Deutsche Bank

EU Transparency Register ID Number 271912611231-56

Deutsche Bank AG
Winchester House
1 Great Winchester Street
London EC2N 2DB

Tel +44 20 75458000
Direct Tel +44 20 75451903

Basel Committee on Banking Supervision
Centralbahnplatz 2
CH-4002 Basel
Switzerland

20 June 2018

DB Feedback on the revisions to the minimum capital requirements for market risk

Dear Mr Nesbitt,

Deutsche Bank (DB) welcomes the opportunity to provide feedback on the Basel Committee's consultation on further revisions to the capital requirements for market risk.

We very much appreciate the Committee's open and constructive dialogue with the industry. We welcome the changes introduced in this consultation, in particular those relating to data alignment, sample and frequency, and the test metrics used within the P&L attribution. The additional improvements to the standardised approach and further development of the framework on non-modellable risk factors (NMRF) are also welcome.

DB has contributed to the wider industry response of ISDA/IIF and our individual response should be read in conjunction with those. In this response, we therefore focus on providing more detailed insight and feedback on the metrics of the P&L test, the framework on NMRFs and the newly introduced Annex D. Our conclusions and recommendations can be summarised as follows:

- **The Kolmogorov-Smirnov (KS) test is preferable to Chi-Squared test:** our analysis shows that desks where RTPL differs significantly from HPL and which are therefore not expected to pass the KS test, have a higher likelihood of passing the Chi-Squared compared to KS test. This outcome shows that the KS test is more accurate and therefore better aligned with the objective of the P&L test itself, which is laid out in more detail in section 1.1. The stricter KS test, however, has to be accompanied by an increase in the thresholds of the Red and Amber zones, as the existing testing set up gives rise for a significantly high failure rate even for desks where HTP and RTPL are generally in line.

- **The traffic light approach applied in the P&L tests needs adjustment:** The thresholds determining the amber zone of the KS (and Chi-Squared) test are too tight. This has been identified based on a representative analysis where Red / Amber / Green zones are calibrated in line with confidence levels as used in regulatory Value-at-Risk backtesting. Details of this analysis can be found in section 1.2. In practice, a too narrow amber zone will mean that the main goal of removing significant variance in RWA calculations on continuous basis will not be achieved. Desks will move from the green to the red zone without experiencing any structural changes in RTPL and HPL, and possibly even without moving to the amber zone before. A necessary widening of the Amber zone should not come at the cost of a resulting tightening of the green zone. This would result in even higher capital implications than the current test set up. Hence, in order to avoid an overly restrictive

test framework for IMA eligible desks, it is recommended to recalibrate the thresholds so that the traffic light approach will meet its goal, but respective thresholds for KS (Chi-Squared) test should be no lower than 0.11, 0.15 (18, 40) for upper Green and Amber Zone thresholds, respectively. Also for the Spearman Rank test the respective thresholds need to be calibrated being no larger than 75%, 65% for upper Green and Amber zone thresholds, respectively.

- **Alternative 2 addresses NMRF observability more appropriately:** We welcome the possibility to decouple the risk factor granularity used in the NMRF observability assessment from the ones used in P&L test allowed by Alternative 2. This is because decoupling removes the unwanted incentive to reduce the risk factor granularity within the risk models. Additional comments and suggestions regarding alternative 2 such as a necessary review of the risk factor granularity proposed in the CP is given in section 2.1. This requires some adjustments as set out in our Annex on NMRFs.

- **Seasonality needs to be addressed:** Seasonal trading activities have a meaningful impact on the overall NMRF charge. We would therefore propose to address this by changing the largest gap requirement to the “3 in 90” rule. Concrete evidence for seasonal trading activities is provided in section 2.2.

- **Principles 6 & 7 in Annex D require further clarification:** Annex D has introduced new guidance for evaluating the sufficiency and accuracy of risk factors for IMA trading desk models. In this light DB appreciates the additional flexibility introduced by Principle 7 to employ the original risk factor time series in the P&L test in cases where a proxy is used for risk modelling given that the resulting basis is properly capitalised as a SES charge.

  However, a detailed specification with regard to the interconnection of risk factor modelling across various IMA components such as Value-At-Risk backtesting, P&L attribution test, ES calculation including Full set and Reduced Set, is still missing in the consultation paper. For example, it is common market practice to use beta-factor models for modelling Equity Spot risk that are calibrated on a pool of names in cases where no adequate historical data exists. The P&L attribution test in such cases is not fully specified, but the answer to this could potentially drive significant business decisions. Another example is the handling of non-modellable risk factors in VaR backtesting which is currently not adequately specified. Taking into account that such topics are critical in the implementation phase across banks, we recommend to add clarity on these topics. For more details see section 3.

A more detailed analysis and recommendation of changes to address these concerns are set out in the Annex to this letter. Please do not hesitate to contact us if you have questions or wish to discuss these issues further.

Yours sincerely,

Matt Holmes
Global Head of Regulatory Policy

1. Detailed comments on the P&L attribution test metrics

Summary and Recommendations

- **KS vs Chi-Squared Test:** By investigating mathematical models capturing key drivers in HPL and RTPL differences, it has been found that the Chi-Squared test is more lenient than KS in various scenarios resulting in a higher Type II error. This implies that desks where RTPL differs significantly from HPL and which are therefore not expected to pass the KS test, have a higher likelihood to pass the Chi-Squared compared to KS test. These scenarios arise either when there is a systematic difference (drift) between HPL and RTPL or when RTPL overestimates the risk. In particular, unlike the KS metric, the Chi-Squared one is not symmetric with respect to under- or over-estimation of the risk. Against this background, the KS test can be viewed as the more appropriate test.

- **RAG Zones – KS and Chi-Squared test:** based on our analysis, it can be concluded that the thresholds for both KS and Chi-Squared test are too strict, and in particular, the amber zone is excessively small. This will still result in a substantially volatile RWA attribution as desks move from green to the red zone without experiencing significant changes between RTPL and HPL. In fact it has been shown that given the Green/Amber zone threshold the resulting Amber/Red zone threshold for the KS (Chi-Squared) test needs to be at least 0.04 (22) larger.

  These results have consistently been found based on a set of simplistic mathematical models which allows to assess the impact of varying Drift, over-/underestimation of RTPL vs HPL and the correlation. In such a framework it is possible to tie the red zone to the 99.99% confidence interval once the threshold that defines the green zone is set to the 95% confidence.

  However, the final determination of the Green Zone threshold should only be done taking results from real desks across banks as otherwise the P&L attribution results could be completely driven by RAG zones that have been calibrated in the environment of a mathematical model. Moreover, the upper bounds of the green, amber zone threshold should be no lower than 0.11, 0.15 for KS test and 18, 40 for Chi-Squared test, respectively. This based on results of the portfolios in our test set up with (intrinsic) drift around +/- 8% [1] or a degree of overestimation of risk of 17% [2].

- **Spearman Rank (SR) Correlation test:** A desk for which the (true) correlation between RTPL and HPL is below 75% will likely fail the Spearman rank correlation as the Red Zone threshold is set to 75%. The distance between Green zone and Red zone has been found to be reasonably calibrated (as opposed to KS and Chi-Squared thresholds) only in case the true intrinsic correlation is around 86%.

  However, proxies are frequently used across desks, especially for Equity and Credit desks, which typically exhibit correlations of only 40% - 45% with the original risk factor that they

[1] See figure 5,6
[2] See figure 3,4

proxy. Therefore, if desks are significantly impacted by proxies, there will be a large risk that such desks will generally fail the SR test.

In conclusion, the final determination of the threshold between Green and Amber zone $T_{GreenAmber}$ should only be done taking results from real desks across banks into account. Based on this amber zone threshold, the respective threshold between Amber and red Zone $T_{AmberRed}$ should satisfy:

$$1 - T_{AmberRed} \geq \frac{1 - AT_{GreenAmber}}{0.7}$$

In order to guarantee an appropriately large amber zone [3]. Moreover, it is recommended to set the upper green, amber zone threshold no higher than 75%, 65%, respectively. In our portfolios this would correspond to a true intrinsic correlation of approx. 80%.[4]

1.1. Detailed Comments on KS vs Chi-Squared test

In this section a model is used that allows to assess differences in KS versus the Chi-Squared test.

Test Concept

For assessing any differences between the KS and Chi-Squared test, a simplistic mathematical toy model has been used which allows a straightforward economic interpretation of its parameters. For this, HPL and RTPL are generated as random Gaussian variables according to the model:

$$\begin{pmatrix} HPL \\ RTPL \end{pmatrix} = N\left(\begin{pmatrix} \mu_{HPL} \\ \mu_{RTPL} \end{pmatrix}, \begin{pmatrix} \sigma_{HPL}^2 & \rho\sigma_{HPL}\sigma_{RTPL} \\ \rho\sigma_{HPL}\sigma_{RTPL} & \sigma_{RTPL}^2 \end{pmatrix}\right) \quad (1)$$

Where:

- $N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1 & \sigma_2 \\ \sigma_3 & \sigma_4 \end{pmatrix}\right)$ refers to a two-dimensional normally distributed variable.
- $\boldsymbol{\rho}$ describes the correlation between HPL and RTPL. This can be driven e.g. by the usage of proxies or the omitting of risk factors when calculating RTPL.
- μ_{HPL}, μ_{RTPL} describe any systematic trend of the respective HPL and RTPL, respectively. In case μ_{HPL} and μ_{RTPL} differ, the model displays a systematic difference between RTPL and HPL. Such a difference can be driven e.g. by using different pricing routines when calculating RTPL versus HPL.

