Credit Research Centre

CRC Working Papers

Unavoidable Model Risk in Expected Credit Loss models under IFRS9 and CECL

By Dr Alan Forrest

September 2024

Keywords: Model Risk, Expected Credit Loss, Banking Regulation, IFRS9 standards, CECL standards, Provisions Accounting, Reserve Accounting.

THE UNIVERSITY of EDINBURGH

Contact
Credit Research Centre
University of Edinburgh Business School
29 Buccleuch Place
Edinburgh, EH8 9JS
credit.research.centre@ed.ac.uk

Unavoidable Model Risk in Expected Credit Loss models under IFRS9 and CECL

Alan Forrest[a]

*Corresponding Author. Email address: v1aforr2@ed.ac.uk

Abstract

This note examines how model risk emerges in Expected Credit Loss (ECL) calculation under IFRS9 and CECL standards, focussing on those risks that are special to ECL models and to unbiased point prediction. It finds that some material model risks are unavoidably created or exacerbated by the standards and guidance, and it describes practical approaches to find, assess and mitigate these risks.

[a] Business Associate at the Credit Research Centre, University of Edinburgh Business School

Introduction

Any regulated implementation and use of models brings model risk, most directly from the models themselves but also paradoxically from the regulatory structure that constrains them[b]. This is because regulation and best-practice, however accommodating and enlightened, do restrict the kinds of models allowed and the ways in which they are deployed, and so constrain in principle the ways in which model risk is managed. Good regulation and practice allow sufficient freedom for good model risk management and minimise these problems; poor regulation does the opposite.

It is therefore instructive to find where IFRS9 and CECL modelling standards[c] do impose negative constraints on model risk management, implicitly or explicitly, and therefore where the model developer, validator and user need to direct special attention to control the model risk.

This paper finds three areas of regulatory constraint for Expected Credit Loss (ECL) calculation that cause increased strategic model risk:

1. The requirement of unbiased central estimates in computing ECL;
2. The requirement of predictive accuracy managed by model adjustment and a short feedback correction loop; and
3. The required network of ECL models and its structure, especially SICR.

This is not intended as a negative criticism of the standards and their accounting aims; rather it is a recognition that those aims and the resulting constraints can compromise modelling practice unintentionally. The positive message of this note is its guidance how to manage the model risk under these regulatory constraints.

On a longer view, however, this note finds that IFRS9 and CECL increase model risk largely by a misunderstanding of the nature of predictive modelling. The ECL calculations treat model outputs as numbers and not as distributions, and most of the risks and difficulties detected in this note flow from this mistake. As a statistician and model risk professional, the author urges that the next iteration of Provisions / Reserve accounting standards is improved by a better understanding of predictive models.

This note first describes precisely the function of statistical models and how the approach to statistical modelling changes as the signal to noise ratio changes: it shows that regulated ECL modelling is intrinsically noisy. From this foundation, the note describes each of the three areas of heightened model risk above, together with model risk management approaches and suggested questions.

---

[b] Regulation is a positive force, of course, ensuring good model risk management by setting suitable standards, validation function etc. Rather, this note is about the constraint of technical options available to the model developer and validator. Experience shows that such constraints can be positive, neutral or negative in their effect.

[c] To clarify: here and throughout the paper "standards" mean the wide framework of ECL regulation, modelling standards and guidance centred on IFRS9 and CECL, including annexes and updates, additional guidance by BCBS [GAECL] and PRA, EBA and other local regulators, as well as internal bank standards and industry best-practice upheld by external Auditors who are the primary arbiter of compliance in setting Provisions / Reserves. This body of practice is not simple, uniform or fixed, but the modelling principles and guidance assumed in this paper are by now (2024) stable accepted best-practice – see [Bellini]. The author's experience is UK-based but, with one exception (SICR in IFRS9), the note is applicable to CECL and IFRS9 in all jurisdictions.

The references given are minimal for this discussion and this note makes no attempt to describe fully the current thinking on these standards. This note is therefore accessible to a reader who has a basic grounding in IFRS9 or CECL and is motivated to manage models better under these standards. For such a grounding see for example [Bellini]. A Glossary clarifies the many acronyms.

