PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 103
Prognozowanie w zarządzaniu firmą 2010

Alicja Wolny-Dominiak
Akademia Ekonomiczna w Katowicach

SZACOWANIE ZMIENNYCH TARYFIKACYJNYCH W UBEZPIECZENIACH MAJĄTKOWYCH Z ZASTOSOWANIEM MIESZANYCH MODELI LINIOWYCH

Streszczenie: Powszechnie stosowana w zakładach ubezpieczeń majątkowych metoda badania wpływu zmiennych taryfikacyjnych na poziom wartości odszkodowań wykorzystuje uogólnione modele liniowe (GLM). W modelach tych zakłada się, iż wszystkie zmienne mają stałe parametry (*fixed effect*). Jednak gdy do modelu wprowadza się zmienną taryfikacyjną charakteryzującą się dużą liczbą przyjmowanych wartości, których nie można uporządkować (np. zmienną taryfikacyjną „model pojazdu"), zastosowanie znajdują tzw. hierarchiczne uogólnione modele liniowe (HGLM), w których wprowadza się parametry losowe (*random effect*). W pracy przedstawione zostanie wykorzystanie modeli typu HGLM do budowy taryf w ubezpieczeniach majątkowych.

Słowa kluczowe: taryfy, liniowy model mieszany, hierarcziczny uogólniony model liniowy.

1. Wstęp

Istotnym problemem w budowie taryf w zakładach ubezpieczeń majątkowych jest badanie wpływu różnorodnych czynników opisujących przedmiot ubezpieczenia oraz osobę ubezpieczającą się na wysokość powstających szkód. W takiej analizie zastosowanie znajdują modele liniowe (LM), w których czynniki wpływu traktowane są jako tzw. zmienne taryfikacyjne. W klasycznym modelu liniowym przyjmowane jest założenie o normalności rozkładu zmiennej objaśnianej Y. W przypadku danych szkodowych najczęściej założenie to nie jest spełnione. Wynika to z charakteru powstawania szkód, dla których rozkładem wartości szkód najczęściej jest rozkład gama [Lee, Nelder, Pawitan 2006]. Dlatego też w procesie budowy taryf zastosowanie znajdują głównie uogólnione modele liniowe (GLM) rozszerzające podejście modeli liniowych następująco: rozkład zmiennej objaśnianej przyjmowany jest z dyspersyjnej rodziny rozkładów wykładniczych oraz μ nie jest wprost równa liniowej kombinacji parametrów i zmiennych objaśniających, a jest powiązana z tą kombinacją funkcją połączenia (*link function*). W przypadku stosowania GLM zakłada się, iż wszystkie wstępujące w modelu zmienne mają stały wpływ na wartość szkód (*fixed* *effects*) ze stałymi parametrami. Jednakże często przyjmuje się, iż część zmiennych taryfikacyjnych ma losowy wpływ na zmienną Y (*random effects*). Wtedy odpowiadajace im parametry mają charakter losowy i traktowane są jako zmienne losowe. W takiej sytuacji uzyskiwany jest mieszany model. Gdy przyjmowane jest założenie o normalności składnika losowego i parametrów losowych, wówczas model nazywany jest liniowym modelem mieszanym (LLM), natomiast po rozszerzeniu rozkładu składnika losowego na dyspersyjną rodzinę rozkładów wykładniczych mamy uogólniony liniowy model mieszany (GLMM). Często jednak w przypadku danych ubezpieczeniowych niezbędne jest również rozszerzenie założenia o rozkładzie parametrów losowych na rozkłady z dyspersyjnej rodziny wykładniczej. Taki model nazywany jest hierarchicznym uogólnionym modelem liniowym (HGLM).

W pracy przedstawiono wykorzystanie modeli mieszanych do budowy taryf w ubezpieczeniach majątkowych w sytuacji, gdy jedna ze zmiennych jest zmienną nominalną wielokategorialną. W celu zobrazowania działania modeli przeprowadzono proces modelowania i budowy taryf dla danych zaczerpniętych ze szwedzkiego rynku ubezpieczeniowego (niestety na rynku polskim nie było to możliwe). Wykorzystano w tym celu program komputerowy R.

