

Poniższe trzy artykuły były prezentowane podczas XV Konferencji Naukowej Młodych Ekonomistów, która odbyła się we wrześniu 2009 r. w Warszawie. Była ona zorganizowana przez WNE UW, SGH, SGGW, Politechnikę Warszawską oraz Instytut Wiedzy i Innowacji.

Modelowanie zamożności polskich gospodarstw domowych metodami statystycznymi

Maciej Jagielski, mgr, Wydział Fizyki, Uniwersytet Warszawski

Ryszard Kutner, prof. dr hab., Zakład Fizyki Biomedycznej, Instytut Fizyki Doświadczalnej, Wydział Fizyki, Uniwersytet Warszawski

Wstęp

Żyjemy w czasach, w których do opisu złożoności otaczającego nas świata nie wystarczy już posługiwanie się metodologiami, jakie oferują poszczególne dyscypliny naukowe. Coraz częściej konieczne jest połączenie metod stosowanych do tej pory w zupełnie różnych, czasem nawet bardzo odległych od siebie dziedzinach nauki. Jedną z takich dyscyplin jest ekonofizyka, która przy użyciu metod i modeli stosowanych w fizyce opisuje wybrane procesy ekonomiczne, a także wzbogaca metodologię ekonomii. Takie interdyscyplinarne podejście prowadzi do głębszego zrozumienia istoty tych procesów.

Na przykład interdyscyplinarne spojrzenie na problemy dzisiejszych gospodarstw domowych daje interesujące rezultaty w przypadku analizy dochodów ludności w Polsce. Temat ten został omówiony w niniejszej pracy. Przedstawiono tutaj analizę empirycznych skumulowanych rozkładów rocznych dochodów gospodarstw domowych w Polsce, przykładowo, w latach 2000 i 2006. Rozkłady te porównano z przewidywaniami praw Pareto, Prawa Efektów Proporcjonalnych, uogólnionego modelu Lotka-Volterra oraz modeli zderzeń.

1. Opis bazy danych

W analizie wykorzystano dane Głównego Urzędu Statystycznego dotyczące rocznych dochodów gospodarstw domowych w latach 2000 oraz 2006. Po usunięciu niewielkiej liczby rekordów (252 w 2000 r. oraz 231 w 2006 r.), dla których dochód był mniejszy bądź równy zero, baza danych zawierała odpowiednio 35 911 i 37 277 obserwacji. Niedodatnie wartości dochodu uznano za obserwacje błędne, gdyż zgodnie z definicją dochód do dyspozycji to:

Dochód rozporządzalny pomniejszony o pozostałe wydatki. Dochód do dyspozycji przeznaczony jest na wydatki na towary i usługi konsumpcyjne oraz przyrost oszczędności [GUS, 2009].

2. Analiza rocznych dochodów gospodarstw domowych w Polsce

Celem niniejszego opracowania jest analiza empirycznych rozkładów skumulowanych rocznych dochodów gospodarstw domowych. Rozkłady te zostały skonstruowane zgodnie z następującą definicją:

$$\Pi(m) = \frac{1}{N} \sum_{i=1}^N I(m_i > m) \quad (1)$$

gdzie:

m — roczny dochód do dyspozycji,

N — liczba obserwacji w próbie,

$I(m_i > m)$ — funkcja indykatorowa postaci:

$$I(m) = \begin{cases} 1 & \text{dla } m_i > m \\ 0 & \text{dla } m_i \leq m \end{cases} \quad (2)$$

A zatem, dla ustalonego m zliczano gospodarstwa domowe indeksowane i ($i = 1, 2, \dots, N$), których dochód jest większy niż m , i dzielono przez liczbę obserwacji w próbie. Wartość m zmieniano co 1000 zł (jest to krok dyskretyzacji histogramu rozkładu), poczynając od 0.

