

**KWANTYFIKACJA RYZYKA  
WYPŁAT KATASTROFICZNYCH  
DLA ZDARZEŃ UBEZPIECZENIOWYCH  
Z WYKORZYSTANIEM TEORII  
WARTOŚCI EKSTREMALNYCH**

ŚLĄSKI  
PRZEGLĄD  
STATYSTYCZNY  
Nr 15(21)

Piotr Dziel, Krzysztof Hrycko

Ubezpieczeniowy Fundusz Gwarancyjny  
e-mails: pdziel@ufg.pl; khrycko@ufg.pl

ISSN 1644-6739  
e-ISSN 2449-9765

DOI: 10.15611/sps.2017.15.02  
JEL Classification: C13, C46, G22

**Streszczenie:** Artykuł poświęcony jest zagadnieniom oceny ryzyka związanego ze zdarzeniami ekstremalnymi skutkującymi wystąpieniem skrajnie wysokich wypłat odszkodowań lub świadczeń w zakładach ubezpieczeń, które prowadzą działalność w zakresie ubezpieczenia odpowiedzialności cywilnej posiadaczy pojazdów mechanicznych. Identyfikację wartości wypłat ekstremalnych, których analizę przeprowadzono z wykorzystaniem uogólnionego rozkładu wartości ekstremalnych oraz uogólnionego rozkładu Pareto, dokonano za pomocą metody bloków oraz metody przekroczeń powyżej progu. W badaniu wykorzystano dane szkodowe przesłane przez zakłady ubezpieczeń do bazy Ośrodka Informacji Ubezpieczeniowego Funduszu Gwarancyjnego. Wyniki analizy mogą być zastosowane w modelach taryfikacji w zakładach ubezpieczeń i reasekuracji oraz w działaniach organu nadzoru nad sektorem finansowym w odniesieniu do tworzenia standardów dotyczących właściwej oceny ryzyka przez podmioty rynku ubezpieczeniowego.

**Słowa kluczowe:** teoria wartości ekstremalnych, metoda Block-Maxima, metoda Excesses over Threshold, uogólniony rozkład wartości ekstremalnych, uogólniony rozkład Pareto.

## 1. Wstęp

Celem pracy jest ocena ryzyka związanego ze zdarzeniami ekstremalnymi będącymi bezpośrednią przyczyną wystąpienia skrajnie wysokich wypłat odszkodowań lub świadczeń w zakładach ubezpieczeń prowadzących działalność w zakresie ubezpieczenia odpowiedzialności cywilnej posiadaczy pojazdów mechanicznych (skrót: OC p.p.m.). Analiza z wykorzystaniem teorii wartości ekstremalnych (*Extreme Value Theory*, skrót: EVT) przeprowadzona zostanie na danych przesyłanych przez zakłady ubezpieczeń do bazy Ośrodka Informacji Ubezpieczeniowego Funduszu Gwarancyjnego (skrót: OI UFG).

W badaniu wskazane zostało, jakich wartości ekstremalnych w zakresie wypłat odszkodowań i świadczeń mogą spodziewać się zakłady ubezpieczeń w ustalonym okresie wraz ze wskazaniem wykorzystywanych parametrów dopasowanych rozkładów. Analizę empiryczną przeprowadzono dla okresów dwunasto-, dwudziestocztero- i trzydziestosześcioletnich. Pozyskana wiedza umożliwi poprawę dokładności oceny ryzyka wystąpienia oraz wolumenu szkód mających znaczny wpływ na obciążenie wyniku finansowego. Z jednej strony, możliwe jest wykorzystanie wyników badania w działalności zakładów ubezpieczeń, np. do poprawy jakości kalkulacji wysokości składki, a z drugiej strony, wyniki te mogą służyć organom sprawującym nadzór nad rynkiem finansowym do tworzenia wytycznych dotyczących właściwej oceny ryzyka towarzyszącego działalności zakładów ubezpieczeń funkcjonujących na rynku ubezpieczeń komunikacyjnych.

W tym drugim kontekście należy postrzegać konieczność posiadania przez te podmioty odpowiednich zasobów kapitałowych będących gwarancją wypłacalności. Zasoby te mają na celu zrównoważenie ryzyka ponoszonego przez ubezpieczycieli i stanowią zabezpieczenie klientów tych instytucji. Fundamentem będącym podstawą takiego rozwiązania jest zasada realności ochrony ubezpieczeniowej zagwarantowana, między innymi, przez uregulowania dotyczące gospodarki finansowej zakładu ubezpieczeń. Nietypowe uwarunkowania działalności ubezpieczycieli, w stosunku do pozostałych instytucji gospodarczych, sprawiają, że szczególnie ważne jest wypracowanie odpowiednich regulacji dotyczących ich funkcjonowania. Wymóg posiadania odpowiednich zasobów kapitałowych przez zakłady ubezpieczeń uregulowany został ustawodawstwem na poziomie Unii Europejskiej. 1 stycznia 2016 roku weszły w życie nowe oraz jednolite dla wszystkich krajów Unii standardy dotyczące wyznaczania wymogów kapitałowych oraz zarządzania ryzykiem w zakładach ubezpieczeń w związku ze zmianą regulacyjną nazywaną Solvency II [Dyrektywa Parlamentu Europejskiego i Rady 2009/138/WE; Dyrektywa Parlamentu Europejskiego i Rady 2013/58/UE]. Formuły standardowe wyznaczania wysokości kapitałowego wymogu wypłacalności w tej dyrektywie zawierają również część związaną z ryzykiem katastroficznym, uwzględniającym wystąpienie szkód ekstremalnych oraz ich akumulację. Niemniej przede wszystkim wymogi te powinny zapewnić, że przychody zakładu, których podstawowym składnikiem jest składka, będą gwarantowały wypłacalność.

W przeszłości teoria wartości ekstremalnych związana była z rozkwiem takich nauk, jak astronomia, inżynieria czy hydrologia. Stosowano ją w analizie procesów fizycznych, np. prędkości wiatru, wielkości opadu atmosferycznego, wysokości fali powodziowej. Tego typu analizy są prowadzone również obecnie i mają na celu tworzenie odpowiednich mechanizmów zabezpieczających przed skutkami występowania wartości ekstremalnych tych zjawisk. Początki teorii datuje się na lata trzydzieste XX wieku. Podstawę stanowią tutaj prace von Bortkiewicza, von Misesa, Dodda, Fishera, Tippetta czy Fréchet'a. Z kolei kluczowe wyniki, które przez lata stanowiły podstawę dla aplikacji EVT w ramach nauk inżynierskich, stanowiła pozycja *Statistics of Extremes* [Gumbel 2004] napisana w 1958 roku przez Emila Juliusa Gumbela, niemieckiego matematyka. Obecnie istnieje wiele monografii dotyczących teorii wartości ekstremalnych [Faldziński 2014, s. 19-20].

W opracowaniu położono szczególny nacisk na aspekt praktyczny wykorzystania informacji w kontekście wskazywanej teorii. W pracy przedstawiono:

- zastosowanie koncepcji tzw. metody bloków (*block maxima*) oraz metody przekroczeń powyżej progu (*excess over threshold* lub *peaks over threshold*),
- estymację rozkładów: uogólnionego rozkładu wartości ekstremalnych (*generalized extreme value distribution*), uogólnionego rozkładu Pareto (*generalized Pareto distribution*).

Uwzględnienie powyższych zagadnień ma na celu próbę jak najlepszego opisu tzw. wypłat ekstremalnych.

## 2. Analiza danych z wykorzystaniem teorii wartości ekstremalnych

### 2.1. Opis danych

W celu przeprowadzenia analizy obróbce poddano dane zawarte w bazie danych Ośrodka Informacji UFG, dotyczące zdarzeń, wypłaconych odszkodowań oraz świadczeń w zakresie ubezpieczenia OC p.p.m. Analizowano zdarzenia, które zaistniały w latach 2013-2015 oraz zostały zgłoszone do momentu przeprowadzanej analizy. Trzyletni okres analizy powinien zapewnić dobrą reprezentatywność danych. Celowe pominięcie zdarzeń z 2016 r. wynika z tego, że dane za ten okres mogą nie uwzględniać wielu szkód zaistniałych i niezgłoszonych przez poszkodowanych do zakładów ubezpieczeń

(*claims incurred but not reported*) lub takich, które „nie rozwinęły się” dostatecznie. Dla celów analizy wybrano wyłącznie zdarzenia, w związku z którymi wypłacone zostało odszkodowanie lub świadczenie niebędące rentą, a zatem mające formę jednorazowej wypłaty lub kolejnych dopłat za szkodę w mieniu lub na osobie. Analizowane zdarzenia nie pochodziły z umów flotowych, a wszystkie wartości odszkodowania lub świadczenia, które dotyczyły jednego zdarzenia, zostały zsumowane (łączna wartość wypłat dla pojedynczego zdarzenia).

