

Research article

Estimating credit default probabilities using stochastic optimisation

Dominic Joseph*

Independent researcher, Kerala, India

* Correspondence: Email: dom.jo@cet.ac.in.

Abstract: Banks and financial institutions all over the world manage portfolios containing tens of thousands of customers. Not all customers are high credit-worthy, and many possess varying degrees of risk to the Bank or financial institutions that lend money to these customers. Hence assessment of default risk that is calibrated and reflective of actual credit risk is paramount in the field of credit risk management. This paper provides a detailed mathematical framework using the concepts of Binomial distribution and stochastic optimisation, in order to estimate the Probability of Default for credit ratings. The empirical results obtained from the study have been illustrated to have potential application value and perform better compared to other estimation models currently in practise.

Keywords: credit risk; default rate; probability of default; calibration; stochastic optimization

JEL Codes: C02, C15, C61, C63, G32

1. Introduction

Probability of default (PD) is a financial risk management term describing the likelihood of a default over a particular time horizon. It provides an estimate of the likelihood that a borrower will be unable to meet its debt obligations. Under Basel II guidelines, formulated by the Basel Committee on Banking Supervision or BCBS (2001), PD is a key parameter used in the calculation of economic capital or regulatory capital for a banking institution. Banks that comply with the new Basel II internal ratings-based approach are obliged to assign a PD value (usually a 1-year PD) to their clients, which are used as the basis for regulatory capital requirements.

A popular method of quantifying Probability of default is through credit ratings, where each entity in a bank's portfolio is assigned a rating grade depending on its past, current and future behaviour.

Entities with same rating grade represents similar credit risk. Some banks have internal score card and rating models that assign ratings based on internal methodologies. On occasions, banks may use ratings supplied by external credit rating agencies such as Moody's and Standard & Poor's (S&P) instead of internal ratings. For example, S&P Global ratings provide a rating system with 22 rating classes and Moody's, which is another popular rating agency classifies entities to 21 rating classes based on their credit worthiness. This is permissible because Basel II accord allows banks to base their capital requirements on internal as well as external rating systems and risk profile.

Large banks employ various mathematical techniques using the historical data of customer or account level ratings assigned by its internal ratings system (or external rating system) to forecast the default probability of each entity. The number of grades used by the bank depends on the bank's individual preference. The "upper" grades represent lower default risk and hence lower probability values are assigned to them while "lower" grades represent higher default risk and consequently, higher probability values are assigned to them.

One of the important concerns in the financial industry is regarding the estimation of true Probability of Default that captures the true default risk associated with a portfolio, to the best extend. Empirically estimated PD values based on sample data often deviate from expectations in that the observed PD values of rating classes are non-monotonic in nature (Pluto and Tasche, 2005). That is, $R' > R$ (R' is "lower" grade) need not imply $\hat{\theta}_{R'} > \hat{\theta}_R$, where $\hat{\theta}_R$ is the estimated default rate for rating grade R . Tasche (2013) discussed this issue which he termed it as the "Inversion of default rates". Non monotonicity is an issue because the observed default rates do not reflect the true default risk associated with the rating class. Krahnen and Weber (2001) mentions monotonicity as one of the most important requirements for a rating system. It is easy to see that without monotonicity, the situation can lead to issues like misclassification of some customers into wrong risk buckets with under-estimation or over-estimation of default probability. It also affects the discriminatory power of a rating system thereby hindering meaningful risk differentiation between less risk and high-risk rating classes.

One reason for this deviation between expected values and observed estimates could be that the portfolio often has an uneven distribution of accounts in the rating classes (Tasche, 2013). It is commonly observed in many empirical data that are analysed, that there is insufficient information in some rating groups, thereby making it impossible to obtain an accurate estimate of its default risk from the observed default rate itself. For example, Table 1 shows the observed default rates based on historical rating assignments (Long-Term Foreign-Currency) of corporate assets by Standard & Poor's Ratings Services for the years 2017 and 2016. Note that out of the two years, there is a break in monotonicity in the empirically obtained default rates for the year 2016, while 2017 maintains an expected trend in default rate. From this observation we can conclude that the default rates in 2016 does not give a true picture of the default risk in 2016, at least for two rating classes. Such inversions need not be a problem stemming from the rank-ordering capacity of the rating methodology, but a result of randomness in the default rate of customers. The issuer weighted long-run average grade-level default rates reported by S&P report (2021) suggests that the observation of default rate inversions as in Table 1 might be an exception.

Table 1. S&P default frequencies and default rates (%) for corporate entities in 2016 and 2017 based on Long Term Foreign-Currency issuer rating.

Rating	2016		2017	
	Default Frequency	Default Rate	Default Frequency	Default Rate
AAA	0	0.0%	0	0.0%
AA	0	0.0%	0	0.0%
A	0	0.0%	0	0.0%
BBB	0	0.0%	0	0.0%
BB	60	4.1%	2	0.1%
B	25	2.0%	14	0.5%
CCC	38	11.6%	31	8.3%
CC	1	3.4%	4	14.3%
C/D	Default Bucket			

In such cases, it is safe to assume that the observed default rate is different from the true default rate. A lack of sufficient data might lead to a difficulty in calculating the true probability of default value for such rating grades. It is not just in simple non-parametric models, that the issue may come up but also in classification models like logistic regression or decision tree models. Such models give us the probability of a customer or entity defaulting, given a set of customer level or entity level factors that influence the default risk (credit rating might be one of the factors here). But once the model is built, there is no guarantee that the aggregate risk of customers when grouped at a rating class level follows a monotonic trend. Hence there is a need to calibrate the estimated default probabilities such that, $R' > R \rightarrow \hat{\theta}_{R'} > \hat{\theta}_R$.

One of the early and most popular techniques used in the industry was developed by Pluto and Tasche (2005), who suggested a low-default portfolio calibration approach, termed as the *most prudent estimation*, using upper confidence bound with confidence level $(1 - \alpha)$, while guaranteeing an ordering of PDs that respects the differences in credit quality as indicated by the rating grades. This method involves optimising an objective function (a cumulative binomial likelihood) by assuming a binomial distribution of defaults and a suitable value for significance level α . Later, some alternatives to the method by Pluto and Tasche was proposed in the form of rating systems or score functions for low default portfolios, see Erlenmaier (2011); Fernandes and Rocha (2011).

