Estimation of the Credibility in the Original Bühlmann Model Illustrated by Introducing a Regression Technique

Conf. dr. Virginia ATANASIU
Catedra de Matematică, Academia de Studii Economice, Bucureşti

In this article we give the mathematical theory of some credibility models. The first section describes the original credibility model of Bühlmann and we derive the best linearized credibility estimator for this model. His original model, involving only one contract, contains the basics of all further credibility models. Section 2 contains a description of the Hachemeister regression model allowing for effects like inflation.
Keywords: *linearized regression credibility premium, the original credibility model of Bühlmann.*

Introducere

Pentru a introduce efectele timpului asupra primei nete de risc a contractului din modelul de credibilitate original al lui Bühlmann sau pentru a introduce efectele inflației asupra acesteia am apelat la tehnica regresiei. Mai precis, pentru a estima credibilitatea din modelul original al lui Bühlmann ce implică un singur tip de contract de asigurare non-viață, în condițiile existenței inflației sau a unei tendințe inflaționiste accentuate (pronunțate), am folosit modelul de regresie a credibilității elaborat de actuarul de marcă Hachemeister. *Secțiunea 1* dezbate din punct de vedere teoretic modelul de credibilitate al lui Bühlmann, relevându-i importanța pentru rezultatele viitoare privind credibilitatea în asigurările non-viață, iar *Secțiunea 2* prezintă modelul de regresie a credibilității ce face obiectul acestui articol.

Secțiunea 1

Modelul original al lui Bühlmann constă dintr-un singur contract de asigurare non-viață, aflat sub observație statistică un număr de $t(\geq 2)$ ani. Riscul implicat de respectivul contract este notat cu X şi asimilat cu o variabilă aleatoare nenegativă. Lucrăm în ipoteza că, riscul pentru polița considerată este stabil în timp, ceea ce înseamnă că poate fi caracterizat prin aceeaşi valoare a parametrului de risc de-a lungul anilor de valabilitate ai contractului; parametrul riscului X este notat cu θ şi este presupus variabilă aleatoare reală; θ este cunoscut şi sub numele de variabilă aleatoare structurală a contractului. Denumirile acordate lui θ sunt justificate de faptul că, acesta reprezintă variabila aleatoare ce descrie caracteristicile riscului luat în considerație. Variabilele observabile ale riscului X se notează cu $X_1, X_2, ..., X_t$; X_r este înregistrarea din anul r pentru contractul de parametru θ, unde $r=\overline{1,t}$; de regulă semnificația lui X_r este aceea de solicitare de despăgubire, din anul r pentru respectivul contract, unde $r=\overline{1,t}$.

Deci, acest model constă din variabila aleatoare structurală θ şi din variabilele aleatoare observabile X_r, unde $r=\overline{1,t}$, a.î. putem afirma despre contractul considerat că este un vector aleator de componente: θ (parametrul de risc aleator neobservabil al contractului) şi $X_1, X_2, ..., X_t$ (observațiile efectuate asupra poliței respective). Aşadar, contractul implicat se identifică cu vectorul $(\theta, X_1, X_2, ... X_t)$, sau simplificând scrierea cu perechea $(\theta, \underline{X}')$; $\underline{X}' \overset{not.}{=}$ (vectorul observațiilor) = (vectorul de componente $X_1, X_2, ..., X_t$).

Introducem prima netă de risc a contractului, dacă parametrul de risc al acestuia este θ, prin următorul şir de egalități:

$\mu(\theta) = M(X_r \mid \theta) = M(X \mid \theta), \forall r = \overline{1,t}$.

Pentru modelul de față teoria credibilității, aşa după cum a fost dezvoltată de Bühlmann are ca scop principal, estimarea lui $\mu(\theta)$. Estimatorul ce urmează a fi folosit în acest sens, sunt funcţii de variabilele observabile, adică de componentele vectorului aleator $\underline{X}' = (X_1, X_2, ..., X_t)$, determinată a.î. să minimizeze eroarea medie pătratică, mai precis, estimarea lui $\mu(\theta)$ necesită rezolvarea problemei de minim:

$$\underset{g(\cdot)}{Min}\, M\left\{[\mu(\theta) - g(X_1, X_2, ..., X_t)]^2\right\},$$

unde $g(\cdot)$ este o funcţie de observaţiile $X_r, r = \overline{1,t}$, arbitrară. În modelul original de credibilitate al lui Bühlmann, clasa de funcţii g este limitată la mulţimea D de funcţii liniare şi neomogene (la mulţimea D de combinaţii liniare şi neomogene) ale variabilelor aleatoare observabile $X_1, X_2, ..., X_t$, de forma:

