A Review on the Probability of Default for IFRS 9

Matthias Bank and Bernhard Eder*

Department of Banking and Finance, University of Innsbruck

December 9, 2021

Abstract

By means of a systematic literature review, we provide a comprehensive overview of 52 contributions on the estimation of the probability of default according to the "Expected Credit Loss Model" (ECLM) in IFRS 9 and the corresponding stage allocation. We distinguish the models discussed to estimate the probability of default with regard to the data they use as well as their time-space and state-space. Thus, we not only consider the merits and limitations of particular models in isolation, but also demonstrate how they are related to each other and where they differ. Due to the methodological diversity alone, it can be expected that loss allowances for expected credit losses will differ significantly across banks. In addition, we discuss the implementation of the stage allocation in the ECLM. Its inherent scope for discretion is likely to further amplify differences in the level of loss allowances across banks.

Keywords: IFRS 9, expected credit loss model, financial instruments, literature review.

1 Introduction

The replacement of the "incurred loss model" in IAS 39 by the "expected credit loss model" (ECLM) in IFRS 9 represents a major paradigm shift in the impairment rules for financial instruments. Previously, loan loss allowances had to be recognized only if there was objective evidence that an impairment had already incurred. However, the financial crisis revealed that impairments were often recognized "too little, too late". In response to political pressure (FCAG 2009; FSF 2009; G20 2009), the IASB has therefore developed a new impairment model for financial instruments. Since 2018 entities have to test whether or not a financial instrument has experienced a significant increase in credit risk since origination. If not, an amount equal to the expected credit loss due to events within the next 12 months has to be recognized. Otherwise, a loss allowance equal to the expected credit loss over the remaining lifetime has to be set aside.

*Corresponding author: bernhard.eder@uibk.ac.at. We are grateful to Christoph Pelger for valuable feedback on our manuscript. Earlier drafts of this paper have been presented at the 32nd Workshop of the Austrian Working Group on Banking and Finance at the University of Innsbruck, the CARF Conference 2018 in Lucerne, and the Credit Scoring and Credit Control Conference 2019 at the University of Edinburgh. We thank the participants for their helpful comments.

Compared to the Basel Accords, where banks may also provide their own estimates of expected losses, IFRS 9 differs mainly with regard to the probability of default (PD). Under the Basel Accords, it is sufficient to estimate the likelihood that a borrower will default within the next year. For IFRS 9, however, the PD over the remaining lifetime of a financial instrument has to be determined. Since the decision whether a financial instrument experienced a significant increase in credit risk since origination or not is to be based on the PD over the remaining lifetime as well, such estimates are required for all financial instruments within the scope of the new impairment model. Yet, IFRS 9 is a principle-based standard, and does not prescribe specific models to be used. Since its publication, a variety of different models have therefore been proposed. But a comprehensive overview and comparison of these models is missing so far.

Against this background, we provide a systematic literature review on estimating the PD for IFRS 9 and the determination of significant increases in credit risk since origination. To create our sample, we conducted a extensive literature search on several academic databases. To ensure a comprehensive sampling, we selected from the search results those contributions that either directly or indirectly refer to IFRS 9. For quality reasons, however, we have mainly included contributions that have been issued by renowned publishers. In total, our final sample comprises 52 contributions.

Our review makes several contributions to the literature. First, we complement previous surveys on estimating the PD according to IFRS 9. These include surveys by Brunel (2016a), Xu (2016), and Skoglund (2017) who already discuss a number of statistical models, benchmark studies by Altman et al. (2020), Breeden and Vaskouski (2020), and Breeden and Crook (2020) who have also compared several models empirically, and the textbooks by Baesens et al. (2017) and Bellini (2019) who provide many practical examples. Yet, unlike these contributions, which have selectively surveyed the previous credit risk literature, we systematically review and synthesize the literature related to IFRS 9. We classify the models under consideration according to their underlying data as well as their state-space and time-space. This allows us not only to discuss the extent to which individual models are suitable for meeting the requirements of IFRS 9, but also to emphasize the relationships and differences between these models. In addition, we also discuss some challenges that arise in connection with the consideration of forward-looking information when estimating the PD. Our review may therefore not only be relevant for researchers, but also for practitioners who wish to replace the first generation of their IFRS 9 models with more sophisticated ones, as well as for auditors and supervisors charged with monitoring these models (EBA 2020).

Second, we provide a critical appraisal of the stage allocation. We emphasize that the assessment of significant increases in credit risk is highly discretionary, which is likely to exacerbate differences in loan loss allowances across banks. Moreover, we find that the stage allocation is occasionally misunderstood as a two-grade rating system. Significant increases in credit risk would thereby only be based on the absolute risk of default on the balance sheet date, whereas IFRS 9 generally requires a comparison with the expected risk of default at origination. In addition, since the stage allocation should be performed at the level of individual financial instruments, we argue that the PD for IFRS 9 should ideally be measured at the account level instead of the portfolio level as well.

This argument is closely related to our third contribution, which emerges from a comparison of IFRS 9 and the Basel Accords. In contrast to previous comparisons, we emphasize two aspects which have direct implications for the measurement of the PD. On the one hand, we highlight differences in the calibration of the PD. While for IFRS 9 the PD should be estimated at the account level, the Basel Accords prescribe to calibrate it either at the rating class or portfolio level, depending on the customer segment under consideration. Approaches to adjust regulatory PDs for purposes of IFRS 9 therefore not only entail significant model risk, but in our view also represent a coarse implementation of IFRS 9. Yet, on the other hand, we also underline parallels between the PD for IFRS 9 and internal ratings according to the Basel Accords. In particular, we advocate that models to estimate the PD over the remaining lifetime may at least be used as a starting point for assigning long-term ratings as required by Basel III. In this respect, we also argue that the common classification that IFRS 9 requires "point-in-time" (PIT) ratings or PDs, while the Basel Accords require "through-the-cycle" (TTC) ratings or PDs, seems both to be misleading and to exaggerate the differences between the two standards.

The remainder of this paper is structured as follows. In the next section, we will provide a short summary of the expected credit loss modell and compare it to the expected loss models in US-GAAP and the Basel Accords. Thereafter, we will outline how we have identified and selected relevant contributions that are included in our literature sample, and introduce some notational conventions. In sections 5 to 9 we provide a structured overview of the models discussed in the literature for estimating the probability of default for IFRS 9 and approaches to adjust regulatory estimates. In sections 10 and 11 we will address the consideration of forward-looking information and the operationalization of the SICR-test. In the final section, we summarize this paper and suggest avenues for future research.

2 Expected Loss Models in Accounting and Regulation

2.1 IFRS 9's Expected Credit Loss Model

The ECLM has to be applied to financial instruments that are either categorized as assets measured at amortized cost or assets that are measured at fair value through other comprehensive income (IFRS 9 5.2.2). Typical products within these categories include loans, credit cards, corporate bonds and treasury bills. The impairment rules also apply to lease and trade receivables, customer contracts, and to specific liabilities such as financial guarantees and loan commitments.[1]

Each financial instrument that is within the scope of the ECLM has to be assigned to one of three stages, as depicted in figure 1. Depending on its performance since origination, an entity has to determine for each financial instrument whether or not it has experienced “a significant increase in credit risk since initial origination”, which is typically abbreviated with “SICR-test”. If the SICR-test is negative, the financial instrument belongs to the first stage. Within this stage, an entity has to recognize a loss allowance equal to the expected credit loss due to a default within the next twelve months (IFRS 9.5.5.5). If the SICR-test is positive, the financial instrument belongs to the second stage and a loss allowance equal to the expected credit loss over its remaining lifetime has to be

1. The impairment rules do not apply to financial assets measured at fair value through profit or loss, such as equity instruments, all types of derivates, insurance contracts, as well as interests in subsidiaries, associates and joint ventures.

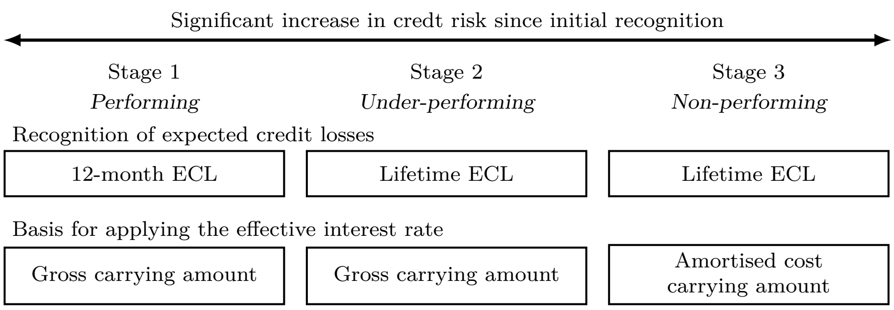

Figure 1: The three stage approach.

recognized (IFRS 9 5.5.3). Within both stages, interest revenue has to be calculated by applying the effective interest rate to the gross carrying amount (IFRS 9 5.4.1).[2] If the instrument is not only significantly underperforming relative to initial expectations, but even credit impaired, it belongs to the third stage (IFRS 9.5.4.4). In this case, a loss allowance equal to the expected credit loss over the remaining lifetime has to be recognized as well, but the effective interest rate has to be applied on the amortized cost carrying amount (IFRS 9 5.4.1). If the reasons for a stage transfer cease to apply at a later balance sheet date, the instrument must be transferred back into the appropriate stage (IFRS 9 5.5.7). Exceptions to this general impairment approach exist for assets that were already credit impaired at purchase or origination (IFRS 9 5.5.13) and for trade receivables, lease receivables and contract assets without a significant financing component (IFRS 9 5.5.15).

Expected credit losses are defined in IFRS 9 (App. A) as the "difference between all contractual cash flows [...] and all the cash flows that the entity expect to receive", and shall fulfil three conditions (IFRS 9 5.5.17). First, they shall be an unbiased and probability-weighted amount. Second, they shall take into account the time value of money, which is done by discounting the expected credit losses by the effective interest rate.[3] And thirdly, they shall reflect information about past events, current conditions, and forecasts of future economic conditions, i.e. forward-looking information. The period over which lifetime expected credit losses have to be measured is in general the remaining time to contractual maturity (IFRS 9 5.5.19). The only exception to this are revolving instruments, such as credit cards, where lenders are exposed to credit risk beyond the contractual notice period. In this case, the remaining economic lifetime is to be used (IFRS 9 5.5.20). In addition, IFRS 9 generally requires to measure expected credit losses at the level of individual instruments. A collective assessment shall only be used as an approximation if reasonable and supportable information is not available without undue cost or effort (IFRS 9 B5.5.4).

It is common to decompose a financial instrument's expected credit loss into its individual com-

2. Exceptions from this apply to financial assets that were already credit impaired at purchase or origination, where the credit-adjusted effective interest rate has to be used.

3. Using the effective interest rate as discount rate distinguishes the expected *credit* losses from the more general concept of expected losses, where the choice of the discount rate is discretionary.

ponents (Xu 2016; EBA 2020):

$$ECL_i = E\left[\sum_{\tau=1}^{M_i} PD_{i\tau}^m(\mathbf{z}_\tau) \cdot LGD_{i\tau}(\mathbf{z}_\tau) \cdot EAD_{i\tau}(\mathbf{z}_\tau) \cdot \frac{1}{(1+k_i)^\tau} | \mathbf{z}_0\right],$$

where the exposure at default ($EAD_{i\tau}$) is the monetary amount outstanding at the time of default, the loss given default ($LGD_{i\tau}$) is the fraction of $EAD_{i\tau}$ that is expected to be lost in the case of default, $PD_{i\tau}^m$ is the marginal probability of default in period τ, and k_i is the effective interest rate. M_i is the shorter of one year and the remaining lifetime if the instrument is allocated to stage 1. Otherwise, it is equal to the remaining lifetime. Conditional on the vector of macroeconomic variables $\mathbf{z}_\tau$, the individual components are usually assumed to be independent. But irrespective of how the values of $\mathbf{z}_\tau$ are predicted or selected, they can only be based on the information available up to the balance sheet date ($\mathbf{z}_0$).

To determine whether a financial instrument has experienced a significant increase in credit risk since initial origination or not, entities have to assess changes in the risk of default (IFRS 9 5.5.9). However, this is subject to a few qualifications. First, other determinants of credit risk, like changes in the LGD, shall explicitly not be considered (IFRS 9 B5.5.12). Second, to assess changes in credit risk, entities have to compare at each balance sheet date the present risk of default over the remaining lifetime to the corresponding risk of default as it was expected at origination. This implies that the initial expectation must be adjusted for the decrease in the risk of default that is solely attributable to the passage of time (IFRS 9 B5.5.11). As a practical expedient, entities are also allowed to use changes in default risk over the next 12 months, if this does not result in a biased assessment, for example due to omitted default concentrations at a specific maturity (IFRS 9 B5.5.13). However, supervisors have clarified that they expect banks to make only limited use of such operational simplifications (BCBS 2015; EBA 2017). Third, it is explicitly clarified in IFRS 9 that it is not permitted to base the SICR-test on an assessment of the absolute value of the default risk at the reporting date, unless a financial instrument has a low credit risk (IFRS 9 5.5.10). This can for example be assumed if the instrument has an investment grade rating (IFRS 9 B5.5.22–24). And finally, it has also been clarified that a financial instrument cannot be in multiple stages at the same reporting date (ITG 2015). This implies that if several scenarios are used, the weighted average of the scenario-specific PDs must be determined before the stage assignment is performed.

However, what constitutes a "significant" increase in credit risk is only vaguely defined. The IASB has only stipulated that there will be already a significant increase before a financial instrument becomes credit-impaired, and introduced the rebuttable presumption that credit risk significantly increased if an obligor is more than 30 days past due (IFRS 9 5.5.11). In addition, a comprehensive but non-exhaustive list of indicators is provided in IFRS 9 which may be relevant to assess changes in credit risk (IFRS 9 B5.5.17). Furthermore, the BCBS has clarified that "significant" should not be equated to statistical significance. Instead, it has to be interpreted as a "substantial" increase in credit risk (BCBS 2015). But apart from these broad guidelines, it is not further specified what constitutes a "significant" increases in credit risk. It is therefore at the discretion of banks to concretize this term.

2.2 Differences to the Basel Accords and US-GAAP

The expected credit loss model in IFRS 9 is very similar to the current expected credit loss (CECL) model in US-GAAP. The most striking difference is that the latter does not require a stage allocation. Instead, lifetime expected credit losses have to be recognized from the onset. This implies that it is not necessary to estimate expected losses at the level of individual exposures, and less complex portfolio models can be used (Breeden and Vaskouski 2020). Other material differences are related to the scope of the impairment models and the treatment of financial instruments that are already credit-impaired at purchase (Hashim et al. 2016), which are, however, beyond the scope of our paper.[4]

The calculation of expected losses for credit risk is also necessary under the Basel Accords, among other things as part of the more extensive requirements for calculating regulatory capital. There are basically two approaches available to banks for this purpose (BCBS 2019, CRE 20.1). In the standardised approach, each exposure is assigned a predefined risk weight, on the basis of which the minimum requirements for regulatory capital are calculated. Under the internal ratings-based (IRB) approach, which is subject to regulatory approval, banks are allowed to base the risk weights on their own risk component estimates, like the PD and the LGD (CRE 30.1).[5] Various supervisors have indicated that the models developed for regulatory purposes may also be used as a starting point for IFRS 9 (BCBS 2015; EBA 2017) and some contributions in our sample also address the adjustment of regulatory default probabilities for purposes of IFRS 9 (e.g. Rubtsov and Petrov 2016; Miu and Ozdemir 2017; Bellini 2019, p. 148). Hence, we briefly discuss key differences between IFRS 9 and the IRB approach in determining the PD.

The most discernible difference is the time horizon of the PD.[6] In regulation, it is limited to one year (CRE 32.3), whereas in accounting it covers the remaining lifetime, provided there has been a substantial increase in credit risk since origination. Another difference concerns the calibration of the PD. In the Basel Accords, this takes place in a two-step process. In the first step, exposures have to be grouped. Retail exposures have to be pooled according to suitable segmentation criteria, while corporate, sovereign and bank exposures are assigned a internal rating (CRE 36.16–21). However, when assigning internal ratings, banks are expected to consider a time horizon that exceeds the one-year horizon of the PD (CRE 36.29) and to assess the borrowers' performance in adverse economic conditions (CRE 36.30). In the second step, each rating class or pool is mapped to a PD, which shall ideally be based on a long-run average of annual default rates (CRE 36.64). In contrast, IFRS 9 does not explicitly address the PD calibration. In line with the general principles for the measurement of expected credit losses, it shall just be measured "in a way that reflects [...] reasonable and supportable information [...] about past events, current conditions and forecasts of future economic conditions"

4. See ESRB (2019), Giner and Mora (2019) and Yang and Chen (2018) for more detailed comparisons of both impairment models, and Hashim et al. (2016) for a review of their development.

5. A further distinction is made between the foundation IRB approach, in which banks are only allowed to determine the PD internally, and the advanced IRB approach, in which they also provide the LGD and EAD (CRE 30.35).

6. Under the Basel Accords, a default is assumed to have occurred if either the obligor is unlikely to pay its credit obligations or if the obligor is more than 90 days past due (CRE 36.69). According to IFRS 9, an entity shall use its internal definition of default, which is constrained by the rebuttable presumption that a financial instrument is in default if it is 90 days past due (IFRS 9.B5.5.37).

(IFRS 9.5.5.17). However, as described above, IFRS 9 requires to determine both expected credit losses and the SICR test at the instrument level. Hence, in order to determine whether a instrument has experienced a significant increase in its risk of default, a PD model should ideally be able to measure such an increase in the first place. In our view it would therefore be consistent to estimate the PD directly at the account level and not at the aggregated level of a rating class or a portfolio, as prescribed in the Basel Accords.

