7

Model Risk Management in Credit Risk Models

Tasneem Chherawala

7.1. Introduction

Over the last few decades, major advances have been made in methodologies for assessment of credit risk. Depending upon the complexity and materiality of credit portfolios, banks and financial institutions (FIs) have implemented a variety of models to quantify multiple dimensions of credit risk. Banks require model-driven internal estimates of regulatory capital under the advanced approaches of the Basel framework and also to determine Expected Credit Loss (ECL) based provisions under the IFRS accounting framework. Internal models allow banks to align risk buffers more closely to their own exposures and portfolios, thereby reducing the potential for capital arbitrage that is inherent in the regulator prescribed risk estimation formulae. Furthermore, credit risk models can be adapted to changes in the macro-economic environment for reliable stress testing under the Internal Capital Adequacy Assessment Process (ICAAP) of the Basel framework and estimation of forward-looking ECL.

More importantly, banks have started incorporating the outputs of internal risk models in ever-widening tactical and strategic decision-making processes, including, but not limited to, loan origination, risk-based pricing, customer and portfolio level risk-adjusted performance measurement, economic capital allocation to business lines, credit limit setting and active portfolio management.

The more recent past has seen significant enhancement of risk analytics capabilities at banks. This has come about due to the digitization of lending business; availability of large volumes of transactional and behavioral customer data from internal and external sources; and, through deployment of advanced Artificial Intelligence (AI) and Machine Learning (ML) techniques. Banks have started integrating risk-based algorithms into automated lending processes for near-real-time decision-making in terms of underwriting, setting pre-approved credit limits and loan pricing. A recent survey conducted by SAS and GARP (2019) found that, over the next three to five years, banks expect to significantly increase adoption of AI & ML models to support key risks.

However, all risk estimation models are themselves a potential source of risk. As defined by the Federal Reserve, (2011), model risk is '*….the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports*'. One type of model risk arises because of the assumptions, data flexibility and methodological choices in model construction, which if untested, can create inaccurate or undesired outcomes. Such flaws in Basel-compliant credit risk models can generate underestimates of capital and reserve requirements. When applied to management decision-making, unauthenticated models can lead to underpricing of credit risk, mis-selection of credit customers and creation of unintended risky portfolios. Model risk may also result from misuse of credit risk models. This may either be because the model design is intentionally manipulated to underplay estimated risks in order to boost short-term business growth and profitability; or because complex models are "black boxes" for management and their use in decision-making is therefore inadvertently compromised.

Thus, while quantitative methods will play an increasingly important role in the future of credit risk management, they must also be approached with caution. The Basel Internal Ratings Based (IRB) Approach prescribes minimum standards for internal credit risk models used for regulatory capital computation. Despite this, there have been instances of model exploitation by some global banks that have led to reputation loss and imposition of regulatory penalties. As a consequence, model risk management (MRM) has become a key concern for banking supervisors the world over. Regulators in US, Europe and UK have established principles for model governance that have to be complied with. Beyond compliance, the top management of banks also has to have internal assurance of the reliability of such models in on-going credit risk assessment.

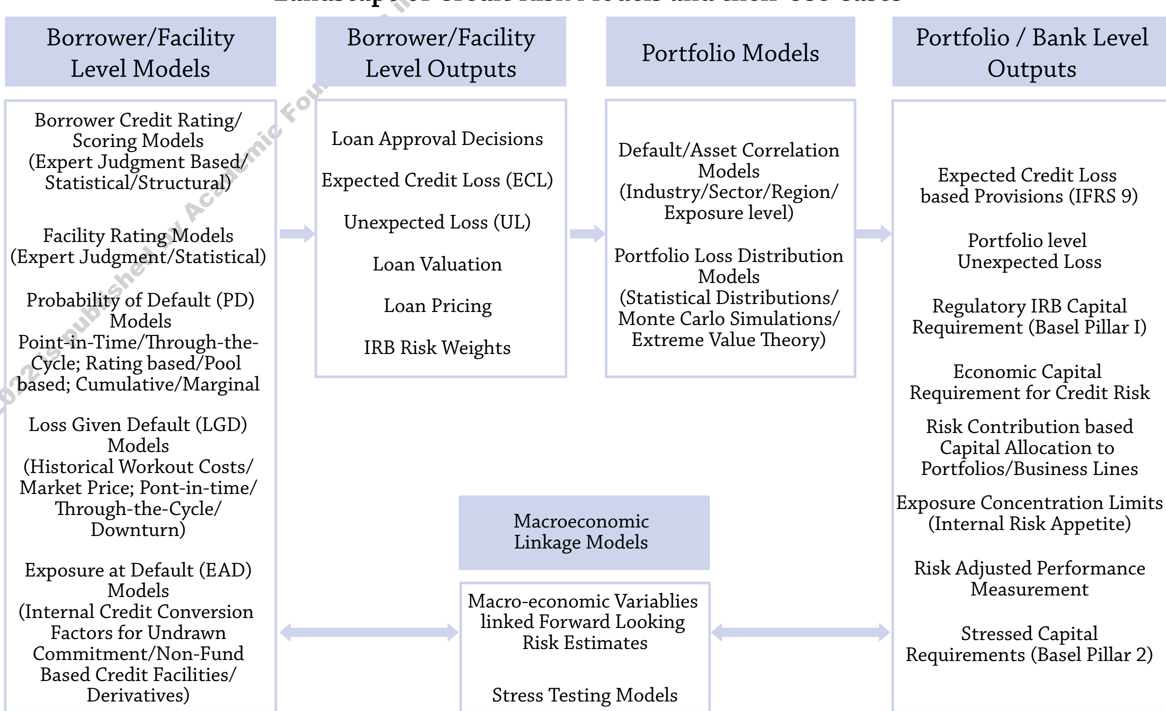
In light of the above concerns, this chapter aims to comprehensively describe the evolving MRM framework and highlight its relevance in the context of credit risk models used by financial institutions. The rest of the chapter is organized as follows. The second section of this chapter focusses on the range of credit risk models that are currently in-use for various purposes. The third section summarizes the regulatory expectations and best practices in MRM. In the fourth section, the qualitative and quantitative aspects of credit risk model validation are described. The fifth section indicates key challenges associated with the MRM discipline in the context of credit risk. The final section concludes with implications for Indian banks and supervisors.