Models, distributions, predictions and bias

The IFRS9 and CECL standards require ECL estimates to be unbiased, predictively accurate, and forward looking (See [GAECL] Principle 5 passim); and so the models used in this estimation must also have these qualities, or at least be adjustable to this standard. This seems a natural requirement for the models' use in estimating ECL in Provisioning or Reserve, and seems obviously to be what quantitative modelling is designed to do. Consequently, banks have responded to IFRS9 and CECL standards with data-driven modelling and sophisticated forecasting, supported by teams of skilled quants and complex ECL modelling systems.

Given this heavy investment in statistical sophistication, it may come as a surprise to non-specialists that unbiased predictive accuracy is not the primary aim of statistical modelling. Rather, statistical models are built to describe and control error distribution, which helps greatly with point estimation of course but is not the same objective. As we will see below, this subtle distinction has major implications for how we should use statistical models to meet IFRS9 and CECL requirements. This shift of understanding may not be easy for those who view models simply as prediction algorithms, but is necessary if the models are to be interpreted and used correctly.

Statistical Models

Statistical models predict uncertain events under given conditions and describe the uncertainty in terms of probability distributions. Therefore the output of a statistical model is a distribution, not a number (see [Cox & Hinkley] for instance). This makes it fundamentally different from a mechanics model or an accounting calculation.

For practical applications, we often use only single number summaries of these distributions – mean or median for a central estimate, and range or variance for its uncertainty etc. - but the model's output distribution always sits behind these summaries. For example, the mean value of the Loss Given Default (LGD) distribution is used as the LGD value in Credit Risk management, but also downturn LGD values are used which are points on the tail of an LGD distribution. More sophisticated risk models assess the distribution of losses through Value at Risk and other statistics of loss distributions. Even Probability of Default (PD) comes from a distribution – a binary Binomial distribution that happens to be defined completely by a single number, the estimated PD.

The fact that a model output is a distribution has obvious implications for model monitoring and for interpreting the model's output or the explanation of the modelled phenomenon. It also has implications for how the model is designed and optimised. The way a model's output distribution is built up – the error structure and error distribution - is the key to quantifying how well the observed data fit the proposed model. Therefore the choice of error structure is critical to the design of the model and to the calibration of the parameters, and can influence all features of the model, including expected outputs and point estimates. Experience and theory show that the choice of error structure is the principal influence on bias in an otherwise well-optimised model.

Model Error structure and noise

So, the model's specified error structure dictates how that model is designed, built and tested.

Some real-life phenomena (e.g. well controlled industrial quality testing) allow a highly predictive model to be built, with easily measured input features, combined to the right degree of complexity with narrow error distributions or tolerances. Here the signal-to-noise ratio is high.

But not every observed phenomenon can be measured so cleanly: a model of an essentially random phenomenon (e.g. radioactive decay of a few atoms) is simple but has error distributions that dominate the outcome – it is almost all noise, by definition. Other human phenomena (e.g. health outcomes or loan defaults) have so many social and other unknown influences that their predictive models also have necessarily low signal to noise, and need sophisticated statistics to disentangle the two.

Modelling under noisy conditions is not a simple variation on high-signal modelling. The presence of noise above a small degree triggers a fully statistical approach to modelling where the choice of error structure becomes an influential part of the model build. Thus the statistical understanding and quantification of error moves centre stage.

Credit Risk Model Error and IFRS9/CECL

Credit Risk is a phenomenon which bridges the signal to noise divide. At a large scale, with good data and economic understanding, the behaviour of a nation's business sector or a bank's portfolio of credit exposures might be realistically predicted by high-signal modelling, on average under close-to-historic market conditions.

It is this stability and predictability that allows Credit securitisation and portfolio valuation to operate within limits – limits, for instance, that were breached during the exogenous and systemic disruption of the Global Financial Crisis (GFC). This predictability also allows General Provisions or Reserves to be assessed reliably across a bank's lending book - and again the limits of this process were broken during the GFC.

