APPLICATION OF EXPECTED CREDIT LOSS MODEL AND MARKOV CHAIN TO CALCULATE NET SINGLE PREMIUM OF UNSECURED CREDIT INSURANCE

Hansen Juni Lieus\(^1\)*, Devin Tedja\(^2\), Vanessa Joewita\(^3\), Agus Sofian Eka Hidayat\(^4\), Alexander R. J. Silalahi\(^5\)

\(^{1,2,3,5}\)Department of Business Mathematics, Faculty of STEM, Prasetiya Mulya University  
St. BSD Raya Utama, Kavling Edutown II, Tangerang, 15339, Indonesia  
\(^4\)Department Actuarial Science, Faculty Business, President University, Jababeka City, 17530, Indonesia

Corresponding author’s e-mail: *hansenjunilieus@gmail.com

ABSTRACT

Transferring credit risk to an insurance company is a way to mitigate risk. Premiums should be calculated accurately to attain economic value for both the lender and the guarantor. The aim of this study was to determine the net single premium (NSP) values for an unsecured credit insurance product using the expected credit loss (ECL) method from IFRS 9. This study used data generated through simulation of insurance policies issued in 2015 or 2016. Their state classifications were monthly observed from 2016 to 2020. The probability of disbursed claim (PDC) parameter replaced the probability of default parameter on the ECL model, whereas the PDC model was constructed based on the components of a state-transition probability matrix, obtained with the Markov chain approach using the cohort method: \( p_{11} = 0.999181, \) \( p_{12} = 0.000130, \) and \( p_{13} = 0.000689. \) The PDC model validation showed relatively decent results, whereas MSE = 2.457% and \( z = 0.608 \) with \( \alpha = 5\% \). These results indicated that the PDC model was a good fit to calculate ECL. 5,000 iterations were done as part of the cash flow simulation process, whereas debtors’ loan amounts were randomly generated during each iteration, and the average NPV of these iterations was Rp564,419,305. Based on model sensitivity analysis, cash flow values were most sensitive to the variable used to construct the PDC model \( (p_{13}). \) Thus, the 5,000-iteration process was repeated with the newly adjusted PDC value, which were \( p_{11} = 0.998924 \) and \( p_{13} = 0.000946. \) The new average NPV of these iterations was Rp409,877,840, indicating that the constructed ECL model was a good fit to calculate NSP values for unsecured credit insurance products.

Received: 3rd July 2023  
Revised: 3rd October 2023  
Accepted: 2nd November 2023

Keywords:  
Cohort;  
Expected Credit Loss; Markov Chain;  
Probability of Disbursed Claim;  
Unsecured Credit Insurance.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite this article:


Copyright © 2023 Author(s)  
Journal homepage: https://ojs3.unpatti.ac.id/index.php/barekeng/  
Journal e-mail: barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id  
Research Article · Open Access
1. **INTRODUCTION**

The extension of credit from savers to borrowers is essential for the growth of the economy [1], [2]. However, there is a risk inherent in extending credit, which is the risk of debtors failing to meet or fulfill their obligation [3]. As a result, this calls for means of risk mitigation, in which insurance is one example.

For an insurer to stay in business, it needs to set a premium that is adequate to cover actual claims as they arise in the future [4]. Moreover, credit insurance is one of the largest classifications of insurance; in 2021, credit insurance is the third largest contributor of premiums and the claim amounts are the largest in the general insurance industry [5]. Hence from a credit insurer’s perspective, determining the amount of premium charged is of utmost importance. Consequently, this study aims to develop a model or framework for unsecured credit insurance premium setting that is based on two concepts: (1) actuarial science, which includes but are not limited to probability, statistics, and financial mathematics; and (2) the expected credit loss (ECL) model as introduced in International Financial Reporting Standards 9 (IFRS 9).