Dwyer (2006), introduced a Bayesian adjustment to Pluto and Tasche's model with the use of a Uniform prior distribution. His approach was later explored in greater detail in Kiefer (2009), Kiefer (2010) where using prior distributions (like Beta distribution) determined by expert judgement was considered. Other authors that considered the application of the Bayesian approach are Tasche (2013), Clifford et al. (2013), Chang and Yu (2014) and Kruger (2015). But most of these methods cannot be implemented on a rated system to ensure monotonic PD trend estimate.

Van der Burgt (2007) suggested a method for estimating low-default portfolio PD curves by using the cumulative accuracy profile (CAP), also known as the power curve or Lorenz curve, and a mathematical function for modelling the CAP, that ensures calibrated monotonic PD curves. This is also called *CAP Curve Calibration* or *VDB Calibration* technique. Tasche Dirk (2009) proposed a two-parameter approach called *Quasi Moment Matching* (QMM) based on Accuracy Ratio and distribution of good accounts. Here, the PD or default probability for each rating is modelled as a

mathematical function of the distribution of non-defaulted population in the portfolio and solved by optimising an objective function of the two user-defined parameters.

A recent work in the field was by Surzhko Denis (2017) who proposed a Bayesian method where the portfolio default rate itself rather than rating level is calibrated to the default rate of another portfolio called the *closest available portfolio* by adjusting the Bayesian prior density of the default rate parameter. Although the method can be adopted for a rating level system, it involves identifying a closely related rated portfolio with reliable and monotonic default statistics and a suitable prior distribution weight.

Almost all the methods which are presently, widely used involve either some level of subjectivity or assumptions in the theoretical framework (discussed briefly in section 3.3). In many cases, these methods are observed to under-predict or over-predict the estimates. Moreover, in some scenarios they have been observed to perform poorly and produce results which are non-intuitive and illogical such as uncalibrated default rates (estimated PDs decrease with decrease in credit quality) in certain types of portfolios. For example, Pluto and Tasche (2005) mentions that the disadvantage of the most prudent approach is that it may lead to non-monotone PD estimates when there are many defaults in high rating grades and less in lower grades.

The main motivation for this study was the gaps in the estimation frameworks of current methods in practise (discussed in brief in section 3.3) and possibility of overcoming the same while improving the performance. In short, the main objective of this study is to;

- A. Introduce an alternative PD estimation technique that;
 - a. Offers better results with minimum under-estimation or over-estimation.
 - b. Performs well under a wide variety of scenarios and different types of portfolios.
 - c. Reduces the subjectivity in the model and hence the sensitivity of estimates due to subjective influences.
- B. Evaluate and compare the performance of the PD technique considered in this study using real data.
- C. Indicate open questions for further research.

This paper proposes an estimation method called stochastic optimisation for achieving the default rate curve estimation and calibration. Using stochastic optimisation principles, we attempt to estimate the default parameter $\theta_R(t)$ for each rating class R and period t given the data. The estimated default rate parameter acts as a best-case estimate of Probability of Default (PD) satisfying all the underlying conditions.

Stochastic optimization refers to a collection of methods for minimizing or maximizing an objective function when randomness is present. They are widely used in the fields of science, business, engineering etc... A stochastic method is chosen assuming that the likelihood function of default distribution of rating classes θ_R subject to a set of given restrictions, is an expression without any closed form solution. Although the idea of stochastic optimisations is not new in general (Kirkpatrick et al., 1983), the method has seen limited application in the field of credit risk. Thus, the idea presented in this paper is novel with promising avenues of further research. The reason for choosing stochastic optimisation is that unlike other techniques which are widely used (like GRG linear) which gives local maxima/minima, stochastic methods give global maxima/minima.

The proposed estimation framework can be divided into two parts;

- A. Defining a mathematical objective function of the default rate parameter θ_R that needs to be optimised.
- B. Optimising the objective function using a suitable technique that gives the best-case estimates for PD.

The theoretical framework for the entire process has been discussed in detail in section 2. The proposed methodology was implemented and tested using two different data sets, which are explained briefly in section 3.1. The results of the study are presented in section 3.2. The results are then

compared to the outputs from other well-known approaches, which is discussed in Section 3.3, where it is explained how the proposed method is better than other methods. Few limitations have been discussed in section 4 and we conclude the article in section 5.

2. Methodology

2.1. Optimisation function and constraints

Binomial distribution is the distribution of the number of successes that occur in N independent trials with the probability of success in each trial is θ . If we consider N to be the number of performing entities at the beginning and k denotes the number of entities at the end who defaulted, then the distribution of number of defaults can be considered as a binomial distributed random variable with likelihood

$$p(\theta, N, k) = \binom{N}{k} \theta^k (1 - \theta)^{N-k}, \quad (1)$$

where θ is the true probability of occurrence of a default event of a single entity.

Given a set of rating grades 1,2,3..m with initial distribution of performing entities $\{N, N_2, \dots, N_m\}$ and distribution of defaulted entities $\{x_1, x_2, \dots, x_m\}$, we need to estimate the default rate parameter $\{\theta_1, \theta_2, \dots, \theta_m\}$ the true value of which follows the condition g given by

$$g: 0 \leq \theta_1 \leq \theta_2 \leq \theta_3 \leq \dots \leq \theta_m \leq 1. \quad (2)$$

Under the assumption that number of defaults is binomially distributed, $x_i \sim \text{Binom}(\theta_i, N_i)$ and default events are independent, i.e. default in one rating grade does not influence another rating grade, the Maximum likelihood Estimate of θ_i^0 for each rating grade is just the ratio of x_i and n_i .

$$\theta_i^0 = \frac{x_i}{N_i}. \quad (3)$$

It is taken as assumption that the reason that the Maximum Likelihood parameter breaks the order is due to limitations discussed earlier. The solution presented through this paper to this problem is to optimise the joint likelihood function of observing the given default data $\{x_1, x_2, \dots, x_m\}$ for rating grades 1,2,3..m, by varying the corresponding default rate parameters $\{\theta_1, \theta_2, \dots, \theta_m\}$.

$$\hat{\theta}_i = \text{argmax}(\text{Likelihood}(x_i | 0 \leq \theta_1 \leq \theta_2 \dots \leq \theta_m \leq 1)). \quad (4)$$

Before we find a way to achieve such a solution, we need to implement an additional constraint on the parameters. If we solve our problem with only constraint g , it can occasionally lead to results which optimise the likelihood function but fails to quantify the true default risk. That is, the optimising algorithm may give unrealistically high or insignificantly low values of PD estimates as the optimal solutions to the objective function, but still satisfying condition g . For example, during trials with one of the data sets under study, it was observed that the solution converged to high PD values even for investment grade ratings (high quality ratings where PD is expected to be < 1%). Figure 1 displays such a case where the estimated PD values begin at around 30% (grade 1) and ends at around 80% (grade 5) while the observed default rate is significantly below these values.

