PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1088 — 2005
Inwestycje finansowe i ubezpieczenia – tendencje światowe a polski rynek

Włodzimierz Szkutnik

Akademia Ekonomiczna w Katowicach

SKŁADKA W ASPEKCIE LINIOWEJ ZALEŻNOŚCI KORELACYJNEJ WIELKOŚCI ROSZCZEŃ I INNYCH CZYNNIKÓW

1. Wstęp

W opracowaniu dotyczącym relacji między wielkością roszczenia a oszacowaniem składki podstawowe znaczenie ma analiza rozkładu prawdopodobieństwa łącznej wartości roszczeń, mająca oczywiście bezpośredni wpływ na ustalanie składki netto dotyczącej danej grupy ryzyka.

Początkowe rozważania będą dotyczyć zależności między wartością składki ζ, traktowanej w modelu jako wielkość zmieniająca się losowo w sposób ciągły, dotyczącej danej polisy ubezpieczeniowej a wielkością roszczenia Y, które może „pojawić się" w danej polisie. Przyjmujemy, że zachodzi następujący związek liniowy

$$\zeta = h \cdot Y + \nu, \tag{1}$$

gdzie h jest ustaloną liczbą, ν – zmienną losową charakteryzującą wpływ różnych czynników na poziom składki i dokładność operacji ustalania poziomu składki na podstawie zależności (1).

Zakładać będziemy, że zmienna losowa ν nie zależy od wielkości roszczenia Y. Wielkość Y może reprezentować informacje historyczne, a zmienna ν – dane teraźniejsze.

Celem jest dokonanie analizy wpływu dokładności szacowania roszczeń na dokładność składki ubezpieczeniowej. Pierwsze podejście do tej analizy przeprowadzimy, zakładając, że zmienne ζ i Y mają rozkłady normalne. Jest oczywiste, że składkę netto P będziemy szacować, przyjmując poziom średni $\bar{\zeta}$ zmiennej losowej z zależności (1). Równość (1) oznacza, że wielkość roszczeń w danym momencie bezpośrednio wpływa na losowy poziom składki ζ w tym samym momencie lub następującym tuż po nim.

Uzasadnienie powyższego sformułowania modelowego dotyczącego szacowania składki netto P wynika stąd, że ryzyko szacowanej składki ma dwa źródła:

- średni poziom $\bar{\zeta}$, przyjęty jako poziom składki netto, może być inny od wielkości roszczeń zrealizowanych w okresie późniejszym,
- poziom $\bar{\zeta}$ może być różny, i tak na pewno jest, od nieznanej wartości $E(Y)$; jest to tym bardziej realne, że zakładamy, że $E(\nu) = 0$.

Istnienie przypadkowych zmian wielkości ν powoduje, że przy jednym i tym samym wymiarze roszczeń Y wysokość składki ulega pewnym przypadkowym zmianom. Mamy tu więc zależność stochastyczną między wielkościami ζ a Y.

Ponieważ między zmiennymi ζ i Y istnieje zależność stochastyczna i każda z nich ma rozkład normalny, więc będziemy mówić, że istnieje między nimi **korelacja normalna**.

Wariancję $D^2(\zeta)$ składki, jako miary ryzyka jej szacowania, można obliczyć bezpośrednio z równości (1):

$$D^2(\zeta) = h^2 \cdot D^2(Y) + D^2(\nu). \tag{2}$$

Jeśli znane są wielkości $D^2(Y)$, $D^2(\nu)$, h, to ocena wpływu dokładności oceny roszczeń na dokładność składki netto P nie nastręcza trudności.

W celu uściślenia dokładności składki P oszacowanej na poziomie

$$E(\zeta) = h \cdot E(Y) + E(\nu) \tag{3}$$

załóżmy obecnie, że w procesie szkód scharakteryzowanym tak samo jak wcześniej, wielkości h i $D^2(\nu)$ nie zmieniają się. Jeśli zmniejszeniu ulega jedynie wariancja roszczenia Y to wariancja $D^2(\zeta)$ z równości (2) także się zmniejszy. Widzimy zatem, że przy takim podejściu do procesu roszczeń, zmianie ulega jedynie pierwszy składnik ze wzoru (2). Będzie tak jednak tylko w pewnych granicach obserwowania wielkości roszczeń. Jest przy tym oczywiste, że uściślanie wariancji składki na podstawie tego samego rozkładu roszczeń Y nigdy nie będzie mniejsze od wariancji składnika przypadkowego ν wpływającego na losowość składki modelowanej w zależności regresyjnej (1).

W powyższych rozważaniach założyliśmy, że dane są wielkości $D^2(Y)$, $D^2(\nu)$ i h. Zauważmy tu, że wariancja $D^2(Y)$ może być określona na podstawie przeszłych danych dotyczących roszczeń lub, jak to przyjęliśmy w kolejnych etapach, na podstawie równości (1) określającej dokładność (mierzoną wariancją) wyznaczania składki, przy założeniu niezmienności warunków (stałe h i $D^2(\nu)$) procesu roszczeń. Sytuacja taka jest możliwa do przyjęcia jedynie w bardzo małym przedziale czasu Δt, chociaż pojawiają się tu pewne wątpliwości dotyczące przyjmowania zmiennego poziomu przez wariancję $D^2(Y)$, co wpływa na szacowanie wariancji $D^2(\zeta)$ w kolejnych iteracjach oznaczanej przez $D^2(\zeta_i)$. Niezależnie od tego (gdyż możemy tu po prostu arbitralnie założyć, że kolejne roszczenia mają właśnie taką zmniejszającą się wariancję równą $D^2(\zeta_i)$) należy stwierdzić, że znacznie trudniej jest określić wariancję $D^2(\nu)$ oraz h.