W pierwszej kolejności do „ogonów” powyższych rozkładów dopasowano słabe prawo Pareto [Richmond i in., 2006] (rys. 1.; skala log-log)¹:

$$\Pi(m) \approx (m/m_0)^{-\alpha} \quad (3)$$

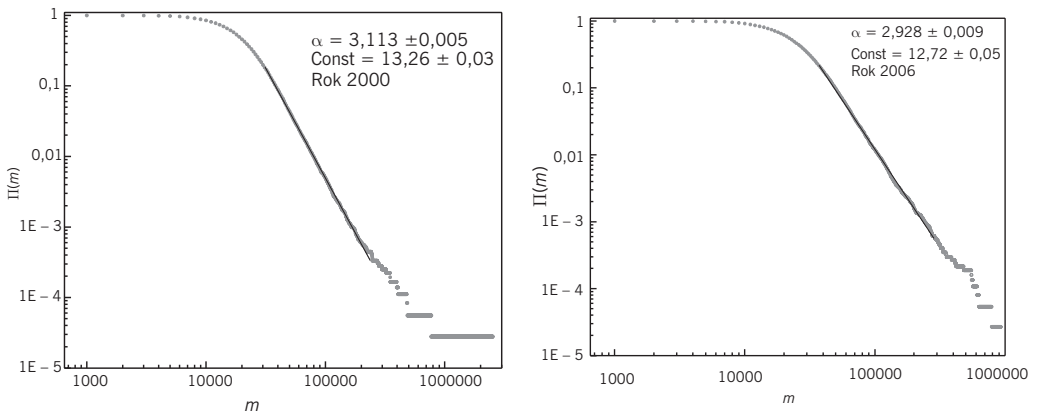
gdzie:

m_0 — czynnik skalujący²: $m_0 = \exp(const/\alpha)$,

α — wykładnik Pareto.

¹ W przypadku wszystkich przedstawionych analiz wyniki dla roku 2003 można znaleźć w pracy [Jagielski, 2009] oraz [Jagielski, Kutner, 2010].

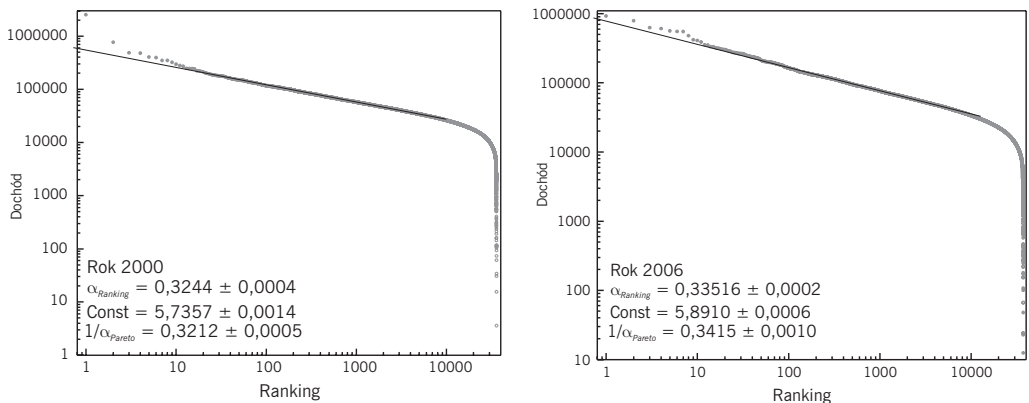
² Dla roku 2000 parametr m_0 wynosi $70,8 \pm 0,8$ zł a dla roku 2006: 77 ± 2 zł.



Rys. 1.

Dopasowanie słabego prawa Pareto (linia ciągła) do empirycznego skumulowanego rozkładu dochodów gospodarstw domowych w Polsce (punkty) z roku 2000 ($\alpha = 3,113 \pm 0,005$) i 2006 ($\alpha = 2,928 \pm 0,009$)

Na rysunku 1. parametr α oznacza wykładnik Pareto (tzn. $\alpha = \alpha_{Pareto}$). Z kolei $const$ jest stałą, pochodzącą z dopasowania prostej (w skali log-log). Możemy zauważyć, że słabe prawo Pareto dobrze opisuje ogony analizowanych rozkładów. Aby uzyskać tak dobre dopasowania konieczne było usunięcie ok. 20–40 obserwacji skrajnych, czyli gospodarstw domowych osiągających największe dochody (szara część krzywej w dolnej części wykresów). Usunięte obserwacje stanowiły ok. 0,06% badanej populacji.



Rys. 2.