To summarize these requirements, it is commonly said that IFRS 9 requires "point-in-time" (PIT) estimates, whereas the Basel Accords require "through-the-cycle" (TTC) estimates (e.g. EBA 2015, 2020; Cohen and Edwards 2017; Restoy and Zamil 2017). However, these terms can also refer to both the borrower-specific PDs and ratings or to the pooled PD, which turns out to be a common source of confusion about the differences and similarities between IFRS 9 and the Basel Accords (Eder 2021). To illustrate this point, it is useful to clarify the relationship between scores, ratings and the two different types of PD. Both scores and ratings are an opinion about a borrower's relative credit quality, while PDs quantify a borrower's absolute risk of default. Aggregated PDs based on (internal) ratings are usually calibrated in three steps. First, logit or Cox models are used to regress observed defaults on a set of explanatory variables (Thomas et al. 2017). The estimated predictor of this regression is subsequently transformed into a score, which is also used by banks in their lending decisions. In the second step, these scores are again regressed against the observed defaults, which creates a mapping between scores and borrower-specific PDs. Based on this regression, different bins are then defined, which are used to determine the rating classes. The bins are thereby usually chosen in such a way that the default risk increases exponentially with the rating classes. In the last step, an aggregated PD is determined for each rating class, for example based on the historical average of its default rates. Thus, even where exposure-specific PDs are available, these are replaced by the aggregated PD of the corresponding rating class (Bellini 2019, p. 45).[7]

PIT and TTC originally referred to the first step of the rating process (Treacy and Carey 1998). In a PIT rating philosophy, the ex-ante estimated PDs, and the ratings derived from them, change according to the state of the economy. The ex-post observed default rates of each rating class should ideally remain constant over time. Under a TTC rating philosophy, a coinage of the rating agencies, an attempt is made to provide long-run ratings that are forward-looking and remain stable over time. The ex-post observed default rates of each rating class should therefore fluctuate around their long-term average (Heitfield 2005). Apart from these general descriptions, however, there is no consensus about the precise meaning of these rating philosophies, especially with regard to TTC ratings (Mayer and Sauer 2017; Varetto 2018). Moreover, PIT and TTC are not only used as rating philosophies. They can also refer to the aggregated PD of a rating class or portfolio. Regardless of the underlying rating philosophy, the long-run average of default rates is referred to as TTC PD, whereas the most recently observed default rate or the one expected for the next year are referred to as PIT PD.

Also in the literature, both readings of PIT and TTC are used to distinguish IFRS 9 and the Basel Accords. On the one hand, some contributions argue that PIT ratings have to be used for IFRS 9

7. In retail banking, exposures are often not rated, but grouped into homogeneous portfolios. Here too, however, the PD of a portfolio should be determined on the basis of the long-run average of observed default rates.

and TTC ratings for the Basel Accords (e.g. Novotny-Farkas 2016; Krüger, Rösch, et al. 2018). On the other hand, there are also a number of proposals to transform aggregated, regulatory "TTC" default probabilities into "PIT" estimates for the purposes of IFRS 9, which are quite common in practice (Rubtsov and Petrov 2016, e.g. Skoglund 2017; Bellini 2019, p. 148). However, as we have argued above, it would be more reasonable for IFRS 9 to estimate the PD directly at the account level, whereas the Basel Accords prescribe an indirect calibration at the portfolio level. For this reason alone, the mere distinction between PIT and TTC seems to be misleading. In the following, we will therefore refrain from using this terminology. Moreover, in section 9, we will argue that it would be more consistent to reverse the widespread process of adjusting regulatory PDs for purposes of IFRS 9. Instead, models should first be developed that estimate a borrower's PD at the account level in accordance with IFRS 9. These models could then be used, at least as a starting point, to determine long-term ratings as required by the Basel Accords.

3 Literature Search

In order to ensure a comprehensive sampling and identification of relevant contributions on the estimation of default probabilities for IFRS 9 and the stage allocation, we started with an extensive search on various databases, including ScienceDirect, Springer, Wiley, Taylor and Francis, Emerald Insight, EBSCO, risk.net and Google Scholar. As keywords, we have used different combinations of IFRS 9, expected (credit) loss, probability of default, and SICR-test. We have preselected contributions based on their title and abstract. In the case of ambiguity we have examined them closer. The last time that we carried out a thorough search was on April 20, 2021.

To create an initial sample, we applied the following criteria. We only included publications in English that appeared since 2013, the year in which the first draft of the final impairment model was published. For quality reasons, we primarily included peer-reviewed articles published by renowned publishers. To a limited extent, we have also included books and working papers, which we have decided on a case-by-case basis. Other types of publications, such as conference papers and white papers, have been excluded. In terms of content, we have only included contributions that mention IFRS 9 or the expected credit loss model either literally or at least in a broader context.[8] Of these contributions, we have further selected only those that are concerned with the estimation of the PD or the implementation of the stage allocation.

We then iteratively added to this sample by conducting a backward search in the bibliographies of the contributions in our sample on the one hand, and a forward search for papers that cite them on the other. Articles that fulfilled the above criteria were also included. This resulted in a final sample of 52 contributions, which are summarized in the appendix in table 4. These include 47 peer-reviewed articles, two books (Baesens et al. 2017; Bellini 2019) and three working papers (Brunel 2016a; Xu 2016; Gaffney and McCann 2018). 45 of these contributions discuss models for estimating the PD. Nine contributions further mention approaches to transforming regulatory default probabilities. And nine contributions also address the implementation of the stage allocation.

8. Specifically, we have examined whether and in which context provisions, impairments or accounting are mentioned, where IFRS 9 was not mentioned directly.

Table 1: Probability of default models

Continuous time survival models	Non-parametric	Kaplan-Meier, Nelson-Aalen, Breslow estimator	(1)
	Semi-parametric	Cox's relative risk model	(2)
	Parametric	Weibull- and exponential models Accelerated failure time models Proportional odds model	(3) (4) (5)
	Extensions	Competing risk models Mixture cure rate models	(6) (7)
Discrete time survival models	Non-parametric	Life-table estimator Vintage models	(8) (9)
	GLM	Logit-, probit- and cloglog-models	(10)
	Extensions	Age-period-cohort models Competing risks models	(11) (12)
Matrix models	Rating-based	Cohort approach Duration approach	(13) (14)
		Roll rate models	(15)
Market-based models		Merton model Bond and CDS based models	(16) (17)
Adjustments of reg. PDs		Scalar approaches, Vasicek model	(18)

4 Structure and Notational Conventions

In the following sections we provide a general classification and comparison of models to forecast the PD for IFRS 9. As a cautionary remark we note that there is no approach that is superior to all other approaches in each and every circumstance. Hence, entities will have to develop separate models for different segments. Appropriate segments that are homogenous with respect to their underlying risk factors may be found, for example, by considering different types of counterparties, geographic regions, and types of financial products. The suitability of particular models is further determined, among other things, by the availability of representative (historical) data, aspects of computational efficiency, the required forecasting horizon, and the ease of communication to internal and external stakeholders.

We structure the models according to the data they use, i.e. whether they are based on historical data or market prices. The former are further classified according to the time-space and state-space they employ. First, we consider survival models (or event history models, respectively) in continuous time, and then in discrete time, where all states are terminating. Terminating states, such as default and complete pre-payment, cannot be left once they are reached. Thereafter, we consider matrix models in discrete time and continuous time, which in addition allow for transient states. Unlike terminating states, transient states can be left again when they are reached. The chapter ends with a discussion of approaches that are used to transform regulatory PDs for purposes of IFRS 9. Table 1 provides an overview of the models that will be outlined below. A list of the corresponding contributions can be found in the appendix in table 4.

To facilitate our exposition, we slightly extend the notational conventions of Crook and Bellotti (2010). Specifically, we let t denote calendar time, v_i the origin or vintage time, and τ_i the duration, respectively age, of exposure i under consideration. Indexes i ($i = 1, \ldots, N$) can refer to both financial instruments or the corresponding counterparties, which we refer to generically as exposures or accounts. Vectors of variables are denoted with $\mathbf{x}_{i\tau}$ if they are exposure-specific and time-varying, with $\mathbf{w}_i$ if they are exposure-specific but constant over time (or at least only observed at origination), and with $\mathbf{z}_t$ if they are not exposure-specific but time-varying, such as macroeconomic variables. Both $\mathbf{x}_{i\tau}$ and $\mathbf{z}_t$ usually involve lags of differing length, but we do not explicitly include them in our notation for brevity's sake. The parameters β_0 and $\beta_{0\tau}$ denote (age-specific) intercepts, and β_1, β_2, and β_3 are vectors of parameters, with the further convention that β_1 is related to $\mathbf{w}_i$, β_2 to $\mathbf{x}_{i\tau}$, and β_3 to $\mathbf{z}_t$. Of course, it will often be possible to include interaction terms of variables as well, but we'll omit them here as well.

5 Continuous Time Survival Models

In order to facilitate our exposition, we'll introduce some basic concepts of multi-period PDs models in continuous time, in addition to the naming conventions above. For the ease of exposition, we will introduce these concepts in a simple, discrete state space, where we only distinguish between default and survival. However, extensions of this framework to more states are possible, and will be introduced where necessary below.

An exposure's time to default T_i can be thought of as a random variable.[9] Its cumulative distribution function is given by $F_i(\tau) = P(T_i \leq \tau)$, which can be interpreted as the probability that the lifetime will not exceed τ. The equivalent survival function gives the probability that the exposure's lifetime T_i is larger than τ, i.e. $S_i(\tau) = P(T_i > \tau)$ with $S_i(0) = 1$ and $S_i(\infty) = 0$. Alternatively, the random variable may be represented by its hazard or intensity function:

$$\lambda_i(\tau) = \lim_{\Delta\tau \to 0} \frac{P(\tau \leq T_i < \tau + \Delta\tau | T_i \geq \tau)}{\Delta\tau} = \frac{f_i(\tau)}{S_i(\tau)} = \frac{f_i(\tau)}{1 - F_i(\tau)},$$

which is strictly positive, and can be interpreted as the instantaneous rate of default, conditional on not having defaulted before. Integrating the intensity function gives the cumulative intensity function:

$$\Lambda_i(\tau) = \int_0^\tau \lambda_i(u) du,$$

from which the survival function can be obtained again:

$$S_i(\tau) = \exp\big(-\Lambda_i(\tau)\big) = \exp\left(-\int_0^\tau \lambda_i(u) du\right).$$

Therefore, if one of these quantities is known, the others can be readily computed. An exposure's conditional probability of default in the period between $\tau - 1$ to τ can then be readily computed by

9. Instead of considering the time to default as a random variable, the default process can also be seen as a counting process that counts the number of events over time (Kiefer and Larson 2015). Important generalizations of classical survival models such as models with time-varying variables or competing risk models are based on this counting process view. For a general understanding, however, this is not decisive, which is why we do not discuss counting processes further.

the following equation:

$$PD^c_{i\tau} = \frac{S_i(\tau-1) - S_i(\tau)}{S_i(\tau-1)},$$

and the marginal probability of default over the same period is given by:

$$PD^m_{i\tau} = S_i(\tau-1) - S_i(\tau-1).$$

However, the estimation of survival functions and the corresponding PDs poses particular challenges for statistical models. This is not only because duration times have to be strictly positive. More challenging is that survival data are censored, meaning that the value of the dependent variable cannot be observed. *Right censoring* occurs with exposures that did not default yet, but still may default in the future. Right censored observations therefore only contribute information on survival, but not about the time of default. Since it is easier to deal with this type of censoring in the survival function than in the cumulative probability function, survival analysis is focused around the former (Box-Steffensmeier and Jones 2004). Moreover, survival models are often expressed in terms of the hazard function, since it allows to assess how exposures' failure rates change as they age (Kalbfleisch and Prentice 2002). Another type of censoring that arises frequently in credit risk applications is *interval censoring*. This means that it is only known that a default happened between two points in time, but not exactly when. In this case, it is still possible to analyse data with continuous-time survival models, but it may be more efficient to use discrete-time models. Taken together, these challenges imply that the usual OLS model will yield biased estimates, and specialized techniques have to be employed. In this section we discuss survival models in continuous time. Starting with non-parametric models, progressively more structure is imposed, leading to fully parametric models. The remaining models in this section represent extensions of these basic survival models.

5.1 Non-parametric Estimators

Non-parametric estimators provide a univariate description of survival data, and three of them have been mentioned within our literature sample. The most well-known among them is the so-called product limit estimator or Kaplan-Meier estimator (Kaplan and Meier 1958) which directly estimates the empirical survival function:

$$S(\tau) = \prod_{\tau_j \leq \tau}^{m} \left(1 - \frac{d_j}{n_j}\right),$$

where $\tau_1 < \tau_2 < \cdots < \tau_m$ are distinct, ordered times of default, d_j is the number of defaults at τ_j, and n_j is the number of active accounts just before τ_j (Bellini 2019, p. 112; Brunel 2016a). Another non-parametric estimator, mentioned in Brunel (2016a), is the Nelson-Aalen estimator of the cumulative intensity function:

$$\Lambda(\tau) = \sum_{\tau_j \leq \tau}^{m} \frac{d_j}{n_j}.$$

It can be transformed into the survival function via the Breslow estimator ($S(\tau) = \exp(-\Lambda(\tau))$).

Non-parametric estimators provide simple and intuitive estimates (McPhail and McPhail 2014).

But they are also quite descriptive as the term-structure is rather an input than an output (Skoglund 2017). Hence, it is also not possible to extrapolate the term-structures beyond the longest maturity observed. Besides this, the capability of these non-parametric estimators to take into account explanatory variables is limited, which makes them susceptible to shifts in internal and external factors. The only possibility to include explanatory variables is to stratify the estimation according to a factor variable, such that a single term structure is estimated for each level of that variable (Baesens et al. 2017, p. 188). The underlying portfolio should therefore be as homogeneous as possible, which can be achieved through appropriate segmentation. However, in order to take macroeconomic variables into account, manual overrides would still be necessary (Bellini 2019, p. 116). In the context of IFRS 9, these models are therefore rather useful as diagnostic tools.

5.2 Semi-parametric Models

The limited ability of purely non-parametric models to incorporate covariates can be overcome with semi-parametric models. The most well-known of these is Cox's relative risk model (Cox 1972) which has the following intensity function:

$$\lambda(\tau|\mathbf{x}_{i\tau}, \mathbf{w}_i, \mathbf{z}_t) = \lambda_0(\tau)\exp(\beta_1\mathbf{w}_i + \beta_2\mathbf{x}_{i\tau} + \beta_3\mathbf{z}_t).$$

$\lambda_0(\tau)$ is thereby an arbitrary, unspecified baseline intensity, which is common to all exposures, and the covariates have a multiplicative effect on it. If only time-invariant covariates are used, they simply shift the baseline intensity up or down. The major contribution of Cox (1972) was the insight that the estimation problem of this model splits into two parts. Since the estimation of the parameter vector β can be separated from the baseline intensity $\lambda_0(\tau)$, the effect of the covariates can be assessed even without estimating the baseline intensity function. However, in order to calculate the survival function, it is necessary to estimate the baseline intensity as well. For this purpose, adapted versions of the aforementioned non-parametric estimators are used.

Cox relative risk model is widely applied in research to study defaults and bankruptcies, and also mentioned in several contributions within our sample. Brunel (2016a) explains the model for the case with time-invariant covariates $\mathbf{w}_i$, whereas Baesens et al. (2017, p. 191–202) and Bellini (2019, p. 116) cover also the case of time-varying covariates.[10] Furthermore, Skoglund (2017) considers the following stratification of Cox's relative risk model:

$$\lambda_r(\tau|\mathbf{z_t}) = \lambda_{0r}(\tau)\exp(\beta_3\mathbf{z}_t),$$

where all subject-specific information is captured by the $r = 1, \ldots, R$ application risk grades, along macroeconomic variables which are contained in the vector $\mathbf{z}_t$. However, Skoglund (2017) also points out that a reliable estimation of β_3 would require data over several economic cycles, which are usually not available in practice.

In summary, Cox relative risk model allows to quantify the impact of explanatory variables,

10. Cox relative risk model is often called proportional hazard model. However, this term is only accurate if the model doesn't contain any time-varying variables (Djeundje and Crook 2019).

which can also be time-varying, without the need to make a potentially erroneous assumption about the functional form of baseline intensity function. This is its major advantage in many research applications. But in practice, this may also be seen as its major limitation, because it does not allow extrapolating survival curves beyond the maximum observed maturity. Besides this, it may not be possible to reliably estimate a smooth baseline intensity function, if only small and short historical samples are available. In addition, the estimates of the PD will generally be biased if a large proportion of the exposures is prepaid. In this case, it is advisable to use competing risk models (Blumenstock et al. 2020).

5.3 Parametric Models

In order to extrapolate the survival curve beyond the maximum observed survival time, it is necessary to resort to parametric regression models, which are commonly categorized into three model families: Cox's relative risk models, accelerated failure time models, and proportional odds models.

Parametric variants of Cox's relative risk model make an explicit assumption about the distribution of the survival time T_i, which restricts the shape of the baseline intensity $\lambda_0(\tau)$. For example, if one assumes that T_i follows a Weibull distribution, one obtains the following Weibull regression model (Kalbfleisch and Prentice 2002):

$$\lambda(\tau|\mathbf{x}_{i\tau}, \mathbf{w}_i, \mathbf{z}_t) = \lambda_0\gamma(\lambda_0\tau)^{\gamma-1}\exp(\beta_1\mathbf{w}_i + \beta_2\mathbf{x}_{i\tau} + \beta_3\mathbf{z}_t),$$

where λ_0 ($\lambda_0 > 0$) is a constant scale parameter, and γ ($\gamma > 0$) is a constant shape parameter. If $\gamma > 1$ ($\gamma < 1$), the baseline function is monotone increasing (decreasing) over an exposure's lifetime. However, the assumption of monotone intensities is often too restrictive in credit risk applications. For $\gamma = 1$, one obtains the exponential model as a special case.[11] However, the latter assumes that the baseline intensity is constant, which is typically too restrictive in credit risk applications.

Most parametric survival models that are used in practice belong to the family of accelerated failure time (AFT) models. These models are usually represented as a linear model on the log-transformed survival time, whereby the covariates have an accelerating or decelerating effect on the survival time (Baesens et al. 2017, p. 202):

$$\ln T_i = \beta_0 + \beta_1\mathbf{w}_i + \sigma\epsilon_i,$$

where β_0 is a constant intercept, σ is a constant scaling parameter, and ϵ_i is a i.i.d. error term which is assumed to follow a particular distribution. The choice of the distribution for the error terms ϵ_i determines the distribution of the survival times T_i, from which different regression models can be obtained. The most common choices are summarized in table 2. For example, the Weibull model (and exponential model) can also be parameterized as an AFT model as long as the model does not contain any time-varying variables. Compared to parametrization as a Cox model, only the regression coefficients change (Lawless 2003). Other models that give rise to either monotonically increasing

11. In case of the exponential regression model, the baseline hazard λ_0 is sometimes included in the vector of regression coefficients, i.e. $\lambda_0 = \exp(\beta_0)$.