$$g(X_1, X_2, ..., X_t) = c_0 + \sum_{r=1}^{t} c_r X_r,$$

cu $c_0, c_1, ..., c_t$ constante reale, iar $X_r, r = \overline{1,t}$ presupuse a avea varianţa finită. Bühlmann introduce aşa-numiţii parametrii de structură, notaţi: m, a, s^2 şi definiţi de actuar după cum urmează:

$$\begin{cases} m = M(X_r) = M(X) = M[\mu(\theta)], r = \overline{1,t}, \\ a = Var[M(X_r \mid \theta)] = Var[M(X \mid \theta)] = Var[\mu(\theta)], r = \overline{1,t}, \\ s^2 = M[Var(X_r \mid \theta)] = M[Var(X \mid \theta)] = M[\sigma^2(\theta)], r = \overline{1,t}, \end{cases}$$

unde:

$$\sigma^2(\theta) = Var(X_r \mid \theta) = Var(X \mid \theta), r = \overline{1,t}$$

(evident sunt îndeplinite ipotezele cerute de aceste notaţii).

În continuare, prezentăm ipotezele de lucru ale modelului şi anume:

(1) pentru θ dat, variabilele aleatoare observabile (observaţiile anuale) $X_r, r = \overline{1,t}$ sunt independente şi identic distribuite condiţional (i.i.d. condiţional);

(2) $0 < Var(X_r) < \infty, r = \overline{1,t}$;

(3) $M(X_r \mid \theta) = \mu(\theta), r = \overline{1,t}$ (prima netă de risc a contractului de parametru θ este definită ca media observaţiilor (riscurilor) condiţionate; am putut reprezenta media observaţiilor anuale condiţionate $(X_r \mid \theta), r = \overline{1,t}$ prin aceeaşi funcţie $\mu(\cdot)$ depinzând de θ, dar independentă de $r, r = \overline{1,t}$, întrucât aceste variabile aleatoare condiţionate sunt i.d.);

(4) $Var(X_r \mid \theta) = \sigma^2(\theta), r = \overline{1,t}$ (am putut reprezenta varianţa observaţiilor anuale condiţionate $(X_r \mid \theta), r = \overline{1,t}$ prin aceeaşi funcţie $\sigma^2(\cdot)$, depinzând de θ, dar independentă de $r = \overline{1,t}$, întrucât aceste variabile aleatoare condiţionate sunt i.d.);

(5) $Cov(X_r, X_q \mid \theta) = 0, r, q = \overline{1,t}$, cu $r < q$ (covarianţa condiţionată de θ dintre observaţiile efectuate asupra contractului de parametru θ în ani diferiţi este nulă, întrucât variabilele aleatoare condiţionate $(X_r \mid \theta), r = \overline{1,t}$ sunt independente; ipoteza (5) exprimă faptul că observaţiile (riscurile) condiţionate din ani diferiţi nu sunt corelate).

Observaţie:
Pentru contractul considerat, rezultă următoarea matrice de covarianţe a observaţiilor (riscurilor) condiţionate din anii $r = \overline{1,t}$ şi anume:

$$Cov(\underline{X}' \mid \theta) = I^{(t,t)} \sigma^2(\theta)$$

Pe baza presupunerilor (1)-(5) şi a parametrilor de structură m, s^2, a, am obţinut folosind m.c.m.m.p. rezultatul liniar şi neomogen de credibilitate, în speţă prima liniară şi neomogenă de credibilitate, pe care o ilustrăm în continuare:

„Estimatorul liniar şi neomogen de credibilitate al primei nete de risc $\mu(\theta)$ pentru contractul de parametru θ bazat pe observaţiile (înregistrările) $\underline{X}'$ efectuate asupra contractului considerat este dat de relaţia:

$\hat{\mu}(\theta) = z\overline{X} + (1 - z)m$, unde $\overline{X} = \frac{1}{t}\sum_{1}^{t} X_r$ este media aritmetică a înregistrărilor individuale pentru respectivul contract, ce indică estimatorul individual pentru $\mu(\theta)$, iar $z = at/(s^2 + at)$ este factorul de credibilitate pentru contractul considerat, ce exprimă o pondere a înregistrărilor individuale pentru contractul de parametru θ, indicând de câtă credibilitate se bucură acestea”.

Secţiunea 2

Modelul de regresie a credibilităţii elaborat de actuarul de marcă Hachemeister vizează presupunerea din cadrul modelului (de credibilitate pentru un singur tip de contract) prezentat în cadrul secţiunii **1.**, conform căreia: prima netă de risc a contractului de parametru θ este independentă de timp, fiind definită prin egalitatea:

$\mu(\theta) = M(X_j \mid \theta) =$ (media observaţiilor (riscurilor) anuale condiţionate), oricare ar fi anul $j \geq 1$

Datorită inflaţiei, prima netă de risc a contractului de parametru θ, adică $\mu(\theta)$ se schimbă (se modifică, variază) în timp, ceea ce înseamnă că este dependentă de timp, deci că trebuie scrisă astfel:

$\mu_j(\theta) = M(X_j \mid \theta), j \geq 1$

După aceste remarci introductive privind necesitatea considerării modelului de regresie a credibilităţii, prezentăm ipotezele de lucru ale modelului lui Hachemeister.