IFRS 9 is published by the International Accounting Standards Board (IASB) as a replacement for International Accounting Standard 39 (IAS 39). This new standard introduced a new approach for measuring loans and receivables, which is the ECL model [6], [7]. The ECL model consists of three quantitative parameters, namely the probability of default (PD), the loss given default (LGD), and the exposure at default (EAD). As these parameters are closely related to the concepts of claim frequency and claim severity in insurance [8], this forms a basis for the objective of this study, which is to develop a premium-setting framework or model for unsecured credit insurance based on the ECL model.

2. **RESEARCH METHODS**

2.1 Research Data and Variables

This research utilizes 10,000 simulated insurance policy data for unsecured credit loans, generated by the researchers using Python. The data consists of the following variables: Policy Start Date, Policy End Date, Loan Amount, Loan Duration, Final State of Debtor, and Default Date.

The Policy Start Date and Loan Duration are randomly generated using a uniform distribution within the specified ranges. The Policy Start Date ranges from 1 January 2015 to 1 December 2016, whereas the Loan Duration ranges from 1 to 5 years. The Loan Amount is generated based on a beta distribution with parameters $\alpha = 2.21749$ and $\beta = 2.31202 \times 10^9$. The beta distribution is chosen to reflect a right-skewed data because typically and naturally, unsecured credit loans are of low ticket sizes as they are assumed to be for consumption purposes. The parameters are determined randomly to take into account the randomness and variations in loan amounts. The state of the debtor is generated based on the probabilities described in Table 1, which determine the likelihood of different states occurring.

<table>
<thead>
<tr>
<th>Loan Duration (In Years)</th>
<th>State</th>
<th>Probability Assumption</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0.9950</td>
<td>2,029</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0045</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0005</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.9815</td>
<td>1,905</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0050</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0135</td>
<td>43</td>
</tr>
</tbody>
</table>
The interest rate used in this research is the average annual guarantee interest rate provided by the Deposit Insurance Agency (LPS) from January 2015 to September 2020, obtained from the lps.go.id website. The average value of the guarantee interest rate obtained is 6.63%, with a 95% confidence interval based on the $t$-distribution of [6.35%, 6.91%]. The researchers then converted this value into a monthly interest rate, resulting in 0.5367% with a 95% confidence interval of [0.5145%, 0.5587%].

2.2 Data Processing

The data processing for this research began with the use of Microsoft Excel to generate borrowers’ profile data. The generated data includes the default status of each borrower, based on the characteristics of an unnamed credit insurance distribution company in Indonesia (referred to as "PT X" in this study).

Next, the process of forming the transition probability matrix was conducted using Microsoft Excel. The classification of policies into states was performed on a monthly basis for a period of 5 years, from 2016 to 2020. The states defined in this study are as follows: state 1 (policies without claim submissions), state 2 (policies with claim submissions, but the claims are rejected or denied, resulting in no claim settlement) and state 3 (policies with claim submissions and successful claim settlements).

Furthermore, the validation of the probability of disbursed claim (PDC) model was conducted using Microsoft Excel by comparing the simulated borrower profile data with the previously established PDC model. Lastly, the researchers utilized Python for calculating net single premiums, claims, data visualizations, simulating loan value data generation, and performing cash flow simulation iterations.

2.3 Calculation of Premiums using ECL

IFRS 9 introduces the expected credit loss (ECL) model for banking, which calculates the expected credit loss on loans and receivables. Ludwig [6] introduced the ECL model, which incorporates parameters such as probability of default, loss given default, and exposure at default. According to IFRS 9 standards, the time value of money must be explicitly accounted for [6]. Therefore, the ECL model calculates the present value by discounting future cash flows. Ludwig [6] states that it is common in literature and practice to describe the ECL equation with a discrete time range. By consolidating all the mentioned parameters, the ECL equation for borrower $i$ with a loan duration of $T$ can be formulated as follows [6], [8]:

$$ECL_i = \sum_{t=1}^{T_i} \frac{PD_{it} \times LGD_{it} \times EAD_{it}}{(1 + r_t)^t}$$

(1)

