[Chief Risk Officer name]
Chief Risk Officer
[Firm name and address]

Sid Malik
Head of Division – Life Insurance and Pensions Risk
Insurance Supervision
T 020 3461 7799
Asad.Malik@bankofengland.co.uk

14 June 2019

PROXY MODELLING SURVEY: BEST OBSERVED PRACTICE

Dear [Chief Risk Officer name]

The purpose of this letter is to share with [Firm name] the results from our recent Proxy Modelling Survey. We would like to thank your team for providing details of [Firm name]'s proxy modelling methodology and engagement with the survey to date. In the remainder of this letter, references to 'you' mean '[Firm name]' for simplicity.

<For internal model firms> This letter is addressed to you in your capacity as [Firm name]'s Chief Risk Officer, because the risk-management function is responsible for the internal model. We have sent similar letters to the Chief Risk Officers of the other firms that participated in the survey.

<For Standard Formula firms> This letter is addressed to you in your capacity as [Firm name]'s Chief Risk Officer, because we have sent similar letters to the Chief Risk Officers of the other firms that participated in the survey. For the other firms, their proxy model formed part of their SII internal model for which the risk-management function is responsible.

The PRA has identified what it considers to be the best practice in the survey responses. This is referred to as 'best observed practice', to emphasise that it refers to the best practice that the PRA has seen across the population of survey respondents. Firms outside this population may apply methods or techniques that the PRA considers to be better. The purpose of this feedback is not to give an assessment of firms' proxy models against the Solvency II internal model tests and standards, or the PRA's expectations with regard to proxy modelling.

The PRA saw a wide range of practice in the survey responses, and no firm had adopted best observed practice in all areas of proxy modelling. The PRA recognises that proxy modelling is an area where thinking and techniques continue to evolve. Given the wide range of practice observed in the survey, the PRA is considering whether to issue a consultation on proposed expectations for how firms can continue to meet internal model tests and standards in respect of proxy modelling.

We note that the most appropriate approach for an individual firm will vary depending on the materiality and complexity of the risks modelled and firms should continue to have regard to these aspects when considering how their method compares to the best observed practice.

<For Internal Model firms> Below we set out the results of the industry survey along with feedback specific to your firm. We have shared the details of best observed practice with all survey participants. The boxed text in each section of the main body of this letter is specific to your firm and has not been shared with other firms. Based on the survey response and documentation provided, we have highlighted where we have concluded that your approach is or is not in line with best observed practice.

We have provided granular feedback in these areas for you to consider as part of your ongoing model development work and we trust that the feedback will be useful.

Although we have attempted to carry out a comparison of your approach against the best observed practice, we expect you to carry out your own assessment against the best observed practice. We note that we have based the below feedback on the documents submitted in response to the survey but we accept that there may be areas where we have misunderstood or misinterpreted your approach. Where this is the case, we would be happy to discuss this with you.

We expect all firms to consider this feedback, and discuss and agree with their supervision team suitable next steps.

<For Standard Formula firms> Below we set out the results of the industry survey, which is information that we have provided to all survey participants. In the interest of transparency, we have also provided firm-specific feedback to other survey participants. We have not however included feedback specific to [Firm name]. This is because [Firm name] uses the Standard Formula and so the proxy model does not fall under the same scope of supervisory review as that of firms that use (or are applying for) an internal model. Even though the information in this letter has been drafted with firms with internal models in mind, we hope that this information may prove useful to you. We do not expect any action on your part, but please let us know if you would like to discuss this further.

Yours sincerely

General comments

In May 2018, the PRA's Proxy Modelling specialism issued a survey to a sample of life insurers with a proxy model. The PRA considered the approaches of eleven firms that use a proxy model (either firms in pre-application for an internal model, with an approved internal model, or using a proxy model for risk management purposes). The specialism designed the survey questions to capture each firm's modelling process along with each firm's views of best practice. The PRA has compared approaches across firms, and considered the input provided by consultancies and other National Supervisory Authorities to identify best observed practice among the survey responses. The team has divided this analysis into eight high-level categories, each representing a key area of the proxy model.