Fie $\underline{X}' = (X_1, X_2, ..., X_t)$ vectorul $(1 \times t)$ aleator al observaţiilor şi θ parametrul de risc aleator neobservabil, ce caracterizează riscul considerat.

În cele ce urmează, formulăm ipotezele de lucru ale modelului de regresie a credibilităţii, aşa după cum au fost ele introduse de actuar şi anume:

(1) dat fiind θ - parametrul de risc aleator neobservabil al contractului, variabilele aleatoare observabile (observaţiile anuale) ale respectivului contract, adică $X_1, X_2, ..., X_t$ sunt independente condiţional (s-a presupus că dispunem de înregistrările solicitărilor de despăgubire pe durata a $t(\geq 2)$ ani pentru contractul de parametru θ);

(2) $\qquad M(X_j \mid \theta) = \mu_j(\theta) = \underline{x}'_j \underline{\beta}(\theta), j = \overline{1,t}$

(prima netă de risc a contractului de parametru θ din anul j, definită ca media observaţiilor (riscurilor) anuale condiţionate este, conform presupunerii de regresie făcute de actuar, de forma: $\underline{x}'_j \underline{\beta}(\theta)$,unde primul factor din presupunerea de regresie: $\underline{x}'_j$ este vectorul $(1 \times q)$ nealeator, cunoscut, ale cărui componente depind de timp (deci de inflaţie), fapt ilustrat de prezenţa indicelui j în notaţia acestui vector ($q \leq 1$), iar al doilea factor din presupunerea de regresie: $\underline{\beta}(\theta)$ este vectorul $(q \times 1)$ aleator de regresie necunoscut, adică vectorul ce conţine q constante în raport cu timpul, necunoscute, numite constante (în raport cu timpul) de regresie); ipoteza (2) scrisă matriceal este echivalentă cu următoarea:

$M(\underline{X} \mid \theta) = \underline{\mu}^{(t,1)}(\theta) = x^{(t,n)} \underline{\beta}^{(n,1)}(\theta), rgx = n(\leq t)$

(vectorul $(t \times 1)$ al primelor nete de risc pentru contractul de parametru θ din anii 1,2,...,t sau vectorul $(t \times 1)$ al primelor nete de risc anuale pentru contractul de parametru θ, definit ca media condiţionată a vectorului observaţiilor anuale efectuate asupra contractului de parametru θ pe durata a $t(\geq 2)$ ani este, conform presupunerii de regresie scrise matriceal, de forma: $x^{(t,n)} \underline{\beta}^{(n,1)}(\theta)$, unde $x^{(t,n)}$ este matricea $(t \times n)$ nealeatoare cunoscută, de rang egal cu $n(\leq t)$, iar $\underline{\beta}^{(n,1)}(\theta)$ este vectorul $(n \times 1)$ aleator de regresie necunoscut (vector ce conţine n constante în raport cu timpul, necunoscute, numite constante de regresie));

(3) $\qquad Cov\left[\underline{\beta}^{(n,1)}(\theta)\right] = a = a^{(n,n)} =$ (matrice $(n \times n)$ pozitiv definită); ipoteza (3) afirmă despre covarianţa vectorului de regresie că este o matrice de tipul $(n \times n)$, presupusă pozitiv definită;

(4) $\qquad M[Cov(\underline{X} \mid \theta)] = \phi = \phi^{(t,t)} =$ (matrice $(t \times t)$ pozitiv definită) $= \begin{pmatrix} M[Var(X_1 \mid \theta)] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & M[Var(X_t \mid \theta)] \end{pmatrix}$;

ipoteza (4) afirmă că matricea de covarianţe a observaţiilor (riscurilor) anuale condiţionate este o matrice diagonală de dimensiune $(t \times t)$, pozitiv definită având forma mai sus prezentată, dacă ţinem seama de faptul că:

$$Cov(\underline{X} \mid \theta) = \begin{pmatrix} Var(X_1 \mid \theta) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Var(X_t \mid \theta) \end{pmatrix},$$

conform ipotezei (1);

(5) $Var(X_r \mid \theta) = \sigma^2(\theta), r = \overline{1,t}$ (s-a presupus că varianțele observațiilor (riscurilor) anuale condiționate se reprezintă prin aceeaşi funcție $\sigma^2(\cdot)$, depinzând de θ, dar independentă de $r, r = \overline{1,t}$).