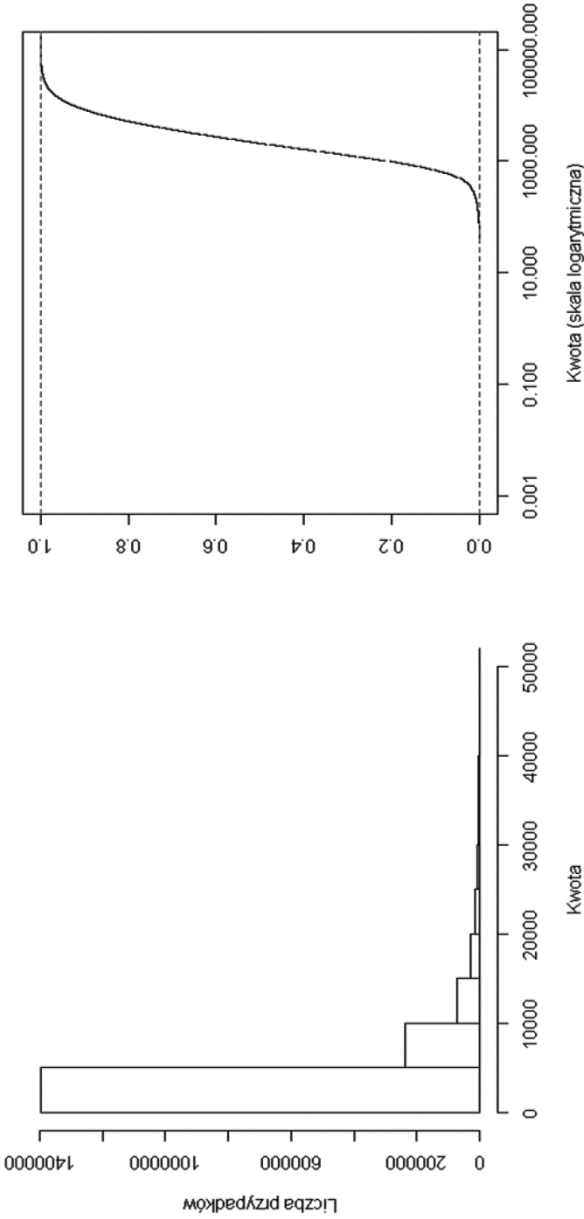
Podstawowe statystyki opisowe dla badanego zbioru wskazano w tab. 1. W badanym przypadku analizie poddano ponad 1,7 miliona obserwacji. Średnia wartość odszkodowania wyniosła 3 827,35 zł, z kolei wartość maksymalna to ponad 2,67 mln zł. Dodatnia wysoka wartość skośności i kurtozy oznacza odpowiednio, że rozkład wypłat ma prawy ogon grubszy od lewego oraz prawy ogon grubszy niż w przypadku rozkładu normalnego. W związku z powyższym w kolejnym podrozdziale nastąpi próba dopasowania kilku rozkładów gruboogonowych oraz asymetrycznych.

**Tabela 1.** Statystyki opisowe badanego zbioru

Liczba obserwacji	1 764 102
Minimum	40 zł
Maksimum	2 672 069 zł
Średnia	3 827,35 zł
Dolny kwartył	1 105 zł
Mediana	2 090 zł
Górny kwartył	4 315 zł
95. percentyl	12 355,83 zł
99. percentyl	25 941,80 zł
Skośność	65,58
Kurtoza	17 918,26

Źródło: opracowanie własne.

Rysunek 1 przedstawia histogram oraz dystrybuantę empiryczną rozkładu wartości wypłat. Na histogramie przede wszystkim uwidacznia się wysoka koncentracja wypłat w lewym ogonie. Wartości w prawym ogonie są tak rzadkie, że ze względu na skalowanie wykresu niezauważalne i w związku z tym oś odciętych została ograniczona do 50 tys. zł. Dystrybuanta empiryczna przedstawia skumulowane częstości obserwacji powyżej zadanego progu – dla kwoty wypłaty wykorzystana została skala logarytmiczna.



**Rys. 1.** Histogram oraz dystrybuanta empiryczna dla wartości wypłat  
Źródło: opracowanie własne.

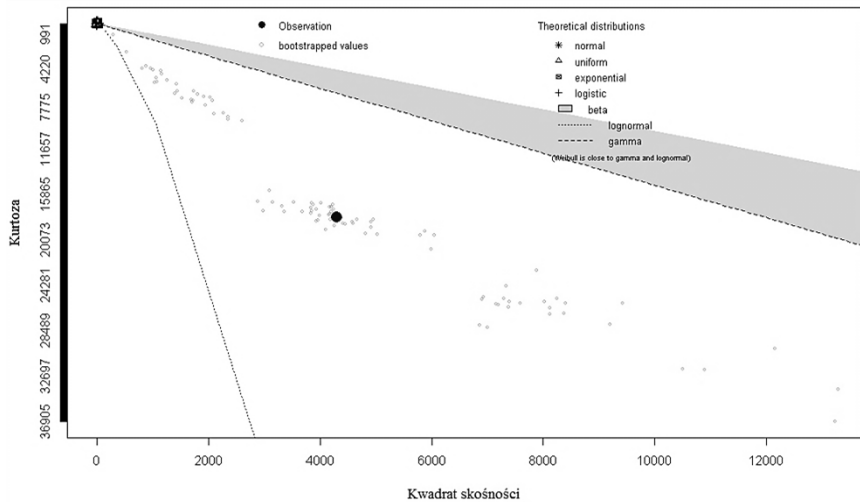
## 2.2. Wstępny dobór rozkładu

Dopasowanie rozkładu analitycznego do badanej zmiennej umożliwia wyznaczenie jej podstawowych charakterystyk. W odniesieniu do analizowanego zagadnienia pozwala to na wskazanie np. średniej wartości wypłat ekstremalnych powyżej zadanego progu czy wyznaczenie prawdopodobieństwa wystąpienia wypłaty większej niż konkretnie obrany poziom. Z kolei wykorzystanie dystrybuanty empirycznej, ze względu na analizę często ograniczonej próby, umożliwia wyłącznie pewne przybliżenie zjawiska. Rozwiązanie analityczne pozwoli na efektywną kwantyfikację ryzyka związanego z ustaleniem wysokości wypłat ekstremalnych.

Ze względu na wnioski z podrozdziału 2.1 dotyczące lewostronnej skośności oraz występowania tzw. grubych ogonów rozkładu nastąpiła próba dopasowania kilku rozkładów gruboogonowych. Dopasowania dokonano dla rozkładów: Weibulla, gamma i lognormalnego z wykorzystaniem metody największej wiarygodności. Estymatory wyznaczone tą metodą mają pewne pożądane własności i stąd ich uniwersalne zastosowanie. Wskazywane dobre własności, przy pewnych technicznych założeniach, które są na ogół spełnione, to przede wszystkim zgodność oraz nieobciążoność estymatora [*Rachunek prawdopodobieństwa...*]. Dodatkowo procedury dotyczące maksymalizacji funkcji wiarygodności mogą również być zastosowane do bardziej złożonych modeli, biorących pod uwagę również trend czy inne efekty [Embrechts, Klüppelberg, Mikosch 1996, s. 321]. Ostateczną jakość dopasowania sprawdzono za pomocą testu Kołmogorowa–Smirnowa, a do weryfikacji wykorzystano dane empiryczne oraz rozkłady wygenerowane dla parametrów uzyskanych metodą największej wiarygodności.

Rysunek 2 został wykonany pomocniczo w celu sprawdzenia, które rozkłady można wykorzystać w celu jak najlepszego dopasowania do danych. Jest to tzw. wykres Cullena–Freya, który przedstawia obszary dopasowania poszczególnych rozkładów w uzależnieniu od kwadratu skośności oraz kurtozy [Delignette-Muller, Dutang].

