

Paweł Cibis

Uniwersytet Ekonomiczny we Wrocławiu

STATYSTYKI POZYCYJNE – SZACOWANIE ROZKŁADU MAKSYMALNYCH STRAT NA PRZYKŁADZIE WARSZAWSKIEJ GIEŁDY PAPIERÓW WARTOŚCIOWYCH

Streszczenie: W artykule opisano historię teorii wartości ekstremalnych wraz z jej matematycznymi podstawami i praktycznymi zastosowaniami. Zaprezentowano również wyniki empirycznego badania możliwości wykorzystania teorii do modelowania poziomu maksymalnych strat na warszawskiej Giełdzie Papierów Wartościowych na podstawie danych czterech najważniejszych indeksów GPW.

Słowa kluczowe: teoria wartości ekstremalnych, zarządzanie ryzykiem, giełda papierów wartościowych, modele rozkładu maksymalnej straty.

1. Wstęp

Teoria wartości ekstremalnych (*Extreme Values Theory* – EVT) jest gałęzią statystyki zajmującą się analizą niezwykle dużych odchyłeń od wartości przeciętnych rozkładu prawdopodobieństwa. Nie jest to temat nowy, lecz jego znaczenie i popularność rośnie szczególnie dziś – w czasach dużych wahań na rynkach finansowych i niepewności co do najbliższej przyszłości w światowej gospodarce.

Celem niniejszej pracy jest prezentacja dotychczasowych dokonań na polu badań nad statystykami ekstremalnymi oraz pokazanie, w jaki sposób teorię wartości ekstremalnych zastosować można do prognozowania ekstremalnych stóp zwrotu z akcji notowanych na warszawskiej Giełdzie Papierów Wartościowych.

W punkcie 2. pokrótce opisano trwającą, jak dotąd, niecałe stulecie historię badań nad wartościami ekstremalnymi oraz praktyczne zastosowania teorii wartości ekstremalnych ze szczególnym uwzględnieniem zastosowań EVT na rynkach finansowych. Punkt 3. przedstawia omawiane modele od strony teoretycznej, 4 – przykładowe zastosowanie EVT dla czterech najważniejszych indeksów GPW, 5. zaś zawiera podsumowanie omawianych zagadnień i wnioski końcowe.

2. Historia i zastosowania teorii wartości ekstremalnych

Pierwszą istotną pracę dotyczącą statystyk ekstremalnych opublikował Dodd [Dodd 1932]. Opisał w niej rezultaty swoich badań nad asymptotycznymi przedziałami zmienności pewnych klas rozkładów prawdopodobieństwa. Następnie w dość krótkich odstępach czasu powstały publikacje Fréchet’a [Fréchet 1927], Fishera i Tippetta [Fishera, Tippett 1928] oraz de Finetti’ego [de Finetti 1930], przy czym największy wkład w ukształtowanie się teorii wartości ekstremalnych miała druga z wymienionych pozycji. Jej autorzy sformułowali fundamentalne twierdzenie EVT, w którym zdefiniowali postać rozkładów granicznych maksimum (rozkładów wartości ekstremalnych). Podkreślili również bardzo powolną zbieżność maksimum rozkładu normalnego do swojego rozkładu granicznego. Gnedenko [Gnedenko 1943] przedstawił warunki konieczne i wystarczające dla zbieżności maksimum ciągu zmiennych losowych do rozkładów wartości ekstremalnych. Dotychczasowe osiągnięcia na polu badań nad wartościami ekstremalnymi zebrał i podsumował Gumbel [Gumbel 1958]. Ważną publikacją w historii teorii była praca de Haana [de Haan 1970], w której wprowadził on regularną zmienność (*regular variation*) jako narzędzie analityczne. Do tego czasu rozkłady statystyk ekstremalnych opisywano za pomocą modeli parametrycznych, jednakże wyniki badań Hilla [Hill 1975] i Pickandsa [Pickands 1975], a także wprowadzenie estymatora momentów przez Dekkersa i de Haana [Dekkers, de Haan 1989] skierowały uwagę badaczy w stronę modeli półparametrycznych (*semiparametric models*). Równocześnie zaproponowano [Balkeema, de Haan 1974; Singpurwalla 1972] alternatywne podejście do modelowania ekstremów – rozkład przekroczeń wysokiego progu, który zbiega do uogólnionego rozkładu Pareto (*Generalized Pareto Distribution – GPD*). Najbardziej znana metodologia estymacji parametrów rozkładu GPD (metodą największej wiarygodności) – *peaks over threshold* (POT) – przedstawiona została w pracach Smitha [Smith 1987] oraz Davisona i Smitha [Davison, Smith 1990]. Dress, Ferreira i de Haan [Dress, Ferreira, de Haan 2004] dowiedli asymptotycznej normalności estymatora największej wiarygodności, tzw. indeksu wartości ekstremalnych (*Extreme Value Index – EVI*) – najważniejszego parametru rozkładu wartości ekstremalnych.

Obszerny przegląd praktycznych zastosowań teorii wartości ekstremalnych można znaleźć w pracy Pactwa [Pactwa 2001]. Za prekursora na tym polu uchodzi Gumbel. W swojej pierwszej głośnej pracy [Gumbel 1935] wykorzystuje EVT w rozważaniach dotyczących maksymalnego czasu trwania życia, a następnie [Gumbel 1941] wskazuje również zastosowanie teorii w statystycznym modelowaniu ekstremalnych poziomów rzek. Gross i inni [Gross i in. 1995] proponują zastosowanie uogólnionego rozkładu Pareto do prognozowania ponadprzeciętnej prędkości wiatru jako alternatywy wobec stosowanego do tej pory modelowania klasycznym rozkładem GEV [Gumbel 1958; Zwiers 1987]. EVT znajduje również zastosowanie w przewidywaniu siły trzęsień ziemi [Lavenda, Cipollone 2000] oraz ekstremalnych temperatur

