

ALICJA WOLNY-DOMINIAK

Możliwości programu R w szacowaniu rezerwy IBNR¹

W zakładach ubezpieczeń majątkowych istotną pozycję w funduszu ubezpieczeniowym zajmuje rezerwa z tytułu zaistniałych szkód niezgłoszonych do dnia tworzenia rezerwy (ozn. IBNR). W literaturze przedmiotu zaproponowano wiele metod szacowania rezerwy IBNR, w których wykorzystano różnorodne techniki modelowania statystycznego, analizy danych czy data mining. Przegląd metod kalkulacji rozpatrywanej rezerwy jest szeroko eksploatowany w literaturze.² Celem niniejszego artykułu jest zapoznanie czytelnika z możliwościami wykorzystania pakietu {ChainLadder} programu komputerowego R³ w analizie rozwoju szkodowości oraz szacowaniu rezerwy IBNR.

Słowa kluczowe: rezerwa IBNR, Chain Ladder, GLM, bootstrap, program R.

Wprowadzenie

Program R jest darmowym programem statystycznym dedykowanym pierwotnie szerokiej grupie naukowców mogących implementować różnorodne algorytmy obliczeniowe w otwartym środowisku informatycznym. Z czasem jednak zyskał popularność również w zastosowaniach komercyjnych, np. w dziedzinie finansów. Od dwóch lat intensywnie rozwijany jest pakiet służący do analizy rozwoju szkodowości oraz szacowania rezerwy IBNR o nazwie {ChainLadder}. W pakiecie

1. Praca częściowo finansowana przez grant Narodowego Centrum Nauki (nr NN 111461540).
2. Por. w pracach: M.V. Wüthrich, M. Merz, *Stochastic Claims Reserving Methods in Non-Life Insurance*, John Willey & Sons, England, 2008; W. Bijak, M. Pawlak, M. Smętek, *Analiza rezerw w oparciu o trójkąty szkód – badanie wstępne dla polskiego rynku ubezpieczeń*, Prace Naukowe Akademii Ekonomicznej Imienia Oskara Langego we Wrocławiu 990, 2003, s. 432–441; S. Wieteska, *Rezerwy techniczno-ubezpieczeniowe zakładów ubezpieczeń. Teoria i praktyka*, Branta Oficyna Wydawnicza, 2004; A. Wolny, *Podejście Aktuarialne do kalkulacji rezerwy szkodowej, Statystyczne zaawansowane metody kalkulacji rezerwy szkodowej*, [w:] *Metody kalkulacji ryzyka rezerw szkodowych w ubezpieczeniach majątkowo-osobowych*, Wydawnictwo Akademii Ekonomicznej w Katowicach, 2005; A. Pობlocka, *Wybrane metody kalkulacji rezerwy IBNR*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu 1197, 2008, s. 368–376.
3. R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Austria 2012.

zaimplementowano metody szacowania IBNR, które można podzielić następująco: metoda *chain ladder* wraz z modyfikacjami⁴, metoda GLM – stosuje uogólniony model liniowy do szacowania wartości wypłaconych odszkodowań⁵. Dane wejściowe wprowadzane są w postaci trójkąta szkód w formie skumulowanej lub nieskumulowanej. Pakiet pozwala na ciekawą wizualizację danych zawartych w trójkącie szkód, jak również wykresy przydatne w analizie wyników estymacji oraz dopasowania modelu do danych.

W pierwszej części pracy zawarto teoretyczne aspekty szacowania IBNR wraz z opisem odpowiednich funkcji w programie R.⁶ W drugiej części skupiono się na opisie technicznej strony stosowania pakietu {ChainLadder} oraz przedstawiono przykładowe ekrany programu R zawierające uzyskane wyniki.

2. Metody szacowania IBNR w pakiecie {ChainLadder} – aspekty teoretyczne

Szacując rezerwę IBNR, wykorzystuje się dane zawarte w trójkącie szkód, w którym wyróżnia się następujące elementy:

- n – liczba okresów uwzględnianych w analizie,
- numer wiersza $i, i=1, \dots, n$ – okres wystąpienia szkody zwany dalej okresem wypadkowym,
- numer kolumny $j, j=1, \dots, n$ – okres opóźnienia w wypłacie odszkodowania,
- X_{ij} – wartości odszkodowań dla szkód, które zaszły w okresie i , wypłaconych z opóźnieniem j ,
- C_{ij} – skumulowana wartość odszkodowań dla szkód, które zaszły w okresie i , wypłaconych z opóźnieniem j : $C_{i,j} = \sum_{k=0}^j X_{i,k}$.

Zazwyczaj wyznacza się rezerwę IBNR w oparciu o dwa rodzaje trójkąta szkód: zawierającego wartości wypłaconych odszkodowań (PLT – *Paid Losses Triangle*) oraz zawierającego wartości wypłaconych i niewypłaconych zaszytych szkód (ILT – *Incurred Losses Triangle*).

Przechodząc do opisu statystycznego, w trójkącie szkód zawarte są obserwacje dla n zmiennych losowych X_{0^*}, \dots, X_{n-1} będących wartościami odszkodowań wypłaconych z opóźnieniem $j=1, \dots, n^2$. Analogicznie można interpretować skumulowane wartości odszkodowań C_{0^*}, \dots, C_{n-1} .

