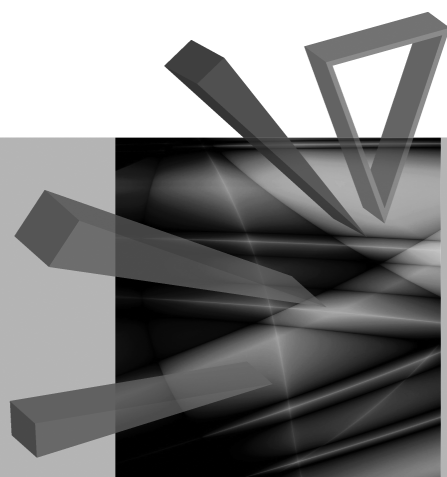


Prognozowanie w zarządzaniu firmą



Redaktorzy naukowi
Paweł Dittmann
Aleksandra Szpulak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2011

Senacka Komisja Wydawnicza

Zdzisław Pisz (przewodniczący),

Andrzej Bąk, Krzysztof Jajuga, Andrzej Matysiak, Waldemar Podgórski,

Mieczysław Przybyła, Aniela Styś, Stanisław Urban

Recenzenci

Włodzimierz Szkutnik, Jan Zawadzki

Redakcja wydawnicza

Barbara Majewska

Redakcja techniczna i korekta

Barbara Łopusiewicz

Skład i łamanie

Comp-rajt

Projekt okładki

Beata Dębska

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2011

ISSN 1899-3192

ISBN 978-83-7695-141-6

Druk: Drukarnia TOTEM

Spis treści

Wstęp	7
Agnieszka Przybylska-Mazur: Optymalne zasady polityki pieniężnej w prognozowaniu wskaźnika inflacji	9
Alicja Wolny-Dominiak: Zmodyfikowana regresja Poissona dla danych ubezpieczeniowych z dużą liczbą zer	21
Andrzej Gajda: Doświadczenia i metody pozyskiwania danych eksperckich na potrzeby badań z wykorzystaniem metod foresight	30
Anna Gondek: Prognozy rozwoju gospodarczego Polski z użyciem metody analogii przestrzenno-czasowych	41
Bartosz Lawędziak: Sekurytyzacja papierów wartościowych opartych na hipotece odwrotnej	50
Filip Chybalski: Prakseologiczne aspekty prognozowania	59
Ireneusz Kuroпка, Paweł Lenczewski: Możliwość zastosowania modeli ekonometrycznych do prognozowania w przedsiębiorstwie Brenntag Polska	69
Jacek Szandula: Wyszukiwanie formacji w kursach giełdowych przy użyciu metod klasyfikacji danych	82
Joanna Perzyńska: Zastosowanie sztucznych sieci neuronowych do wyznaczania nieliniowych prognoz kombinowanych	94
Konstancja Poradowska, Tomasz Szkutnik, Mirosław Wójciak: Scenariusze rozwoju wybranych technologii oszczędności energii w życiu codziennym	102
Maciej Oesterreich: Wykorzystanie pakietu statystycznego R w prognozowaniu na podstawie danych w postaci szeregów czasowych z wahaniami sezonowymi	113
Marcin Błażejowski, Paweł Kufel, Tadeusz Kufel: Algorytm zgodnego modelowania i prognozowania procesów ekonomicznych jako pakiet funkcji <i>Congruent Specification</i> programu Gretl	125
Marcin Błażejowski: Stacjonarność szeregów czasowych o wysokiej częstotliwości obserwowania – implementacja testu stacjonarności Dickeya w programie Gretl	137
Mirosław Wójciak: Wpływ czynników i zdarzeń kluczowych na rozwój nowych technologii – wybrane metody korygowania prognoz na przykładzie technologii energooszczędnych	149
Monika Dyduch: Grupowanie produktów strukturyzowanych	159
Piotr Bernat: Planowanie działalności przedsiębiorstwa wspomagane prognozowaniem	170