[Ferro, Segers 2003; Leadbetter, Lindgren, Rootzén 1983]. Wartości ekstremalne wykorzystywane są także w badaniach nad problemem zanieczyszczeń powietrza [Horowitz 1980; Singpurwalla 1972; Smith 1989]. Jednakże najbardziej popularnym praktycznym wykorzystaniem teorii wartości ekstremalnych do modelowania zjawisk naturalnych jest problem prognozowania wysokości fal morskich w celu konstrukcji odpowiednio wysokich grobli [Dijk, de Haan 1992; de Haan 1990; de Haan 1998; de Haan 1999]. EVT z powodzeniem stosowana jest także w problemach niezwiązanych ze zjawiskami meteorologicznymi. Na przykład Weibull [Weibull 1938], Leadbetter i inni [Leadbetter, Lindgren, Rootzén 1983] oraz Lindgren i Rootzén [Lindgren, Rootzén 1987] dyskutują zastosowania teorii wartości ekstremalnych do badania wytrzymałości materiałów. EVT pojawia się również w analizie rezultatów osiągniętych przez sportowców [Gomes 2008; Strand, Boes 1998].

Spośród wszystkich dotychczas odkrytych zastosowań teorii wartości ekstremalnych pierwsza pozycja bezsprzecznie należy się jednak modelowaniu ryzyka poniesienia znacznych strat finansowych. Straty te dotyczą przede wszystkim inwestycji na rynkach kapitałowych oraz działalności firm ubezpieczeniowych (ryzyko strat katastroficznych).

Podczas modelowania stopy zwrotu z inwestycji finansowych kluczowe założenie, jakie musi przyjąć badacz, dotyczy postaci rozkładu owych stóp zwrotu. Najczęściej zakłada się normalność lub log-normalność stóp zwrotu ze względu na wiele wygodnych własności tych rozkładów statystycznych. O ile jednak rozkłady takie mogą w miarę wiarygodnie przybliżać rzeczywistość w normalnych warunkach rynkowych, o tyle w przypadkach ekstremalnych zawodzą, gdyż ich ogony zanikające w tempie wykładniczym nie są na tyle grube, aby można było za ich pomocą odpowiednio oszacować rozmiary ponadprzeciętnych zysków lub strat.

Na początku lat 90. ubiegłego stulecia pojawiło się sporo publikacji zawierających propozycje wykorzystania rozkładów wartości ekstremalnych w celu wiarygodnego oszacowania rozkładu stóp zwrotu. Hols i de Vries [Hols, de Vries 1991] modelują w ten sposób zmienność kursów walutowych, estymując metodami półparametrycznymi indeks ogona oraz szacując przedział ufności dla prognozy. Problematyką zastosowania EVT w prognozowaniu zmian kursów walut zajmują się później także de Vries [de Vries 1998] oraz Danielsson i de Vries [Danielsson, de Vries 1997].

Dogłębnej analizie ekstremalnych stóp zwrotu z amerykańskich akcji w bardzo szerokim przedziale czasowym (1885-1990) dokonuje za pomocą EVT Longin [Longin 1996]. Za najodpowiedniejszy rozkład ekstremalny dla stóp zwrotu uznaje rozkład Fréchet'a. Przekonuje również o skończoności wartości oczekiwanej i wariancji, pozostawiając jednocześnie tę kwestię nierozstrzygniętą dla momentów wyższych rzędów. Zauważa także znaczną stabilność parametrów rozkładu w czasie (z naciskiem na stałość kształtu ogona) oraz ze względu na sposób agregacji (stopy dzienne, tygodniowe, miesięczne itd.). Podobny pogląd dotyczący istnienia czterech

pierwszych momentów mają Jansen i de Vries [Jansen, de Vries 1991]. Do nieco odmiennych wniosków dochodzą Loretan i Phillips [Loretan i Phillips 1994], odrzucając zarówno hipotezę o skończoności kurtozy, jak i wariancji. Czekala [Czekala 2001] proponuje wykorzystanie EVT w celu weryfikacji założeń modelu Blacka-Scholesa, następnie dokonuje takiej weryfikacji w odniesieniu do wybranych spółek z GPW, dochodząc do wniosku, iż generalnie założenia te nie są spełnione. Przedstawia też procedurę dalszego postępowania w takich przypadkach – zastosowanie rozkładów stabilnych.

Wykorzystanie teorii wartości ekstremalnych w problematyce ubezpieczeniowej opisują Klüppelberg i Mikosch [Klüppelberg, Mikosch 1996] oraz Mikosch i Nagaev [Mikosch, Nagaev 1998], analizując duże odchylenia sum zmiennych losowych o grubych ogonach.

3. Teoretyczne podstawy rozkładów wartości ekstremalnych

Załóżmy, że ciąg X_1, X_2, \dots, X_n jest ciągiem niezależnych zmiennych losowych o takim samym rozkładzie, którego funkcję dystrybuanty oznaczymy jako F . Ponadto niech $M_n = \max(X_1, X_2, \dots, X_n)$, gdzie $n > 2$.

W takim razie prawdopodobieństwo tego, że maksimum M_n nie przekroczy pewnej wartości x wyraża się wzorem:

$$P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = F^n(x),$$

gdzie $x \in \mathbb{R}$ oraz $n \in \mathbb{N}$.

Do postaci granicznych rozkładów zmiennej M_n prowadzi fundamentalne dla teorii wartości ekstremalnych twierdzenie Fishera-Tippetta: jeżeli istnieją pewne stałe normujące $c_n > 0$ i $d_n \in \mathbb{R}$ oraz pewna niezdegenerowana funkcja dystrybuanty H , taka że:

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H, \text{ gdy } n \rightarrow \infty,$$

to H może mieć wyłącznie postać jednego z trzech rozkładów wartości ekstremalnych. W takim razie mówimy, że zmienna losowa X (i jej rozkład o dystrybuancie F) należy do obszaru przyciągania maksimum rozkładu wartości ekstremalnych H , co oznaczamy jako $X \in MDA(H)$ ($F \in MDA(H)$) [Fisher, Tippet 1928].