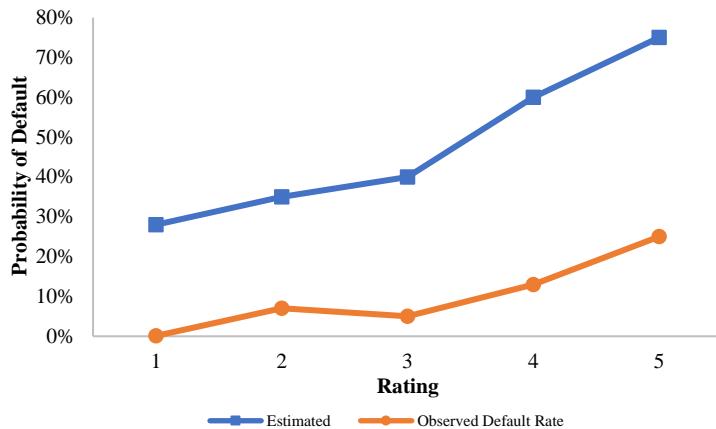


Figure 1. Estimated PD vs Observed Default rate (illustrative).

Remember that we already established an assumption that the observed value of the parameters may be different from the true value, and the variation is caused due to randomness in the default events. It is prudent approach to extend that assumption to further include a condition that, for each rating, the likelihood of data given default parameter $\hat{\theta}_i$ is within $1 - \alpha\%$ confidence interval of the likelihood of data given the maximum likelihood estimate θ_i^0 of the parameter θ_i . The idea of this constraint is to prevent estimates of default rates which maximises our objective function and follows rank order at the same time prevents results which are far off from the expected range of results. Figure 2 shows an illustrative example, where the estimated value of the parameter θ_i is outside the $1 - \alpha\%$ bounds of the MLE of the parameter.

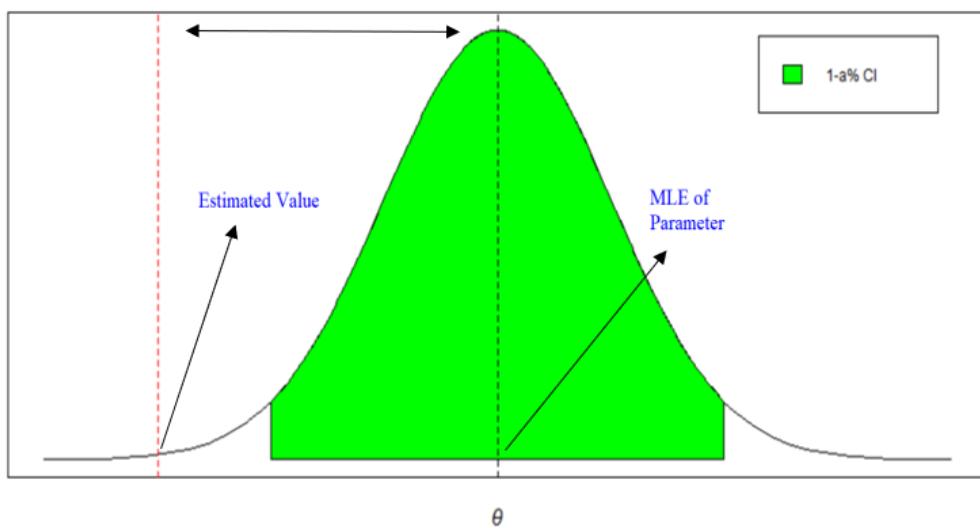


Figure 2. Spread of PD parameter θ showing a $1 - \alpha\%$ confidence interval (illustrative). Black line shows the Maximum Likelihood estimate of θ . Red line shows the estimated PD value for θ .

This new constraint can be implemented using the likelihood ratio test, as shown below,

$$\text{Likelihood ratio} = \lambda = \frac{L(x|\theta_i^0)}{L(x|\theta_i^*)}. \quad (5)$$

where θ_i^* is the candidate θ_i and θ_i^0 is the MLE of θ_i . The test statistic $-2 \log \lambda$ approximates a chi-squared random variable with degree of freedom equal to 1.

$$-2 \log \lambda \sim \chi_1^2(\alpha). \quad (6)$$

So, the estimate should be such that the likelihood ratio should be within pre-defined thresholds given by the condition h given by

$$h: -2 \log \lambda \geq c \sim \chi_1^2(\alpha). \quad (7)$$

So, the problem is redefined as to optimise the joint likelihood function subject to the constraints g and h .

$$\hat{\theta}_i = \text{argmax}(\text{Likelihood}(x_i|g, h)). \quad (8)$$

Likelihood of m binomial distributed random variables is just the product of their binomial probability densities (this follows since the random variables are iid)

$$\text{Likelihood}(x|\theta) = \prod_{i=1}^m \binom{N_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{N_i - x_i}. \quad (9)$$

For mathematical convenience, the product can be converted into addition by using a logarithmic function. So instead of maximising the likelihood function, we may maximise the log of the likelihood.

$$LL(x|\theta) = \text{Log Likelihood}(x|\theta) = \sum_{i=1}^m \left\{ \log \binom{N_i}{x_i} + x_i \log(\theta_i) + (N_i - x_i) \log(1 - \theta_i) \right\}. \quad (10)$$

Since we have a constraint in the form of conditions g and h $g: (0 \leq \theta_1 \leq \theta_2 \dots \leq \theta_m \leq 1)$, and $h: -2 \log \lambda \geq c \sim \chi_1^2(\alpha)$ we need to handle them in the objective function. A variety of constraint handling methods have been suggested by many researchers. Each method has its own advantages and disadvantages. The most popular constraint handling method among users is *penalty function method*. The penalty function assigns a static exterior penalty value (Homaifar et al. 1994) to the likelihood function in such a way that the likelihood is decreased for each instance of breach in constraint. Hence for a number of constraints, our optimisation function may be rewritten as

$$f(x) = LL(x|\theta) + \nu_k \sum_{k=1}^{c1} I_{g_k} + \eta_j \sum_{j=1}^{c2} I_{h_j}, \quad (11)$$

where $\nu_k < 0$ is the exterior penalty value for condition g and I_{g_k} is the indicator that k th constraint in condition g is breached. $\eta_j < 0$ is the exterior penalty value for condition h and I_{h_j} is the indicator that j th constraint in condition h is breached. So, each violation in constraint would lead to a penalty being applied to the likelihood, thereby decreasing the function $f(x)$. Applying the negative penalties would allow the algorithm to optimise values in the region where no violation occurs or where violations are as minimum as possible.