7.2. Landscape of Credit Risk Models

Credit risk is the estimated potential loss that is associated with adverse credit events like standalone and correlated obligor defaults or

FIGURE 7.1

Landscape of Credit Risk Models and their Use Cases



Source: Author's own construction.

downgrades. For banks and FIs, this is a key business risk that needs to be measured, monitored and managed on a continuous basis. Estimates of credit losses can be constructed from models for key credit risk drivers – Probability of Default (PD), Loss Given Default (LGD), Exposure at Default (EAD) and default or asset correlations. In this context, the term 'model' can be defined as *'a quantitative method, system or approach that applies statistical, economic, financial, or mathematical theories, techniques and assumptions to process input data into quantitative estimates. …It also covers quantitative approaches whose inputs are partially or wholly qualitative or based on expert judgment, provided that the output is quantitative in nature'* (Federal Reserve, 2011).

Under the Basel framework, credit risk makes up more than 80 percent of total capital estimates, predominantly driven by internal rating models (Berg et. al., 2017). Models for credit portfolios, developed in-house or outsourced to external vendors, thus constitute the largest proportion of banks' model inventories. The application of credit risk models can therefore have substantial impact on the regulatory and financial performance of a bank. Figure 7.1 summarizes the range of models that can be potentially implemented depending upon the sophistication of credit risk management systems of a bank.

The basic building blocks of credit risk are internal credit rating or scoring models, which assign a rank order to credit customers and loan facilities based on their risk profile. The objective of these models is to maximize the discrimination between good and bad credits to support sanction decisions and price and monitor the loans differentially. These may vary based on the type of borrowers (retail, SME, corporate etc.) to which they are applied. The model design in terms of parameters, weightage and aggregation formula, may be expert-judgement based; or derived through statistical or structural methods using historical data; or a mix of both.

PD, LGD and EAD models quantify three different components for estimation of standalone credit losses. PD measures the chance that an obligor will default on its contractual payment obligations, at some future horizon. Different types of PD models (grade-wise or pool-wise) can be developed based on historical default performance data associated with borrowers in different rating grades or in homogenous pools of shared risk characteristics. The models may be designed to compute a variety of default-risk estimates like those reflecting current conditions (Point-in-Time PD or PIT PD) or average conditions over an economic cycle (Through-the-Cycle PD or TTC PD) and Cumulative PD (over the loan lifetime) or Marginal PD (over incremental future periods). EAD is the nominal amount of future exposure to the obligor at the time of default. EAD models quantify internal credit conversion factors (CCF) based on historical data of defaulted fund and non-fund based credit facilities. LGD represents the fraction of credit exposure that would not be recovered post obligor default. LGD models can be constructed from historical work-out costs of defaulted accounts or from market prices of defaulted credit instruments. PD, LGD and EAD can be translated into Expected Credit Loss for provisioning; credit spreads for loan pricing; and IRB risk weights for credit risk capital charge.

Since credit risk behavior in terms of defaults, recoveries and drawdown of credit lines can be cyclical, banks can also link credit risk drivers with systemic variables to forecast credit losses under different macroeconomic scenarios, including stressed conditions. Furthermore, adverse credit events are often driven by common factors across borrowers from similar sectors, industries or regions. Thus, portfolio models can incorporate estimates of default and asset correlations to capture diversification benefits or concentration risks. Such models enable forecasts of aggregate economic credit loss at a portfolio level, based on techniques like statistical distributions, Monte Carlo simulations or Extreme Value Theory.

7.3. Model Risk Management – Supervisory Expectations and Best Practices

The earliest formal supervisory guidance on model risk management was published by the Federal Reserve (SR Letter 11-7, 2011). The guidance describes three elements of an effective MRM framework – 1) Robust model development, implementation and use; 2) Rigorous model validation process and 3) Effective model governance and control mechanisms. The framework of course needs to be 'customized to be commensurate with a bank's risk exposure, its business activities and the complexity and extent of its model use'.

In 2016, the European Central Bank launched its multi-year project on Targeted Review of Internal Models – TRIM (ECB, 2017) which focused on the internal models for regulatory capital estimation and the concern regarding 'the unwarranted (i.e. non-risk-based) variability of the outputs of models'. This was followed up with detailed instructions for validation of internal credit risk models under the IRB Approach (ECB, 2019a and 2019b). The TRIM project report (ECB, 2021) broadly confirmed that internal models were suitable for regulatory capital charge computation. However, model weaknesses were identified, many of which pertained to low default portfolios and credit scoring and LGD models for retail and SME portfolios. Non-model aspects regarding organization and activities of internal validation function, roll-out and use of models and management of model changes were also found to be deficient. The report projected that the supervisory constraints on model use imposed under TRIM would lead to 12% increase in the aggregated risk weighted assets for systemically important banks.

As a consequence of these results, banking regulators in other nations, like, Bank of England (BoE 2022) and the Financial Services Agency of Japan (FSA 2021) have also published principles for MRM, which encompass governance; model inventory and risk classification; development, implementation and use of models; independent validation; external or vendor products and resources; internal audit; and model risk mitigation. Banks need to follow these principles, not just for their regulatory models, but also for those applied for accounting provisions, internal risk management, pricing and valuation; and more so for models based on complex techniques like ML.

Driven by supervisory expectations, global banks have prioritized the implementation of effective MRM. A recent PWC survey (2022) of 32 FIs from Western Europe, America, Asia, CEE and other nations indicates that bigger banks have larger number of risk-relevant models in place and dedicate larger resources for MRM. Many have already adopted or intend to adopt technology solutions, with the objective of reducing model risk, enhancing operational efficiency and improving regulatory compliance.