Powyższy wywód łatwo przeprowadzić również dla minimów. Wystarczy bowiem zauważyć, że $\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n)$ [Leadbetter, Lindgren, Rootzén 1983].

Wspomniane trzy możliwe rozkłady graniczne znormalizowanych ekstremów to rozkłady:

- Gumbela (typ I) : $\Lambda(x) = \exp(-e^{-x}), x \in \mathbb{R}$,
- Fréchet’a (typ II): $\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases}$, dla $\alpha > 0$,
- Weibulla (typ III): $\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^{-\alpha}), & x < 0 \\ 1, & x \geq 0 \end{cases}$, dla $\alpha > 0$.

Rozkład Gumbela jest asymptotycznym rozkładem ekstremów zmiennych losowych o cienkich ogonach (np. normalny, log-normalny). Cechą odróżniającą go od pozostałych dwóch typów rozkładów ekstremalnych jest jego nieograniczoność zarówno z góry, jak i z dołu. Rozkład Fréchet’a jest ograniczonym z dołu rozkładem granicznym dla ekstremów zmiennych losowych o grubych ogonach (np. t-Studenta, Pareto, Cauchy’ego), natomiast rozkład Weibulla to odbicie lustrzane rozkładu Fréchet’a. Z powodu braku ogona nadaje się on do modelowania zjawisk, których wielkość jest ograniczona z góry. Ponieważ rozkłady stóp zwrotu z inwestycji finansowych mają grube ogony, a ponadto najczęściej nie posiadają górnych ograniczeń, najbardziej odpowiednim do ich modelowania wydaje się rozkład Fréchet’a.

Jenkinson [Jenkinson 1955] połączył trzy typy rozkładów wartości ekstremalnych w jedną zunifikowaną formę zwaną uogólnionym rozkładem wartości ekstremalnych (*Generalized Extreme Value Distribution – GEV*):

$$H_{\xi, \mu, \beta}(x) = \begin{cases} \exp\left\{-\left(1 + \xi \frac{x - \mu}{\beta}\right)^{-\frac{1}{\xi}}\right\}, & \text{dla } \xi \neq 0 \\ \exp\left\{-\exp\left(-\frac{x - \mu}{\beta}\right)\right\}, & \text{dla } \xi = 0 \end{cases},$$

gdzie ξ jest parametrem kształtu (będącego odwrotnością indeksu ogona α , EVI), a μ i β są odpowiednio parametrami położenia i skali. Dodatkowo x jest ograniczone wyrażeniem $(\mu + \beta/\xi)$ z góry dla ξ dodatniego oraz z dołu dla ξ ujemnego. Warto zauważyć, iż dla $\xi = 0$, H jest po prostu dystrybuantą rozkładu Gumbela.

Podstawowym zadaniem w analizie wartości ekstremalnych jest dopasowanie do danych odpowiedniego rozkładu H , a więc estymacja parametrów ξ , μ i β . Kluczowy problem stanowi oszacowanie parametru kształtu ξ , który w największym stopniu decyduje o kształcie rozkładu i grubości ogona. Podejście tradycyjne opisywane przez Smitha [Smith 1987] polega na estymacji parametru ξ metodą największej wiarygodności. Wymaga ono pogrupowania surowych obserwacji w bloki (gdzie 1 blok to np. 1 miesiąc lub 1 rok) i wyznaczenia nowego szeregu czasowego

– maksimum poszczególnych bloków. Zakłada się przy tym, że maksima blokowe są zmiennymi losowymi pochodzącymi z jednego z trzech rozkładów ekstremalnych. Jest to dosyć silne założenie, dlatego prowadzone są również badania (np. [de Haan 1990; Pictet, Dacorogna, Müller 1998]) nad mniej rygorystycznymi estymatorami momentów.

Jeśli znane są oszacowania lub wartości parametrów ξ , μ i β , można łatwo wyznaczyć wartość oczekiwaną i wariancję danego rozkładu wartości ekstremalnych, korzystając ze wzorów:

$$E(X) = \mu - \frac{\sigma}{\xi} + \frac{\sigma}{\xi} g_1 \text{ oraz}$$

$$\text{Var}(X) = \frac{\sigma^2}{\xi^2} (g_2 - g_1^2), \text{ gdzie } g_k = \Gamma(1 - k\xi).$$

4. Zastosowanie teorii wartości ekstremalnych na przykładzie zachowania wybranych indeksów GPW

Możliwości zastosowania technik teorii wartości ekstremalnych zostały zbadane na bazie historycznych poziomów wybranych indeksów warszawskiej Giełdy Papierów Wartościowych. Do analizy wybrano następujące indeksy: WIG, WIG20, MWIG40 (dawniej MIDWIG) oraz SWIG80 (dawniej WIRR). Najważniejsze informacje dotyczące omawianego zbioru danych zamieszczono w tab. 1.

Tabela 1. Podstawowe charakterystyki analizowanego zbioru danych

Nazwa indeksu	WIG	WIG20	MWIG40	SWIG80
Częstotliwość	dzienne	dzienne	dzienne	dzienne
Początek okresu	1991-04-16	1994-04-14	1997-12-31	1994-12-29
Koniec okresu	2008-12-19	2008-12-19	2008-12-19	2008-12-19
Liczba obserwacji	3975	3643	2752	3500

Źródło: opracowanie własne na podstawie [Dom Maklerski...].

