



Adam Pawluczyk, Janusz A. Hołyst
Wydział Fizyki
Politechnika Warszawska
Koszykowa 75, PL-00-662 Warszawa, Polska



Abstrakt

Zgodnie z tytułem poniższa praca ukazuje obserwacje i podobieństwa dwóch typów procesów stochastycznych:

- teoretycznego: *proste błędzenie losowe (symetryczne i Bernoulliego)*
- oraz rzeczywistego: szeregi czasowe utworzone z giełdowych danych rzeczywistych

w świetle Teorii Wartości Ekstremalnych (TWE). W ramach pracy w języku JAVA został utworzony program symulujący *proste błędzenie losowe*. Dokonano teoretycznych opisów rozkładów wartości końcowych i wychyleń maksymalnych w funkcji liczby kroków oraz prawdopodobieństwa pójścia w prawo. Dodatkowo, przedstawiono rozkład wartości maksymalnych będący w zgodzie z podstawowym twierdzeniem TWE. Powyższe rozważania teoretyczne zostały poparte symulacjami dokonanymi w programie oraz potwierdzone testami χ^2 i r^2 .

W drugiej części pracy zajęto się porównaniem *błędzenia losowego* z giełdowymi szeregami czasowymi. Analizowano histogramy dziennych i ciągłych stóp zwrotu oraz maksymalnych stóp zwrotu szeregów czasowych indeksu *WIG-20* oraz spółek *KGHM* i *Żywiec*. Zastosowanie praw skalowania pozwoliło na zaobserwowanie grubych ogonów nie tylko na histogramie stóp zwrotu modelowanym rozkładem normalnym, ale także na histogramie maksymalnych stóp zwrotu modelowanym funkcją Gumbela. Ponadto badano zależność niepewności (odchylenia standardowego) w funkcji horyzontu (przedziału) czasowego zmiennych utworzonych z *prostego błędzenia losowego* oraz z danych rzeczywistych ($\sigma \sim \tau^a$). Dzięki symulacji wyznaczono numerycznie wykładnik teoretyczny $a_{teoria} = 0.5$ i porównywano go z wartościami wyznaczonymi z szeregów rzeczywistych.

Wprowadzenie

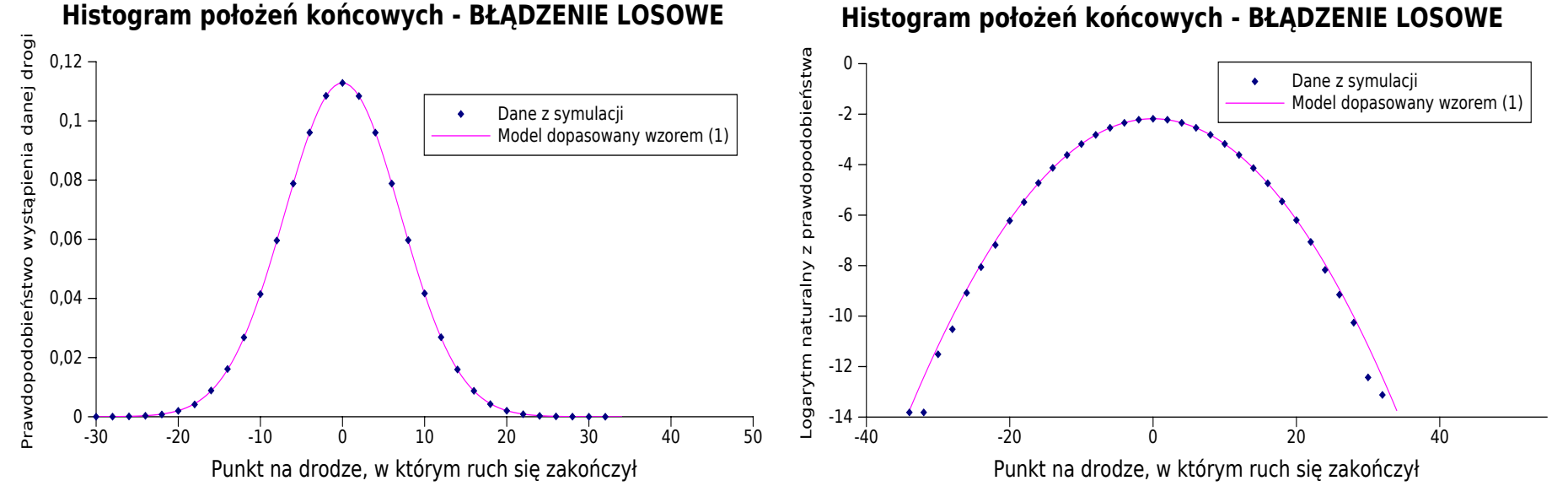
Niepewność ruchów giełdowych od lat była przedmiotem badań naukowców. Jednym z procesów stochastycznych symulujących fluktuacje na giełdzie jest *proste błędzenie losowe*. Ta praca miała na celu porównanie różnych właściwości obu tych procesów, ze szczególnym uwzględnieniem wartości maksymalnych. Pierwszym porównaniem było zestawienie rozkładu wartości (położeń) końcowych prostego błędzenia losowego symetrycznego z histogramem stóp zwrotu. Drugą cechą porównywaną były rozkłady położeń maksymalnych z rozkładem maksymalnych stóp zwrotu. Ostatnim porównaniem było zestawienie niepewności (odchylenia standardowego) odległości położeń w błędzeniu losowym z niepewnością stopy zwrotu w funkcji horyzontu czasowego.