[3] In fact 0.7 equals the current regulatory proposed gap between Green and red Zone:

$$\frac{1 - 82.5\%}{1 - 75\%} = 0.7$$

[4] See figure 9

- $\sigma_{HPL}, \sigma_{RTPL}$ describe the volatility of HPL and RTPL, respectively. Any difference in volatilities can be driven by usage of proxies or by omitting risk factors from the risk model.

The fact that model (1) is using a normal distribution is not expected to impact the results significantly as both KS and Chi-Squared test are non-parametric tests.

Results:

When not otherwise specified, $\mu_{HPL} = \mu_{RTPL} = 0$, $\sigma_{HPL} = \sigma_{RTPL} = 1$, and $\rho = 90\%$.

In figure 1 the RTPL has either a correct or an aggressive estimation of the volatility:

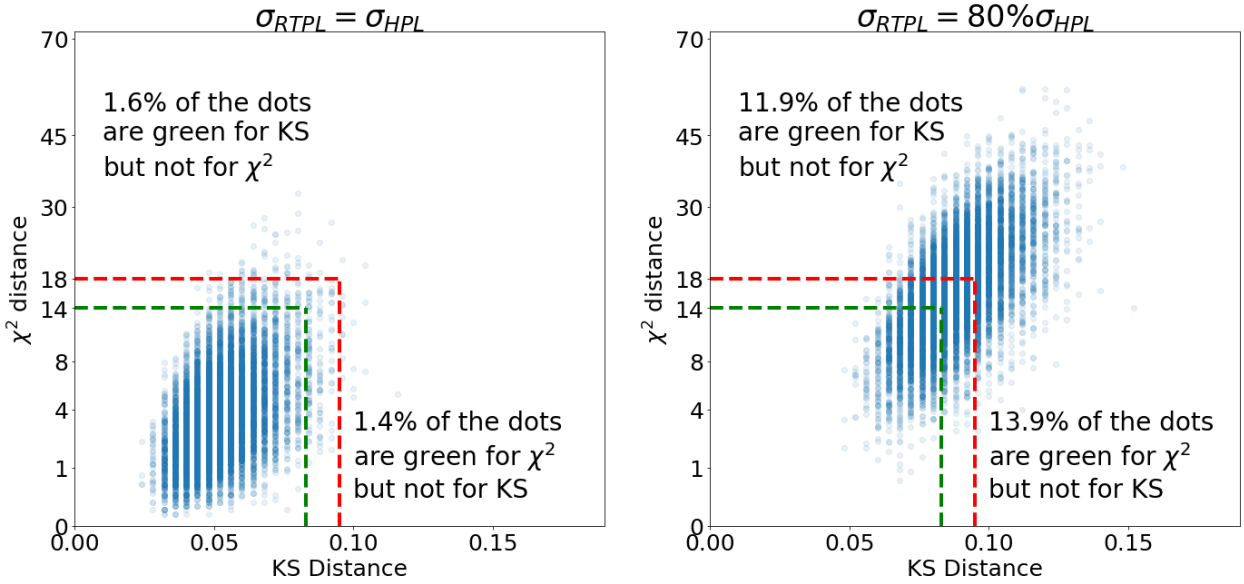

Figure 1: Simulated HPL and RTPL with 90% correlation. Both when RTPL correctly captures the volatility (left) and when it significantly underestimates it (right), KS and Chi-Squared metrics agree on the likelihood of passing the test. In both cases, $\mu_{RTPL} = \mu_{HPL} = 0$ (no drift). Each dot represents a single simulated year out of 10000 simulations. Therefore, different results are only driven by random sampling and the overall cloud represents the distribution of the possible pairs of KS and Chi-Squared metrics. The apparent vertical grid is due to KS metric taking only values which are integer multiple of 1/250=0.004.

However, it is not always the case that KS and Chi-Squared tests give comparable results. The following plots show that, both under the assumption of a deterministic drift, and when RTPL significantly overestimates the volatility, Chi-squared test becomes measurably more lenient:

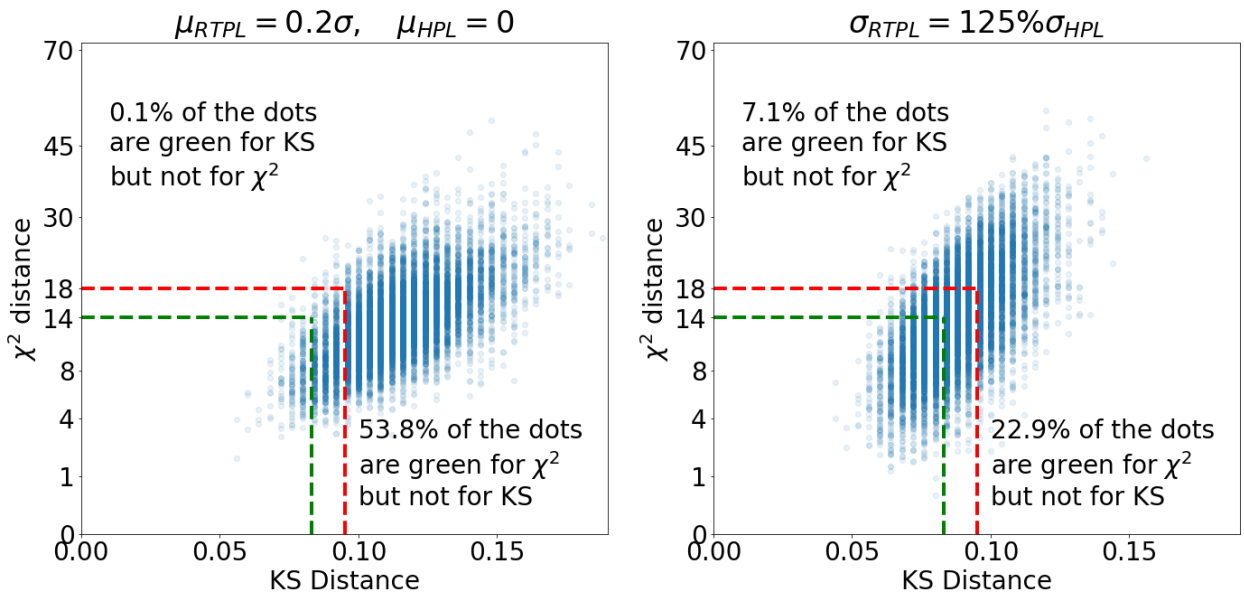

Figure 2: When either there is a significant drift (left) or when the RTPL overestimates the volatility (right), KS metric is measurably more severe than Chi-Squared (i.e. leading to lower Type II error). Correlation is set to 90%. Each dot represents a single simulated year out of 10000 simulations. Therefore, different results are only driven by random sampling and the overall cloud represents the distribution of the possible pairs of KS and Chi-Squared metrics. The apparent vertical grid is due to KS metric taking only values which are integer multiple of 1/250=0.004

We would remark in particular that KS metric is insensitive to whether a lack of accuracy leads to over- or underestimation of the risk, while Chi-Squared favors overestimation over underestimation, as table 3 shows:

	$\sigma_{RTPL} = 125\% \cdot \sigma_{HPL}$			$\sigma_{HPL} = 125\% \cdot \sigma_{RTPL}$		
	Green	Amber	Red	Green	Amber	Red
KS	35%	35%	30%	35%	35%	30%
Chi-Squared	52%	21%	27%	38%	21%	41%

Table 1 - Comparison of KS and Chi-Squared results when RTPL is overestimating the risk (left) or underestimating it (right)), by showing what percentage of the simulated years end up in the green, amber, or red zone. By construction, KS leads to exactly the same results in this case, while Chi-Squared does not.

In summary, the KS test has been identified as the more powerful test compared to the Chi-Squared test in particular in scenarios where:

- There is a significant systematic difference ('Drift') between RTPL and HPL.
- The RTPL overestimates risk compared to the HPL.

In both cases the Type II error of the KS test was significantly lower.

1.2. Detailed Comments on RAG Zones

In this section, a model is used that allows to assess the adequacy of RAG zones as specified in CP with regard to KS, Chi-Squared and Spearman Rank test. Particular focus is on the width of the Amber Zone as the used simplistic models used at this stage cannot be used to identify reliably absolute RAG zones thresholds.

Test Concept

In order to assess whether RAG zones as specified in the CP are conceptually sound, the following approach is followed:

- A simplistic model is used that allows to sample HPL and RTPL dependent on a set of parameters that captures Drift, degree of overestimation of RTPL and correlation between RTPL and HPL. This is the same as has been used in section 1.2 (see equation (1)).
- "Good portfolios" are chosen that are expected to pass the KS, Chi-Squared and SR test. "Good Portfolios" refer to RTPL and HPL samples that refer to model parameters such that a) Drift is reasonably small b) degree of under-/overestimation is reasonably small c) correlation is reasonably close to 1.
- For such portfolios RAG zones can then be calibrated in a convenient way:
  - **Green Zone:** 95% Confidence Interval of relevant statistic
  - **Amber Zone:** Not included in green zone but in 99.99% Confidence interval
  - **Red Zone**: Outside Amber zone

  and compared to RAG zones as specified in CP.