It is important to note the difference in viewpoints inherent to banking and insurance. While banks primarily emphasize the probability of loan delinquency, insurance companies prioritize the likelihood of
successfully-disbursed claim settlements and hence, not all defaults are deemed losses by insurance firms. Consequently, PDC is substituted for probability of default (PD) in the ECL model to represent the probability of both default and claim approval. Probabilities in PDC represents a refined approach from an insurer's standpoint. With reference to the states described in Section 2.2, the formula for the marginal probability of a disbursed claim is as follows:

\[
PDC_t = Pr(X_t = 3, X_{t-1} = 1, \ldots, X_1 = 1 | X_0 = 1) = p_{13} p_{11}^{t-1}
\]  

(2)

There is also a cumulative probability of disbursed claims, which represents the likelihood of a borrower \((i)\) experiencing default and claim settlement from time 0 to a specific time, where the time is determined by the borrower's loan tenure \((t)\).

\[
\pi_i = \sum_{s=1}^{t} PDC(s - 1, s)
\]  

(3)

Researchers determine the probability of claims being paid out using a time-homogeneous Markov chain approach and the cohort method. The steps involve processing debtor data by grouping them into states and constructing a discrete transition probability matrix. The probabilities are obtained using the cohort method, specifically through the following equation [6]:

\[
\hat{p}_{ij} = \frac{\sum_{k=0}^{N-1} n_{ij}(\Delta t_k)}{\sum_{k=0}^{N-1} n_i(t_k)}
\]  

(4)

In the context of insurance, ECL can be seen as the equivalent of insurance premiums received by borrowers as policyholders. By applying the principle of equivalence, the present value (PV) of premiums represents the net single premium of credit insurance [9]. The calculation of the net single premium for an individual borrower in an unsecured loan insurance product incorporates the adjustment of the PD parameter from the ECL equation. The equation for determining the net single premium for borrower \(i\) is as follows:

\[
NSP_i = ECL_i = \sum_{t=1}^{T} \frac{PDC_{i,t} \times LGD_{i,t} \times EAD_{i,t}}{(1 + r_i)^T}
\]  

(5)

Loss given default (LGD) is calculated as the difference between 1 and recovery rate (RR). RR represents the percentage of defaulted loan amount successfully recovered or repaid by the insured, out of the total defaulted loan amount [6], [10]. Thus, LGD represents the proportion of the borrower's outstanding debt that the insurance company, as the insurer, is obliged to bear upon default, considering the percentage of credit recovery. RR is considered to be 0 due to the absence of collateral or guarantees associated with the researched insurance product, which is unsecured loans. As a result, the LGD value is determined to be 1, indicating a complete loss in the event of default.

Exposure at default (EAD) is the loss incurred by the insurance company when a borrower experiences loan default, and it is calculated based on a proportion of the borrower’s outstanding loan balance (OLB) at the time of default [6], [11].

2.4 Model Validation using z-Spiegelhalter Test

The z-Spiegelhalter test is based on using mean squared error (MSE) as its foundation. MSE is the squared difference between the predicted estimation and the observed actual event, averaged over the total number of operations [12]. MSE is also known as the Brier score. The equation for MSE is as follows:
MSE serves as an indicator of model performance, with a lower MSE value indicating superior performance [12]. This is attributed to the smaller average squared error, which signifies closer resemblance between the predicted values and the actual values.

The $z$-Spiegelhalter test employs a hypothesis test, specifically testing the null hypothesis that the observed default rate is equal to the predicted PD ($\pi_i = \hat{\pi}_i$). In this study, the default rate is modified to include both default and disbursed claim rate, with PDC substituted for PD. Spiegelhalter developed a statistical test to assess the significance of the difference between MSE and the expected MSE value. When the predicted default rate matches the observed default rate, the expectations and variances of MSE can be calculated [12].

$$E(MSE_{\pi_i=\hat{\pi}_i}) = \frac{1}{N} \sum_{i=1}^{N} \pi_i (1 - \pi_i)$$

$$Var(MSE_{\pi_i=\hat{\pi}_i}) = \frac{1}{N^2} \sum_{i=1}^{N} (1 - 2\pi_i)^2 \cdot \pi_i \cdot (1 - \pi_i)$$

Under the assumption of independence, this test is based on the central limit theorem. Therefore, the standardized distribution of MSE is as follows [12]:

$$z_s = \frac{MSE - E(MSE_{\pi_i=\hat{\pi}_i})}{\sqrt{Var(MSE_{\pi_i=\hat{\pi}_i})}}$$

The statistical test involves comparing the observed test statistic with the critical values of $z_{\alpha/2}$ and $z_{1-\alpha/2}$ [13]. If the test statistic falls within this range, the null hypothesis ($H_0$) is not rejected. However, if the test statistic exceeds or equals the critical values, the research proceeds by rejecting $H_0$.

