Erik Šoltés
Tatiana Šoltésová
University of Economics
Faculty of Economic Informatics
Bratislava, Slovakia

Application of Crossed Classification Credibility Model in Third-party Auto Insurance in Slovak Republic

Summary
In this article we reviewed the two-way crossed classification credibility model. This model is an extension of the hierarchical models of Jewell and Taylor. When the risk factors are not nested then a hierarchical model is not applicable. In the crossed classification credibility models, the risk factors are modelled without restrictions of a hierarchical structure and that makes them of great practical interest. In the two-way crossed classification credibility model the risks in a portfolio are classified based on two risk factors. In this model the credibility premium for a certain contract is equal to the overall mean for the portfolio plus adjustments for the risk experience within the contract itself and the risk experience within the class of the risk factors to which it belongs. The objective of this article is to show alternatives of an application of crossed classification credibility models in third-party auto insurance in Slovak Republic.

Key words
Crossed classification credibility model, insurance

1. Introduction

In this article we discuss a calculation of the net premium in non-life insurance using crossed classification credibility models. The theory of credibility is based on assumption that the insurance subject wants to set the net premiums to compensate losses from insurance claims, i.e. the net premium amount should equal the average claim amount.

Two types of data are used in the process of estimation of average claim amount using models based on the theory of credibility: data about own insurance risk and data about comparable insurance risks. In the crossed classification credibility models the portfolio is classified by certain number of risk factors. In comparison to hierarchical credibility models, the risk factors do not have to be nested, resulting in a broader range of possibilities of application of these models in real life.

2. Assumptions of the multi-way crossed classification credibility models

Let P be a number of risk factors, according to which the portfolio is classified and let *p*-th ($p = 1, 2, \ldots, P$) risk factor have J_p categories. Then the category $j_p \in \{1, 2, \ldots, J_p\}$ pertaining to the *p*-th factor is characterized by an unknown risk parameter $\theta_{j_p}^{(p)}$. In addition, let's assume that in the cell $j_1, j_2, \ldots, j_P$ there are observed average claim amounts $X_{j_1, j_2, \ldots, j_P, t}$ ($t = 1, 2, \ldots, T_{j_1, j_2, \ldots, j_P}$) for $T_{j_1, j_2, \ldots, j_P}$ periods. The general multi-way crossed classification model consists of the following assumptions:

(i) For each *p* the $\theta_{j_p}^{(p)}$ are i.i.d.; all occurring θ's are independent. The variances of the risks $X_{j_1, j_2, \ldots, j_P, t}$ are finite.

(ii) For certain functions $\mu_{i_1 \ldots i_q}(\cdot)$:

$$E\left[X_{j_1 \ldots j_P\, t} / \theta_{j_{i_1}}^{(i_1)}, \ldots, \theta_{j_{i_q}}^{(i_q)}\right] = \mu_{i_1 \ldots i_q}(\theta_{j_{i_1}}^{(i_1)}, \ldots, \theta_{j_{i_q}}^{(i_q)}),$$

for *q*, $i_1, \ldots, i_q \in \{1, 2, \ldots, P\}$; $j_{i_1} \in \{1, 2, \ldots, J_{i_1}\}$, and so on . Without loss of generality it is assumed that $i_1 < i_2 < \ldots < i_q$.

(iii)
$$E\left[Cov\left[X_{j_1 \ldots j_P\, t}, X_{k_1 \ldots k_P\, u} / \theta^{(1)}, \ldots, \theta^{(P)}\right]\right] = \delta_{j_1 \ldots j_P\, t,\, k_1 \ldots k_P\, u} \cdot \frac{s^2}{w_{j_1 \ldots j_P\, t}},$$

where $\boldsymbol{\theta}^{(q)} = (\theta_1^{(q)}, \ldots, \theta_{J_q}^{(q)})'$, and $\delta_{j_1 \ldots j_P\, t, k_1 \ldots k_P\, u}$ is the Kronecker symbol which is equal to one if $j_1 = k_1, \ldots$ $j_P = k_P$, $t = u$ and zero otherwise.

