The Journal of International Scientific Researches
2023, 8(2)

An R Package for Crossed Classification Credibility Model: Application Regarding Non-Performing Loan

Çapraz Sınıflama Kredibilite Modeli İçin R Paketi: Takipteki Kredilerle İlgili Uygulama

Seda Tugce Altan
PhD student, Gazi University, Ankara , Türkiye, stugce.altan@gazi.edu.tr,
Orcid No: https: //0000-0003-0431-5256

Muhlis Özdemir
Associate Professor, Gazi University, Ankara, Türkiye
muhlisozdemir@gazi.edu.tr ,
Orcid No: https: //0000-0002-4921-8209

Meral Ebegil
Prof. Dr., Gazi University, Ankara, Tükiye, mdemirel@gazi.edu.tr
Orcid No: https: //0000-0003-4798-3422

ArticleType / Makale Türü
ResearchArticle / Araştırma Makalesi

Keywords
Credibility, two-way crossed classification credibility model, the R package, premium estimate.

Anahtar Kelimeler
Kredibilite, iki yönlü çapraz sınıflama kredibilite modeli, R paketi, prim tahmini.

Submitted: 03 / 03 / 2023
Accepted: 14 / 04 / 2023

Abstract

Credibility theory is a calculation method which is used for making weighted estimation of balanced allocation between past and recent period data. Crossed Classification Credibility Model (CCCM) based on credibility theory is introduced by Dannenburg (1995). This model can be used when there is more than one risk factor. The two way CCCM have two risk factors. CCCM offers an alternative method when data are unclassifiable hierarchically. Simultaneously, this model considers the joint and separates the effects of risk factors. To calculate the credibility premiums in this model, variance components which obtained by solving the linear equation system must be calculated. However, this system cannot be solved explicitly. Also, too many parameters must be calculated for the premium estimation. Here, calculation errors can occur, and it is very difficult to find the correct results. Moreover, there is no tool that can easily perform these operations on a computer. In this study, the R package cccm has been developed to calculate the structural parameters easily, quickly, and accurately for CCCM. The R package cccm explained step by step for the users interested in to solve CCCM problems

Öz

Kredibilite teorisi, geçmiş ve yakın dönem veriler arasında dengeli dağılımın ağırlıklı tahminini yapmak için kullanılan bir hesaplama yöntemidir. Kredibilite teorisine dayanan Çapraz Sınıflandırma Kredibilite Modeli (CCCM) Dannenburg (1995) tarafından tanıtılmıştır. Bu model birden fazla risk faktörü olduğunda kullanılabilmektedir. İki yönlü CCCM iki risk faktörüne sahiptir. CCCM, verilerin hiyerarşik olarak sınıflandırılamadığı durumlarda alternatif bir yöntem sunar. Aynı zamanda bu model, risk faktörlerinin etkilerini birlikte ve ayrı ayrı dikkate alır. Bu modelde kredibilite primlerinin hesaplanabilmesi için doğrusal denklem sisteminin çözülmesiyle elde edilecek varyans bileşenlerinin hesaplanması gerekmektedir. Ancak bu sistem açık bir şekilde çözülemez. Ayrıca, prim tahmini için çok fazla parametrenin hesaplanması gerekmektedir. Burada hesaplama hataları oluşabilmekte ve doğru sonuçları bulmak oldukça zor olmaktadır. Üstelik bu işlemleri bilgisayarda kolayca yapabilecek bir araç da bulunmamaktadır. Bu çalışmada, CCCM için yapısal parametreleri kolay, hızlı ve doğru bir şekilde hesaplamak için R paketi cccm geliştirilmiştir. R paketi cccm, CCCM problemlerini çözmek isteyen kullanıcılar için adım adım açıklanmıştır.

Introduction

Credibility theory is a method, which allows an insurer to perform a prospective experience rating a risk or group of risk. Therefore, credibility theory is an effective solution for determining insurance premiums for contracts in a heterogeneous portfolio. The main concern of credibility is to ensure that the premium is sufficiently large to fulfill its obligations. The first approach was proposed by Mowbray (1914). Whitney (1918) conducted a study on the credibility factor value. Perryman (1932) studied partial credibility problems which are relating to limited fluctuation approach. Bailey (1950) rediscovered and advanced Whitney’s ideas. Longley-Cook (1962) tired most treatments both full credibility and partial credibility theory which were belonged to the limited fluctuation approach. Bühlmann (1967) established a greatest accuracy credibility theory. Bühlmann and Straub (1970) extended the Bühlmann's model. Jewell (1975) developed the hierarchical credibility model and the linear regression model which was proposed by Hachemeister (1975). Klugman (1987) used a full Bayesian approach in his analysis of workers' compensation insurance. Dannenburg (1995) introduced the two-way CCCM. Also, Dannenburg et al. (1996) introduced most of CCCM model formulas. Frees, et al., (1999) and Frees, et al., (2001) studied the relationship between credibility models and parametric statistical models. A generalized CCCM was introduced by Goulet (2001). Optimal variance components estimators in CCCM was studied by Wang, X. (2005). Antonio and Beirlant (2005) introduced statistical techniques to model such data within the framework of general linear models in their paper. Fellingham et al. (2005) used BUGS to estimate health insurance claim costs with construct a Bayesian hierarchical model. The application of CCCM in third-party auto insurance was applied by Šoltés and Šoltésová (2006). Rosenberg and Farrell (2008) used version 1.4 of Win- BUGS to predict the incidence and cost of hospitalization for a group of children with cystic fibrosis. Fung, et al. (2008) proposed a linear mixed model, which is relating to two-way CCCM. Wen and Wu (2011) presented a study on the independence between risk parameters. Poon and Lu (2015) proposed conditional cross-sectional covariance with general dependence via the Bühlmann-Straub credibility model. Bozikas (2019) studied the credibility regression model using mortality data which is in a hierarchical form and estimated the period dynamics of mortality under a CCCM framework. Bozikas and Pitselis G. (2020) presented the Lee–Carter method which is formulation of CCCM.

