DE GRUYTER OPEN
Ann. Univ. Paedagog. Crac. Stud. Math. 14 (2015), 37-46
DOI: 10.1515/aupcsm-2015-0003

FOLIA 160

Annales Universitatis Paedagogicae Cracoviensis
Studia Mathematica XIV (2015)

Jakub Szotek

Generalized CreditRisk+ model and applications

Abstract. In the paper we give a mathematical overview of the CreditRisk+ model as a tool used for calculating credit risk in a portfolio of debts and suggest some other applications of the same method of analysis.

In this paper we will give a condensed mathematical overview of the CreditRisk+ model and attempt to generalize it to be applicable in more situations that it was originally created for.

CreditRisk+ model was formed in 1997 by Credit Suisse First Boston bank. It is a tool used to calculate the level of credit risk for a portfolio of debts. The main purpose of the model is to determine the probability distribution for the amount of debt that will not be repaid in a given set of debts in a unit of time, usually one year. Once the probability distribution is estimated it becomes possible to calculate the expected loss, the level of value at risk, the amount of capital necessary to cover an unexpected loss. This information later allows a credit institution to compute the proper interest rate of a granted loan ensuring the maximum profit for the company.

The reasoning presented in this paper that concerns the mathematical structure of the model is based on [3]. The input data necessary to apply the CreditRisk+ model is the following:

i. The number of debts in the portfolio.

ii. The value of each debt (this is the amount that a debtor fails to repay in the event of default).

iii. The probability of default for each debtor.

The model is used as an approximation tool, so it will make sense to employ it in case of a large number of debts in the portfolio and relatively small probabilities of default.

AMS (2010) Subject Classification: 91B28, 91B30.

The analysis in the model is conducted in two steps. In the first one we calculate the probability distribution for the number of debtors failing to repay their liabilities. Since the values of debts are not constant among the debtors we will not be able yet to transmit the number of defaults into the aggregated amount of not paid off debt for the whole portfolio. This is done in the second step of the analysis through a number of approximations and calculations.

The paper is organized as follows. In the first part we present, basing on [3], the line of reasoning in the model that leads us to an applicable formula. Next we give a possible financial example where CreditRisk+ is used. Finally we suggest other areas of life where the model might be of use.

1. The number of defaulted debtors

In this section we will use standard definitions and properties from the probability theory. For further information we refer the reader to [1].

Let N be a positive integer. We will consider a family of independent random variables $\mathcal{X} = \{X_i;\ i = 1, \dots, N\}$ such that $P(X_i = 1) = p_i$ and $P(X_i = 0) = 1 - p_i$ for some $p_i \in (0, 1)$. In CreditRisk+ model the event of $X_i = 1$ signifies the bankruptcy of debtor i. Let $T = \sum_{i=1}^{N} X_i$ be the total number of defaulted debtors in the portfolio. In the first part of the analysis we will calculate an approximation of the probability distribution for T.

Let us denote by F and F_i the probability generating functions of T and X_i, respectively ($i = 1, \dots, N$). Notice that

$$F_i(z) = 1 - p_i + p_i z.$$

Moreover, due to the independence,

$$F(z) = \prod_{i=1}^{N} F_i(z).$$

Theorem 1.1
For any $z \in [0, 1]$ we have

$$\left| F(z) - \exp\left(\sum_{i=1}^{N} p_i(z-1) \right) \right| \leqslant \sum_{i=1}^{N} \frac{p_i^2}{2(1-p_i)^2}.$$

Proof. Since

$$F(z) = \prod_{i=1}^{N} F_i(z) = \prod_{i=1}^{N} (1 - p_i + p_i z),$$

we get

$$\log(F(z)) = \sum_{i=1}^{N} \log(1 + p_i(z-1)), \tag{1}$$

where log is the natural logarithm.