1. Use of the proxy model
2. Fitting
3. Out-of-sample testing
4. Other testing
5. Acceptance criteria
6. Roll-forward
7. Out-of-model adjustments
8. Documentation

The best firms we observed were able to recognise that the proxy models themselves were a 'macro' limitation. Faster and more efficient valuation models in theory would eliminate the need for a proxy model. Firms cited the speed and specification of various modelling software used throughout the end-to-end process as practical constraints in fitting and validating the proxy model. The best observed firms have taken active steps to remove and reduce these constraints.

Use of the proxy model

Best observed practice was for firms to use the proxy model widely in their business. Examples of such uses include

- Economic capital modelling;
- Business management and decision making;
- Sensitivity/stress/scenario testing;
- Setting of risk appetite, capital allocation;
- New business pricing;
- Forecasting;
- Liquidity management.

Firms that had adopted best observed practice demonstrated that the fitting and validation methods were appropriate given the scope of the model. For example, where firms used the proxy model to set risk appetite, best observed practice was to add fitting and validation points in the regions of the loss distribution relating to this use.

Fitting

This section describes 'fitting' the proxy model where firms both determine the form of the loss functions used and parameterise these functions using, for example, regression techniques. This includes techniques such as curve fitting and what the industry commonly refers to as 'Least Squares Monte Carlo'.

There are various approaches to fitting proxy functions in the industry. The 'best' fitting method is dependent on the underlying business and the PRA has no preferred method so long as a firm can justify that its chosen method is appropriate. Best observed practice therefore refers primarily to the quality of justification rather than the methods used. This justification includes both qualitative reasoning and quantitative evidence.

Firms that had adopted best observed practice in this area consider both a wide choice of loss functions and regression techniques with due regard for the specifics of their business. Loss functions considered included ordinary polynomials, Legendre/orthogonal polynomials, piecewise polynomials, splines, and composite functions. Regression techniques considered include linear regression, weighted linear regression, Ridge regression and LASSO regression. Additionally, some firms considered alternative optimisation criteria such as fitting to minimise the maximum residual although the majority of firms seek to minimise the error in the least-squares sense. Firms adopting best observed practice have revisited these choices since the introduction of SII and have improved the accuracy of their proxy model as a result.

The best observed practice was to recognise that the fitting points must provide sufficient coverage of the entire risk space to capture the complexity of the business and the risk profile modelled. A number of firms that use standard curve fitting techniques fit to tens of thousands of fitting points. Firms that had adopted best observed practice in this area selected fitting points that systematically covered the entire risk space, supplemented by fitting points chosen using expert judgement at points of known discontinuities, or other features that are likely to cause fitting errors.

Firms generally acknowledged that interaction effects that introduce the need for cross terms between risks increased the complexity of the model, were more difficult to fit and required additional consideration as to the approach taken to ensure adequate coverage. Features such as the existence of guarantees/options, reinsurance and dependency of valuation on a dynamic Matching Adjustment under stress are all examples which vastly increase the complexity of the model. Best observed practice was to consider sources of interaction terms from first principles, for example by considering interactions among risk factors used in discounting and risk factors used in cashflows. Firms that had considered such interactions generally concluded that there were complex interactions between many risk factors and chose their multivariate points with the aim of capturing material interactions. Although firms surveyed acknowledged that more multivariate points were needed to fit cross-terms for interactions, in practice, many of them focused on univariate points and used more of these than multivariate points.

In addition, best observed practice was to select fitting points considering the various uses of the proxy model beyond the SCR calculation.

In general, the respondents acknowledged that manual intervention in the fitting process should be limited where possible. Best observed practice was to choose the form of the loss function using a combination of data analysis and expert judgement. Where firms applied expert judgements, the best observed practice was to perform sensitivity testing of these judgements. Examples of key expert judgements include the choice of loss function form (including identification or exclusion of cross terms), regression approach, fitting points, fitting range, granularity of loss functions etc. In addition, firms operating in line with best observed practice had established a feedback loop between fitting and validation, using previous validation tests to inform future fits.

