**AFME Consultation Response: EBA consultation regarding the supervisory handbook for the validation of internal ratings-based systems**

October 2022

*The validation handbook makes multiple references to the different regulatory products of the so called ‘EBA repair program’. Institutions are welcomed to provide feedback on whether there are specific aspects or problems that have arisen in practice on any of the topics described in the handbook in connection with the EBA repair program.*

**General Comments:**

AFME welcomes the opportunity to respond to the EBA consultation on the supervisory handbook for the validation of internal ratings-based systems. We would note the following comments which do not directly fall into the questions posed by the consultation:

* With regard to paragraph 11 (page 8)[[1]](#footnote-2)we assume that the adequate tests performed within RoE (and Model Monitoring) are not expected to be repeated for the subsequent validation. For example on p.36, Article 42: the requirement to ‘Assess the materiality of default rating assignments in application sample [number/RWA]’ would not be performed in validation in case already part of RoE/monitoring. It would be useful if the final handbook clarified this.
* Communication of the findings and recommendations (Paragraph 19)[[2]](#footnote-3):While the requirement that Internal Validation communicates its outcomes through the issuing of findings and recommendations is well established in current regulations and practice, we would welcome clarity on how the expectation of issuing recommendations differs from advising the CRCU (Credit risk control unit). We would ask EBA to clarify this point, which we interpret as if the Internal Validation should not have an advisory role towards the CRCU when the letter one is addressing the recommendation issued by Internal Validation. We believe that it is a good practice that Internal Validation is aware of the implication of issues detected to the point of being able to propose solutions for improvement as non-binding suggestions. Furthermore, in the recent IMI, banks have been incentivised by the ECB to improve their validation handbooks introducing, in case of detected deficiencies, the expected remedial actions to be taken by CRCU.
* We would appreciate further clarity regarding the intra-group outsourcing (paragraphs 26 & 150) when it comes to Rating Systems approved to be used by local branches. We understand that the EBA expects the validation of such rating systems to be carried out by the local unit located in the branch (even if outside EU). Validation by a local unit is not considered as outsourcing even when the outcome of the Rating System is used to calculate the RWA at consolidated level, while some validation work performed by the holding (in EU) it is considered as outsourcing.
* In this same vein, we would support greater materiality and proportionality to be introduced for the requirements related to introducing a local validation function in case of banking groups where the models are developed and validated in a centralised manner on the whole group portfolio, especially in cases when subsidiaries share small sub-portfolios (i.e. not significant for the group), which do not enable validation activities locally. In particular, we ask if in case of a centralised approach, the final responsibility could be assigned to the management/governing bodies of the local bank[[3]](#footnote-4).
* Context Box 1: In order to ensure that one material extension or change is not split into several changes or extensions of lower materiality the handbook proposes as good practice “*to compare the own funds requirements resulting from the updated risk parameter estimates with the ones resulting from the last version of the model authorised by the CA (i.e. the version of the model without taking into account non-material model changes)*”. In our understanding, this statement may have many severe implications since:
* It requires the institutions to be able to keep an old version of the model (no matter how many non-material changes have been addressed, i.e. versions of the model from the last approved material change) in the systems that may correspond to several years ago. This may be very burdensome in terms of IT infrastructure and maintenance.
* This implies that, in accordance with the expectation to recalibrate annually the data windows of the models during several years, addressing the corresponding post-notifications and without performing any other model change, a material change may arise from the application of the proposed analysis/criteria. The same issue would happen in case of addressing other non-material changes as pre-notifications. In our opinion, there might be pre-notification changes which need to be performed on a specific year (e.g. due to business needs) and they are not related at all with other pre-notifications performed in different or some years after. Therefore, we believe that this may make it impossible to apply for any non-material change in future.
* Information Box 7 appears to require 3 years of shadow rating processes (i.e. Situation 2) for models not yet live due to outstanding regulatory approval. We see this as overly burdensome.
* Furthermore, Information Box 7 distinguishes between a regular review (conducted at least yearly) and a full review of estimates (conducted with a frequency depending on the materiality of the model, in some cases, every three years regardless of the materiality of the model). However, **Interaction Box 12** suggests the following:

1. *[Type 1 interaction: Leverage on CRCU analyses with independent conclusions]*
2. *[****Type 2*** *interaction: CRCU analyses* ***complemented by own analyses****]*

[…]