Ponieważ zarządzanie ryzykiem skupia się przede wszystkim na zabezpieczeniu portfela inwestycyjnego przed narażeniem na duże straty, a dopiero później rozważana jest możliwość realizacji nadzwyczajnych zysków, w niniejszej analizie skoncentrowano się na dolnym ogonie rozkładu stóp zwrotu. W tym celu obliczone zostały wartości przeciwne do dziennych logarytmicznych stóp zwrotu ze wzoru:

$$-r_{\ln_t} = -(\ln(p_t) - \ln(p_{t-1})),$$

gdzie p_t oznacza poziom danego indeksu na koniec sesji giełdowej o numerze t . Dla uproszczenia wyrażenie $-r_{\ln_t}$ będziemy nazywać dalej po prostu logarytmicznymi

stopami zwrotu. Statystyki danych otrzymanych na podstawie omówionego przekształcenia umieszczono w tab. 2. Dla każdego indeksu skośność jest dodatnia, co wskazuje na prawostronną asymetrię każdego z empirycznych rozkładów. Ponadto dodatnia jest również kurtoza (w każdym przypadku jej wartość przekracza 4), co oznacza, iż rozkład logarytmicznych stóp zwrotu jest bardzo skoncentrowany wokół wartości przeciętnej i ma grube ogony (leptokurtyczny). Odbiega więc mocno od rozkładu normalnego – symetrycznego i mającego kurtozę równą zero (mezokurtyczny).

Histogramy logarytmicznych stóp zwrotu indeksów GPW przedstawiono na rys. 1. Łatwo zaobserwować można, jak bardzo wysmukłe są rozkłady logarytmicznych stóp zwrotu, a także (choć być może nie na pierwszy rzut oka), iż zarysowuje się lekka dodatnia asymetria.

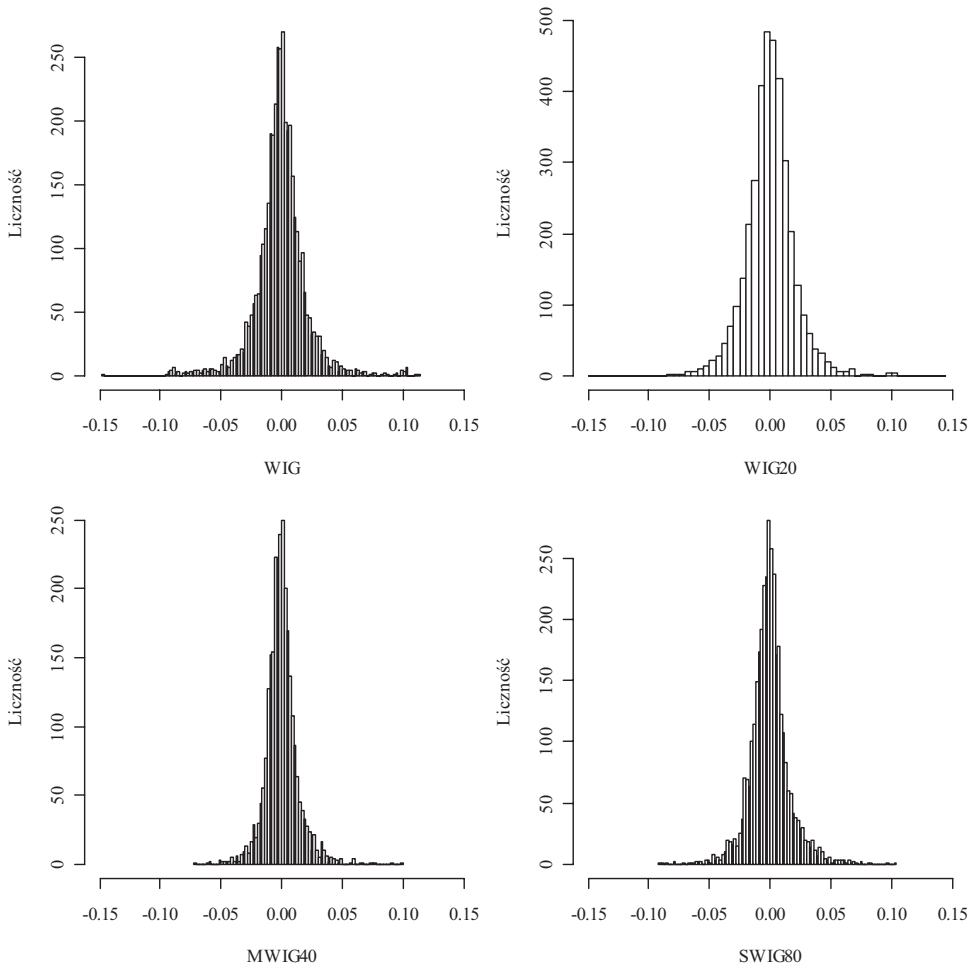
Histogramy modelowanych statystyk – maksimum blokowych logarytmicznych stóp zwrotu – przedstawiono na rys. 2. Wielkość bloku ustalono na poziomie 30 kolejnych sesji giełdowych. Otrzymane histogramy jednoznacznie wskazują na rozkład silnie prawostronnie asymetryczny i ograniczony z dołu. Jak już wspomniano w punkcie 3., spośród rozkładów statystyk ekstremalnych takimi właściwościami charakteryzuje się rozkład Fréchéta.

Tabela 2. Statystyki opisowe logarytmicznych stóp zwrotu z analizowanych indeksów GPW

Nazwa indeksu	WIG	WIG20	MWIG40	SWIG80
Średnia	-0,0008314133	-0,0001579989	-0,0001557777	-0,0005526668
Mediana	-0,000702638	-5,0792e-05	-0,000621562	-0,00110925
Odchylenie stand.	0,02093908	0,02037120	0,01339899	0,01634773
Wariancja	0,0004384449	0,000414986	0,0001795328	0,0002672482
Skośność	0,05204826	0,1455248	0,6529106	0,3457633
Kurtoza	5,58016	4,401608	5,429549	4,096976
Minimum	-0,147831	-0,148356	-0,0705117	-0,0917552
Maksimum	0,113443	0,141608	0,0994172	0,1027110
Rozstęp	0,261274	0,289964	0,1699289	0,1944662

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R.