Pierwszą metodą dostępną w pakiecie {ChainLadder} jest ta bazująca na idei klasycznej metody *chain ladder* (CL). Istotą metody CL jest oszacowanie tzw. współczynnika LDF oznaczonego dalej jako $f_j, j=1, \dots, n-1$, który interpretuje się jako zmianę wartości wypłaconych odszkodowań w danym okresie opóźnienia w stosunku do wcześniejszego okresu opóźnienia. Zakładając zależ-

-
4. Por. w pracach: T. Mack, *Distribution-free calculation of the standard error of chain ladder reserve estimates*, „Astin Bulletin”, 23/2. 1993, s. 213–225; T. Mack, *The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor*, „Astin Bulletin” 29/2, 1999, s. 361–366; G. Quarg, T. Mack, *Blätter der Deutschen Gesellschaft für Versicherungs- und Finanzmathematik*, 26/4 2004, s. 597–630.
 5. Por. w pracach: S. Christofides, *Regression models based on log-incremental payments*, „Claims Reserving Manual” 2/1990, Institute of Actuaries, London; A.E. Renshaw, P. Verrall, *A stochastic model underlying the chain ladder technique*, „British Actuarial Journal”, 4/1998, s. 903–923.
 6. P. Biecek, *Przewodnik po pakiecie R*, Wydawnictwo GIS, 2008.
 7. Standardowo zmienną losową oznacza się wielkimi literami, a jej realizacje małymi literami. Jednak w przypadku opisu metod szacowania IBNR w literaturze przedmiotu stosuje się jedynie wielkie litery.

ność liniową pomiędzy skumulowanymi wartościami trójkąta w kolejnych okresach opóźnienia, do szacowania LDF można zastosować modele liniowe.

Założmy, że zależność regresyjna pomiędzy skumulowanymi wartościami wypłaconych odszkodowań w kolejnych okresach opóźnień opisuje funkcja regresji liniowej:

$$C_{j+1} = f_j C_j + \xi, j=1, \dots, n-1.$$

Przy spełnionych założeniach metody najmniejszych kwadratów, nieobciążonym estymatorem

$$\text{współczynnika LDF jest } \hat{f}_j = \frac{\sum_{i=0}^{n-j-2} C_{i,j} C_{i,j+1}}{\sum_{i=0}^{n-j-2} C_{i,j}^2}, \text{ a jego wariancja } D^2\{\hat{f}_j\} = \frac{\sigma_j^2}{\sum_{i=0}^{n-j-2} C_{i,j}^2}$$

gdzie σ_j^2 jest wariancją resztową. W praktyce ubezpieczeniowej częstym przypadkiem jest występowanie heteroskedastyczności składnika losowego⁸. Możliwe rozwiązanie takiej sytuacji

to wprowadzenie wag $w_{i,j} \in [0,1]$ spełniających warunek $\sum_{i=0}^{n-j-2} w_{i,j} = 1$ oraz przejście do ważonej metody najmniejszych kwadratów. Wtedy nieobciążonym parametrem czynnika LDF jest

$$\hat{f}_j = \sum_{i=0}^{n-j-2} w_{i,j} \frac{C_{i,j+1}}{C_{i,j}} \text{ wraz z wariancją } D^2\{\hat{f}_j\} = \frac{\sigma_j^2}{\sum_{i=0}^{n-j-2} w_{i,j} C_{i,j}^2}, \text{ gdzie } \alpha \in \{0,1,2\}.$$

W przypadku gdy wagi oraz parametr α przyjmowane są jako 1, ważona metoda najmniejszych kwadratów sprowadza się do metody klasycznej.

Funkcja pakietu `{ChainLadder}`, w której zaimplementowana jest powyższa metoda estymacji nosi nazwę `MackChainLadder()`. Główne argumenty funkcji to:

- `Triangle` – skumulowany trójkąt szkód,
- `weights` – wagi, domyślnie przyjmowana jest wartość 1,
- `alpha` – parametr α , domyślnie przyjmowana jest wartość 1,
- `est.sigma` – metoda wyznaczania wariancji resztowej.

Główne wartości zwracane to:

- `FullTriangle` – pełny trójkąt szkód z oszacowanymi wartościami poniżej przekątnej,
- `Models` – model liniowy dla każdego okresu opóźnienia,
- `f` – czynnik LDF,
- `f.se` – odchylenie standardowe dla estymatora czynnika LDF.

W powyższej metodzie rezerwa IBNR szacowana jest dla trójkąta PLT lub trójkąta ILT. W praktyce okazuje się często, iż wysokość wypłaty nie pokrywa się z faktyczną wartością szkody. W metodzie zwanej `MunichChainLadder` (ozn. MCL) nowatorstwem jest uwzględnienie powstałej różnicy poprzez założenie o istnieniu zależności (mierzonej korelacją) pomiędzy dwoma trójkątami. Wprowadzany jest, poza LDF, wskaźnik Q_j , będący stosunkiem wypłaconych odszkodowań do zaistniałych

8. D.M. Murphy, *Unbiased Loss Development Factors*, Proceedings of the Casualty Actuarial Society Casualty Actuarial Society – Arlington, Virginia 1994, LXXXI s. 154–222.

szkód, dla których nie nastąpiła wypłata odszkodowania: $Q_i = \frac{P_i}{I_i}$. Analiza w czasie pozwala na prognozowanie tego wskaźnika, który następnie wykorzystywany jest do korekty rezerwy IBNR.