Ranking gospodarstw domowych (linia ciągła oznacza dopasowanie, a punkty dane empiryczne) — rok 2000 oraz 2006

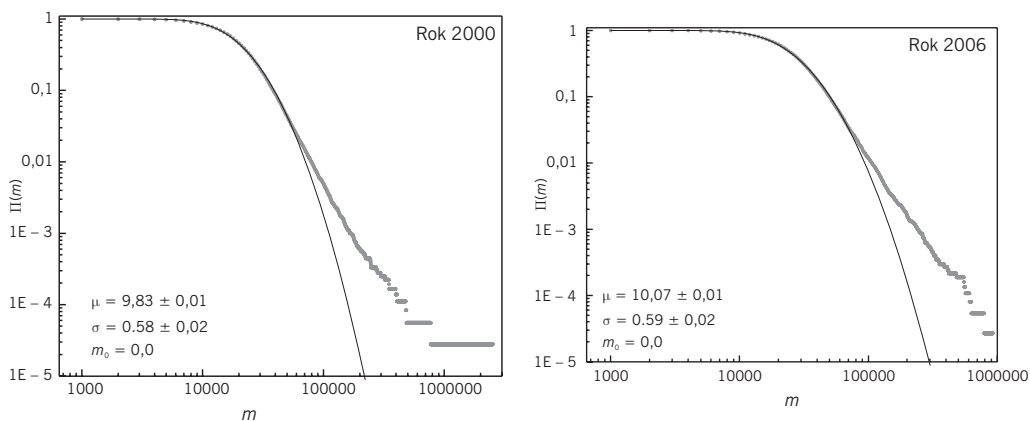
Wartość wykładnika Pareto otrzymujemy także analizując ranking gospodarstw domowych, czyli wykres dochodów gospodarstw domowych w zależ-

ności od ich miejsca w rankingu. Odpowiednie wykresy w skali logarytmicznej zamieszczono na rysunku 2.

Średniozamożne gospodarstwa domowe opisywane są w skali logarytmicznej prostą o nachyleniu $\alpha_{ranking}$. Okazuje się, że odwrotność wykładnika Pareto uzyskanego w przypadku rozkładów skumulowanych rocznych dochodów gospodarstw domowych jest bardzo zbliżona do wyznaczonych wartości $\alpha_{ranking}$. Wskazuje to na zgodność obu zastosowanych metod [Bouchard, Poters, 2001].

Do omawianych rozkładów skumulowanych dopasowano również skumulowany rozkład log-normalny, wynikający z Prawa Efektów Proporcjonalnych [Armatte, 1995]. Wyniki — w skali logarytmicznej — zostały zaprezentowane na rysunku 3.³

$$\Pi(m) = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{\ln(m - m_0) - \mu}{\sqrt{2}\sigma} \right) \right] \quad (4)$$



Rys. 3.

Dopasowanie skumulowanego rozkładu log-normalnego (linia ciągła) do skumulowanego rozkładu dochodów gospodarstw domowych w Polsce (punkty) z roku 2000 i 2006

Skumulowany rozkład log-normalny bardzo dobrze opisuje skumulowane dochody ubogich gospodarstw domowych. Fakt, że skumulowany rozkład log-normalny wykazuje tak dobrą zgodność z danymi empirycznymi wynika bezpośrednio z założeń Prawa Efektów Proporcjonalnych. W modelu tym zakłada się, że zmiany dochodu mogą być niewielkie, co znajduje uzasadnienie w przypadku ubogich gospodarstw domowych [Armatte, 1995]. Z kolei słabe prawo Pareto jest odpowiednie do opisu dochodów średniozamożnych

³ Parametr m_0 wynosi 0, gdyż nie wpływa on na wyniki estymacji pozostałych parametrów; w celu dopasowania krzywych posłużono się algorytmem Levenberga-Marquardta.

gospodarstw domowych, ponieważ w tym przypadku zaczyna odgrywać rolę konkurencja rynkowa. Firmy konkurują między sobą, a ich zyski przekładają się na zarobki pracowników — jednostek tworzących gospodarstwa domowe [Yamamoto i in., 2006]. Dlatego też analiza wykresów na rysunkach 1. oraz 3. nasuwa pytanie o wartość dochodu w punkcie przecięcia się skumulowanego rozkładu log-normalnego i słabego prawa Pareto. Punkt ten może stanowić umowną, lecz zarazem wystarczająco precyzyjną (ze względu na niewielki błąd) granicę podziału między ubogimi a średniozamożnymi gospodarstwami domowymi. Graniczna wartość rocznego dochodu stanowiąca rozróżnienie między ubogimi a średniozamożnymi gospodarstwami domowymi wynosi około:

- 29 000 złotych w roku 2000,
- 36 000 złotych w roku 2006.