Best observed practice was to carry out testing on the loss function fit, before carrying out any out-of-sample tests. This is standard practice in academia and other industries and allows firms to assess objectively the loss function fit, making maximum use of the information available at this stage. This testing helps firms to demonstrate that the form of the proxy function is materially appropriate and helps to identify bias and/or overfitting of the model. Tests conducted include the following:

- Homoscedasticity of residuals;
- Analysis of the number of under/over statements of residuals;
- Normality of residuals;
- Independence of residuals;
- Statistical significance of loss function terms;
- Bootstrapping;
- Jack-knifing;
- Graphical analysis of bivariate fit (sense/reasonableness checks).

Each of the tests outlined above can highlight issues with the fit of the model, which firms could correct at the fitting stage before rolling forward or commencing validation.

Please note the points above only relate to tests on in-sample residuals during the fitting stage. The sections below on **Out-of-sample testing** and **Other testing** contain details of tests applied during the validation stage.

Best observed practice is to fit at a level of granularity that is commensurate with the complexity of the business and the risk profile. Examples of this could include fitting loss functions at a level of granularity that allows the modelling of homogenous risks. Some firms consider it necessary to fit loss functions separately to assets and liabilities, or more granular subsets of these where risk exposures are sufficiently heterogeneous. These firms recognised the potential for loss functions at higher levels of aggregation to contain offsetting fitting errors that would be difficult to isolate and identify when validation tests failed.

Out-of-sample testing

This section describes out-of-sample testing, where firms compare the result of the proxy model to the result of the heavy model at scenarios that are distinct from those used to fit the loss function. What we describe as an 'out-of-sample test' corresponds to a check on the size of this error, where the scenario passes the test if the size of the error is within a certain threshold set by the firm.

Many of the firms surveyed were of the view that the ultimate test of the predictive power of the proxy model is a direct comparison with the heavy model through out-of-sample testing. This is not out of line with academic validation techniques in this space. All firms surveyed carried out some form of out-of-sample testing, typically on-cycle at the reporting date and off-cycle between reporting dates.

Firms and the PRA recognise that the appropriate number of out-of-sample points depends on the risk profile and complexity of business underlying the proxy model, in particular the number of material risk factors and of material interactions between them. Both the quantity and the quality of the points are important, as is the efficient allocation of heavy model runs.

In terms of quantity, the largest number of out-of-sample tests conducted at the reporting date (i.e. 'on-cycle') that the PRA observed was more than 1,000. A number of other firms conduct hundreds of out-of-sample tests on-cycle. Best observed practice was to justify the number of out-of-sample tests with reference to the number of risk factors and material risk interactions modelled.

Firms that had adopted best observed practice in this area recognised that each additional risk dimension generally required an exponential increase in the number of out-of-sample points to test each region of the risk space. Some firms failed to recognise this point and specified the minimum required number of tests as a multiple of the order of the polynomial or the dimension of the risk space. Although no firm has succeeded in testing every area of the risk space (which the PRA considers to be reasonable on proportionality grounds given practical constraints and materiality considerations) those that acknowledged this exponential link to the dimensionality not only tended to have a higher number of out-of-sample points but also displayed more justifiable point selection.

In terms of validation quality, best observed practice was to recognise that the methods for selecting the out-of-sample scenarios are also critical, especially in the context of modelling restrictions limiting the number of points that firms can produce in the available timeframe. Best observed practice was to select points with the clear intention of giving a true indication of the model's goodness of fit across the risk space and not just the loss distribution. Most firms ensure that multiple points on the loss distribution are tested to demonstrate goodness-of-fit for multiple model uses.