Note: In a special case, when there is an equal number of observations in all cells, i.e. $T_{j_1, j_2, \ldots, j_P} = T$ and all weights $w_{j_1, j_2, \ldots, j_P, t}$ equal one, we talk about a balanced model and all relations mentioned in the article can be simplified. In reality, a balanced set of observations is, however, unlikely.

3. Two-way crossed classification credibility model

In the article we focus on the two-way crossed classification model, in which the portfolio is classified according to two qualitative risk factors ($P = 2$). The credibility estimate of an average claim amount $X_{ij,T_{ij}+1}$ ($i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$) is calculated as follows:

$$\begin{aligned}\hat{X}_{ij,T_{ij}+1} &= m + z_{ij}^{(12)}(X_{ijw} - m) + \\ &+ (1 - z_{ij}^{(12)})z_i^{(1)}(Y_{izw} - m) + \\ &+ (1 - z_{ij}^{(12)})z_j^{(2)}(Y_{zjw} - m)\end{aligned}, \qquad (1)$$

where m is a mean of claim amounts in the entire portfolio and its estimate is

$$\text{est } m = X_{www} = \frac{1}{w_{\Sigma\Sigma\Sigma}}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ijw} w_{ij\Sigma}, \qquad (2)$$

where

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

X_{ijw} is an average claim amount on the i-th level of the first factor and simultaneously on the j-th level of the second factor and is calculated as follows:

$$X_{ijw} = \frac{1}{w_{ij\Sigma}}\sum_{t=1}^{T_{ij}} X_{ijt} w_{ijt}. \qquad (3)$$

Y_{izw} is an adjusted average claim amount on the i-th level of the first factor. This statistic represents experience with insurance risk in the i-th class. Similarly Y_{zjw} represents experience with insurance risk in the j-th class of the second factor. These characteristics are determined by the following terms:

$$Y_{izw} = \frac{1}{z_{i\Sigma}^{(12)}}\sum_{j=1}^{J}(X_{ijw} - \hat{\Xi}_j^{(2)}) \cdot z_{ij}^{(12)}, \qquad (4a)$$

$$Y_{zjw} = \frac{1}{z_{\Sigma j}^{(12)}}\sum_{i=1}^{I}(X_{ijw} - \hat{\Xi}_i^{(1)}) \cdot z_{ij}^{(12)}. \qquad (4b)$$

Random variables $\Xi_i^{(1)}$ a $\Xi_j^{(2)}$ are estimated by solving a system of ($I+J$)-normal equations, which are obtained by modification of the following terms

$$\hat{\Xi}_i^{(1)} = z_i^{(1)}\left(\frac{1}{z_{i\Sigma}^{(12)}}\sum_{j=1}^{J}(X_{ijw} - \hat{\Xi}_j^{(2)}) \cdot z_{ij}^{(12)} - \hat{m}\right), \qquad (5a)$$

$$\hat{\Xi}_j^{(2)} = z_j^{(2)}\left(\frac{1}{z_{\Sigma j}^{(12)}}\sum_{i=1}^{I}(X_{ijw} - \hat{\Xi}_i^{(1)}) \cdot z_{ij}^{(12)} - \hat{m}\right). \qquad (5b)$$

$z_{ij}^{(12)}$, $z_i^{(1)}$, $z_j^{(2)}$ are credibility factors and can reach values from the interval $\langle 0; 1\rangle$. The credibility factor $z_{ij}^{(12)}$ represents degree of credibility of insurance risk information in the cell (i, j). $z_i^{(1)}$ represents the credibility provided by information obtained on the i-th level of the first risk factor. Similarly the credibility factor $z_j^{(2)}$ represents the credibility of experiences, obtained from the j-th level of the second factor. The credibility factors are calculated as follows:

$$z_{ij}^{(12)} = \frac{b^{(12)} w_{ij\Sigma}}{b^{(12)} w_{ij\Sigma} + s^2}, \qquad (6)$$

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

$$z_j^{(2)} = \frac{b^{(2)} z_{\Sigma j}^{(12)}}{b^{(2)} z_{\Sigma j}^{(12)} + b^{(12)}}, \qquad (7a) \text{ a } (7b)$$

Where a degree of reliability s^2 is replaced with estimate

$$\hat{s}^2 = \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{J}(T_{ij} - 1)_+}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T_{ij}}(X_{ijt} - X_{ijw})^2 w_{ijt}, \qquad (8)$$

where $(T_{ij} - 1)_+ = T_{ij} - 1$ if $T_{ij} \geq 2$ and zero otherwise.