Using a bibliometric analysis, Cosma and Rimo (2022) show that academic literature on model risk is fairly recent and built around themes of conceptual, computational and organizational aspects of model risk management. De Jongh et. al. (2017) propose a coherent "best practice" framework for model validation, which is founded on strong governance, policies and procedures and comprises three aspects – 1) conceptual soundness and development evidence, 2) process verification and on-going monitoring and 3) outcomes analysis. Garro (2020) describes how MRM practices at financial institutions have evolved, starting with establishing policies and standards and building internal skill-sets to ultimately creating value through automation of the MRM processes. Hill (2019) discuss multiple challenges that banks face in the evolving MRM discipline and the approaches to address these problems.

Berg et. al. (2017) use a supervisory dataset from 40 German banks for more than 17,000 borrowers over quarterly intervals spanning 2008 to 2012, to measure model risk in terms of inconsistency of PD estimates across banks. They conclude that 95 percent variation in internal PDs is idiosyncratic, whereas 5 percent is explained by bank fixed effects. While this evidence supports the use of internal risk models, it also indicates that model arbitrage may be a cause of lower PD estimates, especially for weaker capitalized banks. In the context of market risk, Farkas et. al. (2020) establish that capital estimates, when adjusted for model risk, are less volatile predictors of losses in both normal and stressed conditions and as such, more suited to conservative regulatory use.

Latest academic literature focusses on model risk in the new generation ML algorithms for credit risk, which, while outperforming traditional quantitative models in terms of predictive capabilities, can pose new supervisory risks associated with interpretability, biases, stability and external dependencies. Robisco and Martinez (2022) summarize three broad sources of model risk that can afflict ML-based default prediction models. These include statistics (stability of prediction, goodness-of-fit and calibration), technology (transparency of the ML algorithm and third party infrastructure) and market conduct (latency and interpretability). The relative importance of these factors depends upon whether the model is used for internal credit scoring and monitoring of accounts, regulatory capital estimation or provisioning. Using an open-source database they demonstrate a consistent approach to measuring model risk adjusted performance of alternative ML models.

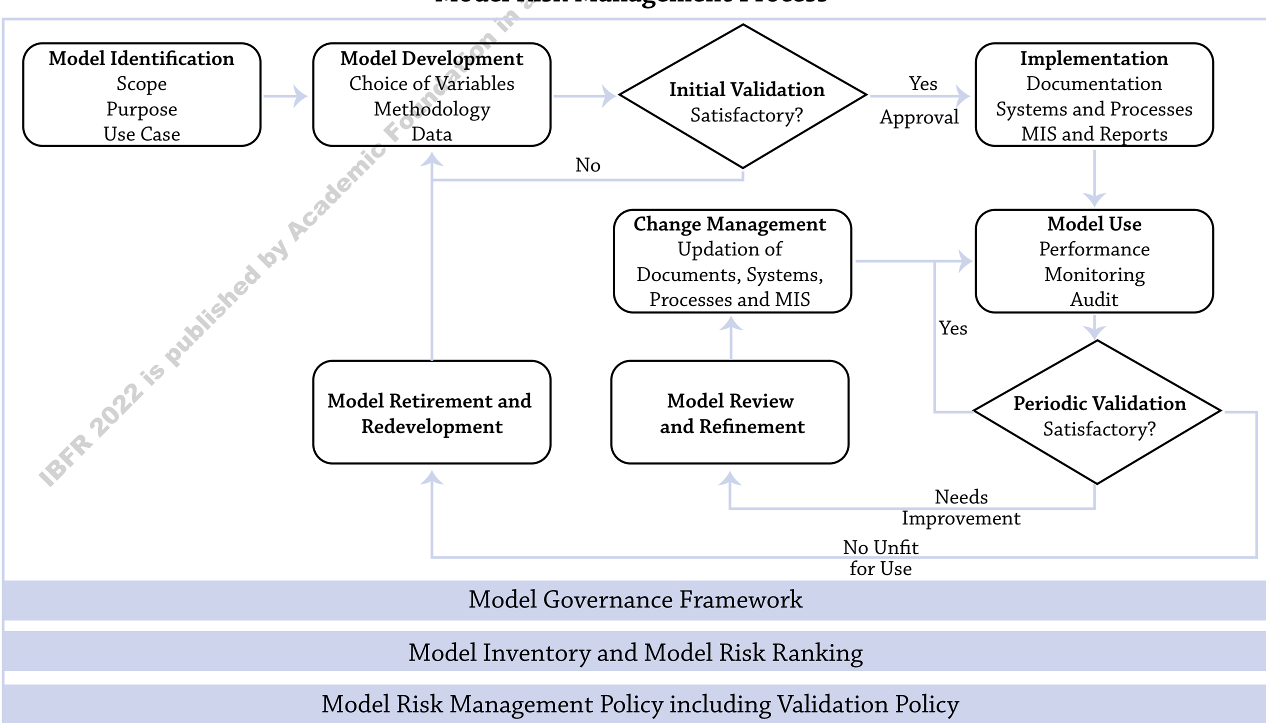
The risk-model lifecycle and its management is an on-going exercise which is expected to continue to grow in relevance and complexity as larger numbers and more sophisticated techniques get ingrained in credit risk analytics. The key steps in the MRM process, as compiled from regulatory guidelines, academic literature and industry practices are summarized in the flowchart (Figure 7.2). This process works best if it is founded on a strong governance framework and good policies encompassing model inventories.

7.4. Validation of Credit Risk Models

The most crucial and technically challenging component of model risk management for credit risk is the process and methodology for validation. The objective of validation is to ensure that 'models are performing as expected, in line with their design objectives and business uses.' (Federal Reserve, 2011).