Za pomocą funkcji dostępnych w bibliotece QRM pakietu statystycznego R do analizowanych danych metodą największej wiarygodności dopasowano modele GEV. Rezultaty estymacji zamieszczono w tab. 3-5. Stopień dopasowania modelu do danych empirycznych zbadano następnie za pomocą testu Andersona-Darlinga (statystyka A^2) i jego modyfikacji (zaproponowanej w [Zempléni 2004]) koncentrującej



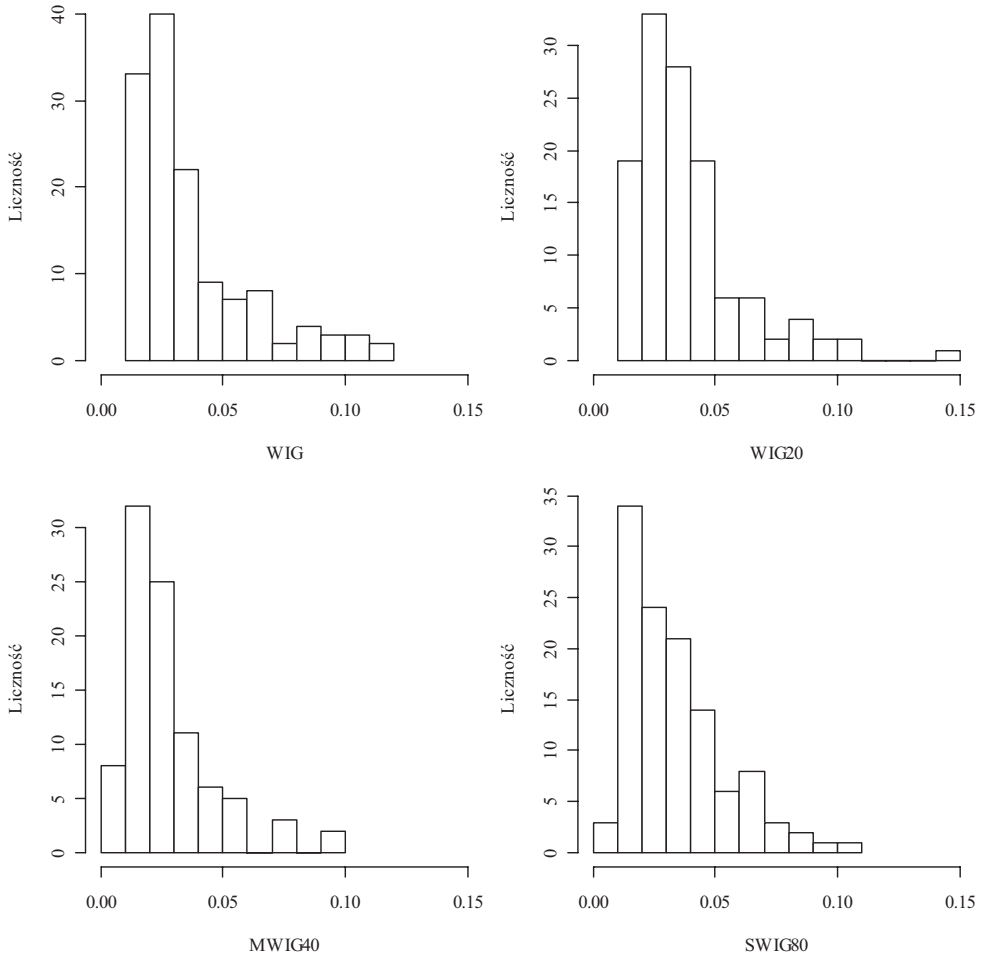
Rys. 1. Histogramy logarymicznych stóp zwrotu z analizowanych indeksów GPW

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R.

się w większym stopniu na prawym ogonie rozkładu¹ (statystyka B^2) oraz za pomocą współczynnika korelacji liniowej r między linią regresji punktów na wykresie kwantylowym i linią 45 stopni. W celach porównawczych badanie przeprowadzono dla maksimów blokowych o 30, 60 i 90 obserwacjach w każdym segmencie. Wyniki

¹ Istotnym problemem przy stosowaniu testu Andersona-Darlinga i jego zmodyfikowanej wersji jest brak stabilizowanych wartości krytycznych dla rozkładu GEV. W przeprowadzonym badaniu posłużono się przybliżonymi wartościami krytycznymi (lub ich przedziałami) wyznaczonymi symulacyjnie przez Zempléni [Zempléni 2004].

testów zgodności (na poziomie istotności $\alpha = 0,01$) opierających się na statystykach A^2 i B^2 zamieszczono w tab. 6-7. W tabeli 8. przedstawiono wartości współczynników korelacji liniowej Pearsona (r).



Rys. 2. Histogramy 30-sesyjnych maksimów blokowych z logarytmicznych stóp zwrotu z analizowanych indeksów GPW

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R z wykorzystaniem biblioteki QRM [McNeil, Frey, Embrechts 2009].

Tabela 3. Estymatory parametrów kształtu (ξ) rozkładu GEV dopasowanego do empirycznych rozkładów analizowanych indeksów GPW przy różnych rozmiarach bloku (m).

W nawiasach podano oceny błędu estymacji parametru

Indeks	$m = 30$	$m = 60$	$m = 90$
WIG	0,398 (0, 094)	0,322 (0, 122)	0,295 (0, 137)
WIG20	0,275 (0, 088)	0,209 (0, 107)	0,310 (0, 162)
SWIG80	0,259 (0, 102)	0,177 (0, 167)	0,141 (0, 240)
MWIG40	0,306 (0, 100)	0,307 (0, 162)	0,376 (0, 247)

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R z wykorzystaniem biblioteki QRM [McNeil, Frey, Embrechts 2005].