Best observed practice was to use a combination of points selected randomly, systematically, and statistically across the risk space, with carefully chosen points informed by expert judgement. Firms that had adopted best observed practice in this area ensured that across the full range of points, key percentiles of material risk factors were tested along with key interactions between risk factors. Some firms use statistical techniques to ensure good coverage of the risk space. Points selected in an automatic or systematic way have the advantage that they are entirely independent of the modeller and are more likely to test a broader range of the risk space. However, they may not always capture key risks or interactions. Similarly, manually selected points suffer due to bias but can benefit from the experience of the modeller. Therefore, best observed practice was to use a combination of both approaches.

Most firms focus validation tests on the same region or plane of the risk space; either in one validation cycle e.g. points focussed on the SCR window, or across multiple cycles e.g. the same percentiles targeted each quarter. Particularly where firms had limited run budget, this made it more difficult for firms to demonstrate the best observed practice of ensuring that a broad area of the risk space is

validated. It also made it more difficult for firms to demonstrate the best observed practice of validating points for alternative uses of the proxy model.

Although we recognise the importance of testing the SCR region, we note that the purpose of validation could also aim to identify regions where the model does not fit well. These regions may be relevant if they could produce errors that might alter the risk profile of the SCR region, for example by changing the ranking of scenarios. This could be important even if it does not materially affect the monetary amount of the SCR at a given point in time since it can impair the model's ability to rank risks and its role in the capital allocation process and other Use Test requirements.

Other testing

The previous section describes out-of-sample testing, where firms compare the result of the proxy model to the result of the heavy model at scenarios that are distinct from those used to fit the loss function. What we describe as an 'out-of-sample test' corresponds to a check on the size of this error, where the scenario passes the test if the size of the error is within a certain threshold set by the firm. This section discusses other tests that are not a direct comparison of the size of out-of-sample errors.

In addition to the out-of-sample testing described above, the best observed practice was to carry out a range of additional tests to support the goodness-of-fit assessment of the loss function. Some of this testing is carried out on-cycle and some is performed off-cycle and used to inform future fits. Most of these tests could, in theory, be carried out using the existing suite of out-of-sample scenarios on-cycle and hence provide a richer view of the model fit. The most commonly applied tests are:

- Homoscedasticity of errors;
- Analysis of the number of under/over statements of errors;
- Normality of errors;
- Independence of errors;
- Graphical analysis of bivariate fit (sense/reasonableness checks, examination of turning points/discontinuity points, behaviour at the extremes etc.);
- Ranking tests of the loss distribution (relationship between heavy and proxy models);
- Quantification of mis-estimation of the SCR;
- Tests for overfitting;
- Analysis of change of the form of the loss function;
- Backtesting the loss function fit;
- Sensitivity/stability testing of key parameters/expert judgements.

Given that the main purpose of the proxy model is to allow firms to rank risk, the firms with best observed practice recognised that the quality of a proxy model relies on its ability to rank scenarios. These firms therefore conducted statistical ranking tests in addition to looking at the out-of-sample error under any individual scenario, or group of scenarios.

All firms use a large amount of expert judgement in their proxy models. The best observed practice we identified was to provide quantitative evidence to test these expert judgements, and not rely solely on a priori reasoning.

Acceptance criteria

The previous sections describe various tests that firms were observed to carry out in order to validate the proxy model. This section discusses what criteria firms applied to the results of these tests to determine whether the proxy model fit was acceptable. The comments in this section mostly relate to out-of-sample tests, since this is the most prevalent approach of surveyed firms. We have not highlighted best observed practice for tests other than out-of-sample tests as approaches taken to the tests described under ***Other testing*** *were too inconsistent to allow a meaningful comparison.*

The firms that had adopted best observed practice in this area selected acceptance criteria to highlight material errors on an individual basis for each out-of-sample point. In addition, the best observed practice was to have acceptance criteria that are specific to each layer of aggregation, where relevant, to ensure a good fit locally as well as overall.