Table 2: AFT model specifications

Regression model	Error term distribution
Exponential regression	Extreme value (Gumbel)
Weibull regression	Extreme value (Gumbel)
Log-normal regression	Standard normal
Log-logistic regression	Logistic
Generalized gamma regression	Log-generalized gamma

or decreasing intensities are the log-normal regression and the log-logistic regression. Even more flexible is the generalized gamma regression, which allows for hump-shaped intensities that often arise in retail lending (Batiz-Zuk et al. 2020).

Within our literature sample, AFT models are featured in the books by Baesens et al. (2017, p. 202) and Bellini (2019, p. 121). Several AFT-models are also applied in the studies by Batiz-Zuk et al. (2020) and Krüger, Oehme, et al. (2018). The latter model for example the joint distribution of default times and recovery rates for corporate bonds, which is only possible with parametric survival models. However, a particular limitation of AFT models is that they can only accommodate time-invariant covariates $\mathbf{w}_i$, i.e. information which is already available at the time of origination. In principle, it is also possible to include time-varying covariates in AFT models, but this is not as straightforward and seldom implemented in readily available software packages (Baesens et al. 2017, p. 208). In contrast, Cox's relative risk model can easily incorporate time-varying variables as well, and therefore seems to be more suitable for purposes of IFRS 9.

The last family of parametric survival models are proportional odds (PO) models. In these models, the covariates shift the odds of default by a given age up or down:

$$\frac{S(\tau|\mathbf{w}_i)}{1 - S(\tau|\mathbf{w}_i)} = \frac{S_0(\tau)}{1 - S_0(\tau)} \exp(\beta_1 \mathbf{w}_i),$$

where $S_0(\tau)$ is a baseline survival function, for which an appropriate distributional assumption has to be made. The models owe their name to the fact that the difference in the log-odds of the survival rate between any two exposures remains constant over time, i.e.:

$$\ln \frac{S(\tau|\mathbf{w}_1)}{1 - S(\tau|\mathbf{w}_1)} - \ln \frac{S(\tau|\mathbf{w}_2)}{1 - S(\tau|\mathbf{w}_2)} = \beta_1(\mathbf{w}_2 - \mathbf{w}_1).$$

In the literature on IFRS 9, the proportional odds model has so far only been applied by Orth (2013) who assumes that the conditional distribution of an exposure's lifetime follows a log-logistic distribution. In this case, the hazard function is given by:

$$\lambda(\tau|\mathbf{x}_{i\tau}, \mathbf{w}_i, \mathbf{z}_t) = \frac{\gamma \exp(\beta_1 \mathbf{w}_i + \beta_2 \mathbf{x}_{i\tau} + \beta_3 \mathbf{z}_t)^\gamma \tau^{\gamma - 1}}{1 + [\exp(\beta_1 \mathbf{w}_i + \beta_2 \mathbf{x}_{i\tau} + \beta_3 \mathbf{z}_t)\tau]^\gamma},$$

where γ determines its shape. Although proportional odds models can also be parameterized with other distributions, the log-logistic distribution is by far the most common choice. In the survival literature, the terms are therefore often used synonymously. Furthermore, as long as the model does not include time-varying variables, the log-logistic regression can also be parameterized as an AFT model. Orth (2013) motivates the choice of the PO model instead of Cox's relative risk model by empirical evidence that the hazard ratios across obligors are not constant.

Overall, it is widely acknowledged that parametric survival models can provide reasonable forecasts of default probabilities. Baesens et al. (2017, p. 209) argue that they are more accurate than discrete-time hazard models, which are introduced in the next section, but more complex and less transparent. Besides this, sufficient historical data has to be available in order to obtain reasonable forecasts, which may be a challenge especially for low-default portfolios (McPhail and McPhail 2014; Skoglund 2017). However, such remarks apply to almost any credit risk model, and are not a particular limitation of parametric survival models.

5.4 Competing Risk Models

The models covered so far have in common that the only events they consider are default and survival. Other terminating events, which are called competing risks, are not taken into account. Examples for such competing risks are complete pre-payments or mergers and acquisitions. However, estimates of survival models will be strongly biased, if these competing risks are correlated with the risk of default. One possibility to account for competing risks is to expand Cox relative risk model by estimating a cause-specific intensity function for each event (Baesens et al. 2017, p. 347; Blumenstock et al. 2020). However, this approach will still lead to biased estimates of default probabilities, as long as the competing risks are not independent of each other. To resolve this bias, a so-called sub-distribution hazard function has to be defined (Fine and Gray 1999), which is based on a different risk set. Using data on 600,000 single-family US mortgages, Blumenstock et al. (2020) recently found that the Fine-Gray model consistently outperforms the cause-specific Cox model. However, in the credit risk literature, it is more common to model competing risks in discrete time. We therefore postpone a further discussion of competing risk models to the next subsection.

5.5 Mixture Cure Rate Models

Survival models assume that in the long-run every exposure will default. However, if there is a large sub-population of exposures that will never experience a default event, i.e. in the presence of heavy right censoring, it may be more efficient to use mixture cure rate models. These models owe their name to the fact that their unconditional survival function is considered to be a mixture of two sub-distributions, one that describes the probability of being susceptible to default, and the other describing the probability distribution of the time to default. Hence, it is not only determined at which age an exposure will leave the portfolio, but also whether it will leave the portfolio before maturity at all or not. In a mixture cure rate model, the unconditional survival function is given by the following equation:

$$S(\tau|\mathbf{w}_{i1}, \mathbf{w}_{i2}, \mathbf{x}_{i\tau}, \mathbf{z}_t) = \pi(\mathbf{w}_{i1}) S(\tau|Y_i = 1, \mathbf{w}_{i2}, \mathbf{x}_{i\tau}, \mathbf{z}_t) + 1 - \pi(\mathbf{w}_{i1}),$$

where Y_i is a Bernoulli variable indicating whether a subject is still at risk of default or not, and $\mathbf{w}_{i1}$ and $\mathbf{w}_{i2}$ are vectors of static, explanatory variables, which may coincide or not. Furthermore, $\pi(\mathbf{w}_{i1}) = P(Y = 1|\mathbf{w}_{i1})$ is the fraction of exposures or borrowers that is susceptible to default, which is usually modelled using a logit regression, if it is made contingent on a set of explanatory variables, i.e.:

$$\pi(\mathbf{w}_{i1}) = \frac{\exp(\beta \mathbf{w}_{i1})}{1 + \exp(\beta \mathbf{w}_{i1})}.$$

Moreover, the conditional survival function $S(\tau|Y = 1, \mathbf{w}_{i2}, \mathbf{x}_{i\tau}, \mathbf{z}_t)$ is usually modelled using Cox relative risk model with time-fixed covariates, but other models are also available in the literature. For example, Dirick et al. (2019) recently extended the model to the case of time-varying covariates, both in the presence and absence of competing risks. However, solving mixture cure rate models is not straightforward, since it is unobservable whether an exposure belongs to the subpopulation of defaulting accounts or not. Hence, expectation maximization (EM) algorithms have to be used, which optimize the model iteratively.

Within our literature sample, mixture cure rate models are only applied by Dirick et al. (2019) and briefly explained by Baesens et al. (2017, p. 208) and Brunel (2016a). They may be especially applied to retail portfolios, where it is a stylized fact that the majority of exposures does not default, such that ordinary event history models would overestimate the default probabilities of exposures with a very low credit risk. However, the assumption that an exposure is immune to default could quickly turn out to be wrong if economic conditions worsen. Hence, these models may be scrutinized by auditors and regulators.[12]

6 Discrete Time Survival Models

Although the event of default can happen at any point in time, its exact timing is usually not observable in practice. Only the interval of its occurrence is known. Discrete-time data may also arise for technical reasons, for example, when default is defined only at certain repayment dates. Irrespective of the underlying reasons, it may therefore be more efficient to use discrete time models instead of continuous time models. The difference between continuous time models and discrete time models is that the former focus on the duration until some event happens, whereas the latter model the probability that the event of interest happens during a specific time interval. Let the time until exposure i defaults be a non-negative, discrete random variable T_i whose cumulative density function is given by $F_i(\tau) = P(T_i \leq \tau)$. In the discrete time framework, the hazard function is than given by:

$$h_i(\tau) = PD^c_{i\tau} = P(T_i = \tau|T_i \geq \tau),$$

which can be interpreted as the conditional probability of default within period τ ($PD^c_{i\tau}$), given that exposure i did not default before. The cumulative hazard function is given by $\Lambda_i(\tau) = \sum_{s=1}^{\tau} h_i(\tau)$,

12. For the sake of completeness, we mention that there also exist mixture cure rate models in discrete time (De Leonardis and Rocci 2014). However, these have not been mentioned in the literature on IFRS 9 yet. Since the motivation for these models is basically the same as for their continuous-time counterparts, we will restrain from further explanations of these models below.

and the corresponding survival function is:

$$S_i(\tau) = P(T_i > \tau) = \prod_{s=1}^{\tau} (1 - h_i(s)),$$

which is again the probability that the time of default is larger than τ. The cumulative PD up to period τ is therefore $1 - S_i(\tau) = F_i(\tau)$ and the marginal PD in period τ is $PD^m_{i\tau} = F_i(\tau) - F_i(\tau - 1) = S_i(\tau - 1)h_i(\tau)$.

6.1 Non-parametric Estimators

The simplest discrete time model is the life-table estimator, which is analogous to the Kaplan-Meier estimator in continuous time. The life-table estimator is a non-parametric estimator of the survival function, which assumes that censoring occurs in the mid of each period:

$$S(\tau) = \prod_{j=1}^{\tau} \left(1 - \frac{d_j}{[n_j + (n_j - w_j)]/2}\right) = \prod_{j=1}^{\tau} \left(1 - \frac{d_j}{n_j - w_j/2}\right),$$

where d_j is the number of defaults during period τ, n_j is the number of active accounts at the beginning of period τ, and w_j the number of accounts that have been withdrawn during this period. The life-table estimator is recommended by Baesens et al. (2017, p. 183) as an alternative to the Kaplan-Meier estimator, if there are many unique event times. Otherwise, it is subject to the same limitations as the non-parametric estimators is continuous time. In particular, it offers only limited possibilities to take into account explanatory variables, and is therefore mainly useful as a diagnostic tool (Schutte et al. 2020).

Another type of non-parametric models in discrete time are vintage models, which are popular in retail banking. Intuitively, these models plot for each cohort (vintage) of exposures the fraction of defaulted exposures against time since origination, which is usually measured on a monthly time scale (McPhail and McPhail 2014). Vintage models therefore provide a direct estimate of the empirically observed cumulative default probability of each cohort. A cohort is thereby defined as a group of exposures that were issued in the same period, like a given calendar year. In this respect, vintage models are closely related to the life-table estimator and non-parametric survival models in continuous time. If one were to stratify the latter according to the cohort, one would obtain the survival function, which is just the opposite of the cumulative default probability. However, they will not be perfectly equivalent, because unlike the survival models, simple vintage models ignore censoring, and they are based on portfolio data instead of account level data. Moreover, in order to obtain forecasts of the PD from vintage models, the historical term structures can be extrapolated, for instance by taking their average. Yet, these have to be adjusted again with manual overrides in order to take macroeconomic developments into account.

A more robust variant of vintage models are so-called age-period-cohort (APC) models. These are traditionally based on aggregated portfolio data (Brunel 2016a; Breeden and Vaskouski 2020), but loan-level implementations are possible as well (Breeden 2016; Breeden and Crook 2020). The latter are more relevant for IFRS 9 and will be explained below.

6.2 Parametric Models

Survival models in discrete time that can accommodate explanatory variables can be embedded in the framework of generalized linear models (GLM) (McCullagh and Nelder 1989). Given an appropriate data structure (Bellini 2019, p. 105), the conditional probability of default at age τ can be estimated by the following specification of the hazard function:

$$h(\tau|\mathbf{w}_i, \mathbf{x}_{i\tau}, \mathbf{z}_t) = f(\beta_{0\tau} + \beta_1\mathbf{w}_i + \beta_2\mathbf{x}_{i\tau} + \beta_3\mathbf{z}_t),$$

with response function $f()$. Equivalently, these models can be represented as follows:

$$g(h(\tau|\mathbf{w}_i, \mathbf{x}_{i\tau}, \mathbf{z}_t)) = \beta_{0\tau} + \beta_1\mathbf{w}_i + \beta_2\mathbf{x}_{i\tau} + \beta_3\mathbf{z}_t,$$

where $g() = f^{-1}()$ is the so-called link-function. These models are in close analogy to Cox semi-parametric model above, since the time-varying intercepts ($\beta_{0\tau}$) resemble the baseline intensity. More specifically, if one chooses the complementary log-log (cloglog) link function, one obtains the following model (Baesens et al. 2017, p. 153; Djeundje and Crook 2019):

$$h(\tau|\mathbf{w}_i, \mathbf{x}_{i\tau}, \mathbf{z}_t) = 1 - \exp[-\exp(\beta_{0\tau} + \beta_1\mathbf{w}_i + \beta_2\mathbf{x}_{i\tau} + \beta_3\mathbf{z}_t)],$$

which a discretized version of Cox's relative risk model. However, unlike in the continuous time setting, the parameters of the baseline intensity are now jointly estimated with the other parameters. Another popular choice is the probit link ($g() = \Phi^{-1}()$), which is used in the literature on IFRS 9 among others by Gürtler et al. (2018) and Krüger, Rösch, et al. (2018). Yet, in credit risk applications, most often a logit link is chosen, which gives rise to the following model (Shumway 2001; Chava and Jarrow 2004; Hillegeist et al. 2004; Campbell et al. 2008):

$$h(\tau|\mathbf{w}_i, \mathbf{x}_{i\tau}, \mathbf{z}_t) = \frac{1}{1 - \exp([\beta_{0\tau} + \beta_1\mathbf{w}_i + \beta_2\mathbf{x}_{i\tau} + \beta_3\mathbf{z}_t)}.$$

The logit link is popular both in practice and research, for example to estimate credit scores, and therefore well understood and comparatively easy to communicate. This is also reflected in our literature sample, where the logit model is often mentioned (Altman et al. 2016, 2020; Baesens et al. 2017, p. 152; Bellini 2019, p. 105; Brunel 2016a; Chen et al. 2020; Kalotay and Altman 2017; McPhail and McPhail 2014; Xu 2016). Although parameter interpretations change with the link function, the models are usually found to yield similar predictions with respect to the probability of default. For this reason the choice of the link function is often seen as a matter of personal taste (Baesens et al. 2017, p. 163).

The models above can be flexibly parameterized. For instance, to estimate the baseline intensity as specified above, it would be necessary to include $\tau - 1$ dummy variables, which substantially increases the number of parameters. In order to reduce them, log-transformations, polynomials or spline functions are often used, which simultaneously smooth the baseline intensity.[13] At the same

13. For example, if the dummy variables $\beta_{0\tau}$ are replaced by the natural logarithm of τ, one obtains a discrete accelerated failure time (AFT) model (Shumway 2001).

time, the other parameters need not be to constant over time. Djeundje and Crook (2019) allow them to be a function of duration time (i.e. $\beta_1(\tau), \ldots, \beta_3(\tau)$). To maintain a parsimonious parametrization, they again use polynomials and spline functions. Using discrete Cox models to predict credit card defaults, they find that varying coefficients models outperform those with constant coefficients. Moreover, to predict future default probabilities, discrete survival models do not necessarily require specifying the time-series dynamics of time-varying variables. Instead, it is also possible to estimate as many lagged versions of the model as forecast periods are necessary and to insert the current state of all variables into the corresponding prediction models (Campbell et al. 2008; Xu 2016; Breeden and Crook 2020). Such an approach, however, is computationally intensive and the number of parameters to be estimated is high (Orth 2013).

Given that credit risk data is usually observed in discrete time intervals, it is natural to use discrete-time survival models. In comparison to models in continuous time, they achieve a comparable performance. In the presence of heavy interval censoring, they tend to be even more efficient and accurate. Another advantage of discrete time models over continuous time models is that they are somewhat simpler and more transparent, making them easier to communicate (Baesens et al. 2017, p. 209). Besides this, their ability to easily account for time-varying covariates is often highlighted as another strength (Djeundje and Crook 2019).

6.3 Age-period-cohort Models

Age-period-cohort (APC) models at the loan level can be thought of as a combination of discrete time survival models and vintage analysis. At the portfolio level, these models decompose the log-odds of the observed (monthly) default rate ($DR(\tau, v, t)$) at age τ in period t for exposures with common origination time (cohort) v as follows (Breeden and Vaskouski 2020):

$$\ln\left(\frac{\mathrm{DR}(\tau, v, t)}{1 - \mathrm{DR}(\tau, v, t)}\right) = F(\tau) + G(v) + H(t),$$

where $F(\tau)$, $G(v)$ and $H(t)$ are non-parametric functions of the corresponding time dimensions, which are assumed to be independent. Specifically, $F(\tau)$ is the life-cycle function that is akin to the hazard rate and indicates how the default risk changes with age τ. The vintage function $G(v)$ measures credit quality as a function of origination date v. It captures shifts in the performance across cohorts, which may be driven for instance by changes in internal lending practices. And $H(t)$ is the environment function which captures changes in exogenous, macroeconomic conditions that affect the default risk of all cohorts at a given calendar time t alike.

A loan-level implementation of these models, which also takes into account explanatory variables, has been developed by Breeden (2016). In the first step, an exposure's (transformed) probability of default is again decomposed into the three aforementioned functions:

$$\ln\left(\frac{p_i(\tau, v, t)}{1 - p_i(\tau, v, t)}\right) = F(\tau) + G(v) + H(t).$$

These generic functions can be either estimated with spline functions or a set of dummy variables.

Each of them can be represented in terms of a constant, a linear trend and non-linear terms:

$$
\begin{aligned}
F(\tau) &= \alpha_0 + \alpha_1 \tau + F'(\tau) \\
G(v) &= \beta_0 + \beta_1 v + G'(v) \\
H(t) &= \gamma_0 + \gamma_1 t + H'(t).
\end{aligned}
$$

If one puts these equations into the model above, one would get three intercepts and three linear terms, which results in two well-known specification errors. First, it is only possible to estimate a single constant. Breeden (2016) therefore assumes that the intercepts of the vintage function (β_0) and the environment function (γ_0) are zero, such that these functions are measured relative to the life-cycle function. The second specification error results from the perfect linear dependence of the three time dimensions ($\tau = v + t$). Because of this, only two linear trends can be estimated. In analogy to the usual assumption that a long-run probability of default exists, Breeden therefore assumes that the environment function has no linear trend ($\gamma_1 = 0$).