Let us fix some $i \in \{1, \ldots, N\}$ and expand the function $g_i \colon [0,1] \to \mathbb{R}$, $g_i(z) = \log(1 + p_i(z-1))$ into the first order Taylor polynomial centered at $a = 1$

$$g_i(z) = g_i(1) + g_i'(1)(z-1) + \frac{g_i''(\xi)}{2}(z-1)^2 = p_i(z-1) - \frac{p_i^2}{2(1+p_i(\xi-1))^2}(z-1)^2$$

for a certain $\xi \in [z, 1]$. Owing to the nature of the domain of g_i we can estimate the remainder term

$$\Big|\frac{-p_i^2}{2(1+p_i(\xi-1))^2}(z-1)^2\Big| \leqslant \frac{p_i^2}{2(1-p_i)^2}.$$

We have therefore shown that $g_i(z)$ can be approximated with $p_i(z-1)$ and the error is not greater than $\frac{p_i^2}{2(1-p_i)^2}$.

Getting back to formula (1) we may now estimate

$$\Big|\log(F(z)) - \sum_{i=1}^{N} p_i(z-1)\Big| = \Big|\sum_{i=1}^{N} \log(1+p_i(z-1)) - \sum_{i=1}^{N} p_i(z-1)\Big| \leqslant \sum_{i=1}^{N} \frac{p_i^2}{2(1-p_i)^2}.$$

The above means that $\log(F(z))$ can be approximated with $\sum_{i=1}^{N} p_i(z-1)$ and the error is not greater than $\sum_{i=1}^{N} \frac{p_i^2}{2(1-p_i)^2}$. Since $\log(F(z)) \leqslant 0$, $\sum_{i=1}^{N} p_i(z-1) \leqslant 0$, and $e^t \leqslant 1$ for all $t \in (-\infty, 0]$, the Lagrange mean value theorem yields

$$\begin{aligned}\Big|F(z) - \exp\Big(\sum_{i=1}^{N} p_i(z-1)\Big)\Big| &= \Big|e^{\log(F(z))} - \exp\Big(\sum_{i=1}^{N} p_i(z-1)\Big)\Big| \leqslant \\ &\leqslant \Big|\log(F(z)) - \sum_{i=1}^{N} p_i(z-1)\Big| \\ &\leqslant \sum_{i=1}^{N} \frac{p_i^2}{2(1-p_i)^2}.\end{aligned}$$

The proof is complete.

Let $\mu = \sum_{i=1}^{N} p_i$. Theorem 1.1 implies that

$$F(z) \approx e^{\mu(z-1)} = e^{-\mu} e^{\mu z} = \sum_{n=0}^{\infty} \frac{e^{-\mu}\mu^n}{n!} z^n$$

whenever the probabilities p_i are small. This means that the number of defaulted debtors might be approximated with the Poisson distribution with the above parameter μ. For more detailed information about using the Poisson distribution as an approximation to the binomial distribution with unequal probabilities we refer to [2].

Theorem 1.1 is an essential improvement of the analysis presented in [3]. It allows us to estimate the discrepancy between the actual probability distribution and its approximation.

2. Setup for the main theorem

2.1. Partition of an interval containing all possible individual losses

Let us now consider a family of independent random variables $\mathcal{Y} = \{Y_i;\ i = 1, \dots, N\}$ such that $P(Y_i = y_i) = p_i$ and $P(Y_i = 0) = 1 - p_i$ for some $y_i \in \mathbb{R}_+$ (we denote by $\mathbb{R}_+$ the set of positive real numbers). In the CreditRisk+ model, Y_i represents the loss from the ith debt. The debtor's probability of default is still equal to p_i and the event of default causes the loss of y_i monetary units. Our aim is to determine the probability distribution of $R = \sum_{i=1}^{N} Y_i$, i.e. the aggregated loss for the whole portfolio of credits.

Let M be an arbitrarily chosen positive number such that $y_i \in [0, M]$ for $i = 1, \dots, N$. We also choose the following:

(i) $L \in \mathbb{R}_+$, $m \in \mathbb{N} \setminus \{0\}$,

(ii) $\xi_0, \dots, \xi_m \in \mathbb{R}$, $\nu_1, \dots, \nu_m \in \mathbb{N} \setminus \{0\}$ such that

  (a) $0 = \xi_0 < \ldots < \xi_m = M$,

  (b) $\xi_{j-1} < \nu_j L \leqslant \xi_j$, $j = 1, \dots, m$.

In other words we have picked a real number L and a partition of $[0, M]$ that allows us to approximate all values within any interval $(\xi_{j-1}, \xi_j]$ with a multiple of L.