This approach is used in particular to assess whether Amber Zones as specified in CP are too narrow. The results of obviously depend on the chosen parameters that are used to model HPL and RTPL: Only if a consistent observations results from the model across a wide range of parameters and selected "good portfolios", a reliable conclusion can be drawn. What is out of scope of this approach is the identification of adequate threshold levels that e.g. determine the Green Zone as these vary when parameters are changed.

Results when applying three different case studies

Case 1 – Dependency of RAG zones on degree of Over- / Underestimation of Risk

In this section the impact of varying degree of Overestimation, i.e. varying $\sigma_{RTPL}/\sigma_{HPL}$ is investigated. Correlation ρ is assumed 0.9 and no Drift is modelled, i.e. $\mu_{RTPL} = \mu_{HPL}$.

Good portfolios are the ones with $\sigma_{RTPL}/\sigma_{HPL}$ close to 1.

Results for KS and Chi-Squared are plotted in figure 3 and figure 4.

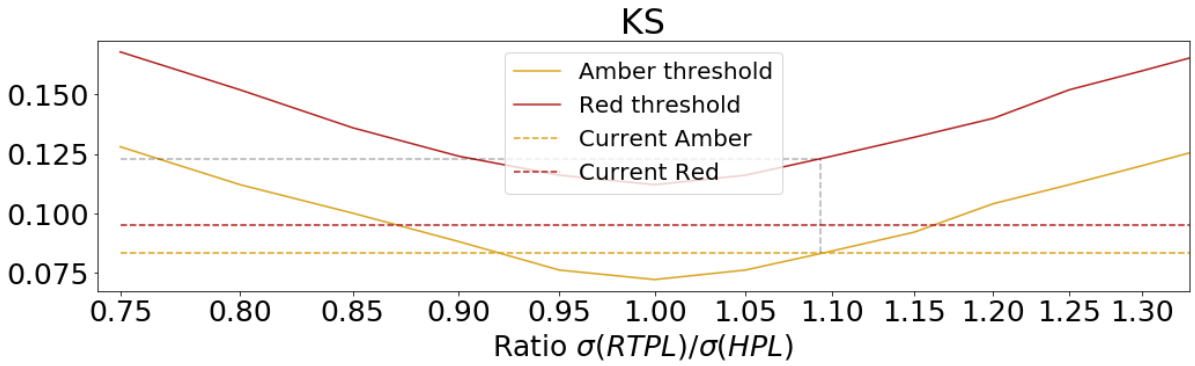

Figure 3 - KS test for varying degree of over- and underestimation of risk. As displayed the gap between the Amber Zone Threshold and the Red Zone Threshold is consistently around 0.04. The Amber zone is significantly too narrow: if one takes $\sigma_{RTPL} = 109\%\sigma_{HPL}$, so that the amber threshold 0.083 of CP corresponds to the 95% confidence interval, then the red zone threshold 0.095 of CP would correspond to 98.9% confidence interval instead of 99.99% ($\alpha = 1.1\%$ is more than 100 times larger than the typical $\alpha = 0.01\%$.)

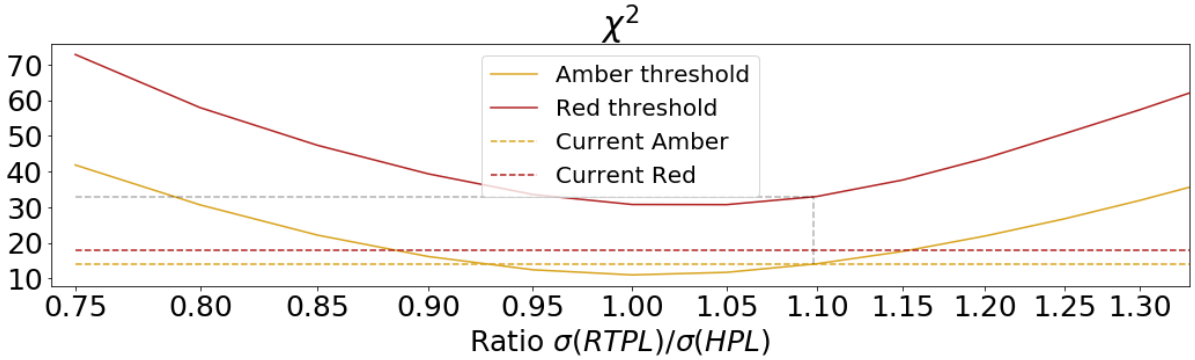

Figure 4 - Chi-Squared test for varying degree of over- and underestimation of risk. As displayed the gap between Amber Zone Threshold and Red Zone Threshold is consistently around 20. The Amber zone is significantly too narrow: if one takes $\sigma_{RTPL} = 109\%\sigma_{HPL}$, so that the amber threshold 14 of CP corresponds to the 95% confidence interval, then the red zone threshold 18 of CP would correspond to 98.8% confidence interval instead of 99.99% ($\alpha = 1.2\%$ is more than 100 times larger than the typical $\alpha = 0.01\%$.)

Outcome: The results show that the gap between amber and red zones for KS should be around 0.04 (vs. the current 0.012) and for Chi-Squared around 22 (vs. the current 4.). These observations are very stable across varied degree of overestimation.

Spearman Rank test was not considered in this case as the correlation is fixed.

Case 2 – Dependency of RAG zones on Drift between RTPL and HPL

In this section the impact of varying degree of drift, i.e. varying $\mu_{RTPL} - \mu_{HPL}$ is investigated. Correlation ρ is assumed 0.9 and volatility of RTPL and HPL are assumed to be equal (e.g. 1 although this doesn't impact the results really).

Good portfolios are the ones with Drift close to 0, $\mu = \mu_{RTPL} - \mu_{HPL} \sim 0$.

Results for KS and Chi-Squared are plotted in figure 5 and figure 6.

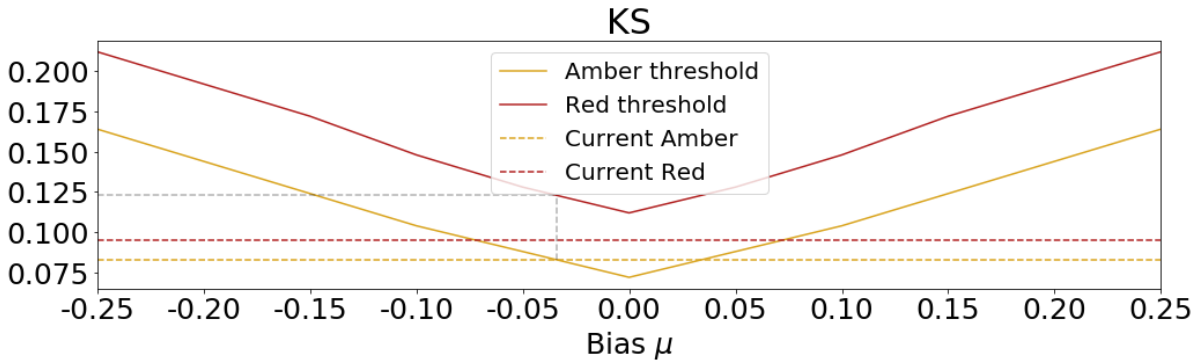

Figure 5 - KS test for varying degree of Drift. As displayed the gap between Amber Zone Threshold and Red Zone Threshold is consistently around 0.04. The Amber zone is significantly too narrow: if one takes $\mu = -0.037$, so that the amber threshold 0.083 of CP corresponds to the 95% confidence interval, then the red zone threshold 0.095 of CP would correspond to 98.9% confidence interval instead of 99.99% ($\alpha = 1.1\%$ is more than 100 times larger than the typical $\alpha = 0.01\%$.)

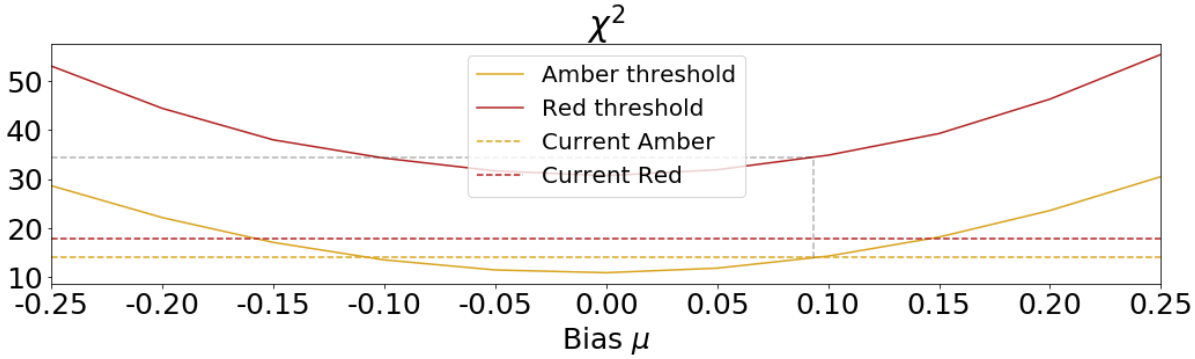

Figure 6 - Chi-Squared test for varying degree of Drift. As displayed the gap between Amber Zone Threshold and Red Zone Threshold is consistently around 20. The Amber zone is significantly too narrow: if one takes $\mu = 9\%\sigma$, so that the amber threshold 14 of CP corresponds to the 95% confidence interval, then the red zone threshold 18 of CP would correspond to 98.7% confidence interval instead of 99.99% ($\alpha = 1.3\%$is more than 100 times larger than the typical $\alpha = 0.01\%$.)

