

Katedra pravděpodobnosti a matematické statistiky



MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

Stanislav Svoboda

Pravděpodobnostní modely příjmů

7. března 2024

- V této části čerpáme ze článku [Łukasiewicz a Orłowski(2004)].
- Máme data příjmů domácností s 1 a se 2 pracujícími lidmi (individuální a dvojí příjem).
- Cílem práce je zjistit, zda existuje nějaký vztah mezi rozdělením individuálních příjmů a rozdělením dvojích příjmů.
- Data jsou z roku 2000. Ze 2 států USA a Polsko.

Rozdělení individuálních příjmů

- Pro určení nejlepšího možného modelu pro rozdělení individuálních příjmů v USA uvažujeme následující 4 rozdělení s hustotami:

- Dagum model:

$$f_D(x) = \frac{abcx^{b+1}}{(1 + ax^{-b})^{c+1}}, a > 1, b > 0, c > 0.$$

- Singh–Maddala model:

$$f_{SM}(x) = \frac{abcx^{b-1}}{(1 + ax^b)^{c+1}}, a > 0, b > 0, c > 0, bc > 1.$$

- Exponenciální model:

$$f_E(x) = \frac{1}{a} e^{-\frac{x}{a}}, a > 0.$$

- Weibull model:

$$f_W(x) = \frac{b}{a} e^{-\frac{x^b}{a}}, a > 0, b > 0.$$

- Pro Polsko uvažujeme rozdělení Dagum a Singh-Maddala. Dále se uvažovala rozdělení log-normal a gamma, ale ta se pro malou shodu s daty zavrhla.
- Parametry se odhadly pomocí metody maximální věrohodnosti.

- Pro určení shody mezi teoretickým a empirickým rozdělením použijeme součet čtvercových chyb (*SSE*) a součet absolutních chyb (*SAE*).
- Dále pro určení shody použijeme kvadrát korelačního koeficientu ρ teoretického a empirického kvantilu. Protože za předpokladu, že jsme zvolili skutečné rozdělení, by se tyto kvantily měly hodně shodovat.

Table 1
Individual incomes in the USA in the year 2000

Model	\hat{a}	\hat{b}	\hat{c}	<i>SSE</i>	<i>SAE</i>	W_r	W_p	ρ^2
Exponential	28.591	—	—	0.0008	0.1268	93.66	93.63	0.9843
Weibull	26.535	0.9806	—	0.0010	0.1392	93.04	92.97	0.9856
Dagum	29474.0	2.7434	0.3138	0.0003	0.0709	96.45	96.17	0.9987
Singh–Maddala	0.0053	1.1228	5.1245	0.0005	0.0899	95.50	95.67	0.9956

Results of estimation of parameters and adjustment of models to empirical data.

Table 2
Individual incomes in Poland in the year 2000

Model	\hat{a}	\hat{b}	\hat{c}	<i>SSE</i>	<i>SAE</i>	W_r	W_p	ρ^2
Dagum	12.7312	2.8498	1.4202	0.0008	0.0992	95.04	95.04	0.9987
Singh–Maddala	0.0369	1.6582	0.6955	0.0010	0.1039	94.81	94.80	0.9983

Results of estimation of parameters and adjustment of models to empirical data.

- Všechny testované hodnoty (degree of adjustment) vyšly nejlépe u rozdělení Dagum.
- Proto přijímáme rozdělení Dagum za skutečné u USA i Polska.
- Z následujících obrázků vidíme, že teoretické hustoty s odhadnutými parametry dobře odpovídají empirickým.

Rozdělení individuálních příjmů

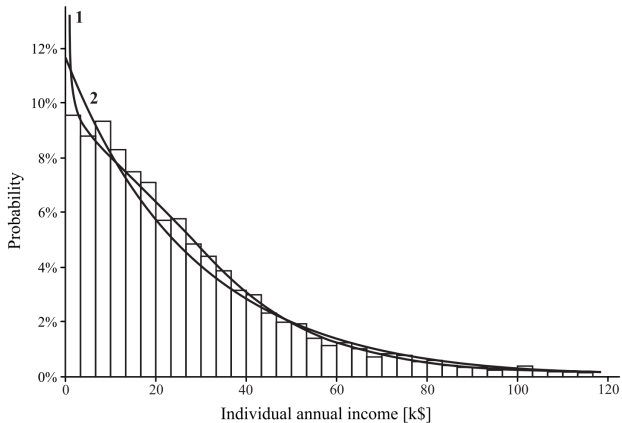


Fig. 1. Probability distribution of individual income (USA in 2000). Dagum's model (1) and exponential model (2).