Rozkłady stóp zwrotu

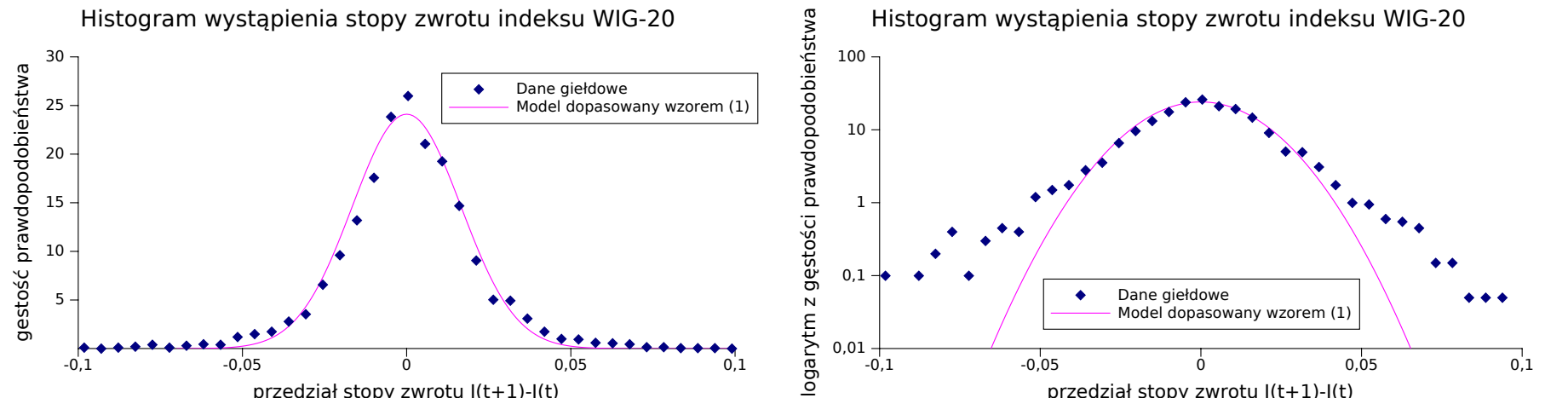
W ramach pracy powstał program w języku JAVA, generujący rozkłady wartości końcowych oraz wychyleń maksymalnych (zasiegu) prostego błędzenia losowego Bernoulliego. W programie można było regulować prawdopodobieństwo pójścia w prawo, liczbę kroków oraz liczbę iteracji. Zgodnie z założeniami teoretycznymi, potwierdzonymi przez symulację w programie, do rozkładu wartości końcowych błędzenia Bernoulliego można dopasować gęstość rozkładu normalnego w funkcji liczby kroków, prawdopodobieństwa pójścia w prawo oraz liczby obserwacji:

$$\mathbf{P}(S_n = x) = \begin{cases} \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(x-n(2p-1))^2}{8np(1-p)}\right) & \text{dla } x \in \{-n, -n+2, \dots, n\} \\ 0 & \text{dla } x \in \{-n+1, -n+3, \dots, n-1\}. \end{cases} \quad (1)$$

gdzie S_n to suma kroków, p to prawdopodobieństwo pójścia w prawo, n to liczba iteracji. Poniżej widnieje histogram wartości końcowych położenia w skali proporcjonalnej i półlogarytmicznej dla prawdopodobieństwa pójścia w prawo $p = 0.5$, 50 kroków i 10^6 iteracji.