Outcome: The result shows that the gap between amber and red zone threshold for KS should be around 0.04 (vs. the current 0.012) and for Chi-Squared around 22 (vs. the current 4.). Results are stable across varying Drift.

Spearman Rank test was not considered in this case as the correlation is fixed.

Case 3 – Dependency of RAG zones on correlation between RTPL and HPL

In this section the impact of varying degree of correlation, i.e. varying ρ, is investigated. Zero Drift is assumed between RTPL and HPL and volatility of RTPL and HPL are assumed to be equal (e.g. 1 although this doesn't impact the results really).

Good portfolios are the ones with correlation close to 1.

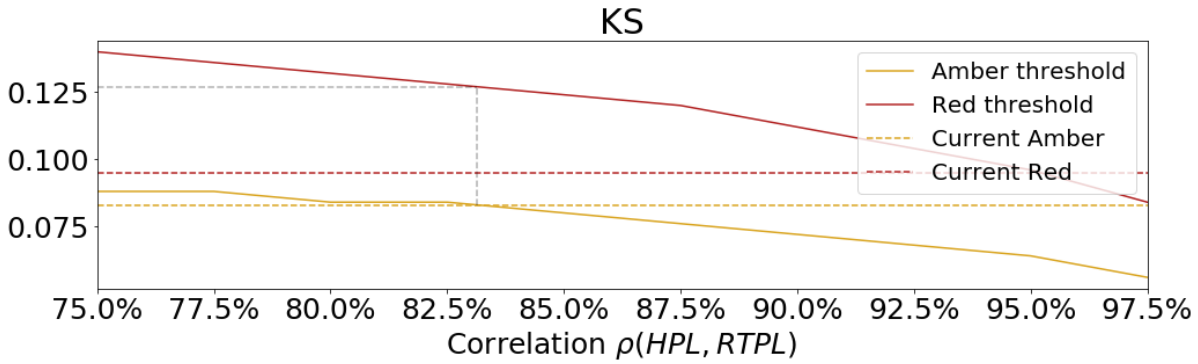

Figure 7 - KS test for varying degree of Correlation. As displayed the gap between Amber Zone Threshold and Red Zone Threshold is consistently around 0.04 (albeit slightly larger for small ρ and smaller for ρ close to 1). The Amber zone is significantly too narrow: if one takes $\rho = 83\%$, so that the amber threshold 0.083 of CP corresponds to the 95% confidence interval, then the red zone threshold 0.095 of CP would correspond to 98.7% confidence interval instead of 99.99% ($\alpha = 1.3\%$ is more than 100 times larger than the typical $\alpha = 0.01\%$.)

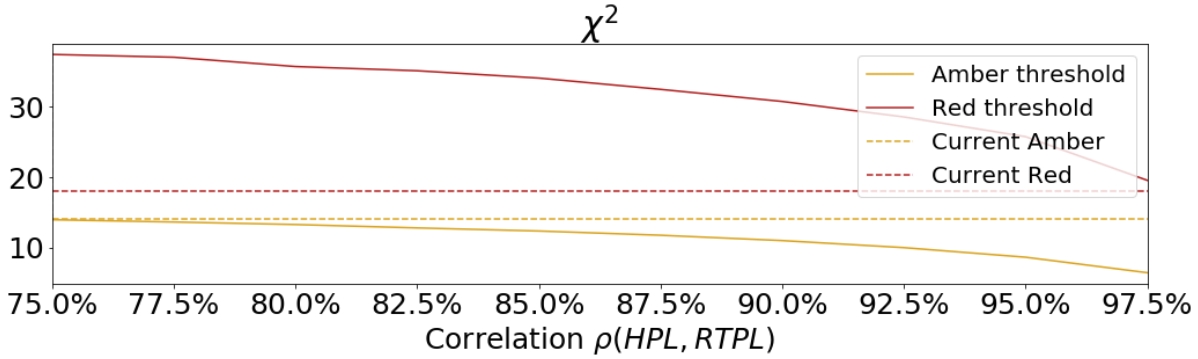

Figure 8 - Chi-Squared test for varying degree of Correlation. As displayed the gap between Amber Zone Threshold and Red Zone Threshold is consistently around 20 (albeit slightly larger for small ρ and smaller for ρ close to 1). The Amber zone is significantly too narrow: if one takes $\rho = 75\%$, so that the amber threshold 14 of CP corresponds to the 95% confidence interval, then the red zone threshold 18 of CP would correspond to 98.3% confidence interval instead of 99.99% ($\alpha = 1.7\%$ is more than 150 times larger than the typical $\alpha = 0.01\%$.)

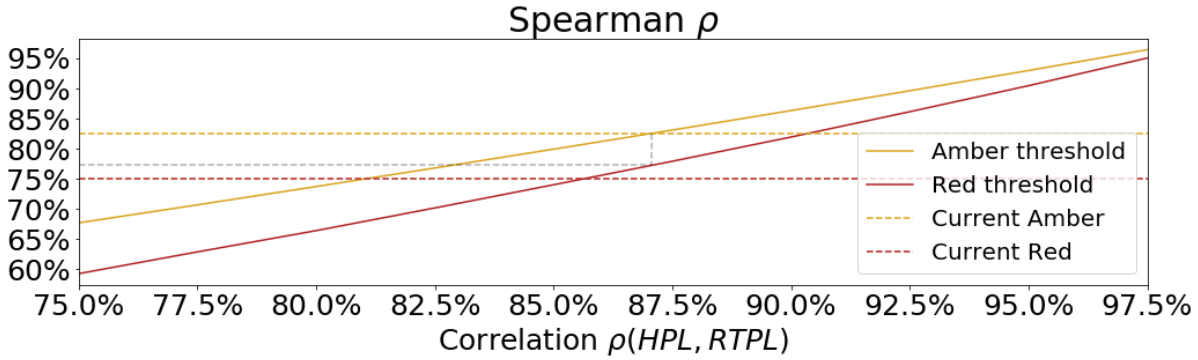

Figure 9 - SR test for varying degree of Correlation. As displayed the Amber zone is not too tight when compared to Amber zone as specified in CP

The result for KS and Chi-Squared test shows that the gap between amber and red zone threshold for KS should be around 0.04 (vs. the current 0.012) and for Chi-Squared around 22 (vs. the current 4.). Results only vary slightly with different correlations.

For Spearman Rank test the result shows that the gap between amber and red zone threshold is more or less aligned to the one specified in the CP (i.e. size of amber zone is not identified as too small) in case of an intrinsic 'true' correlation of ~86%. This gap is actually seen to be quite stable around the value 0.7 which is based on the thresholds provided in the CP:

$$0.7 = \frac{1 - 82.5\%}{1 - 75\%}$$

The analysis suggests therefore that the red zone threshold should be at least 0.7 larger than the respective amber zone threshold.

In summary it has been identified that:

- Amber Zones for KS and Chi-Squared tests as specified in CP are consistently found to be too narrow. Across a wide range of changes in Drift, Degree of overestimation and Correlation, it has been shown that gap between Green and red zone threshold should be at least 0.04 for KS and 22 for Chi-Squared.
- For Spearman Rank Correlation test, the red zone threshold should be at least 0.7 larger than the respective amber zone threshold.

1.3. Additional Comments on Spearman Rank test

In order to assess representative correlations between HPL and RTPL:

1. A distribution of typical correlations between (original) risk factor and its proxy time series is derived. For this a proxy is identified as the time series having the highest correlation in the most recent 1y period and the original time series.

2. Based on step 1 and an assumption on the typical risk factor population that is based on proxies, a likelihood on pass/fail rates is obtained for the existing Spearman Rank test

Results show:

- (1) Typical proxies have generally low correlation and on average around 40-45%. This shows that, for most equities, there exists no possible proxy that has more than 30% to 60% correlation.