Therefore, the optimisation problem is formulated as

$$\hat{\theta}_i = \operatorname{argmax} \left(LL(x|\theta) + \nu_k \sum_{k=1}^{c1} I_{g_k} + \eta_j \sum_{j=1}^{c2} I_{h_j} \right). \quad (12)$$

This expression has no bounded solution and normal optimisation techniques may not converge to a proper solution that satisfies all constraints. Hence, one can use simulation-based optimisation techniques, also called stochastic optimisation, to estimate the calibrated parameters that maximise our objective function.

2.2. Optimisation using Simulated Annealing

Stochastic optimization refers to a collection of methods for minimizing or maximizing an objective function when randomness is present. Unlike deterministic optimisation techniques like gradient descent, which give local maximum or minimum, stochastic methods can give global maximum or minimum. It is suitable for problems where finding an approximate global optimum is more important than finding a precise local optimum in a fixed amount of time. There are multiple models for stochastic optimisation, each of which may be suitable for specific problem. The solution discussed in this paper has been achieved through a method called Simulated Annealing.

Simulated annealing is a general probabilistic local search algorithm, proposed 20 years ago by Kirkpatrick et al. (1983) to solve difficult optimization problems. The algorithm is based upon Annealing process used in material science field. Annealing is the process of heating up a material (mostly metals) until it reaches a fixed temperature and then it will be cooled down slowly in order to change the material to a desired structure. When the material is hot, the molecular structure is weaker and is more susceptible to change. When the material cools down, the molecular structure is harder and is less susceptible to change.

Simulated Annealing (SA) mimics the Annealing process but is used for optimizing parameters in a model. This process is very useful for situations where there are a lot of local minima such that algorithms like Gradient Descent would be stuck at. Simulated Annealing is similar to an MCMC process in that the next value in the iteration depends on the current sample generated. A neighbouring solution is found as a new candidate solution by applying a random perturbation to the current solution using a candidate generator function $\pi(\cdot)$. This randomness helps in preventing from getting stuck in “local minima”. If the selected move improves the solution, then it is always accepted. Otherwise, the algorithm makes the move with some probability less than 1. The probability decreases exponentially with the “badness” of the move, given by the following Equation.

$$P = e^{-\frac{L(\theta_{new}) - L(\theta_{old})}{T}} \quad (13)$$

where, $L(\theta)$ is the objective function to minimise. The parameter T used in the Equation is analogous to temperature in an actual annealing process. At higher values of T , uphill moves are more likely to occur. As T tends to zero, they become more and more unlikely, until the algorithm behaves more or less like an optimisation near local minima. The steps have been described in section 2.3.

The main advantages of choosing Simulated Annealing are the following;

- It ensures a “good” global optimal solution with decent computation time.

- It is easy to implement using a software language such as R.
- It can deal with complex cost functions.

The R code used for the process has been provided in Appendix B.

2.3. Sampling algorithm

The optimisation process is performed in an iterative and step by step manner. Note that instead of trying to maximise the objective function (Equation 12), we are trying to minimise the negative value of the same.

- For each rating we first start with an initial solution $\theta_i = \theta_{i,j=0}$. We also start with an initial temperature $t = t_0$.
- For $j = 1$ to N , propose new θ_i^* from $\pi(\theta_{i,j-1})$ where π is a candidate distribution.
- Check if constraint h is satisfied. Replace θ_i^* with $\theta_{i,j-1}$ for ratings where likelihood ratio test fails.
- Calculate the negative of the objective function given by Equation 12 for the new and old parameters θ_i^* and $\theta_{i,j-1}$ respectively.
- Calculate the acceptance probability function $P = e^{-\frac{L(\theta_i^*) - L(\theta_{i,j-1})}{T}}$.
- Generate u from $U(0, 1)$.
- Accept θ_i^* as $\theta_{i,j}$ if $P > u$, or if the objective function value at θ_i^* < objective function value at θ_i^* . Else reject.
- Decrease t by a factor given by $\left(\frac{T_f}{T_0}\right)^{\frac{i-1}{N}}$ and repeat from step ii.

Because the algorithm is sensitive to initial conditions, it can have an impact on the results as the algorithm may not converge to an optimal solution. So, these steps were performed multiple times to gather the optimal results from each run, which were then compared to obtain the final results. This would help in maintaining the objectivity of the output to a huge extend.

3. Implementation

3.1. Data

“Standard & Poor’s Ratings Services” issues credit ratings for the debt of public and private companies, and other public borrowers such as governments and governmental entities. It rates customers into 22 ratings based on their performance, with AAA being the best rating and minimal chance of default and D being the worst performance rating which stands for Default. Intermediate ratings are offered at each level between AA and CCC (such as BBB+, BBB, and BBB-). For the purpose of the modelling, Long-Term issuer ratings (Foreign-Currency) data of corporate customers rated by S&P was obtained¹. The data consisted of ratings for the years 2016 & 2017.

¹Under SEC Regulation 17g-7, Nationally Recognized Statistical Rating Organizations (NRSRSOs) are required to report their historical rating assignments, upgrades, downgrades and withdrawals since 2010. Rating data are generally reported on a one year delay. CSV format of this data was sourced from the website www.ratingshistory.info.

For the purpose of simplicity and overcoming data challenges like lack of enough defaults in some rating grades, a binning process² was carried out where in adjacent ratings were clubbed together to form a new rating group with 9 ratings. For example, ratings BBB+, BBB and BBB- were merged to form a new group BBB. Also, Ratings C and D were considered to be the default bucket with rating order 9.

December snapshots for the years 2011 to 2017 were analysed and the year 2016 was chosen for the modelling. A 12-month window was considered as an ideal observation window because banks and financial institutions are usually interested in measuring probability of default over a 1-year time horizon. For each snapshot period, the number of performing entities in the beginning and the number of default cases (from the performing entities) at the end of the 12 month window was measured for each rating class separately. See Table 2 for details of obligor and default counts.