2. Ogólna charakterystyka liniowych modeli mieszanych

Modele mieszane stosowane są w budowie taryf zwykle w sytuacji, gdy uzależnia się wartość szkód od zmiennej nominalnej wielokategorialnej. Przykładem takiej zmiennej jest np. region czy model samochodu w ubezpieczeniu komunikacyjnych. W przypadku uwzględniania tego typu zmiennej w modelu proces taryfikacji może być przeprowadzany dla każdej kategorii tej zmiennej osobno. Takie podejście ma niewątpliwie wadę związaną z czasem obliczeń, gdyż konieczne jest przeprowadzanie algorytmu taryfikacyjnego (stosując np. procedurę minimalnego obciążenia lub GLM) dla każdej kategorii osobno. W celu skrócenia czasu obliczeń możliwe jest wykorzystanie tzw. mieszanych modeli liniowych.

Przyjmijmy następujące oznaczenia: Y − zmienna objaśniana, w procesie taryfikacji: wartość szkód, X_t − zmienne objaśniające o stałym wpływie na zmienną objaśnianą, v_z − zmienne objaśniające wielokategorialne mające losowy wpływ na zmienną objaśnianą. Dla powyższych oznaczeń mieszany model liniowy ma postać ogólną [Lee, Nelder, Pawitan 2006]:

$$Y = X\beta + Zv + e,$$

gdzie: X jest macierzą współczynników dla zmiennych o stałym wpływie, β − wektorem parametrów stałych, Z − macierzą współczynników dla zmiennych o losowym wpływie, v − wektorem parametrów losowych.

W klasycznym modelu mieszanym LMM przyjmuje się założenie o rozkładach normalnych: $v \sim WN(0,\Sigma)$, $e \sim WN(0,D)$, gdzie macierze Σ, D są uzależnione od nieznanego parametru τ zwanego komponentem wariancji. W celu estymacji parametrów β oraz v w pracy [Lee, Nelder, Pawitan 2006] przedstawiono procedurę maksymalizującą tzw. połączoną funkcję wiarygodności (*joint likehood function*) o postaci:

$$\log f(y,v) = -\frac{1}{2}(y - X\beta - Zv)^T \Sigma^{-1}(y - X\beta - Zv) - \frac{1}{2}v^T D^{-1}v.$$

Widać, że gdy $D^{-1} = 0,$ parametr losowy v jest traktowany jak parametr stały. Inną metodą jest metoda GEE (*Generalized Estimating Equations*) przedstawiona w pracy [McCulloch, Searle 2001].

W przypadku modelowania danych ubezpieczeniowych zastosowanie znajduje przede wszystkim model HGLM. Model ten, w ogólnym przypadku, przedstawia się następująco [Lee, Nelder, Pawitan 2006]:

(i) Dla zmiennej objaśnianej Y zachodzi

$$E(y\,|\,u) = \mu \text{ oraz } \operatorname{var}(y\,|\,u) = \varphi V(\mu),$$

gdzie ϕ jest parametrem dyspersji, $V(\mu)$ jest funkcją wariancji z modelu GLM. Jądro log-wiarygodności oraz predyktor liniowy mają postać:

$$\sum\left(\frac{y\theta - b(\theta)}{\varphi}\right),$$

$$\eta = g(\mu) = X\beta + Zv,$$

gdzie: θ jest funkcją μ znaną jako parametr kanoniczny, $\phi -$ znanym parametrem dyspersji, $v = v(u)$ − wektorem parametrów losowych, β − wektorem parametrów stałych.

(ii) Czynnik losowy u ma rozkład pochodzący z dyspersyjnej rodziny rozkładów wykładniczych z parametrem λ.