Roman Pawlukowicz: Informacje prognostyczne w rynkowych sposobach wyceny nieruchomości – identyfikacja i pozyskiwanie	182
Wojciech Zatoń: Uwarunkowania psychologiczne w prognozowaniu	189

Summaries

Agnieszka Przybylska-Mazur: Optimal monetary policy rules in forecasting of inflation rate	20
Alicja Wolny-Dominiak: Zero-inflated Poisson Model for insurance data with a large number of zeros	29
Andrzej Gajda: Experience and methods of data collection from experts for research using foresight methods	40
Anna Gondek: Economic growth forecasts for Poland using the time-space analogy method	49
Bartosz Lawędziak: Securitization of survivor bonds based on the reverse mortgage	58
Filip Chybalski: Praxiological aspects of forecasting	68
Ireneusz Kuropka, Paweł Lenczewski: Econometric models usage feasibility in Brenntag Poland forecasting	81
Jacek Szandula: Searching for technical analysis formations in stock prices with the use of cluster analysis methods	93
Joanna Perzyńska: Application of artificial neural networks to build the nonlinear combined forecasts	101
Konstancja Poradowska, Tomasz Szkutnik, Mirosław Wójciak: The scenarios of development of selected technologies related to energy saving in everyday life	112
Maciej Oesterreich: The R application in forecasting unsystematic lacks in seasonal time series	124
Marcin Błażejowski, Paweł Kufel, Tadeusz Kufel: Congruent modelling and forecasting algorithm as function package Congruent Specification in GRETL	136
Marcin Błażejowski: Stationarity of high-frequency time series – implementation of Dickey’s stationarity test in GRETL	148
Mirosław Wójciak: The influence of key and events factors on the development of new technologies – selected methods of forecast correction on the example of energy-saving technologies	158
Monika Dyduch: Ranking of structured products	169
Piotr Bernat: Forecasting assisted business management planning	181
Roman Pawlukowicz: Prognostic data in market ways of property valuation – identification and acquisition	188
Wojciech Zatoń: Psychological aspects of forecasting	199

Alicja Wolny-Dominiak

Uniwersytet Ekonomiczny w Katowicach

ZMODYFIKOWANA REGRESJA POISSONA DLA DANYCH UBEZPIECZENIOWYCH Z DUŻĄ LICZBĄ ZER

Streszczenie: W modelowaniu liczby szkód dla portfeli ubezpieczeniowych zastosowanie znajduje klasyczna regresja Poissona. Jednak specyfika zjawiska wystąpienia szkody charakteryzuje się tym, iż dla wielu polis w okresie ubezpieczenia nie występuje żadna szkoda. Powoduje to, że dane zawierają dużą liczbę zer. W takim przypadku lepsze wyniki daje zmodyfikowana wersja regresji Poissona, uwzględniająca taką sytuację. W pracy przedstawione jest zastosowanie regresji zmodyfikowanej z wykorzystaniem programu komputerowego R oraz pakietu {pscl}.

Słowa kluczowe: ubezpieczenia, szkody, modelowanie.

1. Wstęp

Jednym z problemów występujących w analizie danych ubezpieczeniowych jest modelowanie liczby szkód. Portfele ubezpieczeniowe charakteryzują się tym, że w przypadku wielu polis w okresie ubezpieczenia nie występuje żadna szkoda. Zatem dane zawierają dużą liczbę zer, przez co klasyczna regresja Poissona nie daje zadowalających wyników. W pracy przedstawiona zostanie zmodyfikowana wersja regresji Poissona, uwzględniająca sytuację występowania dużej liczby zer w danych (*zero-inflated Poisson regression*), i jej zastosowanie w modelowaniu liczby szkód w ubezpieczeniach majątkowych. Przedstawione modele regresyjne pozwalają między innymi na prognozowanie oczekiwanej liczby szkód dla nowo zawieranej polisy, jednakże nie chodzi o to, co się wydarzy w przyszłości, lecz o to, jaka będzie liczba szkód w danej klasie polis opisanej przez zmienne objaśniające.