Table 2. S&P corporate credit rating (Long-Term Foreign-Currency) and corresponding numerical order.

Rating	Rating Order	Year = 2016	
		Defaults at End	Performing at Start
AAA	1	0	14
AA	2	0	153
A	3	0	934
BBB	4	0	1814
BB	5	60	1470
B	6	25	1225
CCC	7	38	329
CC	8	1	29

S&P data had rich default rates in many rating grades. The overall number of defaults was 124 for all the 8 rating grades combined, so a low default sample data was created by simulation (assuming a portfolio of only 8 ratings), for the purpose of testing the algorithm on a low default portfolio. The manually generated data has been shown in Table 3 given below.

Table 3. Sample data created by simulation.

Rating	Defaults at End	Performing at Start
1	0	1
2	0	6
3	1	20
4	0	21
5	2	16
6	1	11
7	0	20
8	0	2

²Note that this step is irrelevant from the point of view of actual modelling process and was chosen only for the purpose of convenience.

Note that there are only 4 default instances in this portfolio and ratings 7 and 8 with highest default risk has no default instances. Such scenarios may seem far-fetched but are plausible in extremely low default portfolios such as sovereign portfolios.

The method discussed in section 2 was successfully implemented using R programming language and the two data sets, and the results are discussed in the following section.

3.2. Experimental results

Using the method of simulated annealing with constrained parameters, our objective function is maximised for the two data considered for the experiment. Significance level α for testing condition h was taken as 2.5%. Candidate distribution $\pi(\cdot)$ for determining the next sample was chosen as $Normal(mean = \theta, variance = \sigma^2)$. The variance of the candidate distribution was chosen as the one that gave optimal results, based on trial and error. It was observed during trial runs that high variance would often lead to failure in convergence to an optimal solution, as the range of search space was quite narrow (θ in (0,1)) and on top of that additional constraints have been applied. The calibrated results for the two scenarios are as shown in the Table 4 below.

Table 4. S&P data—estimated PD values for 50,000 simulations.

$N = 5 \times 10^5, \alpha = 2.5\%, v = -1000, \eta = -900, \sigma^2 = 0.0005, T_0 = 100$												
Rating	Default Rate	PD	Upper Bound Confidence Intervals			PD	Upper Bound Confidence Intervals					
			Estimate	$\alpha = 5\%$	50%	75%	90%	Estimate	$\alpha = 2.5\%$	50%	75%	90%
1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2	0.00%	0.00%	0.00%	0.01%	0.03%	0.00%	0.00%	0.00%	0.01%	0.01%	0.03%	0.03%
3	0.00%	0.01%	0.02%	0.04%	0.06%	0.00%	0.02%	0.04%	0.07%	0.04%	0.07%	0.07%
4	0.00%	0.02%	0.06%	0.07%	0.11%	0.01%	0.05%	0.09%	0.11%	0.05%	0.09%	0.11%
5	4.08%	3.07%	3.04%	3.05%	3.06%	3.06%	3.04%	3.06%	3.07%	3.04%	3.06%	3.07%
6	2.04%	3.07%	3.06%	3.07%	3.08%	3.07%	3.06%	3.07%	3.08%	3.06%	3.07%	3.08%
7	11.55%	10.87%	10.82%	11.12%	11.61%	11.00%	10.87%	11.22%	11.80%	10.87%	11.22%	11.80%
8	3.45%	11.07%	11.16%	12.05%	14.21%	11.04%	11.17%	12.59%	14.03%	11.04%	11.17%	12.59%

We can see from the table that the estimates from the code gives monotonic PD values from the default rates for $\alpha = 2.5\%$ and 5% . For $\alpha = 5\%$, the algorithm did not converge to an optimal solution and resulted in rejections for majority of the samples. There appears to be significant variation in the estimated PD values in the case of the 8th rating grade. For the rest of the ratings, the algorithm seems to give relatively close estimates. The results can further be visualised from Figure 3.

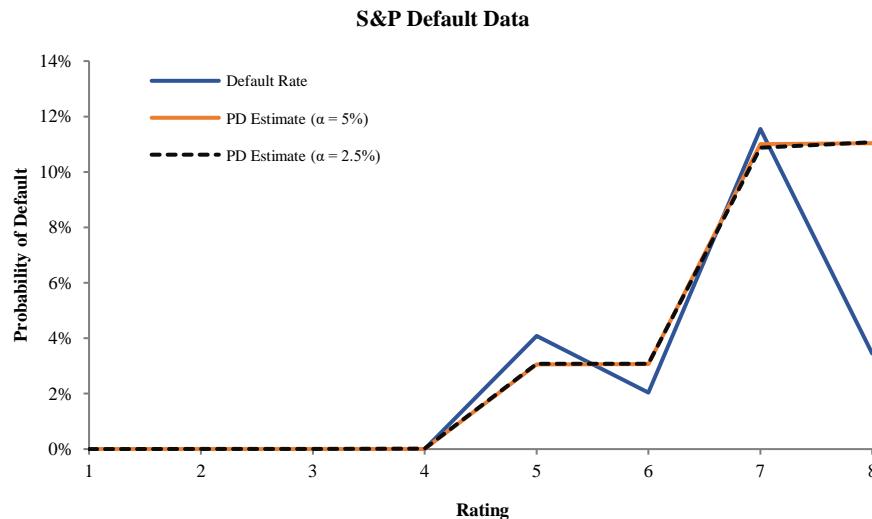


Figure 3. Calibrated PDs vs observed default rates for S&P data.

The estimates of PD for the 2nd data set for $\alpha = 2.5\%$ and 5% are shown in Table 5. The trend can be visualised from Figure 4.

Table 5. Simulated portfolio—data estimated PD values for 50,000 simulations.