2. Wyznaczenie regresji składki względem wielkości roszczeń

Przeprowadzimy obecnie rozważania, analogiczne do znanych rozstrzygnięć w statystyce i rachunku prawdopodobieństwa dotyczących linii regresji.

Uwzględniając znaną równość $D^2(\nu) = E(\nu^2) - [E(\nu)]^2$, składkę P obliczoną na podstawie zasady

$$P = E(Y) + \beta \cdot \sigma(Y),$$

możemy zapisać w postaci:

$$P = E(\zeta) = h \cdot E(Y) + \sqrt{E(\nu^2) - D^2(\nu)}. \tag{4}$$

Otrzymana postać, podobnie jak te ze znanych zasad szacowania składki, uwzględnia zmienność wielkości roszczeń Y wyrażoną tu przez czynnik przypadkowy ν.

W celu ustalenia zależności między wielkościami roszczeń Y a poziomami ζ składki badacz musi dokonać obserwacji, w których wielkości roszczeń odpowiada poziom składki ubezpieczeniowej (dane z przeszłości). Uzyskane pary liczb (Y, ζ) charakteryzują zarówno wielkości poprzednich, jak i następnych operacji szacowania składki netto P na podstawie danej zasady. Możliwe też jest (chociaż nie jest to zadanie proste ze względu na koszty) obserwowanie wielkości roszczeń i zarazem szacowanie składki netto na podstawie nowych danych, bez względu na to, na jakim poziomie była oszacowana składka netto dla celów praktyki ubezpieczeniowej.

Z formalnego punktu widzenia pary liczb (Y, ζ) można rozpatrywać jako realizacje dwuwymiarowej zmiennej losowej, którą w celu uproszczenia zapisu będziemy oznaczać tak samo jak jej realizacje. Z punktu widzenia ilościowego istnieje pewna zależność między operacyjnymi wartościami (w poszczególnych okresach) zmiennych Y i ζ. Tę ilościową zależność można scharakteryzować za pomocą współczynnika korelacji i równania korelacyjnego.

W analizie tych zależności istotne znaczenie ma zbadanie własności zmiennych losowych Y i ζ. Wysokość roszczenia przy dokonanych założeniach (lub bez nich) możemy przedstawić jako sumę większej liczby niezależnych zmiennych losowych – roszczeń z indywidualnych polis (lub jednej polisy w ustalonym okresie czasu), z których każda składa się na łączne roszczenie w portfelu lub danej polisie

$$Y = Y_1 + \ldots + Y_m.$$

Zmienna losowa ν ze wzoru (1) również jest sumą kilku, np. m zmiennych losowych, z których każda jest sumą błędów uwzględniających „niepewność" w obliczaniu składek w operacji $Y \to \zeta$. Suma ta wyraża część ryzyka związanego z szacowaniem składki i przybiera postać

$$\nu = \nu_1 + \nu_2 + \ldots + \nu_m.$$

Zgodnie z wcześniejszymi założeniami roszczenia Y_i są niezależne od składowych rodzajów ryzyka ν_j.

Zatem z równości (1) otrzymujemy, że zmienna ζ z modelu składki losowej wyraża się wzorem

$$\zeta = h \cdot Y_1 + h \cdot Y_2 + \ldots h \cdot Y_m + \nu_1 + \ldots + \nu_m,$$

a zatem także jest sumą niezależnych zmiennych losowych.

W stochastycznej analizie zależności między dwiema zmiennymi losowymi, z których każdą można zapisać w postaci sumy zmiennych losowych, można zastosować twierdzenie S.N. Bernsteina.

Niech zmienne losowe

$$Z_1 = U_1 + U_2 + \ldots U_n, \text{ i } Z_2 = \zeta_1 + \zeta_2 + \ldots \zeta_n$$

będą sumami bardzo dużej liczby składników spełniających założenia twierdzenia Lapunowa, mianowicie

$$\lim_{k \to \infty} \frac{\sum_{i=1}^{k} E\left[X_i - E(X_i)\right]^3}{\left[D^2(Z)\right]^{\frac{3}{2}}} = 0,$$

gdzie $Z = X_1 + \ldots + X_k$.

Wtedy (teza twierdzenia Lapunowa):

$$\lim_{k \to \infty} F_k(x) = \Phi\left(\frac{Z - EZ}{D^2 Z}\right),$$

a zatem suma Z asymptotycznie ma rozkład normalny.

Dla regresji składki jako wartości oczekiwanej warunkowej zmiennej ζ wartości roszczeń Y mamy:

$$\bar{\zeta}(Y) = a_2 + \rho \frac{\sigma_\zeta}{\sigma_Y}(Y - a_1), \tag{5}$$

gdzie $a_1 = E(Y)$, $a_2 = E(\zeta) = h \cdot E(Y) + E(\nu)$.

Korzystając z postaci współczynnika ρ, mamy

$$\bar{\zeta}(Y) = h \cdot Y + E(\nu). \tag{6}$$

Zatem przeciętna składka $\bar{\zeta}(Y)$ rośnie wraz z wartością roszczenia, przy czym wzrost ten jest zintensyfikowany przez wartość h.

Równanie (6) można otrzymać oczywiście wprost, gdy założymy, że zmienna Y jest ustalona (w próbie). Wtedy z równości (1) otrzymujemy:

$$E(\zeta/Y) = h \cdot Y + E(\nu) = \bar{\zeta}(Y).$$

Uzyskaliśmy zatem uogólniający schemat zależności występującej w klasycznym liniowym modelowaniu ekonometrycznym.

Zakładając niezmienności wielkości roszczeń Y dla pewnego okresu, mamy składkę netto:

$$P = h \cdot Y + E(\nu),$$

gdzie

$$h = \rho \cdot \frac{\sigma_\zeta}{\sigma_Y}. \tag{7}$$

składnik ν, i odpowiednio $E(\nu)$, może być interpretowany jako dodatkowe roszczenie mogące się pojawić przy nietypowym, realizowanym w losowy sposób procesie szkód.