Wykres należy traktować nie jako ostateczne wskazanie możliwych rozkładów, a wyłącznie jako jedną z metod rekomendacji rozkładu opisującego zmienną. Wartości kurtozy oraz kwadratu skośności dla badanych danych zostały oznaczone większym punktem. Dodatkowo sprawdzono 100 bootstrapowych prób losowanych z danych poddanych analizie, dla których odpowiadające im wartości kurtozy oraz skośności



**Rys. 2.** Wykres Cullena–Freya

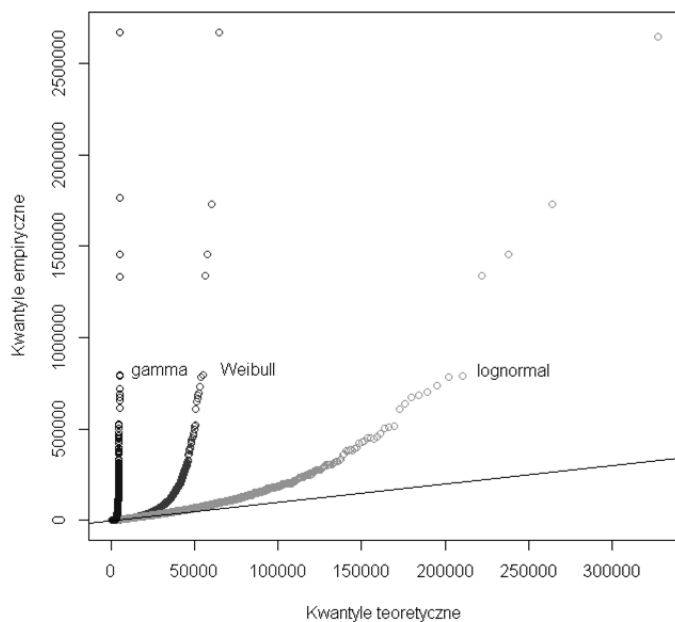
Źródło: opracowanie własne.

zostały oznaczone mniejszymi punktami<sup>1</sup>. Z powyższego wykresu można odczytać, że badany rozkład empiryczny mógłby zostać przybliżony rozkładem lognormalnym, gamma lub Weibulla. Podobne wnioski można wyciągnąć w odniesieniu do prób bootstrapowych.

Rysunek 3 pokazuje dopasowanie rozkładów dla parametrów wyznaczonych metodą największej wiarygodności. Linia  $y = x$  wskazuje na równość kwantyl empirycznych oraz teoretycznych. Wykres kwantyl–kwantyl wskazuje, że kwantyle teoretyczne dopasowanych rozkładów znacznie różnią się od kwantyl empirycznych. Charakterystyczne wypukłe odchylenia od kwantyl teoretycznych wskazują, że mamy do czynienia z rozkładami gruboogonowymi. Kwantyle teoretyczne dla dopasowanych rozkładów przyjmują wartości maksymalne dla wypłat powyżej 300 tys. zł, z kolei obserwacje wskazują, że maksymalna wartość wypłaty wynosi ponad 2,5 mln zł.

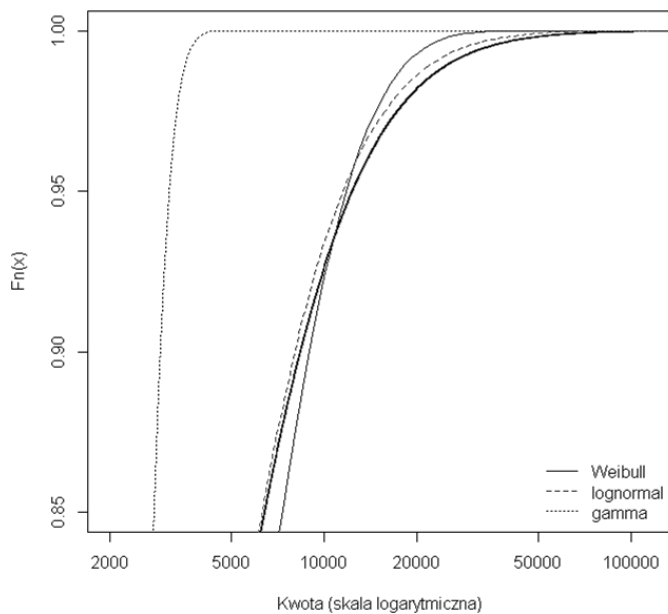
Na rysunku 4 przedstawiono dystrybuantę empiryczną rozkładu wraz z dopasowanymi dystrybuantami teoretycznymi. Szczególnie interesują nas w tym przypadku wartości w prawym ogonie rozkładu. Z tego względu wykres uwzględnia dopasowanie od 85. do 100. percentyla. Na podstawie wykresu można stwierdzić, że dopasowane

<sup>1</sup> Próba bootstrapowa to próba, która jest uzyskana w wyniku losowania ze zwracaniem z pierwotnej próby danych. Zob. [Efron, Tibshirani 1993, s. 12].



**Rys. 3.** Wykres kwantyl–kwantyl dla dopasowanych rozkładów

Źródło: opracowanie własne.



**Rys. 4.** Dystrybuanta empiryczna rozkładu oraz dopasowane dystrybuanty teoretyczne

Źródło: opracowanie własne.



rozkłady nie opisują dobrze ogona rozkładu wypłat. Najlepiej dopasowany do danych jest rozkład lognormalny. Niemniej również w przypadku tego rozkładu wypłaty w ogonie są niedoszacowane.

Weryfikacji tak dopasowanych rozkładów dokonano również za pomocą testu Kołmogorowa–Smirnowa. W teście tym porównywane są dystrybuanty: empiryczna z rozkładem dopasowanym. Najmniejszą wartość statystyka  $D$  testu przyjmuje w przypadku porównania rozkładu empirycznego z rozkładem lognormalnym ( $D = 0,0313$ ), co wskazuje na najlepsze dopasowanie rozkładów. Dla wszystkich badanych rozkładów hipoteza zerowa testu, mówiąca o tym, że obserwacje pochodzą z takiego samego rozkładu, została odrzucona już na poziomie istotności równym 0,01. Wyniki zostały przedstawione w tab. 2.

**Tabela 2.** Test Kołmogorowa–Smirnowa dla rozkładu empirycznego oraz rozkładów dopasowanych

Rozkład teoretyczny	Wartość $D$ testu	$p$ – wartość
Weibulla	0,0945	<0,001
Gamma	0,3365	<0,001
Lognormalny	0,0313	<0,001

Źródło: opracowanie własne.

Rozkłady Weibulla, gamma oraz lognormalny, których parametry oszacowano metodą największej wiarygodności, nie przedstawiają dobrego dopasowania do danych empirycznych. Wstępnego potwierdzenia tego faktu dokonano za pomocą wykresów kwantyl–kwantyl oraz dystrybuanty rozkładu dla danych empirycznych i teoretycznych. Ostatecznego potwierdzenia dokonano za pomocą testu Kołmogorowa–Smirnowa. Należy stwierdzić, że niezbędne jest dopasowanie rozkładów, które będą lepiej opisywały prawy ogon wartości wypłat. Zostanie to przedstawione w dalszych podrozdziałach i umożliwi wskazanie charakterystyk wysokości wypłat ekstremalnych, na które będą narażone zakłady ubezpieczeń oferujące ubezpieczenie OC p.p.m.

### 2.3. Teoria wartości ekstremalnych i jej zastosowanie do badania wartości wypłat

Podobnie jak centralne twierdzenie graniczne jest podstawą dla klasycznej teorii statystyki, tak filarem teorii wartości ekstremalnych jest twierdzenie Fishera i Tippeta [Fisher, Tippet 1928, s. 180-190], które

wskazuje, jaki rozkład asymptotyczny mają ekstrema zmiennych losowych.

### Twierdzenie 1 (twierdzenie Fishera i Tippeta)

Niech  $X_n$  będzie ciągiem zmiennych losowych i.i.d. i  $M_n = \max(X_1, \dots, X_n)$ . Jeśli istnieją stałe normujące (*norming constants*)  $c_n > 0$  i  $d_n \in R$  i pewna niezdegenerowana dystrybuanta rozkładu  $G$  taka, że:

$$\frac{M_n - d_n}{c_n} \xrightarrow{d} G \text{ jak } n \rightarrow \infty,$$

wtedy  $G$  należy do jednej z poniższych trzech dystrybuant rozkładów:

$$\text{Fréchet } \Phi_\alpha(x) = \begin{cases} 0 & x \leq 0 \\ \exp\{-x^{-\alpha}\} & x > 0 \end{cases} \alpha > 0,$$

$$\text{Weibull } \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\} & x \leq 0 \\ 1 & x > 0 \end{cases} \alpha > 0,$$

$$\text{Gumbel } \Lambda(x) = \exp\{-e^{-x}\}, x \in R.$$

Rozkłady  $\Phi_\alpha, \Psi_\alpha, \Lambda$  nazywane są standardowymi rozkładami wartości ekstremalnych, a zmienne losowe tych rozkładów nazywane są standardowymi ekstremalnymi zmiennymi losowymi. Należy zauważyć, że twierdzenie to nie zapewnia istnienia niezdegenerowanej dystrybuanty rozkładu ani nie wskazuje, który typ rozkładu zastosować, jeśli dystrybuanta istnieje. W praktyce zatem, ze względu na to, że dopasowanie trzech różnych rozkładów wiąże się z dużą niedogodnością, zastępuje się je jednym rozkładem ze wskazaną poniżej parametryzacją.