Parameters $b^{(12)}$, $b^{(1)}$, $b^{(2)}$ are obtained by solving the system of three equations of three unknowns. The first one is

$$E\left[\sum_{i=1}^{I}\frac{g_i^{(1)}}{g_\Sigma^{(1)}}\left(\frac{1}{w_{i\Sigma\Sigma}}\sum_{j=1}^{J}(X_{ijw} - X_{iww})^2 w_{ij\Sigma} - \frac{s^2(J-1)}{w_{i\Sigma\Sigma}}\right)\right] = \qquad (9a)$$

$$= (b^{(2)} + b^{(12)}) \cdot \left(1 - \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{g_i^{(1)}}{g_\Sigma^{(1)}} \left(\frac{w_{ij\Sigma}}{w_{i\Sigma\Sigma}} \right)^2 \right),$$

where

$$X_{iww} = \frac{1}{w_{i\Sigma\Sigma}} \sum_{j=1}^{J} X_{ijw} w_{ij\Sigma} \qquad (w_{i\Sigma\Sigma} = \sum_{j=1}^{J} w_{ij\Sigma}) \tag{10}$$

is an average claim amount in the *i*-th class of the first factor.

The second equation (9b) is analogical and is obtained by consistent replacement of indices *i* a *j*. The third equation has form

$$E\left[\frac{1}{w_{\Sigma\Sigma\Sigma}} \sum_{i=1}^{I} \sum_{j=1}^{J} (X_{ijw} - X_{www})^2 w_{ij\Sigma} - \frac{s^2(IJ-1)}{w_{\Sigma\Sigma\Sigma}} \right] = \\ = b^{(1)} \left(1 - \sum_{i=1}^{I} \left(\frac{w_{i\Sigma\Sigma}}{w_{\Sigma\Sigma\Sigma}} \right)^2 \right) + b^{(2)} \left(1 - \sum_{j=1}^{J} \left(\frac{w_{\Sigma j\Sigma}}{w_{\Sigma\Sigma\Sigma}} \right)^2 \right) + \\ + b^{(12)} \left(1 - \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\frac{w_{ij\Sigma}}{w_{\Sigma\Sigma\Sigma}} \right)^2 \right). \tag{11}$$

It can be proved, that term (1) can be expressed in the form

$$\hat{X}_{ij, T_{ij}+1} = \hat{m} + \hat{\Xi}_i^{(1)} + \hat{\Xi}_j^{(2)} + \hat{\Xi}_{ij}^{(12)}, \tag{12}$$

where

$$\hat{\Xi}_{ij}^{(12)} = z_{ij}^{(12)} (X_{ijw} - \hat{m} - \hat{\Xi}_i^{(1)} - \hat{\Xi}_j^{(2)}). \tag{13}$$

4. Application

We applied the two-way crossed classification model to a database of claim amounts and a number of insured vehicles from 2002 to 2004 in third-party auto insurance of an anonymous insurance company in Slovak Republic. In this article we consider two risk factors that impact claim amounts – engine power (kW) and region (districts of Slovak Republic that reflect regional classification at the level NUTS 2). The first risk factor consists of six categories $(i = 1, 2, .., 6)$ and the second factor consists of eight categories $(j = 1, 2, \ldots, 8)$. For individual categories of vehicles we calculated the average claim amount per one vehicle and the total number of insured vehicles, which are listed in Table 1. The column Total represents the calculated average claim amount per one vehicle and the total number of insured vehicles at individual levels of engine power. The row Total represents these statistics calculated for vehicles from the same district of Slovak Republic. The bottom right corner of Table 1 shows the calculated average claim amount per one vehicle for the entire portfolio $(X_{www} = 2\,370 \text{ SKK})$ and the total number of insured vehicles $(w_{\Sigma\Sigma\Sigma} = 48\,306)$.

