

Mateusz Goc

MODELOWANIE ROZKŁADÓW STOPY BEZROBOCIA ZA POMOCĄ KRZYWYCH PEARSONA

MODELLING THE DISTRIBUTION OF THE UNEMPLOYMENT RATES USING PEARSON FAMILY DISTRIBUTIONS

Katedra Zastosowań Matematyki w Ekonomii, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
ul. Klemensa Janickiego 31, 71-270 Szczecin

Abstract. The papers present a trial of use of the Pearson family distributions in modeling the monthly distributions of the unemployment rates by the regions in 2004-2007 period. Results of fitting the Pearson distributions were compared with approximants given by Generalized Extreme Value distribution and Gamma distribution. In many cases Pearson distributions were better fitted than "classical" distributions.

Słowa kluczowe: bezrobocie, modelowanie rozkładów, system krzywych Pearsona.
Key words: distribution modeling, Pearson family distributions.

WSTĘP

W modelowaniu rozkładów empirycznych zmiennych ekonomicznych dominuje podejście polegające na ich aproksymowaniu za pomocą określonej klasy rozkładów teoretycznych zmiennej losowej ciągłej charakteryzujących się np. prawostronną asymetrią. Alternatywne podejście polega na wykorzystaniu do tego celu systemu krzywych Pearsona obejmujących XII typów rozkładów.

Dorobek publikacyjny poświęcony zastosowaniu tego systemu krzywych w modelowaniu i prognozowaniu rozkładów empirycznych jest stosunkowo ubogi. Najobszerniejszą w tym zakresie pracę, poświęconą modelowaniu rozkładów płac, opublikowali Kordos i Stroińska (1973 b).

Celem pracy jest sprawdzenie przydatności krzywych Pearsona w modelowaniu rozkładów na przykładzie miesięcznych rozkładów empirycznych stopy bezrobocia według miast i powiatów w latach 2004–2007.

MATERIAŁ I METODY

Pod koniec XIX wieku Karl Pearson opracował system rozkładów zmiennej losowej, który został wyprowadzony z równania różniczkowego o postaci (Rose i Smith 2002):

$$\frac{dp(x)}{dx} = -\frac{a+x}{c_0 + c_1x + c_2x^2} p(x) \quad (1)$$

Kształt krzywej zależy od czterech parametrów a , c_0 , c_1 , c_2 , które mogą być wyrażone w zależności od pierwszych czterech momentów rozkładu. Pearson uzyskał w ten sposób 12 typów rozkładów, jednak najszersze praktyczne zastosowanie mają typy I–VII.

Przyporządkowanie krzywej rozkładu do odpowiedniego typu zależy od wartości współczynnika κ , który jest wyrażony wzorem (Stuart i Ord 1994):

$$\kappa = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3)(2\beta_2 - 3\beta_1 - 6)} \quad (2)$$

gdzie:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} - \text{kwadrat współczynnika skośności,}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} - \text{współczynnik kurtozy,}$$

μ_i – momenty centralne i -tego rzędu.

W zależności od wartości parametru κ Pearsona wyznaczył trzy typy podstawowe oraz cztery typy przechodnie. Do typów podstawowych zaliczamy (Kordos 1973 a):

- typ I, gdy $\kappa < 0$,
- typ IV, gdy $0 < \kappa < 1$,
- typ VI, gdy $\kappa > 1$.

Typy przechodnie to:

- typ II, gdy $\kappa = 0$,
- typ III, gdy $\kappa = \pm\infty$ (w praktyce stosuje się, gdy $|\kappa| > 4$),
- typ V, gdy $\kappa = 1$,
- typ VII, gdy $\kappa = 0$.

Jeśli oba pierwiastki trójmianu są rzeczywiste i mają różne znaki, to otrzymuje się typ I rozkładu Pearsona. Oznaczając pierwiastki jako a_1 i a_2 ($a_1 < 0 < a_2$), otrzymujemy funkcję gęstości o postaci:

$$f(x) = K(x - a_1)^{m_1} (a_2 - x)^{m_2}, \quad a_1 < x < a_2 \quad (3)$$

gdzie:

$$m_1 = \frac{a + a_1}{c_2(a_2 - a_1)} \text{ oraz } m_2 = -\frac{a + a_2}{c_2(a_2 - a_1)}.$$

Jeżeli oba wykładniki potęg są ujemne, to rozkład jest U-kształtny, natomiast jeśli jedna z potęg jest ujemna, a druga dodatnia – rozkład jest J-kształtny. Typ ten jest ogólną postacią rozkładu beta.

