**PRACE NAUKOWE**
Uniwersytetu Ekonomicznego we Wrocławiu
**RESEARCH PAPERS**
of Wrocław University of Economics

**254**

# Inwestycje finansowe i ubezpieczenia – tendencje światowe a rynek polski

Redaktorzy naukowi
**Krzysztof Jajuga**
**Wanda Ronka-Chmielowiec**

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Diarmuid Bradley, Jan Czekaj, Marek Gruszczyński, Jacek Lisowski, Paweł Miłobędzki, Włodzimierz Szkutnik, Mirosław Szreder, Adam Szyszka, Waldemar Tarczyński, Stanisław Wieteska, Tomasz Wiśniewski

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:
www.ibuk.pl, www.ebscohost.com,
The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon
http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl



**ISSN 1899-3192**
**ISBN 978-83-7695-293-2**

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

## Spis treści

## Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 254 • 2012
Inwestycje finansowe i ubezpieczenia – tendencje światowe a rynek polski ISSN 1899-3192

**Alicja Wolny-Dominiak**
Uniwersytet Ekonomiczny w Katowicach

# MODELOWANIE LICZBY SZKÓD W UBEZPIECZENIACH KOMUNIKACYJNYCH W PRZYPADKU WYSTĘPOWANIA DUŻEJ LICZBY ZER

**Streszczenie:** Istotnym problemem w procesie taryfikacji oraz klasyfikacji polis w portfelu polis jest modelowanie liczby szkód. Portfele ubezpieczeniowe charakteryzują się tym, iż dla znacznej liczby polis nie występuje żadna szkoda. Zatem zmienna licznikowa będąca liczbą szkód dla polisy przyjmuje istotnie dużą liczbę zer (*Zero-Inflated* (ZI) *effect*), co powoduje, że klasyczna regresja Poissona nie daje zadowalających wyników. W pracy przedstawiono zmodyfikowaną regresję Poissona uwzględniającą efekt ZI. W tego typu regresji występuje „podwójny" problem doboru zmiennych do modelu. Jeden zbiór zmiennych niezależnych wpływa na liczbę szkód, natomiast drugi zbiór zmiennych (może być ten sam) wpływa na wystąpienie zerowej szkody. W pracy przedstawiono procedurę wyboru obu zbiorów zmiennych niezależnych w procesie szacowania liczby szkód.

**Słowa kluczowe:** liczba szkód, regresja Poissona, ZIP, ZINB.

## 1. Wstęp

Jednym z problemów występujących w analizie danych ubezpieczeniowych jest modelowanie liczby szkód występujących w danym portfelu polis z zastosowaniem regresji przy założeniu rozkładu Poissona. Jednak portfele ubezpieczeniowe charakteryzują się tym, iż dla wielu polis w okresie ubezpieczenia nie wystąpiła żadna szkoda. Oznacza to, iż dane zawierają dużą liczbę zer, co powoduje, że klasyczna regresja Poissona nie daje zadowalających wyników. W pierwszej części pracy przedstawiono regresję Poissona dla zmiennej licznikowej oraz zmodyfikowaną wersję regresji Poissona uwzględniającą sytuację występowania dużej liczby zer w danych (ZIP). W drugiej części pracy przeprowadzono przykład empiryczny, w którym zastosowano i porównano modele Poissona oraz ZIP do portfela ubezpieczeń komunikacyjnych, a szczególnie procedurę wyboru zbioru zmiennych wpływających na zmienną zależną, czyli liczbę szkód, oraz zbioru zmiennych wpływających na wystąpienie zerowej szkody w portfelu polis. Jako kryterium wyboru układu zmiennych zależnych w obu zbiorach przyjęto kryterium Akaika AIC.