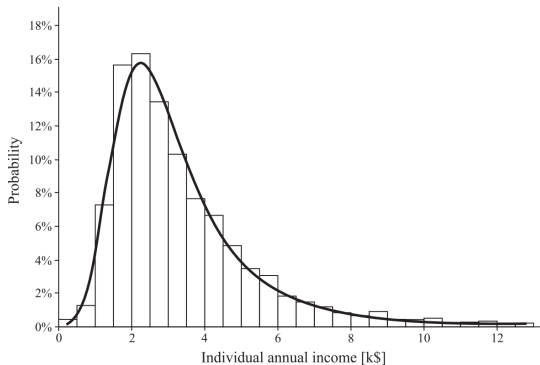


Fig. 2. Probability distribution of individual income (Poland in 2000) and fit of Dagum's model.

- Příjem domácnosti s 2 pracujícími lidmi je součet 2 individuálních příjmů.
- Tedy dvojí příjem Z se rovná součtu individuálních příjmů X a Y .
- Pro X a Y nezávislé můžeme pomocí konvoluce určit hustotu $f_Z(z) = (f_X * f_Y)(z) = \int_{\mathbb{R}} f_X(t)f_Y(z - t)dt$.
- V následujícím obrázku vidíme, že v USA se konvoluce individuálních rozdělení dobře shoduje s empirickým rozdělením a naznačuje to, že by příjmy členů domácnosti mohly být statisticky nezávislé.
- Poslední obrázek říká, že v Polsku to tak není a vypadá to, že příjmy členů domácnosti jsou silně statisticky závislé.

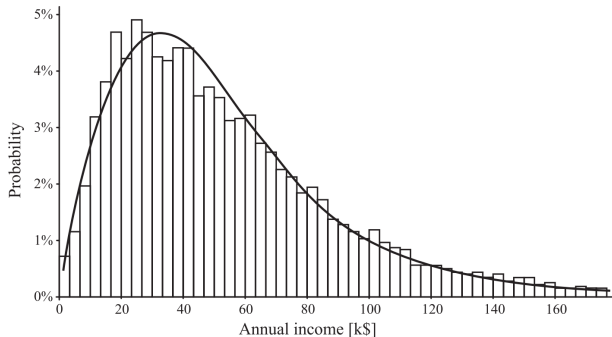


Fig. 3. Probability distribution of income for households with two earners (USA in 2000). Convolution of individual Dagum's distributions.

Rozdělení dvojích příjmů

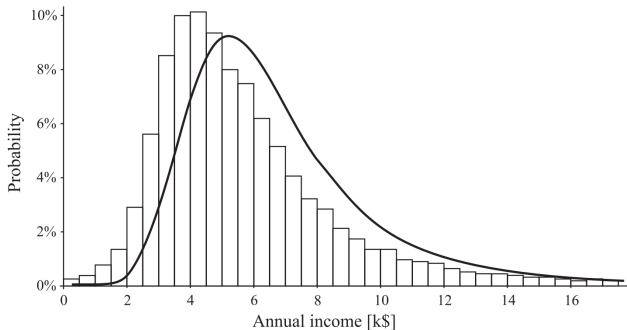


Fig. 4. Probability distribution of income for households with two earners (Poland in 2000). Convolution of individual Dagum's distributions. 1 USD = 4.1432 PLN (December 29, 2000).

- V této části čerpáme ze článku [Brzeziński(2013)].
- Máme data příjmů domácností ze 4 států: Česko, Polsko, Slovensko a Maďarsko. Data jsou z let 1990-2010.
- Cílem práce je zvolit správné rozdělení příjmů, porovnat změnu parametrů v průběhu let a interpretovat tuto změnu.

- Zobecněné beta rozdělení

$$f(x, a, b, p, q) = \frac{ax^{ap-1}}{b^{ap}B(p, q)[1 + (\frac{x}{b})^a]^{p+q}}, x > 0$$

- Vliv změny parametrů:

- Menší a znamená těžší chvosty.
- Menší p znamená těžší levý chvost.
- Menší q znamená těžší pravý chvost.

- Zároveň menší ap a aq znamená těžší oba chvosty.

Dá se interpretovat jako evidence větší příjmové bi-polarizace (větší vzdálenost mezi bohatými a chudými, větší nahromadění u nich).

- b je spojené se střední hodnotou.