Parameter s^2 was estimated using the term (8):

$$\hat{s}^2 = 149\,898\,715.43.$$

Parameters $b^{(12)}$, $b^{(1)}$, $b^{(2)}$ were obtained by solving the system of three equations (9a), (9b) and (11):

$$\hat{b}^{(12)} = 161\,508.98,$$

$$\hat{b}^{(1)} = 211\,348.95, \; \hat{b}^{(2)} = 19\,657.53.$$

We used these estimates for the calculation of credibility factors $z_{ij}^{(12)}$ using the term (6). Based on these factors using the terms (7a) and (7b) we calculated credibility factors $z_i^{(1)}$ and $z_j^{(2)}$ for individual classes of the first and the second risk factors. All credibility factors are summarized in Table 2.

Data from the category $(i, j) = (6, 7)$ have the least credibility (11.2 %). This is due to the lowest number of insured vehicles in this category $(w_{ij\Sigma} = 117)$. On the other hand, when calculating net premiums the own information about vehicles with engine power from 41 kW to 55 kW from Bratislava district are the most credible (90 %) where we also maintain the largest database – 8 337 insured vehicles.

Within the first factor, the most credible (87.7%) are data from its second class (engine power from 41kW to 55kW) and the least credible (73.5%) are data from last class of engine power (engine power over 112 kW). These results correspond with the number of insured vehicles $(w_{i\Sigma\Sigma})$ in individual classes of the engine power. In the last row of Table 2 are listed credibility factors for individual classes of the second risk factor – districts of Slovak Republic. According to the results when calculating net premiums the data about vehicles from the same district are less credible (17.1 % to 34.7 %) than the data about vehicles with the same class of engine power (73.5% to 87.7%). This is mainly due to a greater variability of claims within

Table 1

Average claim amounts (X_{ijw} - in SKK) and total number of insured vehicles ($w_{ij\Sigma}$) in individual categories of vehicles

Engine power (kW)	District								Total X_{iww} ($w_{i\Sigma\Sigma}$)
	BA ($j=1$)	TT ($j=2$)	TN ($j=3$)	NR ($j=4$)	ZA ($j=5$)	BB ($j=6$)	PO ($j=7$)	KE ($j=8$)	
27-40 ($i=1$)	2 310 (1 277)	910 (501)	1 426 (421)	1 477 (549)	748 (577)	1 618 (791)	1 489 (307)	1 642 (428)	**1 585 (4 851)**
41-55 ($i=2$)	2 484 (8 337)	1 948 (1 720)	1 814 (1 703)	2 389 (1 935)	1 620 (1 759)	2 195 (2 247)	1 687 (1 419)	2 549 (1 583)	**2 221 (20 703)**
56-67 ($i=3$)	2 788 (2 500)	1 712 (434)	1 475 (501)	2 069 (506)	2 513 (606)	2 823 (650)	1 274 (334)	3 929 (484)	**2 525 (6 015)**
68-89 ($i=4$)	2 675 (4 893)	1 626 (651)	2 391 (836)	1 531 (685)	4 236 (699)	2 066 (736)	1 567 (475)	2 929 (589)	**2 525 (9 561)**
90-111 ($i=5$)	2 745 (1 945)	2 335 (250)	2 013 (333)	1 745 (248)	2 638 (298)	2 296 (270)	3 126 (131)	2 690 (277)	**2 555 (3 750)**
112 - ($i=6$)	3 984 (1 645)	5 236 (255)	1 770 (305)	1 529 (343)	2 482 (276)	3 642 (308)	4 730 (117)	3 258 (173)	**3 470 (3 422)**
Total X_{wjw} ($w_{\Sigma j\Sigma}$)	**2 700 (20 597)**	**1 975 (3 811)**	**1 863 (4 099)**	**1 989 (4 266)**	**2 191 (4 214)**	**2 261 (5 002)**	**1 790 (2 783)**	**2 737 (3 534)**	**2 370 (48 306)**