Tabela 4. Estymatory parametrów położenia (μ) rozkładu GEV dopasowanego do empirycznych rozkładów analizowanych indeksów GPW przy różnych rozmiarach bloku (m).

W nawiasach podano oceny błędu estymacji parametru

Indeks	$m = 30$	$m = 60$	$m = 90$
WIG	0,024 (0, 001)	0,030 (0, 002)	0,034 (0, 002)
WIG20	0,027 (0, 001)	0,035 (0, 002)	0,039 (0, 002)
SWIG80	0,022 (0, 001)	0,029 (0, 002)	0,034 (0, 003)
MWIG40	0,018 (0, 001)	0,023 (0, 002)	0,026 (0, 003)

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R z wykorzystaniem biblioteki QRM [McNeil, Frey, Embrechts 2005].

Tabela 5. Estymatory parametrów skali (β) rozkładu GEV dopasowanego do empirycznych rozkładów analizowanych indeksów GPW przy różnych rozmiarach bloku (m).

W nawiasach podano oceny błędu estymacji parametru

Indeks	$m = 30$	$m = 60$	$m = 90$
WIG	0,012 (0,001)	0, 014 (0, 002)	0, 015 (0, 002)
WIG20	0, 012 (0, 001)	0, 014 (0, 002)	0, 014 (0, 002)
SWIG80	0, 012 (0, 001)	0, 015 (0, 002)	0, 017 (0, 003)
MWIG40	0, 010 (0, 001)	0, 013 (0, 002)	0, 014 (0, 003)

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R z wykorzystaniem biblioteki QRM [McNeil, Frey, Embrechts 2005].

Tabela 6. Podsumowanie wyników testu zgodności Andersona-Darlinga (statystyka A^2) rozkładu empirycznego blokowych maksimów stóp zwrotu z wybranych indeksów GPW z dopasowanym rozkładem GEV przy różnych rozmiarach bloku (m) dla poziomu istotności $\alpha = 0,01$

Indeks	$m = 30$	$m = 60$	$m = 90$
WIG	zgodność	zgodność	zgodność
WIG20	zgodność	zgodność	zgodność
SWIG80	zgodność	zgodność	zgodność
MWIG40	zgodność	zgodność	zgodność

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R z wykorzystaniem biblioteki QRM [McNeil, Frey, Embrechts 2005].

Tabela 7. Podsumowanie wyników zmodyfikowanego testu zgodności Andersona-Darlinga (statystyka B^2) rozkładu empirycznego blokowych maksimów stóp zwrotu z wybranych indeksów GPW z dopasowanym rozkładem GEV przy różnych rozmiarach bloku (m) dla poziomu istotności $\alpha = 0,01$

Indeks	$m = 30$	$m = 60$	$m = 90$
WIG	zgodność	zgodność	zgodność
WIG20	zgodność	brak zgodności	zgodność
SWIG80	zgodność	zgodność	zgodność
MWIG40	zgodność	zgodność	zgodność

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R z wykorzystaniem biblioteki QRM [McNeil, Frey, Embrechts 2005].

Tabela 8. Współczynniki korelacji liniowej Pearsona (r) pomiędzy linią regresji punktów na wykresie kwantylowym dla rozkładu GEV dopasowanego do empirycznych rozkładów analizowanych indeksów GPW i linią 45-stopni przy różnych rozmiarach bloku (m)

Indeks	$m = 30$	$m = 60$	$m = 90$
WIG	0,943	0,959	0,955
WIG20	0,980	0,987	0,976
SWIG80	0,987	0,975	0,962
MWIG40	0,971	0,976	0,974

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R z wykorzystaniem biblioteki QRM [McNeil, Frey, Embrechts 2005].

Tabela 9. Prawdopodobieństwo poniesienia maksymalnej dziennej straty na analizowanych indeksach GPW w wysokości przekraczającej 10% przy różnych rozmiarach bloku (m)

Indeks	$m = 30$	$m = 60$	$m = 90$
WIG	0,0403	0,0476	0,0594
WIG20	0,0286	0,0407	0,0627
SWIG80	0,0242	0,0341	0,0435
MWIG40	0,0166	0,0311	0,0512

Źródło: opracowanie własne na podstawie obliczeń w pakiecie statystycznym R z wykorzystaniem biblioteki QRM [McNeil, Frey, Embrechts 2005].

Dla zadanego poziomu istotności we wszystkich rozpatrywanych przypadkach nie było podstaw do odrzucenia hipotezy zerowej o zgodności empirycznego rozkładu maksimumów blokowych z dopasowanym rozkładem wartości ekstremalnych dla testu Andersona-Darlinga. W przypadku zmodyfikowanej wersji testu hipoteza zerowa została odrzucona wyłącznie dla danych WIG20 przy bloku składającym się z 60. obserwacji. W pozostałych przypadkach wykazano zgodność rozkładów. W każdym z analizowanych wariantów współczynnik korelacji liniowej r między linią regresji punktów na wykresie kwantylowym i linią 45. stopni był bardzo wysoki (0,943-0,987).

Jeśli parametry rozkładu GEV zostały wyestymowane, można oszacować, z jakim prawdopodobieństwem maksymalna dzienna strata przekroczy pewną określoną wartość. Na przykład dla $m = 30$ maksymalna dzienna strata przekroczy 10% z prawdopodobieństwem 4,0% dla indeksu WIG, 2,9% dla WIG20, 2,4% dla SWIG80 oraz ok. 1,7% dla SWIG80. Analogiczne oszacowania dla wszystkich rozpatrywanych wartości bloku m przedstawiono w tab. 9.