*When it comes to the validation activities and their interaction with* ***the analyses performed in a recent full review of estimates conducted by the CRCU****, the validation function is expected to perform two activities:*

[…]

*2. Perform an* ***assessment of the performance*** *of the rating system* ***solely based on a Type 2*** *interaction for all the analysis as described in section [4* First validation of newly developed rating systems or changed aspects of a rating system*].*

[…]

Therefore, for a regular (annual) review of estimates nothing is prescribed on the handbook with respect what type of review (Type 1 or Type 2) that should be applied for assessing the performance of the models. Since for a full review of estimates a Type 2 review is suggested, it seems that for a regular review a Type 2 would not be required regarding the assessment of model performance. However, according to the “ECB Instructions for reporting the validation results of internal models”, backtesting exercises are requested to be submitted to the ECB annually after performing independently such assessment by the Internal Validation function. For this reason we believe that, although the EBA and ECB indications are not contradictory given that for a regular review of estimates there is no prescription in the Handbook regarding the type of validation that should be considered, there seems to be a misalignment since including only prescriptions for a full review of estimates can implicitly suggest that the regular review of estimates may not require a Type 2 validation.

* **Paragraph 35** – Type of analyses to assess the accuracy and consistency of rating systems: With regards to different frequency and depth of the analyses commensurate with the complexity and materiality of the rating systems under the scope of validation, we think it’s appropriate for the Handbook to consider the opportunity to prescribe that every financial institution should implement an internal methodology (e.g. model risk tiering methodology) both to rank and discriminate complexity and materiality of internal models with a clear and objective approach. The implementation of a risk inherent attribute will facilitate and allow commensurate analyses with respect to the rating systems under validation.
* **Paragraph 88** states that “*it is expected that the validation function reviews the documentation that will be submitted to CAs*”. It is our understanding that the internal validation function should review only the documentation for which they should perform a challenge (methodological and technical documentation, regulatory compliance of methodological choices, rating assignment process, capital impacts, etc) or the documentation for which they are owners (internal validation reports), but not all the documentation submitted to the CAs (e.g. Internal Audit reports, methodological standards, etc).
* In respect of Article 107 & 133 we note translating business and functional requirements into IT specifications is a very technical task and, in our opinion, too far away from validation work, we see this as a pure IT and Business topic. Furthermore, if for instance a pure IT change was reviewed by e.g. internal audit an additional requirement to validation to check proper reflection in business/functional requirements seem unduly burdensome with limited value add (and perhaps better placed completely with internal audit as review function). The IT infrastructure and model development and implementation within that infrastructure in banks is subject to detailed governance and security requirements. From our point of view, the guidelines would not need to include additional references for this area as it is seen redundant.
* We would ask the EBA to confirm that the separation between the development sample and the OOT/OOS sample is supposed to reflect the risk differentiation analysis of the model but not for the risk quantification of parameter estimate where all available data / defaults are to be used. We would welcome clarification on the approach to be used for the risk quantification of parameters estimates and validation. (References: Interaction Box 5[[4]](#footnote-5) and Interaction Box 1/ CRR Article 179 (1) c) states that “*An institution's estimates shall reflect the implications of technical advances and new data and other information, as it becomes available. Institutions shall review their estimates when new information comes to light but at least on an annual basis*” and CRR Article 181 (a) states “*institutions shall estimate LGDs by facility grade or pool on the basis of the average realised LGDs by facility grade or pool using all observed defaults within the data sources*”.
* We would also welcome clarification from the EBA on the amount of time/sample that is expected to be left apart for the out-of-time (OOT) and out-of-sample (OOS) analyses. For the OOT analysis, while a 1-year period would be practicable at least for PD (but probably still not sufficient for LGD) both from a validation and a CRCU perspective, longer time period for OOT samples would imply developing the model on older time horizon, with potential shortcomings on representativeness and also the necessity of an update of the time series (through ex-ante or authorization) suddenly after the authorization of the model change itself. For the OOS analysis we ask the EBA to clarify if a statistical approach (such as stratified random sampling approach or bootstrapping techniques) could be considered as appropriate. (References: Art. 50 and 51 state requirements to *challenge the methodological choices used to derive the PD and LGD best estimates)*.
* Could the EBA confirm its expectation regarding “best estimate” parameters is to establish a link between the PD and LGD “best estimate” in IRB modelling with the PD and LGD models for IFRS9. If this link is not an EBA expectation, we would ask for a different terminology to be used for IRB PD and LGD models before prudential correction (e.g. “LRA PD estimates” and “LRA LGD estimates”) in order to avoid any potential ambiguity. We note Internal Validation aims to assure that the risk parameters are adequate and consistent with the RWEA calculation, thus we deem more appropriate to ground the final validation assessment on the quantitative tests performed on the final risk parameters used for the RWEA calculation, in alignment with paragraph 55 and also with the ECB Supervisory Validation Reporting. Furthermore, developing both types of analyses (i.e. on “best estimate” parameters and on “LRA estimate” parameters) would add complexity when dealing with non-coherent outcomes between the two analyses – for example in terms of recommendations and remedial actions.
* A clarification of the interrelation between the Handbook and both the ECB guide to internal models and the ECB instructions for reporting the validation results of internal models would also be appreciated. In particular, although broadly speaking they are all very much aligned, the validation task required in paragraph 65(h)(vi)(a) of the ECB guide (analysis of model design stability) does not seem to be included in the Handbook. Does this mean that the previously mentioned analysis is no longer required? On the other hand, in relation to the ECB instructions for reporting validation results, these only allow for the possibility of considering model changes that happened within the observation period when these are material, otherwise (i.e. if no changes or only non-material changes have occurred within that period) the parameters to be used in the analysis must be those that were in force at the time of the initial reference date. In contrast, in the Handbook it seems that performance tests must be carried out also when non-material changes are presented.