W celu oszacowania rezerwy IBNR metodą MCL, w pakiecie {ChainLadder} zaimplementowano funkcję MunichChainLadder(). Funkcja ta posiada następujące argumenty:

- Paid – skumulowany trójkąt PLT,
 - Incurred – skumulowany trójkąt ILT,
- Główne wartości zwracane przez funkcję to:
- MCLPaid – prognozowane wartości w trójkącie PLT,
 - MCLIncurred – prognozowane wartości w trójkącie ILT,
 - MackPaid – prognozowane wartości w trójkącie PLT metodą CL,
 - MackIncurred – prognozowane wartości w trójkącie ILP metodą CL,
 - q.f – wskaźnik Q .

W metodzie CL oraz jej rozszerzeniu MCL nie przyjmuje się zasadniczo żadnych założeń odnośnie rozkładów dla zmiennych. Kolejna metoda dostępna w pakiecie {ChainLadder} bazuje na uogólnionym modelu liniowym (ozn. GLM). Danymi wyjściowymi w modelu GLM są wartości z nieskumulowanego trójkąta szkód $[X_{i,j}]_{n \times n}$. Zmienną objaśnianą jest wartość wypłaconych odszkodowań $X_{i,j}$, na którą wpływa okres wystąpienia szkody (α_i) oraz okres opóźnienia (β_j). Model GLM ma następującą postać⁹:

$$\ln(x_{ij}) = \mu + \alpha_i + \beta_j + \xi_{ij},$$

gdzie składnik losowy ξ ma rozkład z podrodziny rozkładów Tweedie, zależny od parametru p [Tweedie, 1984]. W przypadku gdy $p=1$, uzyskiwany jest rozkład od-Poisson, natomiast dla $p=2$ – rozkład gamma. Istotą metody GLM jest oszacowanie wartości μ , co łatwo uzyskać jeśli się korzysta z techniki uogólnionych modeli liniowych¹⁰.

W pakiecie {ChainLadder} występuje funkcja glmReserve(), której głównymi argumentami są:

- var.power – parametr p ,
- link.power – definiowanie funkcji połączenia,
- mse.method – wskazanie metody wyznaczania średniego błędu szacunku.

W wyniku działania funkcji glmReserve() uzyskiwane są wartości:

- FullTriangle – pełny trójkąt szkód z oszacowanymi wartościami poniżej przekątnej,
- model – parametry strukturalne modelu GLM.

Poza tym, jako że w funkcji glmReserve() wykorzystuje się wbudowaną funkcję glm(), dostępne są również wszystkie zmienne wynikowe generowane przez glm().

Kolejna metoda szacowania IBNR, którą można znaleźć w pakiecie {ChainLadder}, wykorzystuje technikę symulacyjną bootstrap (ozn. BCL). Metoda BCL korzysta z dwuetapowej procedury obliczeniowej.¹¹ W pierwszym etapie szacowana jest rezerwa IBNR metodą CL i wyznaczane są reszty modelu. Następnie z uzyskanego ciągu reszt następuje losowanie ze zwracaniem i sy-

9. A.E. Renshaw, P. Verrall, *A stochastic*, op. cit.

10. P. McCullagh, J.A. Nelder, *Generalized Linear Models*, 2nd ed., Chapman and Hall, 1989.

11. P. England, R. Verrall, *Analytic and bootstrap estimates of prediction errors in claim reserving*, „Insurance: Mathematics and Economics” 25/1999, s. 281–293.

mulowany jest ciąg trójkątów szkód, dla których szacowane są rezerwy IBNR. Ostateczna wartość IBNR przyjmowana jest jako średnia wartość tych rezerw. Uzyskany ciąg rezerw służy do określenia rozkładu rezerwy.

Metoda BCL zaimplementowana jest w funkcji `BootChainLadder()` wraz z argumentami:

- `Triangle` – trójkąt szkód,
- `R` – liczba powtórzeń w losowaniu.

W wyniku otrzymywane są wartości:

- `simClaims` – ciąg trójkątów szkód,
- `IBNR.Total` – ciąg rezerw.

3. Analiza rezerwy IBNR w pakiecie `{ChainLadder}` – aspekty obliczeniowe

Do przedstawienia możliwości obliczeniowych w szacowaniu rezerwy IBNR wykorzystano dane zaczerpnięte z literatury przedmiotu¹². Trójkąty szkód PLT oraz ILT o nazwie `MCLpaid` oraz `MCLincurred` przedstawia Rysunek 1.