After the functions have been estimated, the values of the environmental function are regressed against macroeconomic variables. Additionally, the life-cycle function and the vintage function are adjusted for linear trends. Finally, another logit model is estimated in which the vintage function is replaced by standard scoring attributes w_i and the values of the life-cycle function and the environmental function are used as offsets:

$$
\ln\left(\frac{p_i(\tau, v, t)}{1 - p_i(\tau, v, t)}\right) = \text{offset}(\tau, t) + \beta_1 \mathbf{w}_i.
$$

To forecast the conditional PD, the exposures are shifted along the life-cycle function and the environment function is updated on the basis of macroeconomic scenarios.

A slightly different loan-level implementation of APC models has been proposed by Breeden and Crook (2020). Unlike in the specification above, they also take into account prepayments as a competing risk and include days past due as a time-varying variable. They find that taking into account the days past due increases the short-term forecasting accuracy of the model, but also point out that it complicates the model estimation.

According to Breeden (2016), APC models have a more robust out-of-sample performance than other survival models. This is achieved by explicitly addressing the linear specification error that is implicit in all survival models which include both static and time-varying variables. Moreover, if days past due are taken into account as explanatory variable, it is possible to achieve the short-term accuracy of roll rate models while maintaining the long-term performance of vintage models (Breeden and Crook 2020).

6.4 Competing Risk Models

As already mentioned above, ordinary survival models do not take into account other terminating events, like complete pre-payments or mergers and acquisitions. Hence, the estimates will in general be biased, if these competing risks are not accounted for. The most common competing risk model in discrete time are multinomial logit models, which are covered in the literature on IFRS 9 by

Baesens et al. (2017, p. 340), Bellini (2019, p. 231), and Lee et al. (2021). A generic specification of this model is given by:

$$h_r(\tau|\mathbf{w}_i, \mathbf{x}_{i\tau}, \mathbf{z}_t) = \frac{\exp(\beta_{0\tau r} + \beta_{1r}\mathbf{w}_i + \beta_{2r}\mathbf{x}_{i\tau} + \beta_{3r}\mathbf{z}_t)}{1 + \sum_{j=1}^{R} \exp(\beta_{0\tau j} + \beta_{1j}\mathbf{w}_i + \beta_{2j}\mathbf{x}_{i\tau} + \beta_{3j}\mathbf{z}_t)},$$

which gives the conditional probability of risk r for $j = 1, \ldots, R-1$ competing risks. The parameters $\beta_{0\tau j}$ constitute a cause-specific baseline hazard function, and β_j are cause-specific vectors of coefficients. The R-th risk, the so-called reference category, is conditional survival, whose probability is given by one minus the sum over the conditional probabilities of all other risks $(1 - \sum_{r=1}^{R-1} h_r(\tau|\mathbf{w}_i, \mathbf{x}_{i\tau}, \mathbf{z}_t))$. Other competing risk models in discrete time are discussed by Xu (2016), Breeden and Vaskouski (2020), and Breeden and Crook (2020), and qualitative discussions can be found in Skoglund (2017) and Brunel (2016a). In practice, the number of competing risks is usually low, and it is only distinguished between default, (complete) prepayment, and survival.

7 Matrix-based Models

The models presented so far have in common that they directly specify the default process and that all states are terminating. This means that default, early repayments as well as mergers and acquisitions cannot be left once an exposure has reached them. In contrast, matrix models specify the default process only indirectly and the state space is extended by transient states. Matrix models can therefore also be understood as an extension of competing risk models. However, unlike terminating states, transient states can be left again once they have been reached. Exposures can move back and forth between the different transient states until they reach a terminating state.

Basically, two types of matrix models can be distinguished. First, there are roll rate models, where various categories of days past due are used as state space. These models are mainly employed in retail banking, especially to monitor credit card portfolios. Second, there are rating-based matrix models. In these, the transient states correspond to the various rating categories. Default and exits for other reasons are treated as terminating states, which are also called absorbing states. Like survival models, rating-based matrix models can be further differentiated in terms of their time-space. In discrete time, these are also referred to as the "cohort approach" and in continuous time as the "duration approach". They are particularly useful in wholesale banking, where there is often a lack of historical default observations for the better rating classes. Thus, the measured probability of default for these rating classes would be zero, although it is intrinsically positive. Yet, due to the possibility of migrating to lower rating classes, matrix models can generate positive default probabilities even without historical default observations.

7.1 Rating-based Models: Cohort Approach

In practice, rating models in discrete time are among the most popular approaches to estimate multi-period default probabilities. This is also reflected in our literature sample, where they is mentioned in numerous references. We begin our exposition with the simplest estimator of these models, before we move on to discuss extensions that ease some of its restrictive assumptions. The *naïve* estimator computes p_{qr}, the probability to migrate from the current rating category q to rating category r, simply by dividing the number of exposures N_{qr} that moved from q to r during the period of interest by the number of exposures in rating category q at the beginning of the period, N_q, i.e.:

$$p_{qr} = \frac{N_{qr}}{N_q}.$$

In a system with R rating categories (where default is one of these categories), the single transition probabilities p_{qr} are conveniently summarized in a $(R \times R)$ transition matrix:

$$\mathbf{M}(1) = (p_{qr}) = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1R} \\ p_{21} & p_{22} & \dots & p_{2R} \\ \dots & \dots & \dots & \dots \\ p_{R-1,1} & p_{R-1,2} & \dots & p_{R-1,R} \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

In such a transition matrix, each element p_{qr} has to be on the interval $[0,1]$, and each row has to sum up to one ($\sum_{r=1}^{R} p_{qr} = 1$). Thus, the elements on the diagonal (p_{qq}) yield the probability to stay in the current rating class. Ideally, the entries in the transition matrix would smoothly decrease as the distance to the diagonal increases, since minor rating adjustments should be more likely than substantial rating revisions. Furthermore, the last column of the transition matrix represents the PD for each rating class over the period under consideration. As mentioned before, it is assumed that default is an absorbing state. Hence, once a subject is in default, it cannot leave this state any more. This is reflected in the last row of the transition matrix, where all entries are zero except the last one, which is equal to one.[14] Exits for other reasons can also be taken into account by adding another absorbing state to the transition matrix. However, all other rating classes are transient, such that every rating class can be reached from any other non-default rating class. In order to smooth the estimates, an average migration matrix is often formed in practice.

With such a one-period transition matrix at hand, multi-period transition matrices – and the corresponding cumulative PDs – can be easily computed by a matrix multiplication ($\mathbf{M}(\tau) = \prod_1^\tau \mathbf{M}(1)$). The resulting term-structures of the PDs will exhibit a non-linear shape. However, it should be borne in mind that these stem from the underlying model and do not coincide with the empirically observed term structures. Hence, the predictive accuracy of these models will decrease sharply after only two to three years. Moreover, the "naive estimator" above is subject to a number of restrictive assumptions. First, it is assumed that the migration probabilities are static and do not change over time. And second, it is assumed that they represent a first-order Markov chain. This implies that the current rating is the only determinant of future migrations. Hence, the exposure's past rating history and the time that it has already spent in the current rating class are disregarded. Within our literature sample, several contributions address these assumptions.

With regard to the first assumption, there are many approaches that link the time-variation in

14. In practice, if an exposure recovers from default and returns to the rating universe, it is usually treated like a new exposure.

migration probabilities to macroeconomic fluctuations (Trück 2008). These and extensions thereof are also considered to implement IFRS 9. Brunel (2016a) and Perilioglu et al. (2018) for example mention the one-factor model of Belkin et al. (1998). In the original model, the transition probabilities are assumed to be conditional on a latent, i.e. unobservable, credit cycle index Z_t:

$$p_{qrt}(z_t) = \Phi\left(\frac{\bar{c}_{qr+1} - \sqrt{\rho}z_t}{\sqrt{1-\rho}}\right) - \Phi\left(\frac{\bar{c}_{qr} - \sqrt{\rho}z_t}{\sqrt{1-\rho}}\right),$$

where $\bar{c}_{qr}$ ($\bar{c}_{qr+1}$) are constants that are obtained through a transformation of the corresponding long-run average transition probabilities ($\bar{c}_{qr} = \Phi^{-1}(p_{qr})$), and ρ is the so-called asset correlation. For a given ρ, Belkin et al. (1998) estimate the values of the latent credit cycle by minimizing the weighted mean-squared error between the model transition probabilities and the observed transition probabilities. This is repeated for many values of ρ, and the final estimates selected are those where the values of the credit cycle index have unit variance. In contrast, Brunel (2016a) allows for time-varying asset correlations and uses a maximum likelihood estimator. In addition, both the latent credit cycle and the asset correlations are made explanatory in terms of macroeconomic variables, and the updated values thereof are used to estimate future rating migrations. Yet another parametrization is discussed by Perilioglu et al. (2018) who estimate the latent credit cycle per rating class and use a multivariate GARCH model to estimate the correlations between these.[15]

Instead of using latent factor models, it is also possible to explain the time-variation in migration probabilities by macroeconomic factor models, which are perhaps more intuitive. The model of Belkin et al. (1998) above essentially is an ordered probit model (Skoglund and Chen 2017), which takes into account the natural ordering of ratings. Baesens et al. (2017, p. 160) and Pfeuffer and Fischer (2016) discuss another parametrization of the ordered probit model, which was first applied to credit migration matrices by Nickell et al. (2000). Thereby, a separate model is estimated for each row of the migration matrix (q), and the probabilities to migrate from the present rating class q to rating class r are directly explained in terms of observable macroeconomic variables $\mathbf{z}_t$:

$$p_{qrt}(\mathbf{z}_t) = P(Y_{it} = r | Y_{it-1} = q, \mathbf{z}_t) = \begin{cases} \Phi(\beta_{0,1} + \beta_3 \mathbf{z}_t) & \text{if } r = 1, \\ \Phi(\beta_{0,r} + \beta_3 \mathbf{z}_t) - \Phi(\beta_{0,r-1} + \beta_3 \mathbf{z}_t) & \text{if } 1 < r < R, \\ 1 - \Phi(\beta_{0,r-1} + \beta_3 \mathbf{z}_t) & \text{if } r = R, \end{cases}$$

where $\beta_{0,r}$ is an intercept that is specific to each target rating class r, while the sensitivity to macroeconomic factors β_3 is common to all of them.[16] As an alternative to the standard normal distribution ($\Phi()$), the logistic distribution function can also be used, in which case one obtains an ordered logit model (Pfeuffer and Fischer 2016). A disadvantage of these models, however, is that the predicted migration probabilities can become negative. Pfeuffer and Fischer (2016) therefore advocate to use multinomial models as an alternative. For instance, in the multinomial logit model,

15. The one-factor model of Belkin et al. (1998) can be extended to a multi-factor model (Wei 2003) and it can also be calibrated to equity data instead of historical migration data only (Skoglund and Chen 2016, 2017; Skoglund 2017; Bellini 2019, p. 139).

16. Note that a strict notation would require writing $\beta_{q,3}$ ($\beta_{0,q,r}$) instead of β_3 ($\beta_{0,r}$), since a separate model is estimated for each initial rating class q.

the probability to migrate from the current rating class q to another rating class r is given with respect to a reference rating class R by the following equation:

$$p_{qrt}(\mathbf{z}_t) = \frac{\exp(\beta_{0,r} + \beta_{3,r}\mathbf{z}_t)}{1 + \sum_{j=1}^{R-1} \exp(\beta_{0,j} + \beta_{3,j}\mathbf{z}_t))}$$

where all parameters β_r (β_j) are allowed to vary with the target rating category r. The probability to end up in the reference category is then given by $p_{qrt}(\mathbf{z}_t) = 1 - \sum_{j=1}^{R-1} p_{qjt}(\mathbf{z}_t)$. However, this additional flexibility over ordered probit and logit models comes at the cost of an increase in the number of parameters. A reliable estimation of multinomial logit models therefore requires significantly more data (Skoglund 2017; Skoglund and Chen 2016, 2017). Moreover, they do not take into account the natural ordering of the rating categories, which can lead to migration probabilities falling off wobbly.[17]

Categorical regression models, such as ordered models and multinomial models, can be estimated from exposure-specific migration histories (Baesens et al. 2017, p. 160). But if some covariates are shared, as it is the case with macroeconomic variables, the models can also be calibrated to absolute transition frequencies (e.g. Pfeuffer and Fischer 2016; Skoglund and Chen 2017; Yang 2017a). For the case that only relative transition frequencies are available, but the number of absolute migrations is unknown, Pfeuffer and Fischer (2016) in addition discuss a log-ratio transformation approach and the Dirichlet generalized linear model. Yet, they argue that the latter is too restrictive to be useful, and find that the log-ratio transformation approach yields results that are similar to the multinomial logit model.

Some papers have also addressed the second assumption of simple migration matrices that rating migrations can be described by a first-order Markov chain. According to this common assumption, the past rating history of an exposure should have no influence on its future migration probability. Empirically, however, it is well documented that issuers who experienced a recent downgrade by a rating agency are more likely to experience another downgrade in the follow-up period (Kavvathas 2000; Bangia et al. 2002; Lando and Skødeberg 2002; dosReis et al. 2020). This stylized fact is also known as rating momentum, but the empirical results on its effect that are reported in our sample are somewhat conflicting. Skoglund and Chen (2017) conclude that losses can be significantly underestimated if rating momentum is not taken into account, especially in severe downturn scenarios and for the lower rating classes. In contrast, dosReis et al. (2020) find default probabilities of speculative grades to decrease if rating momentum is taken into account. However, it should also be mentioned that their approaches to incorporate rating momentum are hardly comparable, and that they also use different data. In general, it is difficult to take rating momentum into account, as the number of parameters to be estimated increases rapidly and a parsimonious parametrization must be found for a reliable estimation. In addition, rating momentum may also be explained by

17. Both ordered models and multinomial models are categorical regression models. Yet another class of categorical regression models that does take into account the natural ordering of rating categories are so-called continuation ratio models. A variant of these models has been suggested by Yang (2017a). However, continuation ratio models additionally impose the assumption that the various categories can only be reached successively, which is typically not the case for ratings. Nevertheless, they could be applied to describe roll rates, whose state space follows a sequential ordering (see below).

the rating agencies' methodologies, which aim to avoid frequent rating reversals and update ratings only to a limited extent. The extent to which banks' internal ratings also exhibit such momentum effects would have to be investigated on a case-by-case basis.

Another obstacle with migration matrices is that they tend to be noisy, when the underlying sample is noisy as well. Ideally, the diagonal entries should have the largest values, and the values to the right and left should decrease smoothly. Moreover, lower rating grades should have a higher risk of default, and all elements in non-default rating classes should be larger than zero. However, in practice the off-diagonal elements often fall off wobbly, and many of them are equal to zero. A simple solution would be to calculate a historical average of the one-period transition matrix, which rules out the consideration of macroeconomic information, however. Within our sample, other solutions have therefore been discussed. For example, Pfeuffer and Fischer (2016) mention Bayesian methods and generator matrix approximations (see below) as a solution. Blümke (2020) puts forward a sequential Bayesian updating model to estimate default probabilities for no-default and low-default portfolios but does not address other transition probabilities. And Yang (2018) proposes several smoothing algorithms for both default probabilities and migration probabilities that are based on a constrained maximum likelihood estimation.

Other articles in our literature survey that mention the cohort approach have addressed the following issues. Möstel et al. (2020) compare several approaches to derive confidence intervals for migration matrices. Brezigar-Masten et al. (2021) propose a novel rating approach. Traditional rating models use a binary default indicator as the dependent variable, which indicates whether an obligor is at least 90 days past due. Brezigar-Masten et al. (2021) propose instead to directly use the days past due as dependent variable. Since these cannot be less than zero, they employ a Tobit model, which takes this truncation into account. Moreover, Štěpánková (2021) investigates various approaches to aggregate transition matrices of different banks. Based on this, she also compares bank-sourced transition matrices to those of the credit rating agencies, and estimates industry-specific transition matrices. Georgiou et al. (2021) have further developed a method to reduce the state space of Markov chains. Their method could generally be helpful to collapse a fine-grained rating migration matrix to the major rating classes. However, contrary to their primary motivation, their method seems to be unsuitable for the stage allocation in IFRS 9. This is because they essentially interpret the ECLM as a two-tier rating system, whereby stage allocation would be based solely on absolute default risk, while ignoring relative changes in credit risk since origination.

7.2 Rating-based Models: Duration Approach

Another solution to ensure strictly positive transition probabilities is to use rating-based models in continuous time. Instead of focusing on the transition probabilities within a given period, these models attempt to specify the intensity with which subjects migrate between the different rating classes. In other words, the duration approach focuses on the time until the next transition instead of the likelihood of the event itself. Similar to the cohort approach, the intensities λ_{qr} with which subjects move from state q to r are summarized in a $(R \times R)$ generator matrix:

$$\Lambda = (\lambda_{qr}) = \begin{pmatrix} -\lambda_{11} & \lambda_{12} & \dots & \lambda_{1R} \\ \lambda_{21} & -\lambda_{22} & \dots & \lambda_{2R} \\ \dots & \dots & \dots & \dots \\ \lambda_{R-1,1} & \lambda_{R-1,2} & \dots & \lambda_{R-1,R} \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

In a valid generator matrix, all off-diagonal elements λ_{qr} (with $q \neq r$) are larger than zero, and all elements at the diagonal λ_{qq} are the negative sum of the other entries, such that each row sum is equal to zero ($\lambda_{qq} - \sum_{q \neq r} \lambda_{qr} = 0$). The last column represents the default intensity of each rating class, in close analogy to the intensity function of continuous-time survival models. Moreover, default is again defined as an absorbing state, which implies that all entries in the last row of the generator matrix are equal to zero.