**Question 1:**

**1a) How is the split between the first and the subsequent validation implemented in your institution?**

|  |
| --- |
| **Members have indicated the following approaches to validation:**   1. A clearly defined two-step approach, with an **initial validation**, which applies for new models and models on which changes have been carried out. A complete validation is performed over all dimensions, or only over those dimensions or aspects subject to any type of modification. This is followed with a **recurrent validation which is** performed over all those previously validated models. For IRB models, this validation occurs annually, and all dimensions are reviewed in accordance with the regulation in force. 2. In the second approach there is no explicit definition for “first” and “subsequent” validation, while the internal framework clearly defines what an “initial” vs “ongoing” validation is. The main difference is the prescription for qualitative tests in the initial validation (e.g., on variable selection or model design), while quantitative tests are performed both in initial and in ongoing validation. In case of material model change or extension to a rating system, the framework foresees an initial validation. In case of non-material model change or extension to a rating system, when no report by the independent review or validation is expected to be part of the application package, the framework in any case foresees in the relevant ongoing validation that qualitative tests must be performed in case of changes in the steps under assessment or in the relevant requirements. The understanding is that this framework is already compliant to footnote 36. 3. A framework that provides a differentiation between “full” and “standard” validation, with the latter providing a simplified approach in terms of depth of analysis excluding model design checks. In this regard, starting from 2021 the mentioned Internal Validation framework has been updated implementing the Model Risk Tier as a relevant driver to apply different frequency and depth of the analysis “commensurate with the complexity and materiality of the rating systems under the scope of validation”. The methodology assigns Tier 1 and Tier 2 to the most relevant models of the Group, on the other hand Tier 3 and Tier 4 are assigned to less significant models. The full validation is applied at first adoption, in case of material model changes and extensions and in case of non-material model changes and extensions subject to ex-ante notification for Tier 1 models identified as such within the Model Risk Management framework and in particular in the Model Tiering Methodology implemented by the Model Risk Management sub-department since 2019 to classify and rank models allowing the tiers-based streamlining of the main model related processes. Ex-ante notifications for Tier 2 models are instead supported by a standard validation. Subsequent validations are subject to the standard validation on a regular (annual) basis, but subject to the full validation on a periodical basis as provided within the ECB Guide to Internal Models. In case of model changes, either material or non-material subject to ex-ante notifications, model design checks are limited to those components directly or indirectly effected by the change.   Furthermore, in respect of the Handbook specifically, it should be noted regarding the first validation of non-material changes (those not requiring approval of the CA), since paragraphs 9 and 26 do not distinguish between material (requiring approval) and non-material (requiring notification, either ex-ante or ex-post) changes, it may be understood that all changes need to be validated prior to the change notification to the CA, while the CDR on model changes only requires this for changes requiring approval. Paragraph 105 seems to align with this approach, but it may be useful to include some clarification on both paragraphs 9 and 26 to avoid any doubt. |