Pe baza ipotezelor (1)-(5) şi a parametrilor structurali definiți de actuar, după cum urmează şi anume:

$$\begin{cases} \underline{b} = \underline{b}^{(n,1)} = M\left[\underline{\beta}^{(n,1)}(\theta)\right] \\ s^2 = M\left[\sigma^2(\theta)\right] \\ a \text{ (a se vedea ipoteza (3))} \end{cases}$$

(presupuşi independenți de timp, adică de j, cu $j = \overline{1,t}$), am obținut folosind m.c.m.m.p. şi tehnica regresiei, prima liniară şi neomogenă de regresie a credibilității, pe care o ilustrăm în continuare:

„Estimatorul liniar şi neomogen de credibilitate al primei nete de risc $\mu_j(\theta)$ pentru contractul de parametru θ din anul j, bazat pe observațiile anuale $\underline{X}' = (X_1, X_2, ..., X_t)$ efectuate asupra respectivului contract pe durata a $t(\geq 2)$ ani, adică $\hat{\mu}_j(\theta)$ este dat de relația:

$\hat{\mu}_j(\theta) = \underline{x}_j'\left[Z\hat{\underline{b}} + (I - Z)\underline{b}\right]$, cu $j = \overline{1,t}$, unde $\hat{\underline{b}} = \hat{\underline{b}}^{(q\times 1)} = \left(x'\phi^{-1}x\right)^{-1} x'\phi^{-1}\underline{X}$ este un vector $(q \times 1)$ având expresia menționată şi semnificația estimatorului liniar şi neomogen de credibilitate θ- nedeplasat pentru vectorul aleator necunoscut de regresie $\underline{\beta}(\theta)$, ceea ce, conform definiției, înseamnă că estimatorul liniar şi neomogen de credibilitate $\hat{\underline{b}}$ îl estimează aproape sigur (a.s.) în medie condiționată de θ pe $\underline{\beta}(\theta)$, adică $M\left(\hat{\underline{b}} \mid \theta\right) = \underline{\beta}(\theta)$ a.s. (deci un estimator foarte bun pentru vectorul aleator necunoscut de regresie $\underline{\beta}(\theta)$), iar $Z = Z^{(q\times q)} = ax'\phi^{-1}x\left(I + ax'\phi^{-1}x\right)^{-1}$ este o matrice de tipul $(q \times q)$ având expresia menționată şi semnificația factorului de credibilitate rezultat, adică a ponderii lui $\hat{\underline{b}}$, ce arată de câtă credibilitate se bucură acesta".

Concluzii

1) Subliniem faptul că în rezultatul pe care l-am obținut mai sus, factorul reprezentat de paranteza dreaptă şi anume: „$Z\hat{\underline{b}} + (I - Z)\underline{b}$" indică estimatorul liniar şi neomogen de credibilitate pentru vectorul aleator necunoscut de regresie pentru $\underline{\beta}(\theta)$.

2) Factorul la care facem referire în cadrul observației 1), apare ca o combinație liniar-convexă sau ca o medie aritmetică ponderată a estimatorului liniar şi neomogen de credibilitate θ- nedeplasat pentru $\underline{\beta}(\theta)$, reprezentat de $\hat{\underline{b}}$ cu estimatorul colectiv pentru vectorul de regresie $\underline{\beta}(\theta)$, reprezentat de $\underline{b}$.

3) Aşadar, estimând prima netă de risc $\mu_j(\theta)$ a contractului de parametru θ din anul j, cu $j = \overline{1,t}$ implicat în modelul de regresie a credibilității elaborat de actuarul Hachemeister, am estimat implicit vectorul aleator necunoscut de regresie $\underline{\beta}(\theta)$ ce figurează în ipoteza de regresie scrisă pentru această primă.

Bibliografie

[1]. Atanasiu V., *Contributions to the credibility theory,* thesis of doctorate, University of Bucharest-Faculty of Mathematics, 2000.

[2]. Atanasiu V., *Calcule de credibilitate,* Revista „Studii şi Cercetări de Calcul Economic şi Cibernetică Economică" 4 / 1998, XXXII.

[3]. Goovaerts M.J., Kaas R., Van Herwaarden A.E., BauwelinckxT., *Insurance Series, volume 3, Efective Actuarial Methods,* University of Amsterdam, The Netherlands, 1991.

[4]. Pentikäinen T., Daykin C.D., Pesonen M., *Practical Risk Theory for Actuaries,* Université Pierré et Marie Curie, 1990.

[5]. Sundt B., *An Introduction to Non-Life Insurance Mathematics,* Veroffentlichungen des Instituts für Versicherungswissenschaft der Universität Mannheim Band 28 , VVW Karlsruhe, 1984.