Uogólniony rozkład wartości ekstremalnych (*Generalized Extreme Value Distribution, GEV*), oznaczany jako  $G(\mu, \sigma, \xi)$ , ma dystrybuantę zadaną wzorem:

$$G(x; \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]_+^{-\frac{1}{\xi}} \right\},$$

gdzie  $a_+ = \max(0, a)$ . W sytuacji, gdy  $\xi = 0$ , wtedy dystrybuanta nie jest określona. Wówczas wykorzystujemy limit, przy  $\xi \rightarrow 0$  i dystrybuanta jest zadaną wzorem:

$$G(x; \mu, \sigma) = \exp \left\{ -\exp \left( \frac{x - \mu}{\sigma} \right) \right\}.$$

Parametry  $\mu$  ( $-\infty < \mu < \infty$ ),  $\sigma$  ( $\sigma > 0$ ),  $\xi$  ( $-\infty < \xi < \infty$ ) są odpowiednio parametrami położenia, skali oraz kształtu. Wartość para-

metru  $\xi$  odpowiada za jeden z trzech typów rozkładów wymienionych w twierdzeniu: 1:  $\xi = 0$  odpowiada rozkładowi Gumbela,  $\xi > 0$  rozkładowi Fréchet'a,  $\xi < 0$  rozkładowi Weibulla.

Ze względu na to, że  $M_n \xrightarrow{d} G^*(\mu^*, \sigma^*, \xi)$  jak  $n \rightarrow \infty$ , w praktyce estymacji parametrów dokonuje się dla wartości nieznormalizowanych, czyli bezpośrednio dla zbioru wartości ekstremalnych. Przy założeniu, że

$$\Pr\{(M_n - b_n)/a_n \leq z\} \approx G(z), \text{ dla dużych } n,$$

wówczas

$$\begin{aligned} \Pr\{M_n \leq z\} &\approx G\{(z - b_n)/a_n\}. \\ &= G^*(z). \end{aligned}$$

### 2.3.1. Metoda bloków

Należy wskazać, że dopasowania rozkładu GEV dokonuje się dla maksimumów wybranych tzw. metodą bloków. Metoda ta jest najstarszą metodą estymacji wykorzystywaną w teorii wartości ekstremalnych. Kolejne kroki metody przedstawiają się następująco [Faldziński 2014, s. 63]:

1. Daną próbę  $X_1, \dots, X_n$ , dzieli się na  $j$  bloków o długości  $k$  (grupowanie najczęściej odbywa się w ramach obserwacji z jednego miesiąca lub roku).

2. Dla każdego bloku liczone jest maksimum. Uzyskujemy zatem łącznie  $j$  największych wartości ze wszystkich bloków (oznaczymy  $M_j$ ).

3. Do szeregu maksimumów  $M_j$  dopasowywany jest rozkład GEV.

Dopasowania rozkładu GEV dokonamy, podobnie jak we wcześniejszych przypadkach, za pomocą metody największej wiarygodności. Logarytm funkcji wiarygodności jest poniższej postaci:

$$\begin{aligned} l(x; \mu, \sigma, \xi) = & -j \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^j \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]_+ - \\ & \sum_{i=1}^j \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}. \end{aligned}$$

Korzystając z metody bloków, wyznaczono maksymalne wartości wypłat dla kolejnych miesięcy. Statystyki opisowe zbioru zostały wskazane w tab. 3. Wartości przyjmowane przez obserwacje znajdują się w zakresie od około 167 tys. zł do ponad 2,67 mln zł, ze średnią około 551 tys. zł. Dodatnia wartość skośności i kurtozy pozwala na wysnuć takich samych wniosków jak w przypadku analizy zbioru

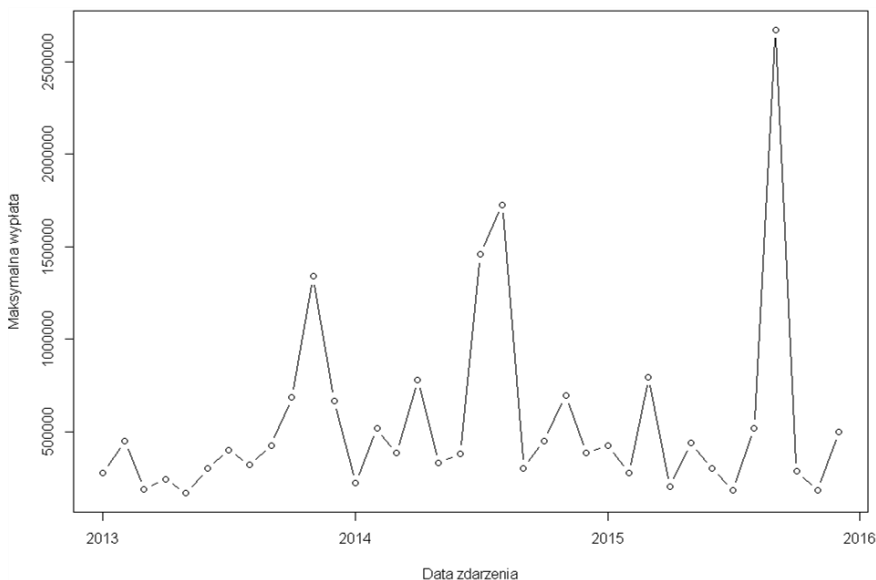
**Tabela 3.** Statystyki opisowe dla zbioru wyznaczonego metodą bloków

Liczba obserwacji	36
Minimum	167 232,30 zł
Maksimum	2 672 069 zł
Średnia	551 881,08 zł
Dolny kwartyl	283 277,94 zł
Mediana	392 092,50 zł
Górny kwartyl	556 974,94 zł
95. percentyl	1 525 064 zł
99. percentyl	2 341 250 zł
Skośność	2,52
Kurtoza	6,68

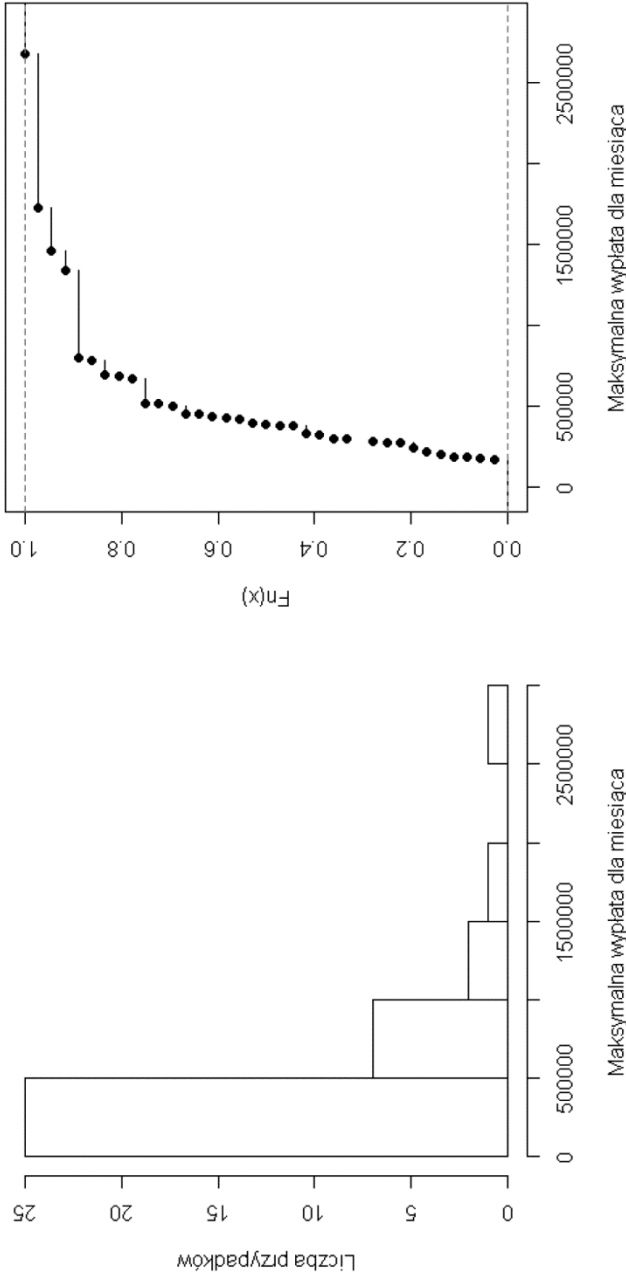
Źródło: opracowanie własne.

wszystkich obserwacji. Niemniej wartości tych statystyk nie są tak wysokie jak w przypadku analizy wszystkich obserwacji.

Na rysunku 5 przedstawiono wykres ze wszystkimi obserwacjami wyznaczonymi metodą bloków. W tym momencie warto wskazać, że jedną z wad metody bloków jest pomijanie wartości ekstremalnych, które znajdują się w tym samym bloku, ale mają mniejszą wartość. Dla przykładu w sytuacji, gdyby we wrześniu 2015 r., oprócz skrajnie wysokiej wartości odszkodowania równej około 2,67 mln zł wypłacono

**Rys. 5.** Obserwacje wyznaczone metodą bloków

Źródło: opracowanie własne.