In the sequel we will assume that if $y_i \in (\xi_{j-1}, \xi_j]$, then y_i can be approximated with $\nu_j L$. Let us therefore introduce a family of independent random variables

$$\overline{\mathcal{Y}} = \{\overline{Y_i};\ i = 1, \dots, N\}$$

such that $P(\overline{Y_i} = \nu_j) = p_i$ and $P(\overline{Y_i} = 0) = 1 - p_i$ whenever $y_i \in (\xi_{j-1}, \xi_j]$.

2.2. Calculation within $(\xi_{j-1}, \xi_j]$

We take into consideration the interval $(\xi_{j-1}, \xi_j]$ for some $j \in \{1, \dots, m\}$ and define

$$\overline{\mathcal{Y}}_j = \{\overline{Y_i};\ y_i \in (\xi_{j-1}, \xi_j]\}.$$

We have therefore identified those debts within the portfolio for which the amount of potential loss falls into $(\xi_{j-1}, \xi_j]$.

We can now perform the same analysis as in Section 1 and obtain the Poisson distribution as an approximation of the number of defaults, say Z_j, within the chosen subset of credits. The parameter is equal to

$$\mu_j = \sum_{\overline{Y_i} \in \overline{\mathcal{Y}}_j} p_i.$$

Since we assume a constant amount of loss for each debt within the subset, we can easily translate the number of defaults into the aggregated loss suffered from this subset. Namely, if we denote the aggregated loss by R_j, we have $R_j = \nu_j L Z_j$.

Next we apply the same reasoning to every other interval $(\xi_{j-1}, \xi_j]$. Thanks to the above simplification we reduce the number of random variables that sum up to the total loss from N to m.

3. Probability distribution of the total loss

Now, our goal is to find the probability distribution for the random variable $R = \sum_{i=1}^{N} Y_i$. It will be approximated with the probability distribution of

$$\overline{R} := \sum_{i=1}^{N} L\overline{Y_i} = \sum_{j=1}^{m} R_j.$$

Since we have assumed that random variables $\overline{Y_1}, \dots, \overline{Y_N}$ are independent, we can conclude that so are $R_1, \dots, R_m$.

Let us consider the probability generating function G of the random variable $\frac{1}{L}\overline{R}$. Recall that $G(z) = \sum_{n=0}^{\infty} q_n z^n$ with $q_n = P(\overline{R} = nL)$. In other words q_n is the probability of the event that the aggregated loss from the entire portfolio is equal to n units L.

We are in a position to state and prove the main result of the paper.

Theorem 3.1
Under the above assumptions and notations the following recursion formula holds true

$$\begin{cases} q_0 = e^{-\mu}, \\ q_n = \sum\limits_{\nu_j \leqslant n} \dfrac{\mu_j \nu_j}{n} q_{n-\nu_j} \qquad \textit{for } n \geqslant 1, \end{cases} \tag{2}$$

where $\mu = \sum_{j=1}^{m} \mu_j$.

Proof. As we concluded in Section 2.2, we have $R_j = \nu_j L Z_j$ and $Z_j \sim \mathrm{Pois}(\mu_j)$, where $j = 1, \dots, m$. The probability generating function G_j of the random variable $\frac{1}{L} R_j$ has the following form

$$G_j(z) = \sum_{n=0}^{\infty} z^{n\nu_j} P(Z_j = n) = \sum_{n=0}^{\infty} \frac{e^{-\mu_j}\mu_j^n}{n!} z^{n\nu_j} = e^{-\mu_j} \sum_{n=0}^{\infty} \frac{(\mu_j z^{\nu_j})^n}{n!} = e^{-\mu_j + \mu_j z^{\nu_j}}.$$

The independence of $R_1, \dots, R_m$ yields

$$G(z) = \prod_{j=1}^{m} G_j(z) = \prod_{j=1}^{m} e^{-\mu_j + \mu_j z^{\nu_j}} = \exp\left(-\mu + \sum_{j=1}^{m} \mu_j z^{\nu_j} \right).$$

Defining

$$f(z) = \frac{1}{\mu} \sum_{j=1}^{m} \mu_j z^{\nu_j}$$

leads us to the following simplification

$$G(z) = \exp\left(-\mu + \mu \frac{1}{\mu} \sum_{j=1}^{m} \mu_j z^{\nu_j} \right) = e^{\mu(f(z)-1)}.$$