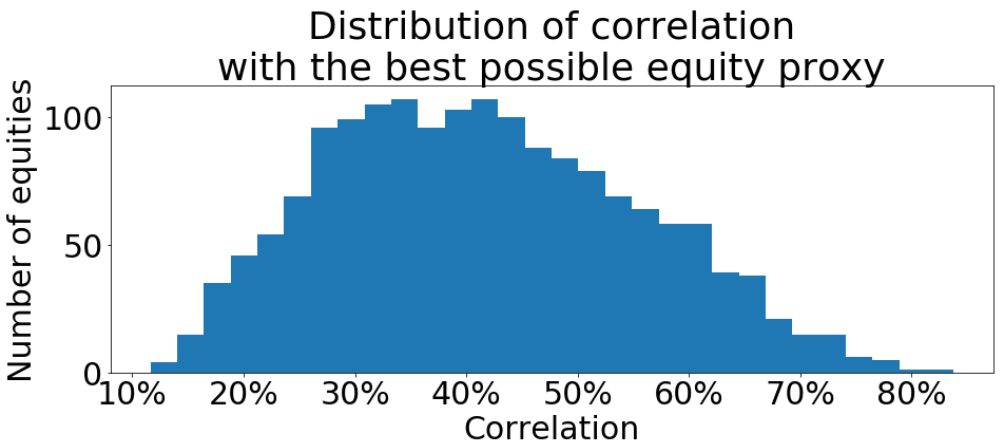

Figure 10 - Distributions of correlations of Equity Spot time series.

- (2) A trading desk where more than 20% (resp. 30%) of its materiality using proxies will fail Spearman test with 82.5% (resp. 75%) thresholds with more than 50% probability.

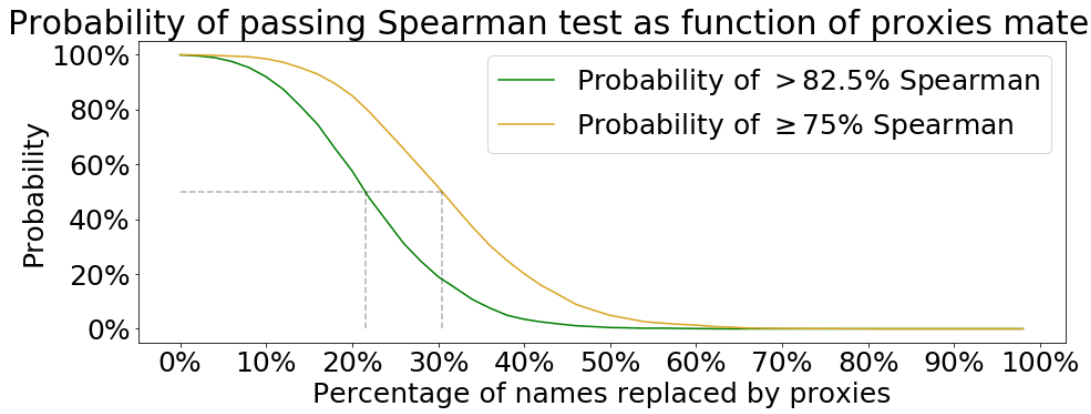

Figure 11 - Probability of passing Spearman Test

These results support the conclusion that Spearman Rank threshold between Green Zone and Amber Zone needs to be calibrated to realistic correlation levels that are displayed by real portfolios – otherwise Equity and Credit desks with significant use of proxies will generally fail.

2. Detailed comments on Non-modellable risk factors (NMRF)

Summary and Recommendations

- **Alternative 2 is preferable, but needs adjustments.** Alternative 2 is currently tightly linked to the risk factor definition of the standardized approach (SBA) for curves. However for some asset classes the SBA granularity seems to be excessive (e.g. commodity risk factors). Adjustments to the alternative should therefore focus on aligning the buckets to SBA where sensible (e.g. IR swaption vols), and a review of granularity for edge cases. In addition, OTM buckets should be simplified to a high strike and a low strike bucket based on moneyness measures commonly used for risk management purposes.

  Lastly, to avoid bank-specific and complex mappings from the standardized observability buckets to bank-specific internal model risk factor sets, the standardized bucket should be used as a benchmark model while banks can define equivalent bucketing schemes for the internal risk factor set. In that case the benchmark model would set an effective minimum quality criterion for banks internal bucketing approaches and facilitate the comparability of modellability results across banks

- **Addressing seasonality:** Changing the largest gap requirement to the "3 in 90" rule would best address the seasonality issue

2.1. Bucketing

The consultative paper introduces two alternatives of risk factor bucketing for the RFET allowing to use same RPOs to assess modellability of all risk factors in a bucket.

Alternative 1 enforces the same buckets for observability and P&L attribution. Banks are allowed to define these buckets in the similar way as banks can currently define their risk factor granularity.

Pro	Con
▪ Banks have control regarding their risk factor set and the observability assessment. ▪ No mapping from standardized buckets necessary.	▪ No real benefit compared to original paper. The flexibility to adjust risk factor granularity within the bounds of PLA was explicitly available in CRR2 and hinted at in original BCBS paper.

Alternative 2 explicitly decouples the risk factors used in the internal model and the PLA from the buckets used for the observability assessment. Observability buckets will be standardized across banks.

Pro	Con
▪ Decoupled granularity between ES/PLA and observability. ▪ Observability assessment somewhat standardized.	▪ Complexity and additional efforts: Buckets need to be maintained over time and mapped to the bank-specific risk factor set. This mapping will be non-

trivial and might introduce misalignment between banks.

- Excessive buckets granularity for some risk factors, especially in moneyness dimension.

We welcome the possibility to decouple the risk factor granularity used in the NMRF observability assessment from the ones used in PLA allowed by Alternative 2. This decoupling removes the unwanted incentive to simplify the risk factor set used in internal models. On the other side it should be noted that additional complexity will be introduced when mapping the bank specific risk factor set to the standardized buckets. For some risk factor classes the suggested granularity of alternative 2 is deemed to be excessively granular.

Alternative 2 is currently tightly linked to the risk factor definition of the standardized approach (SBA) for curves. However for some asset classes the SBA granularity seems to be excessive (e.g. commodity risk factors) which would than translate to challenges for the observability assessment. We also noticed that some 2-dimensional risk factors are not perfectly aligned to SBA buckets (e.g. swaption vols) and further clarification on these deviations would be welcome.

Based on these observations, we would like to bring the following proposals forward, which should reduce complexity, while at the same time still aloow for decoupling risk factor granularity.

1) **Align buckets to SBA where sensible (e.g. IR swaption vols), review granularity for moneyness and other edge cases.**

While the majority of the observability buckets are defined, are in line with the risk factors definition of the standardized approach, there are some deviations that should be resolved. For two dimensional risk factors (such as swaption volatilities where the risk factor space is spanned based on the expiry of the swaption and the tenor of the underlying swap) are treated differently in the standardized approach.

For credit risk factors, the standardized approach allows the netting of sensitivities for names within the same sector. We would propose to define the observability buckets per for each underlying issuer.

2) **Simplify OTM buckets to high strike, low strike buckets based on moneyness measures used for risk management purposes.**

The Alternative 2 proposal stipulates a bucketing scheme in moneyness dimension with a very high granularity that includes 8 buckets. It is industry practice though to model smile and skew risk with a significantly smaller number of risk factors (e.g. skew and smile based on risk reversals and butterflies, SABR vol and correlation, etc). The moneyness measure introduced in the CP is generally not in line with how moneyness is measured for risk management purposes across asset classes.

3) **Use standardized observability results as a benchmark model allowing banks to define their own observability buckets tailored to their risk factor set as long as the quality of the bucketing scheme is at least as good as the benchmark model**

Risk factor definitions used in a bank's internal model are often evolving over time to match the particular trading strategies. This has can lead to significant deviations in the definition and granularity of risk factors across banks. When using standardised observability buckets, these standardised buckets need to be mapped to the bespoke risk factor definition of banks, which is introducing additional complexity and subjectivity. We see this additional complexity as one of the most severe drawbacks of alternative 2. To avoid the unnecessary mapping step, banks could define bucketing schemes in line with their risk factor definitions that is equivalent to the buckets suggested in alternative 2:

- Define a quality measure for bucketing measuring the intra-bucket similarity of risk factors.
- Calculate quality criterion for buckets Alternative 2 buckets and allow internal bucketing scheme that is at least as good as Alternative 2 proposal.
- An example for such a quality criterion based on the Pearson correlation distance can be found in the Appendix 0**.**

The proposed approach has two main benefits. Standardised buckets do not have to be mapped to the internal model risk factor set and the quality and comparability of the customized bucketing scheme is ensured via an objective benchmark given by the alternative 2 buckets.

2.2. Seasonality

Similar to the views from the wider industry, we also witness the impact of seasonal trading patterns on relatively liquid risk factors that fail the largest gap requirement. Therefore, we propose to change the one-month gap rule to an alternative rule that requires at least 3 days with observations in any 90 consecutive days over previous 12 months. The analysis in this section describes rationale of the change and explains how it helps to mitigate the seasonality impact.

Number of observations in the previous year as well the largest gap between any two consecutive real price observations are shown as a scatter plot in **Figure 12** for FX and Rates risk factors[5]. The bucketing approach from the Alternative 2 with adjusted moneyness definition was used. Both RFET requirements, 24 RPOs in 12 months and 30 days for the largest gap are shown by horizontal and vertical dashed lines correspondently. These lines divide plot into 4 sectors:

- Upper left corner: Modellable section containing very liquid risk factors (30%),
- Lower right corner: Non-modellable section containing very illiquid risk factors that fail both requirements (60%),
- Upper right corner: Non-modellable due to the largest gap only with relatively risk factors liquidity (10%) risk factors that fail the largest gap requirement

[5] These asset classes were chosen because pooled data is available for them from publicly available trade repositories. These risk factors should therefore give a good indication of future modellability results.