Rysunek 1. Trójkąty szkód stosowane w obliczeniach

```
> PersonalAutoPaid
dev
origin  1      2      3      4      5      6      7      8      9     10
1  101125 209921 266618 305107 327850 340669 348430 351193 353353 353584
2  102541 203213 260677 303182 328932 340948 347333 349813 350523   NA
3  114932 227704 298120 345542 367760 377999 383611 385224   NA   NA
4  114452 227761 301072 340669 359979 369248 373325   NA   NA   NA
5  115597 243611 315215 354490 372376 382738   NA   NA   NA   NA
6  127760 259416 326975 365780 386725   NA   NA   NA   NA   NA
7  135616 262294 327086 367357   NA   NA   NA   NA   NA   NA
8  127177 244249 317972   NA   NA   NA   NA   NA   NA   NA
9  128631 246803   NA   NA   NA   NA   NA   NA   NA   NA
10 126288   NA   NA   NA   NA   NA   NA   NA   NA   NA

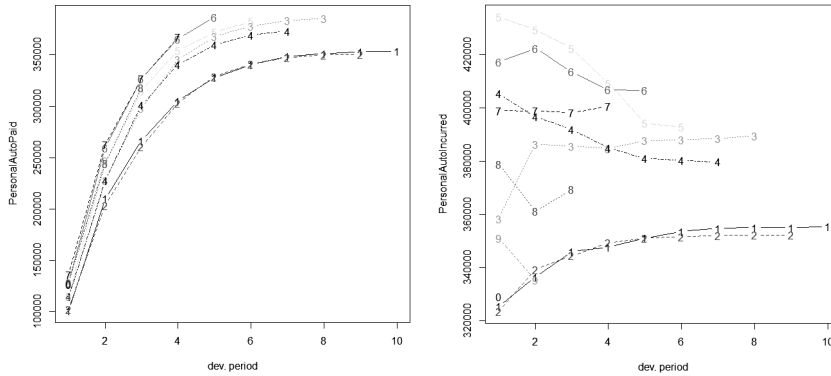
> PersonalAutoIncurred
dev
origin  1      2      3      4      5      6      7      8      9     10
1  325423 336426 346061 347726 350995 353598 354797 355025 354986 355363
2  323627 339267 344507 349295 351038 351583 352050 352231 352193   NA
3  358410 386330 385684 384699 387678 387954 388540 389436   NA   NA
4  405319 396641 391833 384819 380914 380163 379706   NA   NA   NA
5  434065 429311 422181 409322 394154 392802   NA   NA   NA   NA
6  417178 422307 413486 406711 406503   NA   NA   NA   NA   NA
7  398929 398787 398020 400540   NA   NA   NA   NA   NA   NA
8  378754 361097 369328   NA   NA   NA   NA   NA   NA   NA
9  351081 335507   NA   NA   NA   NA   NA   NA   NA   NA
10 329286   NA   NA   NA   NA   NA   NA   NA   NA   NA
```

Źródło: ekran programu R.

W pierwszym etapie przykładu wyznaczono współczynniki LDF i wykonano prezentację graficzną, co przedstawiają Wykres 1, Tabela 1 oraz Wykres 2. Do obliczenia LDF wykorzystano funkcję `ata()`.

12. Y. Zhang, *A general multivariate chain ladder model*, „Insurance: Mathematics and Economics” 46/2010, s. 588–599.

Wykres 1. Zmiana wartości trójkątów szkód PLT oraz ILT w okresach opóźnienia



Źródło: ekran programu R.

Rysunek 2. Wartości współczynników LDF dla trójkątów szkód PLT oraz ILT

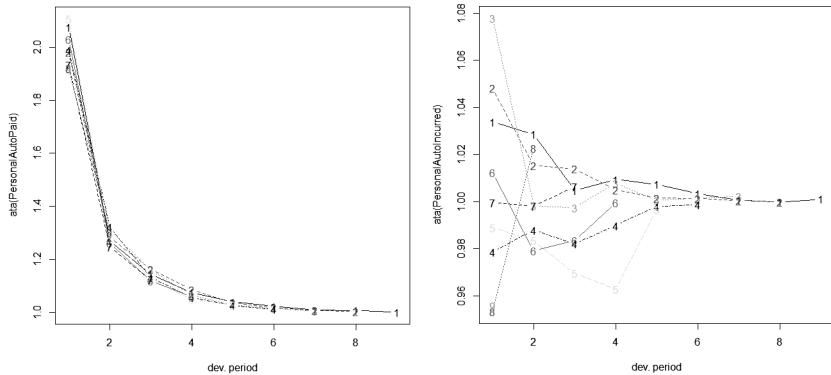
```
> ata(PersonalAutoPaid)
dev
origin 1-2 2-3 3-4 4-5 5-6 6-7 7-8 8-9 9-10
1 2.076 1.270 1.144 1.075 1.039 1.023 1.008 1.006 1.001
2 1.982 1.283 1.163 1.085 1.037 1.019 1.007 1.002 NA
3 1.981 1.309 1.159 1.064 1.028 1.015 1.004 NA NA
4 1.990 1.322 1.132 1.057 1.026 1.011 NA NA NA
5 2.107 1.294 1.125 1.050 1.028 NA NA NA NA
6 2.030 1.260 1.119 1.057 NA NA NA NA NA
7 1.934 1.247 1.123 NA NA NA NA NA NA
8 1.921 1.302 NA NA NA NA NA NA NA NA
9 1.919 NA NA NA NA NA NA NA NA NA
smpl 1.993 1.286 1.138 1.065 1.031 1.017 1.006 1.004 1.001
vwtd 1.990 1.285 1.137 1.064 1.031 1.017 1.006 1.004 1.001

> ata(PersonalAutoIncurred)
dev
origin 1-2 2-3 3-4 4-5 5-6 6-7 7-8 8-9 9-10
1 1.034 1.029 1.005 1.009 1.007 1.003 1.001 1 1.001
2 1.048 1.015 1.014 1.005 1.002 1.001 1.001 1 NA
3 1.078 0.998 0.997 1.008 1.001 1.002 1.002 NA NA
4 0.979 0.988 0.982 0.990 0.998 0.999 NA NA NA
5 0.989 0.983 0.970 0.963 0.997 NA NA NA NA
6 1.012 0.979 0.984 0.999 NA NA NA NA NA
7 1.000 0.998 1.006 NA NA NA NA NA NA
8 0.953 1.023 NA NA NA NA NA NA NA NA
9 0.956 NA NA NA NA NA NA NA NA NA
smpl 1.005 1.002 0.994 0.996 1.001 1.001 1.001 1 1.001
vwtd 1.004 1.000 0.993 0.995 1.001 1.001 1.001 1 1.001
```