2. Modele regresji z licznikową zmienną objaśnianą przyjmującą dużą liczbę wartości zerowych

W tej części pracy przedstawione zostaną modele regresji, w których zmienną objaśniającą jest zmienna licznikowa przyjmująca wartości całkowite nieujemne.

W ubezpieczeniach majątkowych modele takie mają zastosowanie w szczególności w modelowaniu oraz prognozowaniu liczby szkód. Stosowany jest najczęściej model regresji Poissona, w którym przyjmuje się założenie, że zmienna objaśniana Y ma rozkład Poissona $Y \sim Pois(\lambda)$ warunkowany wartościami zmiennych objaśniających [Denuit i in. 2007]:

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

W powyższym wzorze Y_i oznacza liczbę szkód dla i -tej osoby ubezpieczonej. Parametr λ_i uzależniony jest od pewnych zmiennych zależnych X_j , $j = 1, \dots, k$, 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, że 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, \dots, y_n są niezależne o stałej wariancji.

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

Powyższy model nie uwzględnia przypadku, w którym zmienna licznikowa przyjmuje dużą liczbę wartości zerowych. Taka sytuacja występuje często, gdy modelowana jest liczba szkód. Analizując portfele ryzyk, można zauważyć, że dla wielu polis nie wystąpiła żadna szkoda, a jeśli wystąpiły szkody, 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 ϖ_i lub wartości z rozkładu Poissona: $Y_i \sim Pois(\lambda_i)$ z prawdopodobieństwem $1 - \varpi_i$, co można zapisać następująco [Lambert 1992]:

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

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

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

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

gdzie Z_1, \dots, Z_l są zmiennymi zależnymi dla równania pierwszego, natomiast X_1, \dots, 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, że średnia liczba szkód jest równa wariancji. Jeśli wariancja jest wyższa od średniej, występuje problem nadmiernej dyspersji, który często charakteryzuje zmienne licznikowe. Powoduje on, że statystyki χ^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*).

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 [Jong, Heller 2008]. Baza danych zawiera następujące zmienne uwzględnione w modelu:

- a) zmienna objaśniana licznikowa:
numclaims – liczba szkód;
- b) zmienne objaśniające:
veh_value – wartość samochodu,

veh_age – wiek kierowcy: 1 (najmłodszy), 2, 3, 4,
gender – płeć kierowcy: 0 (kobieta), 1,
agecat – wiek kierowcy: 1 (najmłodszy), 2, 3, 4, 5, 6;

c) ekspozycja na ryzyko (mierzona w okresie trwania polisy w stosunku do całego okresu uwzględnianego w bazie):

exposure – wartości z przedziału [0,1].

Obliczenia wykonano w programie komputerowym R.

Rozkład liczby szkód w analizowanym portfelu przedstawia się następująco:

Tabela 1. Rozkład liczby szkód

Liczba szkód	Liczba szkód	Częstość (w %)	Średnia ekspozycja na ryzyko
0	63 232	93,19	0,45
1	4333	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.

Z tabeli 1 wynika, że liczba szkód charakteryzuje się bardzo dużą liczbą zer – 93% polis nie wygenerowało żadnej szkody w portfelu. Wartość wariancji przewyższa wartość średniej i indeks nadmiernej dyspersji jest na poziomie:

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

co oznacza słaby efekt nadmiernej dyspersji w portfelu. Do modelowania liczby szkód zastosowano w pierwszej kolejności regresję Poissona. W modelu M1 badano wpływ poszczególnych zmiennych na liczbę szkód, natomiast w modelu M2 wprowadzono interakcję pomiędzy płcią i wiekiem osoby ubezpieczonej.

