as consisting of measuring discriminatory power, predictive power and historical robustness of a methodology because this would mean Article 8 provides an exemption from the requirement to carry out each of these items (where quantitative evidence is limited) only to impose the same obligations under the heading of back-testing.

Furthermore, where there is “limited quantitative evidence to support predictive power” because of a small sample of ratings and/or a small number of defaults, there is likely to be limited evidence of discriminatory power as well.

We interpret ‘back-testing’ as the application of the methodology to credits or obligations for which historical data is available and the outcome (default or no default) is also known.

We note also that Article 8 (c) requires a CRA to have “processes in place to ensure systemic credit rating anomalies highlighted by back-testing are identified and are appropriately addressed.”

<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>

Our rating process involves the application of human judgement and a committee process so it is not possible to replicate it fully in an algorithm. As such, the true test of a rating methodology as applied is the performance of the actual ratings. Our understanding is that the reference to predictive power in Article 7 is a reference to the predictive power of historical ratings assigned by the CRA (paragraph 2(a) of Article 7 makes reference to the “predictive power of credit ratings issued using the relevant methodology”). This is different to the predictive power of the methodology when applied without analytical input to historical information on credits that the rating agency did not in fact rate. Expanding data sets and creating hypothetical transactions will not increase the number of actual ratings so will not enhance the evidence of the predictive power of credit ratings issued by the CRA.

<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>

As noted in response to question 18, we do not interpret Article 8 as requiring evidence of the discriminatory power of actual ratings. We do, though, agree in general with ESMA’s views in paragraph 58.

We agree with ESMA’s view that a CRA should consider combining rating categories. We have typically studied whole rating categories (i.e. AA, A, BBB etc.) rather than individual notched ratings (A (high), A, A (low) etc.) for this reason.

We agree that using an extended time period is a potential solution. This has its limits though because methodologies change over time, so that the performance of ratings assigned several years ago will have been the result of a different methodology than the one being validated.

We see the use of a ‘relaxed’ default definition as more akin to studying the migration rates of ratings. While these can be informative, they do not necessarily permit the study of the ratings’ discriminatory power.

<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>

As noted in responses above, we see merit in considering migration matrices when validating a methodology. We agree that these are relevant when there is limited evidence of predictive power.

Back-testing, the study of transition matrices, and others, are useful techniques even in the case of sparse historical data sets, as they can highlight potential anomalies, even without attaching statistical significance, for which richer data is typically necessary. In our view, this is the intent of the reference to back-testing in Article 8(c). Historical robustness measures, which presumably measure the robustness of a methodology in the face of substantial historical evidence (the absence of which would challenge the determination of robustness), will likely indicate lack of ability to determine the methodology’s robustness in the case of limited quantitative evidence.

<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>

DBRS agrees that migration matrices are useful in validating a methodology.

DBRS believes benchmarking exercises can be informative. We would note that in our experience of such exercises, when benchmarking ratings in a given sector, CRAs’ differing opinions regarding sovereign risk or banking system risk often dominate the ratings assigned to entities or obligations within that sovereign or banking system, obscuring the relative effects of the primary methodology for the sector.

<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>

As per answers to questions 15 and 21, DBRS finds transition matrices to be useful information in validating methodologies.

<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>

DBRS has no suggestions.

<ESMA\_QUESTION\_VR\_CRA\_24>

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

<ESMA\_QUESTION\_VR\_CRA\_25>

DBRS agrees that quantitative testing is most useful when there is some understanding of what constitutes reasonable or expected results versus anomalous results.

<ESMA\_QUESTION\_VR\_CRA\_25>

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

<ESMA\_QUESTION\_VR\_CRA\_26>

TYPE YOUR TEXT HERE

<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>

DBRS agrees that quantitative testing is most useful when there is some understanding of what constitutes reasonable or expected results versus anomalous results.

DBRS would caution though that all of the measures contemplated in this consultation need to be considered with care: we understand that Accuracy Ratio measurements are not comparable when carried out on different datasets; binomial tests of predictive power make the unrealistic assumption that credits in a sector behave independently, and the Vasicek test, which allows the introduction of correlation, requires an assumption regarding the correlation which will likely be difficult to substantiate.

It therefore seems difficult to define ahead of time what needs to be done if a measurement appears anomalous beyond requiring a reviewer to investigate the reasons why ratings have behaved in an unforeseen manner to determine whether the cause is a weakness in the methodology or in its implementation or whether the performance is down to an unlikely but unforeseeable set of events.

<ESMA\_QUESTION\_VR\_CRA\_27>

1. Methodology includes models and key rating assumptions. [↑](#footnote-ref-2)