Most firms specify acceptance criteria for individual out-of-sample tests which are of the form 'Test passes if the maximum absolute error observed is less than £Xm', where various methods for calculating '£Xm' were observed across firms. The best observed practice was to specify this tolerance with reference to the heavy model loss in that scenario i.e. 'Test passes if the absolute error observed is within X% of heavy model loss in that scenario'. This method had the advantage of avoiding the circularity of specifying this tolerance as a percentage of SCR, which is an output of the proxy model. It also avoided the risk that a tolerance set in fixed £m amount is too high or low relative to the area of the loss distribution being tested. For example, setting a tolerance of £Xm because this relates to Y% of SCR gives a criterion which may be a reasonable margin of error in the SCR region, but which allows for substantial errors to be accepted at lower percentiles.

Some firms investigated aggregate statistics of the fit and set acceptance criteria for these. Firms considered that these statistics could usefully summarise the overall fit. Firms that demonstrated best observed practice supplemented individual out-of-sample tests with aggregate statistics, selected aggregate statistics that did not mask offsetting errors (for example using mean absolute error or mean squared error, thereby avoiding the potential to offset large positive and negative errors), and monitored aggregate statistics over time.

Firms that have adopted the best observed practice in this area have a range of remediation actions planned in response to testing failures, which can be carried out on-cycle to ensure that the reported SCR is in line with Calibration and Statistical Quality requirements.

Roll-forward

Firms operating in line with best observed practice have invested in the efficiency of their modelling processes, allowing them to re-calibrate and validate the proxy model in full each quarter (i.e. on-cycle). This means they do not have to rely on a roll-forward process. Roll-forward methods can be a source of significant proxy model error, particularly if they are applied over a long time period, or are not thoroughly validated.

However, we recognise that some firms may favour the use of a roll-forward where their IT infrastructure does not allow a re-calibration on-cycle or for other uses (e.g. solvency projection). Best observed practice in this area is to:

- Validate loss functions after they have been rolled forward, on-cycle before the SCR has been reported.
- Carry out additional off-cycle validation of the roll-forward model and use the results of this to iteratively improve the roll-forward process. This includes comparison of the rolled forward loss function against re-calibrated loss functions at the same date (i.e. backtesting).
- Identify trigger points that would invalidate the use of a roll-forward approach and have outlined remediation actions that could be employed in the event these triggers were breached.

Out-of-model adjustments

The proxy model is by definition approximate. Despite the best model fit and validation results, any error, regardless of whether this is an over- or under-statement of the heavy model is evidence of model error. Observed errors add to the uncertainty in the proxy model's ability to rank the scenarios and hence select the correct SCR biting scenario. This means that the loss given by the heavy model under this scenario does not necessarily correspond to the SCR that would result from a perfect proxy model free of error. Best observed practice in this area was to recognise the inherent uncertainty in the model output.

For most of the firms surveyed, the primary use of the proxy model was to calculate the SCR. These firms typically compare the proxy model loss and the heavy model loss in the SCR region. Firms typically either

i. Made no adjustment to the proxy model implied SCR if errors observed were within a stated threshold.
ii. Made out-of-model adjustments to the proxy model implied SCR based on test results at the biting scenario only.
iii. Made out-of-model adjustments to the proxy model implied SCR based on test results over the wider SCR region. Firms used a variety of approaches to quantify the out-of-model adjustment in this case, largely dependent on the relevant validation test conducted.

In some cases, adjustments had the potential to be negative, giving a lower SCR than the proxy model implied. In all cases seen in the survey responses, adjustments focussed only on the SCR window and no other regions of the loss distribution.

Firms that demonstrated best observed practice applied a strictly positive adjustment to the SCR recognising the risk of model error and the prudence of this approach. Firms that applied a strictly positive adjustment were at relatively lower risk of understating the SCR as a result of model error, than if they had applied no adjustment or a negative adjustment.

Best observed practice is to ensure that out-of-model adjustments made to the final loss functions, such as allowances for the loss absorbing capacity of deferred taxes or management actions, are robustly tested to ensure that the ranking of scenarios is not affected. This includes assessing the impact under each risk scenario. Calculating these adjustments only for the biting scenario would disregard how the impact may vary for different scenarios and hence how the ranking of scenarios may be affected.