Biorąc pod uwagę powyższe do rzeczywistych danych także dopasowano rozkład normalny. Obserwowano dzienne stopy zwrotu indeksu WIG-20 z lat 1994-2009, a następnie zgodnie z zasadami statystycznymi utworzono histogramy tych stóp.



Porównując histogram rzeczywistych stóp zwrotu z tym wygenerowanym z prostego błędzenia losowego daje się zauważyć leptokurtyczne „grube ogony”, czyli wartości maksymalne o wiele bardziej prawdopodobne niż zakłada to model dany równaniem (1).

Rozkłady maksymalnych stóp zwrotu

Praca korzystała z Teorii Wartości Ekstremalnych (TWE). Przy omawianiu TWE warto przytoczyć tu podstawowe dla tej teorii twierdzenie. Jeżeli zmienne losowe $X_1, X_2, \dots, X_N$ są typu i.i.d. (są niezależne oraz pochodzą z identycznego rozkładu), to przy $N \rightarrow \infty$ do empirycznej dystrybucji utworzonej z wartości ekstremalnych tych zmiennych można dopasować jedną z trzech funkcji:

$$F_\mu(x) = \exp(-x^{-\mu}) \text{ dla } x \geq 0 \text{ i } 0 \text{ dla } x < 0 \ (\mu > 0); \quad (2)$$

$$W_\mu(x) = \exp(-(x)^{-\mu}) \text{ dla } x \leq 0 \text{ i } 0 \text{ dla } x > 0 \ (\mu < 0); \quad (3)$$

$$G(x) = \exp(-e^{-x}) \text{ dla } x \in \mathbf{R}. \quad (4)$$

Dystrybuanty niosą nazwy odpowiednio: Frecheta, Weibulla oraz Gumbela. Jako, że w pracy wykorzystuje się tylko trzecią funkcję, trzeba wyprowadzić jej gęstość prapodobieństwa. Różniczkując (4) otrzymuje się:

$$g(x) = \exp(-e^{-x}) \cdot e^{-x}. \quad (5)$$

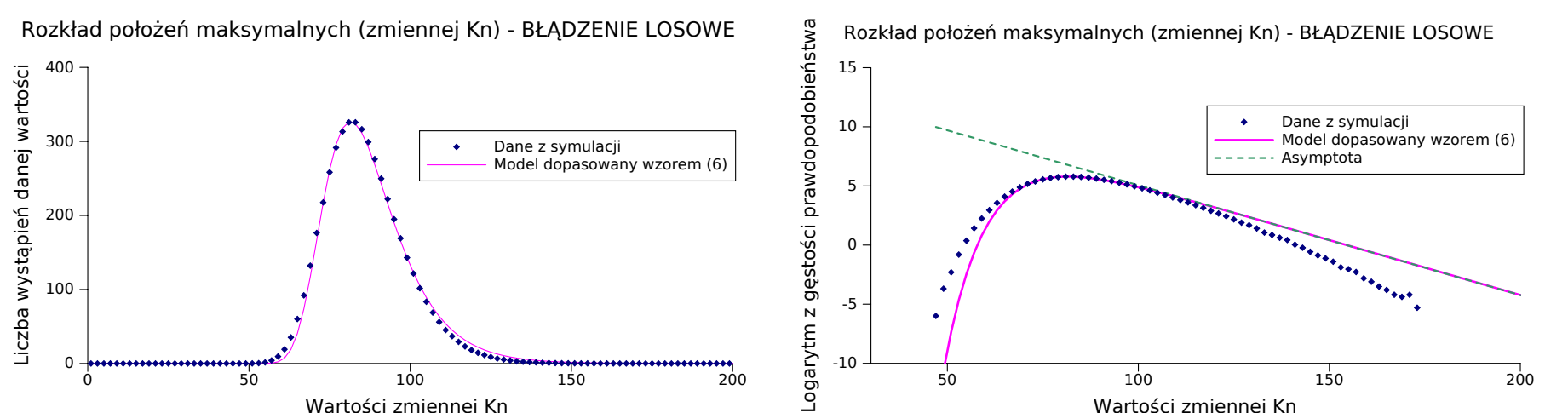
Ponadto, aby móc modelować tą funkcją rozkłady rzeczywiste należy wzbogacić powyższy wzór o parametry: położenia μ i skali σ :

$$g_{\mu,\sigma}(x) = \frac{1}{\sigma} \exp(-e^{-(x-\mu)/\sigma}) \cdot e^{-(x-\mu)/\sigma}. \quad (6)$$