Consequently,

$$q_0 = G(0) = e^{\mu(f(0)-1)} = e^{-\mu}.$$

Now, pick any $n \geqslant 1$. It is obvious that

$$q_n = \frac{1}{n!}\frac{d^n G(z)}{dz^n}\bigg|_{z=0}.$$

Let us calculate the nth derivative

$$\begin{aligned}\frac{d^n G(z)}{dz^n} &= \frac{d^{n-1}}{dz^{n-1}}\frac{dG(z)}{dz} = \frac{d^{n-1}}{dz^{n-1}}\frac{d\left(e^{-\mu+\mu f(z)}\right)}{dz} \\ &= \frac{d^{n-1}}{dz^{n-1}}\left(\mu e^{-\mu+\mu f(z)}\frac{df(z)}{dz}\right) = \frac{d^{n-1}}{dz^{n-1}}\left(\mu G(z)\frac{1}{\mu}\frac{d}{dz}\sum_{j=1}^{m}\mu_j z^{\nu_j}\right) \\ &= \frac{d^{n-1}}{dz^{n-1}}\left(G(z)\frac{d}{dz}\sum_{j=1}^{m}\mu_j z^{\nu_j}\right).\end{aligned}$$

Therefore, using the Leibniz rule, we obtain

$$q_n = \frac{1}{n!}\frac{d^n G(z)}{dz^n}\bigg|_{z=0} = \frac{1}{n!}\sum_{k=0}^{n-1}\binom{n-1}{k}\frac{d^{n-k-1}G(z)}{dz^{n-k-1}}\frac{d^{k+1}}{dz^{k+1}}\sum_{j=1}^{m}\mu_j z^{\nu_j}\bigg|_{z=0}.$$

Observe that

$$\frac{d^{k+1}}{dz^{k+1}}\sum_{j=1}^{m}\mu_j z^{\nu_j}\bigg|_{z=0} = \begin{cases}(k+1)!\mu_j, & \text{if } k+1=\nu_j \text{ for some } j,\\ 0, & \text{otherwise.}\end{cases}$$

We also know that

$$\frac{d^{n-k-1}G(z)}{dz^{n-k-1}}\bigg|_{z=0} = (n-k-1)!q_{n-k-1}$$

for every non-negative integer $k \leqslant n-1$. Finally,

$$\begin{aligned}q_n &= \frac{1}{n!}\sum\binom{n-1}{k}(n-k-1)!q_{n-k-1}(k+1)!\mu_j \\ &= \sum\frac{1}{n!}\frac{(n-1)!}{k!(n-k-1)!}(n-k-1)!(k+1)!q_{n-k-1}\mu_j,\end{aligned}$$

where the sum ranges over all non-negative $k \leqslant n-1$ such that $k+1=\nu_j$ for some $j \in \{1,\dots,m\}$. Hence

$$q_n = \sum_{\nu_j \leqslant n}\frac{\mu_j\nu_j}{n}q_{n-\nu_j}.$$

The above proof is a more detailed version of the proof presented in [3]. Let us point out that (2) is an approximate formula. It may be used when the probabilities p_i are small.

4. An example of application of CreditRisk+

Let us present a possible real-life situation of using CreditRisk+. Suppose we have a particular sample portfolio of 1000 debts. Due to the significant amount of data let us just describe the main characteristics. Each debt's amount ranges between 0 and 110 000 monetary units and each individual probability of default lies between 0 and $0,25$. In the first stage of the analysis we approximate the probability distribution of the number of defaulted debtors with the Poisson distribution whose parameter μ is equal to the sum of individual probabilities. In our set of data we have calculated that $\mu = 111,4$. The graph of the probability distribution is presented in Fig. 1.