### 2.5 Cash Flow Simulation

In this phase, discounted cash flow simulations were conducted using the ECL model to calculate the net present value (NPV) of unsecured loan insurance. The revenue of the insurance company was generated from unsecured insurance premiums, while expenditure calculations for credit insurance involved subtracting the remaining principal debt from subrogated assets. However, as the focus was on unsecured insurance without asset recovery, the insurance company's expenditure only considered the remaining principal debt. Claim calculations for defaulted debtors are determined based on a proportion of their OLB (covered by the insurance company) at the time of default. The formula for calculating NPV is as follows [14], [15]:

$$NPV = C_0 + \sum_{t=1}^{T} \frac{C_t}{(1 + r)^t}$$
3. RESULTS AND DISCUSSION

3.1 Calculating the Predicted Probability of Disbursed Claim Value

The research process was initiated by recording the state condition of each debtor every month, continued by observing the monthly movement of shifting of states 1 (not default), 2 (default but not disbursed), and 3 (default and disbursed), from the beginning of January 2016 until the beginning of January 2021. Then, using the cohort method, the transition probability matrix \( P \) obtained is as follows:

\[
P = \begin{bmatrix}
0.999181030378059 & 0.000130146335269657 & 0.0006888232866711 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

Figure 1. State Transition Diagram Illustrating The Different Stages In The Borrower's Journey.

Based on matrix \( P \), \( \hat{p}_{22} \) and \( \hat{p}_{33} \)’s values are equal to 1, indicating that states 2 and 3 are absorbing states. This result aligns with what normally happens in the world of credit insurance, whereas it is impossible for a debtor to still be classified as an active debtor if he or she has defaulted on his or her loan.

Subsequently, the \( \hat{p}_{13} \) and \( \hat{p}_{11} \) values were used to compute the marginal probability of disbursed claim (\( PDC_i \)) and create the predicted cumulative probability of disbursed claim (\( \hat{\pi}_i \)) model for each debtor \( i \), using Equation (2) and Equation (3).

3.2 Model Validation

\( \hat{\pi}_i \) values for all observed debtors were then exerted in the mean squared error (MSE) formula, resulting in an MSE value of 2.457%. This result was essential to conduct the \( z \)-Spiegelhalter test with \( \alpha = 5\% \), by which the details of the hypothesis test were as follows:

- \( H_0 \): The observed rate of disbursed claim is equal to the predicted probability of disbursed claim (\( \pi_i = \hat{\pi}_i \))
- \( H_1 \): The observed rate of disbursed claim is not equal to the predicted probability of disbursed claim (\( \pi_i \neq \hat{\pi}_i \)).

Obtaining \( E(MSE) \) value of 2.367% and \( Var(MSE) \) value of 0.0002159%, the calculated \( z \)s based on those two results has a value of 0.60837552. As \( |z| < z_{1-\alpha/2} \) (1.96), there was a failure to reject the null hypothesis, meaning that there was not sufficient evidence to support the claim that there is a difference between the observed rate of disbursed claim and the predicted probability of disbursed claim. As a result, it was concluded that the PDC Markov chain model constructed using the cohort method is eligible to be used for the next research stage (\( p_{11} = \hat{p}_{11} = 0.999181030378059, \ p_{13} = \hat{p}_{13} = 0.0006888232866711 \)).