IFRS9 and CECL are designed in part to address this GFC breakdown of simple estimates, and they achieve this in many ways including new modelling principles and requirements[d], adding to the intricacy, scope and depth of ECL modelling. First, to account for exogenous variance, ECL must take account of how the models perform in realistic non-stress forward-looking scenarios. Second, as noted above, predictive forward-looking unbiased accuracy should always be sought. Third, this accuracy should be assured at a granular level, aware that some segments of exposures can be specially disrupted by particular credit risks or exogeneity. Finally the credit behaviour may need to be predicted over the contracted duration, a period that may be long (consumer mortgages) or indefinite (revolving exposures). IFRS9 (not CECL) also adds a distinction between Stage 1 and Stage 2 ECL and specifies flags (SICR), which may be partly modelled, to govern movement of exposures between the Stages.

[d] This note considers only the modelling of individual exposure ECL=PDxEADxLGD by statistical models. This is the approach taken in larger banks for all consumer lending exposures and most corporate lending, so is by far the greater part of bank business. However we note that IFRS9 and CECL allow other approaches to ECL estimation, at portfolio level or by other benchmarks, whenever supporting data are not sufficiently strong.

Even listing the features here shows that the ECL context has moved the modelling well to the noisy end of the scale.

- Forward-looking views and scenarios are a clear source of judgemental uncertainty and approximation, speculation and assumption about their effects on hypothetically reacting portfolios;
- the finer granularity automatically raises sampling- and standard-error and a greater reliance on topical data and detail, more likely to be incomplete with short history and poor quality;
- the model has large prediction error when it projects over a full term of contract: an uncertain and long period of market, social and operational changes; and
- overall the intricate model interdependency brings structural non-linearity and possible cliff-effects that can be anticipated only by an accurate appreciation of the output distributions and not just the central values.

To conclude, ECL modelling under the refined conditions of IFRS9 and CECL is unavoidably noisy in the technical sense described above. Therefore a strong statistical appreciation of model prediction error and systems of models is critical to achieving the ECL goals of predictive accuracy and bias control.

The next sections describe three places where the IFRS9 and CECL standards themselves have made it difficult to achieve and exploit such a statistical understanding of models. This circumstance in turn generates model risk and each section describes these risks and how they might be managed.

Model Risk area 1: Central estimates rather than distributions

The previous section made clear how noisy phenomena, such as ECL, need a statistical, distributional approach to modelling. Presenting error structures and distributions correctly helps model management, giving:

- a model specification in which model error, complexity, signal and noise have been weighed up correctly, allowing better accuracy, understanding, monitoring and greater robustness over time; and
- a mature modelling culture of error awareness, challenge and repair, leading to prudent model use and good communication.

Unfortunately, with a focus on single number unbiased outputs, ECL modelling is in danger of missing both these points. This may seem an abstract concern but it affects directly the practical management of ECL model risk.

For instance, the requirement for unbiased output in ECL modelling disables one of the most useful and powerful model risk management tools – conservatism.

In Basel 2 Internal Ratings Based (IRB) Credit Risk for capital setting, a modeller can (indeed must) use conservatism to quantify and manage actively the weaknesses and assumptions in the model. This is done by applying a conservatism adjustment to cover (i.e. over-capitalise) the model uncertainties identified by the weaknesses and assumptions; and as those weaknesses are repaired and assumptions made more definite, so the conservatism adjustment is managed down. The financial pressure to lower capital motivates the bank to repair model weaknesses, and sets up a positive conversation between validator and modeller that can be escalated to senior managers as a quantified trade-off between capital and model quality. This creates a useful tension in IRB between regulated model output and bank capital setting, which is healthy for model risk management and for the financial system.

This process is replaced in IFRS9 / CECL by model adjustment (MA) to correct recognised quantified biases or model inaccuracies [GAECL 61b]. Like IRB conservatism, ECL standards expect MAs to be managed down as a benefit of improved modelling [PRA Annex 3], but there are two important differences between MA and conservatism:

- There is not always a financial benefit to the bank in making MAs smaller, because MAs, like bias, can be positive or negative;
- Bias is not about the model's output (as a distribution) but about the central estimate of that output. This subtle point implies that MAs, which correct bias, are indifferent to the width of the model error distribution. But model risk *is* concerned with the error width, which represents model uncertainty and which conservatism covers robustly.

Together these are enough to make MA materially different from conservatism: ECL model adjustment is a model management action that incompletely addresses the model weaknesses and model risk, and which has ambiguous motivation for the bank as whole.