FIGURE 7.2

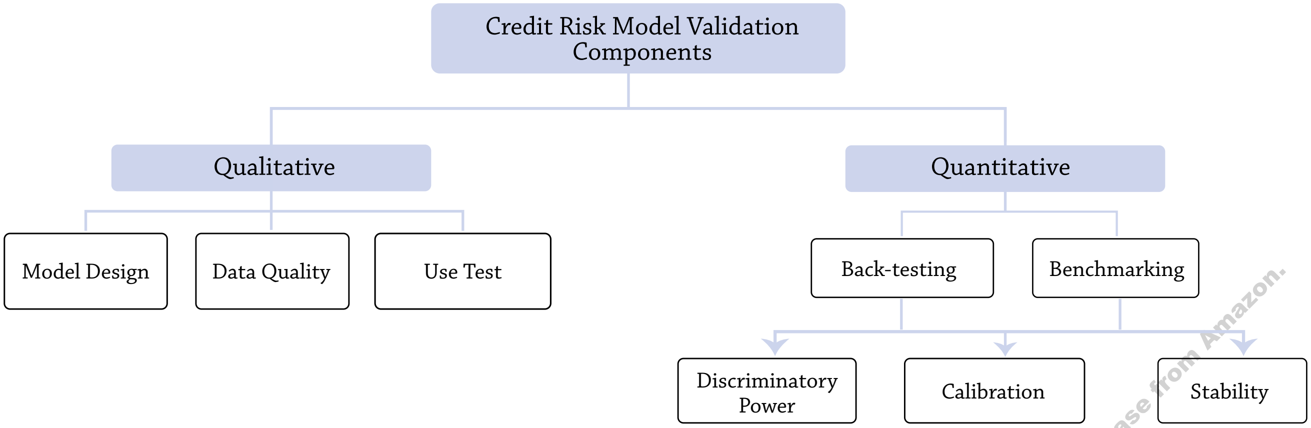
Model Risk Management Process



Source: Author's own construction.

FIGURE 7.3

Components of Credit Risk Model Validation



Source: Author's own; adapted from OeNB and FMA, 2004.

Validation is also a minimum requirement of the Basel IRB approach, which states that 'the institutions shall have a regular cycle of model validation that includes monitoring of model performance and stability; review of model relationships; and testing of model outputs against outcomes.' (BIS, 2019).

In order for model validation to be effective, there are some essential prerequisites. Firstly, the validation exercise should have a unique "owner", and be independent from model development and use, so as to provide an unbiased opinion on model performance (De Jongh, 2017). This can be achieved by establishing a distinct internal model validation division within the risk organization (McKinsey, 2017) or by involving external validators. Secondly, the validation team should have strong knowledge of model building, high level of technical expertise in statistical methods and a good understanding of the business lines to which the model is applicable (Federal Reserve, 2011). Finally, the validation process should not just be compliance driven but result in meaningful improvement of model robustness over time.

A comprehensive validation of credit risk models comprises both qualitative and quantitative aspects, which are depicted in Figure 7.3.

Qualitative Validation of Credit Risk Models

The first aspect of qualitative validation is to ensure that each model has a well-defined statement of purpose and scope of application. Very often, more than one methodology may be applied to measure a single credit risk driver for a particular credit portfolio. For example, both a pool-PD model and a score-based PD model may be concurrently applied for retail loans. In such cases, proper use-case delineation of the two models has to be ensured, where for instance, the bank employs the pooled-PD model for IRB capital charge and ECL-based provisioning and uses the score-based PD model for loan origination and pricing. Such clear-cut demarcations reduce the potential of model arbitrage and also have different implications for quantitative validation.

Verifying the strength of model design is another crucial component of qualitative validation, which needs to be established and evidenced at the model development stage. Credit risk models must be based on proven theory, sound industry practice and logic and supported by empirical validity. The input and output variables must be precisely defined, based on objective criteria and suitable to the model purpose. A credit risk model may fail the test of robust design due to multiple reasons. For example, a single, one-size-fit-all credit scorecard for retail loan customers may fail to include product-specific variables which may be critical for evaluation of credit quality (known as the problem of missing variables). A heuristic credit rating model may include qualitative input variables like business and management risk, whose sensitivity to credit worthiness may not be statistically significant. Historical LGDs may not be adjusted for collection costs and time value of money, thereby leading to an LGD model that underestimates economic losses. Another case may be when a linear regression technique, where the dependent variable takes continuous values has been used to develop a credit scorecard which classifies good and bad borrowers on probabilistic basis. An inaccurate model design can have cascading repercussions on credit risk quantification.

Furthermore, internal credit risk models used for IRB capital are subject to minimum regulatory design standards. For example, a Basel compliant internal credit rating model has to be two–dimensional so as to capture both borrower and facility risk and must have at least 7 grades. PD estimates for credit portfolios or pools have to be based on long-run average spanning a minimum historical observation period of 5 years. LGD models have to generate downturn-LGD estimates reflecting stressed economic conditions and have to be derived empirically, using minimum 7 years of historical data. The validation process has to thus confirm that there is adherence with regulatory guidance, else the model may get disqualified at the time of supervisory review.

Data quality is another important aspect of model verification. Most credit risk models are highly data intensive and data drawn from multiple loan systems is used over the entire model lifecycle. If the data used at the time of model building is not sanitized for missing or incorrect values or not representative of the credit portfolio, it may lower the model's predictive power. Ensuring data sanctity can be especially problematic if the historical records are obtained from legacy loan systems or manually collected. If the data entered by users of credit risk models, say for the purpose of borrower rating, is incomplete, biased or unreliable, it can result in mis-reporting of credit quality. Also, if adequate historical and cross-sectional granular data is not available, the assessment of the models on quantitative basis may be compromised.

Finally, under qualitative validation, all credit risk models must be well-documented and pass the use-test criteria. A precise model workflow, user acceptance and incorporation into credit functions are essential use-test criteria that must be established at the time of model development. If discretionary overrides based on individual interpretation are frequently used in a credit rating process, it may weaken the objectivity of the rating model. For unbiased credit risk evaluation, the application of the model should be independent of the business line, and at the same time, the model output should figure in some aspect of credit business decisions and strategies in a consistent manner. For instance, if rating based limits for loan approvals are disregarded by credit officers in sanctioning loans, it may result in a skewed risk profile of exposures, which breaches the bank's credit risk appetite. The models also need to be assessed on how seamlessly their functionality is integrated with the broader credit information systems of the bank. Finally, the reporting framework on the outputs of the models needs to be well-defined and suitable for top management to derive meaningful and timely signals of the changing risk profile of different credit portfolios.