Notes: BA – Bratislava district, TT – Trnava district, TN - Trenčin district, NR - Nitra district, ZA – Žilina district, BB – Banska Bystrica district, PO – Prešov district, KE – Košice district

Source: Authors' own calculation

Table 2

Credibility factors $z_{ij}^{(12)}$ for individual categories of vehicles

Engine power (kW)	District								$z_i^{(1)}$
	BA ($j=1$)	TT ($j=2$)	TN ($j=3$)	NR ($j=4$)	ZA ($j=5$)	BB ($j=6$)	PO ($j=7$)	KE ($j=8$)	
27-40 ($i=1$)	0,579	0,351	0,312	0,372	0,383	0,460	0,249	0,316	**0,798**
41-55 ($i=2$)	0,900	0,650	0,647	0,676	0,655	0,708	0,605	0,630	**0,877**
56-67 ($i=3$)	0,729	0,318	0,351	0,353	0,395	0,412	0,264	0,343	**0,806**
68-89 ($i=4$)	0,841	0,412	0,474	0,425	0,429	0,442	0,339	0,388	**0,831**
90-111 ($i=5$)	0,677	0,212	0,264	0,211	0,243	0,225	0,124	0,230	**0,741**
112 - ($i=6$)	0,639	0,216	0,247	0,270	0,230	0,249	0,112	0,157	**0,735**
$z_j^{(2)}$	**0,347**	**0,208**	**0,218**	**0,219**	**0,221**	**0,233**	**0,171**	**0,201**	X

Source: Authors' own calculation

Table 3

Estimates of components $\Xi_{ij}^{(12)}$ of credibility premiums for individual classes of vehicles

Engine power (kW)	District								$\hat{\Xi}_i^{(1)}$
	BA ($j=1$)	TT ($j=2$)	TN ($j=3$)	NR ($j=4$)	ZA ($j=5$)	BB ($j=6$)	PO ($j=7$)	KE ($j=8$)	
27-40 ($i=1$)	275	-253	-47	-41	-358	-33	-31	-43	**-694**
41-55 ($i=2$)	171	-91	-144	232	-341	32	-233	195	**-235**
56-67 ($i=3$)	170	-203	-288	-84	44	170	-280	490	**26**
68-89 ($i=4$)	93	-302	40	-333	784	-157	-263	163	**35**
90-111 ($i=5$)	109	-9	-82	-124	50	-32	94	37	**55**
112 - ($i=6$)	495	481	-292	-386	-132	144	194	16	**681**
$\hat{\Xi}_j^{(2)}$	**160**	**-46**	**-99**	**-90**	**6**	**15**	**-63**	**104**	X

Source: Authors' own calculation

Table 4

Credibility premiums $\hat{X}_{ij,2005}$ (in SKK) for individual categories of vehicles

Engine power (kW)	District							
	BA ($j=1$)	TT ($j=2$)	TN ($j=3$)	NR ($j=4$)	ZA ($j=5$)	BB ($j=6$)	PO ($j=7$)	KE ($j=8$)
27-40 ($i=1$)	2 110	1 377	1 530	1 545	1 324	1 657	1 582	1 736
41-55 ($i=2$)	2 465	1 998	1 892	2 277	1 800	2 182	1 839	2 434
56-67 ($i=3$)	2 725	2 146	2 008	2 222	2 446	2 580	2 052	2 990
68-89 ($i=4$)	2 658	2 057	2 346	1 982	3 195	2 263	2 079	2 672
90-111 ($i=5$)	2 693	2 369	2 243	2 211	2 481	2 407	2 456	2 566
112 - ($i=6$)	3 705	3 485	2 659	2 574	2 924	3 209	3 182	3 171

Source: Authors' own calculation

one district in comparison to the variability of claims related to vehicles of the same engine power.

Estimates of credibility net premiums for individual categories of vehicles were determined using the term (12). Components $\Xi_i^{(1)}$ and $\Xi_j^{(2)}$ were estimated by solving a system of fourteen normal equations which are obtained using the terms (5a) and (5b).