Pointers to manage Model Risk area 1 – central estimates

The first goal of the validator and model risk manager is to make sure the whole of model risk is managed by the model validation and model management actions. They should not let the ECL focus on bias and correction cause material parts of model risk to be ignored or treated more lightly.

As far as is consistent with ECL principles, the conservatism philosophy of IRB Credit modelling should be adopted in model adjustments: each model weakness should be brought to bear in the model adjustment process, ideally as a specific adjustment. Even if a specific weakness does not lead to an adjustment, and it remains unaddressed, it may still be used to set a pre-condition for removing MAs.

<u>Questions to ask:</u>

1. What is the model prediction error distribution? How is it estimated and chosen?
2. What is the model's error structure? What parts are considered uncorrelated, or implicitly without error? Do other error structures cause variation in model fit and output bias?
3. Is each model adjustment the correct, compelling response to a declared and well-understood model weakness?
4. After considering the declared adjustments, are there further model risks and model weaknesses that need to be managed, and can they also be tracked and managed in the ECL approval and adjustment process? If not, then how will they be managed?
5. What is the management plan for making the adjustments smaller and for removing the model weaknesses? How does adjustment reduction demonstrate benefit for the bank even if financials are affected adversely?

Model Risk area 2: Predictive accuracy and the feedback correction loop

A model is correctly specified if its prediction error matches the observed error in its outcomes, even if that error is large. Abstractly this is a basic law of information theory or thermodynamics, and not a question of technique or regulatory expectation: a model calibrated within a noisy environment is simply incorrect if it underestimates the noise and claims too much precision.

This is true in practice too: an over-precise claim makes it difficult to communicate the credibility and usefulness of an estimated model output. Further such over-precision impairs the setting of monitoring thresholds and launches inappropriate model adjustment and management actions.

Therefore to achieve the most accurate and unbiased predictive models, noise should be presented to the correct degree as prediction error and should not be underestimated nor modelled as if it were signal.

In contrast, the standards' focus on unbiased point ECL estimates carries the danger that prediction error is seen as a defect of the model, to be removed somehow by better segmentation, modelling or technique. However, such removal is simply impossible; it is as firm as the 3rd law of thermodynamics. Granted, fresh data or new data sources can narrow prediction error distributions, but such incremental improvement has its limits and diminishing returns. Therefore, for hard scientific reasons, any data-driven model adjustment process will fail to make material improvements in prediction accuracy, unless it includes a corresponding contribution of fresh informative data.

In its focus on accuracy, the ECL guidance appears to miss this point but instead relies on model intricacy and a rapid monitoring, control and recalibration cycle, not just for management of the model lifecycle (which is good practice), but to seek greater accuracy for the ECL model outputs by continual model adjustment [GAECL, 60b]. In practical implementations banks apply this adjustment at least quarterly.

Certainly, there are benefits from a process that uses recent data dynamically to granulate and update the model, but they are unlikely to lead to increased accuracy (see above). In addition, such adjustment needs care, technique and validation to avoid causing more problems than it solves – it is in itself a new model risk.

Model adjustment methodology

- is an algorithm that requires its own validation and assessment for self-stability, accuracy etc;
- requires skilled resource to build and maintain;
- is backward looking, however recent the data, so needs continued expert intervention to remain forward-looking; and
- will not, in the end, narrow the prediction error (see above); in fact naive feedback loops will almost certainly cause more volatility and error.

To illustrate these points and to make the link with noise and model error more clear, imagine the outcomes of an experiment that is all noise – the roll of a 6-sided die. We've predicted correctly that 3.5 is the unbiased central value of the outcome. At the first roll we see "2" come up, so we adjust the model to predict 2 (or 2.75?), but then "5" comes up so we adjust our next prediction cautiously to 4 or more aggressively to 6 because there's clearly an upward trend and we must look forward… and so on. Clearly this is an inappropriate adjustment process, but it can be put right only when we understand the proportion of signal to noise.

For ECL likewise we are in danger of imposing a similarly defective adjustment process unless we get the prediction error distributions right.