Quantitative Validation of Credit Risk Models

Quantitative testing of credit risk models aims to check model accuracy and performance through a wide variety of statistical tests. These tests may vary substantially depending upon the purpose for which the model is designed, the underlying methodology and data on which the model is constructed and the nature of the model output. The broad areas of quantitative tests include verifying the discriminatory power, calibration and stability of the model. Furthermore, the test parameters should be back-tested to ensure that predictive ability is not decaying and benchmarked with other models for comparative performance checks. A BCBS working paper (2005) provides a detailed description of quantitative validation techniques used for internal rating systems. Wu and Olsen (2010) demonstrate the application of these methods on predictive credit scorecards of a large Canadian bank. In the following subsections, we discuss some of these techniques.

Discriminatory Power

Credit rating or scoring models, which typically attempt to rank order customers or facilities based on credit quality, need to be assessed for classification accuracy or discriminatory power, which is ability to differentiate between "good" and "bad" credit. Borrower or counterparty rating models try to forecast whether a borrower will default (bad) or not default (good) and therefore reflect PD. Facility ratings, on the other hand, classify the extent of post-default loss, that is, LGD. The rating scale thus implies an ordinal ranking of the credit risk driver (PD or LGD as the case may be).

At the development stage, credit rating models are usually tested for their classification accuracy based on sample data prior to selecting the final model. Post implementation, periodic analysis and monitoring of models' discriminatory power is an important component of quantitative validation. The classification accuracy tries to measure the extent to which the rating model correctly differentiates good and bad credit and more importantly, the extent to which accounts are mis-classified, thereby leading to Type 1 or α-error (bad accounts classified as good) and Type 2 or β-errors (good accounts classified as bad). As summarised in Table 7.1, a variety of graphs or statistical measures can be used to test model discriminatory power.

To apply these tests, performance data of accounts rated or scored by the model is required over a pre-defined observation period. Table 7.2 depicts account-wise summary data of good (non-defaulted) and bad (defaulted) accounts over a one year observation period post application of a 10-grade credit rating model. From this data, we can construct a ROC curve as shown by the red arc in Figure 4 and measure the AUROC, which is 73.63%. An ideal rating model, that is, one which has perfect discriminatory power, would be graphically represented by the green line in Figure 7.4, such that AUROC = 1. On the other hand, for a model which cannot differentiate between good and bad cases, the black diagonal line would represent the ROC curve, for which AUROC = 0.5. Thus, the greater the convexity of the ROC Curve, the higher the value of AUROC, the better the discriminatory power of the model. Models where AUROC < 0.5 would be ones where the rating system classifies the cases at least partly in the wrong order.

TABLE 7.1

Summary of Discriminatory Power Analysis Methods

Graphical Representation	*Statistical Measures*
• *Cumulative Frequency Distributions* of good and bad cases across credit scores	• *Kolmogorov-Smirnov (K-S) Statistic:* Maximum difference between cumulative proportions of good and bad accounts across credit scores
• *Receiver Operating Characteristic (ROC) curve:* the plot of cumulative proportion of bad accounts vis-à-vis cumulative proportion of good accounts across credit scores	• *AUROC:* Total area under the ROC Curve
• *Cumulative Accuracy Profile (CAP) curve:* the plot of cumulative proportion of bad accounts vis-à-vis cumulative proportion of all accounts across credit scores	• *Accuracy Ratio / Gini Coefficient* = 2*AUROC - 1
• α-β *Error curve:* the plot of β error vis-à-vis the α error across credit scores	• *Bayesian Error Rate:* Minimum weighted average of α and β errors across credit scores

Source: Author's own construction, based on literature.

TABLE 7.2

Grade-wise Distribution of Good and Bad Accounts of an Internal Rating Model

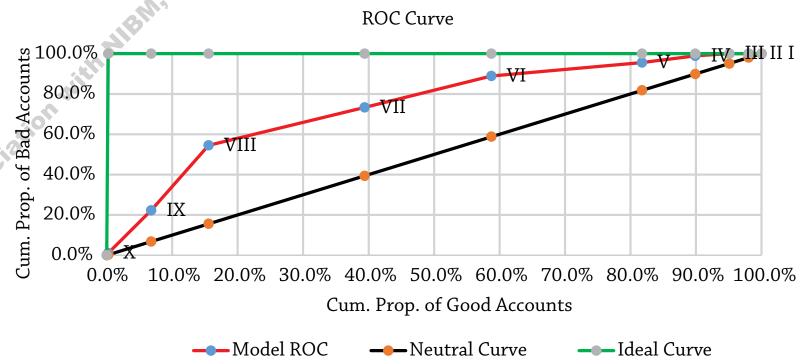
Rating Grades	*No. of Good Accounts*	*No. of Bad Accounts*	*Total No. of Accounts*	*Proportion of Total Accounts*	*Proportion of Bad Accounts*	*Grade-wise Default Rate*
X (worst)	8	1	9	0.26%	1.11%	11.11%
IX	219	19	238	6.92%	21.11%	7.98%
VIII	294	29	323	9.40%	32.22%	8.98%
VII	797	17	814	23.68%	18.89%	2.09%
VI	648	14	662	19.26%	15.56%	2.11%
V	770	6	776	22.58%	6.67%	0.77%
IV	273	3	276	8.03%	3.33%	1.09%
III	173	1	174	5.06%	1.11%	0.57%
II	99	0	99	2.88%	0.00%	0.00%
I (best)	66	0	66	1.92%	0.00%	0.00%
All Grades	3347	90	3437	100.00%	100.00%	2.62%

Source: Illustrative Data.

Whether a model's classification accuracy is acceptable or not can be verified with respect to its statistical significance and also by comparing the model metrics with benchmarks of similar models – for example external rating agency models. Discriminatory power of the model should also be within a threshold value of the validation metric specified in the bank's model risk appetite. Finally, whether there is a substantial reduction of discriminatory power from that which was established at the time of model development, should also be examined through back-testing.

FIGURE 7.4

The ROC Curve for a 10-Grade Rating Model



Source: Illustrative Diagram constructed from data in Table 7.2.