W modelach HGLM metodą estymacji jest maksymalizacja tzw. funkcji *h*-wiarygodności. Klasyczna funkcja wiarygodności ma najczęściej postać:

$$L(\theta; y) = f_\theta(y),$$

gdzie f jest funkcją gęstości zmiennej Y z parametrem $\theta.$ W przypadku modeli mieszanych taka postać funkcji wiarygodności pozwala na szacowanie parametrów jedynie dla stałych czynników wpływu. Aby jednocześnie szacować parametry dla stałych i losowych czynników wpływu, konieczne jest rozszerzenie funkcji wiarygodności:

$$L(\theta, v; y, v) = L(\theta, v; y\,|\,v)L(\lambda, v),$$

gdzie $L(\theta, v; y \mid v)$ jest warunkową funkcją wiarygodności. Ta rozszerzona funkcja wiarygodności jest nazywana funkcją *h*-wiarygodności, jeżeli v jest kanoniczny, tzn. spełnia warunek:

$$\frac{L(\theta_1, \hat{v}_1; y \mid v)}{L(\theta_2, \hat{v}_2; y \mid v)} = 1.$$

Estymatory $\hat{v}_1, \hat{v}_2$ są estymatorami największej wiarygodności v dla arbitralnie przyjętych wartości parametrów θ_1 oraz θ_2. Przechodząc do estymacji stałych parametrów β oraz parametru losowego v, maksymalizowana jest funkcja logarytmiczna *h*-wiarygodności o postaci:

$$h = \log L(\beta, v; y \mid v) + \log L(v).$$

Algorytm procedury maksymalizacji dla dowolnych konfiguracji rozkładów losowego czynnika wpływu oraz składnika losowego jest złożony; jego dokładny opis przedstawiono w pracy [Lee, Nelder, Pawitan 2006]. Implementacja komputerowa tego algorytmu dla większości rozkładów zmiennych przeprowadzona została w programie statystycznym *GenStat*, gdzie począwszy od wersji 9.0, występuje specjalistyczny moduł *HGLM GenStat*. Jest to obecnie jedyne oprogramowanie, w którym występuje moduł estymacji parametrów w modelu HGLM. Podejmowane są również kroki w celu implementacji metody *h*-wiarygodności, co jest również przedmiotem prac autorki.

3. Zastosowanie modeli HGLM w procesie taryfikacji

Proces taryfikacji w ubezpieczeniach majątkowych można zapisać w postaci modelu HGLM, przyjmując następującą interpretację zmiennych: Y − wartość szkód; X_t − zmienne taryfikacyjne mające stały wpływ na wartości szkód; v_k − zmienne taryfikacyjne wielokategorialne mające losowy wpływ na wartości szkód. Najczęściej w procesie taryfikacji wprowadza się jedną zmienną o losowym wpływie na wartości szkód i różnicującą portfel. Zatem w modelu postaci:

$$Y = X\beta + Zv + e,$$

gdzie $\beta = [\beta_1, ..., \beta_I]$ oraz $v = [v(u_1), ..., v(u_k)]$, parametry mają następującą interpretację:

– parametr β_i, $i = 1, ..., I$ mierzy poziom wpływu *i*-tej zmiennej taryfikacyjnej na wartość szkód (stały dla wszystkich kategorii),
– parametr $v(u_k)$, $k = 1, ..., K$ mierzy poziom ryzyka w obrębie kategorii (zmienny dla każdej kategorii).

Oszacowane wartości parametrów stałych służą najczęściej do wyznaczenia tzw. wskaźników taryfikacyjnych *t*. Wyznaczanie tych wskaźników zależy od liczby zmiennych taryfikacyjnych w modelu. W przypadku np. założenia modelu multiplikatywnego i dwóch zmiennych taryfikacyjnych wskaźniki mają postać $t_{ij} = \exp(\beta_i + \beta_j)$. Ostateczny poziom taryfy to skorygowany wskaźnik poziomem ryzyka $v(u)$.

Jak już wspomniano, algorytm maksymalizacji *h*-wiarygodności w modelu HGLM jest złożony ze względu na możliwość przyjmowania różnych rozkładów dla składnika losowego i zmiennej o wpływie losowym, nie tylko rozkładów normalnych. W przypadku procesu taryfikacji w ubezpieczeniach majątkowych możliwe jest uproszczenie tego algorytmu, przyjmując pewne założenia [Nelder, Verral 1997].