**1b) Do you see any constraints in implementing the proposed expectations (i) as described in section 4 for the first validation for a) newly developed models; and b) model changes; and (ii) as described in section 5 for the subsequent validation of unchanged models?**

|  |
| --- |
| Both in an initial and a recurrent validation, some constraints could be found in the validation or challenge of models where the human judgment or expert criteria play an important role, since they make it difficult to apply and define homogeneous tests and thresholds (e.g. when analysing the consistent application of the human judgement in the rating assignment).  Additionally, there may be constraints on the validation or challenge of data that are not typically found in standardized databases (e.g. cash flow projections used for specialized lending, intermediate components of the ranking method, qualitative information, etc.) or for which access to databases of portfolio/business management teams are required (e.g. some collateral or guarantee information).  On the other hand, regarding out-of-time (OOT) and out-of-sample (OOS) testing, usually the most recent available information is taken into account in the development of the models and the calculation of the proposed values (especially for credit risk parameters). Therefore, no or few data will be available for the OOT or OOS-testing in these cases.  Additionally, with regard to the analysis performed on external data sources, it should be taken into account that in the case of comparisons against benchmarks provided on the EBA benchmarking portfolios, they may be conditioned to structural differences between portfolios, since the entity is not able to ensure the comparability of the portfolios with the public information available.  As regards the first validation specifically, we see constraints in fulfilling Article 94(a) related to overrides, as the Rating Desk would be required to simulate the usage of the new model before its implementation and while the previous is still in production, to understand whether overrides are needed. This implies an unplanned effort from Rating Desk and, according to the size of the portfolio, the task could need several months, hindering the compliance on the recentness of data. Furthermore, Article 94(b) requires a backward simulation of ratings for testing stability; while to compare the ratings at performance and use test samples is quite straightforward, adding further years would be cumbersome.  Finally, as regards the subsequent validation, we see that the prescription of Articles 113 and 115 should be part of the review of estimates performed by CRCU, and that the internal validation should leverage on it, as per Type 1 interaction in Interaction Box 12. |

**Question 2: For rating systems that are used and validated across different entities, do you have a particular process in place to share the findings of all relevant validation functions? Do you apply a singular set of remedial action across all the entities or are there cases where remedial actions are tailor-made to each level of application?**

|  |
| --- |
| For one member regulatory models with a consolidated use are validated in a centralised manner, in accordance with the model validation policy. The conclusions and findings resulting from the validation process are transparent and available to all validation teams in the Group.  Similarly for another member all the regulatory models with a consolidated use covering different entities are consistently applied across the group and centrally validated. As for model design aspects, conclusions and findings resulting from the validation process as well as remedial actions, are not tailored to each level of application, unless a local remedial action is required (e.g. a data gap affecting only one legal entity) since the models are centrally developed. In the case of processes and application portfolios specific analyses on the legal entities (if not included in the development samples and with ad-hoc processes), these are centrally performed by Internal Validation and recommendations are addressed to competent (centralised or local) functions and shared with them to define the proper remedial actions. Gap severity is expressed both at individual and at consolidated level.  A third member notes, the only situation where single models are applied differently at the level of their legal group Entities for Regulatory purposes falls under the so-called “Group-Wide Models”. These models cover customers whose nature goes beyond the domestic dimension and whose risk profile is defined independently from the Legal Entity of the Group that practically manages the relationship. The development of these models is centralised, based on samples covering all the intended application perimeters; the validation is performed consistently, i.e., focusing on the entire scope of application. In case a given model suffers any gap, this is detected at overall level. Additional analyses are foreseen at the level of specific Legal Entities, but these tests have a purely descriptive purpose, and no recommendations are generally issued. Indeed, in case a model shows a good performance at overall level, any given issue identified on local sub-portfolios is not related to a methodological deficiency. In case a correction on the model is needed, this has to be defined and set at overall level: an adjustment driven by a local gap would bias the results on other Legal Entities; the alternative path of defining corrective measures at local level would lead to the violation of the rating uniqueness principle and would introduce the possibility of unwarranted arbitrage among Legal Entities (i.e., the same credit given to the same multinational customer would have different RWs based on the Legal Entities materially disbursing the loan). |