Rys. 6. Histogram oraz dystrybuanta empiryczna obserwacji wyznaczonych metodą bloków  
Źródło: opracowanie własne.

odszkodowanie w wysokości równej 1 mln zł, przypadek taki nie byłby brany pod uwagę. Może się zatem okazać, że w sytuacji wyznaczenia maksymalnych wartości metodą bloków część ważnych obserwacji zostanie pominięta.

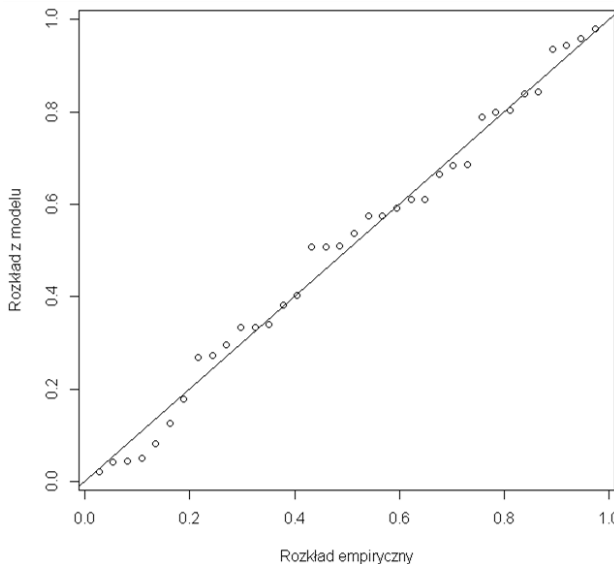
Rysunek 6 przedstawia histogram oraz dystrybuantę empiryczną dla obserwacji wyznaczonych metodą bloków. Poszczególne wartości obserwacji nie są rozłożone równomiernie – najwięcej wypłat wybranych tą metodą znajduje się kwocie do 500 tys. zł.

Z wykorzystaniem metody największej wiarygodności wyznaczone zostały wartości poszczególnych parametrów rozkładu GEV wraz z odpowiadającymi im błędami standardowymi:

$$\hat{\mu} = 314\,266,4 \quad (30\,980,59) \quad \hat{\sigma} = 156\,547,7 \quad (30\,895,59) \quad \hat{\xi} = 0,58 \quad (0,2).$$

Należy zauważyć, że próba dopasowania jednego z rozkładów GEV nastąpiła już w podrozdziale 2.2. Zgodnie z informacją na temat rozkładu GEV jest on rozkładem Weibulla, gdy parametr kształtu  $\xi$  przyjmuje wartość mniejszą od 0, co nie ma miejsca w tym przypadku i stąd słabe dopasowanie do danych empirycznych. Z drugiej, strony próba dopasowania nastąpiła do wszystkich danych, a rozkład GEV jest tzw. rozkładem dla maksimumów.

Na rysunku 7 przedstawiono dopasowanie dystrybuanty rozkładu empirycznego oraz rozkładu z modelu dla parametrów wskazanych



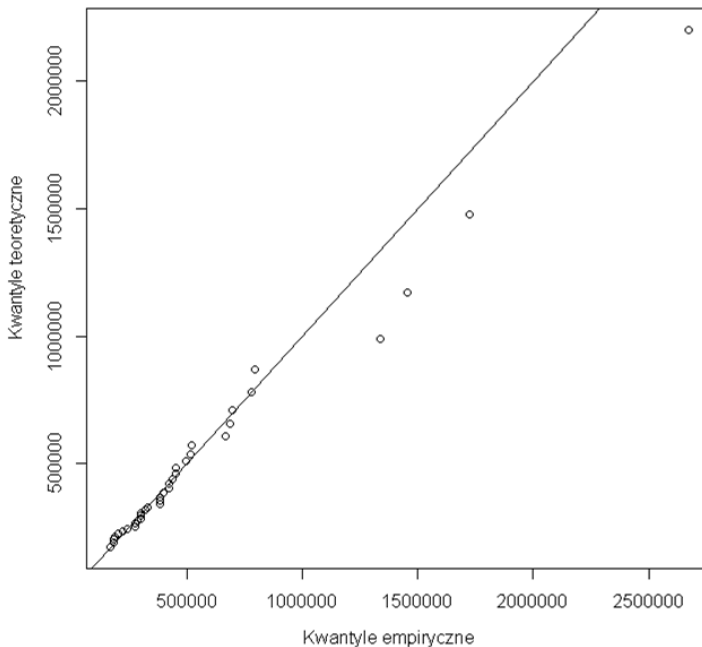
**Rys. 7.** Dopasowanie dystrybuanty empirycznej oraz dystrybuanty z modelu rozkładu GEV

Źródło: opracowanie własne.

powyżej. Wartości dystrybuanty z modelu układają się wzdłuż linii  $y = x$ , można zatem przyjąć, że rozkład GEV, którego parametry zostały wyznaczone z wykorzystaniem metody największej wiarygodności, przedstawia dobre dopasowanie do danych empirycznych.

Dodatkowo sprawdzono dopasowanie, wykorzystując test Kołmogorowa–Smirnowa, który przeprowadzono, porównując rozkład empiryczny danych z rozkładem teoretycznym dla wygenerowanych 1000 obserwacji w 1000 replikacji. Średnia  $p$  – wartość testu wyniosła 0,91, z odchyleniem standardowym wynoszącym 0,08. Daje to dobre podstawy do nieodrzućania hipotezy zerowej mówiącej o równości badanych rozkładów.

Na rysunku 8 znajduje się wykres przedstawiający kwantyle z rozkładu teoretycznego oraz rozkładu empirycznego. Na rysunku widoczna jest dobre dopasowanie modelu do danych empirycznych dla wartości wypłat poniżej 1 mln zł i słabe dopasowanie dla wypłat powyżej tego progu.



Rys. 8. Wykres kwantyl–kwantyl dla dopasowanego rozkładu GEV

Źródło: opracowanie własne.

Dopasowany rozkład GEV stanowi źródło informacji, na podstawie którego można wskazać ważną z punktu widzenia zastosowania

charakterystykę, nazywaną poziomem zwrotu (*return level*) [Faldziński 2014, s. 64]. Określony poziom zwrotu, oznaczany  $R_t$  z okresem zwrotu  $t$ , jest kwantylem rozkładu maksimów  $M_j$ , który spełnia własność:

$$Pr(M_j > R_t) = \frac{1}{t}, \text{ gdzie } \frac{1}{t} = p \in (0,1).$$

$R_t$  jest to poziom zwrotu, który oczekujemy, że zostanie przekroczony średnio tylko raz co  $t$  okresów. Blok, w którym nastąpi przekroczenie poziomu zwrotu, nazywany jest okresem stresu (*stress period*)<sup>2</sup>.

W analizowanym przypadku do weryfikacji wykorzystano miesięczne okresy, dla których wyznaczono maksima. Należy zatem założyć, że jeśli  $t = 12$ , to w takiej sytuacji wyznaczony zostanie poziom zwrotu, który zostanie przekroczony raz w okresie 12 miesięcy. Przedział ufności dla tak wyznaczonej wartości można oszacować, stosując metodę delta [Coles 2001, s. 32-33].  $R_t$  jest kwantylem rozkładu GEV, który można wyrazić wzorem analitycznym:

$$R_t = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[ \left( -\log \left( 1 - \frac{1}{t} \right) \right)^{-\hat{\xi}} - 1 \right], & \hat{\xi} \neq 0, \\ \hat{\mu} + \hat{\sigma} \left[ -\log \left( 1 - \frac{1}{t} \right) \right], & \hat{\xi} = 0. \end{cases}$$

Wartości poziomu zwrotu dla kilku okresów analizy wraz z oszacowaniami błędu statystycznego wskazano w tab. 4. Interpretacja tych wyników jest następująca: biorąc pod uwagę dane wszystkich zakładów ubezpieczeń prowadzących działalność w zakresie ubezpieczenia OC p.p.m. w Polsce, można stwierdzić, że raz do roku zdarzy się wypadek ubezpieczeniowy, którego konsekwencje finansowe przekroczą około 1,15 mln zł. Analogicznie w ciągu 36 miesięcy co najmniej raz wystąpi wypadek, dla którego wartość wypłaty przekroczy 2,16 mln zł. Wraz ze wzrostem długości okresu, dla którego wyznaczana jest ta wartość, wzrasta wartość błędu statystycznego. W odniesieniu do sytuacji praktycznej możemy wskazać, że wraz ze wzrostem odległości horyzontu czasowego analizy wzrasta niepewność związana z wypłatą ekstremalnie wysokiego odszkodowania lub świadczenia. Należy

<sup>2</sup> Należy wskazać, że pojęcie to związane jest również z zagadnieniami tzw. stress testów wykorzystywanych przez regulatorów rynku finansowego do pomiaru ryzyka towarzyszącego instytucjom ekonomicznym w sytuacji wystąpienia skrajnie negatywnych warunków rynkowych. Zob. również [Sokół 2012, s. 434-439].



również zauważyć, że otrzymane poziomy wypłat ekstremalnych mogą być niedoszacowane, na co wskazywałby wykres kwantyl–kwantyl na rys. 8.