Subsequently we estimated components $\Xi_{ij}^{(12)}$ using the term (13). The values of the estimated components are tabulated in Table 3.

The resulting credibility premium is then sum of four elements - $\hat{m}$, $\hat{\Xi}_i^{(1)}$, $\hat{\Xi}_j^{(2)}$ and $\hat{\Xi}_{ij}^{(12)}$. For example, the credibility premium for the vehicles in the category (1, 1), i.e. vehicles with engine power from 27 kW to 40 kW and from BA district, is:

$$\hat{X}_{1,1,2005} = 2\,370 - 694 + 160 + 275 = 2\,111 \text{ SKK}.$$

The first component $(\hat{m} = 2\,370 \text{ SKK})$ is the overall average of claim amounts as an estimate of *m*. Vehicles with engine power from 27 kW to 40 kW get a bonus of 694 SKK, vehicles from Bratislava district are penalized by amount of 160 SKK and finally the vehicles with the above mentioned engine power and from Bratislava district get an additional malus of 275 SKK. From the last row of Table 3 we can see that drivers from Bratislava, Košice, Banská Bystrica and Žilina districts should pay above average premiums. The results in the last row confirm that the claim amount caused by a vehicle rises with the increasing engine power of this vehicle.

The credibility net premium can be also estimated using the credibility formula (1), which uses adjusted average claim amount for the *i*-th class of the first factor or the *j*-th class of the second factor. Statistics for the category (1, 1) are as follows

$$Y_{1zw} = \frac{4\,531.339}{3.021} \doteq 1\,500 \qquad \text{a}$$

$$Y_{z1w} = \frac{12\,352.240}{4.365} \doteq 2\,830.$$

According to the credibility formula (1) the credibility premium for the vehicles from the category (1, 1) is then:

$$\begin{aligned}\hat{X}_{1,1,2005} &= 2\,370 + 0.579 \cdot (2\,310 - 2\,370) + \\ &+ (1 - 0.579) \cdot 0.798 \cdot (1\,500 - 2\,370) + \\ &+ (1 - 0.579) \cdot 0.347 \cdot (2\,830 - 2\,370) = 2\,110 \text{ Sk}.\end{aligned}$$

The credibility premiums for individual categories of vehicles are calculated in Table 4.

References

1. Daannenburg, D.R. (1996), Basic actuarial credibility models, Amsterdam.
2. Chajdiak, J. (2005), Štatistické úlohy a ich riešenie v Exceli, STATIS, Bratislava.
3. PACÁKOVÁ, V.(2000), Aplikovaná poistná štatistika, ELITA, Bratislava.
4. Wanat, S. (1998), "Regresyjne, ewolucyjne i hierarchiczne modele wiarogodnosci w ubezpieczeniach", Zborník príspevkov. IV. slovensko - poľský odborný seminár Štatistické metódy v sociálno-ekonomickom výskume - teória a aplikácie, EKONÓM, Bratislava, pp.71-83.

Mgr. Erik Šoltés PhD.

Katedra štatistiky FHI EU
Dolnozemská cesta 1
852 35 Bratislava
Slovakia
e-mail: soltes@euba.sk
Tel.: 02/672 95 717

Mgr. Erik Šoltés, PhD. serves as a lecturer at the University of Economics, Department of Statistics in Bratislava since 1998. His research activities are focused on exploiting the Theory of Credibility (Credibility Regression Models in particular) in non-life insurance. He also lectures in Introduction to Statistics and Regression Analysis.

Mgr. Tatiana Šoltésová

Katedra štatistiky FHI EU
Dolnozemská cesta 1
852 35 Bratislava
Slovakia
e-mail: soltesova@euba.sk
Tel.: 02/672 95 808

Mgr. Tatiana Šoltésová serves as a lecturer at the University of Economics, Department of Mathematics in Bratislava since 1998. At the present she is finishing her PhD. Her dissertation focuses on the Theory of Martingales in Stochastic Processes and their particular use in life insurance. She lectures in Introduction to Mathematical Analyses and Algebra.