**Question 3:**

**3a) Do you deem it preferential to split the review of the definition of default between IRB-related topics and other topics?**

|  |
| --- |
| The scope of the validation activities should be limited to any issue regarding the IRB model development or risk quantification (including the corresponding appropriate adjustments and margin of conservatism) related to the definition of default. In this regard, the entity agrees with the division of analyses between Internal Audit and Internal Validation functions as described in interaction Box 3 of the Handbook. The Internal Audit is expected to verify the compliance to external regulation and the IT implementation of default detection, while Internal Validation is expected to verify the proper implications on IRB models and relevant RDS. An explicit split would be beneficial. |

**3b) If you do prefer a split in question 3a, which topics of the definition of default would you consider to be IRB-related, and hence should be covered by the internal validation function?**

|  |
| --- |
| Regarding this division of analyses, non-modelling issues (such as the compliance of the internal criteria used for the identification of defaulted exposures with the regulation and the correct implementation of the default definition) are expected to be assessed by the Internal Audit function, whereas issues in the IRB model development or risk quantification are under the scope of Internal Validation activities.  In this respect it would be helpful if the Handbook confirmed:   * For changes related to IRB portfolios, but for which the Credit Risk Control Unit (CRCU) demonstrate that no change is needed to the rating system, internal validation is expected to verify the CRCU’s assessment and, in case it is confirmed, no further validation activity is needed. * For changes related to IRB portfolios for which the CRCU identified the need for change to the rating system (recalibration or redevelopment), internal validation activity is expected on the classification of the change and, consequently, on the rating system. * In both cases, all process-related aspects are expected to be covered by IA as per Content box 1. |

**Question 4: Which approach factoring in the rating philosophy of a model into the back-testing analyses should be considered as best practices?**

|  |
| --- |
| When long-run estimates are back-tested using realised observed default frequencies (ODFs) in short periods of time, non-representative of the likely range of variability (LRoV), results need to be nuanced considering the economic environment in the chosen period of time and the rating philosophy nature of the model. Isolation of the effects of the changes in the portfolio mix for the chosen period for back-testing purposes and trying to capture only the effect of the economic cycle is advisable. Realised migrations between scores and rating grades can be checked for coherence with the expected direction according to the declared rating philosophy and the economic environment in the chosen period. In the same way, results of the comparison of realized ODFs vs estimates can be checked to be coherent with the economic environment.  At this stage of development, where standards on metrics on rating philosophy are not broadly defined in the industry, it is very difficult to define and perform pure quantitative tests. It is also hard to define adequate thresholds to determine whether the rating is point-in-time (PiT) or through-the-cycle (TTC). Further supervisory guidance on the expectations around this topic would be more welcome - e.g. including a list of the type of approaches that would be expected).  We would also note other members provided specific feedback for consideration (though this is not representative of all members approaches):   * *“In our opinion the back-testing of estimates for IRB purposes shall follow a Through-the-cycle calibration philosophy: at each moment in time, PD estimates shall be representative of a quantification of riskiness that reflects an entire economic cycle. In other words, in each and every moment the quantification of the PD shall reflect the LRADR of relevant portfolio. This is to ensure the stability of capital requirements and, in consequence, to limit as much as possible any procyclical effect of the Basel framework. This assumption reflects in the need for a back-testing that is, as a first and most important step, based on the entire portfolio and comparing the PD estimate against the LRADR. Once verified that, at the validation date, the portfolio as a whole has a PD reflecting the LRADR, one can say that the overall quantification is sound and that capital requirements are not distorted by the use of the PD model (of course, limited to the back-testing). It is important to stress, in addition, that generally speaking a series of comparisons between estimates of PD at a given reference date and corresponding 1-year DRs cannot provide a complete assessment of the risk quantification capabilities of a model: this approach, indeed, would return satisfactory results only in case of PIT-calibrated models, returning PD always aligned to most figures current at that date. TTC-calibrated models instead will be aligned with the 1-year DR only on few points of the time series.*   *After this preliminary and general assessment, the Validation shall perform additional tests aimed at better understanding and evaluating the risk-quantification capabilities of the model, in particular focusing on the back-testing at rating class level. In this context, the rating philosophy of a model plays a fundamental role: in case of a PIT rating assignment, the Validation should expect a DR of the given class rather stable through time; in contrast, a TTC rating assignment will reflect in classes showing more volatile DR. In this vein, a failure of a comparison between PD and DR at grade level will give rise to more severe findings in the assessment of a PD model featured by a PIT rating philosophy while in case of a TTC one the misalignment is somehow expected and should be seen as a gap only in case of systematical misalignments observed on several years, typically with the same direction. The results of the comparisons between each 1-year PD and relevant observed DR can also provide additional insight on the rating philosophy, in addition to the analyses required by Article 47(c).”*   * For another member, backtesting PD estimates at rating class level against realised default rates in the short-term is only appropriate for purely point-in-time models, while purely through-the-cycle models would require to be backtested considering the entire likely range of variability across the entire economic cycle. As a consequence, the timeframe considered for the validation sample used for backtesting analyses shall be set consistently with the degree of PIT-ness in practical situations, with the effect of potentially making a purely OOT backtesting of limited significance.   Finally, we would note that within the present handbook, no mention is made to the calibration philosophy; does this mean that the approach to back-testing should depend only on the rating philosophy (and that TTC B-E estimations should be tested in the same way)? |