Pointers to manage Model Risk area 2 – model correction feedback

Here the key concern is the correctness and robustness of the model adjustment feedback loop. Such a process can appear stable and deterministic on the drawing board, but the effects on the models it adjusts can be unexpected and stochastic: they need measurement and control. Especially we need to understand the proportion of signal to noise. Some instabilities could be the unwitting artefacts of the feedback process itself, so we need to understand that as a dynamic process as well. To control properly unwanted volatility and an explosion of errors requires technical sophistication or intense management (e.g. Kalman Filtering or manual interventions), with skills to match in the modelling and validation teams.

Questions to ask:

1. Do the model developers understand and quantify the amount of prediction error and model error? Has that analysis informed their design of the model adjustment / model correction feedback loop?
2. Do we know the proportion of signal to noise in the models and is that correct?
3. Does documentation cover in sufficient detail how feedback, monitoring and corrections are processed and controlled? Is it reproducible, tested and auditable?
4. Is this feedback process theoretically stable? And is it demonstrated stable in practice –with testing and performance or response characteristics assessed?
5. What conditions would break the feedback and correction process? What are the stop and escape procedures?

Model Risk area 3: The network of models and its structural weaknesses

Even at its most basic, ECL is a calculation that involves, say for retail mortgages, 10's or even 100's of data and model elements in a network of many layers and connections. Many firms have repurposed much of the (already complex) Basel IRB Credit suite to provide base PD, EAD and LGD risk parameters, but that is just the start. From here PD must be adjusted into Lifetime PD, supported by a model for the length of that lifetime; PD, EAD and LGD must also be modified to take adjustments from scenarios; the scenarios themselves need to be built up from time series data and expert judgement and then weighted; for IFRS9, SICR needs to be defined and assessed; on all of these elements, model adjustments could be applied, to correct for known biases; finally all model outputs are combined into a final number, approved by experts and managers at all levels, and embedded in a robust auditable financial account system.

System Complexity

The diagram that follows is a (simplified) graph of how these models and components are related in an IFRS9 ECL calculation. Each dot represents a data source, model or aggregating calculation, arranged in 8 layers and multiply connected. In the top left corner are the classic risk parameter models PD, EAD, LGD; in the bottom left are the econometric scenario models. These mix in the middle to inform lifetime PD, scenario-adjusted PD, EAD and LGD, and the SICR decision. They are then gathered in the right-most 2 layers as scenario-adjusted ECLs and weighted and combined to give the overall ECL.