**Tabela 4.** Poziomy zwrotu oraz błędy statystyczne dla poszczególnych okresów zwrotu wyznaczone za pomocą metody bloków

Okres analizy	12 miesięcy	24 miesiące	36 miesięcy
Wartość zwrotu	1 152 128	1 718 028	2 167 794
Błąd statystyczny	312 084	645 930	960 651

Źródło: opracowanie własne.

Jak zostało wcześniej wskazane, metoda bloków wykorzystuje wartości maksymalne w obrębie jednego okresu. W takiej sytuacji część obserwacji ekstremalnych może zostać pominięta. W kolejnym podrozdziale zostanie przedstawiona metoda alternatywna, w przypadku której to zagadnienie zostało wyeliminowane.

### 2.3.2. Metoda przekroczeń powyżej progu

Metoda zakłada, że dany jest szereg badanych obserwacji  $X_1, \dots, X_n$  i.i.d. z nieznaną dystrybuantą  $F$ . Interesują nas przekroczenia powyżej zadanej wysokiej wartości progowej  $u$ .

#### Twierdzenie 1 (twierdzenie Pickandsa–Balkemy–de Hanna)

Dla szerokiej klasy rozkładów danych rzeczywistą dystrybuantą  $F$  warunkowy rozkład przekroczeń  $F_u(y)$ , dla dużej wartości  $u$ , jest dobrze aproksymowany za pomocą  $F_u(y) \approx G_{\xi, \sigma}(y)$ ,  $u \rightarrow \infty$ , gdzie:

$$G_{\xi, \sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0, \end{cases}$$

oraz  $y \in [0, (x_F - u)]$ , jeżeli  $\xi \geq 0$ , i  $y \in \left[0, -\frac{\sigma}{\xi}\right]$ , jeżeli  $\xi < 0$  ( $x_F$  jest skończonym lub nieskończonym kresem górnym dystrybuanty  $F$ ).  $G_{\xi, \sigma}(y)$  jest uogólnionym rozkładem Pareto.

Schemat metody przekroczeń powyżej progu przedstawia się następująco [Fałdziński 2014, s. 81-82]:

1. Dla danego szeregu zmiennych  $X_1, \dots, X_n$  należy wybrać wartość progową  $u$ . Niech  $N_u$  będzie liczbą obserwacji przekraczających  $u$  ( $X_{i_1}, \dots, X_{i_{N_u}}$ ). Obserwacje należy zapisać jako  $Y_j = X_{i_j} - u \geq 0$ .

2. Kolejnym etapem jest dopasowanie rozkładu  $G_{\xi, \sigma}(y)$  do przekroczeń  $Y_{i_1}, \dots, Y_{N_u}$ , aby otrzymać oszacowanie parametrów  $\xi$  i  $\sigma$ .

W celu estymacji parametrów rozkładu, podobnie jak we wcześniejszych przypadkach, wykorzystana zostanie metoda największej wiarygodności. Logarytm funkcji wiarygodności jest postaci:

$$l(\mathbf{y}; \sigma, \xi) = -N_u \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{N_u} \log\left(1 + \frac{\xi y_i}{\sigma}\right)_+.$$

Podobnie jak w przypadku metody bloków również tutaj dopasowania dokonuje się do ogona rozkładu. Należy zatem w odpowiedni sposób wybrać wartość progu  $u$ . Wszystkie wartości obserwacji, które będą poniżej tego progu, zostaną zignorowane, a pozostałe posłużą do estymacji parametrów. W omawianym przypadku do wskazania progu  $u$  wykorzystana zostanie metoda, która w literaturze anglojęzycznej określana jest jako *Mean Excess Plot*<sup>3</sup> (wykres średnich powyżej progu). Zgodnie z tą metodą:

$$E(X - u | X > u) = \frac{\sigma_{u_0} + \xi u}{1 - \xi},$$

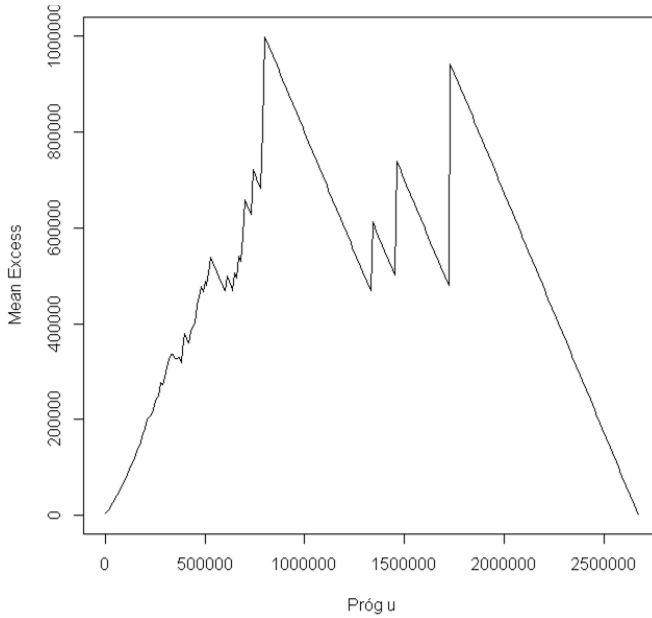
gdzie:  $\sigma_{u_0}$  oznacza parametr skali, dla obserwacji powyżej progu  $u$  ( $u > u_0$ ).

Jak wynika z powyższego równania, wartość oczekiwana dla obserwacji pomniejszych o próg  $u$  i tych, które jednocześnie przekraczają próg  $u$ , jest liniowa względem tego progu.

Rysunek 9 przedstawia wykres średnich powyżej progu. Wykres jednoznacznie nie wskazuje, jaką wartość progu  $u$  można byłoby wybrać, by dopasować odpowiedni uogólniony rozkład Pareto. Dodatkowo należy pamiętać, że wybór zbyt niskiego progu będzie powodował, iż dopasowanie rozkładu odbędzie się do danych, które w obserwacji nie będą przedstawiały wartości ekstremalnych. Z drugiej strony, zbyt wysoki próg będzie powodował, że liczba obserwacji wykorzystywanych do dopasowania rozkładu będzie zbyt mała, co może powodować wysoką wariancję parametrów.

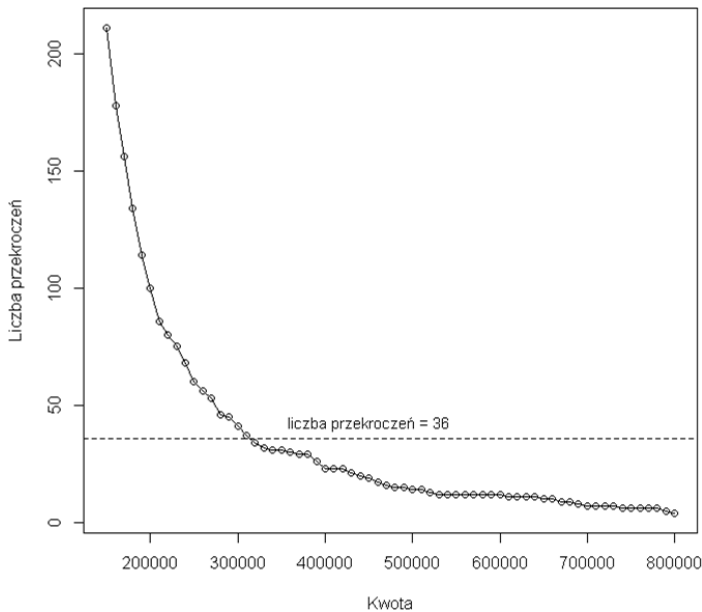
Na rysunku 10 przedstawiono łączną liczbę przypadków, która przekracza dany próg  $u$ , z oznaczeniem liczby obserwacji występujących w zbiorze wyznaczonym metodą bloków – odgórnie założono, że dalsza analiza będzie odbywała się dla progu, który będzie miał co najmniej tyle samo obserwacji co w przypadku metody bloków dla okresów miesięcznych. Biorąc pod uwagę warunek wstępny określony

<sup>3</sup> Spotykane są również inne metody wyznaczenia progu  $u$ : wykres estymatora parametru kształtu oraz wykres Gertensgarbe. Zob.: [Cebrián, Denuit, Lambert 2003, s. 9].



Rys. 9. Mean Excess Plot dla poszczególnych progów kwotowych  $u$

Źródło: opracowanie własne.