## 2. Modele regresji z licznikową zmienną objaśnianą typu ZI

W tej części pracy przedstawione zostaną modele regresji, w których zmienną zależną jest zmienna licznikowa przyjmująca wartości całkowite nieujemne oraz występuje duża liczba wartości zerowych (*Zero-Inflated* – ZI). W ubezpieczeniach majątkowych modele takie mają zastosowanie przede wszystkim w modelowaniu oraz prognozowaniu liczby szkód. Stosowany jest najczęściej model regresji Poissona, w którym przyjmuje się założenie, iż zmienna objaśniana $Y$ ma rozkład Poissona $Y \sim Pois(\lambda)$ warunkowany wartościami zmiennych objaśniających [Denuit i in. 2007]:

$$P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \; i = 1, ..., n .$$

W powyższym wzorze $Y_i$ oznacza liczbę szkód dla $i$-tej osoby ubezpieczonej. Parametr $\lambda_i$ uzależniony jest od pewnych zmiennych zależnych $X_j$, $j = 1, ..., k$ charakteryzujących ubezpieczonego oraz pojazd, którego dotyczy ubezpieczenie, np. płci, wieku, pojemności silnika samochodu. Najczęściej przyjmowana jest logarytmiczna funkcja połączenia:

$$\ln \lambda_i = \sum_{j=1}^{k} \beta_{ji} X_{ji} .$$

Przechodząc do wartości oczekiwanej, mamy:

$$\lambda_i = \mu_i = e^{\sum_{j=1}^{k} \beta_{ji} X_{ji}} .$$

Zatem widać, iż dla każdej kombinacji zmiennych objaśniających uzyskiwana jest zawsze dodatnia oczekiwana liczba szkód. W modelu przyjmuje się założenia, że zmienna $Y$ ma rozkład Poissona, średnia wartość zmiennej jest równa wariancji oraz $y_1, ..., y_n$ są niezależne o stałej wariancji.

Parametr $\lambda_i$ może być wykorzystywany do rangowania polis ze względu na liczbę szkód. Jednak niezbędna jest korekta tego parametru wskaźnikiem czasu ekspozycji na ryzyko dla $i$-tej polisy $d_i$, który pokazuje najczęściej w przypadku ubezpieczeń majątkowych, jaką część badanego okresu obejmowała polisa:

$$\lambda_i = d_i e^{\beta_0 + \sum_{j=1}^{k} \beta_{ij}} .$$

Powyższy model nie uwzględnia przypadku, w którym zmienna licznikowa przyjmuje dużą liczbę wartości zerowych. Taka sytuacja występuje często w przypadku modelowania liczby szkód. Analizując portfele ryzyka, można zauważyć, że dla wielu polis nie wystąpiła żadna szkoda, natomiast w przypadku wystąpienia szkód są to jedna, dwie, trzy i rzadko więcej. Dlatego w przypadku analizy liczby szkód w zakładzie ubezpieczeń zasadniejsze wydaje się stosowanie zmodyfikowanej regresji Poissona, gdzie uwzględnia się dużą liczbę wartości zerowych w danych zwanej modelem ZIP (*Zero-Inflated Poisson*). W modelu ZIP niezależne zmienne $Y_i$ przyjmują wartości zerowe: $Y_i \sim 0$ z prawdopodobieństwem $\varpi_i$ lub wartości z rozkładu Poissona: $Y_i \sim Pois(\lambda_i)$ z prawdopodobieństwem $1 - \varpi_i$, co można zapisać następująco [Lambert 1992]:

$$P(Y_i = y_i) = \begin{cases} \varpi_i + (1-\varpi_i)e^{-\lambda_i}, & y_i = 0 \\ (1-\varpi_i)\dfrac{e^{(-\lambda_i)}\lambda_i^{y_i}}{y_i!}, & y_i > 0 \end{cases}, \ i = 1, ..., n.$$