Documentation

Firms operating in line with best observed practice evidenced (as part of their survey response) documentation that was clear and detailed enough to understand the full operation of the model. These firms produced up-to-date documentation covering separately the fitting and validation methodology.

Where the same team carried out these activities, a separate independent review report was produced. Best observed practice was for documentation to contain appropriate details of each of the following:

- Purpose and scope of the model;
- Limitations of the model;
- An overview of the methodology employed;
- Triggers for review;
- Clear articulation and justification of acceptance criteria;
- Results of the fitting exercise;
- Results of the validation exercise;
- Areas for off-cycle investigation;
- Ongoing monitoring of fit/backtesting etc.

Best observed practice was to have clearly documented procedures in the event that the model validation process identified quantitative or qualitative problems with the reliability of the model, and to have put in place a decision-making process and governance to resolve these problems on-cycle.

Best observed practice in results reports included results of the tests outlined in the above sections on **Out-of-sample testing** and **Other testing**. This included:

- testing at multiple layers of aggregation
- separate line by line analysis of errors
- graphical analysis of errors
- details on the number of risk factors modelled
- the number of fitting points per univariate/multivariate terms
- the number and nature of out-of-sample tests per loss function etc.

This level of detail allows the rationale for why the number of tests is sufficient for the risk space modelled to be ascertained, and for the risk function to independently review the model fit and/or validation.

Appendix A: Glossary of terms

Term	Meaning
Acceptance Criteria	The criteria used to assess the validation errors and determine whether or not they are acceptable and, if not, what subsequent actions to take.
Basis functions	See 'Loss function' definition.
Biting window / scenario	The window/scenario around/at the relevant percentile (e.g. for the SCR this is the 99.5th). Also known as the critical scenario or killer scenario. Window refers to a range of scenarios either side of the biting scenario itself, ordered by losses.
Burnthrough	Where With-Profits funds are unable to meet their liability cash flows using resources within the fund and the shortfall is met by the Shareholder fund or remainder of the entity, the losses are said to 'burn through' the ring-fenced nature of the With-Profits fund and the amount of loss is referred to as a 'burnthrough cost'.
Calibration Cycle	The cycle over which the proxy model fitting (and some validation) processes occur. This may coincide with the use of the proxy model for reporting or it may take place in advance of the reporting cycle, with roll-forward or other techniques used to adapt the calibrated proxy model for use at the balance sheet date.
Cross-Term	A loss function term involving more than one risk factor, which captures the interaction between two or more risks. For example, a cross-term for equity risk and lapse risk might be 'equity*lapse'.
Curve fitting	The process of constructing a function, usually using polynomials, that has the best fit to a series of data points. Most often, this function is determined using least squares regression. Where we refer to 'Standard' Curve Fitting this corresponds to the approach where the data points used to fit the function are accurate heavy model results. This is to draw a distinction between this approach and the Least Squares Monte Carlo approach, which fits to approximate heavy model results. However, both approaches could be considered to be Curve Fitting.
Empirical loss function	A loss function based on interpolating losses from the heavy model at a number of fitting points, as distinct from fitting a polynomial or other function to the fitting point losses.
Fitting Points	A fitting point is a combination of risk factor realisations evaluated using the heavy model to give a loss that the loss function is fit to. Each fitting point can be thought of as a 'scenario test'. Proxy models are typically calibrated using appropriate sets of fitting points. Any point used during the fitting phase is classed as 'in-sample'.
Goodness of Fit	The goodness of fit of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. In proxy modelling, goodness of fit measures the extent to which the proxy model matches the heavy model.
Heavy model	The heavy model is the asset and/or liability valuation model which is approximated by the proxy model. Also referred to as Valuation model, Cashflow model, or Full model.