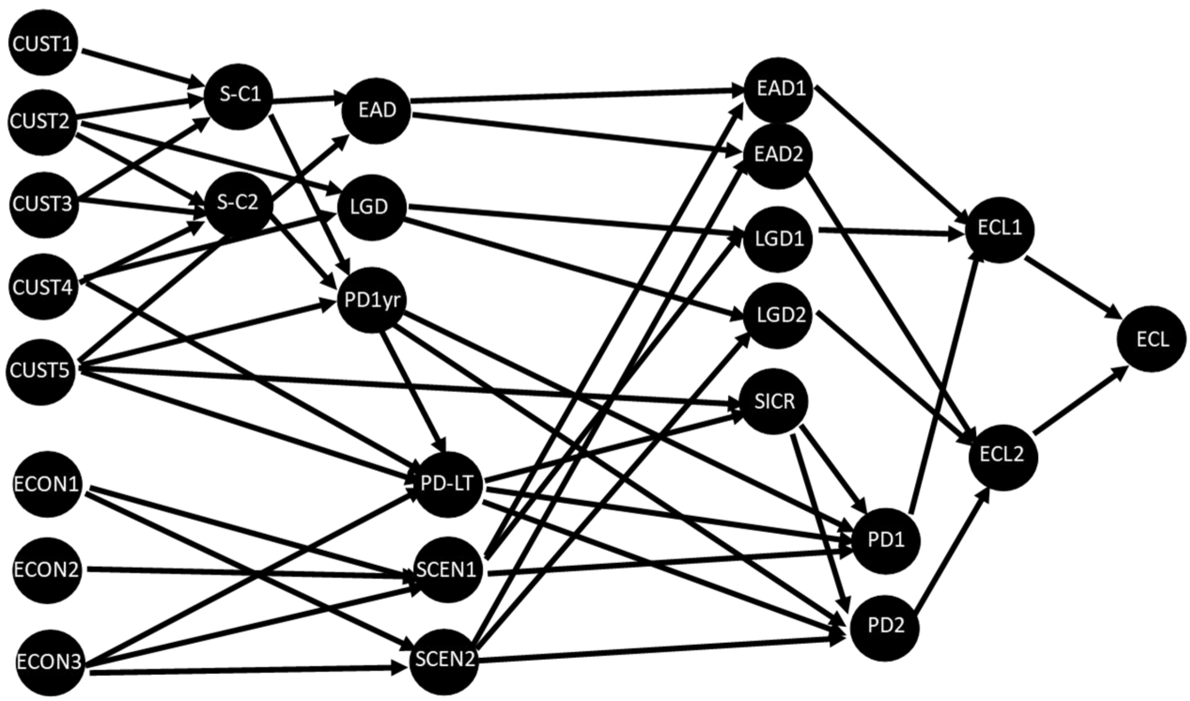

Figure 1: A schematic diagram of information flow among data sources and models in an ECL calculation for IFRS9

The complexity and size of this system is in itself a model risk, and model risk managers are alert to the instabilities added by using highly interconnected models. This point underlies validators' concern with multilayer neural nets, for example. In that case, good model risk practice seeks simplification and analysis of the system as a whole to guard against unexpected consequences and instabilities. This concern about neural nets applies also to this ECL calculation network, with two variations:

- In ECL the network complexity cannot be simplified - it is embedded in the standards.
- The model risks of the ECL network are mitigated somewhat by the active management and well-understood meaning of each node. All the data, models and calculations are assessed by experts in a process of constant inspection and impact analysis, and so the intense governance and manual oversight compensates for the complexity to some degree.

Uncertainties about SICR specification

Another example of structural model risk is found in the IFRS9 standards about Stage 1 and Stage 2 and significant increase in credit risk (SICR). Each exposure originates in Stage 1, but may subsequently move to Stage 2 under SICR, a decision rule that determines when the credit risk of an exposure has become significantly worsened since origination. It is reasonable to expect that default rates and PDs will generally be greater in Stage 2 exposures than in Stage 1, and the risk assessment is correspondingly more cautious and wide-ranging (Lifetime) in Stage 2. Moreover, as a precaution, the recovery from Stage 2 to Stage 1 is put on Probation for a period before that recovery move is made. For a good overview of SICR and quantitative approaches to setting it, see [Cleary & Macnee].

IFRS9 guidance is aware of the obvious specification risk in the setting of the SICR decision rule, and gives helpful guidelines for what is acceptable here, but it still leaves a wide choice which banks can interpret or game arbitrarily. It is particularly difficult to quantify the Model Risk generated by this because SICR is a model processing step not a predictive model - it simply directs exposures to models in Stage 1 or in Stage 2. Therefore the model is not "predicting" and many metrics of model performance and validation must be modified accordingly. Further the conditions that move exposures from Stage 1 to 2 are complex and open to interpretation. As a consequence, SICR model validation allows wide variation, with corresponding variation in Stage 2 volume.

The following chart from a recent Deloitte report [Tedder] shows wide variance in UK banks' interpretation of Stage 2; far in excess of the (relative) variance seen among the default rates (Stage 3) of UK banks.