Experience rating system, which is used to determine the premium for next year by taking into account both the individual and the collective experience, are difficult to use when data are countless and heterogeneous. This time hierarchical credibility model, which is developed by Jewell (1975) can be proposed as an alternative method to experience rating system. The hierarchical credibility model, successfully allows complex tree-like classification structures. However, the hierarchical credibility model is not suitable for use when there is an interaction risk factors are non-nested. To illustrate the approach, a portfolio of automobiles that includes risks categorized by the gender and age of the insured could be considered. There is no reason gender should be modelled at under level than age or vice versa. When risk factors non-nested like this, CCCM can be used.

In CCCM, an insurance portfolio is subdivided by two qualitative risk factors which is modeled in symmetrical way. Also, CCCM offers an alternative method when data are unclassifiable hierarchically. Similarly, this model allows for the joint and separate effects of risk factors. To calculate the premiums in this model, variance components which are obtained by solving the system of linear equations should be calculated. Generally, the system cannot be solved explicitly. Simultaneously, too many parameters must be calculated for the premium estimation. Here, calculation errors can occur, and it is very difficult to find the correct results. Moreover, there is no tool that can easily perform these operations on a computer. This study introduces an R package has been developed to easily calculate the structural parameters and variance components in CCCM. Additionally, the R package presented in this paper enables CCCM to be used in different sciences.

When the literature is examined, studies that related to credibility theory and CCCM classified by publishing years, their methods and the used data are given in table 1.

In section 2, two ways CCCM is introduced step by step. Also, the formulas related to the model and calculations of credibility premium are given.

In section 3, the application of CCCM with using the real data is given. The data consist of two categories, which are bank type and type of loan. The main aim of this study is to calculate the credibility premium using CCCM.

Table 1. Studies of Credibility Theory and CCCM

Study	Method	Data
(Whitney, A. W., 1918)	Experience rating	Auto insurance
(Bailey, A. L., 1950)	Simulation	Hypothetical
(Longley-Cook, L. H., 1962)	Simulation	Hypothetical
(Hachemeister, C. A., 1975)	Credibility for regression models	Auto insurance
(Jewell, W. S., 1975)	Hierarchical credibility model	Hypothetical
(Klugman, S., 1987)	Hierarchical normal linear model.	Hypothetical
(Dannenburg, D. R., 1995)	Crossed classification credibility models	Auto insurance
(Frees, E. W., Young, V. R., and Luo, Y., 1999)	Longitudinal data modeling	Auto insurance
(Frees, E. W., Young, V. R., and Luo, Y., 2001)	Panel data analysis	Health care
(Goulet, V., 2001)	Crossed classification credibility models	Theorical
(Wang, X., 2005)	Simulation	Hypothetical
(Fellingham, G. W., Dennis Tolley, H, and Herzog, T. N., 2005)	A linear mixed model and a Bayesian hierarchical model	Medical
(Antonio, K. and Beirlant, J., 2005)	Generalized linear mixed models	Theorical
(Šoltés, E., and Šoltésová, T., 2006)	Crossed Classification Credibility Model	Auto insurance
(Rosenberg, M. A., and Farrell, P. M., 2008)	A Bayesian statistical model	Health care
(Fung, W. K., and Xu, X., 2008)	Simulation	Hypothetical
(Wen, L. and Wu, X., 2011)	Regression credibility modeling	Hypothetical
(Poon, J. and Lu, Y., 2015)	Simulation	Crop insurance
(Bozikas, A. E., 2019)	Multi-level hierarchical credibility regression model	Mortality data
(Bozikas, A., and Pitselis, G., 2020)	Crossed classification credibility formulation of the Lee–Carter	Mortality data

1. The Two Way Crossed Classification Model

In the two-way crossed classification model, an insurance portfolio is grouped by two risk factors. Rows represent the first risk factor which, is defined as I *(i =1,2…,I)*, columns represent the second risk factor which is defined as *J* (*j*=*1,2,.., J*). (I, J) corresponds to row *I* and column *J*. T_{ij} shows the observations for available risks.

The process and the solving steps of the two-way CCCM with *R package* cccm are given in figure 1.