Generator matrices can be either obtained from discrete transition matrices or from transition data of individual subjects. One possibility to obtain a generator matrix from a given discrete transition matrix relies on the exponential function, i.e. $M(\tau) = \exp(\tau \cdot \Lambda)$, such that its entries are given by $\lambda_{qq} = log(p_{qq})$ and $\lambda_{qr} = p_{qr} \log(p_{qr})/(p_{rr} - 1))$ (Israel et al. 2001). In this approach, however, it is assumed that a firm migrates at most once during the period under consideration, which means that the resulting generator matrix is only a crude approximation and can deviate significantly from the discrete migration matrix (Trück 2008). This problem can be circumvented if the generator matrix is approximated by a Taylor series:

$$\Lambda \approx \sum_{k=1}^{\infty} (-1)^{k+1} \cdot \frac{(M - I)^k}{k},$$

where I is the identity matrix (Israel et al. 2001). Yet, this popular approach does not guarantee that all off-diagonal entries of the resulting generator matrix are non-negative, which may be solved heuristically, however. Besides this, the approximation may not always converge and even when it does, it is not guaranteed that the solution is unique. Several statistical and simulation algorithms have therefore been developed to address these problems, which are further studied in the literature on IFRS 9 by dosReis and Smith (2018) and Pfeuffer et al. (2019). Another drawback of generator matrices obtained from discrete transition matrices is that the assumption of constant transition intensities is violated in practice. The estimated cumulative default probabilities will therefore significantly deviate from the observed term structures, especially for longer maturities. Bluhm et al. (2007) have thus extended the static generator matrix by a time-dependent diagonal matrix, which allows to calibrate the generator matrix to historically observed cumulative default frequencies. Recently, this approach has been slightly generalized by Filusch (2021). However, both approaches address only that the likelihood of a transition changes with the time spent in the current rating class. In order to account for macroeconomic fluctuations as well, they still require manual overrides.

If data on individual rating transition is available, it is also possible to estimate the entries of the generator matrix directly. For example, Bellini (2019, p. 136) mentions the following maximum likelihood estimator:

$$\lambda_{qr}(\tau) = \begin{cases} \frac{N_{qr}(\tau)}{\int_0^\tau Y_q(s)ds} & \text{for } q \neq r \\ -\sum_{q \neq r} \lambda_{qr}(\tau) & \text{for } q = r, \end{cases}$$

where $Y_q(s)$ is the number of subjects in rating class q at time s, and N_{qr} is the number of subjects that migrated from rating class q to r during period τ. But this estimator assumes that the transition intensities are constant, i.e. time-homogenous.[18] To alleviate this assumption, one can use the Aalen-Johansen estimator, which is an extension of the Kaplan-Meier estimator (Lando and Skødeberg 2002). Another possibility of estimating non time-homogeneous transition intensities is based on Cox's semi-parametric model. This model is discussed in the literature on IFRS 9 by Skoglund and Chen (2017) and Bellini (2019, p. 143), as it also allows for example to include information on time-varying macroeconomic variables:

$$\lambda_{qr}(\tau) = Y_{q\tau} \lambda_{0qr}(\tau) \exp(\beta_{qr} z_t),$$

where $Y_{q\tau}$ is an indicator process, which is equal to one if the process is in rating category q, and zero otherwise (Lando and Skødeberg 2002; Leow and Crook 2014). However, in order to compute discrete transition matrices from a time-varying generator matrix, it is no longer possible to use the exponential function or the Taylor-series approximation. Instead, the product integral has to be computed.

The major advantage of the duration approach over the cohort approach is that it can be used to estimate PDs over any period, whereas the latter only allows to determine default probabilities at the end of discrete periods. Moreover, it is more robust and assigns positive PDs even over short periods of time (Brunel 2016a). In principle, the duration approach thus seems to be more suitable for IFRS 9. However, in practice ratings are often only updated once a year, which makes it difficult to reliably estimate the models. In addition, the duration approach may not be easy to communicate and also mentioned less frequently in our sample.

Compared to survival models, rating-based models provide a richer description of the process leading to default, because they do not only embody migrations to default, but also migrations between other states. However, this comes at the cost of requiring further restrictions such as the Markov assumption, and the need to estimate many more parameters. They are therefore less efficient than survival models, which approximate the data-generating process more closely. Moreover, rating-based models require a sound understanding of the underlying rating methodology as well. In particular, the stability of ratings affects migration probabilities. The more stable the ratings are, the lower will be the variability of migrations between non-default states and thus their sensitivity to changes in economic conditions. In contrast, the observed default rate per rating class behaves exactly the other way round. The more stable the ratings are held, the more the default rate and the PDs derived from it will fluctuate.

Yet, when it is not possible to reliably estimate survival models due to a lack of historical data,

18. dosReis et al. (2020) have extended this estimator by adding a self-exiting intensity with exponential decay in order to capture the effect of rating momentum, with the effect that observed downgrades increase future downgrades for a while.

rating-based models may provide a practical solution. This is especially the case for low-default portfolios in wholesale banking, for which only limited default observations are available. In this vein, rating-based models may also be useful when internal default data is scarce, because they allow to leverage the default experience of the rating agencies. An obstacle thereby, however, is that not all exposures posses an external rating. For such cases, various shadow rating approaches have been discussed in the literature on IFRS 9, which aim to replicate external ratings. For example, the textbook by Baesens et al. (2017, p. 223–225) presents a common approach consisting of two steps. In the first step, a cumulative logistic regression is used to regress the external ratings against the corresponding firm-specific variables. In the second step, this model is used to predict ratings for firms that do not have an external rating. However, Chawla et al. (2015) criticize such approaches for assuming that the rating classes are equidistant with respect to the risk of default. They therefore propose to first estimate the probability of default for rated companies. Specifically, they use a binary default indicator as the dependent variable in a regression on both the default rate of the corresponding rating class and information on the credit cycle. The output of this regression corresponds to the PD or the distance-to-default for a rating class. For a sample of rated borrowers, this output then serves as dependent variable in another regression on obligor-specific factors. Using the estimated model equation, they eventually predict the distance-to-default for non-rated obligors, from which they derive their implied ratings.

7.3 Roll Rate Models

Roll rate models are another class of discrete-time matrix models, which are well-established in financial accounting to estimate loss allowances. Whereas rating-based matrix models are traditionally applied to wholesale portfolios, roll rate models are foremost applied to retail portfolios. Instead of rating classes, the state space of roll rate models is based on days past due, which are grouped into 30-day intervals. This leads to transition matrices of the following form (Skoglund 2017):

$$\mathbf{M}(1) = (p_{qr}) = \begin{pmatrix} p_{11} & p_{12} & 0 & 0 & p_{15} & p_{16} \\ p_{21} & p_{22} & p_{23} & 0 & p_{25} & p_{26} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & p_{36} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} & p_{46} \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where the first state is being current, the second is 1–30 days past due (DPD), the third is 31–60 DPD and the fourth is 61–90 DPD. The fifth and sixth state correspond to default and prepayment, respectively. The entries on the diagonal (p_{qq}) indicate the probabilities to remain in the current state and the off-diagonal entries represent the probabilities of migrating from the current state to another state. The reason that some entries of the transition matrix are zero is that some adjacent states cannot be skipped. For example, an exposure that is not yet past due would first have to be up to 30 days past due before it can be up to 60 days past due. Otherwise, the transition matrix of roll rate models is similar to that of rating-based matrix models. Default and prepayment are treated as absorbing states, row sums have to be equal to one, and multi-period transition matrices are obtained through matrix multiplication.

The simplest possibility to estimate the parameters of this transition matrix is a non-parametric calculation of historical roll rates. After appropriate segmentation, the number of exposures who migrated from state q to state r during the current month ($N_{qr}(t)$) is simply divided by the number of all exposures who occupied state q at the end of the previous month ($N_q(t-1)$). To smooth these estimates, it is possible to average recent transition probabilities (Brunel 2016a). To lift the assumption that roll rates remain constant, they can be modelled either with time series models or as a function of transformed, macroeconomic variables (Breeden and Crook 2020; Breeden and Vaskouski 2020). Alternatively, the transition probabilities may also be estimated based on account-level data. To this end, logit models (Breeden and Vaskouski 2020), multinomial logit models (Skoglund 2017; Skoglund and Chen 2016) and Cox relative risk model (Djeundje and Crook 2018; Gaffney and McCann 2018) have been proposed, which can also take into account exposure-specific variables. However, it must be ensured that the column and row sums are always equal to one.

Roll rate models are also mentioned as a practical expedient in both the application guidance of IFRS 9 (B5.5.35) and the corresponding implementation guidance (example 12), where they are named provision matrix, to estimate expected credit losses of trade receivables. Roll rate models are overall subject to the same strengths and caveats as their rating-based counterparts. Since they are based on delinquencies, which are an early indicator of elevated credit risk, they achieve a high short-term performance, but their long-term performance is rather poor (Breeden and Vaskouski 2020).

8 Models based on Market Prices

Models based on market prices allow to assess the PD in the absence of historical default data, which is especially useful in wholesale banking. Obviously, they can only be applied to exposures of obligors for which market prices are available. Although the absolute number of such obligors is limited, they can still account for a sizeable share of the total portfolio. These models are considered to be inherently forward-looking because, according to the efficient market hypothesis, market prices should contain all available information about an obligor's risk of default. Such models will therefore also react more promptly to changes in the risk of default than models based solely on historical data. However, their reliance on market prices is also seen a major caveat. First, the default definition of these models has to be reconciled with the internal definition of default (Bellini 2019, p. 70). Second, the volatility of market prices will ultimately also be mirrored in the PDs. Especially in times of crisis, these models may therefore also overestimate the PD. And third, market prices do not only reflect the risk of default, but also other factors such as liquidity premiums for assets that are rarely traded which are difficult to disentangle (Gubareva 2020, 2021).

8.1 Merton Model

A widely used market-based model is the famous Merton (1974) model, where default occurs if a firm's asset value V_T falls below the nominal value of its debt B at the end of a given time-period

T.[19] The probability of default at time T evaluated at time t is given by:

$$P(V_T \leq B) = \Phi\left(- \frac{\ln(\frac{V_t}{B}) + (\mu_V - \frac{1}{2}\sigma_V^2)(T-t)}{\sigma_V \sqrt{T-t}} \right),$$

where $\Phi()$ denotes the cumulative density function of the standard normal distribution, μ_V is the (annualized) drift-rate of the asset-value process, and σ_V is the corresponding standard deviation. However, since the neither the asset value V_t nor its drift rate μ_V and standard deviation σ_V are observable, they have to be estimated. This is done by solving the Black and Scholes (1973) equation, where a firm's equity S_t is interpreted as European call option on its asset value:

$$S_t = V_t \Phi(d_t) - Be^{-k(T-t)} \Phi(d_t - \sigma_V \sqrt{T}),$$

with

$$d_t = \frac{\ln(\frac{V_t}{B}) + (\mu + \frac{1}{2}\sigma_V^2)(T-t)}{\sigma_V \cdot \sqrt{T-t}},$$

where k is the instantaneous risk-free rate. To solve this, one can either use a maximum likelihood approach (Duan 1994), an iterative procedure (Vassalou and Xing 2004) or solve a system of two equations by minimizing the sum of squared errors (Crosbie and Bohn 2003). In practice the model is usually applied over a one-year horizon, which leads to a mismatch between the option maturity and the debt maturity. This mismatch is typically resolved on an ad-hoc basis, for example by adjusting the default threshold B such that only half of the long-term debt is taken into account.

Within our literature sample, only Bellini (2019, p. 70) briefly mentions the application of Merton's model to low default portfolios. Besides this, Xu (2016) considers extensions of Merton's model, where default may occur at any time before maturity.[20] However, he also acknowledges that reduced-form models may be more suitable for purposes of IFRS 9. Yet, even if the Merton model and its extensions are not directly used, the distance to default (d_t) may be used as a powerful explanatory variable in other models (e.g. Bharath and Shumway 2008; Campbell et al. 2008; Kalotay and Altman 2017; Blöchlinger and Leippold 2018).

8.2 Spread-based Models

Spread-based models exploit the direct relationship between the price of credit risky products and default probabilities. One possibility to measure a borrower's cumulative PD resorts to bond spreads:

$$P(T_i \leq \tau) = 1 - \exp\left(- \frac{(y_{i\tau} - k_\tau) \cdot \tau}{1 - R_i} \right),$$

where $y_{i\tau}$ is the annual yield of a bond with maturity τ, k_τ is the risk-free interest rate, respectively the yield of a risk-free bond with the same maturity, and R_i is the recovery rate ($R_i = 1 - LGD_i$)

19. For the ease of exposition, we follow the usual notation conventions and omit subscript i in this section.

20. Bellini (2019, p. 139), Skoglund (2017), and Skoglund and Chen (2016, 2017) also discuss threshold models that are multivariate generalizations of the Merton model. These models are similar to the one-factor model of Belkin et al. (1998) and are used to model rating migrations of large corporations (Gupton et al. 1997). See Skoglund and Chen (2015) for details on model calibration.

(Hull et al. 2005). Alternatively, spreads on credit default swaps (CDS) can be used to construct the PD term-structure:

$$P(T_i \leq \tau) = 1 - \exp\left(-\frac{s_{i\tau} \cdot \tau}{1 - R_i}\right),$$

where $s_{i\tau}$ is the credit spread (Gubareva 2021). When multiple instruments with different maturities are available, the PD term-structure can be interpolated. Otherwise, it is usually assumed that the default intensity is constant over time.

However, it is well-known that spread-based models give rise to risk-neutral PDs that reflect the market price of default risk. Only a small part of the spreads is due to the actual, real-world PD. Other determinants include for instance the expected recovery rate, a premium for liquidity risk, and investors' risk aversion. It is therefore necessary to separate the default risk from other determinants. A simple method to isolate the default risk in CDS spreads for purposes of IFRS 9 is proposed by (Gubareva 2020, 2021). First, the cumulative default rates reported by the rating agencies are used to calculate a constant average default spread for each maturity, assuming that the recovery rate is 40 %. Then, using a historical sample of observed CDS spreads for each obligor, its average CDS spread is determined. The default spreads from step one are then divided by the corresponding average CDS spread to calculate the weight of default risk in the CDS spreads of each obligor. In the last step, these shares are multiplied by the CDS spread observed on the balance sheet date, from which the probability of default is finally derived. However, the approach relies on the assumption that the long-run average default spread can be identified from the rating agencies' cumulative default rates, that it is the same for all obligors, and that the weight of the default component remains constant over time. As for spread-based models in general, it is also necessary to make an assumption about the recovery rate. The industry standard is to assume a constant value of 40 %. Alternatively, it would also be possible to use the framework of Pan and Singleton (2008) to estimate both the default component and the recovery rate from the term structure of CDS spreads. This approach is used in the literature on IFRS 9 by Chamizo and Novales (2020) who do not use CDS spreads to estimate PDs but to construct a global credit factor that may be used for scenario analysis. Besides the conflation of default risk with other risk premiums, however, the use of spread-based models is further aggravated by limitations in terms of data availability and quality, which applies in particular to CDS spreads. These models can therefore only be applied to a small number of obligors, and will at most provide a complementary source of information.

9 Adjustments of Regulatory Default Probabilities

As outlined in section 2.2, IFRS 9 requires to estimate a financial instrument's PD over its remaining lifetime under consideration of current and forward-looking information. In contrast, banking regulation requires to group exposures into rating classes or segments, and the PD over the next 12 months has to be calibrated to the long-run average of annual default rates. Notwithstanding these differences, it has already been discussed during the development of IFRS 9 that it might be possible to adjust regulatory PDs for accounting purposes (EAP 2010). Similarly, regulators have also indicated that regulatory models may be used as a starting point for the ECLM (BCBS 2015;

EBA 2017). Accordingly, a number of approaches have since been proposed in the literature on how to adjust regulatory PDs for IFRS 9. In practice, banks also frequently use such approaches in order to save development costs and to ensure consistency between regulatory and accounting reporting (ECB 2017).

9.1 Scalar Approaches

The simplest approach to adjust regulatory PDs for forward-looking information is the so-called *variable scalar approach* (VSA). It was initially proposed by the FSA (2006) to transform volatile PD estimates into long run average PDs.[21] However, for purposes of IFRS 9 this approach is reversed such that the regulatory PD of an exposure or rating class (PD_{i1}^{Reg}) is multiplied by a scalar that varies with the state of the economy. Popular is for instance to multiple the regulatory PD by the forecasted default rate of the portfolio (DR_τ), divided by its long-run average ($\bar{DR}$):

$$PD_{i\tau}^{c} = PD_{i1}^{Reg} \frac{\hat{DR}_\tau}{\bar{DR}}.$$

In order to obtain forecasts of the portfolio default rate DR_τ, one may regress for example macroeconomic variables against past values of the portfolio default rate, and update the latter based on forecasts of these macroeconomic variables (Bellini 2019, p. 148). Alternatively, macroeconomic variables may be used to scale the regulatory PD. However, irrespective of the scaling factor used, a major drawback of the VSA is that the estimated $PD_{i\tau}^{c}$ is not bounded, and may exceed 100 %. Alternatively one may therefore use the *Bayesian scalar approach* (BSA), whereby $PD_{i\tau}^{c}$ is computed as follows (Skoglund 2017):

$$PD_{i\tau}^{c} = \frac{(1-\bar{DR}) \cdot \hat{DR}_\tau \cdot PD_{i1}^{Reg}}{\bar{DR} \cdot (1-\hat{DR}_\tau) \cdot (1-PD_{i1}^{Reg}) + (1-\bar{DR}) \cdot \hat{DR}_\tau \cdot PD_{i1}^{Reg}}.$$

It is generally acknowledged that both the VSA and the BSA are attractive because of their simplicity. However, both approaches suffer from a lack of theoretical foundation and the choice of the long-term default rate tends to be quite ad hoc (Rubtsov and Petrov 2016). Moreover, both approaches also require to separate systematic risks from obligor-specific risks. Ideally, the regulatory PD should be independent of macroeconomic variables, such that they are solely accounted for by the projected default rate (Skoglund 2017).

9.2 Vasicek's Single Risk Factor Model

Some of these challenges can be overcome with *Vasicek's single risk factor model* (VSRF model), which can be seen as a special case of Belkin et al.'s (1998) approach to adjust rating migration probabilities. Its variants are by far the most frequently mentioned adjustment of regulatory default probabilities. In its simplest form, which holds for large homogenous portfolios, the VSRF model is

21. In the meantime, the FSA (2007, 2009) and its successor (PRA 2017) have repeatedly released more information on the appropriate application of the variable scalar approach.

given by the following expression:

$$PD_{i\tau}^{c}(z_t) = \Phi\left(\frac{\Phi^{-1}(PD_{i1}^{Reg}) - \sqrt{\rho}z_t}{\sqrt{1-\rho}}\right),$$

where $\Phi()$ denotes the cumulative density function of the standard normal distribution, ρ is the asset-correlation, and z_t is a latent variable, representing the state of the credit cycle. The unknown model parameters ρ and z_t can be inferred from historical time series of the default rate, whereby $PD_{i\tau}^{c}$ is replaced by the observed default rate (Skoglund 2017; Bellini 2019, p. 139). Forward-looking $PD_{i\tau}^{c}$ may then be obtained by forecasting the latent factor z_t. To this end, García-Céspedes and Moreno (2017) have proposed a closed form solution based on a Taylor series expansion that takes into account the serial correlation of the latent factor. Another option is to regress the estimated values of z_t on a set of explanatory macroeconomic variables, and to update z_t based on the forecasted values of these variables (Skoglund 2017).