**Question 5: What analyses do you consider to be best practice to empirically assess the modelling choices in paragraph [76] and, more generally, the performance of the slotting approach used (i.e. the discriminatory power and homogeneity)?**

|  |
| --- |
| The following analyses can be performed to assess either the aggregation methodology or the discriminatory power:   * 1. **Qualitative assessment** on justification provided by the CRCU for the selected sub-factors, their category, sensitivity analysis to the defined weights.   2. **Assessment of distribution of scores** by most and least relevant factors/sub-factors in each type of specialized lending exposures (potential non expected concentration of scores). Concentration metrics on scores on these risk factors can be computed.   3. Observed historical relationship between regulatory slot and default/delinquency status or percentage of historical realized losses. Coherent monotonicity should be expected.   It has to be highlighted that, although these analyses are conceptually sound, the results are conditioned by the low number of defaults typically observed in many sub-exposures classes in the specialized lending portfolio. In addition, if enough data are available, discriminatory power statistics could be computed and assessed.  Furthermore, as per Article 76, the validation assessment covers all items from a) to d). In particular, the modelling choices are evaluated considering the regulatory requirements, the adherence to the internal processes, the market conditions, the products characteristics as well as all expert-based evidence reported in the documentation provided by the CRCU. In case the cut-off values or the weights (used for aggregation purposes) are obtained by means of statistical/empirical approaches, the appropriateness of the methodology adopted is investigated considering, for example, the size of samples used, the type of target variable. In case the weights have been expertly defined, the adequacy and meaningfulness of the justifications are evaluated from a qualitative point of view. In any case, the validation function requires to have a weight equal to 100% for each level of aggregation (i.e., sub-factor components, sub-factor, factor).  With reference to the performance of the slotting approach, the validation function assesses the discriminatory power by means of the Somers’ D metric at different level of analysis (e.g., final score and factor level) considering the default event occurred in the relevant observation period (flag 0/1) as target variable compared to the score obtained at each level of assessment. Also, the representativeness between the modelling samples and the application portfolio is verified by means of the PSI metric considering the most relevant drivers.  On top of these quantitative assessment, the validation function performs additional qualitative/empirical analyses on ongoing basis in order to verify potential concentration or instability phenomena among the final categories. Moreover, an excessive potential frequency of overrides is verified together with their impact on the final categories. In case of not fully positive results, the validation function deeper investigates the evidence in order to verify if the weaknesses could be related to the weights assignment or to the aggregation methodology at sub-factor/factor levels. Finally, for the default event occurred in the relevant observation period, the validation function back-tests the appropriateness of the categories assigned when the position was in performing (e.g., at the beginning of the observation period), considering the expectation that the worse categories (i.e., average and weak) should have been assigned. It is worth noting that all results from both quantitative tests and empirical analyses are evaluated taking into consideration the potential low size of the samples as well as the availability of the default event within the relevant observation period.  Another member further notes that the approach to the validation of slotting approaches covers all the areas provided in par. 76 under model design analyses, assessing the choices done during model development, verifying alignment to regulatory requirements and reviewing critically the choice of aggregation. However, quantitative tests on low-default-portfolios, such as portfolios under the slotting approach, typically do not necessarily provide robust outcomes either for assigning categories at model development stage or for backtesting purposes. Also, as the SA-scale is pre-defined around broad categories, it is not surprising that the portfolios may concentrate around the class consistent with the risk appetite of the bank towards specialised lending businesses, making an assessment of risk differentiation or homogeneity barely interpretable with standard measures.  For these reasons, Validation assessment is typically more grounded on a qualitative assessment of the key modelling choices and on justifications/rationales provided by the CRCU, supported by standard quantitative metrics (distribution, concentration, accuracy, representativeness, calibration, etc.) that do not typically drive Internal Validation conclusions. In the qualitative assessment, the involvement of business expert and the grounding of the slotting criteria on sound economic rationales is a key element. |