$N = 5 \times 10^5, v = -1000, \eta = -900, \sigma^2 = 0.00005, T0 = 100$										
Rating	Default Rate	PD Estimate	Upper Bound Confidence Intervals			PD Estimate	Upper Bound Confidence Intervals			
			Intervals				Intervals			
			$\alpha = 5\%$	50%	75%	90%	$\alpha = 2.5\%$	50%	75%	90%
1	0.00%	0.00%	0.00%	0.05%	0.36%	0.36%	0.00%	0.00%	0.00%	0.32%
2	0.00%	0.17%	0.28%	0.58%	1.05%	0.74%	0.32%	0.65%	0.93%	
3	5.00%	2.51%	2.58%	2.79%	3.02%	2.74%	2.48%	2.67%	2.84%	
4	0.00%	2.51%	2.69%	2.87%	3.19%	2.74%	2.55%	2.76%	2.96%	
5	12.50%	6.06%	6.20%	6.50%	6.83%	6.09%	6.34%	6.77%	7.04%	
6	9.09%	6.28%	6.29%	6.65%	6.96%	6.13%	6.42%	6.82%	7.18%	
7	0.00%	6.28%	6.38%	6.76%	7.06%	6.17%	6.52%	6.91%	7.31%	
8	0.00%	6.84%	13.55%	22.60%	31.64%	6.18%	16.57%	24.77%	32.10%	

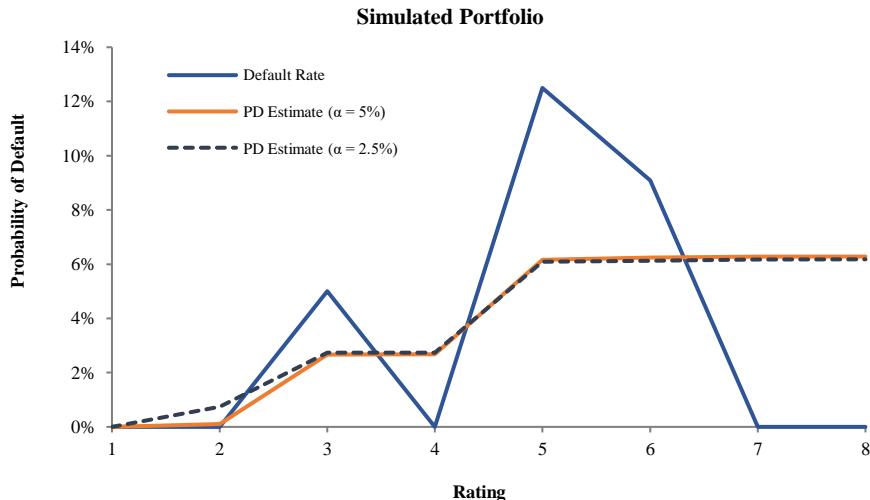


Figure 4. Calibrated PDs vs observed default rates for simulated data.

3.3. Comparison with other methods

The results obtained from the method were compared with other popular calibration techniques in the industry to get a comparison of estimates. The following are the three major popular methods used currently for estimation and calibration of default rates.

A. The *most prudent estimation* PDs suggested by Pluto and Tasche (2005).

This technique uses a cumulative binomial distribution function to estimate the probability of default for each rating class, as shown below:

$$1 - \alpha \leq F_{Bin}(N_{i \geq R}, D_{i \geq R}, \theta_R). \quad (14)$$

where F_{Bin} is the cumulative binomial distribution function, $N_{i \geq R}$ is the cumulative number of customers in rating class R and $D_{i \geq R}$ is the cumulative number of defaults in rating class R . α is the level of significance which is subjectively chosen. θ_R is the PD value which solves the Equation for given α .

One major drawback is that while estimation, this binomial probability is solved using cumulative distribution of customers and defaults in each bucket rather than actual number of customers and defaults. The results also depend on the value of α and there is no formal rule for an appropriate value that can give an optimum result and this is of large impact on the PD estimates. Also, this method has been observed to produce non-intuitive results such as trend reversals, in many cases.

Generally, this method is used for estimating PDs for low default portfolios only.

B. CAP Curve Calibration or VDB Calibration suggested by Van der Burgt (2008).

This method assumes an arbitrary functional form assigned to the cumulative PD curve given by

$$\theta_R = \frac{1 - e^{-kR}}{1 - e^{-k}}. \quad (15)$$

where, k is a factor which depends on the AUC value of the cumulative accuracy profile. θ_R is the cumulative theoretical default rate at rating R . The objective function in this method is given by the sum of squared error between cumulative PD estimates and cumulative actual default rates. Note that

Equation 15 is just arbitrarily selected based on the general trend of a cumulative accuracy profile, unlike the objective function defined in Equation 11.

$$\sum_{R=1}^m (y_R - \theta_R)^2. \quad (16)$$

Where, y_R is the cumulative actual default rate, m is the number of ratings.

Using this objective function, the method tries to fit a best possible cumulative accuracy profile by minimising the objective function, thereby calibrating the PD estimates. The drawback of this method is that it gives poor estimates if the accuracy profile of the actual portfolio itself is poor.

C. *Quasi Moment Matching* (QMM) proposed by Tasche Dirk (2009).

Similar to *CAP Curve Calibration*, this method uses cumulative distribution of non-defaults and defines the PD as a mathematical function of it. But unlike the former, it does not assign a functional form to the cumulative distribution. It rather gives a functional form to the PD value for each rating.

$$\theta_R = \frac{1}{1 + e^{\alpha + \beta \phi^{-1}(G_R)}}. \quad (17)$$

where, G_R is the cumulative actual non-default rate and θ_R is the PD value at rating R . β and α are parameters of the functional form which needs to be optimised. The optimisation function also differs in that the sum of squared error between PD estimates and actual default rates are used in this method.

$$\sum_{R=1}^m (y_R - \theta_R)^2. \quad (18)$$

Where, y_R is the actual default rate, m is the number of ratings.

Similar to previous method, there is a drawback that it gives poor estimates if the accuracy profile of the actual portfolio itself is poor.

3.3.1 S&P default data

The values of the PD estimates from each estimation technique, have been displayed in the Table 6. From the table, we can observe a significant variation among the different methods, especially in the worst rating grades. A visual comparison of PD estimate trend can be observed from Figure 5.

Table 6. Comparison of PD results for S&P data.

Rating	Pluto Tasche ($\alpha=75\%$)	CAP Calibration	QMM	SA ($\alpha = 5\%$)	SA ($\alpha = 2.5\%$)
1	2.22%	0.17%	0.03%	0.00%	0.00%
2	2.22%	0.19%	0.09%	0.00%	0.00%
3	2.28%	0.32%	0.27%	0.01%	0.00%
4	2.72%	0.76%	0.71%	0.02%	0.01%
5	4.33%	2.23%	1.63%	3.07%	3.06%
6	4.43%	5.47%	3.68%	3.07%	3.07%
7	12.22%	7.79%	9.82%	10.87%	11.00%
8	9.02%	8.23%	24.89%	11.07%	11.04%

For S&P data, the comparison exercise performed using *Pluto-Tasche Calibration*, *QMM* and *CAP calibration method* suggests that simulation-based approach provides reasonable results in line with other industry-best practise techniques. The values are neither under-predicting nor over-predicting in comparison to the other models. Both $\alpha = 5\%$ and $\alpha = 2.5\%$ gave close estimates.