5. Wnioski

Wyniki przeprowadzonego badania pozwalają stwierdzić, iż teoria wartości ekstremalnych jest dobrym narzędziem do modelowania maksymalnych strat dla GPW. Zyskuje ona na znaczeniu szczególnie w okresach niestabilności na rynkach finansowych – takich jak obserwowane w rozpatrywanym okresie. Wystarczy zauważyć, iż w ciągu ostatnich 90. dni każdy z analizowanych indeksów poniósł co najmniej jedną dzienną stratę wysokości 7,5% lub więcej. Rysunki 1 i 2 pokazują, iż każda taka obserwacja bezsprzecznie należy do ogona rozkładu.

Pewnym mankamentem zastosowanych metod jest trudne do zweryfikowania w praktyce założenie o niezależności i takim samym rozkładzie logarytmicznych stóp zwrotu z indeksów giełdowych. Ewentualna zależność ciągu zmiennych losowych nie stanowi jednak problemu przy spełnieniu przez nie warunków $D(u_n)$ i $D'(u_n)$ Leadbettera (asymptotyczna niezależność) [Leadbetter, Lindgren, Rootzén 1983]. W takiej sytuacji rozpatrywany rozkład wartości ekstremalnych jest rozkładem granicznym także dla maksimum z takiego ciągu zmiennych losowych.

Problem, który wciąż stanowi przedmiot badań, ilustrują tab. 3-5. O ile dla parametrów położenia i skali względny błąd oceny parametrów jest dość niski, o tyle wariancja estymatora parametru kształtu znajduje się na bardzo wysokim poziomie. W skrajnych przypadkach można nawet podać w wątpliwość, czy parametr ten rzeczywiście jest dodatni (a co za tym idzie, czy na pewno mamy do czynienia z postulowanym wcześniej rozkładem Fréchéta). Dlatego też najnowsze prace dotyczące EVT koncentrują się przede wszystkim na badaniu alternatywnych metod estymacji parametru kształtu rozkładu wartości ekstremalnych mających zapewnić znacznie mniejszą jego wariancję przy niewielkim obciążeniu. W ciągu ostatnich kilku lat pojawiło się kilka nowych propozycji estymatorów. Araújo Santos i inni [Araújo

Santos, Fraga Alves, Gomes 2006] opisali klasę estymatorów PORT (*peaks over random threshold*), Gomes i Pestana [Gomes, Pestana 2007a] wprowadzili ważony estymator Hilla (*weighted Hill estimator*), a Fraga Alves i inni [Fraga Alves 2007] – mieszany estymator momentów (*mixed moment estimator*). Efektem tych badań są estymatory o asymptotycznej wariancji zbliżonej do klasycznego estymatora Hilla, ale zarazem o mniejszym asymptotycznym obciążeniu [Gomes, Pestana 2007b].

Literatura

- Araújo Santos P., Fraga Alves M.I., Gomes M.I., *Peaks over random threshold methodology for tail index and quantile estimation*, "Revstat" 2006, 4(3), s. 227-247.
- Balkeema A.A., Haan L. de, *Residual life time at great age*, Ann. Probab. 1974, 2, s. 792-804.
- Czekała M., *Statystyki pozycyjne w modelowaniu ekonometrycznym: wybrane problemy*, AE, Wrocław 2001.
- Daniëlsson J., de Vries C.G., *Tail index and quantile estimation with very high frequency Data*, "Journal of Empirical Finance" 1997, 4.
- Davison A.C., Smith R.L., *Model for exceedances over high thresholds*, J. Royal Statist. Soc. 1990, B 52, s. 393-442.
- Dekkers A.L.M., Haan L. de, *On the estimation of the extreme-value index and large quantile estimation*, "The Annals of Statistics" 1989, 17, s. 1795-1832.
- Dijk V., Haan L. de, *On the estimation of the exceedance probability of a high level*, [w:] *Order statistics and nonparametrics: theory and applications*, Elsevier, Amsterdam 1992, s. 79-92.
- Dodd E.L., *The greatest and the least variate under general laws of error*, Trans. Amer. Math. Soc. 1923, 25, s. 525-539.
- Dom Maklerski Banku Ochrony Środowiska (bossa.pl).
- Drees H., Ferreira A., Haan L. de, *On maximum likelihood estimation of the extreme value index*, Ann. Appl. Probab. 2004, 14, s. 1179-1201.
- Ferro C.A.T., Segers J., *Inference for clusters of extreme values*, "Journal of the Royal Statistical Society. Series B (Statistical Methodology)" 2003, vol. 65, no 2, s. 545-556.
- Finetti B. de, *Funzione Caratteristica di un fenomeno aleatorio*, Atti. Accad. Naz. Lincei Rend. Cl. Sci. Fiz. Mat. Nat. 1930, 4, s. 85-133.
- Fisher R.A., Tippett L.H.C., *Limiting forms of the frequency distribution of the largest of smallest members of a sample*, Proc. Cambridge Philos. Soc. 1928, 24, s. 180-190.
- Fraga Alves M.I., Gomes M.I., Haan L. de, Neves C., *Mixed moment estimator and location invariant alternatives*, "Notas e Comunicações" CEAUL 14/2007.
- Fréchet M., *Sur la loi de probabilité de l'écart maximum*, Ann. de la Soc. Polonaise de Math. 1927, 6, s. 93-116.
- Gnedenko B.V., *Sur la distribution limite du terme maximum d'une série aléatoire*, Ann. Math. 1943, 44, s. 423-453.
- Gomes M.I., Pestana D., *A simple second order reduced bias' tail index estimator*, J. Stat. Comput. Simul. 2007a, 77(6), s. 487-504.
- Gomes M.I., Pestana D., *A sturdy reduced-bias extreme quantile (VaR) estimator*, J. Am. Stat. Assoc. 2007b, 102(477), s. 280-292.
- Gomes M.I., *Estatística de Extremos e Desporto – como Estimar alguns Parâmetros Úteis*, Centro de Estatística e Aplicações da Universidade de Lisboa, 11/08, Lisboa 2008.