Calibration

Internal rating models need to be calibrated to a relevant and cardinal measure of credit risk. Thus, default-risk scoring models should map a forecasted PD value to each rating or score band. Similarly, facility rating models should assign predicted LGD values to facility grades. Logistic or Probit regression based credit scoring models have outputs which are in the form of statistical default probabilities. Structural credit risk models which use price data from equity, bond or credit default swap markets generate risk-neutral default probabilities. Expert-judgment based rating models, on the other hand, produce discrete risk grades. Each of these models requires the scaling of the a-priori default probability bands or rating grades to projected TTC PD or PIT PD as suitable. Tasche (2013) and Nehrebecka (2016) describe a variety of calibration techniques and the conditions under which they are fit for purpose.

An ideal calibrated PD curve should preferably demonstrate monotonicity (that is, default rate for a better rating should be lower than default rate for an adjacent worse rating). Furthermore, forecasted PDs, especially if used for regulatory capital charge, should be strictly positive. Column 2 in Table 7.3 provides an illustrative example of the actual default rates across rating grades in the observation period. It includes some instances of inversions of observed default rates (1.09% for Grade 4 > 0.77% for Grade 5), as may arise in real-life data. Such deviations may indicate a problem with the rank ordering capacity of the rating methodology or might just be an exception to the long-run pattern. Column 3 shows the calibrated PDs for the forecast period, estimated using the Quasi Moments Matching Method (Tasche, 2013) and subject to the constraints of monotonicity and strictly positive values.

Since calibrated credit risk outputs forecast the performance of the credit portfolio in quantitative terms, they should also be back-tested against realized outcomes. Thus, calibrated grade-wise or pool-wise 1-year PDs at time t, which predict the default rates at time t+1 year, need to be verified against observed default rates at the end of the 1-year horizon. Various statistical tests like Jeffrey's Test (ECB, 2019a), Brier Score (OeNB and FMA, 2004), t-test, standard normal test can be used to determine if there is significant deviation of actual default rates from the calibrated PDs both at a portfolio level and for individual rating grades or portfolio segments.

TABLE 7.3

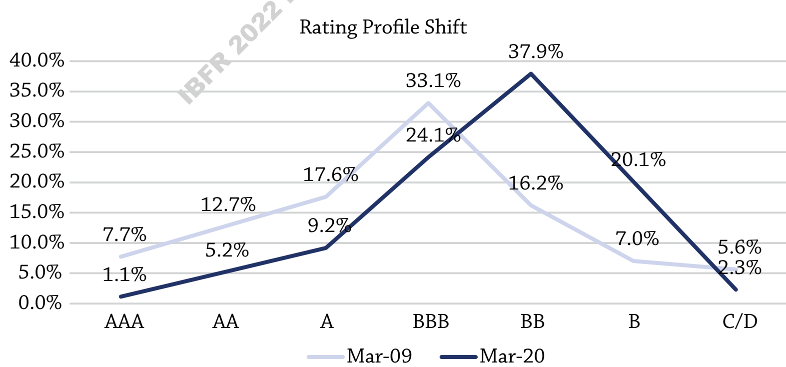
Calibrating PDs to Internal Rating Grades

Rating Grades	*Observed Default Rates (time t)*	*Calibrated Forecasted PDs (time t+1)*
10 (worst)	11.11%	27.55%
9	7.98%	17.60%
8	8.98%	10.71%
7	2.09%	6.31%
6	2.11%	3.64%
5	0.77%	2.08%
4	1.09%	1.18%
3	0.57%	0.67%
2	0.00%	0.21%
1 (best)	0.00%	0.03%
All Grades	2.62%	4.98%

Source: Illustrative Data.

FIGURE 7.5

Shift in the Rating Profile of Borrowers Over Time



Source: Illustrative Depiction.

Stability

The stability of credit risk models is typically examined from three perspectives - change in the portfolio profile, stability of the rating migration matrix and concentration in rating grades. If there has been a significant change in the customer profiles between the time a credit risk model was first developed and at the time of its latest validation, it could be a possible cause of the underperformance of the model. The Population Stability Index (PSI) provides a relevant quantitative measure of whether a large shift in the portfolio's rating profile has occurred and the same can be detected visually as shown in Figure 7.5, where the distribution of accounts across ratings has deviated in 2020 from that in 2009.

For credit rating models where the customer grade is reviewed and revised periodically upon availability of fresh information, the rating migration matrix captures the movement of accounts from one grade to another over a predefined horizon. Such matrices need to be scrutinised for their stability in terms of proportion of accounts remaining at their original rating grade, vis-à-vis transitioning to higher or lower rating grades. Point-in-time rating models which use current borrower information, can be expected to have a higher degree of variability in terms of upgrades and downgrades over different phases of the macro-economic cycle. This feature is desirable especially for models used for ECL provisions under IFRS. However, through-the-cycle rating models capture the average credit quality of borrowers over business cycles, and are intended to ensure a steady prediction of regulatory capital. Such models must demonstrate a fair degree of stability of rating transitions at a portfolio level.

TABLE 7.4

1-Year Rating Migration Matrix

	CR1	*CR2*	*CR3*	*CR4*	*CR5*	*CR6*	*CR7*	*CR8*	*CR9*	*D*
CR1	60.0%	7.06%	17.65%	5.89%	2.35%	5.88%	1.18%	0.00%	0.00%	0.00%
CR2	5.11%	59.71%	22.70%	7.16%	5.11%	0.00%	0.00%	0.00%	0.00%	0.20%
CR3	0.70%	5.26%	55.03%	19.56%	11.99%	2.86%	1.73%	0.64%	0.55%	1.67%
CR4	0.15%	2.42%	8.08%	63.09%	14.47%	5.03%	3.06%	1.21%	0.79%	1.63%
CR5	0.18%	1.93%	4.78%	14.22%	69.70%	4.12%	2.41%	1.02%	0.73%	0.91%
CR6	0.63%	0.21%	1.90%	9.49%	20.89%	54.96%	7.38%	1.27%	1.16%	2.10%
CR7	0.00%	0.00%	1.01%	3.04%	21.27%	10.85%	54.27%	4.05%	1.88%	3.62%
CR8	0.00%	3.03%	0.00%	10.10%	10.10%	7.07%	21.21%	38.38%	6.06%	4.04%
CR9	0.00%	0.00%	0.00%	0.00%	9.86%	0.00%	8.45%	8.45%	52.11%	21.13%

Source: Illustrative Data.