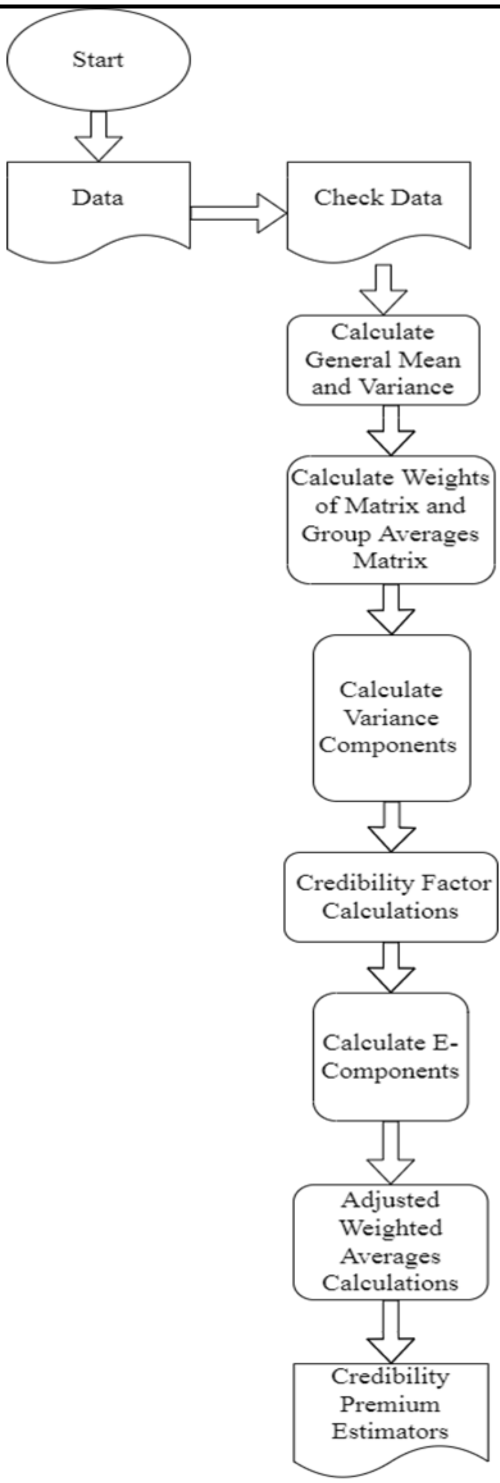

Figure 1. The Process of The Two-Way CCCM

The credibility premium estimator in two-way CCCM (Dannenburg at all. 1996: 41, 60) can be written as follows,

$$X_{i,T_{ij}+1} = m + E_i^{(1)} + E_j^{(2)} + E_{ij} + E_{ijt} \qquad t=1,2, \ldots, T_{ij} \tag{1}$$

In this formula E-components are all independent with mean zero and represent a deviation of the collective mean. $E_i^{(1)}$, $E_j^{(2)}$can be interpreted as a deviation of the collective mean for first and second risk factors respectively. The variables E_{ij} represent the interaction between risk factors, E_{ijt} represent the deviation in the cell for the period of observation t from the expected claim size given,

$E_i^{(1)}$, $E_j^{(2)}$and E_{ij}. For $E_i^{(1)}$, $E_j^{(2)}$ the variances are $b^{(1)}$ and $b^{(2)}$; respectively. Also, the variances of E_{ij} and E_{ijt} are equal to $b^{(12)}$ and $\frac{s^2}{w_{ijt}}$ respectively.

Weighted averages and the weights can be calculated from equality (2-4),

$$w_{ij\Sigma} = \sum_{t=1}^{T_{ij}} w_{ijt}, \qquad w_{\Sigma\Sigma\Sigma} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T_{ij}} w_{ijt} \tag{2}$$

$$X_{ijw} = \frac{1}{w_{ij\Sigma}}\sum_{t=1}^{T_{ij}} X_{ijt}w_{ijt} \tag{3}$$

$$m = X_{www} = \frac{1}{w_{\Sigma\Sigma\Sigma}}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ijw}w_{ij\Sigma} \tag{4}$$

The terms m, X_{ijw} and $w_{ij\Sigma}$ are called the structure parameters. These parameters are generally unknown and must be estimated from the entire data in the portfolio. m is a collective mean of weighted average in the entire portfolio. X_{ijw} is a weighted average on the I^{th} level of the first factor and simultaneously on the J^{th} level of the second factor. In addition to this $w_{\Sigma\Sigma\Sigma}$ represents the sum of the weights in the portfolio.

$b^{(1)}$, $b^{(2)}$ and $b^{(12)}$ are variance components and can be obtained from equality (5-7) by solving liner equation systems.

$$E\left[\sum_{i=1}^{I}\frac{1}{I}\left(\frac{1}{w_{i\Sigma\Sigma}}\sum_{j=1}^{J}\left(X_{ijw}-X_{iww}\right)^2 w_{ij\Sigma} - \frac{s^2(J-1)}{w_{i\Sigma\Sigma}}\right)\right] = \left(b^{(2)}+b^{(12)}\right) * \left(1-\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{w_{ij\Sigma}}{w_{i\Sigma\Sigma}}\right)^2\right) \tag{5}$$

$$E\left[\sum_{j=1}^{J}\frac{1}{J}\left(\frac{1}{w_{\Sigma j\Sigma}}\sum_{i=1}^{I}\left(X_{ijw}-X_{wjw}\right)^2 w_{ij\Sigma} - \frac{s^2(I-1)}{w_{\Sigma j\Sigma}}\right)\right] = \left(b^{(2)}+b^{(12)}\right) * \left(1-\sum_{j=1}^{J}\sum_{i=1}^{I}\left(\frac{w_{ij\Sigma}}{w_{\Sigma j\Sigma}}\right)^2\right) \tag{6}$$

$$E\left[\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{w_{ij\Sigma}}{w_{\Sigma\Sigma\Sigma}}\left(X_{ijw}-X_{www}\right)^2 - \frac{s^2(IJ-1)}{w_{\Sigma\Sigma\Sigma}}\right] = b^{(1)}\left[1-\sum_{i=1}^{I}\left(\frac{w_{i\Sigma\Sigma}}{w_{\Sigma\Sigma\Sigma}}\right)^2\right] + b^{(2)}\left[1-\sum_{i=1}^{I}\left(\frac{w_{\Sigma j\Sigma}}{w_{\Sigma\Sigma\Sigma}}\right)^2\right] + b^{(12)}\left[1-\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{w_{ij\Sigma}}{w_{\Sigma\Sigma\Sigma}}\right)^2\right] \tag{7}$$