Przyjmijmy następujące oznaczenia: u_k − zmienna nominalna wielokategorialna, $k = 1, ..., K$, y_{ik} − wartości szkód dla i-tego ryzyka oraz *k*-tej kategorii zmiennej *u*, $i = 1, ..., I$. Ponadto zakłada się, że zmienne u_k są niezależne o tych samych rozkładach i mają taką samą stałą wariancję σ_U^2, zmienna $y_{ik} \mid u_k$ ma rozkład *Tweedie*, pomiędzy zmiennymi o stałym i losowym wpływie występuje zależność multiplikatywna. Wtedy

$$E(y_{ik} \mid u_k) = \mu_i u_k, \qquad \mathrm{var}(y_{ik} \mid u_k) = \varphi V(\mu_i).$$

Procedura iteracyjna dla jednoczesnego szacowania parametrów stałych i losowych przedstawia się następująco (PROC1) [Ohlsson, Johansson 2003]:

KROK 1. Przyjęcie założenia dla parametru ryzyka $\hat{u}_k = 1$ dla $k = 1, ..., K$.

KROK 2. Estymacja μ_i dla $i = 1, ..., I$, stosując model GLM z logarytmiczną funkcją połączenia oraz $\log(\hat{u}_k)$ jako zmienną *offset*.

KROK 3. Estymacja $\varphi\alpha$:

$$\varphi\alpha = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I} w_{ik}\mu_i^{2-p}\left(\frac{y_{ik}}{\mu_i} - \bar{u}_k\right)^2}{\sum_{k=1}^{K}(I_k - 1)} \frac{\sum_{k=1}^{K}\sum_{i=1}^{I} w_{ik}\mu_i^{2-p}(\bar{u}_k - 1)^2}{\sum_{k=1}^{K}\sum_{i=1}^{I} w_{ik}\mu_i^{2-p}}.$$

KROK 4. Estymacja $\bar{u}_k$:

$$\bar{u}_k = \frac{\sum_{i=1}^{I} \frac{w_{ik}\mu_i^{2-p} y_{ik}}{\mu_i}}{\sum_{i=1}^{I} w_{ik}\mu_i^{2-p}}, \ k = 1, ..., K.$$

KROK 5. Estymacja $\hat{u}_k$:

$$\hat{u}_k = \frac{\sum_{i=1}^{I} \frac{w_{ik} y_{ik}}{\mu_i^{p-1}} + \varphi\alpha}{\sum_{i=1}^{I} w_{ik} \mu_i^{2-p} + \varphi\alpha}, \quad k = 1, ..., K.$$

KROK 6. Powrót do **KROK 2** z nową zmienną *offset* jako $\log(\hat{u}_k)$.

W przykładzie obliczeniowym przedstawionym w dalszej części opracowania przyjęte zostało kryterium stopu w kolejnych iteracjach jako $\varepsilon < 0,0000001$.

Z procedury PROC1 widać, iż w pierwszej fazie procesu taryfikacji szacowane są parametry o stałym wpływie. Parametry te są takie same w obrębie jednej kategorii zmiennej wielokategorialnej. W drugiej fazie natomiast szacowane są wartości zmiennej wielokategorialnej dla każdej kategorii. Zatem parametr u_k pokazuje, o ile należy skorygować taryfę dla zmiennych taryfikacyjnych o stałym wpływie w k-tej kategorii, dla każdego $k = 1, ..., K$.

Inną możliwością estymacji parametrów w modelu HGLM jest wykorzystanie wspomnianego wcześniej specjalistycznego modułu *HGLM* w oprogramowaniu *GenStat*, w którym stosowana jest maksymalizacja *h*-wiarygodności, w szczególności nie musi być spełnione założenie o stałej wariancji σ_U^2 dla każdej kategorii zmiennej wielokategorialnej.