- Lower left corner: Non-modellable section caused by a lack of 24 RPOs while passing the largest gap requirement (showing that this requirement is way less restrictive then the largest gap one)

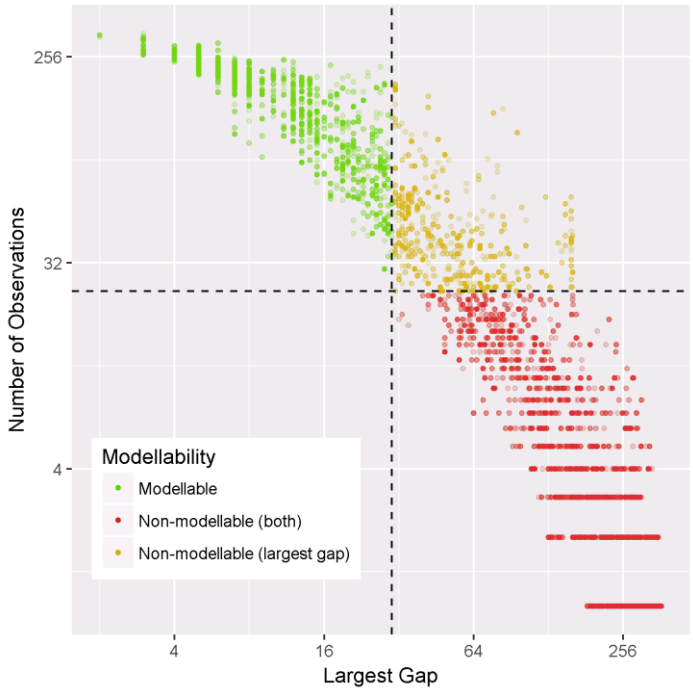

Figure 12. RFET results for FX and Rates (curves, spot and vols) on the SDRs data in the period between July 2016 and June 2017 using bucketing approach from the Alternative 2 with adjusted moneyness definition. Dotted lines correspond to the current RFET requirements: at least 24 observations in previous year and the largest gap not more than 30 days.

The 'yellow' sector in the **Figure 12** contains risk factor that satisfy the 24 observations per year requirement but have a largest gap bigger than 30 days. Because the likelihood of trades occurring is not constant over time a substantial amount of risk factors are non-modellable even though they are fairly liquid. This can be explain by:

- **Seasonal trading activities**: Some months show persistently low trading activities across asset classes and products.
- **Non-constant trade arrival times:** The amount of trading activities changes over time according to demand and supply of individual risk factors. It is quite common that periods of frequent trade occurrences are followed by calmer markets. As a single largest gap can already be sufficient to fail the modellability assessment, a substantial number of risk factors that are fairly liquid most of the year are still non-modellable.

The following figures provides concrete evidence of seasonal trading patterns and their manifestation as gap bigger than 30 days. Risk factors that pass the "3 in 90" rule, but fail the "1 in 30" rule are used as a basis for the analysis and are hence referred to as "seasonal risk factors".

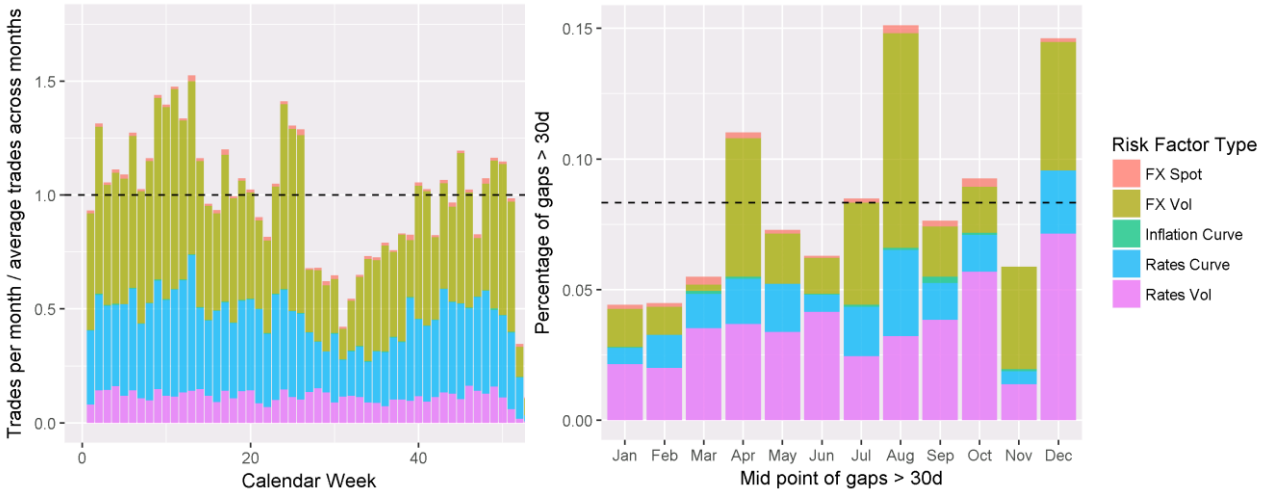

Figure 13. Seasonal patterns are evident in the typical vacation months when a drop in trading activities (left plot) push relatively liquid "seasonal risk factors" beyond the 30d gap (right plot).

Declining volumes of trades in the summer months and in December shown on the left side of **Figure 13** lead to concentration of largest gaps in these months on the right plot where the monthly distribution of middle point of gaps bigger than 30 days is shown for the "seasonal risk factors" (that pass 3 in 90, but fail 1 in 30). This effect can generally be observed across all asset classes that were investigated in this analysis.

Several material vanilla products are impacted by seasonal risk factors:

- For FX and IR volatilities: Standard options are the most material product category associated to seasonal NMRFs
- For Rates curves: Swaps (plain-vanilla, OIS and basis swaps)
- Credit: Government and corporate bond and CDSs
- EQ repo: Futures and options
- EQ volatility: Standard options (long-term index options)

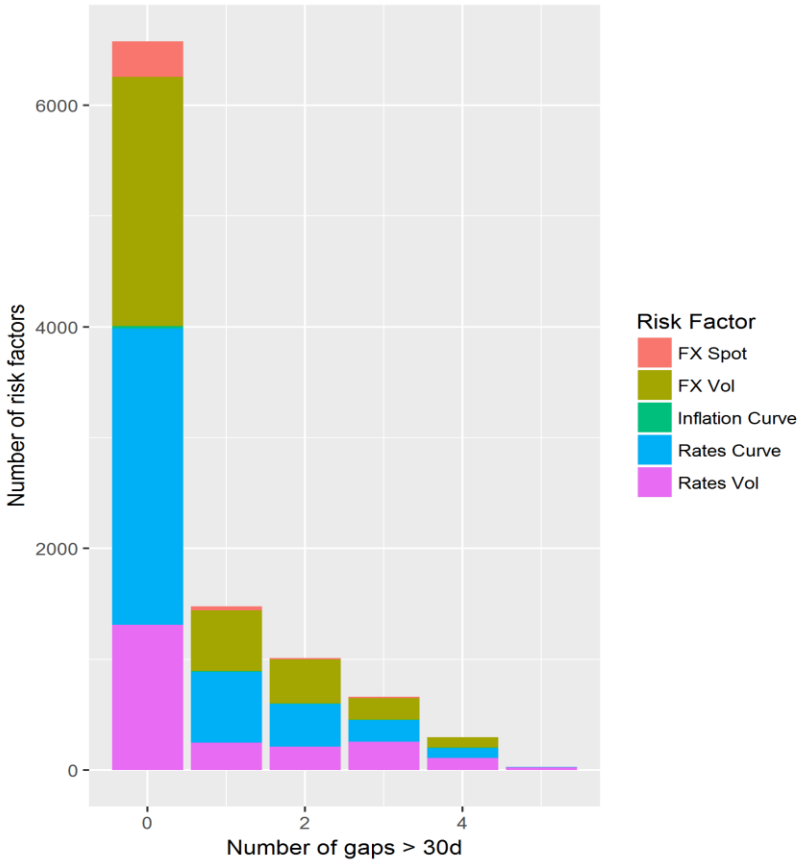

Figure 14. Distribution of gaps bigger than 30 days for the "liquid" risk factors that pass 24 observations criterion. A significant number of risk factors have exactly one gap bigger than 30 days but are still relative liquid with on average more than 80 trading days p.a. (i.e. more than 1 trade per week). They are heavily overlap with risk factors that pass 3 in 90.

Distribution of gaps bigger than 30 days for the risk factors that pass 24 observations criterion across different risk factor types is plotted in **Figure 14**. Thus the first column represents modellable risk factors while the rest of the columns show liquid but non-modellable risk factors. The second column corresponds to the risk factors that have exactly one large gap that is bigger than 30 days. These risk factors are fairly liquid having on average more than 80 observations per year or more than 1 observations per week but would still fail the largest gap requirement due to a single less liquid time period. These risk factors exist in all asset classes and there is a significant overlap with those risk factors that pass "3 in 90" proposal. Making these risk factors modellable would lead to 25% more modellable risk factors.