Zatem w modelu ZIP występują dwa parametry: $\lambda_i$ oraz $\varpi_i$. Oba parametry, podobnie jak w przypadku regresji Poissona, połączone są ze zmiennymi objaśniającymi następującymi funkcjami połączeń:

$$\ln\left(\frac{\varpi_i}{1-\varpi_i}\right) = \sum_{j=1}^{t} \gamma_{ji} Z_{ji},$$

$$\ln \lambda_i = \sum_{j=1}^{k} \beta_{ji} X_{ji},$$

gdzie $Z_1, ..., Z_l$ są zmiennymi zależnymi dla równania pierwszego, natomiast $X_1, ..., X_k$ zmiennymi dla równania drugiego. Oczekiwana liczba szkód oraz wariancja liczby szkód *i*-tej polisy w modelu ZIP wynosi:

$$E(Y_i) = \mu_i(1-\varpi_i),$$

$$D(Y_i) = (1-\varpi_i)(\mu_i + \varpi_i \mu_i^2).$$

Podobnie jak w przypadku regresji Poissona w modelu ZIP zakłada się, iż średnia liczba szkód jest równa wariancji. W przypadku, gdy wariancja jest wyższa od średniej, występuje problem nadmiernej dyspersji, który często charakteryzuje zmienne licznikowe. Powoduje on, iż statystyki $\chi^2$ testujące istotność parametrów strukturalnych modelu są przeszacowane, natomiast nie zmienia zgodności estymatorów parametrów. W celu uniknięcia nadmiernej dyspersji można zastosować skorygowane błędy standardowe lub przejść do modelu, w którym wprowadzany jest rozkład negatywny dwumodalny [Kopczewska i in. 2009]. Model ten oznaczany jest najczęściej ZINB (*Zero-Inflated Negative Binomial*):

$$P(Y_i = y_i) = \begin{cases} \varpi_i + (1-\varpi_i)(1+\alpha\lambda_i^c)^{\frac{-\lambda_i^{1-c}}{\alpha}}, & y_i = 0 \\ (1-\varpi_i)\dfrac{\Gamma(y_i + \frac{\lambda_i^{1-c}}{\alpha})}{y_i!\Gamma(\frac{\lambda_i^{1-c}}{\alpha})}(1+\alpha\lambda_i^c)^{\frac{-\lambda_i^{1-c}}{\alpha}}(1+\dfrac{\lambda_i^{-c}}{\alpha})^{-y_i}, & y_i > 0 \end{cases},$$

gdzie $\alpha \geq 0$ jest parametrem dyspersji. W przypadku, gdy $\alpha -> 0$, model ZINB sprowadza się do modelu ZIP (bez nadmiernego rozproszenia).

## 3. Przykład empiryczny

Proces modelowania i prognozowania liczby szkód w zakładzie ubezpieczeń przeprowadzono z wykorzystaniem bazy danych szkód komunikacyjnych (*third party motor insurance claims*) zaczerpniętej z pozycji [de Jong, Heller 2008] http://www.acst.mq.edu.au/research/books/GLMsforInsuranceData/data_sets. Baza danych zawiera następujące zmienne uwzględnione w modelu:

a. Zmienna objaśniana licznikowa:
*numclaims* – liczba szkód.
b. Zmienne objaśniające:
*veh_body* – kształt samochodu,
*veh_age* – wiek pojazdu: 1 (najmłodszy), 2, 3, 4,
*gender* – płeć kierowcy: 0 (kobieta), 1,
*agecat* – wiek kierowcy: 1 (najmłodszy), 2, 3, 4, 5, 6.
c. Czas ekspozycji na ryzyko (mierzony w okresie trwania polisy w stosunku do całego okresu uwzględnianego w bazie):
$d_i$ – wartości z przedziału $[0,1]$.

Obliczenia wykonano w programie komputerowym R.

Rozkład liczby szkód w analizowanym portfelu przedstawiono w tab. 1.

**Tabela 1.** Rozkład liczby szkód

| Liczba szkód | Liczba polis | Częstość | Średni czas ekspozycji na ryzyko |
|---|---|---|---|
| 0 | 63 232 | 93,19% | 0,45 |
| 1 | 4 333 | 6,39% | 0,6 |
| 2 | 271 | 0,40% | 0,71 |
| 3 | 18 | 0,03% | 0,7 |
| 4 | 2 | 0,00% | 0,88 |

Źródło: opracowanie własne.