- Gross J.L., Heckert N.A., Lechner J.A., Simiu E., *A study of optimal extreme wind estimation procedures*, [w:] *Proceedings of the Ninth International Wind Engineering Conference*, New Delhi 1995.
- Gumbel E.J., *Les valeurs extrêmes des distribution statistiques*, Ann. Inst. Henri Poincaré 1935, 5, s. 115-158.
- Gumbel E.J., *Statistics of extremes*, Columbia University Press, New York 1958.
- Gumbel E.J., *The return period of flood flows*, Ann. Math. Stat. 1941, 12, s. 163-190.
- Haan L. de, Ronde J. de, *Sea and wind: multivariate extremes at work*, "Extremes" 1998, vol. 1, no.1, s. 7-45.
- Haan L. de, Sinha A.K., *Estimating the probability of a rare event*, "The Annals of Statistics" 1999, vol. 27, no 2, s. 732-759.
- Haan L. de, *Fighting the Arch-enemy with mathematics*, "Statistica Neerlandica" 1990, vol. 44, s. 45-68.
- Haan L. de, *On regular variation and its application to the weak convergence of sample extremes*, [w:] *Mathematical Centre Tracts*. 32. Math Centrum, Amsterdam 1970.
- Hill B., *A simple general approach to interference about the tail of a distribution*, "The Annals of Statistics" 1975, 3(5), s. 1163-1174.
- Hols M.C.A.B., Vries C.G. de, *The Limiting distribution of extremal exchange rate returns*, "Journal of Applied Econometrics" 1991, 6, s. 287-302.
- Horowitz J., *Extreme values from a nonstationary stochastic process: an application to air quality analysis*, "Technometrics" 1980, vol. 22, no 4, s. 469-478.
- Jansen D.W., Vries C.G. de, *On the frequency of large stock returns: putting booms and busts into perspective*, "The Review of Economics and Statistics" 1991, s. 18-24.
- Jenkinson A.F., *The frequency distribution of the annual maximum (or minimum) values of meteorological elements*, "Quarterly Journal of the Royal Meteorological Society" 1955, 81, s. 158-171.
- Klüppelberg C., Mikosch T., *Self-normalized and randomly centered spectral estimates*, [w:] *Proceedings of the Athens International Conference on Applied Probability and Time Series*, vol. 2. Time Series, Springer, Berlin 1996, s. 259-271.
- Lavenda B.H., Cipollone E., *Extreme value statistics and thermodynamics of earthquakes: large earthquakes*, [w:] "Annali di Geofisica" 2000, vol. 43, no 3.
- Leadbetter M.R., Lindgren G., Rootzén H., *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York 1983.
- Lindgren G., Rootzén H., *Extreme values: theory and technical applications*, "Scandinavian Journal of Statistics" 1987, vol. 14, no 4, s. 241-279.
- Longin F., *The asymptotic distribution of extreme stock market returns*, "Journal of Business" 1996, 63, s. 383-408.
- Loretan M., Phillips P.C.B., *Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial datasets*, "Journal of Empirical Finance" 1994, 1(2), s. 211-248.
- McNeil J., Frey R., Embrechts P., *Quantitative risk management: concepts, techniques and tools*, Princeton Press, 2005.
- Mikosch T., Nagaev A.V., *Large deviations of heavy-tailed sums with applications in insurance*, "Extremes" 1998, 1(1), s. 81-110.
- Pactwa T.E., *Using extreme values theory to value stock market returns*, [w:] *ETD Collection for Florida International University*, January, 2001.
- Pickands J., *Statistical inference using extreme order statistics*, "The Annals of Statistics" 1975, 3(1), s. 119-131.
- Pictet O.V., Dacorogna M.M., Müller U.A., Hill, *bootstrap and jackknife estimators for heavy tails*, [w:] R.J. Adler, R.E. Feldman, M. Taquq (red.), *Practical guide to heavy tails: statistical techniques and applications*, Birkhäuser, Boston 1998, s. 283-309.

- Singpurwalla N.D., *Extreme values from a lognormal law with applications to air pollution problems*, "Technometrics" 1972, vol. 14, no 3, s. 703-711.
- Smith R.L., *Estimating tails of probability distributions*, Ann. Statist. 1987, 15, s. 1174-1207.
- Smith R.L., *Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone*, "Statistical Science" 1989, vol. 4, no 4, s. 367-377.
- Strand M., Boes D., *Modeling road racing times of competitive recreational runners using extreme value theory*, "The American Statistician" 1998, vol. 52, no 3, s. 205-210.
- Vries C.G. de, *Stylized facts of nominal exchange rate returns*, [w:] F. Van Der Ploeg (red.), *The handbook of international macroeconomics*, Blackwell, Cambridge 1994, MA, s. 384-389.
- Weibull W.A., *The phenomenon of rupture in solids*, [w:] Ingeniörvetenskapsakademiens 149, Handlingar 1938.
- Zempléni A., *Goodnes-of-fit tests in extreme value applications*, "Colaborative Research Center 386, Discussion Paper 383", München 2004.
- Zwiers F.W., *An extreme-value analysis of wind speeds at five canadian locations an extreme-value analysis of wind speeds at five Canadian locations*, "The Canadian Journal of Statistics" 1987, vol. 15, no 4, s. 317-327.

ORDER STATISTICS – ESTIMATION OF THE DISTRIBUTION OF THE MAXIMAL LOSSES ON AN EMPIRICAL EXAMPLE THE WARSAW STOCK EXCHANGE

Summary: The article reviews the history of the Extreme Values Theory (EVT) with its mathematical background and empirical applications. There are also presented the results of an empirical study of the possibility of modeling the maximal stock market loss on the Warsaw Stock Exchange based on the data of its four most important indices.