Rys. 10. Liczba obserwacji przekraczająca próg kwotowy  $u$

Źródło: opracowanie własne.

na podstawie wykresu średnich powyżej progu oraz warunków powyższych, ustalono, że dopasowanie rozkładu może nastąpić dla progu  $u$  do maksymalnej wartości około 310 tys. zł. Uwzględniając dodatkowo wyniki testu dopasowania Kołmogorowa–Smirnowa (podobnie jak w przypadku rozkładu GEV porównano rozkład empiryczny danych z rozkładem teoretycznym i średnia  $p$  – wartość testu wyniosła 0,77, z odchyleniem standardowym wynoszącym 0,14) oraz weryfikację wykresów dopasowania kwantyl–kwantyl, przyjęto oszacowania parametrów uogólnionego rozkładu Pareto dla progu o wartości 190 tys. zł. Za wyborem tej wartości progu przemawia również zgodność parametru  $\xi$  w tym przypadku z wartością tego parametru w dopasowanym rozkładzie GEV [Coles 2001, s. 75-76].

Statystyki opisowe zbioru wyznaczonego metodą przekroczeń powyżej progu dla 190 tys. zł zostały wskazane w tab. 5. Wartości przyjmowane przez obserwacje znajdują się w zakresie od około 190 tys. zł do ponad 2,67 mln zł ze średnią około 359 tys. zł. Dodatnia wartość skośności i kurtozy może być interpretowana podobnie jak w przypadku analizy zbioru obserwacji wyznaczonych metodą bloków.

**Tabela 5.** Statystyki opisowe dla zbioru wyznaczonego metodą przekroczeń powyżej progu

Liczba obserwacji	114
Minimum	190 011 zł
Maksimum	2 672 069 zł
Średnia	359 267 zł
Dolny kwantyl	211 181 zł
Mediana	256 867,02zł
Górny kwantyl	377 325 zł
95. percentyl	749 857,20 zł
99. percentyl	1 691 892,50 zł
Skośność	4,52
Kurtoza	24,9

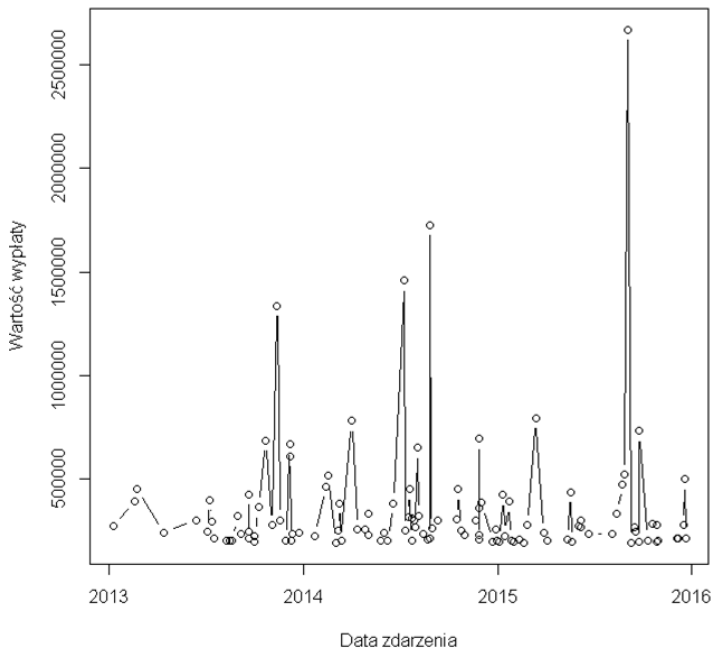
Źródło: opracowanie własne.

Na rysunku 11 przedstawiono poszczególne wartości obserwacji wyznaczone metodą przekroczeń powyżej progu.

Rysunek 12 przedstawia histogram oraz dystrybucję empiryczną obserwacji dla zbioru wyznaczonego metodą przekroczeń powyżej progu. Wraz ze wzrostem wartości wypłaty liczba przypadków systematycznie maleje.

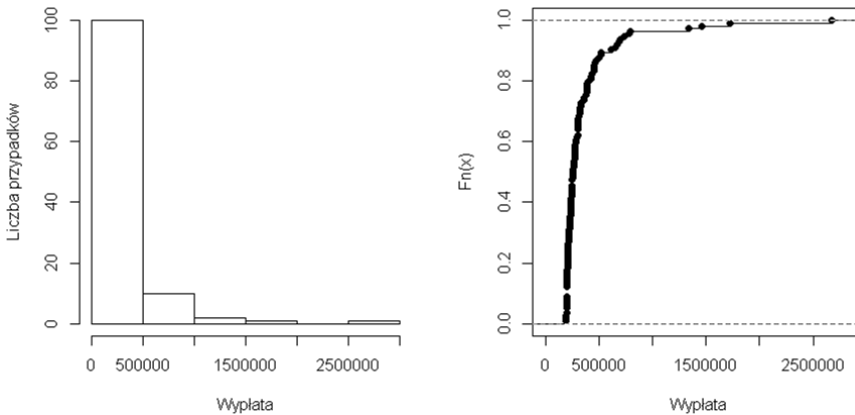
Wykorzystując metodę największej wiarygodności, wyznaczono wartości poszczególnych parametrów uogólnionego rozkładu Pareto wraz z odpowiadającymi im błędami standardowymi:

$$\hat{\sigma} = 79173,16 \quad (3003,52) \quad \hat{\xi} = 0,58 \quad (0,12).$$



Rys. 11. Obserwacje wyznaczone metodą przekroczeń powyżej progu

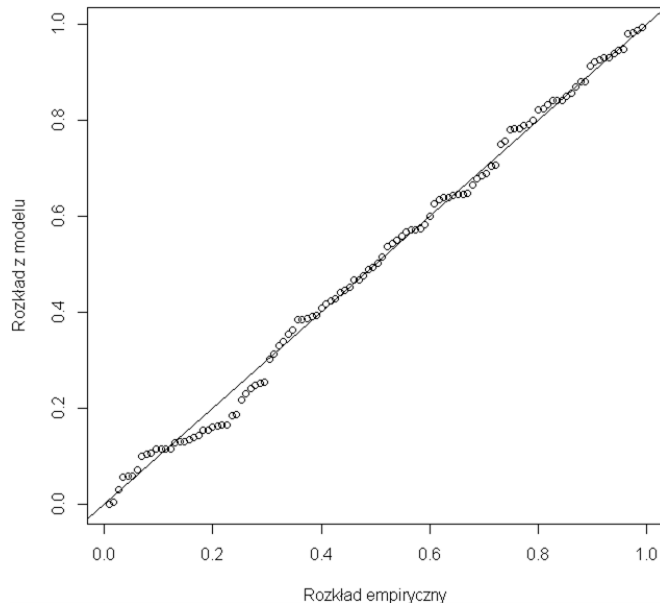
Źródło: opracowanie własne.



Rys. 12. Histogram oraz dystrybuanta empiryczna obserwacji wyznaczonych metodą przekroczeń powyżej progu

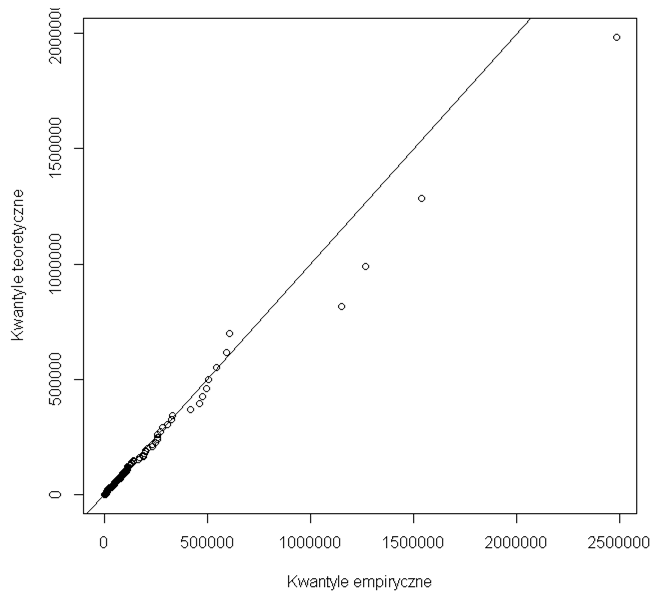
Źródło: opracowanie własne.

Rysunki 13 i 14 wskazują, że dopasowanie uogólnionego rozkładu Pareto do danych empirycznych jest dobre. Największe rozbieżności występują w przypadku wykresu kwantyl–kwantyl dla kilku wypłat



**Rys. 13.** Dopasowanie dystrybuanty empirycznej oraz dystrybuanty z uogólnionego rozkładu Pareto

Źródło: opracowanie własne.



**Rys. 14.** Wykres kwantyl-kwantyl dla dopasowanego uogólnionego rozkładu Pareto

Źródło: opracowanie własne.

o wartości powyżej 1 mln zł. W takiej sytuacji kwantyle teoretyczne z dopasowanego uogólnionego rozkładu Pareto są niedoszacowane.