Widać, iż liczba szkód charakteryzuje się bardzo dużą liczbą zer, gdzie 93% polis nie wygenerowało żadnej szkody w portfelu. Wartość wariancji przewyższa wartość średniej i indeks nadmiernej dyspersji jest na poziomie:

$$O = \frac{wariancja - średnia}{średnia} = 0,0063,$$

co oznacza słaby efekt nadmiernej dyspersji w portfelu. Do modelowania liczby szkód zastosowano w pierwszej kolejności regresję Poissona. W modelu P1 badano wpływ poszczególnych zmiennych na liczbę szkód:

$$\ln \lambda = \beta_0 + \beta_1 veh\_body + \beta_2 veh\_age + \beta_3 gender + \beta_4 agecat\ .$$

Model szacowano, wykorzystując funkcję `glm(){stats}`, przyjmując rozkład Poissona dla liczby szkód. W pierwszej kolejności zbadano istotność poszczególnych zmiennych na liczbę szkód (tab. 2).

**Tabela 2.** Parametry strukturalne regresji Poissona dla modeli P1

| Model P1 | $\hat{\beta}_i$ | Średni błąd szacunku | $p$-wartość |
|---|---|---|---|
| Stała | –1,18 | 0,32 | 0,00 |
| *Veh_body* | –0,95 | 0,39 | 0,05 |
| *Veh_age* | –0,04 | 0,01 | 0,00 |
| *gender* | –0,01 | 0,03 | 0,79 |
| *agecat* | –0,08 | 0,01 | 0,00 |

Źródło: obliczenia własne.

W modelu P1 na poziomie istotności 5% zmienna charakteryzująca płeć jest statystycznie nieistotna. Zatem w dalszej analizie zmienna ta została usunięta z modelu. Ponadto pozostałe zmienne nie są skorelowane. Nowy model P2 przyjął postać:

$$\ln \lambda = \beta_0^1 + \beta_1^1 veh\_body + \beta_2^1 veh\_age + \beta_3^1 agecat\ .$$

Uzyskano parametry strukturalne przedstawione w tab. 3.

**Tabela 3.** Parametry strukturalne regresji Poissona dla modelu P2

| Realizacje zmiennych w modelu P2 | $\beta$ | $e^{\beta}$ | Średni błąd szacunku |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| Stała | –1,35 | 0,26 | 0,32 |
| *veh_age1* | 0,00 | 1,00 | – |
| *veh_age2* | 0,13 | 1,14 | 0,04 |
| *veh_age3* | 0,001 | 1,001 | 0,04 |

**Tabela 3**, cd.

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| *veh_age4* | –0,08 | 0,93 | 0,04 |
| *agecat1* | 0,00 | 1,00 | – |
| *agecat2* | –0,17 | 0,85 | 0,05 |
| *agecat3* | –0,20 | 0,82 | 0,05 |
| *agecat4* | –0,23 | 0,80 | 0,05 |
| *agecat5* | –0,42 | 0,65 | 0,06 |
| *agecat6* | –0,43 | 0,65 | 0,07 |
| *veh_body*_BUS | 1,00 | 1,00 | – |
| *veh_body*_CONVT | –1,75 | 0,17 | 0,66 |
| *veh_body* _COUPE | –0,75 | 0,47 | 0,34 |
| *veh_body* _HBACK | –1,10 | 0,33 | 0,32 |
| *veh_body* _HDTOP | –0,87 | 0,42 | 0,33 |
| *veh_body* _MCARA | –0,46 | 0,63 | 0,41 |
| *veh_body* _MIBUS | –1,15 | 0,32 | 0,35 |
| *veh_body* _PANVN | –0,84 | 0,43 | 0,34 |
| *veh_body* _RDSTR | –0,68 | 0,51 | 0,66 |
| *veh_body* _SEDAN | –1,04 | 0,35 | 0,32 |
| *veh_body* _STNWG | –1,00 | 0,37 | 0,32 |
| *veh_body* _TRUCK | –1,04 | 0,35 | 0,33 |
| *veh_body* _UTE | –1,25 | 0,29 | 0,32 |

Źródło: obliczenia własne.