Interpretacje

Równanie linii regresji nie daje nam niestety możliwości odpowiedzi na pytanie, o ile zwiększy się dokładność oszacowanej składki, jeżeli zwiększy się dokładność prognozy dotyczącej wielkości roszczeń. Możemy jednak na to pytanie odpowiedzieć, jeżeli na podstawie danych z przeszłości o dwuwymiarowej zmiennej losowej (Y, ζ) obliczymy wielkości ρ, $D^2(\zeta)$, $D^2(Y)$.

Kolejność postępowania jest wtedy następująca. Z wzoru $h = \rho \cdot \frac{\sigma_\zeta}{\sigma_Y}$ możemy obliczyć stałą h. Z kolei wariancję możemy obliczyć jako różnicę

$$D^2(\nu) = D^2(\zeta) - h^2 \cdot D^2(Y). \tag{8}$$

Zakładając, że zmienna losowa ν nie zależy od Y, możemy obliczyć wariancję $D^2(\nu)$, która nie zależy od wartości $D^2(Y)$:

$$\begin{aligned} D^2(\nu) = D^2(\zeta) - \rho^2 \frac{\sigma_\zeta^2}{\sigma_Y^2} D^2(Y) = D^2(\zeta) - \rho^2 \sigma_\zeta^2 = \\ = D(\zeta)\left(1-\rho^2\right). \end{aligned} \tag{9}$$

Dlatego, znając parametry ρ, $D^2(\zeta)$, $D^2(Y)$, można określić h i $D^2(\nu)$ występujące we wzorze (2), a mając $D^2(Y)$ – określić wpływ dokładności oszacowania wielkości roszczeń na dokładność oszacowanej składki.

Przykład 1

Na podstawie danych i hipotetycznych obliczeń stwierdzono, że wariancja wielkości roszczeń wynosi $D^2(Y) = 1000$, wariancja składki losowej $D^2(\zeta) = 100$, współczynnik korelacji $\rho = 0{,}3$ między wielkością roszczeń a składką „ciągłą” w takim znaczeniu, jak określiliśmy to na wstępie, czyli jako pewnej wartości traktowanej jako hipotetyczna wielkość losowa.

Określić o ile zwiększy się dokładność oszacowania składki, jeżeli wariancja roszczeń oszacowana będzie na podstawie dodatkowych danych o roszczeniach i przyjmie wartość

$$D^2(Y) = 500.$$

Na podstawie wzoru (7) otrzymujemy

$$h^2 = (0{,}3)^2 \cdot \frac{100}{1000} = 0{,}009.$$

Korzystając z wzoru (9), odpowiadającemu ryzyku szacowania składki lub dodatkowemu roszczeniu, mamy

$$D^2(\nu) = 100\,(1 - 0{,}09) = 100 - 0{,}0911 = 91{,}0$$

lub inaczej

$$D^2(\nu) = 100 - 0{,}09 \cdot 1000 = 100 - 9 = 91.$$

Zgodnie z równaniem (2), równanie zależności między $D^2(\zeta)$ i $D^2(Y)$ przybiera postać

$$D^2(\zeta) = 0{,}009 \cdot D^2(Y) + 91.$$

Zatem, jeśli $D^2(Y) = 500$, to

$$D^2(\zeta) = 95{,}5,$$

co znaczy, że wariancja szacowania składki jest nieco mniejsza od pierwotnie wyznaczonej.

Wniosek 1

Zmniejszenie wariancji roszczeń wpływa na zmniejszenie ryzyka szacowania składki.

Zauważmy jeszcze, że wyraźna poprawa dokładności oszacowania składki nastąpi wtedy, gdy zwiększy się stopień skorelowania między wielkościami ζ i Y.

Jeśli załóżmy bowiem, że $\rho = 0{,}9$, to wtedy $h^2 = 0{,}081$ i $D^2(\nu) = 19$. Następnie obliczamy wariancję $D^2(\zeta)$ składki losowej

$$D^2(\zeta) = 0{,}081 \cdot D^2(Y) + 19,$$

co przy $D^2(Y) = 500$, daje nam oszacowanie dokładności szacowanej składki na poziomie

$$D^2(\zeta) = 59{,}5.$$

Zatem widoczne jest istotne wystąpienie polepszenie dokładności oszacowania.

Wniosek 2

Przy silnym skorelowaniu ζ i Y ($\rho \approx 1$) można twierdzić, że zwiększenie dokładności w szacowaniu roszczeń prowadzi do istotnego zwiększenia dokładności oszacowania składki, natomiast przy słabym skorelowaniu zwiększanie dokładności szacowania wielkości roszczeń (w porównaniu z istniejącą standardową normą tego oszacowania) nie wpływa istotnie na dokładność oszacowania składki.

Trzeba podkreślić, że zmniejszenie dokładności szacowania roszczeń, przy danym współczynniku korelacji między ζ i Y, spowoduje znaczne zaniżenie dokładności szacowanej składki, czyli nie jest wskazane przy wyznaczonej korelacji między ζ i Y, szacując składki, w kolejnych etapach, opierać się na danych bardzo szacunkowych względem wielkości roszczeń. Ma to istotne znaczenie wtedy, gdy współczynnik ρ uzyskano z dużego agregatu, a następnie wprowadzono go w nielicznym portfelu podobnych rodzajów ryzyka o niezbyt długiej historii lub w odmiennych warunkach przejawiania się procesu szkód.