Remark 1. (Determining variance components)

One way to solve the system of linear equation, which is given equality (5-7) is to use matrix calculus. This is given as follows:

$\vec{E}^{(1)} = \vec{C}^{(1)} - D^{(1)}\vec{E}^{(2)}$;
$\vec{E}^{(2)} = \vec{C}^{(2)} - D^{(2)}\vec{E}^{(1)}$;

with $\vec{C}^{(1)}$ the *I x1* vector with its *k*th element equal to

$c_k = z_k^{(1)}(X_{kzw} - m)$,

and the $D^{(1)}$ *I x J* matrix in which the (*k, l*) element equals $z_k^{(1)}\, z_{kl}/\, z_{k\Sigma}$. The *J x 1* vector $\vec{C}^{(2)}$ and the *J x I* matrix $D^{(2)}$are defined in the same way.

Substituting $\vec{E}^{(2)}$ into $\vec{E}^{(1)}$ gives,

$\vec{E}^{(1)} = \vec{C}^{(1)} - D^{(1)}\vec{C}^{(2)} + D^{(1)}D^{(2)}\vec{E}^{(1)}$,

$\vec{E}^{(1)} = \left(I - D^{(1)}D^{(2)}\right)^{-1}\left(\vec{C}^{(1)} - D^{(1)}\vec{C}^{(2)}\right)$

similarly,

$\vec{E}^{(2)} = \left(I - D^{(2)}D^{(1)}\right)^{-1}\left(\vec{C}^{(2)} - D^{(2)}\vec{C}^{(1)}\right)$

can be calculated provided that the displayed inverses exist.

If the variance $b^{(12)}$ between the cells is relatively large, more credibility is given to the observations in cell (*I, J*). Like this, if the variability $b^{(1)}$ between classes is high compared with the variance $b^{(12)}$ between the cells in row *I*, much credence is given to the risk experience within column *I*. Similarly, the credibility factor corresponding to column *J*.

To obtain E-components, credibility factor values must be calculated. For cells (*I, J*) the credibility factor value is $z_{ij}^{(12)}$, for row and column, the credibility factor values are $z_i^{(1)}$ and $z_j^{(2)}$ respectively. To calculate credibility factor values equality (8-10) can be used.

$$z_{ij}^{(12)} = \frac{b^{(12)}w_{ij\Sigma}}{b^{(12)}w_{ij\Sigma} + s^2} \tag{8}$$

$$z_i^{(1)} = \frac{b^{(1)}z_{i\Sigma}^{(12)}}{b^{(1)}z_{i\Sigma}^{(12)} + b^{(12)}} \tag{9}$$

$$z_j^{(2)} = \frac{b^{(2)}z_{\Sigma j}^{(12)}}{b^{(2)}z_{\Sigma j}^{(12)} + b^{(12)}} \tag{10}$$

The formula of E-components is shown in equality (11-13), and is obtained using credibility factor values.

$$E_i^{(1)} = z_i^{(1)}\left(\frac{1}{z_{i\Sigma}^{(12)}}\sum_{j=1}^{J}\left(X_{ijw} - E_j^{(2)}\right).z_{ij}^{(12)} - m\right) \tag{11}$$

$$E_j^{(2)} = z_j^{(2)}\left(\frac{1}{z_{\Sigma j}^{(12)}}\sum_{i=1}^{I}\left(X_{ijw} - E_i^{(1)}\right).z_{ij}^{(12)} - m\right) \tag{12}$$

$$E_{ij} = z_{ij}^{(12)}\left(X_{ijw} - m - E_i^{(1)} - E_j^{(2)}\right) \tag{13}$$

Then, adjusted weighted averages must be calculated to consider the effect of each risk factor. The formula of adjusted weighted averages is shown in equality (14-15),

$$Y_{izw} = \frac{1}{z_{i\Sigma}^{(12)}}\sum_{j=1}^{J}\left(X_{ijw} - E_j^{(2)}\right).z_{ij}^{(12)} \tag{14}$$

$$Y_{zjw} = \frac{1}{z_{\Sigma j}^{(12)}}\sum_{i=1}^{I}\left(X_{ijw} - E_i^{(1)}\right).z_{ij}^{(12)} \tag{15}$$

Finally, the credibility premium estimates, which use adjusted weighted average, are calculated using equality (16),

$$X_{ijt} = m + z_{ij}(X_{ijw} - m) + (1 - z_{ij})z_i^{(1)}(Y_{izw} - m) + (1 - z_{ij})z_j^{(2)}(Y_{zjw} - m) \quad (16)$$

Algorithms for the formulas given in the equations are presented step by step in figure 2 to easily calculate the credibility premium using CCCM.