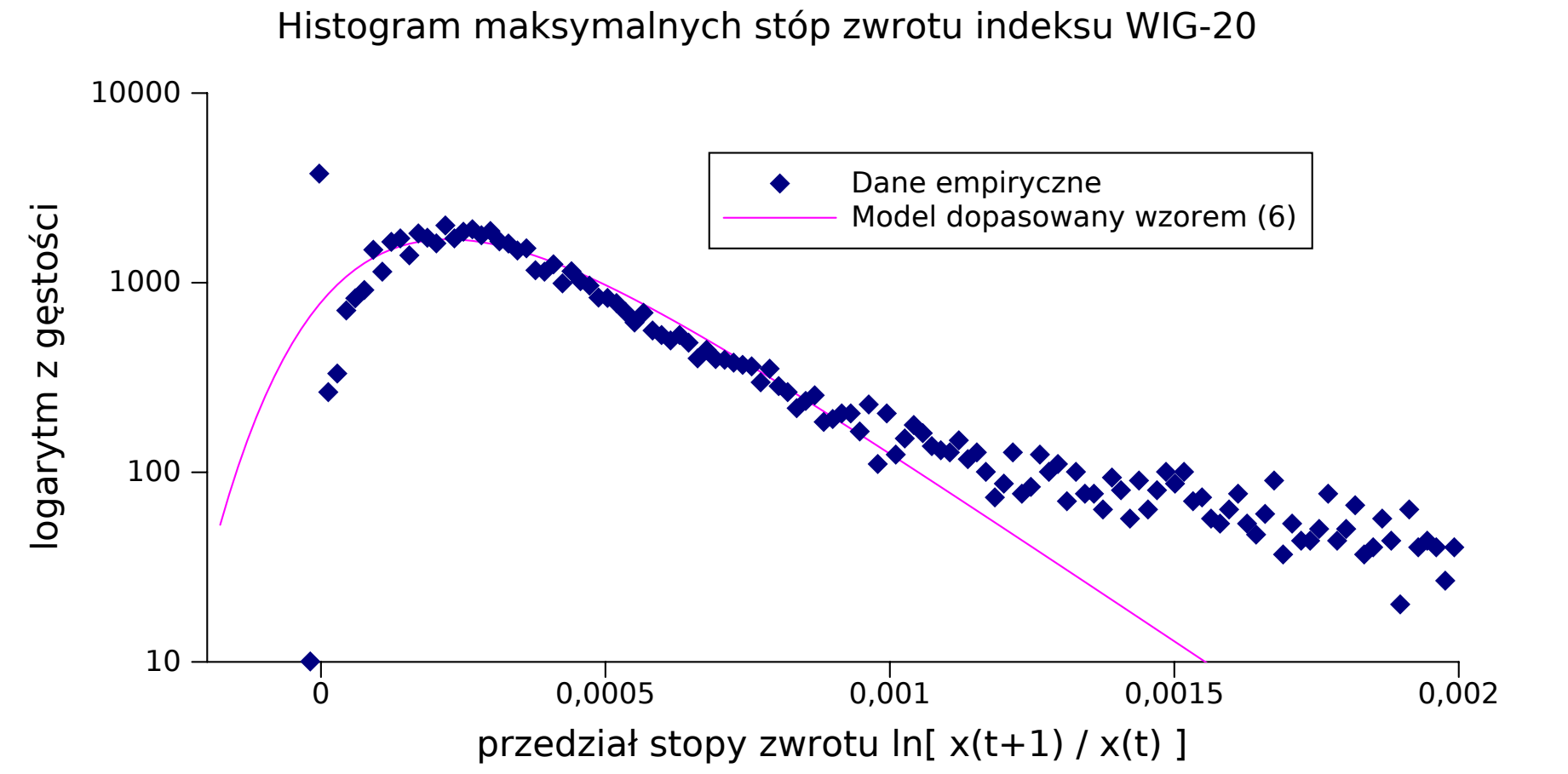
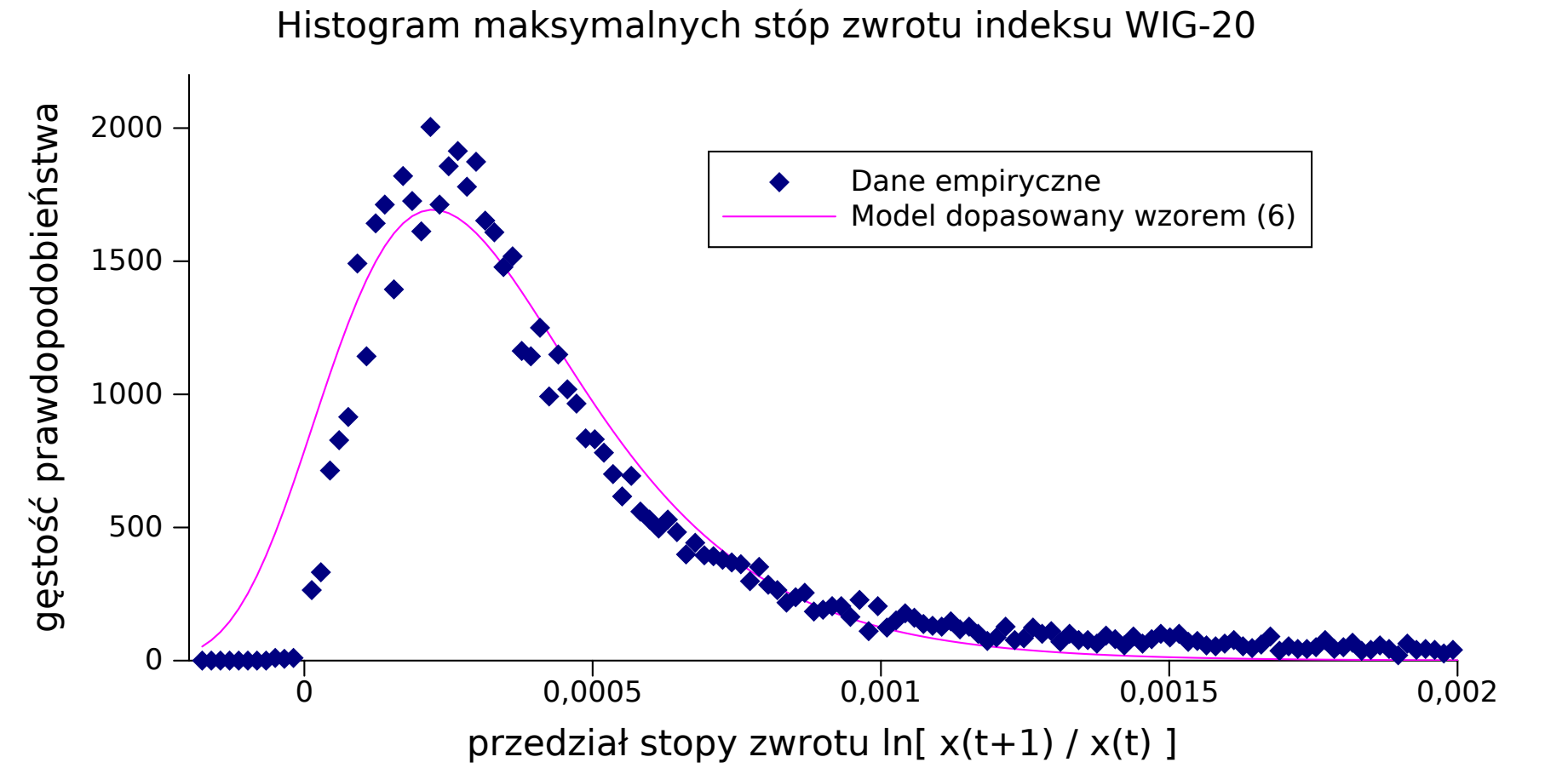
Początkowo użyto tego wzoru do modelowania histogramu wartości maksymalnych zmiennej:

$$K_l = \max\{S_n^i : 1 \leq i \leq l\} \text{ dla } l \in \mathbf{N}, \quad (7)$$

czyli maksymalnego położenia końcowego z l kolejnych iteracji prostego błędzenia losowego. Otrzymane wyniki w skali proporcjonalnej i półlogarytmicznej prezentują się poniżej:



Następnie zbadano rozkład maksymalnych ciągłych (jednostką czasu jest 15 sekund) stóp zwrotu indeksu WIG-20. Pogrupowano je w jednogminutowe okresy (4 wartości co 15 sekund) i wybierano spośród nich wartości maksymalne. Wyniki w skali proporcjonalnej i półlogarytmicznej są przedstawione poniżej:



Powyższe wykresy przedstawiają histogram empiryczny oraz dopasowanie funkcją (6). Na skali podwójnie logarytmicznej wyraźnie widać charakterystyczny "gruby ogon", co oznacza, że wartości maksymalne na Warszawskiej Giełdzie są bardziej prawdopodobne niż zakłada to funkcja Gumbela.

Niepewność w funkcji horyzontu czasowego

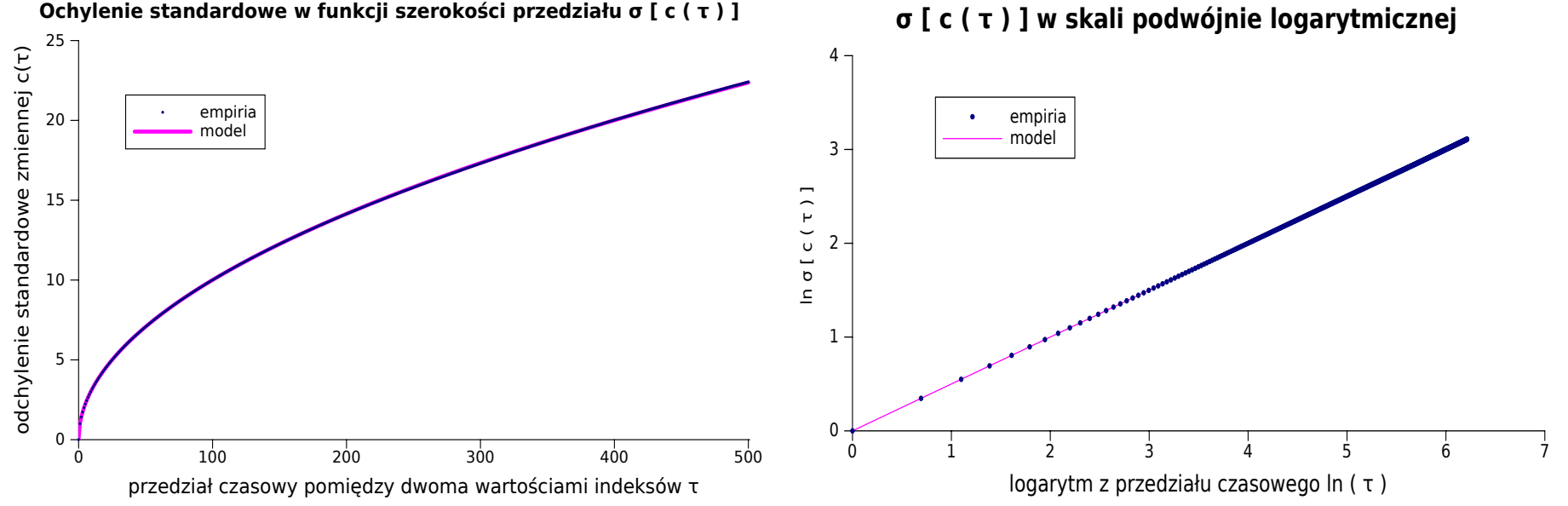
Trzecią porównywaną cechą była niepewność stopy zwrotu (odchylenie standardowe) w funkcji przedziału czasowego. Oznaczając za $s(t)$ położenie punktu w chwili t , a za $s(t+\tau)$ położenie punktu w chwili $t+\tau$, wprowadzamy nową zmienną $c_\tau(t)$:

$$c_\tau(t) \equiv s(t+\tau) - s(t) \text{ dla } t \in \{0, 1, 2, \dots, n-\tau\}, \quad (8)$$

gdzie $\tau \in \{1, 2, \dots, n\}$, a n to liczba kroków. W przypadku prostego błędzenia losowego symetrycznego zależność między odchyleniem standardowym zmiennej c a horyzontem czasowym ma wzór:

$$\ln \sigma[c_\tau] = a \cdot \ln \tau. \quad (9)$$

Przeprowadzono symulację w celu sprawdzenia powyższej zależności - wyniki są przedstawione poniżej w skali logarytmicznej i podwójnie logarytmicznej:



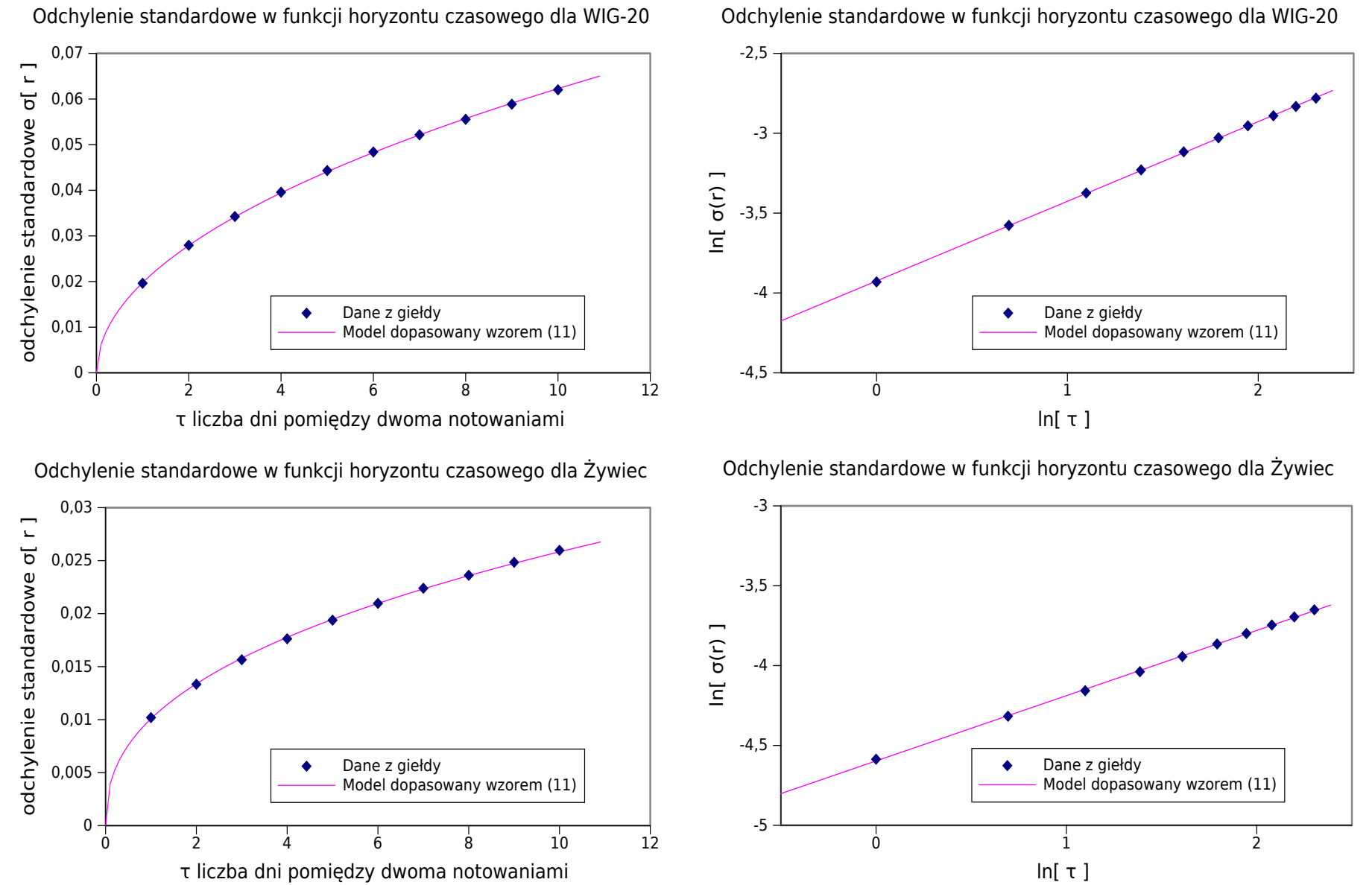
Z wykresu widać, że nachylenie w skali podwójnie logarytmicznej jest równe 0.5, co potwierdza teorię. Ponadto wykonano podobne obliczenia dla rzeczywistych szeregów czasowych. Stopa zwrotu po czasie τ wynosi:

$$r_\tau(t) \equiv \ln \frac{b(t+\tau)}{b(t)} = I(t+\tau) - I(t) \text{ dla } t \in \{1, 2, \dots, n-\tau\}, \quad (10)$$

gdzie $b(t)$ to wartość indeksu w chwili t a $I(t) \equiv \ln(b(t))$. W przypadku giełdowych danych rzeczywistych dopasowano funkcję wzbogaconą o parametr b :

$$\sigma[r_\tau] = e^b \cdot \tau^a, \quad (11)$$

Poniższe wykresy pokazują zależność odchylenia w funkcji horyzontu czasowego:



Parametr a z dopasowań do danych rzeczywistych w przypadku indeksu WIG-20 wyszedł zbliżony do współczynnika teoretycznego ($a = 0.50$), a dla spółki Żywiec odbiega od wykładnika teoretycznego o 20% ($a = 0.41$).

Konkluzje

Powyższa praca udowadnia, że teoria wartości maksymalnych ma zastosowanie do prostego błędzenia losowego. Ponadto udało się zaobserwować "grube ogony" na histogramie wartości maksymalnych stóp zwrotu danych rzeczywistych. Po trzecie potwierdzono, że niepewność w funkcji horyzontu czasowego w skali podwójnie logarytmicznej rośnie liniowo, a dla niektórych spółek wartość nachylenia jest bardzo bliska nachyleniu zaobserwowanemu przy prostym błędzeniu losowym Bernoulliego.

Literatura

- R.D. Reiss, M. Thomas, *Statistical Analysis of extreme values with Applications to Insurance, Finance, Hydrology and Other Fields*, Berlin 2001, 3-7, 66-67
- M. Clusel, E. Bertin, *Global fluctuations in physical systems: a subtle interplay between sun and extreme value statistics*, *Int. J. Mod. Phys. B* 22, 3311 (2008)
- R.N. Mantegna, H.E. Stanley, *Ekonofizyka - wprowadzenie*, Warszawa 2001, 32-33, 114
- J. Józwiak, J. Podgórski, *Statystyka od podstaw*, Warszawa 2009, 239-242, 423-425, 484-485.
- G. Grimmet, D. Stirzaker, *Probability and Random Processes*, Oxford 2001, 78-79
- K. Klimañska, *Indeks ekstremalny stacjonarnych szeregów czasowych względem błędzenia losowego według algorytmu Metropolis'a* [on-line], [dostęp 1 grudnia 2009]. Dostępny w World Wide Web: <http://sites.google.com/site/kklimanska/>
- Freakwave* [on-line], portal *BBC* [dostęp 1 grudnia 2009]. Dostępny w World Wide Web: <http://www.bbc.co.uk/science/horizon/2002/freakwave.shtml>
- E. Alvarado, D. Sandberg, S.G. Pickford, *Modeling large forest fires as extreme events*, *Northwest Science* 72 (1998), 66-75
- J.H.J. Einmahl, *S.G.W.R. Smeets, Ultimate 100m World Records Through Extreme-Value Theory*, *CentER Discussion Paper*, Tilberg University, 57 (2009)
- D. Bobrowski, *Ciągi losowe*, Poznań 2002
- P. Grzegorzewski, K. Bobecka, A. Dembiñska, J. Pusz, *Rachunek prawdopodobieństwa i statystyka, wydanie piąte, poprawione*, Warszawa 2008, 154-155
- B. Malkiel, *Błądząc po Wall Street*, Warszawa 2003, 4-5.
- J. Jakubowski, R. Stencel, *Rachunek prawdopodobieństwa dla (prawie) każdego*, Warszawa 2002, 147
- K. Pearson, *The problem of the Random Walk*, *Nature*. 72, 294 (1905)
- A.A. Griffith, *The Phenomenon of Rupture and Flow in Solids*. *Philosophical Transactions of the Royal Society of London. Series A* 221 (1921), 163-198