The basic version of the VSRF model can be extended in numerous ways. Some authors suggest estimating the model per rating class (Rubtsov and Petrov 2016; Miu and Ozdemir 2017), arguing that different rating classes may react differently to the current state of the macroeconomy. As with rating-based matrix models, however, this requires a solid understanding of the underlying rating methodology. The more sensitive the ratings are to changes in the economic environment, the more stable the default rates of the rating classes will be, which ultimately reduces their sensitivity to the state of the economy. Other authors instead propose to estimate the model per industry, since credit cycles may differ across them (Chawla et al. 2016, 2017; Rubtsov and Petrov 2016). Furthermore, the latent factor variable may be replaced by one or several observable macroeconomic variables already when estimating the model (Rubtsov and Petrov 2016; Miu and Ozdemir 2017; Yang 2017b; Oeyen and Celis 2019).

Unlike the VSA and BSA, the VSRF model and its variants are theoretically underpinned. However, its underlying assumptions are quite restrictive, and the model is not particularly robust to violations thereof. Skoglund (2017) also highlights that the resulting term structure is quite sensitive to the latent risk factor. Since there is a strong convexity between negative values of z_t and the PD, the latter will rise rapidly in economic downturns. Another challenge is that a stable estimation of the model requires time series of default rates that cover several business cycles. However, banks do not usually possess such long time series. As a potential remedy, Miu and Ozdemir (2017) therefore propose to use external time series that are representative for the default rate of the portfolio at hand. Thereby, it is implicitly assumed that the quality of the portfolio remains stable over time, such that all changes in the default rate are only due to calendar time effects.

To the best of our knowledge, there are no studies yet that analyse the quality of such transformations and empirically compare the VSRF model with the VSA and BSA. Common to these approaches is that they use one-year PDs as input, which should ideally be as independent of macroeconomic variables as possible. These are then transformed with the aim of deriving conditional PDs that are sensitive to changes in the economic environment. In this respect, all of these three approaches entail a significant model risk.

Given that the stage allocation should ideally be performed for each financial instrument separately, we argue that it would be more natural and robust to begin with the estimation of multi-period PD models at the level of individual exposures that meet the requirements of IFRS 9. The inputs or outputs of these models could then be adjusted to derive scores and ratings that meet the regulatory requirements as well. To this end, survival models seem to be particularly suitable. On the one hand, they allow to quantify the absolute default risk at the level of individual exposures, taking into account time-varying variables in line with IFRS 9. On the other hand, they also provide a ranking of exposures that can be used to derive scores. Based on these scores, internal ratings could then be assigned in accordance with the Basel Accords. And finally, the regulatory PDs could still be calibrated using the observed default rates of the respective rating classes.

10 Forward-looking Information

As mentioned above, IFRS 9 requires the consideration of forward-looking information, which we have mostly sidestepped so far. Although this requirement is considered to be one of the biggest challenges associated with IFRS 9 (EBA 2020), only a few contributions have explicitly addressed it. For models based on historical data, the usual approach is to first select three to five scenarios describing the potential evolution of the macroeconomic variables, then predict the dependent variable for each of these scenarios, and finally calculate a weighted average thereof. This approach can be used for models based on account-level data as well as for models based on aggregate portfolio data. Yet, besides the selection of relevant macroeconomic variables and suitable time-series models to predict them (see Bellini 2019, p. 257), this approach involves several challenges which have received little attention in the literature to date.

Firstly, for account-level models with time-varying obligor-specific variables, it would be necessary to forecast these variables as well ($\mathbf{x}_{i\tau}$), since they are strictly speaking a stochastic process themselves ($\mathbf{x}_i(\tau)$). This would also entail modelling the dependency between obligor-specific and macroeconomic variables, which is rarely done at all and requires a parsimonious parametrization (Duffie et al. 2007).

Secondly, the number of scenarios has to be determined. The possibilities range from a single scenario to a full Monte-Carlo simulation with potentially thousands of scenarios. However, the ITG (2015) has noted that a single scenario will only be appropriate if there is a linear relationship between the credit risk parameters and the economic scenarios. If not, multiple scenarios have to be used. At the other extreme, the IASB has also clarified that the estimation of expected credit losses "need not be a rigorous mathematical exercise, whereby an entity identifies every possible outcome and its probability." (IFRS 9 BC5.265). In practice, it is rather common to select only a few scenarios, since full Monte-Carlo simulations are computationally intensive. An advantage of full Monte-Carlo simulations is, however, that it becomes obsolete to choose a set of discrete scenarios and to attach probabilities to them, since they are already represented in the sampling process.

If only a few scenarios are used, the third challenge is therefore to determine both the severity and the probability of each scenario. In addition to a baseline scenario, it is common to determine both a pessimistic and an optimistic scenario that are based on qualitative expert judgements. The same applies to the probabilities of these scenarios, which can only be partially rationalized by quantitative approaches (deRitis et al. 2018; Yang et al. 2020).

Given that it is difficult to predict the course of the economy, it is usually also assumed that the explanatory variables converge to their long-term average after two to three years. This convergence can take place either abruptly or gradually (Breeden and Liang 2015; deRitis et al. 2018; Bellini 2019, p. 276). But after the forecast horizon, the PD will only be conditional on the long-term average of the explanatory variables and thus on a single scenario.[22]

Overall, it can be seen that the consideration of forward-looking information by means of a scenario analysis is statistically challenging. At the same time, the process is also highly judgemental, which increases management's scope for discretion. And furthermore, the estimated PDs (and expected credit losses) can hardly be interpreted when using some but few scenarios, as they represent neither a conditional nor an unconditional assessment.

For discrete-time survival models, many of these problems can at least be mitigated with an alternative approach. This approach consists of estimating a separate model for each forecast horizon and then using the most recently observed values of all explanatory variables to forecast the PD (Campbell et al. 2008). Although the number of PD models to be estimated is high with this approach, it eliminates the need to model time-varying variables, which otherwise introduces a model risk of its own kind. Nevertheless, this approach is rarely used (Breeden and Vaskouski 2020) or mentioned (Xu 2016) in the literature on IFRS 9. Future research could therefore further investigate whether and under which circumstances a scenario analysis actually yields an improved out-of-time performance. Results from Breeden and Vaskouski (2020) already suggest that a good model to predict the PD is more crucial than macroeconomic scenarios. In view of this, the IASB could also relax the information set to be used by making scenario analysis only optional. This would be consistent with the otherwise principles-based standard, and also with the IASB's original deliberation to require an entity's best estimate (IASB 2009).

11 Stage Allocation

Since the stage allocation is based solely on an assessment of default risk, it is closely related to the PD models above. The stage allocation is also the most significant feature in which the ECLM in IFRS 9 differs from the CECL in US GAAP. As outlined in chapter 2, a lifetime expected credit loss has to be recognized only when there is a significant increase in credit risk since origination. The stage allocation therefore has a substantial impact on the level of loan loss allowances, especially for portfolios of high default risk (Krüger, Rösch, et al. 2018; Skoglund and Chen 2018). In order to evaluate the contributions on the SICR test, we decompose the stage allocation according to the three essential decisions that have to be made. This includes selecting a measure of default risk, choosing a scale that relates the value of that measure at the balance sheet date to its value at origination, and setting a threshold above which a change in the scaled measure is considered to be significant.

22. If the model is based on aggregated default rates or migration rates, it is assumed that the rates themselves revert to their long-run average (deRitis et al. 2018).

11.1 Risk Measure

Regarding the risk measure, it can be observed that the qualitative risk of default is usually replaced by the quantitative PD. However, there are different views on whether the PD over the remaining lifetime or the one-year PD should be used. By default, IFRS 9 requires to use the cumulative PD. Most authors who advocate its use justify this with the corresponding requirement (e.g. Brezigar-Masten et al. 2021). In addition, Krüger, Rösch, et al. (2018) emphasize that cumulative PDs are less pro-cyclical and allow to identify long-term trends that would otherwise be neglected. By contrast, authors who advocate the use of the 12-month PD, respectively rating downgrades, stress that the estimation of cumulative PDs is highly uncertain which renders them unreliable (Chawla et al. 2016; Gaffney and McCann 2018).[23] Accordingly, it is also criticized that according to IFRS 9 only the suitability of the 12-month PD has to be demonstrated, but not that of the cumulative PD.

Another complicating factor in using the cumulative PD is that a given increase in the absolute PD will be more significant the shorter the remaining maturity. Chawla et al. (2016) and Bellini (2019, p. 278) therefore suggest that the cumulative PD should be divided by the remaining maturity, in order to make the term structures comparable across assets with different maturities. In order to also take into account the increasing estimation uncertainty, Chawla et al. (2016) additionally propose to discount the marginal or conditional PD with the effective interest rate. However, despite being difficult to communicate, it is questionable to what extent the last two measures are still supported by the requirements of IFRS 9.

Besides the PD, it is possible to use the list of indicators in IFRS 9 to assess significant increases in credit risk qualitatively. Among other things, this list mentions changes in internal price indicators, contractual terms and macroeconomic variables. However, Chawla et al. (2016) conclude after a critical appraisal that the majority of indicators in that list lag any changes in the risk of default. They therefore argue that most of these indicators can only serve as benchmarks for the PD or as backstop criteria. In addition, they point out that qualitative criteria are likely to be assessed differently across decision-makers and reporting dates. They therefore suggest to develop internal guidelines, in order to ensure a consistent application.

Independent of which risk measure is used, the ITG (2015) has clarified that an asset cannot be in different stages at the same reporting date. As already mentioned above, this implies that the stage allocation must not be performed for each scenario (Bellini 2019, p. 279). Instead, a weighted average of scenario-specific PDs must be computed beforehand.

11.2 Scaling

Once a risk measure has been set, the scaling must be chosen. It determines how the current PD is compared with the PD that was expected at origination for the present reporting date.[24] A summary of these scales is provided in table 3. The simplest scaling would be to use the *absolute* level of the

23. When using external ratings, however, it should be borne in mind that they often lag changes in credit risk (Löffler 2004; Chawla et al. 2016).

24. In practice the expected PD may not be available. Gaffney and McCann (2018) address this problem by basing their predictions on those values of the explanatory variables that were observed at the time of origination, while using the same model.

Table 3: Scaling alternatives

Scaling	Formula
Absolute level	PD_{it}
Additive scaling	$PD_{it} - E_0(PD_{it})$
Relative scaling	$\frac{PD_{it}}{E_0(PD_{it})}$
Combined scaling	$\frac{PD_{it}-E_0(PD_{it})}{E_0(PD_{it})}$

current PD, which essentially is a ordinary rating system with only two rating classes (Brunel 2016b; Georgiou et al. 2021). However, this absolute scaling lacks a comparison with the expected PD, and is therefore generally not compliant with IFRS 9. Only for financial instruments with a low credit risk, such as those with an investment grade rating, would the use of this scaling be permissible (IFRS 9.5.5.10). Notwithstanding, regulators expect banks to make limited use of this low credit risk exemption (BCBS 2015; EBA 2017).

A first alternative to the absolute level is an *additive* scaling, whereby the expected PD is subtracted from the current PD. Another possibility is to assess *relative* changes, whereby the current PD is divided by the expected PD (Chawla et al. 2016; Gaffney and McCann 2018; Brezigar-Masten et al. 2021). However, as IFRS 9 considers a given absolute change in credit risk to be more significant the lower the credit risk initially was, it would be most natural to use a *combined* scaling. Thereby, the additive scaling is additionally divided by the expected PD. The combined scaling is mentioned in the literature by Miu and Ozdemir (2017) and Krüger, Rösch, et al. (2018), and is also put-forward in regulatory guidelines (EBA 2017).

In addition, Miu and Ozdemir (2017) discuss a double trigger requirement, whereby both an absolute and a combined scaling threshold have to be exceeded, in order to prompt a transfer. In practice, such double trigger requirements are often used with the aim of avoiding high volatility of impairment provisions. However, such double-trigger requirements might again be scrutinized by auditors and regulators (EBA 2020). For one thing, they could conceal significant increases in credit risk. And for another, the low credit risk exemption was introduced by the IASB as an operational simplification rather than a global relief.

11.3 Threshold

The threshold is the value of the scaled measure which triggers a stage transfer. In other words, the threshold operationalizes what a "significant" increase in credit risk since origination is. But although the threshold has a material impact on the absolute level of loan loss allowances and provisions, so far there are no generally accepted thresholds (Chawla et al. 2016; Gubareva 2021). Accordingly, almost all contributions avoid limiting the range of admissible values. If any values are given, they serve only for illustrative purposes. The key challenge is that – due to the qualitative requirements of IFRS 9 – there are no right or wrong stage allocations, which could be used to backtest significant increases in credit risk since origination.

Chawla et al. (2016) discuss various criteria that could be used to determine appropriate thresholds. First, they consider to rely on existing credit processes, such as watch lists and lending criteria, which they reject for being backward-looking, however. Second, they consider to use confidence intervals, which they dismiss for being impractical as well. Moreover, it should be noted that such an approach would conflate significant increases in credit risk with statistical significance. It would therefore be more appropriate to speak of substantial increases in credit risk. As a third option remains to rely on expert opinions. For example, a doubling or tripling of PD is often considered to be substantial, when using a relative scaling (e.g. Brezigar-Masten et al. 2021). Krüger, Rösch, et al. (2018), who use a combined scaling, use thresholds of 5%, 20% and 50%. Likewise, downgrades by one notch on the rating agencies' eight-point scale are considered to be substantial as well. Another possibility to approximate appropriate threshold values might be to rely on the rebuttable presumption that a significant increase in credit risk has occurred if an exposure is 30 days past due. If a high proportion of the stage transfers is only triggered through this presumption, this could indicate that the threshold is set too high.

12 Discussion and Conclusion

Previous surveys on the estimation of default probabilities for IFRS 9 have often described only a limited number of models to implement the new impairment requirements. Against this background, we present the first systematic review on both the estimation of PDs over the remaining lifetime and the ongoing discussion about the stage allocation. In a systematic literature review, we summarize 52 contributions that have been published on these subjects so far.

We have structured the models to estimate the PD according to the data they use, as well as their time-space and state-space. This enabled us not only to discuss the merits and limitations of individual models, but also to demonstrate the relationships and differences between them. Overall, it becomes apparent that banks can draw on a variety of statistical models to determine the PD over the remaining lifetime. As far as can be inferred from the banks' annual reports, in practice mainly transformations of regulatory PDs and rating-based matrix models seem to be used. However, these methods approximate the underlying data-generating process only indirectly and require a sound understanding of the underlying rating methodology. In the academic literature, on the other hand, survival models that directly describe the data-generating process are now predominantly used to model the risk of default or bankruptcy, respectively. Survival models that can take into account time-varying variables also seem to be particularly suitable for IFRS 9. Where, in addition, early repayments play a significant role, competing risk models may be considered. However, the calibration of these models also requires that sufficient default observations are available. In practice, this is likely to be the case only for retail portfolios, if at all. For low-default portfolios in wholesale banking, rating-based matrix models will therefore probably remain the market standard.

Furthermore, we argue that for IFRS 9 a calibration of the PD at the account level is more appropriate than a calibration at the portfolio level as prescribed in the Basel Accords. This is because in order to identify a significant increase in the default risk of a financial instrument, it is necessary that the PD model is able to measure such increases in the first place. Three conclusions follow from this. First, the usual classification of IFRS 9 as "point-in-time" and the Basel Accords as "through-the-cycle" seems to be unwarranted, as it ignores these calibration differences. Second, it calls into question the feasibility of approaches to adjust aggregated regulatory PDs for IFRS 9. In order to maintain only a few models and to ensure consistency between accounting and regulation, we propose that banks could once again resort to survival models. But third, it can also be deduced that account-level models are necessary primarily due to the stage allocation. For the CECL in US GAAP, where there is no stage allocation, simpler models based on aggregated portfolio data can be used (Breeden and Vaskouski 2020).

With regard to the stage allocation, the following conclusions can be made on the basis of the available literature. Most often, the PD is mentioned as a risk measure, but opinions vary on whether annual PDs or their entire term-structure should be used. To scale the measure, relative or combined scalings seem to be more appropriate than absolute scalings or double trigger requirements. But with regard to the threshold, there is still no consensus on appropriate values. Banks can therefore exercise a considerable scope of discretion, which will ultimately limit the comparability of expected credit losses across them. This could be further reinforced by the choice of models to estimate the PD. However, no study has yet empirically investigated how model choice affects the stage allocation. In addition, we find that the stage allocation is susceptible to misinterpretation. Although it basically provides for an assessment of default risk at the balance sheet date relative to initial expectations, it is occasionally interpreted as a two-grade rating system in which exposures are solely assessed according to their absolute risk of default at the balance sheet date.

In closing, we would like to point out some avenues for future research. So far, there are only a few benchmark studies that compare different models empirically. Breeden and Vaskouski (2020) and Breeden and Crook (2020) use single family loan data from Fannie Mae and Freddie Mac to compare age-period cohort models, vintage models and roll rate models, among others. Beyond our sample, Gupta et al. (2018) compare the performance of Cox's relative risk model with discrete-time hazard models to predict bankruptcies of small and medium-sized enterprises (SMEs). Similarly, Dirick et al. (2017) use data on personal loans and loans to SMEs to compare Cox's relative risk model with various AFT models as well as mixture cure rate models. Future benchmark studies could compare other models, whose performance should be validated out-of-time. In particular, there seem to be no studies available yet that empirically assess the quality of transformations of regulatory PDs for purposes of IFRS 9. Moreover, most studies on rating-based models use data from the rating agencies. Little is still known about the properties of banks' internal rating data (Štěpánková 2021).