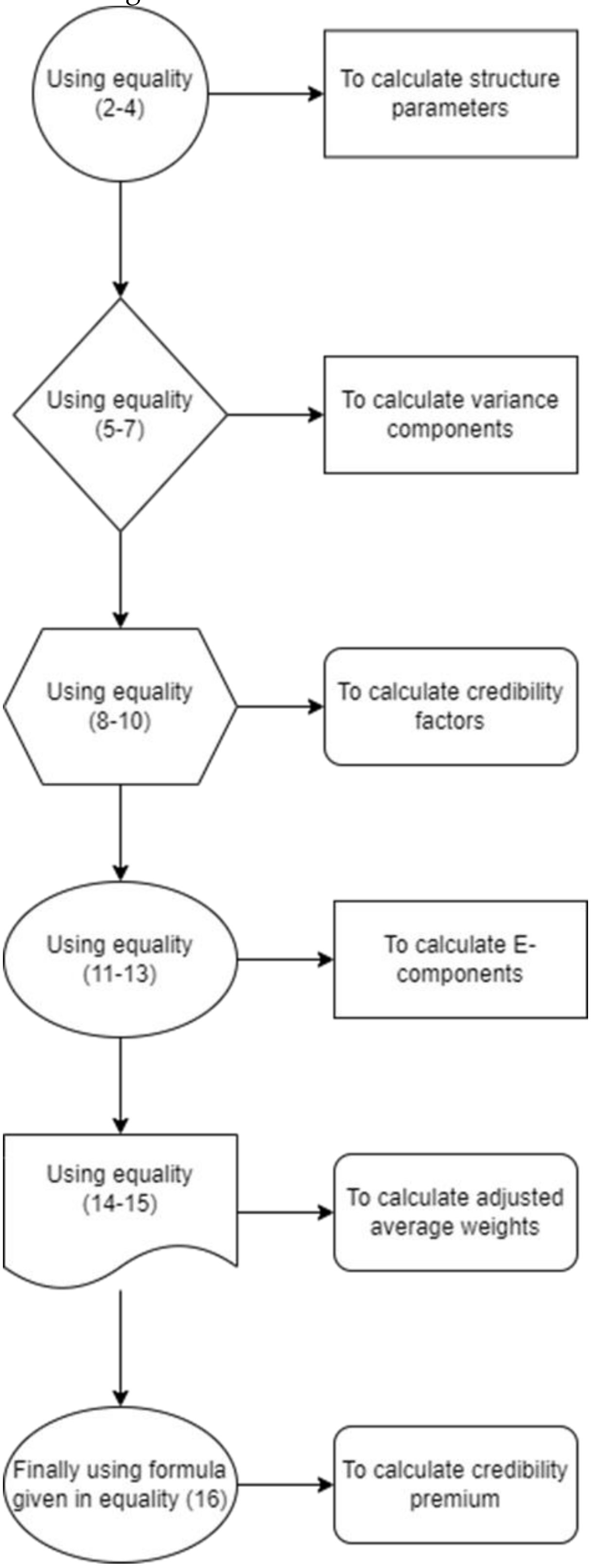

Figure 2. Algorithm of the formulas given in the equations

2. CCCM in Practice

Package *cccm* is developed for crossed credibility classification method and is available on The Comprehensive R Archive Network [https://cran.rproject.org/web/packages/cccm/index.html]. Its usage will be shown in this subchapter step by step as follows.

debt data

debt data is a pre-defined dataset in the *cccm* package to the illustration usage of the *cccm* package. These are a real data published by the Banking Regulation and Supervision Agency [*https://www.bddk.org.tr/BultenAylik/en*] that consists of 106 rows and 4 columns. Features of the data consist of *bank, loan, weights,* and *debt* columns. *Bank* is categorical data of bank type. Bank type includes four subcategories such as State Banks, Deposit Banks, Foreign Banks, and Privately Owned Deposit Banks. *Loan* is categorical data of debt type. Loan type includes three subcategories such as non-performing vehicle, home, and consumer loan. *Weights* consist of Numeric values of the weights. *Debt* consist of Numeric values of debt.

A code block for the data *debt* and its output for the first six rows are given as follows.

```
> install.package("cccm")
> library(cccm)
> head(debt)

      bank   loan      weights debt
1     state  vehicle      1     7675
2     state  vehicle      1     7282
3     state  vehicle      1     7137
4     state  vehicle      1     6917
5     state  vehicle      1     6113
6     state  vehicle      1     6132
```

The R package *cccm* consists of fifteen functions and seven of them are available for the users. Eight of the non-available functions for the users are to handle algorithm's necessities. The Available functions' explanations for the users are as follows.

calculate_generalMean() : The Function calculates weighted average amount for the entire portfolio. The general mean is also known as *so-called collective mean* in the literature.

calculate_group_averages_matrix() : The function calculates the sum of the weighted average amount for all categories for each risk factor.

calculate_obs_and_group_weights() : The function calculates the sum of the weights for all categories for each risk factor.

calculate_variance_and_std() : The function calculates the variation from the collective mean for each risk factor and all categories in the risk factors.

calculate_varianceComponents() : The function calculates the variance of the E components, which are representing deviation from the collective mean for each risk factor and both the risk factors. Also, this function enables calculate variance components easily for users; thus solving the linear equation system is not needed.

calculate_weights_of_obs_matrix() : The function calculates adjusted weighted average. This function uses the effects of each risk factor when calculating.

estimate_credibility() : A function calculates the credibility estimation immediately and reports all the results.

This section introduces cccm package and its usage. As an example, we used real data which published by the Turkey Banking Regulation and Supervisory Board. 3-month period data were used and selected approximately, 2019-03/ 2021-12 periods for Ankara City. There are two risk factors that are bank type and type of non-performing loan. Bank type includes four subcategories such as State Banks, Deposit Banks, Foreign Banks and Privately Owned Deposit Banks. Loan includes three subcategories such as vehicle loan, home loan and consumer loan. Our aim is to calculate the non-performing loan for the next period using the cross-classification credibility model.