Podobnie jak w przypadku rozkładu GEV możliwe jest wyznaczenie za pomocą wzoru analitycznego kwantyli wskazujących na tzw. poziomy zwrotu, opisujące, z jaką częstotliwością dana wartość wypłaty zostanie przekroczona. Przedział ufności dla tak wyznaczonej wartości można oszacować, stosując metodę delta dla rozkładu Pareto [Coles 2001, s. 82-83].  $R_t$  wyraża się następującym wzorem:

$$R_t = \begin{cases} u + \frac{\hat{\sigma}}{\hat{\xi}} \left[ (rn_y \hat{\lambda}_u)^{\hat{\xi}} - 1 \right], & \xi \neq 0, \\ u + \hat{\sigma} \log(rn_y \hat{\lambda}_u), & \xi = 0, \end{cases}$$

gdzie  $n_y$  – liczba obserwacji w zbiorze dla okresu jednego roku (przyjęto średnią liczbę obserwacji w okresie roku),  $\hat{\lambda}_u$  – proporcja liczby przekroczeń do całego zbioru obserwacji,  $r$  – roczny okres zwrotu.

**Tabela 6.** Poziomy zwrotu oraz błędy statystyczne dla poszczególnych okresów zwrotu wyznaczone za pomocą metody przekroczeń powyżej progu

Okres analizy	12 miesięcy	24 miesiące	36 miesięcy
Wartość zwrotu	1 172 199	1 722 724	2 162 942
Błąd statystyczny	301 759	577 925	828 120

Źródło: opracowanie własne.

Informacja o poziomie zwrotu dla dopasowanego uogólnionego rozkładu Pareto została zawarta w tab. 6. Wartości zwrotów w poszczególnych miesiącach są porównywalne do tych wskazanych za pomocą rozkładu GEV. Z kolei błędy statystyczne dla tak wyznaczonych obserwacji przyjmują mniejsze wartości. Interpretując wartość zwrotu dla 36 miesięcy, należy przyjąć, że biorąc pod uwagę dane wszystkich zakładów ubezpieczeń prowadzących działalność w zakresie ubezpieczenia OC p.p.m. w Polsce, średnio raz na trzy lata zdarzy się wypadek ubezpieczeniowy, którego konsekwencje finansowe przekroczą około 2,16 mln zł. Również w tym przypadku należy wnioskować, że wartość ta jest niedoszacowana oraz pod uwagę brać możliwe odchylenia wskazane w ramach błędu statystycznego.

### 3. Podsumowanie

Wysokie wypłaty o charakterze ekstremalnym mogą mieć znaczny wpływ na sytuację finansową zakładu ubezpieczeń. W przypadku ubezpieczenia OC p.p.m. zakład ubezpieczeń ma pośredni wpływ na

wysokość odszkodowań oraz ich częstość. Ubezpieczyciel stara się wpłynąć na jakość ubezpieczonego portfela poprzez odpowiednią selekcję ryzyka. Bezpośrednio wiąże się to ze stosowaniem składek ubezpieczeniowych adekwatnych do ryzyka wprowadzanego do portfela w związku z ubezpieczeniem danego podmiotu lub przedmiotu. Pomimo możliwości pośredniego wpływu na jakość ubezpieczonego portfela, zdarzenia o charakterze ekstremalnym mogą wystąpić w każdym z zakładów ubezpieczeń, który ubezpiecza posiadaczy pojazdów od odpowiedzialności cywilnej związanej z posiadaniem pojazdu. W takim przypadku zakład ubezpieczeń powinien mieć odpowiedni kapitał niezbędny w celu pokrycia powstałych zobowiązań.

W pracy przedstawiono sposób oceny ryzyka katastroficznego towarzyszącego zakładowi ubezpieczeń prowadzącemu działalność w obszarze ubezpieczeń OC p.p.m.. Do zbioru obserwacji dopasowano uogólniony rozkład wartości ekstremalnych oraz uogólniony rozkład Pareto. Estymacja parametrów rozkładu wymagała zastosowania odpowiednich metod wyboru obserwacji o charakterze ekstremalnym. W omawianym przypadku zastosowano dwie metody – metodę bloków oraz metodę przekroczeń powyżej progu. Przeprowadzone badanie wskazuje, że do modelowania wypłat ekstremalnych można z powodzeniem stosować zaproponowane rozkłady. Rozkłady te dobrze dopasowują się do danych będących przedmiotem badania.

Zaletą oszacowań przedstawionych w pracy jest wykorzystanie pełnej informacji o wypłatach za zdarzenia OC p.p.m. dla wszystkich zakładów ubezpieczeń działających w Polsce w segmencie ubezpieczeń komunikacyjnych. Kalkulacja rozkładu wypłat ekstremalnych może być wykorzystana przez poszczególne zakłady ubezpieczeń do lepszego szacowania wysokości składki przez wyznaczenie tzw. narzutu na czynnik ekstremalny. Z drugiej strony, wskazane opracowanie może posłużyć zakładom ubezpieczeń do zapewnienia jeszcze lepszej adekwatności w zakresie posiadanego kapitału wypłacalności. Związane jest to również z właściwym nadzorem nad zapewnieniem modeli statystycznych oddających w sposób najbardziej zbliżony naturę badanego zjawiska i z punktu widzenia regulatora zapewniającego bezpieczeństwo klientów rynku ubezpieczeniowego, a przez to finansowego.

## Literatura

- Cebrián A.C., Denuit M., Lambert P., 2003, *Generalized Pareto fit to the society of actuaries large claims database*, Universidad de Zaragoza.
- Coles S., 2001, *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.



- Delignette-Muller M. L., Dutang C., *fitdistrplus: An R Package for Fitting Distributions*, <https://www.jstatsoft.org/article/view/v064i04/v64i04.pdf> (05.09.2016).
- Dyrektywa Parlamentu Europejskiego i Rady 2009/138/WE z dnia 25 listopada 2009 r. w sprawie podejmowania i prowadzenia działalności ubezpieczeniowej i reasekuracyjnej (Wypłacalność II).
- Dyrektywa Parlamentu Europejskiego i Rady 2013/58/UE z dnia 11 grudnia 2013 r. zmieniająca dyrektywę 2009/138/WE (Wypłacalność II) w odniesieniu do terminu jej transpozycji i daty rozpoczęcia jej stosowania oraz daty uchylecia niektórych dyrektyw (Wypłacalność I).
- Efron B., Tibshirani R.J., 1993, *An Introduction to the Bootstrap*, Chapman & Hall/CRC.
- Embrechts P., Klüppelberg C., Mikosch T., 1996, *Modelling Extremal Events for Insurance and Finance*, Springer Verlag.
- Fałdziński M., 2014, *Teoria wartości ekstremalnych w ekonometrii finansowej*, Wydawnictwo Naukowe UMK, Toruń.
- Fisher R.A., Tippett L.H.C., 1928, *Limiting Forms of the Frequency Distribution of the Largest Or Smallest Member of a Sample*, Proc. Cambridge Phil. Soc. 24.
- Gumbel E.J., 2004, *Statistics of Extremes*, Dover Publications, Inc., Mineola, New York.
- Rachunek prawdopodobieństwa i statystyka/Wykład 12: Metoda największej wiarygodności, [http://wazniak.mimuw.edu.pl/index.php?title=Rachunek\\_prawdopodobieństwa\\_i\\_statystyka/Wykład\\_12:\\_Metoda\\_największej\\_wiarygodności](http://wazniak.mimuw.edu.pl/index.php?title=Rachunek_prawdopodobieństwa_i_statystyka/Wykład_12:_Metoda_największej_wiarygodności) (13.09.2016).
- Sokół J. 2012, *Skuteczność stress-testów jako narzędzie pomiaru ryzyka*, [w:] *Wyzwania gospodarki globalnej*, Prace i Materiały Instytutu Handlu Zagranicznego Uniwersytetu Gdańskiego, nr 31.

## THE QUANTIFICATION OF RISK OF PAYMENTS IN THE FIELD OF CATASTROPHIC INSURANCE EVENTS USING EXTREME VALUE THEORY

**Summary:** In the paper the authors apply Extreme Value Theory in the field of extremely high values of claims paid out by insurance companies conducting motor third party liability insurance. Extremely high indemnities and compensations were identified using Block Maxima and Excesses Over Threshold methods. The value of payments is modelled using Generalized Extreme Value and Generalized Pareto distributions. The empirical part of the work presents the study on claims data held by Information Centre of Insurance Guarantee Fund. The results of analysis can be applied in pricing models used by insurance and reinsurance undertakings as well can be used in the activities of the financial supervisory authority in relation to setting standards for the proper assessment of risk faced by entities of insurance market.

**Keywords:** Extreme Value Theory, Block Maxima method, Excesses over Threshold method, generalized extreme value distribution, generalized Pareto distribution.