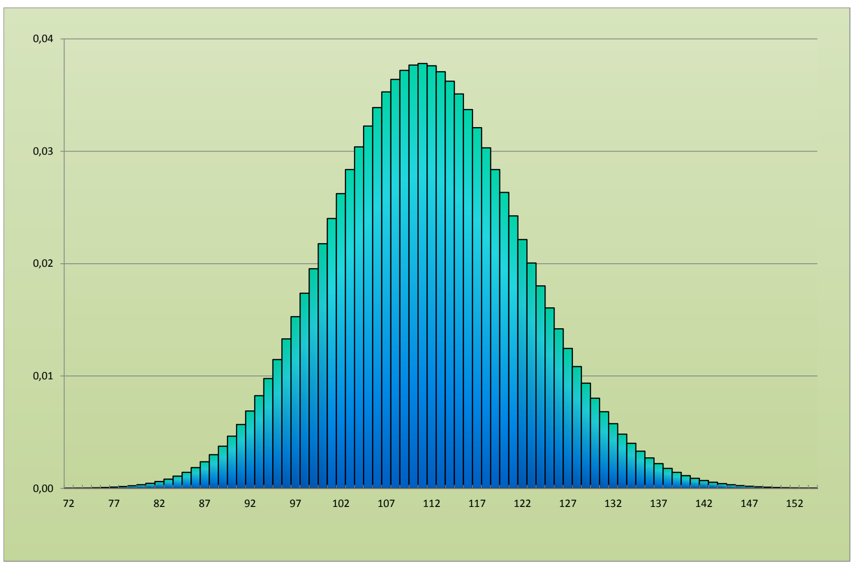

Figure 1: Probability distribution for the number of defaults

Next we will introduce a partition of all possible losses according to Section 2. In the first step we split the whole portfolio into subsets of debts falling into the same interval of exposure. For the sake of simplicity let the length of every interval be equal to 5000. Thus the intervals would be $(0, 5000]$, $(5000, 10000]$ and so on. Within each interval we calculate the sum of the probabilities of default. In the subsequent step we fix the exposure for all debtors in the same interval. For instance we can use the middle value, obtaining 2500, 7500 and so on. We also choose a monetary unit L that we will use in further analysis. A suitable value would be $L = 2500$ ensuring that all the exposures could be expressed as integer multiples of L.

Finally we apply formula (2) to obtain the approximate probability distribution of the sum of losses generated by all the debtors who cannot repay their liabilities. The distribution is presented in Fig. 2. For the purpose of transparency the values have been grouped as one can see on the horizontal axis.

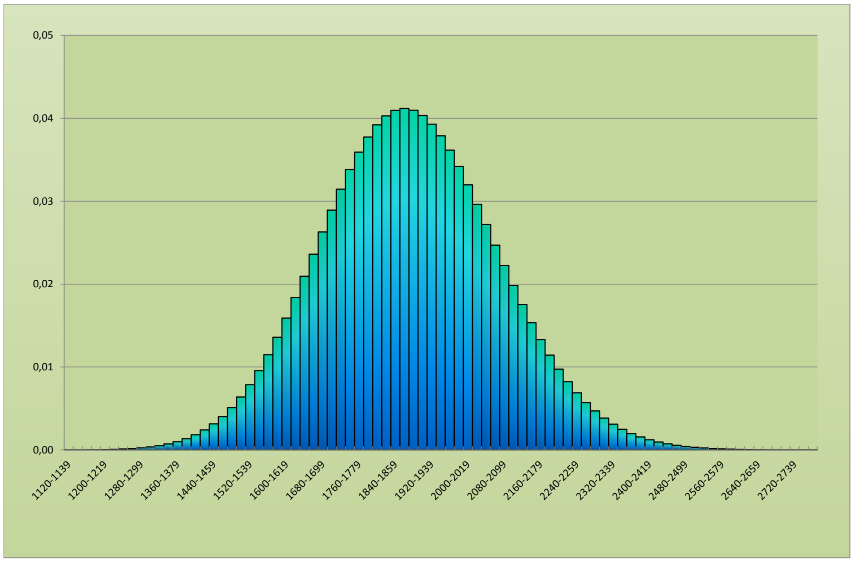

Figure 2: Probability distribution for the aggregated loss

When the approximate probability distribution for the aggregated loss is calculated we can extract all the data necessary from the financial point of view. Some of them could be the expected loss, standard deviation and the percentile of a specified level (e.g., 95%) indicating the so-called Value at Risk. One can observe the results of our calculations in Tab. 1.