In the two-way model the t^{th} loss in the cell corresponding to bank type I and class J of the non-performing loan debt is represented by the random variable X_{ijt}. The weights w_{ijt} are equal to one, but the set of observed payments is not balanced because the number of observations differs from cell to cell.

By using equality (2-4) structure parameters are obtained. These parameters are:

m= 189659, $s^2 = 21704.9$

To obtain these values by using *cccm* package in R, are shown as follows.

The functions in the *cccm* package have for parameters. These parameters are *raw_data*, *categorical_columns*, *weights_column*, and *debt_column* and should be defined before the usage. The values assigned to these parameters are the column numbers of the dataset.

```
> raw_data <- debt
> categorical_columns <- c(1,2)
> weights_column <- 3
> debt_column <- 4
```

To calculate the value *m*, *calculate_generalMean()* function can be used.

```
> calculate_generalMean(raw_data, categorical_columns, weights_column, debt_column)

[1] 189659.1
```

To obtain the variance and standard deviation, *calculate_variance_and_std()* function can be used.

```
> calculate_variance_and_std(raw_data, categorical_columns, weights_column, debt_column)

  variance         std
471102729.8     21704.9
```

In table 2 below average amounts (X_{ijw}) paid by credit insurer are tabulated. The twelve averages in the center of the table are cell averages. In the margins the row and column averages are given; the bottom right corner shows the overall average of amounts. For example, as can be seen from table 2, the realization of the cell average X_{11w} is equal to 6.878 TL. From table 2, we can say that the average amount is higher in consumer loan. Also deposit bank is the highest loan amount.

Table 2. Average Amounts Paid by The Credit Insurer

	Type Of Non-Performing Loan (Thousand TL)			
Bank Type	Vehicle Loan	Home Loan	Consumer Loan	X_{iww}
State Banks	6,878	49,724	375,082	**126,361**
Deposit Banks	15,651	125,006	1,088.596	**370,012**
Foreign Banks	5,912	54,156	401,841	**136,422**
Privately Owned Deposit Banks	5,309	27,816	328,363	**119,662**
X_{wjw}	**8,389**	**65,800**	**547,847**	**189,659**

To calculate the average amount paid by the credit insurer *calculate_group_averages_matrix()* function can be used.

```
> calculate_group_averages_matrix(raw_data, categorical_columns, weights_column, debt_column)
          vehicle      home  consumer
state    6878.000  49724.29  375082.2
deposit 15651.300 125006.11 1088595.8
foreign  5911.833  54155.80  401841.0
local    5308.600  27815.88  328362.9
```

Table 3. Weights For Bank Type and Type of Loan

	Type Of Non-Performing Loan			
Bank Type	Vehicle Loan	Home Loan	Consumer Loan	$w_{i\Sigma\Sigma}$
State Banks	8	7	6	**21**
Deposit Banks	10	9	8	**27**
Foreign Banks	12	10	9	**31**
Privately Owned Deposit Banks	10	8	9	**27**
$w_{\Sigma j\Sigma}$	**40**	**34**	**32**	**106**

In table 3, the total number of observations is 106, varying from 6 to 12 per cell in the two-way table. The weighting process was done in time. For example, for State Bank (*I=1*) and Vehicle Loan (*J=1*) weight is 8. This means that for this cell (*I=J=1*) 8 periods of data, each of which consists of 3-month data, are drawn.

To calculate the group weights *calculate_weights_of_obs_matrix()* function can be used.

```
> calculate_weights_of_obs_matrix(raw_data, categorical_columns, weights_column, debt_column)

        vehicle home consumer
state         8    7        6
deposit      10    9        8
foreign      12   10        9
local        10    8        9
```

Then, using equality (5-7), $b^{(1)}$, $b^{(2)}$ and $b^{(12)}$ parameter estimators can be obtained. These estimators are

$b^{(1)}$ =2006102779, $b^{(2)}$ = 68928817482, $b^{(12)}$= 42516742950.

To calculate parameter estimators *calculate_varianceComponents()* function can be used.

```
> calculate_varianceComponents(raw_data, categorical_columns, weights_column, debt_column)

            [,1]
[1,]  2006102779
[2,] 68928817482
[3,] 42516742950
```

By using these estimators, credibility factors can be calculated using eq. 8-10. In table 4, the credibility factor values are given.

Table 4. Credibility Factor Values

	Type Of Non-Performing Loan			
Bank Type	Vehicle Loan	Home Loan	Consumer Loan	$z_i^{(1)}$
State Banks	0.9986	0,9984	0.9982	**0.1238**
Deposit Banks	0.9989	0.9988	0.9986	**0.1239**
Foreign Banks	0.9991	0.9989	0.9988	**0.1239**
Privately Owned Deposit Banks	0.9989	0.9986	0.9988	**0.1239**
$z_j^{(2)}$	**0.8663**	**0.8662**	**0.8662**	

In table 4, credibility factor values are close to each other; this is caused by the observation values are close to each other. The average of credibility factors z_{ij} that are linked to the risk experience within the cells is 0.9987, which is quite high. We can say more credibility is given to the cell. Also, more credibility is given to the columns than the rows. So that, in the credibility premium estimation, more weight will be given to the type of non-performing loan. The highest credibility factor value is seen in vehicle loan.