Future studies could also compare statistical models with machine learning methods. Such methods have so far only been mentioned in the literature on IFRS 9 by Bellini (2019), Altman et al. (2020), and Blumenstock et al. (2020). Since these contributions already provide a concise summary of the methods they use, but do not go into much technical detail themselves, we have decided not to expand on them in our review. Apart from that, Altman et al. (2020) find that even advanced machine learning techniques such as neural nets and support vector machines do not outperform a logistic regression for predicting financial distress of small and medium-sized enterprises. And given the black-box nature of many machine learning methods, Blumenstock et al. (2020) also acknowledge that they are still far from regulatory approval. In the near future, they will therefore mainly have a complementary role in model validation and as benchmarks. However, given that machine learning currently receives a lot of attention in the credit risk community, further reviews and benchmark studies would certainly be welcome.

Finally, PD estimates should also be compared across banks (EBA 2020). In the context of banking regulation, previous studies have already shown that the outputs of internal models differ significantly (Le Leslé and Avramova 2012). Several studies have also found evidence that banks do use their discretion over internal models strategically (Vallascas and Hagendorff 2013; Mariathasan and Merrouche 2014; Plosser and Santos 2014; Beltratti and Paladino 2016; Firestone and Rezende 2016). For at least three reasons, these problems could be even more serious for IFRS 9. First, it involves inherently more uncertainty to estimate the PD over the remaining lifetime than over one year. Second, as this review has shown, banks can choose from a variety of models and approaches when estimating the PD for IFRS 9, and the stage allocation is also largely at their own discretion. And third, unlike the Basel Accords and despite these additional imponderables, IFRS 9 contains hardly any requirements for the validation and backtesting of internal models, which could further exacerbate differences across banks.

13 Appendix

References

Altman, Edward I., Małgorzata Iwanicz-Drozdowska, Erkki K. Laitinen, and Arto Suvas. 2016. "Financial and nonfinancial variables as long-horizon predictors of bankruptcy." *Journal of Credit Risk* 12 (4): 49–78. DOI: 10.21314/JCR.2016.216.

———. 2020. "A Race for Long Horizon Bankruptcy Prediction." *Applied Economics* 52 (37): 4092–4111. DOI: 10.1080/00036846.2020.1730762.

Baesens, Bart, Daniel Röesch, and Harald Scheule. 2017. *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS.* Hoboken (NJ): Wiley.

Bangia, Anil, Francis X. Diebold, Christian Kronimus Andréand Schagen, and Til Schuermann. 2002. "Ratings Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing." *Journal of Banking & Finance* 26 (2–3): 445–474. DOI: 10.1016/S0378-4266(01)00229-1.

Basel Committee on Banking Supervision. 2015. *Guidance on accounting for expected credit losses.* Guideline. Basel. Accessed 2020-07-31. www.bis.org/bcbs/publ/d350.pdf.

———. 2019. *CRE - Calculation of RWA for Credit Risk.* Basel. Accessed 2021-08-10. www.bis.org/basel_framework/standard/CRE.htm.

Batiz-Zuk, Enrique, Fabrizio López-Gallo, Abdulkadir Mohamed, and Fátima Sánchez-Cajal. 2020. "Determinants of Loan Survival Rates for Small and Medium-sized Enterprises: Evidence from an Emerging Economy." Forthcoming, *International Journal of Finance & Economics,* 1–15. DOI: 10.1002/ijfe.2397.

Belkin, Barry, Stephan Suchower, and Lawrence R. Forest. 1998. "A one-parameter representation of credit risk and transition matrices." *CreditMetrics Monitor* 1998 (3): 46–56.

Bellini, Tiziano. 2019. *IFRS 9 and CECL Credit Risk Modelling and Validation: A Practical Guide with Examples Worked in R and SAS.* London: Academic Press.

Table 4: Contributions in our sample

Study	Models considered	Data	Data provider	Region	Period	Frequency
Altman et al. 2016	(10)	Corporate portfolio	Suomen Asiakastieto Oy	Finland	2003 – 2013	Yearly
Altman et al. 2020	(10)	Corporate portfolio	Suomen Asiakastieto Oy	Finland	2003 – 2013	Yearly
Baesens et al. 2017	(1) (2) (4) (7) (8) (10) (14)	Various	Various	Various	Various	Various
Batiz-Zuk et al. 2020	(1) (3) (4)	SME portfolio	Central Bank of Mexiko	Mexico	2010 – 2017	Monthly
Bellini 2019	(1) (2) (4) (10) (13) (14) (18)	Various	Various	Various	Various	Various
Blumenstock et al. 2020	(6)	Single family loan data	Fannie Mae, Freddie Mac	US	1999 – 2017	Monthly
Blümke 2020	(14)	Simulated				
Breeden 2016	(11)	Auto loan portfolio	Proprietary	US	2007 – 2012	Monthly
Breeden and Crook 2020	(9) (11) (15)	Single family loan data	Fannie Mae, Freddie Mac	US	2001 – 2017	Monthly
Breeden and Vaskouski 2020	(9) (11) (15)	Single family loan data	Fannie Mae, Freddie Mac	US	2001 – 2017	Monthly
Brezigar-Masten et al. 2021	(10) (13)	Non-financial firms	Bank of Slovenia	Slovenia	2007 – 2017	Quarterly
Brunel 2016a	(1) (2) (7) (9) (11) (13) (14) (15)					
Chawla et al. 2015	(13)					
Chawla et al. 2017	(18)					
Chen et al. 2020	(10)	Credit card accounts	Proprietary	China		
Dirick et al. 2019	(7)	Personal loans	Proprietary	Belgium	2004 – 2014	Monthly
Djeundje and Crook 2018	(15)	Credit card accounts	Proprietary	UK	2005 – 2011	Monthly
Djeundje and Crook 2019	(10) (12)	Credit card accounts	Proprietary	UK	2002 – 2011	Monthly
dosReis and Smith 2018	(13)	Simulated				
dosReis et al. 2020	(13) (14)	Corporate credit ratings	Moody's		1987 – 2017	Annual
Filusch 2021	(13) (14)	Corporate credit ratings	S&P	Global	1983 – 2017	Annual
Gaffney and McCann 2018	(15)	Mortgage loans	Proprietary	Ireland	2008 – 2015	Biannual
García-Céspedes and Moreno 2017	(18)	Simulated				
Georgiou et al. 2021	(13)	Souvereign credit ratings	Fitch, Moody's, S&P		1990 – 2006	Annual
Gubareva 2020	(17)	CDS	Bloomberg, CMAN,	US	2007 – 2017	Daily
Gubareva 2021	(17)	Corporate bonds, CDS	Bloomberg, CMAN	US, Europe	2007 – 2017	Daily
Gürtler et al. 2018	(10)	Checking accounts	Proprietary	Europe	2007 – 2014	Monthly
Kalotay and Altman 2017	(10)	Non-financial corporations	Moody's, CRSP, Compustat	US	1987 – 2011	Quarterly
Krüger, Oehme, et al. 2018	(4)	Corporate bonds	Moody's, FRED	US	1982 – 2014	Yearly
Krüger, Rösch, et al. 2018	(10)	Corporate bonds	Moody's, FRED	US	1990 – 2013	Yearly
Lee et al. 2021	(12)	Mortgage loans	Int. Financial Research	US	2002 – 2014	Monthly
McPhail and McPhail 2014	(2) (3) (4) (9) (10) (13) (15)					
Miu and Ozdemir 2017	(18)	Residential mortgages				
Möstel et al. 2020	(13) (14)	Corporate credit ratings	Moody's		1987 – 2017	Annual
Orth 2013	(5)	Corporate portfolio	S&P, CRSP, Compustat	US	1980 – 2010	Monthly
Oeyen and Celis 2019	(18)	Simulated				
Perilioglu et al. 2018	(13)		Moody's	US	1980 – 2016	Quarterly
Pfeuffer and Fischer 2016	(13)	Corporate portfolio	S&P	US	1980 – 2013	Annual
Pfeuffer et al. 2019	(13) (14)	Corporate credit ratings	S&P	Global	2002, 2008	Annulal
Rubtsov and Petrov 2016	(18)	Corporate portfolio	Proprietary			
Schutte et al. 2020	(8)	Retail portfolio,	Proprietary	South Africa	2005 – 2015	Montlhy
Skoglund 2017	(2) (3) (9) (13) (18)					
Skoglund and Chen 2016	(13) (15)	Simulated				
Skoglund and Chen 2017	(13) (18)	Simulated				
Štěpánková 2021	(13)	Large corporations	Credit Benchmark	NOAM, EU, UK	2015 – 2019	Monthly
Xu 2016	(10) (16)					
Yang 2017b	(13) (18)	Simulated				
Yang 2017a	(13)	Simulated				
Yang 2018	(13)	Simulated				
Zhang and Tung 2019	(18)	Large corporations	S&P	Global	1995 – 2015	Annual

Notes: The second column refers to the numbers in table 1. Empty fields indicate that the corresponding information is either not available or does not apply. In addition to the contributions above, our sample also includes Brunel (2016b) and Chawla et al. (2016) who, however, only discuss the stage allocation.

Beltratti, Andrea, and Giovanna Paladino. 2016. “Basel II and regulatory arbitrage. Evidence from financial crises.” *Journal of Empirical Finance* 39:180–196. DOI: 10.1016/j.jempfin.2016.02.006.

Bharath, Sreedhar T., and Tyler Shumway. 2008. “Forecasting Default with the Merton Distance to Default Model.” *The Review of Financial Studies* 21 (3): 1339–1369. DOI: 10.1093/rfs/hhn044.

Black, Fischer, and Myron Scholes. 1973. “The Pricing of Options and Corporate Liabilities.” *Journal of Political Economy* 81 (3): 637–654. DOI: 10.1086/260062.

Blöchlinger, Andreas, and Markus Leippold. 2018. “Are Ratings the Worst Form of Credit Assessment Except for All the Others?” *Journal of Financial and Quantitative Analysis* 53 (1): 299–334. DOI: 10.1017/S0022109017000874.

Bluhm, Cristian, Ludger Overbeck, and Christoph Wagner. 2007. “Calibration of PD term structures: to be Markov or not to be.” *Risk* 20 (11): 98–103.

Blumenstock, Gabriel, Stefan Lessmann, and Hsin-Vonn Seow. 2020. “Deep Learning for Survival and Competing Risk Modelling.” Forthcoming, *Journal of the Operational Research Society,* 1–13. DOI: 10.1080/01605682.2020.1838960.

Blümke, Oliver. 2020. “Estimating the probability of default for no-default and low-default portfolios.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 69 (1): 89–107. DOI: 10.1111/rssc.12381.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511790874.

Breeden, Joseph L. 2016. “Incorporating Lifecycle and Environment in Loan-level Forecasts and Stress Tests.” *European Journal of Operational Research* 255 (2): 649–658. DOI: 10.1016/j.ejor.2016.06.008.

Breeden, Joseph L., and Jonathan Crook. 2020. “Multihorizon discrete time survival models.” Forthcoming, *Journal of the Operational Research Society,* 1–14. DOI: 10.1080/01605682.2020.1777907.

Breeden, Joseph L., and Sisi Liang. 2015. “A Mean-reverting Scenario Design Model to Create Lifetime Forecasts and Volatility Assessments for Retail Loans.” *Journal of Risk Model Validation* 9 (4): 19–30. DOI: 10.21314/JRMV.2015.141.

Breeden, Joseph L., and Maxim Vaskouski. 2020. “Current expected credit loss procyclicality: it depends on the model.” *Journal of Credit Risk* 16 (1): 27–48. DOI: 10.21314/JCR.2020.258.

Brezigar-Masten, Arjana, Igor Masten, and Matjaž Volk. 2021. “Modeling credit risk with a Tobit model of days past due.” *Journal of Banking & Finance* 122:105984. DOI: 10.1016/j.jbankfin.2020.105984.

Brunel, Vivien. 2016a. *Lifetime PD Analytics for Credit Portfolios: A Survey.* Working Paper. Accessed 2016-10-26. https://ssrn.com/abstract=2857183.

———. 2016b. “Sizing the Stage 2 Portfolio for IFRS 9 Provisions.” *Risk* 29 (5): 73–76.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi. 2008. “In Search of Distress Risk.” *Journal of Finance* 63 (6): 2899–2939. DOI: 10.1111/j.1540-6261.2008.01416.x.

Chamizo, Álvaro, and Alfonso Novales. 2020. “Looking through systemic credit risk: Determinants, stress testing and market value.” *Journal of International Financial Markets, Institutions and Money* 64:101167. DOI: 10.1016/j.intfin.2019.101167.

Chava, Sudheer, and Robert A. Jarrow. 2004. "Bankruptcy Prediction with Industry Effects." *Review of Finance* 8 (4): 537–569. DOI: 10.1093/rof/8.4.537.

Chawla, Gaurav, Lawrence R. Forest, and Scott D. Aguais. 2015. "AERB: Developing AIRB PIT-TTC PD Models Using External Ratings." *Journal of Risk Model Validation* 9 (4): 1–18. DOI: 10.21314/JRMV.2015.143.

———. 2016. "Some Options for Evaluating Significant Deterioration under IFRS 9." *Journal of Risk Model Validation* 10 (3): 69–89. DOI: 10.21314/JRMV.2016.158.

———. 2017. "Convexity and Correlation Effects in Expected Credit Loss Calculations for IFRS9/CECL and Stress Testing." *Journal of Risk Management in Financial Institutions* 10 (1): 99–110.

Chen, Ying, Chen Yang, and Chi Zhang. 2020. "Study on the influence of IFRS 9 on the impairment of commercial bank credit card." Forthcoming, *Applied Economics Letters,* 1–6. DOI: 10.1080/13504851.2020.1855298.

Cohen, Benjamin H., and Gerald A. Edwards. 2017. *The new era of expected credit loss provisioning.* Bank of International Settlements, Basel. Accessed 2020-08-28. www.bis.org/publ/qtrpdf/r_qt1703f.pdf.

Cox, David R. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society* 34 (2): 187–220.

Crook, Jonathan, and Tony Bellotti. 2010. "Time varying and dynamic models for default risk in consumer loans." *Journal of the Royal Statistical Society: Series A* 173 (2): 283–305. DOI: 10.1111/j.1467-985X.2009.00617.x.

Crosbie, Peter J., and Jeffrey R. Bohn. 2003. *Modeling Default Risk.* Accessed 2018-08-01. https://business.illinois.edu/gpennacc/MoodysKMV.pdf.

De Leonardis, Daniele, and Roberto Rocci. 2014. "Default risk analysis via a discrete-time cure rate model." *Applied Stochastic Models in Business and Industry* 30 (5): 529–543. DOI: 10.1002/asmb.1998.

deRitis, Cristian, Juan M. Licari, and Gustavo Ordonez-Sanz. 2018. "Macroeconomic Forecasting and Scneario Design for IFRS 9 and CECL." In *The New Impairment Model Under IFRS 9 and CECL,* edited by Jing Zhang. London: Risk Books.

Dirick, Lore, Tony Bellotti, Gerda Claeskens, and Bart Baesens. 2019. "Macro-Economic Factors in Credit Risk Calculations: Including Time-Varying Covariates in Mixture Cure Models." *Journal of Business & Economic Statistics* 37 (1): 40–53. DOI: 10.1080/07350015.2016.1260471.

Dirick, Lore, Gerda Claeskens, and Bart Baesens. 2017. "Time to default in credit scoring using survival analysis: a benchmark study." *Journal of the Operational Research Society* 68 (6): 652–665. DOI: 10.1057/s41274-016-0128-9.

Djeundje, Viani Biatat, and Jonathan Crook. 2018. "Incorporating heterogeneity and macroeconomic variables into multi-state delinquency models for credit cards." *European Journal of Operational Research* 271 (1): 697–709. DOI: 10.1016/j.ejor.2018.05.040.

———. 2019. "Dynamic survival models with varying coefficients for credit risks." *European Journal of Operational Research* 275 (1): 319–333. DOI: 10.1016/j.ejor.2018.11.029.

dosReis, Goncalo, Marius Pfeuffer, and Greig Smith. 2020. "Capturing model risk and rating momentum in the estimation of probabilities of default and credit rating migrations." *Quantitative Finance* 20 (7): 1069–1083. DOI: 10.1080/14697688.2020.1726439.

dosReis, Goncalo, and Greig Smith. 2018. "Robust and consistent estimation of generatiors in credit risk." *Quantitative Finance* 18 (6): 983–1001. doi: 10.1080/14697688.2017.1383627.

Duan, Jin-Chuan. 1994. "Default Risk in Equity Returns." *Mathematical Finance* 4 (2): 155–167. doi: 10.1111/j.1467-9965.1994.tb00055.x.

Duffie, Darrell, Leandro Saita, and Ke Wang. 2007. "Multi-period Corporate Default Prediction with Stochastic Covariates." *Journal of Financial Economics* 83 (3): 635–665. doi: 10.1016/j.jfineco.2005.10.011.

Eder, Bernhard. 2021. *Revisiting the Dualism of Point-in-Time and Through-the-Cycle Credit Risk Measures.* Working Paper. Accessed December 9, 2021. https://ssrn.com/abstract=3981340.

European Banking Authority. 2015. *The EBA's views on the adoption of IFRS 9 Financial Instruments (IFRS 9).* London. Accessed 2020-08-28. www.eba.europa.eu/documents/10180/943157/Letter+to+EFRAG+Board+on+IFRS+9+endorsement.pdf.

———. 2017. *Guidelines on credit institutions' credit risk management practices and accounting for expected credit losses,* EBA/GL/2017/06, London. Accessed 2018-08-05. https://eba.europa.eu/regulation-and-policy/accounting-and-auditing/guidelines-on-accounting-for-expected-credit.

———. 2020. *IFRS 9 Implementation by EU Institutions.* Paris. Accessed 2021-11-28. https://www.eba.europa.eu/eba-notes-significant-efforts-ifrs-9-implementation-eu-institutions-cautions-some-observed.

European Central Bank. 2017. *SSM Thematic Review on IFRS 9 – Assessment of Institutions' Preparedness for the Implementation of IFRS 9.* Frankfurt. Accessed 2020-07-22. www.bankingsupervision.europa.eu/press/letterstobanks/shared/pdf/2017/ssm.reportlsi_2017.en.pdf.

European Systemic Risk Board. 2019. *The cyclical behaviour of the ECL model in IFRS 9.* Frankfurt. Accessed 2019-07-29. doi: 10.2849/054320.