Uwaga 1

Odnosząc się do równania (5), tzn. wartości warunkowej $\bar{\zeta}(Y)$ oczekiwanej składki względem Y, należy zauważyć, że jeżeli zmienne ζ i Y są nieskorelowane, tzn. jeżeli $\rho = 0$, to zmienne ζ i Y są niezależne. Oczywiste jest również, co wynika z własności rozkładu normalnego, że jeżeli przynajmniej jedna ze zmiennych ζ i Y nie ma rozkładu normalnego, to zerowa wartość współczynnika korelacji nie zawsze jest równoznaczna z niezależnością tych zmiennych.

Uwagi uzasadniające przyjęte rozstrzygnięcia modelowe

Istotnie ograniczającym założeniem w powyższych rozważaniach jest założenie normalności odpowiednich zmiennych, chociaż z twierdzenia Lapunowa uzyskujemy uzasadnienie dla tego założenia mające praktyczne podstawy w liczbie obserwowanych roszczeń. Brakuje jednak wyraźnego wyjaśnienia przyczyn powodujących powstanie zależności typu (1). Zwrócić należy także uwagę na traktowanie w równaniu (1) poziomu ζ składki jako zmiennej losowej. Wynika to, z jednej strony, ze względów formalnych, a z drugiej – z autentyczności tego założenia jako procesu ciągłego, który w rzeczywistości się upraszcza. Nie można przecież składki opłacać w sposób ciągły, korygując ją. Ustala się ją na pewien okres, korygując ewentualne rozbieżności między jej faktyczną wartością a wartością ustaloną, co wyraża dodany w modelu składnik ν mający niezerową wartość przeciętną.

3. Związek między rozstępem w poziomie roszczeń a rozstępem w poziomie składki

Założymy tu, co jest zresztą w pełni uzasadnione, istnienie rozstępu między obserwowanymi wartościami wielkości roszczeń. Oznaczymy go jako

$$\zeta' = Y_{max} - Y_{min}.$$

Wielkość ζ' jest zmienną losową, która w dalszej perspektywie może być analizowana z punktu widzenia jej rozkładu, gdy będzie to przedmiotem badań przez zmienne

$$Y_{max} = \max \{Y_1,...,Y_n\},$$

$$Y_{min} = \min \{Y_1,...,Y_n\}.$$

W tym miejscu pominiemy ten aspekt modelowych założeń. W podobny sposób spostrzegamy składkę losową. Spowodowane jest to sztywnością założeń dotyczących systemu warunkującego ryzyko szacowania składki: postaci analitycznej związku ζ i Y, postaci współczynnika regulującego h i roszczenia.

Celem obecnych rozważań będzie zbadanie związku między rozstępem roszczeń a rozstępem ζ składki losowej.

W rozpatrywanym przypadku przejście od rozstępu roszczeń do rozstępu składki losowej jest związane z występowaniem pewnych elementów ryzyka ubezpieczeniowego, jakim jest szacowanie składki, uwzględnionych i występujących w procesie szkód ubezpieczeniowych w sposób stały, które nazwiemy czynnikami wpływu zależnymi od stopnia wspomnianej sztywności założeń i od siły wpływu postaci analitycznej na szacowaną składkę ubezpieczeniową.

Siłę wpływu oznaczymy przez S, a „sztywność" założeń scharakteryzujemy pewną stałą γ. Wtedy czynniki wpływu oznaczymy przez Δ i zapiszemy multiplikatywnie w postaci

$$\Delta = \gamma \cdot S.$$

Wprowadzimy obecnie pojęcie intensywności wpływu związanego z siłą wpływu S. Przyjmiemy, dla jasności wprowadzonego pojęcia, że siła wpływu S zależy od wielkości W-tego wpływu mierzonego w jednostkach roszczeń; może to być spowodowane możliwością ustalania składki na wyższym lub niższym poziomie, co jest uzależnione od bieżącej polityki zakładu ubezpieczeń i oczywiście od kondycji finansowej zakładu ubezpieczeń.

Można to zapisać następująco

$$S = C \cdot W^{\omega},$$

gdzie C i ω są pewnymi stałymi niezależnymi od W. Wielkość W siły wpływu będzie największa wtedy, gdy wielkość roszczeń jest największa, a najmniejsza wówczas, gdy wielkość roszczenia jest najmniejsza.

W przybliżeniu tę wielkość wpływu zapiszemy w następującej postaci:

$$W = \frac{1}{2}(Y - a),$$

gdzie: Y – wielkość roszczenia, a – ustalona norma dla kształtowanego poziomu stawki, ustalona np. w wyniku różnego typu ekspertyz lub przewidywanej wielkości roszczeń.

W każdym konkretnym przypadku dla roszczenia Y wielkość wpływu W będzie się zmieniała od postaci

$$W_{min} = \frac{1}{2}(Y_{\min} - a)$$

do postaci

$$W_{max} = \frac{1}{2}(Y_{\max} - a).$$

Powoduje to oczywiście, że czynniki wpływu, wyrażone w wielkości Δ, w procesie ustalania składki względem tego samego roszczenia będą się zmieniały od wartości $\Delta_{\max}$ do wartości $\Delta_{\min}$,
gdzie

$$\Delta_{\max} = \gamma \cdot S_{max} = \gamma \cdot C \cdot W^{w}_{\max} = \gamma \cdot \mathrm{C} \cdot \left[\frac{1}{2}(Y_{\max} - a)\right]^{w}$$

$$\Delta_{\min} = \gamma \cdot \mathrm{S}_{\max} = \gamma \cdot \mathrm{C} \cdot W^{w}_{\min} = \gamma \cdot \mathrm{C} \cdot \left[\frac{1}{2}(Y_{\min} - a)\right]^{w}. \tag{10}$$

Zauważyć należy, że wielkość Δ czynników wpływu ma bezpośrednie przełożenie na poziom składki (losowy). Mianowicie, jeśli wielkość wpływu wynosi Δ, to losowa wartość składki zwiększa się o $2 \cdot \Delta$. Wynika to z przyjętych w tym modelowym schemacie omawianej zależności czynników wpływu Δ, powodujących „rozchwianie" rzeczywistego procesu obserwowanego w realnie ustalanej składce. Zostanie tu wprowadzone założenie mające na celu zwiększenie bezpieczeństwa lub zmniejszenie ryzyka szacowania składki.