Table 7.4 provides an example of a rating migration matrix, estimated from one-year rating movements of borrower accounts, averaged over 5 years. The diagonals represent the average proportion of accounts in each grade where the rating remained same as at the start of each year, which should ideally be the highest vis-à-vis other transition probabilities. The last column indicates the long-run average annual default rate (or TTC PD) for each grade which should optimally increase with worsening grades. The green cells and orange cells respectively show the proportion of upgrades and downgrades from each rating, which should fall for migrations further away from the diagonal. Transition matrices derived from real-life applications of rating models may show some violations of these features, as is seen in Table 7.4. Persistently deviant outcomes may be indicative of defective model design. For example, if in the rating model design, certain grades have been mapped to very broad score bands, those grades may not be sensitive to small changes in the underlying score.

The more stable the rating model, the lower would be the variability in transition probabilities over time. To back test the stability of the rating model (grade-wise and overall), the latest one-year migration data pertaining to the validation period can be compared with the average migration percentages from the past data to identify whether there has been a sharp deviation, which could be indicative of potential model instability.

The final feature of stability that needs to be proved is the homogeneity of accounts and exposures across grades or pools created for the estimation of the credit risk parameter. If there is a significant concentration of accounts in a few grades, the risk output can become highly variable within those grades. The Hirschman Herfindahl Index (HHI) is a useful metric to capture the uniformity of accounts across grades or pools. Where xi represents the percentage of accounts in the i-th grade, HHI is estimated as

$$\text{HHI} = \sum_{i=1}^{10} x_i^2$$

TABLE 7.5

Grade-wise Concentration of Accounts

Rating Grades	*No. of Rated Accounts*	*% of Accounts (xi)*
I	26	0.4%
II	513	8.7%
III	630	10.7%
IV	728	12.4%
V	935	15.9%
VI	949	16.1%
VII	1322	22.4%
VIII	606	10.3%
IX	176	3.0%
X	8	0.1%
Total	5893	100.0%

Source: Illustrative Data.

As an illustration, Table 7.5 shows the distribution of accounts across borrower grades, where proportion of accounts is relatively higher in grades VI and VII. If there was an exactly even distribution of accounts, the HHI would take a value of 10% since there are 10 grades. However, the actual HHI for this profile is 14.7%. While perfect homogeneity of rating profile may not be possible for any risk model, higher exposure concentration in a few rating grades needs to be examined for its potential causes, implications and remedial measures.

7.5. Challenges in Model Risk Management

Ensuring an effective MRM framework at a bank throws up many challenges. On the governance side, bank boards need to prioritise MRM as an important pillar of risk management, so that their credit strategies and regulatory performance are not undermined. This requires an independent and dedicated team of skilled professionals, which may be difficult to build and retain.

The validation exercise can be especially difficult for legacy credit risk models which suffer from inadequate documentation, poor data quality and untested methodologies and assumptions. Thus, it is important for banks to establish high standards for these aspects, which should be equally applicable for in-house and outsourced models from the time of model development.

Credit risk models used for regulatory purposes usually build in a degree of conservatism in either the inputs or the design or through adjustment of outputs. Expert-judgment based models often incorporate some management overlays. The Covid-19 pandemic, which involved government and central bank interventions to protect stressed borrowers may have entailed banks applying subjective adjustments to their credit risk models to reflect the pandemic shock (BCBS, 2022). The challenge in validating such models is to ensure that the margin of conservatism or discretion does not bias the risk estimates to an extent where they are way off the mark from realized credit losses.

The pandemic shock also impacted historical default trends and correlations, which have disrupted the use of the data pertaining to this aberrant period in validating credit risk models. Engelmann (2022) suggests a framework that measures the pandemic intervention effects and how they can be included as scenarios in credit risk models, to improve the accuracy of their short-term predictions and evaluate long-term consequences.

Credit portfolios which constitute a low number of accounts or low number of defaults can present unique hurdles in applying standard statistical validation methods. Risk models developed for such portfolios therefore need to be benchmarked to comparable external models or tested with representative data from external sources.

Because of the large landscape of credit risk models, there can be a number of complex interactions among upstream and downstream models. Identifying these dependencies and ensuring that adverse outcomes of any one model do not spill over to other connected models can become a cumbersome and unwieldy task (Hill, 2019).

The most challenging task in managing model risk is to reach a conclusion about whether the model is fit for use, needs minor adjustments to improve performance, requires material changes or has to be entirely re-developed. Such change management decisions can be costly for the bank since all associated processes, policies and systems would need to be altered and should be based on careful interpretation of the validation results. Thus, it is imperative for banks to put in place a model tiering or ranking system which holistically assesses model risk and to establish appropriate model risk appetite thresholds based on the ranking.

7.6. Concluding Observations

Banks in India have implemented many credit risk models, either through regulatory compulsion or to improve business management. While internal credit rating models for commercial loans were already in use since the first decade of the twenty first century, these were primarily developed for competitive pricing and as such, had many limitations and were not subject to regulatory scrutiny. After the issuance of the Guidelines on the Basel IRB Approach (RBI, 2011), most banks retired their legacy rating systems and developed Basel-compliant models for internal ratings and estimation of PD, LGD and EAD for different regulatory credit portfolios. While no bank in India has yet migrated to the IRB Approach, these models continue to be maintained and leveraged for credit decisions. With the imminent transition to IFRS accounting standards, banks have also taken steps to augment their credit risk models for the purpose of computation of ECL based provisions in line with IFRS 9 guidance. Furthermore, a PwC and FICCI Survey (2022) indicates that AI & ML based credit risk management is an emerging focus area for Indian banks and FIs who are using tools like decision trees and neural networks for developing credit scoring models, for deploying new credit and for collections and recovery optimization.