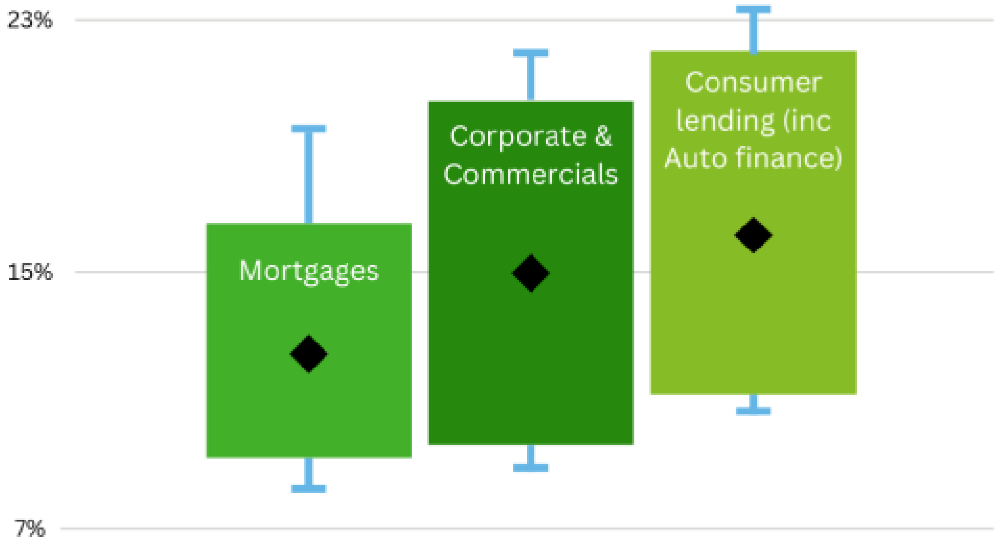

Chart 1 Proportion of loans in Stage 2 for the largest UK banks at 30 Dec 2023 (range and mean). Source: [Tedder]

Taking this range of interpretation to quantify roughly the model specification risk arising from SICR, we find that the rate of Stage 1 to Stage 2 migration is specified to within roughly a factor of 2, from which each bank might deduce a resulting variance on ECL caused by the systemic specification risk of SICR.

Simpson’s Paradox

A further model risk on PD emerges from the SICR structure itself, caused precisely by SICR’s correlation with the default event. It is an example of Simpson’s Paradox [Simpson], a rich source of counterintuitive statistical behaviour.

Table 1 illustrates changes in the staging and behaviour of a hypothetical portfolio on 3 consecutive years. In year 2 the default rates (DR) drop in each Stage, yet the overall portfolio default rate stays the same. In year 3, the default rates return to year 1 levels in each Stage, but the portfolio default rate exceeds the year 1 value. This seems paradoxical, but is explained by the migration of high risk exposures in Stage 1 to become low risk exposures of Stage 2.

Note further that probation periods only prolong the effects of Simpson’s Paradox. In the example, the year 1 to year 2 paradox is caused by a migration of 500 customers to Stage 2 in year 2; the year 1 to year 3 paradox is caused by worsening conditions and the prevention of migration back to Stage 1 in year 3.

Year 1	Stage 1	Stage 2	Total
Total	1500	1000	2500
Defaults	15	50	65
DR	1.00%	5.00%	2.60%

Year 2	Stage 1	Stage 2	Total
Total	1000	1500	2500
Defaults	5	60	65
DR	0.50%	4.00%	2.60%

Year 3	Stage 1	Stage 2	Total
Total	1000	1500	2500
Defaults	10	75	85
DR	1.00%	5.00%	3.40%

Table 1a-1c : Illustration of Simpson's Paradox for a hypothetical portfolio over 3 years. Default Rates (DR) by Stage and by Total show conflicting measures of movement.

From this example it is clear that, depending whether we measure by the whole portfolio or by each Stage separately, we see a conflicting picture of default behaviour. This in turn would be reflected in conflicting monitoring outcomes for PD and for PD bias. Further, this is not a small difference, but it could influence directly the monitoring and management of the model: a model could be adjusted or even rejected on one level of granularity, and unchanged on the other.

The paradox is caused by the structural correlation between Staging and PD, meaning that as the SICR process gets better at its function, the model risk of the PD model increases. This paradox – improving one part of the model causes an increase in model risk – is unavoidable in the IFRS9 SICR structure.

Note that other models such as EAD and LGD are also open to this effect, particularly if they have a different specification in each Stage.

Pointers to manage Model Risk area 3 – model networks

For networks the key concerns are the instability and unintended consequences of concatenated models and the systemic amplification of errors. This is not an easy technical area and a full analysis is a sophisticated and demanding effort. On the assumption that the skills are not readily available in modelling and validation teams, more local and approximate explorations might be all that can be achieved.

If one model's output (upstream) is an input for the next model (downstream), we can assess directly the influence by applying credible upstream variations and seeing how these affect the downstream model. This leads naturally to sensitivity analysis and a systematic investigation of how a small perturbation in an input flows through to distant model outputs. Such testing in turn needs a sophisticated modelling platform and an investigation team with resource, time and skill, but it is at least simple and scalable.

The assessment of SICR's impact needs sensitivity analysis or validation by simulation, as it is a procedure not a predictive model. Its connection with Simpson's Paradox noted above, means that metrics to measure SICR effectiveness and PD performance, must be calculated at a consistent and fixed level of granularity. This should be maintained even when the alternative granularity shows a conflicting picture.