Term	Meaning
Hetero- / homoscedastic	A set of random variables is homoscedastic (heteroscedastic) if all the variables have the same (varying) finite variance. In the context of proxy modelling this relates to the variance of each of the in-sample residuals or out-of-sample errors.
On-cycle	Actions are described as carried out 'on-cycle' when they are completed within the reporting deadlines for a particular reporting date. This is also referred to as 'in-cycle'.
In-sample Points	See Fitting points.
In-sample Residual	The difference between the loss function and the heavy model when both are evaluated at a fitting point.
Joint stress/ multivariate point – either fitting or validation	A point taken from the risk space where more than one risk is being stressed. Also called an off-axis point.
Least Squares Monte Carlo	A proxy modelling technique used in cases where the heavy model for valuation is based on Monte Carlo simulation. Such models are time-consuming to execute and this tends to limit the number of fitting points that can be examined. Least Squares Monte Carlo uses a large number of possible fitting points with relatively small numbers of scenarios used for the heavy model valuation, and then fits a loss functions through the fitting point results using least squares regression.
Loss distribution	The loss distribution arises from applying loss functions to the risk space (see below) to give monetary losses. It can be thought of as 'collapsing' the risk distribution into a single dimension, and as such many combinations of risk factors can give the same loss.
Management Actions	Management Actions has the same meaning as in Articles 23 and 236 of the Solvency II Delegated Regulation. For example, trading assets or reducing discretionary bonuses to reduce losses in stressed conditions.
Order of polynomial	The highest exponent of the terms of a polynomial. A polynomial of form $Y=a+bx+cx^2$ has order two. A polynomial of the form $Y = a +bx + cx^2 + dxz^2$ has order three from the xz^2 term which is equal to x^1z^2.
Off-cycle	Actions are described as carried out 'off-cycle' when they are completed outside the reporting deadlines for a particular reporting date. This is also referred to as 'Out-of-cycle'.
Out-of-Model Adjustments	Any adjustment made to loss functions or the output of the loss functions that is not part of the fitting process (either manual or otherwise).
Out-of-sample Errors	The difference between the loss function and the heavy model when both are evaluated at an out-of-sample point.
Out-of-sample Points	Any point which represents a combination of risk factor realisations evaluated using the heavy model to give a loss that is not used in the loss function fitting phase. Also sometimes referred to as validation points.

Term	Meaning
Over-fitting	A modelling error which occurs when a function is too closely fit to a limited set of points. Typically this manifests as loss functions which contain too many terms or terms of too high order. Overfitting introduces the risk that the model will fail to give reasonable results when applied to points outside the original fitting set.
Loss function	A function representing the value (or change in the value) of an asset, liability or other variable given changes in the value of underlying risk factors. This is also referred to as a 'proxy function' or a 'basis function'.
Proxy Model	A simplified model used to represent another model (typically called the 'full' or 'heavy' model) without incorporating full and explicit calculations. A key use of proxy models is to represent changes in the Solvency II balance sheet over a one-year time horizon as functions of simulated risk factors. A proxy model consists of a collection of loss functions as defined below. Also referred to as Lite model.
Replicating portfolio	A particular type of proxy model in which a portfolio of assets is designed to replicate the features (sometimes the cash flows, or the values) of a particular block of liabilities. The value of the liabilities are determined under various scenarios by determining the value of the replicating portfolio under the same scenarios.
Risk Factor	A factor upon which the value of the predicted variable depends.
Risk Space	The risk space is the n dimensional space where each of the n dimensions corresponds to the range of risk factor of the heavy model. Loss functions map points in this space (collections of risk factor realisations) to values or changes in value of the predicted variable (losses). Sometimes the term risk space is used to refer to the n+1 dimensional space in which the first n dimensions are risk factors and the last dimension is losses.
RMSE	Root mean squared error – the square root of the average of squared errors. It is a measure of the differences between values predicted by a model, and the values observed.
Roll-Forward Process	The process by which the proxy model is altered in order to be fit for purpose at a date after the initial calibration, without full re-calibration. For example, to account for actual experience or market movements which have occurred between calibration and the reporting date.
Single stress / univariate point	A scenario of risk factor realisations where all risk factors are held constant at their base value except for one risk factor, which is varied i.e. a point along an axis in the risk factor space.
True-up / true-down	A true-up (true-down) is a positive (negative) adjustment to the proxy model calculated SCR, for example where the adjustment is equal to the difference between the proxy and heavy model at the biting scenario.