Conclusion

Changing the largest gap requirement to the "3 in 90" rule would best address the seasonality issue. This is confirmed by the analysis below where we define "seasonal risk factors" as risk factors that fail the largest gap requirement but satisfy both 24 observations and "3 in 90" criteria.

Figure 15 shows the observability results for Interest Rates ATM volatilities based on pooled data from trade repositories. The risk factors are mostly non-modellable outside of G4

currencies. The most material products impacted by these risk factors are swaptions. A change of the requirement from “1 in 30” to “3 in 90” leads to the modellability of some reasonably liquid risk factors that are highlighted in yellow on the left side of the figure, but are considered modellable on the right side of the figure. Typical examples are GBP volatilities and several tenor/expiry pairs of the swaption surfaces for EUR, USD and JPY.

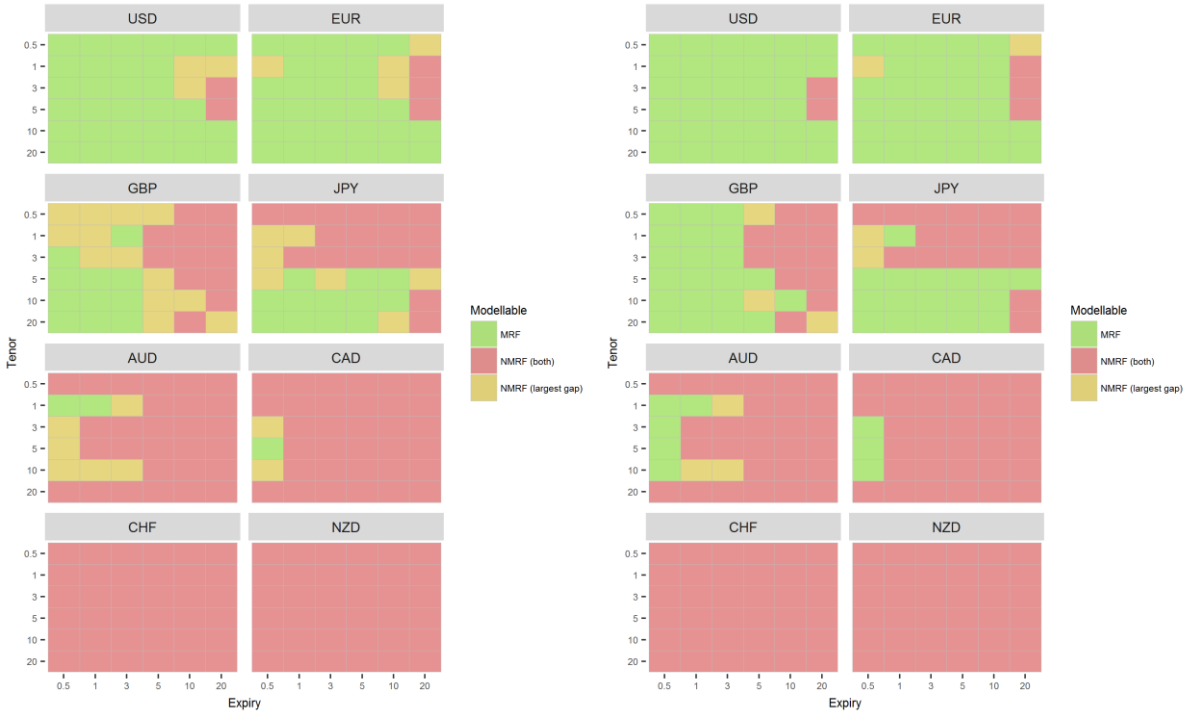

Figure 15. Observability assessment based on current “1 in 30” (left plot) and on proposed “3 in 90” (right) requirement for IR ATM Vols. The main products impacted are European and Bermudan swaptions, caps and floors. “3 in 90” slightly relief observability results impacting only a fraction of “yellow” risk factors. Real price observation data was obtained from publicly available trade repositories and is based on pooled data from a large number of contributing banks. The data does not reflect DB specific positions.

3. Detailed comments on Annex D

Summary and Recommendations

A detailed specification with regard to the interconnection of risk factor modelling across various IMA components such as Value-At-Risk backtesting, P&L attribution test, ES calculation including Full set and Reduced Set is still missing in this Annex and consultation paper. For example, it is common market practice to use beta-factor models for modelling Equity Spot in the risk model that are calibrated on a pool of names if no adequate historical time series exist. The P&L attribution test in such cases is not fully specified, but the answer to this could potentially drive businesses. Another example is the handling of non-modellable risk factors in VaR backtesting which is currently not adequately specified. Taking into account that such topics are critical in the implementation phase across banks, we recommend to add clarity to this topic.

Principle 6: This principle provides guidelines about risk factor modelling and proxying in relation to the Reduced Risk Factor set. Name-specific risk factors that do not have any history in stressed period shall be excluded from the reduced according to the CP. The understanding here is that 'risk factor' refers to the economic risk factor to which a security is exposed to as opposed to the time series that is used to model the risk. In this context, the CP prevents the full alignment of Reduced and Full Risk Factor Set in case banks are exposed to such name-specific risk factors: In fact in many cases such banks would use the original risk factor time series for modelling the risk related to the Full Current Set but be forced to exclude this name-specific time series for any calculations related to the Reduced Set. This has the following consequences:

- The computational burden significantly increases as ES(F,C) and ES(R,C) would have to be calculated additionally for any run,
- The risk calculation of ES is based on the regulatory prescribed scaling approach, which is based on its own methodological assumptions and limitations. A direct approach on the other hand can naturally circumvent these limitations.

The CP is not completely clear in its specifications about how name-specific risk factors without history should be modelled that are part of the Reduced Set. We therefore propose to relax the risk factor modelling of such name-specific risk factors, in case banks can show to regulators that the used modelling techniques are prudent. In this way, banks may decide to target for a direct modelling of ES.

Principle 7:

In this principle additional proxy guidelines and their implications on P&L Attribution test are prescribed. It should be stressed in the CP that this principle refers to risk factor modelling as part of the Full Set as opposed to the Reduced Set. This means that whenever proxies are used for risk factor modelling of modellable risk factors as part of the Reduced Set this doesn't impact P&L Attribution test as long as the original, non-proxied risk factor time series is used for the modelling in the Full Set.

In this set up it is well anticipated though, that the regulators give banks the option to capitalize any proxy basis as NMRF for the sake of using the non-proxied original risk factor time series in the P&L Attribution test.

Detailed Comments on Annex D

Principle 6

In case a bank has exposure to SnapChat Spot, banks currently face a limited number of possible modelling options if they would like to include SnapChat Spot in the Reduced Risk factor Set. In principle the following options exist:

Table 2 - Use case for Risk Factor modelling of SnapChat as part of the Reduced Set

Approach	Description	Example	Allowed by CP ?
Best Proxy (1)	Use best proxy that was available at given time	Use Facebook between today and 2015. Stitch together with Google between 2015 and 2009 and use S&P500 before 2009. "Stitching" means that respective returns between dates of name-shift is zero by construction.	**No** – Models idiosyncratic behaviour
Global Proxy (2)	Use best proxy that existed globally between 2007 and today	Use S&P500	**Yes** – mapped to global proxy that is part of Reduced set for risk calculation related to ES(R,S) or ES(R,C)
Beta Factor Model (3)	Model risk according to beta factor model that covers general and specific risk: $r_x = \beta * r_{bmk} + \epsilon$ Where: • r_x: Risk factor returns of original name • r_{bmk}:Risk factor returns of suitable benchmark • β: beta scaling factor • ϵ: Idiosyncratic noise term, e.g. calibrated on name-specific data or a suitable pool.	Use $\beta = 1$, r_{bmk} = log returns of S&P500, $\epsilon \sim N(0, \sigma)$ and σ calibrated on US technology pool time series	**No** – Models idiosyncratic behaviour

The modelling technique (3) is an approach that banks would use for risk modelling for Equity and Credit risk factors where no appropriate historical data exists. As this technique is excluded by CP, only option (2) is left. As a consequence:

- The risk factor modelling as used in Reduced Set (where e.g. S&P500 is used) is different to the one in Full Set (where SnapChat is used directly). This was Full set will be different from Reduced set leading to an increased computational burden for any bank targeting an alignment.
- The universe of "global" proxies that existed throughout 2007 until today is quite limited, resulting in potentially non-suitable proxy choices and a higher correlation across those risk

factors (e.g. in cases where a significant portion of names is mapped to the same proxy). This may result in distorted capital numbers.

Modelling technique (3) has some advantages over using a global proxy, as it aims at modelling the name-specific risk directly (i.e. if possible calibrated on name-specific data). Excluding such an approach via CP doesn't seem desirable. Moreover, the beta factor Model would not naturally lead to a discrepancy between reduced and Full Risk Factor set.

It should also be pointed out that the scaling approach, i.e. the approach that is based on the scaling of ES(R,S) by ES(F,C)/ES(R,C) has various strong assumptions as it assumes that the correlation structure between risk factors in the current period is similar to the one in the stressed period (if these would exist there).