Dlatego też różnica

$$2\Delta_{max} - 2\Delta_{min}$$

określa rozstęp składki, którą otrzymuje się w przypadku, gdy rozstęp roszczeń wynosił ζ'.

Z (10) wynika, że

$$2(\Delta_{max} - \Delta_{min}) = \frac{1}{2} \cdot \gamma \cdot c \left\{ \left[\frac{1}{2}(Y_{max} - a) \right]^{\omega} - \left[\frac{1}{2}(Y_{min} - a) \right]^{\omega} \right\}.$$

Jeżeli przyjmiemy, że wykładnik $\omega = 1$, to

$$2(\Delta_{max} - \Delta_{min}) = \gamma \cdot c \cdot (Y_{max} - Y_{min}) = \gamma \cdot c \cdot \zeta'.$$

Z modelowych uwarunkowań wynika także, że na skutek wpływu niepewności w procesie ustalania składki powstaje dodatkowy rozstęp ν, który nie zależy od ζ'. Dlatego możemy zapisać, że rozstęp ustalonej składki jest równy

$$\zeta = 2(\Delta_{max} - \Delta_{min}) + \nu,$$

a dla $\omega = 1$:

$$\zeta = \gamma \cdot c \cdot \zeta' + \nu.$$

Wprowadzając oznaczenie

$$h = \gamma \cdot c,$$

otrzymamy zależność

$$\zeta = h \cdot \zeta' + \nu, \tag{11}$$

czyli otrzymamy wzór analogiczny do wzoru (1). Jednak w odróżnieniu od modelowego zagadnienia, w którym podstawą był wzór (1), stochastycznie zależne wielkości ζ i ζ' nie muszą tu mieć założonego z góry rozkładu normalnego.

Zależność liniowa (11) między wielkościami ζ i ζ' wynika z założenia o liniowym związku między wielkością czynników wpływu Δ na składkę losową i między intensywnością wpływu S a wielkością wpływu wynikającą z rozmiaru roszczeń W. Jeżeli nie zmienimy dwóch ostatnich założeń, to przy analizowaniu zależności między wymiarem składki ζ a wysokością roszczeń Y otrzymamy równość (1).

4. Uogólnienie założeń modelowych szacowania składki względem roszczeń. Zastosowanie pojęcia rozstępu przy szacowaniu składki

Przyjmiemy obecnie silne, ale uzasadnione założenie, że wielkość h, intensyfikująca wpływ losowych roszczeń na losowość składki ze wzoru (11), jest wielkością losową. Takie założenie jest uzasadnione, ponieważ intensywność wpływu S, a więc i wielkość c, zależą od charakterystyk rozkładu roszczeń uwarunkowanych przez proces szkód, procedury szacowania szkód i innych różnorodnych czynników. Zakładając choćby tylko to, że h zależy od uwarunkowań procesu roszczeń, ich jakościowej formy, mamy już uzasadnienie dla przyjęcia założenia o losowości współczynnika h.

Analityk opracowujący związek między wielkościami ζ i ζ', podobnie jak w poprzednim układzie założeń, może badać dwuwymiarową zmienną losową (ζ, ζ') i na podstawie obserwacji próbować wyjaśnić wpływ dokładności oszacowania roszczeń, czyli ryzyka szacowania roszczeń, tj. jego miary, którą jest wielkość $D^2(\zeta')$, na ryzyko szacowania (dokładność) składki, tj. na $D^2(\zeta)$. Odmienność występujących w tym przypadku uwarunkowań formalnych, w porównaniu z tą z części 1, jest teraz oczywista.

Obliczymy, korzystając z modelu (11), wariancję rozstępu $D^2(\zeta)$ składki. Ponieważ wielkości ν i $h \cdot \zeta'$ są niezależne, więc

$$D^2(\zeta) = D^2(h \cdot \zeta') + D^2(\nu).$$

Stąd, korzystając z zależności

$$D^2(h \cdot \zeta') = D^2(h^2 \cdot \zeta'^2) - E^2(h \cdot \zeta')$$

i twierdzenia o wartości oczekiwanej iloczynu niezależnych zmiennych losowych i niezależności h i ζ', otrzymamy, że

$$E(h^2 \cdot \zeta'^2) = E(h^2) \cdot E(\zeta'^2), \quad E(h \cdot \zeta') = E(h) \cdot E(\zeta').$$

Otrzymamy więc wzór

$$D^2(h \cdot \zeta') = E(h^2) \cdot E(\zeta'^2) - E^2(h) \cdot E^2(\zeta') = D^2(h) \cdot D^2(\zeta') + E^2(\zeta')\, D^2(h) + E^2(h^2) \cdot D^2(\zeta'^2).$$

Dla sprawdzenia jego prawdziwości wykonamy poniższe przekształcenia

$$(D^2(h) + E^2(h)) \cdot (D^2(\zeta') + E^2(\zeta')) - E^2(h) \cdot E^2(\zeta') =$$
$$= D^2(h) \cdot D^2(\zeta') + D^2(h) \cdot E^2(\zeta') + E^2(h) \cdot D^2(\zeta') - E^2(h) \cdot E^2(\zeta') +$$
$$+ E^2(h) \cdot E^2(\zeta') = D^2(h) \cdot D^2(\zeta') + D^2(h) \cdot E^2(\zeta') + E^2(h) \cdot D^2(\zeta').$$

Dlatego

$$D^2(\zeta) = D^2(h) \cdot D^2(\zeta') + D^2(h) \cdot E^2(\zeta') + E^2(h) \cdot D^2(\zeta') + D^2(\nu). \tag{12}$$

Ze wzoru (12) wynika, że przy wyznaczaniu wariancji $D^2(\zeta)$ należy znać wariancje $D^2(h)$, $D^2(\zeta')$ oraz wartości oczekiwane $E(h)$ i $E^2(\zeta')$. Należy tu zauważyć, że oprócz $D^2(\zeta')$ i $E^2(\zeta')$ pozostałe parametry są dość trudne do wyznaczenia na podstawie badania wartości zmiennej losowej (ζ, ζ'). Można je jednak uzyskać w sposób niebezpośredni, korzystając z krzywych regresji zmiennej ζ względem ζ'.