Typ II jest szczególnym przypadkiem rozkładu typu I, dla którego $m_1 = m_2$, a pomiędzy pierwiastkami zachodzi zależność $a_1 = -a_2$. Rozkład tego typu jest symetryczny, a jego funkcja gęstości ma postać:

$$f(x) = K(x^2 - a_1^2)^{-\frac{1}{2c_2}}, \quad -a_1 < x < a_1 \quad (4)$$

Szczególnym przypadkiem tego typu dla $\frac{1}{2c_2} = 0$ jest rozkład prostokątny.

W sytuacji, gdy $c_2 = 0$ (oraz $c_1 \neq 0$), otrzymuje się typ III rozkładu systemu Pearsona, którego funkcję gęstości można zapisać w postaci:

$$f(x) = K(c_0 + c_1x)^m \exp\left(\frac{-x}{c_1}\right), \quad x > -\frac{c_0}{c_1} \text{ dla } c_1 > 0 \text{ oraz } x < -\frac{c_0}{c_1} \text{ dla } c_1 < 0 \quad (5)$$

gdzie:

$$m = \frac{1}{c_1} \left(\frac{c_0}{c_1} - a \right).$$

Szczególnymi przypadkami rozkładów tego typu są rozkłady gamma i χ^2 .

Typ IV rozkładu systemu Pearsona otrzymujemy, gdy trójmian kwadratowy $c_0 + c_1x + c_2x^2$ nie ma rzeczywistych pierwiastków. Jego funkcja gęstości wyraża się wzorem:

$$f(x) = K \exp \left[\frac{(c_1 - 2ac_2) \arctg\left(\frac{c_1 + 2c_2x}{\sqrt{4c_0c_2 - c_1^2}}\right)}{c_2 \sqrt{4c_0c_2 - c_1^2}} \right] (c_0 + x(c_1 + c_2x))^{-\frac{1}{2c_2}}, \quad -\infty < x < \infty \quad (6)$$

Gdy trójmian $c_0 + c_1x + c_2x^2$ ma tylko jeden pierwiastek, otrzymujemy typ V o funkcji gęstości:

$$f(x) = Ke^{\frac{2ac_2 - c_1}{c_2(c_1 + 2c_2x)}} (c_1 + 2c_2x)^{-\frac{1}{c_2}}, \quad x \geq 0 \quad (7)$$

Szczególnymi przypadkami rozkładów tego typu jest rozkład normalny odwrotny.

Gdy pierwiastki trójmianu są rzeczywiste i mają ten sam znak, otrzymuje się typ VI rozkładu. Jego funkcja gęstości ma postać:

$$f(x) = K(x - a_1)^{m_1} (x - a_2)^{m_2}, \quad x > a_2 \text{ dla } a_1 < a_2 < 0 \text{ oraz } x < a_1 \text{ dla } a_2 > a_1 > 0 \quad (8)$$

Szczególnym przypadkiem rozkładu typu VI jest rozkład typu VII. Jest to rozkład symetryczny; otrzymuje się go, gdy $c_1 = a = 0$ dla $c_0 > 0$; jego funkcja gęstości wyrażona jest wzorem:

$$f(x) = K(c_0 + c_2x^2)^{-\frac{1}{2c_2}}, \quad -\infty < x < \infty \quad (9)$$

Szczególnym przypadkiem tej krzywej jest rozkład t-Studenta.

Aproksymanty uzyskane na podstawie krzywych Pearsona zostaną porównane z aproksymantami uzyskanymi na podstawie uogólnionego rozkładu wartości ekstremalnych (GEV) oraz rozkładu gamma (G). Przydatność powyższych funkcji w modelowaniu rozkładów stopy bezrobocia została udowodniona m.in. w pracy Goca (2009). Funkcja gęstości uogólnionego rozkładu wartości ekstremalnych wyrażona jest wzorem (10) – Kotz i Nadarajah (2001):

$$f(x) = \begin{cases} \frac{1}{\sigma} \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\xi} - 1} \exp \left(- \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\xi}} \right) & -\infty < x \leq \mu - \frac{\sigma}{\xi} \quad \text{dla } \xi < 0 \\ \frac{1}{\sigma} \exp \left(- \frac{x - \mu}{\sigma} \right) \exp \left(- \exp \left(- \frac{x - \mu}{\sigma} \right) \right) & \mu - \frac{\sigma}{\xi} \leq x < \infty \quad \text{dla } \xi > 0 \\ \frac{1}{\sigma} \exp \left(- \frac{x - \mu}{\sigma} \right) \exp \left(- \exp \left(- \frac{x - \mu}{\sigma} \right) \right) & -\infty < x < \infty \quad \text{dla } \xi = 0 \end{cases} \quad (10)$$

Natomiast funkcję rozkładu gamma opisuje wzór (11) – Johnson i in. (1994):

$$f(x) = \frac{x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{x}{\beta}\right), \quad x \geq 0 \quad (11)$$

Do testowania zgodności rozkładów empirycznych wykorzystano test λ -Kolmogorowa oraz test zgodności χ^2 .