Although the function *estimate_credibility()* function reports all the results at once, it is possible to calculate the credibility factor values by using the followed R code.

```
> estimate_credibility(raw_data, categorical_columns, weights_column,
debt_column)$credibility_factor_values

          vehicle      home  consumer
state   0.9986169 0.9984196 0.9981567
deposit 0.9988932 0.9987704 0.9986169
foreign 0.9990775 0.9988932 0.9987704
local   0.9988932 0.9986169 0.9987704
```

Based on the credibility factor values and the solution of the linear equation system which is given equality (11-13) we can obtain E-components. The calculated E-components are listed in table 5.

Table 5. Estimations of the E Components

	Type Of Non-Performing Loan			
Bank Type	Vehicle Loan	Home Loan	Consumer Loan	$E_i^{(1)}$
State Banks	-17,772	-24,646	-116,678	**-9,981**
Deposit Banks	-39,154	20,397	565,662	**20,194**
Foreign Banks	-19,984	-21,470	-91,262	**-8,741**
Privately Owned Deposit Banks	-18,508	-45,693	-162,575	**-10,818**
$E_j^{(2)}$	**-155,004**	**-105,269**	**312,297**	

In table 5, among banks Deposit Banks has the highest difference from the overall mean. Like that, among type of non-performing loan consumer loan has the highest difference from the collective mean. Also, non-performing consumer loan drawn from deposit banks is has the highest difference from the collective mean. The lowest difference from the collective mean in rows is Foreign Banks, in column is home loan.

The estimation of the E components followed by an R code can be used.

```
> estimate_credibility(raw_data, categorical_columns, weights_column,
debt_column)$Eij_values_for_cells
          vehicle      home   consumer
state   -17771.52 -24646.45 -116678.04
deposit -39154.14  20396.57  565661.91
foreign -19983.81 -21470.12  -91262.29
local   -18507.56 -45693.12 -162575.17
```

Using equations 14-15 we can calculate the adjusted weighted averages.

Finally, we can calculate the credibility premium estimates using equation (16). The credibility premium estimates are given in table

Table 6. The Credibility Premium Estimates

	Type Of Non-Performing Loan (Thousand TL)		
Bank Type	Vehicle Loan	Home Loan	Consumer Loan
State Banks	6,906	49,765	375,303
Deposit Banks	15,700	124,985	1,087.820
Foreign Banks	5,932	54,180	401,957
Privately Owned Deposit Banks	5,329	27,877	328,564

Credibility premium estimates can be calculated using the following R code

```
> estimate_credibility(raw_data, categorical_columns, weights_column,
debt_column)$credibility_predictions

          vehicle      home  consumer
state    6905.670  49764.50  375302.6
deposit 15700.172 124985.32 1087819.8
foreign  5932.328  54180.44  401956.7
local    5329.268  27877.36  328563.8
```

If the user wants to see all the results at once, then *estimate_credibility()* function can be used.

```
> estimate_credibility(raw_data, categorical_columns, weights_column, debt_column)

$general_mean
[1] 189659.1

$variance_and_std
   variance         std
471102729.8     21704.9

$weights_of_obs_matrix
        vehicle home consumer
state         8    7        6
deposit      10    9        8
foreign      12   10        9
local        10    8        9

$group_averages_matrix
          vehicle      home  consumer
state    6878.000  49724.29  375082.2
deposit 15651.300 125006.11 1088595.8
foreign  5911.833  54155.80  401841.0
local    5308.600  27815.88  328362.9

$variance_components
            [,1]
[1,]  2006102779
[2,] 68928817482
[3,] 42516742950

$credibility_factor_values
          vehicle      home  consumer
state   0.9986169 0.9984196 0.9981567
deposit 0.9988932 0.9987704 0.9986169
foreign 0.9990775 0.9988932 0.9987704
local   0.9988932 0.9986169 0.9987704

$first_risk_factor_Zi
    state   deposit   foreign     local
0.1238251 0.1238645 0.1238812 0.1238645

$second_risk_factor_Zj
  vehicle      home  consumer
0.8662661 0.8662435 0.8662323
```

```
$Eij_values_for_cells
          vehicle      home   consumer
state   -17771.52 -24646.45 -116678.04
deposit -39154.14  20396.57  565661.91
foreign -19983.81 -21470.12  -91262.29
local   -18507.56 -45693.12 -162575.17

$first_risk_factor_Ei
           [,1]
[1,]  -9980.537
[2,]  20194.147
[3,]  -8740.577
[4,] -10818.008

$second_risk_factor_Ej
              [,1]
vehicle  -155004.5
home     -105268.9
consumer  312297.1

$adj_weighted_avg_for_first_risk_factor
         vehicle     home  consumer
state   161658.6 154748.2  62669.35
deposit 170466.9 229991.8 775224.94
foreign 160767.9 159248.2  89433.81
local   160135.6 132900.7  16046.05

$adj_weighted_avg_for_second_risk_factor
          vehicle      home  consumer
state   16835.219  59610.46  384352.9
deposit -4537.818 104683.08 1066923.9
foreign 14638.894  62826.76  410076.7
local   16108.759  38580.45  338763.8

$credibility_predictions
          vehicle      home  consumer
state    6905.670  49764.50  375302.6
deposit 15700.172 124985.32 1087819.8
foreign  5932.328  54180.44  401956.7
local    5329.268  27877.36  328563.8
```

Conclusion

In this paper, a real data set was used to apply crossed classification credibility method with the new developed R package *cccm*. The implementation of the R package in this study is presented to introduce the package and evaluate its results. The obtained results show that the presented R package is working quickly and give the results correctly and has high sensitivity. In this way, the results can be evaluated in detail through every perspective and can be instantly checked. The presented R package can be applied to any crossed classification credibility method in any risk-related field.