$E(\overline{R})$	4 702 325 $(1880, 93\,L)$
$D(\overline{R})$	484 575 $(193, 83\,L)$
VaR 95%	5 515 000 $(2206\,L)$

Table 1: Parameters

5. Other applications of CreditRisk+ model

Now we will describe other possible applications of the model. Our task will consist in defining the input data in terms of random variables Y_i. Next it would be possible to apply formula (2) to approximate the probability distribution of their aggregate.

5.1. Insurance

As the first example let us consider an insurance company. Suppose the company offers vehicle insurance policies and attempts to estimate the aggregated amount of compensation paid to its customers within a period of one year. This way the company can calculate the expected level of indemnities paid or the amount of capital to be held as a reserve to cover an unexpected value of compensation. It is also possible to compute the price of a policy and the expected profit from this part of business activity.

Let us assume that the company has a portfolio of N policies. Each of them is sold to a separate driver who owns a car of a certain model. Judging by the driving history, age, health condition and possibly other criteria the company assigns to each driver the probability of causing a car accident within a one-year period. Let us denote this probability as p_i $(i = 1, \ldots, N)$. Next the company can calculate the historical average cost of damage that a car of a certain model suffers in an accident. This way we obtain the values y_i, thus completing the form of random variables Y_i.

5.2. Sales revenue

Let us suppose there is a company that hires sales representatives to sell products. Each one of the representatives has his own sales effectiveness, which is a fraction of successful offer presentations. The company has a base of potential clients to whom the representatives will offer one of the company's products. Let us denote by N the number of offers the company is hoping to present within a specified period of time. The company assigns representatives to potential clients and selects the product each client is most likely to buy. Each successful sale brings the company an expected revenue depending on the product. Let Y_i $(i = 1, \ldots, N)$ be the random variable of the sales revenue from one product offer. For each offer we know which representative is trying to make a sale, thus we know the probability p_i of success and we also know the potential revenue y_i being the result of a successful sale. This way we have described the sales process in mathematical terms of random variables Y_i. Now we are able to approximate the probability distribution of the aggregated sales revenue.

5.3. Courier company

Suppose we have a company that delivers packages. They are indexed with $i \in \{1, \ldots, N\}$. Each of them has its own value y_i to be paid by the company to a client in the event of delivering the defected package. The probability of being defected is the same for every package (and equal to, say, p). In this case although the probability is constant the values y_i are not, so we cannot use any well known distribution. However, we still can apply formula (2) to approximate the probability distribution of aggregated amount of compensation paid to the customers within a specified period of time, for instance a year.

5.4. The cost of flu

During the year there are some periods of high incidence of flu or a cold. Many of us have our own methods of fighting the illness, so in many cases it is possible to estimate the expenses caused by sickness. For a person under investigation we could denote the expenses as y_i $(i = 1, \ldots, N)$. Next either could we assume a constant probability p of catching a flu, or attempt to appraise personal probabilities p_i. Such information could be obtained in a survey of some sort. Finally we could try to approximate the probability distribution for the aggregated cost of flu in the examined population.

5.5. Harvest

Following the lead from the previous example we could consider an agricultural territory. Each farmer has his own field where he cultivates some kind of crop. Suppose the crops are susceptible to a specific type of disease. Each field, depending on the crop farmed, soil condition, fertilization etc., is characterized by its own probability p_i of the disease occurrence. If the disease develops the harvest is on average diminished by a certain amount y_i (expressed either in natural or financial units). The amount depends on the area of the field or the type of crop. Given these data we can apply formula (2) to obtain the approximate probability distribution for the aggregated loss caused by the disease in the whole territory.

References

[1] W. Feller, *An introduction to probability theory and its applications. Vol. I.*, Third edition John Wiley & Sons, Inc., New York-London-Sydney 1968. Cited on 38.

[2] J.L. Hodges, Jr., L. Le Cam, *The Poisson approximation to the Poisson binomial distribution*, Ann. Math. Statist. **31** (1960), 737–740. Cited on 39.

[3] *CreditRisk+. A Credit Risk Management Framework*, Credit Suisse First Boston, 1997. Cited on 37, 38, 39 and 42.

State Street Bank GmbH
ul. Podgórska 36
Warszawska 24
31-536 Kraków
Poland
E-mail: jakubszotek@gmail.com

Received: May 22, 2014; final version: January 7, 2015;
available online: , 2015.