In a world of digitalized bank credit, where increasingly financial decisions and regulatory compliances are getting automated through implementation of multiple and complex models, model risk management has become a critical oversight requirement for supervisors and top management. The importance of assessing the adequacy and robustness of credit risk models is slowly gaining traction in India, as banks wake up to the reality and benefits of model based measurement of their capital and provisions. Currently, Indian banks are focused more on validation of individual models and that too, primarily those which are within the regulatory oversight. In this context, it is important that the Reserve Bank of India, like banking supervisors in other countries, formalises comprehensive MRM guidelines. This will provide Indian banks with a consistent and holistic framework for evaluating model performance and managing model risk. It will also ensure regulatory confidence in internal credit risk estimates in line with global best practices.

References

BCBS. (2005). Studies on the validation of internal rating systems, Working Paper No. 14, Available at: https://www.bis.org/publ/bcbs_wp14.htm.

BCBS. (2022). BCBS Newsletter on Covid-19 related credit risk issues, Basel Committee on Banking Supervision, March 2022. Available at: https://www.bis.org/publ/bcbs_nl26.htm.

Berg, T., and Koziol, P. (2017). "An analysis of the consistency of banks' internal ratings". *Journal of Banking & Finance*, 78, pp.27-41.

BIS. (2019). CRE 36 – IRB Approach: Minimum Requirements to use IRB Approach, Bank for International Settlements, December 2019. Available at: https://www.bis.org/basel_framework/chapter/CRE/36.htm.

BOE. (2022). CP/6/22-Model Risk Management Principles for Banks, Bank of England, Prudential Regulation Authority, June 2022. Available at: https://www.bankofengland.co.uk/prudential-regulation/publication/2022/june/model-risk-management-principles-for-banks.

Cosma, Simona and Rimo, Giuseppe. (2022). Model risk in banking studies: A bibliometric analysis, Conference Paper, September 2022, Available at: https://www.researchgate.net/publication/363470261_Model_risk_in_banking_studies_a_bibliometric_analysis.

De Jongh, P. J., Larney, J., Mare, E., Van Vuuren, G. W., and Verster, T. (2017). "A proposed best practice model validation framework for banks". *South African Journal of Economic and Management Sciences*, 20(1), 115.

ECB. (2017). Guide for the Targeted Review of Internal Models (TRIM), European Central Bank, February 2017. Available at: https://www.bankingsupervision.europa.eu/ecb/pub/pdf/trim_guide.en.pdf.

ECB. (2019a). Instructions for reporting the validation results of Internal models IRB Pillar I models for credit risk, European Central Bank, February 2019. Available at: https://www.bankingsupervision.europa.eu/banking/tasks/internal_models/shared/pdf/instructions_validation_reporting_credit_risk.en.pdf.

ECB. (2019b). ECB Guide to Internal Models, European Central Bank, October 2019. Available at: https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.guidetointernalmodels_consolidated_201910~97fd49fb08.en.pdf.

ECB. (2021). Targeted Review of Internal Models, Project Report, European Central Bank, April 2021. Available at: https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.trim_project_report~aa49bb624c.en.pdf.

Engelmann, Bernd. (2022). "Modeling credit risk in the presence of central bank and government intervention", *Journal of Risk Model Validation*, 16(1), March, pp.1- 22.

Farkas, W., Fringuellotti, F., and Tunaru, R. (2020). "A cost-benefit analysis of capital requirements adjusted for model risk". *Journal of Corporate Finance*, 65, 101753.

Federal Reserve. (2011). Supervisory guidance on model risk management. Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency, SR Letter 11-7, April 2021. Available at: https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf.

FSA of Japan. (2021). Principles for Model Risk Management, Financial Services Agency of Japan, November 2021. Available at: https://www.fsa.go.jp/common/law/ginkou/pdf_03.pdf.

Garro, M. (2020). "The evolution of model risk management processes". *Journal of Risk Management in Financial Institutions*, 13(1), pp.16-23.

Hill, J. R. (2019). "The top 14 challenges for today's model risk managers: Has the time come to think about going beyond SR11-7?", *Journal of Risk Management in Financial Institutions*, 12(2), pp.145-167.

McKinsey. (2017). The evolution of model risk management. Available at: https://www.mckinsey.de/businessfunctions/risk-and-resilience/our-insights/the-evolution-of-model-risk-management.

Nehrebecka, Natalia. (2016). Probability of default curve calibration and validation of internal rating systems, Eighth IFC Conference on "Statistical Implications of the New Financial Landscape", Basel 8-9, September 2016, Available at: https://www.bis.org/ifc/publ/ifcb43_zd.pdf.

OeNB and FMA. (2004). Guidelines on credit risk management – rating models and validation, November. Available at: https://www.oenb.at › rating_models_tcm16-22933.

PWC. (2022). Model Risk Management Survey Report, April 2022. Available at: https://www.pwc.com/cz/en/temata/model-risk-management-survey.html.

PwC and FICCI. (2022). Uncovering the ground truth: AI in Indian financial services, February 2022. Available at: https://www.pwc.in/assets/pdfs/research-insights/2022/ai-adoption-in-indian-financial-services-and-related-challenges.pdf.

RBI. (2011). Implementation of the Internal Rating Based (IRB) Approaches for Calculation of Capital Charge for Credit Risk, Reserve Bank of India RBI/2011-12/311, December 2011. Available at: https://rbidocs.rbi.org.in/rdocs/notification/PDFs/31ICB211211F.pdf.

Robisco, Andrés Alonso and Martínez, José Manuel Carbó. (2022). Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction, Financial Innovation, Springer; *Southwestern University of Finance and Economics*, 8(1), pp.1-35.

SAS and GARP. (2019). Artificial Intelligence in banking and risk management: keeping pace and reaping benefits in a new age of analytics, Available at: https://www.sas.com/content/dam/SAS/documents/marketing-whitepapers-ebooks/third-party-whitepapers/en/artificial-intelligence-banking-risk-management-110277.pdf.

Tasche, Dirk. (2013). "The art of probability-of-default curve calibration", *Journal of Credit Risk*, December 2013, 9(4), pp.63-103.

Wu, D., and Olson, D. L. (2010). "Enterprise risk management: coping with model risk in a large bank". *Journal of the Operational Research Society*, 61(2), pp.179-190.