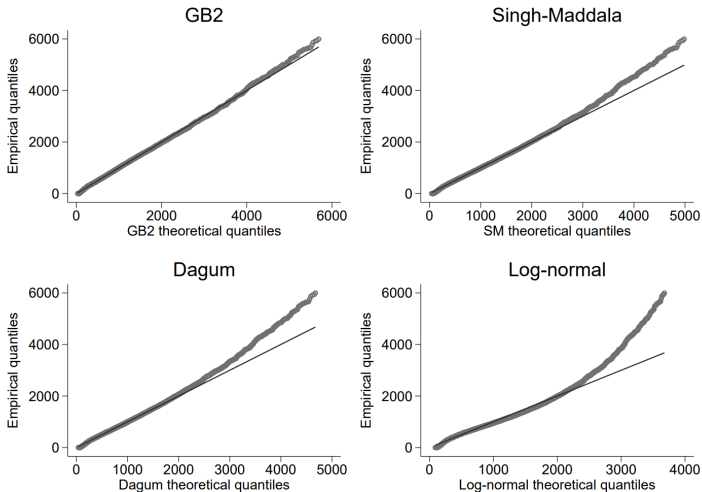
Table 4: Maximum likelihood estimates of models' parameters for Czech Republic

Parameter estimates	Singh-Maddala	Dagum	GB2
1992			
<i>a</i>	5.373 (0.064)	4.811 (0.055)	5.823 (0.274)
<i>b</i>	90938.49 (709.664)	91353.39 (877.08)	91574.01 (757.937)
<i>p</i>	-	1.157 (0.034)	0.885 (0.060)
<i>q</i>	0.845 (0.022)	-	0.762 (0.048)
Log-likelihood	-192443.57	-192450.99	-192441.99
1996			
<i>a</i>	4.146 (0.040)	3.782 (0.033)	3.776 (0.133)
<i>b</i>	129775.2 (1080.466)	128804.7 (1198.58)	128810 (1206.311)
<i>p</i>	-	1.151 (0.026)	1.153 (0.061)
<i>q</i>	0.882 (0.019)	-	1.002 (0.052)
Log-likelihood	-350202.42	-350198.63	-350198.63
2004			
<i>a</i>	3.902 (0.093)	3.711 (0.083)	3.864 (0.372)
<i>b</i>	152841 (3406.206)	153019.8 (3587.802)	152764.1 (3513.3)
<i>p</i>	-	1.072 (0.060)	1.014 (0.140)
<i>q</i>	0.929 (0.051)	-	0.941 (0.131)
Log-likelihood	-54971.207	-54971.296	-54971.202

Table 5: Likelihood ratio test for Czech Republic

	Singh-Maddala vs. GB2		Dagum vs. GB2	
	LR	p-value	LR	p-value
1992	3.16	0.075	18.0	0.000
1996	7.58	0.006	0.000	1
2004	0.01	0.920	0.188	0.665

Figure 1: Quantile-quantile plots, Poland, 1993



Data set	Inequality index			
	Gini	GE(2)	P90/P10	P75/P25
Czech Republic				
1992	0.206	0.112	2.360	1.548
1996	0.256	0.163	2.974	1.765
2004	0.267	0.184	3.212	1.801
Hungary				
1991	0.283	0.186	3.355	1.873
1994	0.321	0.273	4.138	1.970
1999	0.292	0.195	3.432	1.888
2005	0.291	0.248	3.311	1.845
Poland				
1993	0.284	0.239	3.312	1.808
1998	0.286	0.220	3.469	1.856
2004	0.313	0.259	4.000	1.981
2010	0.319	0.670	3.847	1.955
Slovak Republic				
1992	0.189	0.081	2.251	1.519
1996	0.250	0.131	3.038	1.716
2004	0.268	0.179	3.286	1.810
2010	0.265	0.213	3.253	1.814

Table 15: Test results for equality of the Lorenz curves

Combinations of estimated parameters and test statistics							
	a	p	q	ap	aq	χ^2	p -value
Czech Republic							
1992	5.823	0.885	0.762	5.153	4.437	63.49	0.000
2004	3.864	1.014	0.941	3.918	3.636		
Hungary							
1991	5.096	0.525	0.676	2.6754	3.445	-	-
2005	5.065	0.609	0.603	3.085	3.045		
Poland							
1993	5.463	0.575	0.564	3.141	3.081	114.63	0.000
2010	4.014	0.752	0.726	3.019	2.914		
Slovak Republic							
1992	5.734	0.901	0.906	5.166	5.195	278.24	0.000
2010	4.811	0.554	0.868	2.665	4.176		

- U všech států kromě Maďarska je pokles u *a*, *ap*, *aq*. To znamená, že rozdělení příjmu v 1990 je lorenzovsky dominantní oproti rozdělení příjmu v 2010. Neboli je větší nerovnost příjmu.
- *p*-hodnoty z testu potvrzují, že tato zjištění jsou signifikantní.
- Pokles v *aq*, *ap* také znamená, že státy zažily vzestup příjmové bi-polarizace.



BRZEZIŃSKI, M. (2013).

Parametric modelling of income distribution in central and eastern europe.

Central European Journal of Economic Modelling and Econometrics, (3).



ŁUKASIEWICZ, P. a ORŁOWSKI, A. (2004).

Probabilistic models of income distributions.

Physica A: Statistical Mechanics and its Applications, **344**(1), 146–151.