Appendix B: Survey questions

Area	No.	Question
Purpose and scope of proxy model	1	Please explain how your **proxy model** is used as part of determining your SCR, and outline how **out-of-model adjustments** affect this calculation. For what purposes do you use your proxy model other than determining your SCR?
	2	What is the scope of your proxy model in terms of Major Business Units, Material Lines of Business or any other relevant way of classifying exposures? Where a Major Business Unit or Material Line of Business is excluded from your proxy model, please explain why.
Fitting methodology	3	What is your overall methodology for proxy modelling? Specifically: i) At what granularity are proxy functions calibrated? By granularity we mean combinations of Major Business Unit, Material Line of Business, and any other relevant way of classifying exposures within your model's scope, and whether assets and liabilities are modelled separately or whether basic own funds are modelled as a whole. **For the remainder of this survey, where relevant, please give details for each exposure type where the approach varies.** ii) Do you model changes in the balance sheet directly, or the balance sheet itself and subtract the opening position at the end to give changes over a 1-year horizon? iii) What is the form of the proxy functions and how are those forms determined? Please include a description of cross-terms. iv) What method do you use to fit the proxy functions? v) How do you assess **goodness of fit** when fitting proxy functions and how do you strike a balance between **goodness of fit** and **over-fitting**?
	4	How many **fitting points** do you use to fit the **proxy functions** and how are these selected? In particular, please outline: i) How you select the **risk factors** to be modelled; ii) How you determine that the number of points is sufficient in relation to number of risk factors modelled; iii) How you determine that the points adequately cover the **risk space** being modelled in the context of determining the SCR and any other purpose for which your proxy model is used; iv) What each fitting point represents; and v) Whether you allow for **management actions** within the fitting process, and if so, how.
Validation	5	What methods do you use to validate the fit of your proxy functions and the SCR that results from applying the proxy functions? Please distinguish between testing which is carried out **on-cycle** vs **off-cycle**. How often are your proxy model process and results reviewed by your internal/external auditors?

Area	No.	Question
	6	Where you use **'out-of-sample' points**, how many do you use and how are these selected? Please outline: i) Who selects these points and the methods used to perform the selection; ii) How any conflicts of interest or lack of independence between staff performing fitting and staff performing validation are identified and managed; iii) How you determine that the points adequately cover the risk space being modelled; and iv) What percentiles of the overall loss distribution the points relate to, if relevant.
	7	What are the **acceptance criteria** for your validation tests? Please include details of: i) How these criteria are set and why they are appropriate; ii) What actions you undertake in the event that **validation errors** breach acceptance criteria; and iii) How these results feed into the next **calibration cycle**.
Timeline & roll-forward	8	How much effort (in person-days) and how long (in elapsed time) does it take to fit and validate the proxy model? Please break your answer down into key stages, as a minimum fitting and validation. At what points in your reporting cycle do you conduct re-calibration and validation?
	9	If relevant, please outline the **roll-forward process** that you undertake. Please include details of the validation you carry out after the roll-forward.
Miscellaneous	10	What are the major weaknesses and limitations in your proxy model and what future improvement work is planned? How has your Board been made aware of these limitations? How have you ensured that decision makers understand the range of uncertainty around the SCR?
	11	Please outline any external resources you have found helpful regarding proxy modelling, for example conference talks, industry surveys, IFoA working party materials, etc.
	12	What do you consider to be 'best practice' in relation to proxy model fitting and validation? Is there any other information you can provide on your proxy model process which you consider to be relevant?