**Question 6:**

**6a) Which of the above-mentioned approaches do you consider as best practices to assess the performance of the model in the context of data scarcity?**

|  |
| --- |
| As a general rule, in the context of data scarcity, the qualitative approach is more adequate and the preferred option, as it is more adaptable to the particularities of each portfolio.  Quantitative tests can be computed, however, their results are conditioned by the number of cases, and they have to be put into context of the low number of cases.  Furthermore, consideration for validation could be conducted on data which is not used for the purpose of development. In the context of data scarcity, it is common practice to perform OOT analyses leveraging on the additional time series that becomes available between the beginning of development a validation activity. In case, for any give reason, the Validation assessment relies on time series overlapping (even partially) with the samples used for development, quantitative tests will consider a stricter set of thresholds. |

**6b) More in general, which validation approaches do you consider as best practices to assess the performance of the model in the context of data scarcity?**

|  |
| --- |
| As a general rule, in the context of data scarcity, the qualitative approach is more adequate and the preferred option, as it is more adaptable to the particularities of each portfolio.  Quantitative tests can be computed, however, their results are conditioned by the number of cases, and they have to be put into context of the low number of cases.  Validation samples based on OOT observations becoming available after the beginning of development activities can be integrated with multi-year samples build by random sampling a given portion of the development samples, integrated with the additional year used for validation. For instance, assuming a model developed on the 2010-2020 time series and validated based on 2021 observations, an additional layer of analysis could be the (e.g.) 30% random sample extracted from the overall 2010-2021 time series. This approach could bring additional insight on model performance, of course to be read taking into consideration its overlap with the information underlying estimates. Nonetheless, given data scarcity, it may not always be feasible to be performed, since given that there is lack of data, extracting a random sample may reduce even more the available information.  Indeed, we do not deem the OOT/OOS framework as depicted in the interaction Box 5 appropriate in the context of data scarcity - especially when techniques still do not allow to draw conclusive evidence on the performance of the model, the interpretation of the tests, as well as the overall performance assessment which should mostly leverage on a qualitative approach. |

**Contact:**

|  |
| --- |
| **Constance Usherwood** |
| Director Prudential  [Constance.usherwood@afme.eu](mailto:Constance.usherwood@afme.eu) |

|  |  |
| --- | --- |
| Direct | +44 (0)20 3828 2719 |
| Mobile         +44 (0) 7785 623 439 | |

1. **Specifically:** *“validation function can review and challenge the analyses performed by the CRCU but, in contrast to the general expectations during the first validation, may decide not to perform additional tests if the CRCU tests are deemed adequate in terms of input data and specifications”* [↑](#footnote-ref-2)
2. **Specifically in paragraph 192:** “*it is nonetheless expected to have a good understanding on the issues detected and on the possibilities on how these might be remediated (findings and recommendations). However, in order to maintain its independence, the validation function is expected not to advise CRCU on how to improve certain aspects or to rectify deficiencies, and in any case should always remain critical on any changes implemented on the rating system* [↑](#footnote-ref-3)
3. **Art 142:** Non-transferability of the responsibility it is reported as the following: “*Hence, the validation function of the respective institution will always retain responsibility for the opinion it should get on the rating system (and the related components of this opinion mentioned in paragraph [2.2]). As such, the validation function remains responsible of its validation policy (in particular its soundness), of the correct implementation of the validation methodology and of the final assessment on the rating system (including the follow-up of the validation function’s findings by the institution and, where relevant, of the findings raised by the CA as mentioned in paragraph [105.c]).*”. [↑](#footnote-ref-4)
4. Specifically in interaction box 5: “*Institutions shall establish a rigorous statistical process including out-of-time (OOT) and out-of-sample (OOS) performance tests for validating the model. In particular, institutions need to develop robust models to allow for stable model use across time (and thus to a certain extent across potentially changing environment or even economic conditions). As such, these tests are based on two different samples, as illustrated in a schematic view in Figure 2*.” [↑](#footnote-ref-5)