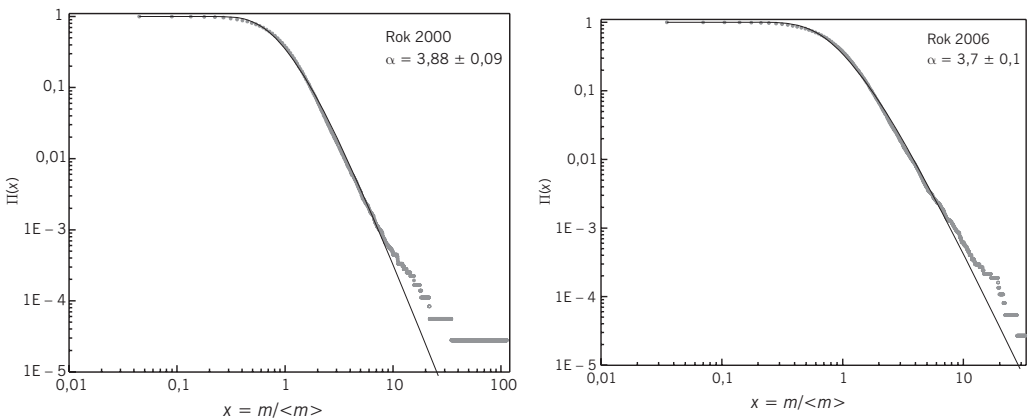
Do rozważanych rozkładów empirycznych dopasowano także rozkład skumulowany pochodzący z uogólnionego modelu Lotka-Volterra — rys. 4.⁴ [Richmond, Solomon, 2001; Solomon, Richmond, 2002]:

$$\Pi(x) = 1 - \frac{\Gamma\left(\alpha, \frac{\alpha - 1}{x}\right)}{(\alpha)} \quad (5)$$

gdzie:

α — wykładnik kształtu opisujący dopasowaną funkcję,

$x = m/\langle m \rangle$ — względny dochód gospodarstwa domowego ($\langle m \rangle$ — średnia wartość dochodu gospodarstwa domowego w analizowanej próbie).



Rys. 4.

Dopasowanie uogólnionego modelu Lotka-Volterra (linia ciągła) do skumulowanego rozkładu dochodów gospodarstw domowych w Polsce (punkty) z roku 2000 ($\alpha = 3,88 \pm 0,09$) i 2006 ($\alpha = 3,7 \pm 0,1$)

⁴ Skala log-log; w celu dopasowania krzywych posłużono się algorytmem Levenberga-Marquardta.

Choć uogólniony model Lotka-Volterra nie opisuje tak dobrze skumulowanego rozkładu rocznych dochodów gospodarstw domowych jak skumulowany rozkład log-normalny i Pareto, to jednak różnice te są niewielkie. Istotną zaletą tego modelu jest możliwość scharakteryzowania analizowanych rozkładów za pomocą pojedynczej funkcji. Oferuje on także cenne podejście teoretyczne na poziomie mikroskopowym, w którym dochód gospodarstwa domowego determinowany jest przez dochody dotychczas uzyskiwane i przez wysokość świadczeń pomocy społecznej (ogólniej mówiąc redystrybucję dochodów w społeczeństwie) oraz ogólny stan gospodarki [Richmond, Solomon, 2001; Solomon, Richmond, 2002].

W modelu Lotka-Volterra można wyznaczyć efektywną minimalną roczną wartość dochodu m_{\min} :

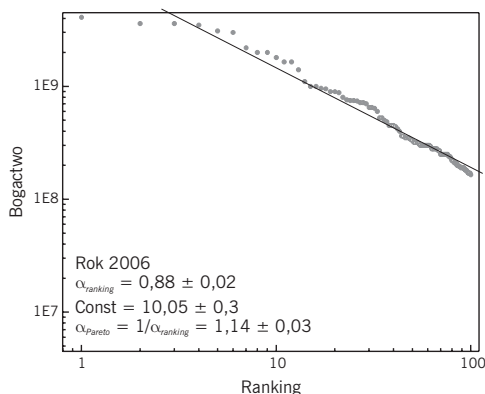
$$m_{\min} = \frac{\alpha - 1}{\alpha + 1} \langle m \rangle, \quad (6)$$

poniżej której prawdopodobieństwo osiągnięcia mniejszych dochodów gwałtownie maleje

Na podstawie powyższego wzoru otrzymujemy minimalne wartości rocznego dochodu:

- w roku 2000: $m_{\min} = 13\,183,88$ zł dla $\langle m \rangle = 22\,339,35$ zł;
- w roku 2006: $m_{\min} = 16\,373,32$ zł dla $\langle m \rangle = 28\,501,70$ zł.