Równanie korelacyjne (5), które otrzymaliśmy przy założeniu normalnej korelacji, może być wykorzystane jako ogólne równanie korelacyjne przy istnieniu zależności liniowej między odpowiednimi zmiennymi. Można się o tym przekonać, konstruując metodę szacowania rozstępu.

Przedstawimy ten sposób postępowania w etapach.

Etap I

Obliczymy współczynnik korelacji między ζ i ζ'. Z definicji

$$\mathrm{cov}\,(\zeta, \zeta') = E(\zeta \cdot \zeta') - E(\zeta') \cdot E(\zeta).$$

Uwzględnimy w tym miejscu następującą zależność wynikającą z równania (11):

$$E(\zeta \cdot \zeta') = E(h \cdot \zeta'^2 + \zeta' \cdot \nu) = E(h) \cdot E(\zeta'^2) + E(\zeta') \cdot E(\nu) =$$
$$= E(h) \cdot (D^2(\zeta') + E^2(\zeta') + E(\zeta') \cdot E(\nu)) =$$
$$= E(h) \cdot D^2(\zeta') + E(h) \cdot E^2(\zeta') + E(\zeta') \cdot E(\nu)$$

oraz

$$E(\zeta) = E(h \cdot \zeta' + \nu) = E(h) \cdot E^2(\zeta') + E(\nu).$$

Stąd

$$\operatorname{cov}(\zeta, \zeta') = E(h) \cdot D^2(\zeta') + E(h) \cdot E^2(\zeta') + E(\zeta') \cdot E(\nu) - E(h) \cdot E^2(\zeta') -$$
$$- E(\zeta') \cdot E(\nu) = E(h) \cdot D^2(\zeta').$$

Zatem współczynnik korelacji wyraża się wzorem

$$\rho = \frac{\operatorname{cov}(\zeta', \zeta)}{\sigma_{\zeta'} \sigma_{\zeta}} = E(h) \cdot \frac{\sigma_{\zeta'}}{\sigma_{\zeta}}. \tag{13}$$

Jeżeli przy obliczaniu współczynnika korelacji między ζ i ζ' skorzystamy z równania korelacyjnego (5), to otrzymamy zależność

$$\bar{\zeta}(\zeta') = a_2 + \rho \frac{\sigma_{\zeta}}{\sigma_{\zeta'}} (\zeta' - a_1), \tag{14}$$

gdzie

$$a_1 = E(\zeta'),$$
$$a_2 = E(\zeta) = E(h) \cdot E(\zeta') + E(\nu).$$

Podstawiając do równania korelacyjnego wyrażenia określające współczynnik ρ i a_1, a_2, otrzymamy

$$\bar{\zeta}(\zeta') = E(h) \cdot E(\zeta') + E(\nu) + E(h) \cdot [\zeta' - E(\zeta')] =$$
$$= E(h) \cdot \zeta' + E(\nu). \tag{15}$$

Taką samą zależność można otrzymać, jeżeli obliczenia wykonamy, wychodząc z równości (11). Zatem przyjmując za podstawę wzór (15), aby otrzymać krzywą regresji, wystarczy znać wartości parametry rozkładu $E(\zeta')$, $E(\zeta)$, ρ, $D^2(\zeta')$, $D^2(\zeta)$, które zawsze mogą być znalezione na podstawie danych dotyczących (ζ', ζ).

Etap II

Wariancja zmiennej $\bar{\zeta}(\zeta')$

Pamiętając o celu obecnych rozważań, a mianowicie o otrzymaniu oszacowania dla $D^2(h)$ i $D^2(\nu)$, zauważmy najpierw, że sama krzywa regresji nie określa dokładnie zależności między $D^2(\zeta)$ i $D^2(\zeta')$.

Wprowadzimy obecnie pewne dodatkowe pojęcia. Wielkość $\bar{\zeta}(\zeta')$, zgodnie z (14), jest funkcją zmiennej losowej ζ, czyli wielkości roszczeń. Wielkość tę możemy, a nawet powinniśmy, traktować jako zmienną losową. Należy zatem obliczyć jej wariancję.

Z (15) wynika, że

$$D^2(\bar{\zeta}) = D^2[E(h)\cdot\zeta' + E(\nu)] = E^2(h)\cdot D^2(\zeta') + 0. \qquad (16)$$

Ponieważ dla $\bar{\zeta}(\zeta')$ przyjęliśmy nazwę „warunkowa wartość oczekiwana", naturalne jest, że wariancję $D^2\left(\bar{\zeta}(\zeta')\right)$ trzeba nazwać **wariancją średnich warunkowych**.

Jeżeli założymy, że ζ' jest ustaloną obserwacją wielkości roszczenia, a celem jest obliczenie na podstawie wzoru (10) wariancji zmiennej ζ, to w wyniku tego otrzymamy wariancję jej rozkładu warunkowego. Korzystając z oznaczeń wprowadzonych przy wyprowadzeniu wzoru (14), możemy zapisać, że

$$D^2(\zeta/\zeta') = \zeta'^2\cdot D^2(h) + D^2(\nu)$$

i $D^2(\zeta/\zeta')$ będziemy nazywać **wariancją warunkową**.