Źródło: ekran programu R.

Dodatkowe wartości wynikowe dla funkcji ata() to: smpl – średnia wartość LDF dla okresu opóźnienia, vwtd – średnia ważona wartość LDF dla okresu opóźnienia.

Wykres 2. Współczynniki LDF dla trójkątów szkód PLT oraz ILT



Źródło: ekran programu R.

Następnie oszacowano rezerwę IBNR metodą CL dla danych zawartych w trójkącie PLT. W pierwszym przypadku założono, iż parametr σ szacowany jest z pomocą modelu logarytmiczno-normalnego. W drugim przypadku założono estymator Macka dla parametru σ . Jak widać na Rysunku 4, poziom rezerwy IBNR nie zmienił się, jednak zdecydowanie zwiększył się poziom współczynnika wariancji CV w każdym roku wypadkowym. Błąd modelu Mack S.E. wzrósł natomiast nieznacznie.

Rysunek 3. Wyniki dla metody CL – model log-normal

```
> CL
MackChainLadder(Triangle = PersonalAutoPaid, weights = 1, alpha = 1,
  est.sigma = "log-linear")

```

	Latest	Dev.To.Date	Ultimate	IBNR	Mack.S.E	CV(IBNR)
1	353,584	1.000	353,584	0	0	NaN
2	350,523	0.999	350,752	229	680	2.9689
3	385,224	0.995	387,054	1,830	1,522	0.8316
4	373,325	0.989	377,481	4,156	1,720	0.4140
5	382,738	0.973	393,454	10,716	2,760	0.2575
6	386,725	0.943	409,932	23,207	3,758	0.1619
7	367,357	0.887	414,305	46,948	6,351	0.1353
8	317,972	0.780	407,609	89,637	9,141	0.1020
9	246,803	0.607	406,593	159,790	12,506	0.0783
10	126,288	0.305	414,021	287,733	19,067	0.0663

```

Totals
Latest: 3,290,539.00
Dev: 0.84
Ultimate: 3,914,785.82
IBNR: 624,246.82
Mack S.E.: 29,858.91
CV(IBNR): 0.05

```

Źródło: ekran programu R.

Rysunek 4. Wyniki dla metody CL – model Mack

```
> CL1
MackChainLadder(Triangle = PersonalAutoPaid, weights = 1, alpha = 1,
  est.sigma = "Mack")

```

	Latest	Dev.To.Date	Ultimate	IBNR	Mack.S.E	CV(IBNR)
1	353,584	1.000	353,584	0	0	NaN
2	350,523	0.999	350,752	229	998	4.3544
3	385,224	0.995	387,054	1,830	1,713	0.9360
4	373,325	0.989	377,481	4,156	1,885	0.4537
5	382,738	0.973	393,454	10,716	2,872	0.2680
6	386,725	0.943	409,932	23,207	3,847	0.1658
7	367,357	0.887	414,305	46,948	6,405	0.1364
8	317,972	0.780	407,609	89,637	9,177	0.1024
9	246,803	0.607	406,593	159,790	12,532	0.0784
10	126,288	0.305	414,021	287,733	19,085	0.0663

```

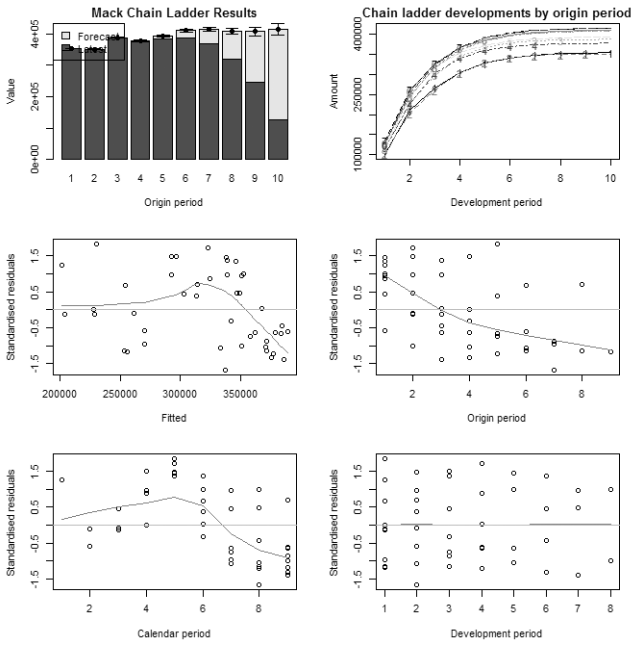
Totals
Latest: 3,290,539.00
Dev: 0.84
Ultimate: 3,914,785.82
IBNR: 624,246.82
Mack S.E.: 30,358.21
CV(IBNR): 0.05