3.3 Expected Credit Loss Model Construction

As mentioned in Section 2.3, the present value of premiums indicates the net single premium (NSP) of credit insurance, whilst the calculation of NSP for an individual debtor in an unsecured loan insurance plan takes into account the adjustment of the probability of default (PD) parameter based on the ECL
equation. Therefore, using Equation (5), the NSP value for each distinct debtor \((i)\) could be calculated using the ECL model as follows:

\[
NSP_i = ECL_i = \sum_{t=1}^{T} \left( \frac{p_{13} p_{11}^{t-1}}{1 + 0.005367} \times 1 \times EAD_{i,t} \right)
\]

where \(T = \) loan tenure (in months), \(LGD_{i,t} = 1 - RR = 1\) (whereas \(RR = 0\), as mentioned in Section 2.3), and monthly interest rate \(\approx 0.005367\), as mentioned in Section 2.1. This ECL model was then used to generate all NSP values for all observed debtors.

3.4 Cash Flow Simulation

All insurance policies observed in this research were issued in either 2015 or 2016 and had a maximum loan duration of 5 years. Since there is a possibility of policies being issued in 2016 and defaulting in 2021 (which falls outside the scope of this research observation), the researchers only considered policies issued in 2015 so that the cash flow simulation could better portray the profitability through the ECL model implementation from the perspective of an insurance company.

After calculating the net single premiums and claim amounts (in case of debtor default) for all policies, these values were discounted to January 1, 2015. The resulting simulation values represented the net present value (NPV), as shown in Table 2.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Policies</th>
<th>Number of Defaults</th>
<th>Total Premium (Rp)</th>
<th>Total Claims (Rp)</th>
<th>Cash Balance (Rp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>4,959</td>
<td>33</td>
<td>2,701,126,628</td>
<td>1,067,436,619</td>
<td>1,633,690,009</td>
</tr>
<tr>
<td>2016</td>
<td>4,926</td>
<td>46</td>
<td>0</td>
<td>1,070,752,956</td>
<td>562,937,053</td>
</tr>
<tr>
<td>2017</td>
<td>4,880</td>
<td>33</td>
<td>0</td>
<td>467,487,320</td>
<td>95,499,733</td>
</tr>
<tr>
<td>2018</td>
<td>4,857</td>
<td>23</td>
<td>0</td>
<td>268,522,194</td>
<td>-173,072,461</td>
</tr>
<tr>
<td>2019</td>
<td>4,824</td>
<td>13</td>
<td>0</td>
<td>61,562,517</td>
<td>-234,634,977</td>
</tr>
<tr>
<td>2020</td>
<td>4,811</td>
<td>2</td>
<td>0</td>
<td>12,983,935</td>
<td>-247,618,912</td>
</tr>
</tbody>
</table>

The table above shows that the insurance company has an NPV of cash flow of -Rp247,618,912 at the end of 2020. The results of the cash flow simulation were highly dependent on the loan amounts of each debtor as the loan values played a significant role in determining the magnitude of premiums and claims (in the event of debtor default). However, the loan amounts for each debtor were generated randomly based on a distribution, hence drawing conclusions from the above results would not be a wise thing to do. Hence, iterative cash flow simulations were conducted where the loan values of all debtors were randomly generated in each iteration, thereby providing an overview or range of various possible simulation outcomes. The researchers continued the iteration process until the 5,000th iteration, where the results of the 5,000th iteration showed that the average and standard deviation of NPV of cash flow exhibited a converging pattern.

3.5 Model Sensitivity Analysis

4,992 out of 5,000 iterations resulted in negative final cash flows, with an average of -Rp564,419,305 and a standard deviation of Rp207,660,566. These negative results indicate that the constructed ECL model has not yet produced an NPV value of zero, even though it is desirable to obtain an NPV value of zero as it ideally indicates that the insurance company neither incurs losses nor gains profits. Despite this, the researchers believed that insurance companies could still utilize the ECL model using the obtained values of \(p_{11}\) and \(p_{13}\) as the base rate for determining net single premiums. Nonetheless, adjustments and modifications based on each individual debtor’s risk profile are required for more accurate premium calculations.
The last section of this research process is to conduct model sensitivity analysis to assess the impact of \( p_{13} \) and \( r \) towards the NPV of cash flows of an unsecured credit insurance company. The values of \( p_{13} \) and \( r \) used were obtained from the 95% confidence interval of each variable, where the lowest value of \( p_{13} \) and \( r \) represents the lower bound of the confidence interval, and the highest value is taken from the upper bound of the confidence interval. The researchers used the results in Table 2 as the base case for this analysis, and Table 3 below represents the outcomes of sensitivity analysis.