Generally, we can say that the credibility factor values in table 3 are close to each other. The main reason is that the weights in the cells in table 2 are close. Also in table 3, the credibility factor values for the second risk factor are higher than those for the first risk factor. This is because the variance components of the second risk factor are higher than those of the first risk factor. So that, when calculating the credibility premium second risk factor has more effect than the first risk factor. When we look E-components which are given in table 4, the collective mean of the second risk factor is higher than that of the first risk factor. Considering all the results, we can say that the credibility premium for the second risk factor will be high even without calculating the credibility premium.

When we look at the credibility estimation values in table 6, among the bank types, deposit banks have more loans than the other bank types for the next period. Among the loan types, consumer loans are higher than the other loan types. Furthermore, the lowest non-performing loan is seen in vehicle loan. So that, we can say that for the first risk factor, deposit banks have riskier and for the second risk factor, consumer loans are riskier. When both risk factors were considered, consumer loans drawn from deposit banks are riskier.

As a conclusion, deposit banks can take measures against the risk of non-repayment while giving a consumer loan and the package can be used as a tool.

References

Altan S.T., Özdemir M. & Ebegil M. cccm: An R Package for Crossed Classification Credibility Model, 2022. *Title*. [Avaible online at: https://cran.r-project.org/web/packages/cccm/index.html], R package version 1.2.

Antonio, K., & Beirlant, J. (2007). Applications of generalized linear mixed models in actuarial statistics. Insurance: Mathematics and Economics. 40, 58–76

Bailey, A.L., (1950). Credibility Procedures - La Place's Generalization of Bayes' Rule and the Combination of Collateral Knowledge with Observed Data. New York State Insurance Department. 37, 7-23, 94-115.

Bozikas, A., & Pitselis, G. (2020). Incorporating crossed classification credibility into the Lee–Carter model for multi-population mortality data. Insurance: Mathematics and Economics.. 93, 353-368.

Bozikas, A. E. (2019). Actuarial models in demography. (Doctoral dissertation, Piraeus University).

Bühlmann, H. (1967). Experience rating and credibility. ASTIN Bulletin: The Journal of the IAA. 4(3), 199-207.

Bühlmann, H. & Straub, E., (1970) Glaubwürdigkeit für Schadensätze. Bulletin of the Swiss Association 70, 111-133. English translation by C.E. Brooks.

Dannenburg, D., (1995). “Crossed classification credibility models”, In. Trans. 25th Int. Congress of Actuaries. 4: 1–36

Dannenburg, D. R., Kaas, R., & Goovaerts, M. J. (1996). Practical actuarial credibility models. Institute of Actuarial Science and Econometrics of the University of Amsterdam.

Fellingham, G. W., Dennis Tolley, H., & Herzog, T. N. (2005). Comparing credibility estimates of health insurance claims costs. North American Actuarial Journal. 9(1), 1-12.

Frees, E. W., Young, V. R., & Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. Insurance: Mathematics and Economics. 24(3), 229-247.

Frees, E. W., Young, V. R., & Luo, Y. (2001). Case studies using panel data models. North American Actuarial Journal. 5(4), 24-42.

Fung, W. K., & Xu, X. (2008). Estimation of Structural Parameters in Crossed Classification Credibility Model Using Linear Mixed Models. In COMPSTAT (241-251). Physica-Verlag HD.

Goulet, V. (2001). A generalized crossed classification credibility model. Insurance: Mathematics and Economics. 28(2), 205-216.

Hachemeister, C. (1975). Credibility for regression models with application to trend (reprint). Credibility: Theory and Applications. Edited by P. Kahn. New York: Academic Press, Inc. 307-48.

Jewell, W.S. (1975) The Use of Collateral Data in Credibility Theory: A Hierarchical Model. Giornale dell'Istituto Italiano degli Attuari. 38:1-16.

Klugman, S. (1987). Credibility for classification ratemaking via the hierarchical normal linear model. In Proceedings of the Casualty Actuarial Society. Vol. 74, 272-321.

Longley-Cook, L. H. (1962). An introduction to credibility theory. Casualty Actuarial Society. 49, 194-221.

Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium. In Proceedings of the Casualty Actuarial society. Vol. 1, No. 1, 24-30.

Perryman, F.S., (1932). Some Notes on Credibility. PCAS, 19, 65-84.

Poon, J., & Lu, Y. (2015). A spatial cross-sectional credibility model with dependence among risks. North American Actuarial Journal. 19(4), 289-310.

Rosenberg, M. A., & Farrell, P. M. (2008). Predictive modeling of costs for a chronic disease with acute high-cost episodes. North American Actuarial Journal. 12(1), 1-19.

Šoltés, E., & Šoltésová, T. (2006). Application of Crossed Classification Credibility Model in Third-party Auto Insurance in Slovak Republic. Management information systems. 59.

Turkey Banking Regulation and Supervisory Board Banking Sector Data by Province *Title*. [Available online at: https://www.bddk.org.tr/BultenFinTurk], Retrieved on October 6, 2022

Wang, X. (2005). On the numerical evaluation of optimal variance components estimators in crossed classification credibility (Doctoral dissertation, Concordia University).

Wen, L. & Wu, X. (2011). The credibility estimator with general dependence structure over risks. Communications in Statistics-Theory and Methods, 40:1893-1910.

Whitney, A., The Theory of Experience Rating. Proceedings of the Casualty Actuarial Society, 4: 274–292 (1918).