W dotychczasowych rozważaniach nie analizowano dochodów najbogatszych gospodarstw domowych, gdyż obserwacje takie były zbyt mało liczne, aby można było je poddać opisowi statystycznemu. Problem ten można pośrednio rozwiązać analizując np. wartość majątku 100 najbogatszych Polaków w postaci rankingu⁵ (rys. 5.; skala logarytmiczna).



Rys. 5.

Ranking 100 najbogatszych Polaków z roku 2006 ($\alpha_{\text{Pareto}} = 1,14 \pm 0,03$; linia ciągła oznacza dopasowanie, a punkty — dane empiryczne)

⁵ Dane o majątku 100 najbogatszych Polaków w roku 2006 zostały uzyskane ze strony internetowej tygodnika „Wprost” [Rankingi Wprost, 2009].

Zgodnie z tym, co pokazano wcześniej, jeśli ranking najbogatszych (krezusów) można w skali logarytmicznej opisać zależnością liniową, to wówczas dochody krezusów podlegają słabemu prawu Pareto o wykładniku $\alpha_{Pareto} = 1/\alpha_{ranking}$. Choć w tym przypadku zamiast o dochodach mówimy o bogactwie, to jednak z własności rozkładu Pareto wynika, że nie wpływa to na obserwowane uniwersalności [Mandelbrot, 1960].

Otrzymane wartości wykładnika Pareto są bliskie jedności, co jest zgodne z wynikiem teoretycznym uzyskanym na podstawie modelu zderzeń ze zróżnicowaną skłonnością jednostek do oszczędzania [Bhattacharyya i in., 2005; Chatterjee, Chakrabarti, 2007 i referencje tamże]. Zatem można oczekiwać, że bogacenie się klasy średniej będzie polegało (z formalnego punktu widzenia) na maleniu wykładnika Pareto. Z kolei mechanizm zdobywania dochodów przez najbogatszych, podobnie jak w przypadku średniozamożnych gospodarstw domowych, opiera się na istnieniu konkurencji rynkowej [Yamamoto i in., 2006]. W tym przypadku jednostki tworzące gospodarstwa domowe są przeważnie właścicielami (konkurującymi) firm, z których czerpią swój dochód. Dochody tych firm opisywane są prawem Zipfa [Okuyama i in., 1999], czyli właśnie słabym prawem Pareto z wykładnikiem $\alpha_{Pareto} = 1$.

Zakończenie

W pracy dokonano analizy empirycznych rozkładów skumulowanych dotyczących rocznych dochodów gospodarstw domowych w Polsce w latach 2000 i 2006. Okazało się, że mogą one być bardzo dobrze opisywane zarówno skumulowanym rozkładem log-normalnym w przypadku ubogich gospodarstw domowych, jak również słabym prawem Pareto dla gospodarstw średniozamożnych. Punkt przecięcia obu rozkładów może wyznaczać granicę podziału między ubogimi, a średniozamożnymi gospodarstwami domowymi. Fakt, że skumulowany rozkład log-normalny wykazuje tak dobrą zgodność z danymi empirycznymi wynika bezpośrednio z założeń Prawa Efektów Proporcjonalnych. W modelu tym zakłada się, że zmiany dochodu mogą być niewielkie, co znajduje uzasadnienie w przypadku ubogich gospodarstw domowych.

Wśród rozważanych modeli obiecujące wyniki uzyskano także dla uogólnionego modelu Lotka-Volterra. Model ten w dobrym przybliżeniu odzwierciedla analizowane rozkłady oraz oferuje cenne podejście teoretyczne na poziomie mikroskopowym, w którym dochód gospodarstwa domowego jest określony jest przez dochody dotychczas uzyskiwane i przez wysokość świadczeń pomocy społecznej (redystrybucję dochodów w społeczeństwie) oraz aktualny stan gospodarki. Dzięki temu, iż wykładnik Pareto opisujący dochody średniozamożnych gospodarstw domowych można uzyskać zarówno poprzez analizę rozkładów skumulowanych, jak i rankingu, udało się także wyznaczyć jego wartość dla najbogatszych jednostek społeczeństwa polskiego. A zatem skumulowane roczne dochody gospodarstw domowych mogą być opisane w całym zakresie za pomocą rozkładów:

- skumulowanego log-normalnego w przypadku ubogich gospodarstw domowych,
- skumulowanego słabego rozkładu Pareto z wykładnikiem równym około 3 w przypadku średniozamożnych gospodarstw domowych,
- skumulowanego słabego rozkładu Pareto z wykładnikiem równym około 1 w przypadku najbogatszych gospodarstw domowych.