<u>Questions to ask</u>:

1. Is an assessment made of the whole system of ECL calculation, in addition to the validation of each individual model?
2. Are individual models assessed routinely for sensitivity to input variation, and is there an appreciation of any combinations of inputs that would break or make invalid the model's output?
3. Is the sensitivity analysis of individual models joined up to explore sensitivity across (wide sections of) the ECL calculation network? What are the hot spots in the system, where strong volatility or sensitivity builds up most easily or destructively?
4. How robust and effective is SICR under simulation or sensitivity analysis? What does it take to cause SICR to produce results that fail to make business sense? How does the SICR outcome compare is to peer experience, as seen e.g. through Company Reporting?
5. What level of granularity is used to monitor model performance across the portfolio – by Stages or all together? Is this used consistently across all monitoring and correction feedback loops?
6. How big an effect has Simpson's Paradox on the PD, EAD and LGD model outcomes? This itself could be an informative metric of interaction between upstream and downstream models.

Conclusion

The IFRS9 / CECL standards for ECL calculation include assumptions, principles and structure that cause or exacerbate model risks as noted above. These risks could cause significant model errors that require intense model adjustment to manage. The control of these risks can be improved, first by being alert to their causes and effects, then by following a suitably focussed validation process to challenge and ask appropriate questions, as suggested in this note. Such a model validation and challenge process applied early and often, allows the best chance of improvement in strategic model risk management of ECL models, and should lead to more quantifiable and better managed model adjustments.

References

[Bellini] Bellini, T. (2019). *IFRS 9 and CECL Credit Risk Modelling and Validation: A Practical Guide with Examples Worked in R and SAS*. Academic Press.

[Cleary & Macnee] "IFRS 9: Significant Increase in Credit Risk" Katie Cleary and Kathleen Macnee. Presentation to Business School at University of Edinburgh, Oct 2020. _https://www.crc.business-school.ed.ac.uk/sites/crc/files/2020-10/D27-IFRS-9-Significant-Increase-in-Credit-Risk-Cleary.pdf

[Cox & Hinkley] Cox, D.R. and Hinkley, D.V. (1974) Theoretical statistics. Chapman and Hall, London.

[GAECL] " Guidance on credit risk and accounting for expected credit losses" Basel Committee on Banking Supervision, December 2015.

[PRA] "Letter sent to Chief Financial Officers of selected deposit-takers" Victoria Saporta, Executive Director, Prudential Policy, Prudential Regulation Authority, 15 April 2019 https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/letter/2019/written-auditor-reporting-update-and-main-thematic-findings.pdf?la=en&hash=9A6CBCFD97212529151EBCDBA66AAE6D8DB1259C

[Simpson] Simpson, Edward H. (1951). "The Interpretation of Interaction in Contingency Tables". *Journal of the Royal Statistical Society, Series B*. **13**: 238–241.

[Tedder] Richard Tedder 2024 "The IFRS 9 Staging dilemma. One requirement, different interpretations, multiple solutions." Deloitte publication. https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-ifrs-9-staging.pdf

Glossary

IFRS9	International Financial Reporting Standard (IFRS) issued by the International Accounting Standards Board (IASB), effective in non-US banks from 2018.
CECL	Current Expected Credit Losses - an accounting standard issued by the Financial Accounting Standards Board (FASB) – published 2016, effective in US banks
PD	Probability of Default
EAD	Exposure at Default
LGD	Loss given Default
ECL	Expected Credit Loss- broadly PD x EAD x LGD
SICR	Significant Increase in Credit Risk – a formal status attached to an exposure, which then undergoes an enhanced, lifetime ECL modelling approach. This is assessed in IFRS9, but in CECL all exposures are lifetime ECL so SICR is not assessed.
Stage 1	Exposures that are not in default and have not flagged SICR
Stage 2	Exposures that have flagged SICR but are not in default
Stage 3	Exposures in Default
GFC	Global Financial Crisis
IRB	Internal Ratings Based approach to setting Credit Risk Capital under the Basel 2 directive
MA	Model adjustment - which can be a recalibration of the model itself or a change to the model output (overlay or post-model adjustment)