Wariancja warunkowa jest funkcją zmiennej losowej ζ' i dlatego ona sama także jest zmienną losową. Obliczając jej wartość oczekiwaną, otrzymamy

$$E[D^2(\zeta/\zeta')] = E[\zeta'^2\cdot D^2(h)] + D^2(\nu) = D^2(\zeta')\cdot D^2(h) + E^2(\zeta)\cdot E^2(D(h)) + D(\nu) =$$
$$= D^2(\zeta')\cdot D^2(h) + E^2(\zeta)\cdot D^2(h) + D(\nu). \qquad (17)$$

Z drugiej strony zauważmy, że

$$D^2(\bar{\zeta}) + E[D^2(\zeta/\zeta')] = E^2(h)\cdot D^2(\zeta') + D^2(\zeta')\cdot D^2(h) +$$
$$+E^2(\zeta)\cdot D^2(h) + D(\nu) = D^2(\zeta). \qquad (18)$$

Ostatnie składniki równości (18) są wynikiem następującego, bardziej ogólnego twierdzenia [Feller 1996]:

Jeśli Z_2 jest zmienną losową zależną od Z_1, tak by średnia warunkowa $\bar{Z}_2/Z_1$ była funkcją ciągłą zmiennej losowej Z_1, to wariancja $D^2(Z_2)$ jest sumą wariancji średnich warunkowych i wartości oczekiwanej wariancji warunkowej.

$$D^2(Z_2) = D^2\left(\bar{Z}_2/Z_1\right) + E\left[D^2\left(Z_2/Z_1\right)\right].$$

Ponieważ $D^2(\bar{\zeta})$ i $E[D^2(\zeta/\zeta')]$ można obliczyć na podstawie danych z obserwacji, twierdzenie to ma praktyczne i istotne zastosowanie przy określaniu dokładności $D^2(\zeta)$ (ryzyka) szacowania składki $E(\zeta) = P$.

Interpretacje

Zgodnie ze wzorem (16) wariancja średnich warunkowych $D^2(\bar{\zeta})$ jest wprost proporcjonalna do wariancji $D^2(\zeta')$, tzn. wariancji rozstępu roszczeń. Dlatego zmniejszając $D^2(\zeta')$, np. trzykrotnie, zmniejszamy tym samym trzykrotnie wielkość $D^2(\bar{\zeta})$. Poza tym, jak wynika ze wzoru (17), zmniejsza się także wartość oczekiwana warunkowej wariancji, a więc wielkość liczbowa charakteryzująca przeciętnie losowe zmienności $D^2(\bar{\zeta})$. Na podstawie równości (17) łatwo stwierdzić, że zmniejsza się także i wartość oczekiwana warunkowej wariancji. Ponieważ z (17) wynika, że zależność między $E[D^2(\zeta/\zeta')]$ a $D^2(\bar{\zeta})$ i $E(\zeta')$ ma charakter bardzo złożony, trudno jest określić wielkość tych zmian. We wszystkich przypadkach można jednak twierdzić, że jeżeli dokładność oszacowań roszczeń będzie się zwiększać, to wielkość oczekiwana ryzyka $E[D^2(\zeta/\zeta')]$ będzie się zmniejszać, bo

$$E[D^2(\zeta/\zeta')] = D^2(\zeta) - D^2(\bar{\zeta}).$$

Przypuśćmy teraz, że w wyniku obserwacji dla pewnej dokładności oszacowania roszczeń $D_1^2(\zeta')$ otrzymano wielkości $D_1^2(\bar{\zeta})$, $E_1[D^2(\zeta/\zeta')]$. Tak więc

$$D_1^2(\zeta) = D_1^2(\bar{\zeta}) + E_1[D^2(\zeta/\zeta')] \tag{19}$$

jest wariancją rozrzutu wynegocjowanych i wypłaconych odszkodowań.

Zakłada się, że w wyniku dodatkowych obserwacji dokładność zwiększa się następnie do wielkości $D_2(\bar{\zeta})$, gdzie

$$D_2(\bar{\zeta}) = \varepsilon \cdot D_1^2(\bar{\zeta}) \quad (0 < \varepsilon < 1).$$

Należy tu wyjaśnić, jak się przekłada to na dokładność oszacowania składki czystej, przełożonej na wynegocjowane odszkodowania.

Można dla uproszczenia przyjąć, że ogólne uwarunkowania się nie zmieniają, więc wielkości

$$E(h), D^2(h), D^2(\nu)$$

także się nie zmieniają. Wtedy możemy twierdzić, że jeśli oszacowanie składki są przeprowadzane na podstawie roszczeń z polis – kontraktów o dokładności $D_2^2(\zeta')$, to wielkości $D_2^2(\bar{\zeta})$, $E_2[D^2(\zeta/\zeta')]$ będą określane na podstawie (17) następującymi wzorami:

$$D_2^2(\bar{\zeta}) = E^2(h) \cdot D_2^2(\zeta') = \varepsilon \cdot D_1^2(\bar{\zeta})$$

oraz – z (19):

$$E_2[D^2(\zeta/\zeta')] = \varepsilon \cdot D_1^2(\zeta') \cdot D^2(h) + E_2^2(\zeta') \cdot D^2(h) + D^2(\nu). \tag{20}$$

Zatem

$$D_2^2(\bar{\zeta}) = \varepsilon \cdot D_1^2(\bar{\zeta}), \tag{21}$$

co wynika z równości (17)

oraz

$$E_2[D^2(\zeta/\zeta')] \leq E_1[D^2(\zeta/\zeta')],$$

gdyż przy zmniejszeniu rozstępu spełniona jest nierówność

$$E_2(\zeta') \leq E_1(\zeta').$$

Wprowadzimy obecnie następujące oznaczenie:

$$D_2^{*2}(\zeta) = D_2^2(\bar{\zeta}) + E_1[D^2(\zeta/\zeta')].$$

Zatem $D_2^{*2}(\zeta)$ wyraża, jako suma, błąd średni przy starej informacji i błąd – wariancję średnich warunkowych z nowej informacji.