1) Model P1:

$$\ln \lambda = \beta_0 + \beta_1 \text{veh_value} + \beta_2 \text{veh_age} + \beta_3 \text{gender} + \beta_4 \text{agecat}.$$

2) Model P2 (zagnieżdżony):

$$\ln \lambda = \beta_0 + \beta_1 \text{veh_value} + \beta_2 \text{veh_age} + \beta_3 \text{gender} + \\ + \beta_4 \text{agecat} + \beta_5 \text{gender} * \text{agecat}.$$

Modele szacowano, wykorzystując funkcję `glm()` {stats} i przyjmując rozkład Poissona dla liczby szkód. Oszacowane parametry regresji przedstawia tabela 2.

W obu modelach na poziomie istotności 5% zmienne charakteryzujące wiek samochodu oraz płeć są statystycznie nieistotne. Ponadto kryterium AIC przyjmuje nieznacznie niższą wartość w modelu P1 (AIC = 36130) niż w modelu P2

(AIC = 36132). Zatem do dalszej analizy zastosowano model P1 z usuniętymi zmiennymi *veh_age* oraz *gender*.

Tabela 2. Parametry strukturalne regresji Poissona dla modeli P1

Model P1	$\hat{\beta}_i$	Średni błąd szacunku	<i>p</i> -wartość
Stała	-2,396	0,068	0,000
<i>veh_value</i>	0,043	0,013	0,001
<i>veh_age</i>	-0,012	0,016	0,453
<i>genderM</i>	-0,020	0,029	0,493
<i>agecat</i>	-0,077	0,010	0,000

Źródło: obliczenia własne.

Tabela 3. Parametry strukturalne regresji Poissona dla modeli P2

Model P2	$\hat{\beta}_i$	Średni błąd szacunku	<i>p</i> -wartość
Stała	-2,398	0,074	0,000
<i>veh_value</i>	0,043	0,013	0,001
<i>veh_age</i>	-0,012	0,016	0,452
<i>genderM</i>	-0,016	0,074	0,831
<i>agecat</i>	-0,077	0,014	0,000
<i>gend:agec</i>	-0,001	0,020	0,950

Źródło: obliczenia własne.

3) Model P1.1:

$$\ln \lambda = \beta_0 + \beta_1 \text{veh_value} + \beta_2 \text{agecat} .$$

Oszacowane parametry modelu P1.1 kształtują się następująco:

Tabela 4. Parametry regresji Poissona dla modeli P1.1

Model P1.1	$\hat{\beta}_i$	Średni błąd szacunku	<i>p</i> -wartość
Stała	-2,4421	0,0421	0,000000
<i>veh_value</i>	0,0473	0,0106	0,000008
<i>agecat</i>	-0,0777	0,0101	0,000000

Źródło: obliczenia własne.

W zmodyfikowanym modelu wszystkie zmienne są statystycznie istotne na poziomie istotności 5%. Ponadto test ilorazu wiarygodności pokazuje, że również cały model P1.1 jest statystycznie istotny.

Interpretując uzyskane wyniki na podstawie wartości parametrów strukturalnych zawartych w tabeli 4 można jedynie stwierdzić kierunek wpływu zmiany wartości samochodu oraz wieku kierowcy na liczbę szkód na podstawie znaku. Tak

więc wzrost wartości samochodu zwiększa oczekiwaną liczbę szkód, natomiast wzrost wieku kierowcy zmniejsza tę liczbę. W celu określenia jednostkowego wpływu zmiennych objaśniających na liczbę szkód niezbędne jest wyznaczenie eksponentów parametrów strukturalnych modelu, co wynika z postaci funkcji połączenia.

Tabela 5. Test ilorazu wiarygodności dla modelu P1.1

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
Model P1.1	3	-18060,698	NA	NA	NA
Model – tylko stała	1	-18101,501	-2,000	81,605	0,000

Źródło: obliczenia własne.