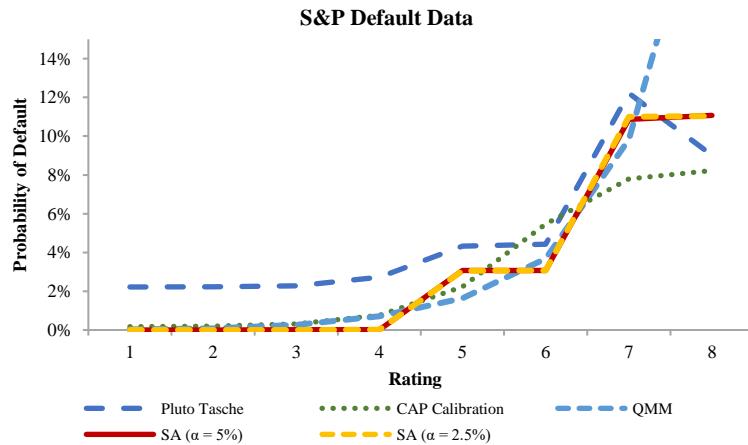


Figure 5. Comparison of results with estimates obtained from other methods for S&P Data.

We can quantitatively analyse the performance of PDs generated from Simulated Annealing framework using few statistical tests. Two tests have been conducted to determine if the observed default rate estimate is statistically different from the estimated PD value.

- The Binomial Test is a hypothesis test for measuring if the theoretically expected parameter is significantly different from the observed value of the parameter. It assumes the Null hypothesis that the estimated value of the parameter is the true value. If the Null hypothesis is supported by the observed data, then the observed parameter should fall within a $1-\alpha\%$ confidence interval of the estimated parameter. This test has no test statistic.
- The Likelihood Ratio Test is a simple hypothesis test based on the ratio of likelihood between Null distribution and Alternate distribution. The Null Hypothesis is that the observed default rate is the true parameter and the Alternate Hypothesis is that the estimated value is the true parameter. If the null hypothesis is supported by the observed data, the two likelihoods should not differ by more than sampling error and the test statistic would be small. The test statistic $-2 \log \lambda$ approximates a chi-squared random variable with degree of freedom equal to 1.

The number of ratings which failed both the tests are as shown in Table 7.

Table 7. Number of Ratings which failed performance tests.

Method	Binom Test (95% CI)	Binom Test (90% CI)	Lik Ratio Test ($\alpha = 2.5\%$)	Lik Ratio Test ($\alpha = 5\%$)
Pluto Tasche	3	4	4	4
CAP Calibration	4	4	5	5
QMM	4	4	4	5
SA ($\alpha = 2.5\%$)	2	2	0	2
SA ($\alpha = 5\%$)	2	2	0	2

It can be inferred that the proposed method passes the performance criteria, better than other competing techniques. Note that even though *CAP calibration* and *QMM* produced estimates in line with expected trend, the results does not pass the success criteria for 4–5 number of rating groups. The results of Binomial test and Likelihood Ratio test have been provided in the Appendix A.

It is not entirely surprising that the proposed method performed this way, as we had already defined through one of the constraints of the objective function, that the estimate should be within a boundary of the MLE of the default rate parameter.

3.3.2 Simulated portfolio data

The values of the PD estimates from each estimation technique, have been displayed in the Table 8. From the table, we can observe a significant variation among the different methods. A visual comparison of PD estimate trend can be observed from Figure 6.

Table 8. Comparison of PD results for simulated portfolio.

Rating	Pluto Tasche ($\alpha=75\%$)	CAP Calibration (AR = 50%)	QMM	SA ($\alpha = 5\%$)	SA ($\alpha = 2.5\%$)
1	3.17%	0.36%	5.23%	0.00%	0.00%
2	3.20%	0.40%	4.84%	0.10%	0.74%
3	3.41%	0.62%	4.49%	2.67%	2.74%
4	3.56%	1.33%	4.22%	2.68%	2.74%
5	5.06%	3.38%	4.03%	6.16%	6.09%
6	3.94%	7.37%	3.89%	6.25%	6.13%
7	3.03%	10.03%	3.68%	6.27%	6.17%
8	24.80%	10.52%	3.29%	6.28%	6.18%

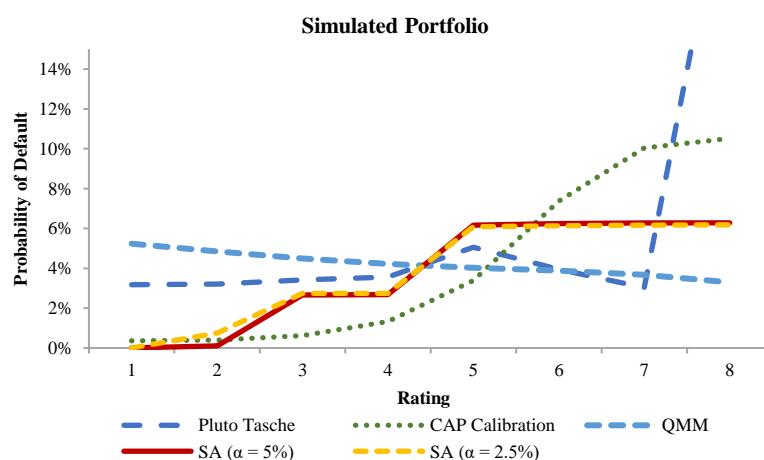


Figure 6. Comparison of results with estimates obtained from other methods for the simulated imaginary portfolio

For the simulated data with low default cases, it can be observed that both *Pluto Tasche* and *QMM* provided estimates with reversal of trend (decreasing default rates with increasing risk). In this

scenario, the proposed method using stochastic optimisation gave proper results with monotonic trend. Results for both $\alpha = 5\%$ and $\alpha = 2.5\%$ gave close estimates. *CAP calibration method* delivered expected trend, assuming an accuracy ratio of 50%. This was subjectively chosen given the low default nature of the portfolio.

The number of ratings which failed both the tests are as shown in Table 9.

Table 9. Number of Ratings which failed performance tests.