```

Źródło: ekran programu R.

Dodatkowym wygodnym narzędziem służącym do analizy dopasowania modelu do danych jest funkcja `plot()` pozwalająca na graficzną analizę reszt.

Rysunek 5. Wykresy generowane dla metody CL



Źródło: ekran programu R.

W dalszej części przykładu wykonano szacowanie IBNR z zastosowaniem metody MCL. W tym celu korzystano z danych zawartych w trójkącie PLT oraz ILT. Wyniki przedstawia poniższy rysunek.

Rysunek 6. Wyniki dla metody MCL

```
> MCL
MunichChainLadder(Paid = PersonalAutoPaid, Incurred = PersonalAutoIncurred,
  est.sigmaP = 0.1, est.sigmaI = 0.2)

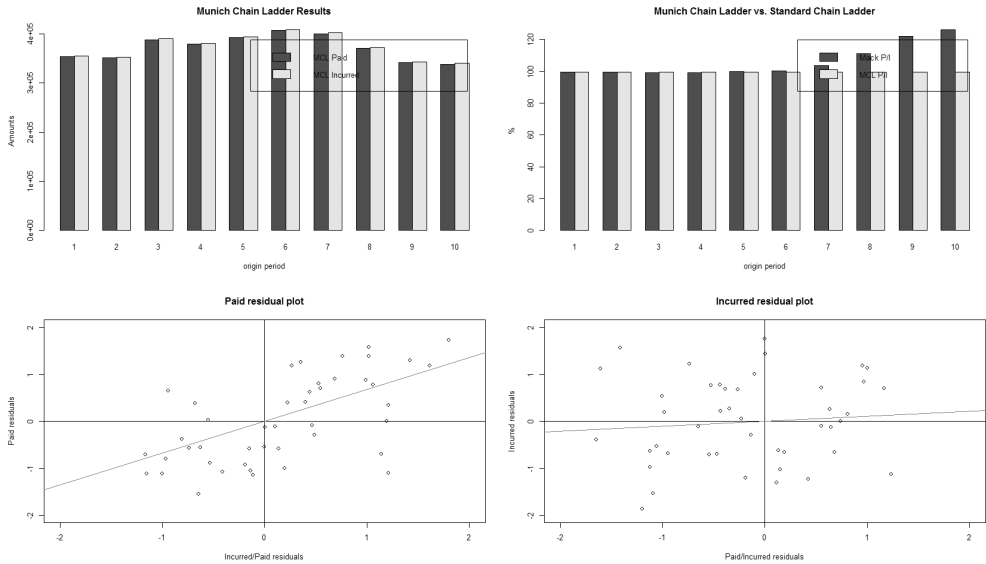
  Latest Paid Latest Incurred Latest P/I Ratio Ult. Paid Ult. Incurred Ult. P/I Ratio
1    353,584    355,363    0.995 353,584    355,363    0.995
2    350,523    352,193    0.995 350,781    352,558    0.995
3    385,224    389,436    0.989 387,875    389,681    0.995
4    373,325    379,706    0.983 378,609    380,371    0.995
5    382,738    392,802    0.974 392,252    394,124    0.995
6    386,725    406,503    0.951 406,450    408,393    0.995
7    367,357    400,540    0.917 400,212    402,144    0.995
8    317,972    369,328    0.861 370,281    372,098    0.995
9    246,803    335,507    0.736 341,172    342,875    0.995
10   126,288    329,236    0.384 338,227    339,913    0.995

Totals
      Paid Incurred P/I Ratio
Latest: 3,290,539 3,710,614    0.89
Ultimate: 3,719,441 3,737,520    1.00
```

Źródło: ekran programu R.

Za pomocą funkcji plot() można porównać metody CL oraz MCL, jak również wykonać analizę reszt modelu.

Rysunek 7. Wykres generowany dla metody MCL



Źródło: ekran programu R.

Kolejną stosowaną metodą była metoda GLM. Wyniki dla funkcji `glmReserve()` dla trójkąta PLT przedstawia Rysunek 8.