4. Przykład obliczeniowy dla modelu HGLM

W celu prezentacji zastosowania mieszanych modeli liniowych w procesie taryfikacji wykorzystano bazę danych szkód komunikacyjnych w Szwecji (*third party motor insurance claims*) zaczerpniętą z bazy internetowej o adresie `http://www.statsci.org/data/general/motorins.html`. Baza zawiera następujące dane:

a) *Kilometers* (X_1) – liczba kilometrów przebytych w ciągu roku:

1: < 1000,
2: 1000-15000,
3: 15000-20000,
4: 20000-25000,
5: > 25000,

b) *Zone* (X_2) – strefa geograficzna w Szwecji:

1: Stockholm, Goteborg, Malmo,
2: pozostałe duże miasta,
3: mniejsze miasta z południowej Szwecji,
4: obszary wiejskie w południowej Szwecji,

5: mniejsze miasta z północnej Szwecji,

6: obszary wiejskie w północnej Szwecji,

7: Gotlandia,

c) *Bonus* (X_3) – bonus za bezszkodową jazdę równy liczbie lat plus 1, począwszy od ostatniej szkody,

d) *Make* (X_4) – liczby od 1 do 8 reprezentujące osiem najczęściej występujących modeli samochodów. Pozostałe modele zawarte są w kategorii 9,

e) *Insured* (X_5) – liczba ubezpieczonych w danym roku,

f) *Claims* – liczba szkód,

g) *Payment* (X_6) – całkowita wartość szkód w Skr.

Tabela 1. Estymacja parametrów w modelu GLM

Zmienne	Parametry strukturalne modelu	Odchylenie standardowe	*t-value*	Stopa taryfy
Stała	8,3946	0,0234	359,4743	.
X11	0,0000	.	.	1,0000
X12	0,0245	0,0129	1,9032	1,0248
X13	0,0212	0,0149	1,4287	1,0215
X14	0,0431	0,0207	2,0771	1,0440
X15	0,0394	0,0220	1,7893	1,0402
X21	0,0000	.	.	1,0000
X22	0,0229	0,0163	1,4017	1,0231
X23	0,0479	0,0166	2,8786	1,0490
X24	0,1287	0,0149	8,6388	1,1374
X25	0,0517	0,0250	2,0703	1,0531
X26	0,1465	0,0204	7,1717	1,1578
X27	0,0228	0,0699	0,3258	1,0230
X31	0,0000	.	.	1,0000
X32	0,0435	0,0208	2,0911	1,0444
X33	0,0692	0,0232	2,9779	1,0716
X34	0,0568	0,0251	2,2674	1,0585
X35	0,0336	0,0240	1,4007	1,0342
X36	0,0699	0,0200	3,4928	1,0724
X37	0,1163	0,0149	7,7784	1,1233
X41	0,0000	.	.	1,0000
X42	–0,0352	0,0365	–0,9650	0,9654
X43	0,0844	0,0431	1,9551	1,0880
X44	–0,1643	0,0414	–3,9660	0,8485
X45	–0,0872	0,0348	–2,5082	0,9165
X46	–0,0393	0,0298	–1,3182	0,9614
X47	–0,1194	0,0401	–2,9767	0,8875
X48	0,2135	0,0543	3,9337	1,2381
X49	–0,0549	0,0171	–3,2181	0,9466

Źródło: obliczenia własne w programie R

Tabela 2. Estymacja parametrów w modelu GLMM

Zmienne	Parametry strukturalne modelu	Odchylenie standardowe	*t-value*	Stopa taryfy
Stała	8,4178	0,0843	99,8198	.
X11	0,0000	.	.	1,0000
X12	0,0138	0,0588	0,2345	1,0139
X13	0,0849	0,0596	1,4247	1,0886
X14	0,0178	0,0621	0,2868	1,0180
X15	0,1167	0,0631	1,8491	1,1238
X31	0,0000	.	.	1,0000
X32	0,1220	0,0736	1,6581	1,1297
X33	0,0136	0,0736	0,1852	1,0137
X34	0,0853	0,0740	1,1518	1,0890
X35	0,0667	0,0735	0,9082	1,0690
X36	0,0666	0,0719	0,9256	1,0688
X37	0,1060	0,0704	1,5060	1,1119
X41	0,0000	.	.	1,0000
X42	–0,0428	0,0791	–0,5416	0,9581
X43	0,1589	0,0823	1,9308	1,1722
X44	–0,0334	0,0860	–0,3881	0,9672
X45	–0,0186	0,0790	–0,2356	0,9816
X46	0,0253	0,0784	0,3234	1,0257
X47	–0,0625	0,0799	–0,7814	0,9394
X48	0,2621	0,0830	3,1595	1,2997
X49	–0,0454	0,0757	–0,6000	0,9556

Źródło: obliczenia własne w programie R.