Method	Binom Test (95% CI)	Binom Test (90% CI)	Lik Ratio Test ($\alpha = 2.5\%$)	Lik Ratio Test ($\alpha = 5\%$)
Pluto Tasche	0	0	0	0
CAP Calibration	0	0	0	1
QMM	0	0	0	0
SA ($\alpha = 2.5\%$)	0	0	0	0
SA ($\alpha = 5\%$)	0	0	0	0

Comparison of methods using hypothesis tests yielded no significant breaches in any of the rating grades. This is not unexpected as the data only has small sample size and consequently leads to wider confidence intervals.

From the comparison, it is evident that the proposed method is more suitable or at par when compared with the other methods currently used in the industry. It has the added advantage of providing proper results without unintuitive trend in case of extremely low default instances.

4. Limitations

Although this approach was simple and straightforward to implement, one drawback that plagued the research was the accuracy of results. Rarely, the algorithm threw results which did not converge to the maximum likelihood. It was also observed to show some moderate sensitivity (significant increase/decrease in negative log-likelihood) near the neighbourhood of some points. It was suspected that the initial values at the beginning had a significant effect on the solution seeking capability of the algorithm.

Also, as already widely known, the performance of the algorithm strongly depends on the choice of the cooling schedule and the neighbourhood structure of the objective function. To alleviate the problem to an extent, multiple independent iterations were performed from and the results from each independent run were combined and compared to arrive at the results. Although it does not solve the problem completely, it does provide reasonably solid framework for further development to be carried out.

Values of starting temperature, the penalties of each constraint, variance of the candidate distribution etc all can have significant impact on the output. Optimal values of all these variables were identified through trial and error and kept fixed for the purpose of achieving results. There might be a different combination from the one that has been presented here, that might provide a better optimisation.

5. Conclusions

In this paper, we have discussed in length how the observed default rates are sometimes, not an accurate representation of the actual risk that is present. In case of rating systems, this can lead to issues like non-monotonous trend in the estimate of Probability of Default, where, a rating which is

supposed to represent a higher risk has a lower PD estimate than ratings with lower risk (or vice-versa). This severely affect the institution's risk management capabilities by hindering its ability to identify high risk and low risk customers. We have presented an alternate method for estimation of Probability of Default for rating systems, which solves the issue of non-monotonicity which creeps in when default rates are empirically determined. The method introduced in the document is implemented by assuming that the default rate parameter is a random variable and that the defaults follow a Binomial Distribution. A simulation algorithm is designed using the concepts of stochastic optimisation to obtain the maximum value of likelihood function which is our objective function. In this paper, the method is demonstrated for two cases. One, for a portfolio of corporate entities rated by Standard and Poor's credit rating agency and, two for an imaginary portfolio with simulated defaults.

One advantage of this method is that it is conceptually simple and straight forward and has minimum number of assumptions. The assumptions are also logical and easy to justify. The results obtained have also been observed to be comparable or better to that of other popular calibration methods, especially in situations of data scarcity or low default cases.

Although this approach is simple and straightforward, for higher precision and for large number of ratings, more simulations may be required which may be limited by the computational power of the user's computer system. Reasonably good amount of measures have been adopted to tackle problems like sensitivity to initial conditions which in turn affect the accuracy of the estimates. But presence of large number of model parameters and constraints makes the problem solving a step more difficult. These are general issues plaguing any simulation-based technique and further research needs to be conducted on how to improve the framework by using faster sampling with improved accuracy of the estimates. Using heuristic algorithms which are more problem – specific and making best use of extra information about the system, but taking advantage of the model framework as laid out in section 2.1 may lead to improved results.

Conflict of interest

I declare no conflicts of interest in this paper.

References

Basel Committee on Banking Supervision (2001) The Internal Ratings-Based Approach Consultative Document. Available from: <https://www.bis.org/publ/bcbsca05.pdf>.

Chang YP, Yu CT (2014) Bayesian confidence intervals for probability of default and asset correlation of portfolio credit risk. *Comput Stat* 29: 331–361.

Clifford T, Marianski A, Sebesteyn K (2013) Low default portfolio modelling: Probability of default calibration conundrum. In Credit Scoring and Credit Control XIII Conference.

Dwyer DW (2007) The distribution of defaults and bayesian model validation. *J Risk Model Valid* 1: 23–53.

Erlenmaier U (2011) The shadow rating approach: Experience from banking practice. *The Basel II Risk Parameters*, Springer, 37–74.

Fernandes G, Rocha CA (2011) Low default modelling: A comparison of techniques based on a real Brazilian corporate portfolio. In Credit Scoring and Credit Control XIII Conference, 24–26.

Homaifar A, Qi CX, Lai SH (1994) Constrained optimization via genetic algorithms. *Simulation* 62: 242–254.

Kiefer NM (2009) Default estimation for low-default portfolios. *J Empir Financ* 16: 164–173.

Kiefer NM (2010) Default estimation and expert information. *J Bus Econ Stat* 28: 320–328.

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by Simulated Annealing. *Science* 220: 671–680.

Krahnen JP, Weber M (2001) Generally accepted rating principles: A primer. *J Bank Financ* 25: 3–23.

Kruger M (2015) Validation and monitoring of pd models for low default portfolios using proc mcmc. In SAS Global Forum 2015.

Pluto K, Tasche D (2005) Thinking Positively. *Risk*, 72–78.

S&P (2021) Default, Transition, and Recovery: 2020 Annual Global Corporate Default And Rating Transition Study. Available from: <https://www.maalot.co.il/Publications/TS20210408160139.PDF>.

Surzhko D (2017) Bayesian Approach to PD Calibration and Stress-testing in Low Default Portfolios. *J Appl Financ Bank* 7: 1–6.

Tasche D (2002) Remarks on the monotonicity of default probabilities. Available from: <https://arxiv.org/pdf/cond-mat/0207555.pdf>.

Tasche D (2009) Estimating discriminatory power and PD curves when the number of defaults is small. Working paper, Lloyds Banking Group.

Tasche D (2013) Bayesian estimation of probabilities of default for low default portfolios. *J Risk Manage Financ Inst* 6: 302–326.

Tasche D (2013) The art of probability-of-default curve calibration. *J Credit Risk* 9: 63–103.

Van der Burgt M (2008) Calibrating low-default portfolios, using the cumulative accuracy profile. *J Risk Model Valid* 1:17–33.

Yeniay Ö (2005) Penalty function methods for constrained optimization with genetic algorithms. *Math Comput Appl* 10: 45–56.



© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)