Otrzymane rezultaty dają impuls do dalszych badań. Przede wszystkim interesującym byłoby dokonanie analogicznej analizy w przypadku województw i porównanie jej z wynikami uzyskanymi w skali całego kraju. Cennych informacji dostarczyłyby również studia nad dochodami społeczeństw innych krajów Europy. Zebrane w ten sposób informacje mogłyby posłużyć budowie bardziej zaawansowanych modeli teoretycznych uwzględniających zarówno zróżnicowany terytorialnie charakter zamożności społeczeństw, jak też dynamikę zmian ich dochodów.

Podziękowania

Składamy podziękowanie Głównemu Urzędowi Statystycznemu za udostępnienie nieidentyfikowalnych danych źródłowych dotyczących rocznych dochodów gospodarstw domowych w Polsce w latach 2000, 2003 oraz 2006.

Bibliografia

- Armatte M., 1995, *Robert Gibrat la loi de l'effet proportionnel*, „Mathématiques et sciences humaines” nr 129.
- Bhattacharyya P., Chatterjee A., Chakrabarti B.K., 2005, *A common mode of origin of power laws in models of market and earthquake*, „Physica A” nr 381.
- Bouchard J.-P., Potters M., 2001, *Theory of financial risks: from statistical physics to risk management*, Cambridge University Press, Cambridge.
- Chatterjee A., Chakrabarti B.K., 2007, *Kinetic exchange models for income and wealth distributions*, „The European Physical Journal B” nr 60.
- Główny Urząd Statystyczny, 2009, <http://www.stat.gov.pl/>.
- Jagielski M., 2009, *Badanie zamożności gospodarstw domowych w Polsce metodami egzotycznej i tradycyjnej fizyki statystycznej*, praca magisterska dostępna w bibliotece IFD Wydziału Fizyki Uniwersytetu Warszawskiego.
- Jagielski M., Kutner R., 2010, *Study of households' income in Poland by using the statistical physics approach*, „Acta Physica Polonica A” t. 117, z. 4.
- Mandelbrot B., 1960, *New Methods in Statistical Economics*, „The Journal of Political Economy” nr 71.
- Okuyama K., Takayasu M., Takayasu H., 1999, *Zipf's law in income distribution of companies*, „Physica A” nr 269.
- Rankingi Wprost, 2009, <http://100najbogatszych.wprost.pl/>.
- Richmond P., Hutzler S., Coelho R., Repetowicz P., 2006, *A Review of Empirical Studies and Models of Income Distributions*, w: Chakrabarti B.K., Chakraborti A., Chatterjee A., *Society in Econophysics and Sociophysics: Trends and Perspectives*, WILEY-VCH, Weinheim.
- Richmond P., Solomon S., 2001, *Power laws are disguised Boltzmann laws*, „International Journal of Modern Physics C” nr 12.

- Solomon S., Richmond P., 2002, *Stable power laws in variable economies; Lotka-Volterra implies Pareto-Zipf*, „The European Physical Journal B” nr 27.
- Yamamoto K., Miyazima S., Yamamoto H., Ohtsuki T., Fujihara A., 2006, *The power-law exponent and the competition rule of the high income model*, w: Takayasu H., *Practical Fruits of Econophysics: Proceedings of the Third Nikkei Econophysics Symposium*, Springer, Tokyo.

A b s t r a c t **Wealth Modeling of Polish Households Using Statistical Methods**

A

The aim of this work is to analyze the empirical cumulative layouts of annual household income in Poland between 2000 and 2006. Data were collected from the Central Statistical Office. The layouts were compared with predictions of Pareto rights, Law of Proportionate Effect, generalized Lotka-Volterra model and models of collisions. It turned out that middle class and wealthy Polish households are described very well by the cumulative distributions of Pareto, which differ by the value of the Pareto exponent. On the other hand the income of poor households is described by a cumulative log-normal distribution. To describe the poor and middle class households the generalized Lotka-Volterra model can be used, which provides the theoretical interpretation of the level of individual households.