Oczywiście z pierwszej z równości (20) wynika, że

$$D_2^{*2}(\zeta) = \varepsilon \cdot D_1^2(\bar{\zeta}) + E_1[D^2(\zeta / \zeta')], \tag{22}$$

co na podstawie równości (21) oznacza, że

$$D_2^{*2}(\zeta) \geq D_2^2(\bar{\zeta}) + E_2[D^2(\zeta / \zeta')] = D_2^2(\zeta). \tag{23}$$

Zatem mając wielkość $D_2^{*2}(\zeta)$ (wzór (22)), możemy twierdzić, że wariancja wynegocjowanego odszkodowana w każdym przypadku nie będzie od niej większa.

Znak równości w (23) zachodzi tylko wówczas, gdy wartość oczekiwana $E_1[D^2(\zeta / \zeta')]$ nie zależy od $D^2(\zeta')$. Jak wynika z (17), taka niezależność ma miejsce tylko wtedy, gdy h jest wielkością ustaloną. W tym przypadku $D^2(h) = 0$, a

$$E_1[D^2(\zeta / \zeta')] = D^2(\nu).$$

Taki przypadek był rozpatrywany w pierwszej części artykułu. Jeżeli h = const, to zmiany dokładności oszacowań roszczeń nie powodują zmiany wartości oczekiwanej wariancji warunkowej, czyli obliczenie dokładności jest tu o wiele prostsze.

Jeżeli zaś h jest zmienną losową, to oszacowanie dokładności wynegocjowanego odszkodowania przy zmieniających się dokładnościach wielkości roszczeń można otrzymać tylko w postaci nierówności (23).

Przykład 2

Przyjmijmy, że dla ryzyka szacowania roszczeń $D_1^2(\zeta') = 10$ otrzymano, że $D_1^2(\zeta) = 6$, $D_1^2(\bar{\zeta}) = 4$, $E[D^2(\zeta / \zeta') = 2]$. Oszacować zmiany ryzyka szacowania składki, jeżeli ryzyko szacowania roszczeń zmniejszy się do $D_2^2(\zeta') = 7$.

Obliczymy najpierw wielkość ε:

$$\varepsilon = \frac{D_2^2(\zeta')}{D_1^2(\zeta)} = \frac{7}{10} = 0,7.$$

Z wzorów (22) i (23) otrzymujemy:

$$D_2^{*2}(\zeta) = 0,7 \cdot 4 + 2 = 4,8 \geq D_2^2(\zeta).$$

Wynik ten oznacza, że wariancja rozstępu $D_2^2(\zeta)$ szacowania składki nie będzie większa niż 4,8. Równość zaistnieje wtedy, gdy h będzie stałą, tzn. wtedy, gdy wahania wpływu różnych czynników na szacowanie roszczeń są małe w porównaniu z wartością średnią tych roszczeń.

Znaczenie współczynnika korelacji ρ jest, tak jak już wcześniej zauważyliśmy, takie samo, tzn. że dla małych wartości tego współczynnika zmniejszenie wartości $D^2(\zeta')$ powoduje niewielkie zmiany wartości $D^2(\zeta)$.

Przykład 2 kończy rozważania dotyczące zastosowania teorii korelacji przy szacowaniu ryzyka szacowania składki ubezpieczeniowej. Obejmowały one tylko liniowe zależności korelacyjne, i to nie wszystkie. Analiza zależności nieliniowych, które mogą w szczególnym przypadku powstać przy nieliniowym wpływie czynników na wielkości roszczeń, jest znacznie trudniejsza. W podejściu do rozpatrywanego zagadnienia od strony teorii korelacji można także uwzględnić korelację wieloraką, co możliwe jest wtedy, gdy na jakość oszacowań roszczeń, a zatem i składki netto, ma wpływ kilka cech.

Literatura

David A., *Controling Premia by Repackaging Asset-Backet Securities*, „The Journal of Risk Insurance" 1997, vol. 64, nr 4, s. 619-648.

Downs D.H., *Monitoring, Ownership, and Risk-Taking: the Impact of Guaranty Funds*, „The Journal of Risk Insurance" 1997, vol. 66, nr 3, s. 477-497.

Feller W., *Wstęp do rachunku prawdopodobieństwa*, PWN, Warszawa 1966.

Fisz M., *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN, Warszawa 1969.

PREMIUM IN THE ASPECT OF LINEAR DEPENDENCE CORRELATIVE CLAIMS' SIZE AND OTHER FACTOR

Summary

The theme in the article is to show, in its main dimension, that various approaches may by applied to the analysis of the issue concerning relation between estimated premium and distribution of insurance claims'probability.

The models presented in the article, in spite of some simplified scheme of its exemplification in the linear form, are far-reaching departure from the assumptions assumed, for instance, in econometric modeling.

The standpoint proposed in the article allows to define the estimation of premium P in the form of sum (total) of claims'size, assuming their invariability, measured with premium regression coefficient in relation to claims and other factors affecting the claim risk and the second ingredient which may be interpreted as an additional claim.

The far-reaching generalization was made as well, assuming that the quantity intensifying the influence of random claims on the random claim is random as well, it means that the constant from the models considered earlier is now a random variable.

The purpose of the considerations is also achievement of both estimation of variance of the ingredient additionally affecting (apart from claims) premium and of variance of this "randomnised" constant. The notion of the so-called variances of conditional averages. It is shown that variance of negotiated compensation, what implies a certain α-tipicality of the considered insurance procedures, will be in all cases not bigger then the sum of average error at the old information and the error expressed by the variance of conditional averages the new information.