Principle 7

It should be stressed in the formulation of principle 7, that the usage of proxies as part of risk factor modelling of modellable risk factors, as part of the Reduced Set, should not lead to using this proxy also in the P&L Attribution test. This will be the case as long as the non-proxied, original risk factor time series is used as part of the risk factor modelling of the Full Set. Example use case:

Table 3 - Use case of risk factor modelling in Full and Reduced Set and Implications for P&L Attribution Test

Economic Risk Factor	Modellability Assessment of economic Risk Factor	Time Series used in Full Set	Time Series used in reduced Set	P&L Attribution Test
SnapChat Equity Spot	Modellable	SnapChat Equity Spot	Google	• RTPL:SnapChat Equity Spot • HPL: SnapChat Equity Spot
Korean Won Cross-Currency Spreads (with regard to USD)	Modellable	Korean Won Cross-Currency Spreads (with regard to USD)	EUR Cross-Currency Spreads (with regard to USD)	• RTPL: Korean Won Cross-Currency Spreads (with regard to USD) • HPL: Korean Won Cross-Currency Spreads (with regard to USD)

We fully appreciate the possibility CP provides - in case risk factors are proxied - to either use the proxy in the P&L Attribution test directly or capitalize the basis as an NRMF and use the original, non-proxied risk factor in the P&L Attribution. This is illustrated as part of the following use cases:

Table 4 - Use case of proxies in the Full Set and relation to P&L Attribution test

Economic Risk Factor	Modellability Assessment of economic Risk Factor	Time Series used in Full Set	Capitalize Basis as NRMF?	P&L Attribution Test
SnapChat Equity Spot (A)	Modellable	Google	Yes: SnapChat-Google basis	• RTPL: SnapChat • HPL: SnapChat
Korean Won Cross-Currency Spreads (with regard to USD) (B)	Modellable	EUR Cross-Currency Spreads (with regard to USD)	No	• RTPL: EUR Cross-Currency Spreads (with regard to USD) • HPL: Korean Won Cross-Currency Spreads (with regard to USD)
Small Tech Equity Spot (C)	Modellable	Beta-Factor Model: $\beta * r_{S\&P500} + \epsilon,$	No[6]	• RTPL: Small Tech Equity Spot • HPL: RTPL: Small Tech Equity Spot

In the last use-case (C) the modelling via a beta-factor model should not be regarded as a proxy as the model is calibrated on the name, hence all name-specific components are modelled. As a consequence the (original) Small Tech Equity Spot can be used directly for P&L Attribution test.

[6] In case the bank can demonstrate that idiosyncratic risks are adequately captured.

4. Detailed comments on SBA

In general, we fully subscribe to the wider industry comments on SBA, focussing on FX Curvature, the Positive Gamma / Delta-Curvature disconnect, calibration of the risk weights and RRAO on interest rate yield curve options and variance derivatives. In addition we would like to further emphasise the need to correctly reflect the capitalisation of distress debt positions.

Defaulted positions are generally traded price based (as opposed to rate based). It is observed in the market that for defaulted issues, price changes may not be driven by credit spreads or swap rates. This means that the sensitivities to a change in credit spreads or swap rates may no longer be appropriate for modelling the P&L. Since the product is already defaulted the outstanding risk relates to the recovery of the cash flows and the timing of such recovery. The necessity of a default risk charge is questionable given that the nature of the product lead only to a recovery risk which could be dealt via a residual risk charge only. In addition, the SA rules are not very specific around capitalisation of defaulted/ price based positions:

- FRTB SA DRC rules prescribe a:
  - 100% risk weight for Defaulted exposures, equivalent to 100% PD for defaulted exposure
  - Either 75% LGD for Senior exposure or 100% LGD for non Senior exposures

- FRTB SA Sensitivity Based Method doesn't provide any guidance with respect Price based positions nor on how to treat recovery risk

To illustrate the issue let's consider the below example:

Notional: 100
Price: 10
Seniority: Non-Senior/Equity
Rating: Defaulted

JTD = Max(LGD * Market Value, 0) = Max(100% * (100*0.1), 0) = 10
DRC = Risk Weight * JTD = 100% * $10 = $10
SBM (Equity) = Risk Weight * Delta = 70% * $10 = $7 Possible regulatory interpretation that would lead to a total charge exceeding the max loss.
Total Capital = $17 or 170% capitalized compared to a market value

Recommendation:
Defaulted products are generally only subject to recovery risk. Such positions are charged via DRC 100% RW and 75%/100% LGD, so only DRC charge should be applied. The capital charge for an individual position should therefore be capped at the maximum loss that can potentially incurred.

While for the subordinated exposures DRC SA is already providing adequate capitalisation (which would cover any residual risk and losses), for senior exposures it can be argued that some residual risk exist due to recovery risk. For this reason such exposures could be charged a residual risk add on charge of 0.1% to cover for such residual recovery risk.

Similar to the Basel text change in Paragraph 161 for securitization, the SA framework should include wording around capping the capital charge such that: “The SA capital charge for an individual position may be capped at the maximum loss that can be incurred.”

Appendix Details on proposed bucketing approach based on Pearson correlation distance

The Pearson distance is a correlation distance based on Pearson's product-momentum correlation coefficient of the two sample vectors $x$and y:

$$d_{Pearson}(x,y) = 1 - \mathrm{corr}(x,y) = 1 - \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} = 1 - \frac{(x-\bar{x})\cdot(y-\bar{y})}{\|(x-\bar{x})\|_2 \, \|(y-\bar{y})\|_2}$$

It lies between 0 and 2 and shows how closely time series are correlated distinguishing between positive and negative correlations. Tightly connection to the correlation coefficient simplifies its interpretation in the way that lower distance says about stronger correlation between two vectors. Thus the minimum distance 0 corresponds to fully correlated time series while the maximum distance 2 is obtained for ideally anticorrelated time series.

When applied to the risk factors with a single dimension (curve) the Pearson distance can be used to select maturity buckets with the vertices that have strongest correlations across history. An example of such application to IR swaps is shown in **Figure 16**.

months	1	2	3	6	9	12	18	24	36	48	60	84	120	180	240	360
1	0.000	0.091	0.205	0.427	0.545	0.638	0.694	0.751	0.794	0.822	0.842	0.861	0.880	0.891	0.904	0.918
2	0.091	0.000	0.076	0.278	0.400	0.503	0.573	0.640	0.696	0.731	0.758	0.790	0.819	0.842	0.861	0.880
3	0.205	0.076	0.000	0.174	0.294	0.397	0.478	0.553	0.620	0.663	0.696	0.739	0.780	0.813	0.835	0.854
6	0.427	0.278	0.174	0.000	0.058	0.159	0.238	0.317	0.401	0.457	0.499	0.558	0.616	0.661	0.684	0.704
9	0.545	0.400	0.294	0.058	0.000	0.076	0.131	0.198	0.281	0.340	0.388	0.457	0.526	0.578	0.601	0.623
12	0.638	0.503	0.397	0.159	0.076	0.000	0.033	0.092	0.170	0.230	0.280	0.357	0.436	0.494	0.519	0.545
18	0.694	0.573	0.478	0.238	0.131	0.033	0.000	0.029	0.087	0.142	0.192	0.269	0.353	0.416	0.444	0.475
24	0.751	0.640	0.553	0.317	0.198	0.092	0.029	0.000	0.036	0.080	0.126	0.200	0.286	0.351	0.381	0.415
36	0.794	0.696	0.620	0.401	0.281	0.170	0.087	0.036	0.000	0.023	0.055	0.117	0.196	0.263	0.295	0.335
48	0.822	0.731	0.663	0.457	0.340	0.230	0.142	0.080	0.023	0.000	0.017	0.065	0.136	0.201	0.235	0.277
60	0.842	0.758	0.696	0.499	0.388	0.280	0.192	0.126	0.055	0.017	0.000	0.032	0.092	0.154	0.187	0.231
84	0.861	0.790	0.739	0.558	0.457	0.357	0.269	0.200	0.117	0.065	0.032	0.000	0.035	0.086	0.117	0.161
120	0.880	0.819	0.780	0.616	0.526	0.436	0.353	0.286	0.196	0.136	0.092	0.035	0.000	0.029	0.054	0.095
180	0.891	0.842	0.813	0.661	0.578	0.494	0.416	0.351	0.263	0.201	0.154	0.086	0.029	0.000	0.013	0.045
240	0.904	0.861	0.835	0.684	0.601	0.519	0.444	0.381	0.295	0.235	0.187	0.117	0.054	0.013	0.000	0.018
360	0.918	0.880	0.854	0.704	0.623	0.545	0.475	0.415	0.335	0.277	0.231	0.161	0.095	0.045	0.018	0.000

Figure 16. Example of pairwise distances between 10d returns of vertices averaged over a set of IR swaps.

Proposed buckets as in the Alternative 2 can be now tested to estimate average correlation distance inside the proposed buckets. The obtained average correlation plays a role of a benchmark showing the maximum allowed average distance in further analysis. A bank may now chose any bucket split under condition that the average Pearson distance does not exceed obtained on the prescribed buckets. Our analysis shows that this procedure allows to use less amount of buckets without lost in quality of bucketing in most of the cases. Shows results for the IR swaps using clustering algorithm for selecting of buckets.

Figure 17. Clustering algorithm usually allows smaller amount of buckets then proposed by the Alternative 2 ensuring the same level of average correlations in clusters. The example is given for IR swaps.