Expert Advisory Panel. 2010. *Expert Advisory Panel - Summary of all EAP meetings.* International Accounting Standards Board, London, July. Accessed 2020-10-14. www.ifrs.org/content/dam/ifrs/project/fi-impairment/exposure-draft-2009/expert-advisory-panel/eap-meeting-summary-june-2010.pdf.

Filusch, Tobias. 2021. "Risk assessment for financial accounting: modeling probability of default." *Journal of Risk Finance* 22 (1): 1–15. doi: 10.1108/JRF-02-2020-0033.

Financial Crisis Advisory Group. 2009. *Report of the Financial Crisis Advisory Group.* Report. Accessed 2019-10-01. https://cdn.ifrs.org/-/media/feature/groups/fcag/report-of-the-fcag.pdf.

Financial Services Authority. 2006. *Use of Long Run Probabilities of Default, Counter-Cyclical Scaling Factors, and the Interaction of these with Economic Cycle Stress Testing.* Mimeo. London. Accessed 2019-02-18. www.fsa.gov.uk/pubs/international/crsg_procyclicality3.pdf.

———. 2007. *Use of Variable Scaling Factors to Derive Long Run Probabilities of Default for Retail Portfolios.* Mimeo. London. Accessed 2019-02-18. www.fsa.gov.uk/pubs/international/crsg_variable_scalars.pdf.

———. 2009. *Variable Scalar Approaches to Estimating Through the Cycle PDs.* Mimeo. London. Accessed 2019-02-18. www.fsa.gov.uk/pubs/international/variable_scalars.pdf.

Financial Stability Forum. 2009. *Report of the Financial Stability Forum on Addressing Procyclicality in the Financial System.* Report. Accessed 2019-10-01. www.fsb.org/wp-content/uploads/r_0904a.pdf.

Fine, Jason P., and Robert J. Gray. 1999. "A Proportional Hazards Model for the Subdistribution of a Competing Risk." *Journal of the American Statistical Association* 94 (446): 496–509.

Firestone, S., and M. Rezende. 2016. "Are Banks' Internal Risk Parameters Consistent? Evidence from Syndicated Loans." *Journal of Financial Services Research* 50:211–242. DOI: 10.1007/s10693-015-0224-z.

G20 Leaders. 2009. *London Summit – Leaders' Statement.* Report. Accessed 2019-10-01. www.imf.org/external/np/sec/pr/2009/pdf/g20_040209.pdf.

Gaffney, Edward, and Fergal McCann. 2018. *The cyclicality in SICR: mortgage modelling under IFRS 9.* Research Technical Papers. Dublin. Accessed 2019-02-05. www.centralbank.ie/publications/Documents/16RT14.pdf.

García-Céspedes, Rubén, and Manuel Moreno. 2017. "An approximate multi-period Vasicek credit risk model." *Journal of Banking & Finance* 81:105–113. DOI: 10.1016/j.jbankfin.2017.05.002.

Georgiou, K., G. N. Domazakis, D. Pappas, and A.N. Yannacopoulos. 2021. "Markov chain lumpability and applications to credit risk modelling in compliance with the International Financial Reporting Standard 9 framework." *European Journal of Operational Research* 292 (3): 1146–1164. DOI: 10.1016/j.ejor.2020.11.014.

Giner, Begoña, and Araceli Mora. 2019. "Bank loan loss accounting and its contracting effects: the new expected loss models." *Accounting and Business Research* 49 (6): 726–752. DOI: 10.1080/00014788.2019.1609898.

Gubareva, Mariya. 2020. "IFRS 9 compliant economic adjustment of expected credit loss modeling." *Journal of Credit Risk* 16 (2): 29–66. DOI: 10.21314/JCR.2020.260.

———. 2021. "How to estimate expected credit losses – ECL – for provisioning under IFRS 9." *Journal of Risk Finance* 22 (2): 1–19. DOI: 10.1108/JRF-05-2020-0094.

Gupta, Jairaj, Andros Gregoriou, and Tahera Ebrahimi. 2018. "Empirical comparison of hazard models in predicting SMEs failure." *Quantitative Finance* 18 (3): 437–466. DOI: 10.1080/14697688.2017.1307514.

Gupton, Greg M., Christopher C. Finger, and Mickey Bhatia. 1997. *CreditMetrics.* Technical Document. J.P. Morgan & Co.

Gürtler, Marc, Martin T. Hibbeln, and Piet Usselmann. 2018. "Exposure at default modeling – A theoretical and empirical assessment of estimation approaches and parameter choice." *Journal of Banking & Finance* 91 (1): 176–188. DOI: 10.1016/j.jbankfin.2017.03.004.

Hashim, Noor, Weijia Li, and John O'Hanlon. 2016. "Expected-loss-based Accounting for Impairment of Financial Instruments: The FASB and IASB Proposals 2009-2016." *Accounting in Europe* 2 (13): 229–267. DOI: 10.1080/17449480.2016.1210179.

Heitfield, Erik. 2005. *Dynamics of rating systems: Studies on the Validation of Internal Rating Systems.* Working Paper, 14. Basel Committee on Banking Supervision, Basel. Accessed 2020-07-30. www.bis.org/publ/bcbs_wp14.pdf.

Hillegeist, Stephen A., Elizabeth K. Keating, Donald P. Cram, and Kyle G. Lundstedt. 2004. "Assessing the Probability of Bankruptcy." *Assessing the Probability of Bankruptcy* 9 (1): 5–34. DOI: 10.1023/B:RAST.0000013627.90884.b7.

Hull, John, Mirela Predescu, and Alan White. 2005. "Bond prices, default probabilities and risk premiums." *Journal of Credit Risk* 1 (2): 53–60. DOI: 10.21314/JCR.2005.007.

IFRS Transition Resource Group for Impairment of Financial Instruments. 2015. *Significant increases in credit risk.* Staff papper. London: International Accounting Standards Board. Accessed 2016-09-11. www.ifrs.org/Meetings/MeetingDocs/Other%20Meeting/2015/September/AP1-Significant-increases-in-credit-risk.pdf.

International Accounting Standards Board. 2009. *Request for Information.* London. Accessed 2020-09-29. www.ifrs.org/content/dam/ifrs/project/fi-impairment/request-for-information/published-documents/rfi-expected-loss-model.pdf.

Israel, Robert B., Jeffrey S. Rosenthal, and Jason Z. Wei. 2001. "Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings." *Mathemtaical Finance* 11 (2): 245–265. DOI: 10.1111/1467-9965.00114.

Kalbfleisch, John D., and Ross L. Prentice. 2002. *The Statistical Analysis of Failure Time Data.* Hoboken, NJ: Wiley.

Kalotay, Egon A., and Edward I. Altman. 2017. "Intertemporal Forecasts of Defaulted Bond Recoveries and Portfolio Losses." *Review of Finance* 21 (1): 433–463. DOI: 10.1093/rof/rfw028.

Kaplan, Edward. L., and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association* 53 (282): 457–481. DOI: 10.1080/01621459.1958.10501452.

Kavvathas, Dimitros. 2000. *Estimating Credit Rating Transition Probabilities for Corporate Bonds.* Working Paper. Accessed 2018-07-30. DOI: 10.2139/ssrn.252517.

Kiefer, Nicholas M., and C. Erik Larson. 2015. "Counting processes for retail default modeling." *Journal of Credit Risk* 11 (3): 45–72. DOI: 10.21314/JCR.2015.194.

Krüger, Steffen, Toni Oehme, Daniel Rösch, and Harald Scheule. 2018. "A copula sample selection model for predicting multi-year LGDs and Lifetime Expected Losses." *Journal of Empirical Finance* 47 (4): 246–262. DOI: 10.1016/j.jempfin.2018.04.001.

Krüger, Steffen, Daniel Rösch, and Harald Scheule. 2018. "The impact of loan loss provisioning on bank capital requirements." *Journal of Financial Stability* 36:114–129. DOI: 10.1016/j.jfs.2018.02.009.

Lando, David, and Torben M. Skødeberg. 2002. "Analyzing rating transitions and rating drift with continuous observations." *Journal of Banking & Finance* 26 (2–3): 423–444. DOI: 10.1016/S0378-4266(01)00228-X.

Lawless, Jerald F. 2003. *Statistical Models and Methods for Lifetime Data.* Princeton, NJ: John Wiley & Sons, Inc.

Le Leslé, Vanessa, and Sofiya Avramova. 2012. *Revisiting Risk-Weighted Assets: Why Do RWAs Differ Across Countries and What Can Be Done About It?* Working Paper. International Monetary Fund, Washington. Accessed 2021-02-24. www.imf.org/external/pubs/ft/wp/2012/wp1290.pdf.

Lee, Yongwoong, Daniel Rösch, and Harald Scheule. 2021. "Systematic credit risk in securitised mortgage portfolios." *Journal of Banking & Finance* 122:105996. DOI: 10.1016/j.jbankfin.2020.105996.

Leow, Mindy, and Jonathan Crook. 2014. "Intensity models and transition probabilities for credit card loan delinquencies." *European Journal of Operational Research* 236 (2): 685–694. DOI: 10.1016/j.ejor.2015.05.030.

Löffler, Gunter. 2004. "An anatomy of rating through the cycle." *Journal of Banking & Finance* 28 (3): 695–720. DOI: 10.1016/S0378-4266(03)00041-4.

Mariathasan, Mike, and Ouarda Merrouche. 2014. "The manipulation of basel risk-weights." *Journal of Financial Intermediation* 23 (3): 300–321. DOI: 10.1016/j.jfi.2014.04.004.

Mayer, Manuel, and Stephan Sauer. 2017. "Are Through-the-Cycle Credit Risk Models a Beneficial Macro-Prudential Policy Tool?" In *Monetary Policy, Financial Crises, and the Macroeconomy: Festschrift for Gerhard Illing,* edited by Frank Heinemann, Ulrich Klüh, and Sebastian Watzka, 201–224. Cham: Springer International Publishing.

McCullagh, Peter, and John A. Nelder. 1989. *Generalized linear models.* 2nd ed. Boca Raton, Fla.: Chapman & Hall/CRC.

McPhail, Joseph, and Lihong McPhail. 2014. "Forecasting Lifetime Credit Losses: Modelling Considerations for Complying with the new FASB and IASB Current Expected Credit Loss Models." *Journal of Risk Management in Financial Institutions* 7 (4): 375–388.

Merton, Robert C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29 (2): 449–470. DOI: 10.1111/j.1540-6261.1974.tb03058.x.

Miu, Peter, and Bogie Ozdemir. 2017. "Adapting Basel II advanced internal-ratings-based models for International Reporting Standard 9." *Journal of Credit Risk* 13 (2): 53–83. DOI: 10.21314/JCR.2017.224.

Möstel, Linda, Marius Pfeuffer, and Matthias Fischer. 2020. "Statistical inference for Markov chains with applications to credit risk." *Computational Statistics* 35:1659–1684. DOI: 10.1007/s00180-020-00978-0.

Nickell, Pamela, William Perraudin, and Simone Varotto. 2000. "Stability of rating transitions." *Journal of Banking & Finance* 24 (1–2): 203–227. DOI: 10.1016/S0378-4266(99)00057-6.

Novotny-Farkas, Zoltán. 2016. "The Interaction of the IFRS 9 Expected Loss Approach with Supervisory Rules and Implications for Financial Stability." *Accounting in Europe* 13 (2): 197–227. DOI: 10.1080/17449480.2016.1210180.

Oeyen, Brent, and Oliver Salazar Celis. 2019. "On Probability of Default and its Relation to Observed Default Frequency and a Common Factor." *Journal of Credit Risk* 15 (3): 41–66. DOI: 10.21314/JCR.2019.253.

Orth, Walter. 2013. "Multi-period credit default prediction with time-varying covariates." *Journal of Empirical Finance* 21:214–222. DOI: 10.1016/j.jempfin.2013.01.006.

Pan, Jun, and Kenneth J. Singleton. 2008. "Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads." *Journal of Finance* 63 (5): 2345–2384. DOI: 10.1111/j.1540-6261.2008.01399.x.

Perilioglu, Ahment, Karina Perilioglu, and Skriye Tuysuz. 2018. “A risk-sensitive approach for stressed transition probability matrixes.” *Journal of Risk Model Validation* 12 (3): 1–24. DOI: 10.21314/JRMV.2018.192.

Pfeuffer, Marius, and Matthias Fischer. 2016. “Connecting Rating Migration Matrices and the Business Cycle By Means of Generalized Regression Models.” *Applied Stochastic Models in Business and Industry* 32 (5): 639–647. DOI: 10.1002/asmb.2184.

Pfeuffer, Marius, Linda Möstel, and Matthias Fischer. 2019. “An extended likelihood framework for modelling discretely observed credit rating transitions.” *Quantitative Finance* 19 (1): 93–104. DOI: 10.1080/14697688.2018.1465196.

Plosser, Matthew C., and João A. C. Santos. 2014. *Banks' incentives and the quality of internal risk models.* Staff Report, No. 704. Federal Reserve Bank, New York, NY.

Prudential Regulation Authority. 2017. *Internal Ratings Based (IRB) approaches.* Supervisory Statement (SS11/13). Bank of England, London. Accessed 2021-03-16. www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/policy-statement/2017/ps1317.pdf.

Restoy, Fernando, and Raihan Zamil. 2017. *Prudential policy considerations under expected loss provisioning: lessons from Asia.* Financial Stability Institute, Basel. Accessed 2020-08-28. www.bis.org/fsi/publ/insights5.pdf.

Rubtsov, Mark, and Alexander Petrov. 2016. “A point-in-time–trough-the-cycle approach to rating assignment and probability of default calibration.” *Journal of Risk Model Validation* 10 (2): 83–112. DOI: 10.21314/JRMV.2016.154.

Schutte, Willem D., Tanja Verster, Derek Doody, Helgard Raubenheimer, and Peet J. Coetzee. 2020. “A proposed benchmark model using a modularised approach to calculate IFRS 9 expected credit loss.” *Cogent Economics & Finance* 8 (1): 1735681. DOI: 10.1080/23322039.2020.1735681.

Shumway, Tyler. 2001. “Forecasting Bankruptcy More Accurately: A Simple Hazard Model.” *Journal of Business* 74 (1): 101–124. DOI: 10.1086/209665.

Skoglund, Jimmy. 2017. “Credit Risk Term-Structures for Lifetime Impairment Forecasting: A Practical Guide.” *Journal of Risk Management in Financial Institutions* 10 (2): 177–195.

Skoglund, Jimmy, and Wei Chen. 2015. *Financial Risk Management: Applications in Market, Credit, Asses and Liablity Management and Firmwide Risk.* Hoboken, NJ: John Wiley & Sons, Inc.

———. 2016. “The application of credit risk models to macroeconomic scenario analysis and stress testing.” *Journal of Credit Risk* 12 (2): 1–45. DOI: 10.21314/JCR.2016.208.

———. 2017. “Rating momentum in the macroeconomic stress testing and scenario analysis of credit risk.” *Journal of Risk Model Validation* 11 (1): 21–47. DOI: 10.21314/JRMV.2017.167.

———. 2018. “Stage Transfer Effect on Impairment Forecasts.” *Risk Management in Financial Institutions* 11 (3): 244–256.

Štěpánková, Barbora. 2021. “Bank-sourced credit transition matrices: Estimation and characteristics.” *European Journal of Operational Research* 288 (3): 992–1005. DOI: 10.1016/j.ejor.2020.06.024.

Thomas, Lyn, Jonathan Crook, and David Edelman. 2017. *Credit Scoring and Its Applications.* 2nd ed. Philadelphia (PA): Society for Industrial and Applied Mathematics.

Treacy, William F., and Mark S. Carey. 1998. “Credit risk rating at large US banks.” *Federal Reserve Bulletin* 84:897–921.

Trück, Stefan. 2008. "Forecasting credit migration matrices with business cycle effects — a model comparison." *The European Journal of Finance* 14 (5): 359–379. DOI: 10.1080/13518470701773635.

Vallascas, Francesco, and Jens Hagendorff. 2013. "The Risk Sensitivity of Capital Requirements: Evidence from an International Sample of Large Banks." *Review of Finance* 17 (6): 1947–1988. DOI: 10.1093/rof/rfs042.

Varetto, Franco. 2018. "Point-in-time vs. through-the cycle: filosofie di rating a confronto." *Working Paper IRCrES* 4 (11): 1–83. DOI: 10.23760/2421-7158.2018.011.

Vassalou, Maria, and Yuhang Xing. 2004. "Default Risk in Equity Returns." *Journal of Finance* 59 (2): 831–868. DOI: 10.1111/j.1540-6261.2004.00650.x.

Wei, Jason Z. 2003. "A multi-factor, credit migration model for sovereign and corporate debts." *Journal of International Money and Finance* 22 (5): 709–735. DOI: 10.1016/S0261-5606(03)00052-4.

Xu, Xin. 2016. *Estimating Lifetime Expected Credit Losses under IFRS 9.* Working Paper. Accessed 2021-12-08. http://ssrn.com/abstract=2758513.

Yang, Bill H. 2017a. "Forward Ordinal Probability Models for Point-in-Time Probability of Default Term Structure: Methodologies and Implementations for IFRS 9 Expected Credit Loss Estimation and CCAR Stress Testing." *Journal of Risk Model Validation* 11 (3): 1–18. DOI: 10.21314/JRMV.2017.181.

———. 2017b. "Point-in-Time Probability of Default Term Structure Models for Multiperiod Scenario Loss Projection." *Journal of Risk Model Validation* 11 (1): 73–94. DOI: 10.21314/JRMV.2017.164.

———. 2018. "Smoothing Algorithms by Constrained Maximum Likelihood." *Journal of Risk Model Validation* 12 (2): 1–14. DOI: 10.21314/JRMV.2018.189.

Yang, Bill H., Biao Wu, Kaijie Cui, Zunwei Du, and Glenn Fei. 2020. "International Financial Reporting Standard 9 expected credit loss estimation: advanced models for estimating portfolio loss and weighting scenario losses." *Journal of Risk Model Validation* 14 (1): 19–34. DOI: 10.21314/JRMV.2020.217.

Yang, Jimmy, and Kenneth Chen. 2018. "Implementing Both IFRS 9 and CECL." In *The New Impairment Model under IFRS 9 and CECL,* edited by Jing Zhang. London: Risk Books.

Zhang, Xin, and Tony Tung. 2019. "On the mathematical modeling of point-in-time and through-the-cycle probability of default estimation/validation." *Journal of Risk Model Validation* 13 (1): 39–61. DOI: 10.21314/JRMV.2018.199.