### Table 3. Sensitivity Table Between \( p_{13} \) and \( r \) (In Million Rupiahs)

<table>
<thead>
<tr>
<th>( p_{13} )</th>
<th>0.0175%</th>
<th>0.0432%</th>
<th>0.0689%</th>
<th>0.0946%</th>
<th>0.1203%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest Rate ( (r) )</td>
<td>0.515%</td>
<td>-2,265.86</td>
<td>-1,252.10</td>
<td>-245.54</td>
<td>753.88</td>
</tr>
<tr>
<td>0.526%</td>
<td>-2,262.52</td>
<td>-1,250.96</td>
<td>-246.58</td>
<td>750.67</td>
<td>1,740.87</td>
</tr>
<tr>
<td>0.537%</td>
<td>-2,259.19</td>
<td>-1,249.82</td>
<td>-247.62</td>
<td>747.48</td>
<td>1,735.55</td>
</tr>
<tr>
<td>0.548%</td>
<td>-2,255.87</td>
<td>-1,248.69</td>
<td>-248.65</td>
<td>744.31</td>
<td>1,730.25</td>
</tr>
<tr>
<td>0.559%</td>
<td>-2,252.55</td>
<td>-1,247.55</td>
<td>-249.67</td>
<td>741.15</td>
<td>1,724.97</td>
</tr>
</tbody>
</table>

Table 3 indicates two things:

- If the final NPV of cash flow is positive, an increase in the interest rate \( (r) \) will lead to a smaller final NPV of cash flow. However, if the final NPV of cash flow is negative, it becomes more positive (approaches zero) as \( r \) increases. Nevertheless, changes in \( r \) have minimal impact on the final NPV of cash flow.

- The final NPV of cash flow generated using the ECL model is highly sensitive to \( p_{13} \) (and it is more sensitive to \( p_{13} \) rather than \( r \)).

With the knowledge that the NPV of cash flow is highly sensitive to \( p_{13} \), the 5,000-iteration process was repeated with an adjusted \( p_{13} \) value of 0.000946 and \( p_{11} \) of 0.998924. This simulation yielded a positive average NPV of Rp409,877,840 with a standard deviation of Rp200,744,073. These results support the previously mentioned analysis that the values of \( p_{11} \) and \( p_{13} \) obtained using the ECL method could be used as base rates to determine premium prices, but they need to be adjusted or slightly increased according to debtors’ respective risk profiles. More importantly, the ECL model in this study is feasible to be used as a basis for calculating the premiums for insurance of unsecured credit.

### 4. CONCLUSIONS

The ECL model that was developed by IASB can be derived further to be used as a premium-setting model for unsecured credit insurance. To more closely relate to the concepts of insurance, the PD parameter is adjusted to become PDC. Results from the PDC model validation indicated that it is a relatively good model to determine the frequency probability of claim disbursements. Hence, the PDC obtained from the development of the ECL model can be used in the next phase of the study.

Using the PDC obtained, the cash flow simulation consisting of 5,000 iterations yielded a negative average NPV. As sensitivity analysis showed that the cash flow values are most sensitive to the PDC value, upward adjustments were made to the PDC value. This adjustment is used to reflect the fact that in reality, borrowers may have varying levels of credit risk and hence, varying levels of PDC. Repeating the simulation with the adjusted PDC value results in a new positive average NPV. Moreover, the initial negative NPV constitutes only 9.17% of total premiums received, which is arguably small. As a result, this concludes that the ECL model derived in this study is feasible and can be used to calculate premiums for unsecured credit insurance.
ACKNOWLEDGMENT

The authors would like to express gratitude to the advisors, mentors, and reviewers who have supported the authors in various ways throughout the research process.

REFERENCES