Rysunek 8. Wyniki dla metody GLM

```
> GLM
  Latest Dev.To.Date Ultimate  IBNR      S.E      CV
2  350523  0.9993471  350752    229  464.3474  2.02771788
3  385224  0.9952720  387054   1830 1187.4583  0.64888432
4  373325  0.9889902  377481   4156 1700.7849  0.40923602
5  382738  0.9727643  393454  10716 2650.8531  0.24737338
6  386725  0.9433882  409932  23207 3867.6316  0.16665797
7  367357  0.8866825  414305  46948 5552.4815  0.11826875
8  317972  0.7800907  407609  89637 8024.2919  0.08951986
9  246803  0.6070026  406593 159790 11982.6054  0.07498971
10 126288  0.3050280  414021 287733 22428.0768  0.07794753
total 2936955  0.8247089  3561202 624247 30832.5263  0.04939155
```

Źródło: ekran programu R.

Poza głównymi informacjami o wielkości rezerwy IBNR i błędach szacunku, w prosty sposób można uzyskać informacje dotyczące wyników działania standardowej funkcji `glm()`.

Rysunek 9. Wyniki dla metody GLM – parametry strukturalne modelu

```
> summary(GLMmodel)

Call:
glm(formula = value ~ factor(origin) + factor(dev), family = fam,
     data = ldaFit, offset = offset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-30.698  -14.131   -2.018   13.057   39.727

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.588524    0.040888 283.421 < 2e-16 ***
factor(origin)2 -0.008041    0.051794  -0.155 0.877489
factor(origin)3  0.090443    0.050637   1.786 0.082506 .
factor(origin)4  0.065399    0.051048   1.281 0.208340
factor(origin)5  0.106844    0.050796   2.103 0.042484 *
factor(origin)6  0.147870    0.050751   2.914 0.006105 **
factor(origin)7  0.158482    0.051518   3.076 0.003990 **
factor(origin)8  0.142187    0.053672   2.649 0.011908 *
factor(origin)9  0.139692    0.058005   2.408 0.021275 *
factor(origin)10 0.157796    0.073551   2.145 0.038740 *
factor(dev)2    -0.010061    0.029810  -0.338 0.737684
factor(dev)3    -0.566599    0.036779 -15.406 < 2e-16 ***
factor(dev)4    -1.051402    0.046362 -22.678 < 2e-16 ***
factor(dev)5    -1.682313    0.064812 -25.960 < 2e-16 ***
factor(dev)6    -2.340274    0.095993 -24.380 < 2e-16 ***
factor(dev)7    -2.933742    0.143040 -20.510 < 2e-16 ***
factor(dev)8    -3.882731    0.263858 -14.715 < 2e-16 ***
factor(dev)9    -4.315592    0.406793 -10.609 1.25e-12 ***
factor(dev)10   -6.146107    1.430109  -4.298 0.000125 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Tweedie family taken to be 472.0577)

Null deviance: 2379116 on 54 degrees of freedom
Residual deviance: 16899 on 36 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 4
```

Źródło: ekran programu R.

Ten wydruk podaje szczegółowe wartości parametrów strukturalnych modelu GLM dla okresu opóźnienia. Ponadto podawany jest średni błąd szacunku dla każdego parametru oraz wyznaczana jest wartość p-value. Widać zatem, które z parametrów są istotne w modelu, a które nie. Algorytm uzyskał zbieżność w czterech krokach.

W ostatnim etapie przykładu zastosowano metodę BCL. Wyniki generowane na ekranie to średnia rezerwa IBNR uzyskana jako średnia z wygenerowanego ciągu 99 rezerw.

Rysunek 10. Wyniki dla metody BCL

```
> BCL
BootChainLadder(Triangle = PersonalAutoPaid, R = 999, process.distr = c("gamma",
"od.pois"))

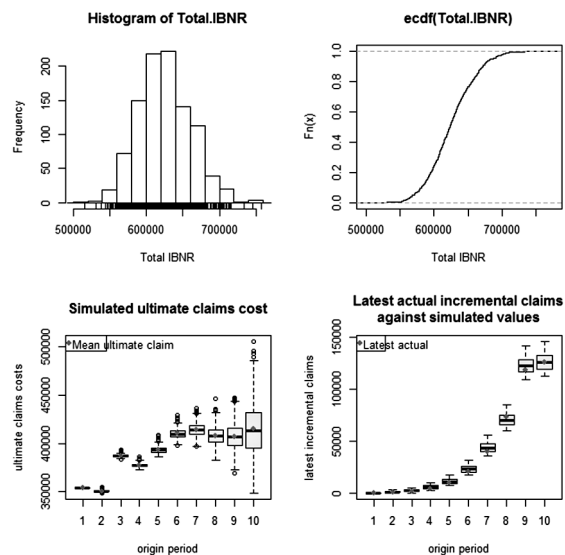
  Latest Mean Ultimate Mean IBNR SD IBNR IBNR 75% IBNR 95%
1  353,584          353,584          0          0          297  1,369
2  350,523          350,749          226          589
3  385,224          387,033          1,809          1,335          2,503          4,412
4  373,325          377,430          4,105          1,924          5,196          7,578
5  382,738          393,387          10,649          3,195          12,631          16,308
6  386,725          410,053          23,328          4,522          26,114          31,091
7  367,357          414,109          46,752          6,465          50,913          57,435
8  317,972          407,931          89,959          9,115          96,205          105,665
9  246,803          407,005          160,202          13,520          169,111          182,904
10 126,288          414,607          288,319          25,538          305,239          333,485

Totals
Latest:          3,290,539
Mean Ultimate:  3,915,887
Mean IBNR:      625,348
SD IBNR:        35,169
Total IBNR 75%: 648,446
Total IBNR 95%: 684,450
```

Źródło: ekran programu R.