Tabela 6. Eksponenty parametrów strukturalnych modelu P1.1

	$\exp(\hat{\beta}_i)$
Stała	0,0870
<i>veh_value</i>	1,0484
<i>agecat</i>	0,9253

Źródło: obliczenia własne.

Zatem jednostkowy wzrost wartości samochodu zwiększa o 4% liczbę szkód, natomiast starszy wiek kierowcy wpływa na zmniejszenie liczby szkód o 8%.

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

4) Model ZIP1:

$$\ln \lambda^{ZIP1} = \beta_0^{ZIP1} + \beta_1^{ZIP1} veh_value + \beta_2^{ZIP1} veh_age + \beta_3^{ZIP1} gender + \beta_4^{ZIP1} agecat,$$

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

5) Model ZIP2:

$$\ln \lambda^{ZIP2} = \beta_0^{ZIP2} + \beta_1^{ZIP2} veh_value + \beta_2^{ZIP2} veh_age + \beta_3^{ZIP2} gender + \beta_4^{ZIP2} agecat,$$

$$\ln\left(\frac{\varpi^{ZIP2}}{1 - \varpi^{ZIP2}}\right) = \gamma_0^{ZIP2} + \gamma_1^{ZIP2} veh_value + \gamma_2^{ZIP2} veh_age + \gamma_3^{ZIP2} gender + \gamma_4^{ZIP2} agecat.$$

Modele szacowano wykorzystując funkcję `zeroinfl()` `{pscl}` i uzyskując następujące wyniki:

Tabela 7. Parametry strukturalne modelu ZIP1

Model ZIP1	β_i^{ZIP1}	Średni błąd szacunku	<i>p</i> -wartość
Stała	-1,819	0,089	0,000
<i>veh_value</i>	0,045	0,014	0,001
<i>veh_age</i>	-0,011	0,016	0,503
<i>genderM</i>	-0,021	0,030	0,491
<i>agecat</i>	-0,077	0,010	0,000
Model ZIP1	γ_0^{ZIP1}	Średni błąd szacunku	<i>p</i> -wartość
Stała	-0,235	0,1227	0,0551

Źródło: obliczenia własne.

W modelu ZIP1 na poziomie istotności 5% zmienne: wiek samochodu, płeć oraz wiek kierowcy są nieistotne w równaniu drugim w modelu ZIP, natomiast stała, wiek samochodu oraz płeć są nieistotne w równaniu pierwszym modelu ZIP. Do wyboru modelu (niezagnieżdżonego) spośród powyższych dwóch zastosowano test Vuonga, w którym statystyką empiryczną jest znormalizowana logarymiczna różnica funkcji wiarygodności obu modeli [Lambert 1992]. Test ten daje jednoznaczne rozstrzygnięcie (w przeciwieństwie na przykład do testu Coxa), nie wymaga żadnych założeń początkowych. Użyto funkcji `vuong()` `{pscl}`. Statystyka testu wynosi -5,105, co oznacza, że model ZIP2 jest lepszy od modelu ZIP1 na poziomie istotności 5% (*p*-wartość = 0,000000165). Zatem dalej rozważano model ZIP2, usuwając z niego parametry nieistotne:

Tabela 8. Parametry strukturalne modelu ZIP2

Model ZIP2	β_i^{ZIP2}	Średni błąd szacunku	<i>p</i> -wartość
Stała	-1,842	0,179	0,000
<i>veh_value</i>	-0,099	0,027	0,000
<i>veh_age</i>	0,020	0,051	0,701
<i>genderM</i>	-0,014	0,074	0,854
<i>agecat</i>	-0,021	0,027	0,431
Model ZIP2	γ_i^{ZIP2}	Średni błąd szacunku	<i>p</i> -wartość
Stała	0,223	0,431	0,605
<i>veh_value</i>	-0,559	0,100	0,000
<i>veh_age</i>	-0,010	0,120	0,934
<i>genderM</i>	0,047	0,166	0,777
<i>agecat</i>	0,125	0,059	0,036

Źródło: obliczenia własne.