Tabela 3. Poziom zmiennej u_k dla $k = 1, ..., 7$

Zone (X2)	GLM	GLMMPQL	PROC1
Stockholm, Goteborg, Malmo	1,0000	0,9711	1,0000
Pozostałe duże miasta	1,0231	0,9326	1,1008
Mniejsze miasta z południowej Szwecji	1,0490	0,9894	0,9600
Obszary wiejskie z południowej Szwecji	1,1374	1,0215	1,0125
Mniejsze miasta z północnej Szwecji	1,0531	1,0395	1,0181
Obszary wiejskie z północnej Szwecji	1,1578	1,0570	1,0135
Gotlandia	1,0230	0,9944	1,0068

Źródło: obliczenia własne.

Dla danych przeprowadzono estymację parametrów stałych oraz losowych w modelu GLM, GLMM oraz PROC1. Wykorzystano program statystyczny R, pakiet `stats` (funkcja `glm`), pakiet `MASS` (funkcja `glmmPQL`) oraz własną implementację procedury PROC1 (załącznik). Uzyskane wyniki dla poziomu składnika losowego przedstawiają tab. 1-3 (w modelu GLM zmienna X_2 była traktowana oczywiście jako zmienna o stałym wpływie).

W modelu GLM uzyskano składkę bazową na poziomie $B = e^{8,3945} = 4422,98$ Skr. Najmniejszym ryzykiem charakteryzują się polisy na samochód modelu o nr 4 (niestety z bazy danych nie wynika, jaki to jest model), w odniesieniu do którego uzyskano taryfę 0,8485. Oznacza to, że dla tej grupy polis składkę bazową należy obniżyć o 15,15% (przy uwzględnianiu jedynie zmiennej X_4). Natomiast największe ryzyko występuje również dla zmiennej X_4, ale dla przypadku modelu nr 8, w odniesieniu do którego powinno nastąpić zwiekszenie składki o 23,81%. Oczywiście, wyznaczając ostateczny poziom taryfy dla konkretnej polisy, należy uwzględnić wartości parametrów dla wszystkich zmiennych uwzględnianych w taryfikacji.

W modelu GLMM uzyskano ogólnie nieco wyższe wartości parametrów. Najniższa taryfa występuje dla regionu „Pozostałe duże miasta” – obniżenie składki bazowej o 6,74%, najwyższa taryfa dotyczy, podobnie jak w przypadku GLM, modelu samochodu nr 8 – wzrost składki o 29,97%.

5. Podsumowanie

W artykule przedstawiono w zarysie ideę liniowych modeli mieszanych, w których występują zmienne objaśniające mające stały oraz losowy wpływ na zmienną objaśnianą. Następnie wskazano procedury służące do estymacji parametrów strukturalnych modeli mieszanych wraz z odnośnikami do literatury, gdzie można znaleźć szczegółowe opisy tych procedur. W dalszej części pracy przedstawiono uproszczoną procedurę estymacji parametrów w modelu HGLM, która znajduje zastosowanie w budowie taryf w ubezpieczeniach majątkowych. Procedura ta została zaimplementowana w programie R. W celu zobrazowania poszczególnych algorytmów estymacji przeprowadzono przykład obliczeniowy. Dalsze prace autorki dotyczą implementacji komputerowej modelu HGLM w programie R dla przypadków różnych rozkładów, co pozwoli na analizę efektywności wykorzystywania tych modeli w taryfikacji.

Literatura

Lee Y., Nelder A.J., Pawitan Y., *Generalized Linear Models with Random Effects*, Monographs on Statistics and Applied Probability 106, Chapman&Hall\CRC, 2006.