Dodatkowo po zastosowaniu funkcji `plot()` widać między innymi histogram oraz dystrybuantę dla rezerwy IBNR.

Rysunek 11. Wykres generowany dla metody BCL



Źródło: ekran programu R.

Podsumowanie

W artykule przedstawiono jedynie fragment możliwości pakietu {ChainLadder}. Projekt związany z rozwojem tego oprogramowania jest kontynuowany. Zasadniczo nie jest on konkurencyjny w stosunku do komercyjnego oprogramowania stosowanego w zakładach ubezpieczeń, mimo iż posiada podobną funkcjonalność. To jednak doskonałe uzupełnienie w analizach, gdyż dodawane są nowe metody szacowania IBNR będące efektem najnowszym prac badawczych naukowców. Przykładem jest analiza zależności pomiędzy trójkątami szkód, dla różnych grup ubezpieczeniowych uwzględniana w metodzie nazwanej wielowymiarową metodą CL.

Wykaz źródeł

- Biecek P., *Przewodnik po pakiecie R.*, Wydawnictwo GIS, 2008.
- Bijak W., Pawlak M., Smętek M., *Analiza rezerw w oparciu o trójkąty szkód – badanie wstępne dla polskiego rynku ubezpieczeń*, Prace Naukowe Akademii Ekonomicznej Imienia Oskara Langego we Wrocławiu 990, 2003, s. 432–441.
- Christofides S., *Regression models based on log-incremental payments*, „Claims Reserving Manual” 2/1990, Institute of Actuaries, London.
- England P., Verrall R., *Analytic and bootstrap estimates of prediction errors in claim reserving*, „Insurance: Mathematics and Economics”, 25/1999, s. 281–293.

- Mack T., *Distribution-free calculation of the standard error of chain ladder reserve estimates*, „Astin Bulletin” 23/2, 1993, s. 213–225.
- Mack T., *The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor*, „Astin Bulletin” 29/2, 1999, s. 361–366.
- McCullagh P., Nelder J.A., *Generalized Linear Models*, 2nd ed., Chapman and Hall, 1989.
- Murphy D.M., *Unbiased Loss Development Factors*, Proceedings of the Casualty Actuarial Society Casualty Actuarial Society – Arlington, Virginia 1994, LXXXI, 154–222.
- Quarg G., Mack T., *Blätter der Deutschen Gesellschaft für Versicherungs- und Finanzmathematik*, 26/4 2004, s. 597–630.
- Pobłocka A., *Wybrane metody kalkulacji rezerwy IBNR*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu 1197, 2008, s. 368–376.
- R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Austria 2012.
- Renshaw A.E., Verrall P., *A stochastic model underlying the chain ladder technique*, „British Actuarial Journal” 4/1998, s. 903–923.
- Tweedie M.C.K., *An index which distinguishes between some important exponential families*, [in:] Ghosh J.K. Roy, J., *Statistics: Applications and New Directions*, Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Calcutta: Indian Statistical Institute, 1984, s. 579–604.
- Wieteska S., *Rezerwy techniczno-ubezpieczeniowe zakładów ubezpieczeń. Teoria i praktyka*, Branta Oficyna Wydawnicza, 2004.
- Wolny A., *Podejście aktuarialne do kalkulacji rezerwy szkodowej, Statystyczne zaawansowane metody kalkulacji rezerwy szkodowej, [w:] Metody kalkulacji ryzyka rezerw szkodowych w ubezpieczeniach majątkowo-osobowych*, Wydawnictwo Akademii Ekonomicznej w Katowicach, 2005.
- Wüthrich M.V., Merz M., *Stochastic Claims Reserving Methods in Non-Life Insurance*, John Willey & Sons, England 2008.
- Zhang Y., *A general multivariate chain ladder model*, „Insurance: Mathematics and Economics” 46/2010, s. 588–599.

Possibilities of R software in estimating the IBNR reserve

The reserve for incurred but not reported losses (IBNR) is an important item in the insurance fund of non-life insurance companies. The literature devoted to this topic abounds in methods in which the IBNR reserve can be estimated, in which various techniques of statistical modelling, data analysis and data mining are used. The review of methods used to calculate the reserve in question is widely discussed in the literature. The aim of this article is to make the reader acquainted with the possibilities of using {ChainLadder} package of R software in analysis of loss ratio development and estimation of the IBNR reserve.

Key words: IBNR reserve, Chain Ladder, GLM, bootstrap, R software.

DR ALICJA WOLNY-DOMINIAK – adiunkt w Katedrze Metod Statystyczno-Matematycznych w Ekonomii Uniwersytetu Ekonomicznego w Katowicach.