Wszystkie parametry modelu są statystycznie istotne. Również test ilorazu wiarygodności pokazuje, iż model jest w całości statystycznie istotny. W wyniku działania funkcji `lrtest{lmtest}` uzyskano bardzo niski, prawie zerowy poziom $p$-wartości.

**Tabela 4.** Test ilorazu wiarygodności dla modelu P2

| | #Df | LogLik | Df | Chisq | Pr(>Chisq) |
|---|---|---|---|---|---|
| Model P1.1 | 21 | –18029,2 | NA | NA | NA |
| Model – tylko stała | 1 | –18101,5 | –20 | 144,5839 | 0,000001 |

Źródło: obliczenia własne.

Kryterium AIC przyjęło wartość na poziomie 36 120. Interpretując uzyskane wyniki, na podstawie wartości parametrów strukturalnych zawartych w tab. 3 można stwierdzić kierunek wpływu zmiany wieku samochodu, wieku kierowcy oraz kształtu samochodu na liczbę szkód na podstawie znaku. Natomiast w celu określenia jednostkowego wpływu zmiennych objaśniających na liczbę szkód wyznaczono eksponenty parametrów strukturalnych modelu (w modelu przyjęto logarytmiczną funkcję połączenia).

Powyższe wyniki można wykorzystać do rangowania polis, wyznaczając wartość $\lambda$. Ze względu na dużą liczbę polis w bazie w pracy nie został przedstawiony pełny ranking. Można jednak stwierdzić, iż najlepsze polisy uzyskały $\lambda = 0,0271$, natomiast najgorsze $\lambda = 0,2567$. Analizując zmienne taryfikacyjne, można następnie korygować stopy taryfy ze względu na „skłonność” do generowania szkód.

Jako że w analizowanej bazie danych występuje duża liczba polis, dla których nie wystąpiła żadna szkoda, dalej zastosowano modele ZIP. W pierwszym modelu założono, iż estymowane jest równanie drugie, zakładając jednakowe prawdopodobieństwo wystąpienia braku szkody na polisie w całym portfelu. Natomiast w modelu drugim przyjęto, iż w obu równaniach występują te same zmienne objaśniające.

**Model ZIP1**

$$\ln \lambda^{ZIP1} = \beta_0^{ZIP1} + \beta_1^{ZIP1} veh\_body + \beta_2^{ZIP1} veh\_age + \beta_3^{ZIP1} agecat$$

$$\ln(\frac{\varpi^{ZIP1}}{1-\varpi^{ZIP1}}) = \gamma_0^{ZIP1}.$$

**Model ZIP2**

$$\ln \lambda^{ZIP2} = \beta_0^{ZIP2} + \beta_1^{ZIP2} veh\_body + \beta_2^{ZIP2} veh\_age + \beta_3^{ZIP2} agecat$$

$$\ln(\frac{\varpi^{ZIP2}}{1-\varpi^{ZIP2}}) = \gamma_0^{ZIP2} + \gamma_1^{ZIP2} veh\_body + \gamma_2^{ZIP2} veh\_age + \gamma_3^{ZIP2} agecat.$$

**Model ZIP3**

$$\ln \lambda^{ZIP3} = \beta_0^{ZIP3} + \beta_1^{ZIP3} veh\_body + \beta_2^{ZIP3} veh\_age + \beta_3^{ZIP3} agecat$$

$$\ln(\frac{\varpi^{ZIP3}}{1-\varpi^{ZIP3}}) = \gamma_0^{ZIP3} + \gamma_1^{ZIP3} veh\_body.$$

**Model ZIP4**

$$\ln \lambda^{ZIP4} = \beta_0^{ZIP4} + \beta_1^{ZIP4} veh\_body + \beta_2^{ZIP4} veh\_age + \beta_3^{ZIP4} agecat$$

$$\ln(\frac{\varpi^{ZIP4}}{1-\varpi^{ZIP4}}) = \gamma_0^{ZIP4} + \gamma_1^{ZIP4} veh\_body + \gamma_2^{ZIP4} veh\_age.$$