McCulloch Ch.E., Searle Sh.R., *Generalized, Linear, and Mixed Models*, Wiley Series in Probability and Statistics, John Wiley & Sons, INC., 2001.

Nelder J.A., Verral R.J., *Credibility theory and generalized linear models,* „ASTIN Bulletin” 1997, vol. 27:1, 71-82.

Ohlsson E., Johansson B., *Credibility theory and GLM revised,* Research Report 2003: 15, „Mathematical Statistics", Stockholm University, 2003.
Walesiak M., Gatnar E., *Statystyczna analiza danych z wykorzystaniem programu R*, PWN, Warszawa 2009.

RATEMAKING IN NON-LIFE INSURANCE WITH HGLM

Summary: Insurance companies specializing in non-life insurance create their own rating systems for setting fair premiums for every risk for different kinds of insurance portfolios. The most popular rating technique is to estimate the relativities of a number of rating factors in a multiplicative or an additive model. Generally, these relativities are discreet (e.g. sex) or continuous (e.g. age, engine capacity). In recent years, the standard practice in insurance companies is to use two types of the estimation methods: minimum bias methods or generalized linear models (GLM). In the paper, the author presents the generalized linear mixed model in which some rating factors are treated as random effects.

Załącznik

```
u_k_macierz=NULL
model.glm=glm(Y~X1+X2+X3+X4+X5+X6+offset(log(of)),  fam-
ily=Gamma(link="log"), data=dane, weights=dane$N)
p=2
mi.glm=fitted.values(model.glm)
K=7
u_k=c()
u_k_1=c()
z_k=c()
mse=(c)
I_M=c(0,295,295,293,306,236,264,108)
sigma2_k=c()
for (k in 1:K){
suma=0
suma1=0
j=1
for (i in ((I_M[k]+1):I_M[k+1])){
      suma=suma+((dane$N[i]*(mi.glm[j]^(2-
p)))*dane$Y[i])/mi.glm[j]
      suma1=suma1+dane$N[i]*(mi.glm[j]^(2-p))
j+1
}
u_k_1=c(u_k_1,(suma/suma1))
}
for (k in 1:K) {
suma=0
j=1
for (i in ((I_M[k]+1):I_M[k+1])){
      suma=suma+(dane$N[i]*(mi.glm[j]^(2-
p)))*((dane$Y[i]/mi.glm[j])-u_k_1[k])^2
j=j+1
}
sigma2_k
=c(sigma2_k,(1/(I_M[k]-1
))*suma)
}
suma=0
suma1=0
for (k in 1:K){
      suma=suma+((I_M[k+1]-1)*sigma2_k[k])
```

```
      suma1=suma1+(I_M[k+1]-1)
}
sigma2=suma/suma1
for (k in 1:K){
j=1
for (i in ((I_M[k]+1):I_M[k+1])){
      suma=suma+dane$N[i]*(mi.glm[j]^(2-p))*((u_k_1[k]-
1)^2)
      suma1=suma1+dane$N[i]*(mi.glm[j]^(2-p))
j=j+1
}
}
sigma2_u=(suma-(K*sigma2))/suma1
psi_alfa=sigma2/sigma2_u
for (k in 1:K){
suma=0
suma1=0
j=1
for (i in ((I_M[k]+1):I_M[k+1])){
      suma=suma+(dane$N[i]*dane$Y[i])/(mi.glm[j]^(p-1))
      suma1=suma1+dane$N[i]*(mi.glm[j]^(2-p))
j=j+1
}
u_k=c(u_k,((suma+psi_alfa)/(suma1+psi_alfa)))
}
for (k in 1:K){
suma=0
suma1=0
j=1
for (i in ((I_M[k]+1):I_M[k+1])){
      suma=suma+dane$N[i]*(mi.glm[j]^(2-p))
      suma1=suma1+dane$N[i]*(mi.glm[j]^(2-p))
j=j+1
}
z_k=c(z_k,(suma/(suma1+psi_alfa)))
}
predict=c()
for (k in 1:K){
for (i in 1:I_M[k]){
      predict=c(predict,mi.glm[i]*u_k[k])

}
}-
```