6) Model ZIP2.1

$$\ln \lambda^{ZIP2.1} = \beta_0^{ZIP2.1} + \beta_1^{ZIP2.1} veh_value$$

$$\ln\left(\frac{\varpi^{ZIP2.1}}{1-\varpi^{ZIP2.1}}\right) = \gamma_1^{ZIP2.1} veh_value + \gamma_2^{ZIP2.1} agecat.$$

Oszacowane parametry modelu ZIP2.1 przedstawia tabela 9.

Tabela 9. Parametry strukturalne modelu ZIP2.1

Model ZIP2.1	$\beta_i^{ZIP2.1}$	Średni błąd szacunku	<i>p</i> -wartość
Stała	-1,8579	0,0455	0,0000
<i>veh value</i>	-0,1044	0,0157	0,0000
Model ZIP2.1	$\gamma_i^{ZIP2.1}$	Średni błąd szacunku	<i>p</i> -wartość
<i>veh value</i>	-0,5014	0,0705	0,0000
<i>agecat</i>	0,1684	0,0179	0,0000

Źródło: obliczenia własne.

Tabela 10. Test ilorazu wiarygodności dla modelu ZIP2.1

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
Model 2.1	4	-18026,5	NA	NA	NA
Model – tylko stała	2	-18052,2	-2	51,3121	0,0000

Źródło: obliczenia własne.

Wszystkie parametry modelu, a także cały model są istotne. W ostatnim kroku przeprowadzono test Vuonga, porównując model P2.1 z modelem ZIP1.2. Statystyka testu wynosi -2,404, co oznacza, że Model ZIP2 jest lepszy na poziomie istotności 5% (*p*-wartość = 0,00808).

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. Analizując dane empiryczne w przykładzie, można zauważyć, że model ZIP daje lepsze rezultaty niż model klasyczny. Zasadniczo wadą obu modeli jest fakt, że w danej klasie polis wszystkie polisy charakteryzują się taką samą oczekiwaną liczbą szkód, co jest założeniem mało realnym. Rozwiązanie tego problemu stanowi przejście do mieszanego modelu Poissona poprzez wprowadzenie czynnika losowego różnicującego polisy.

Literatura

- Jong P. de, Heller G.Z., *Generalized Linear Models for Insurance Data*, Cambridge University Press, 2008. http://www.acst.mq.edu.au/research/books/GLMsforInsuranceData/data_sets.
- Denuit M., Marechal X., Pitrebois S., Walhin J. *Actuarial Modelling of Claims Counts*, John Wiley&Sons Ltd., Chichester 2007.
- Kopczewska K., Kopczewski T., Wójcik P., *Metody ilościowe w R. Aplikacje ekonomiczne i finansowe*, Cedetu.pl, Wydawnictwa Fachowe, Warszawa 2009.
- Lambert D., *Zero-Inflated Poisson Regression, with an application to defects in manufacturing*, „Technometrics”, 1 Feb. 1992, vol. 34.
- Vuong Q., *Likelihood ratio test for model selection and nonnested hypothesis*, „Econometrica” 1989, no. 57.

ZERO-INFLATED POISSON MODEL FOR INSURANCE DATA WITH A LARGE NUMBER OF ZEROS

Summary: In modelling the number of insurance claims the classical Poisson regression is used. Insurance portfolios are characterized by the fact that for many policies in the insurance period there are no claims at all. Thus, the data contain a large number of zeros, so that classical Poisson regression does not give satisfactory results. The work presented in the paper is a modified version of Poisson regression, taking into account the situation of a large number of zeros in the data (called zero-inflated Poisson regression) and its application in modelling the number of claims in insurance assets. Calculations of the actual data taken from the literature are implemented in a computer program R using the package {pscl}.