**Model ZIP5**

$$\ln \lambda^{ZIP5} = \beta_0^{ZIP5} + \beta_1^{ZIP5} veh\_body + \beta_2^{ZIP5} veh\_age + \beta_3^{ZIP5} agecat$$

$$\ln(\frac{\varpi^{ZIP5}}{1-\varpi^{ZIP5}}) = \gamma_0^{ZIP5} + \gamma_1^{ZIP5} veh\_body + \gamma_2^{ZIP5} agecat.$$

**Model ZIP6**

$$\ln \lambda^{ZIP6} = \beta_0^{ZIP6} + \beta_1^{ZIP6} veh\_body + \beta_2^{ZIP6} veh\_age + \beta_3^{ZIP6} agecat$$

$$\ln\left(\frac{\varpi^{ZIP6}}{1-\varpi^{ZIP6}}\right) = \gamma_0^{ZIP6} + \gamma_1^{ZIP6} veh\_age + \gamma_2^{ZIP6} agecat.$$

**Model ZIP7**

$$\ln \lambda^{ZIP7} = \beta_0^{ZIP7} + \beta_1^{ZIP7} veh\_body + \beta_2^{ZIP7} veh\_age + \beta_3^{ZIP7} agecat$$

$$\ln\left(\frac{\varpi^{ZIP7}}{1-\varpi^{ZIP7}}\right) = \gamma_0^{ZIP7} + \gamma_2^{ZIP7} veh\_age .$$

**Model ZIP8**

$$\ln \lambda^{ZIP8} = \beta_0^{ZIP8} + \beta_1^{ZIP8} veh\_body + \beta_2^{ZIP8} veh\_age + \beta_3^{ZIP8} agecat$$

$$\ln\left(\frac{\varpi^{ZIP8}}{1-\varpi^{ZIP8}}\right) = \gamma_0^{ZIP8} + \gamma_3^{ZIP8} agecat.$$

Modele szacowano, wykorzystując funkcję `zeroinfl(){pscl}`. Do wyboru modelu wykorzystano kryterium Akaika AIC.

**Tabela 5.** Wartości kryterium AIC dla modeli typu ZI

| Model | AIC |
|---|---|
| ZIP1 | 35 938 |
| ZIP2 | 35 941 |
| ZIP3 | 35 952 |
| ZIP4 | 35 958 |
| ZIP5 | 35 957 |
| ZIP6 | 35 958 |
| ZIP7 | 35 966 |
| ZIP8 | 35 961 |

Źródło: obliczenia własne.

Widać, iż wszystkie modele typu ZI osiągają niższy poziom AIC niż model nieuwzględniający masowego występowania wartości zerowej. Minimalną wartość AIC uzyskał model ZIP1, gdzie na wartość szkód wpływają zmienne istotne: *veh_body*, *veh_age, agecat,* natomiast liczba zerowych polis jest modelowana wyrazem wolnym. Parametry strukturalne oraz eksponenty parametrów w tym modelu przedstawiono w tab 6.

W celu porangowania polis niezbędne jest wyznaczenie dodatkowego parametru $\varpi$ , który będzie stały dla wszystkich polis. Wynika to z postaci modelu ZIP1, gdzie:

$$\ln\left(\frac{\varpi}{1-\varpi}\right) = -0,27,$$

**Tabela 6.** Parametry strukturalne modelu ZIP2

| Realizacje zmiennych w modelu P2 | $\beta^{ZIP1}$ | $e^{\beta^{ZIP1}}$ | Średni błąd szacunku |
|---|---|---|---|
| Stała | –0,79 | 0,45 | 0,35 |
| Veh_age1 | 0,00 | 1,00 | – |
| Veh_age2 | 0,13 | 1,14 | 0,04 |
| Veh_age3 | 0,00 | 1,00 | 0,04 |
| Veh_age4 | –0,08 | 0,93 | 0,05 |
| agecat1 | 0,00 | 1,00 | – |
| agecat2 | –0,16 | 0,85 | 0,06 |
| agecat3 | –0,20 | 0,82 | 0,05 |
| agecat4 | –0,23 | 0,80 | 0,05 |
| agecat5 | –0,42 | 0,66 | 0,06 |
| agecat6 | –0,43 | 0,65 | 0,07 |
| Veh_body_BUS | 0,00 | 1,00 | – |
| Veh_body_CONVT | –1,75 | 0,17 | 0,68 |
| Veh_body _COUPE | –0,74 | 0,48 | 0,36 |
| Veh_body _HBACK | –1,09 | 0,33 | 0,34 |
| Veh_body _HDTOP | –0,88 | 0,42 | 0,35 |
| Veh_body _MCARA | –0,47 | 0,62 | 0,43 |
| Veh_body _MIBUS | –1,15 | 0,32 | 0,37 |
| Veh_body _PANVN | –0,83 | 0,43 | 0,36 |
| Veh_body _RDSTR | –0,65 | 0,52 | 0,69 |
| Veh_body _SEDAN | –1,04 | 0,35 | 0,34 |
| Veh_body _STNWG | –0,99 | 0,37 | 0,34 |
| Veh_body _TRUCK | –1,04 | 0,35 | 0,35 |
| Veh_body _UTE | –1,24 | 0,29 | 0,35 |
| | $\gamma^{ZIP1}$ | $e^{\gamma^{ZIP1}}$ | Średni błąd szacunku |
| Stała | –0,27 | 0,76 | 0,12 |

Źródło: obliczenia własne.

dając parametr $\varpi = 0,43$. Podobnie jak dla modelu P2 stworzono ranking polis, gdzie najlepsze polisy uzyskały $\tilde{\lambda} = 0,027$, natomiast najgorsze $\tilde{\lambda} = 0,2941$.

## 4. Podsumowanie

W pracy przedstawiono zmodyfikowaną regresję Poissona w przypadku, gdy w danych występuje duża liczba zer dla zmiennej licznikowej, i jej porównanie z klasyczną regresją Poissona. W analizowanym przykładzie model ZIP przyniósł lepsze rezultaty niż model klasyczny, dając niższy poziom wartości kryterium AIC. Zasadniczo wadą obu modeli jest fakt, iż w danej klasie polis wszystkie polisy charakteryzują się taką samą oczekiwaną liczba szkód, co jest założeniem mało realnym. Rozwiązaniem tego problemu jest przejście do mieszanego modelu Poissona, wprowadzając czynnik losowy różnicujący polisy, co leży w zakresie przyszłych prac autorki nad omawianą tematyką.

## Literatura

De Jong P., Heller G.Z., *Generalized Linear Models for Insurance Data*, Cambridge Unversity Press, 2008.

Denuit M., Marechal X., Pitrebois S., Walhin J. *Actuarial Modelling of Claims Counts*, John Wiley&Sons Ltd, 2007.

Kopczewska K., Kopczewski T., Wójcik P., *Metody ilościowe w R. Aplikacje ekonomiczne i finansowe*, Cedewu.pl Wydawnictwa Fachowe, Warszawa 2009.

Lambert D., *Zero-Inflated Poisson regression, with an application to defects in manufacturing*, "Technometrics," vol. 34, no 1, Feb. 1992.

Vuong Q., *Likelihood ratio test for model selection and nonnested hypothesis*, "Econometrica" 1989, nr 57.

## ZERO-INFLATED CLAIM COUNT MODELING IN AUTOMOBILE INSURANCE. CASE STUDY

**Summary:** An important problem in the ratemaking process is to model the number of claims. Insurance portfolios are characterized by the fact that a significant number of policies have zero claims (called zero-inflated (ZI) effect). Thus, the classical Poisson regression does not give satisfactory results for the claim count with ZI effect. This paper presents a modified Poisson regression taking into account ZI effect. There is a „double" problem of the selection of variables into the model in this type of regression. One set of independent variables affects the number of accidents and the second set of variables (which may be the same) affects the occurrence of zero claims. In the paper we present a procedure for the selection of these two sets of independent variables in the claim count estimation. In the investigated case study, the automobile insurance dataset is taken from the literature [de Jong et al. 2008]. All calculations are based on a package {pscl} from R CRAN software.

**Keywords:** claim count, zero-inflation, ZIP, ZINB.