WYNIKI I DYSKUSJA

W pracy dokonano aproksymacji miesięcznych rozkładów stopy bezrobocia według miast i powiatów w latach 2004–2007.

W pierwszym etapie badań, w celu przyporządkowania rozkładów empirycznych do konkretnego typu krzywej, dla każdego miesiąca zostały obliczone współczynniki κ oraz β_1 i β_2 . Wyniki obliczeń zawiera tab. 1.

Po przeprowadzeniu obliczeń okazało się, że we wszystkich badanych miesiącach rozkłady stopy bezrobocia według miast i powiatów powinny być aproksymowane za pomocą krzywej I typu.

W kolejnym etapie dokonano modelowania rozkładów empirycznych za pomocą I typu pearsonowskiego (PEARSON I). Uzyskane wyniki porównano z aproksymantami uzyskanymi na podstawie rozkładów uogólnionej wartości ekstremalnych (GEV) oraz gamma (G). Dla poziomu istotności $\alpha = 0,05$ wartości krytyczne testów: $\lambda = 1,36$ oraz $\chi_{\text{PEARSON I}}^2 = 9,49$, $\chi_{\text{GEV}}^2 = 11,07$ i $\chi_G^2 = 12,59$. Wyniki zgodności rozkładów empirycznych teoretycznych zestawiono w tab. 2.

Z przeprowadzonych obliczeń wynika, że w przypadku rozkładu PEARSON I dla wszystkich miesięcy uzyskano najmniejsze wartości statystyk empirycznych λ_{emp} . W przypadku statystyki χ_{emp}^2 miało to miejsce w ponad 35% przypadków. Jedynie w dwóch miesiącach (w lutym 2006 r. oraz w styczniu 2007 r.) wartości statystyk χ_{emp}^2 nie potwierdziły zgodności rozkładów empirycznych i teoretycznych.

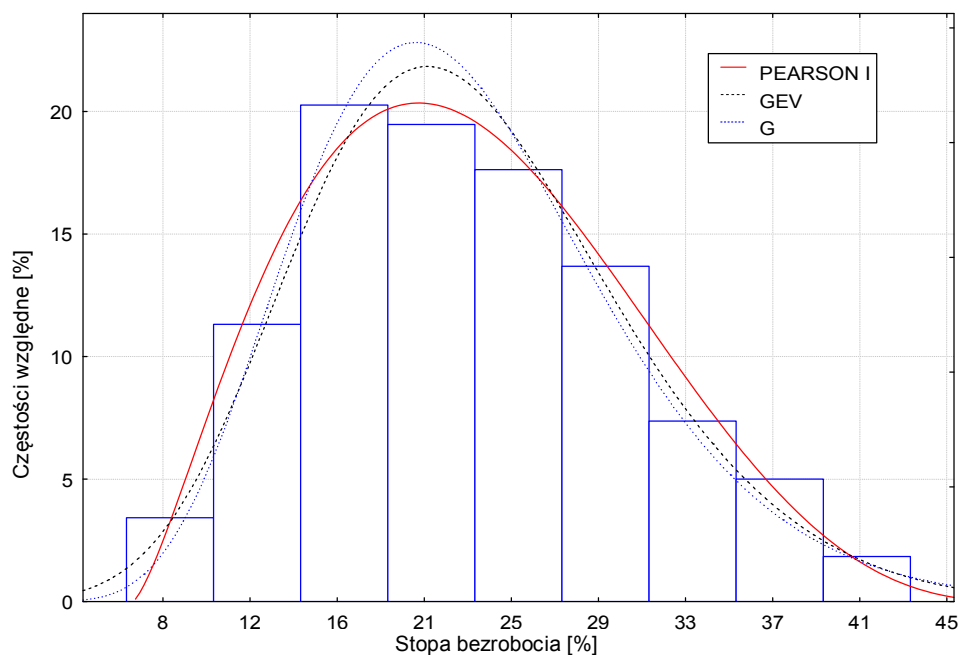
Na rysunkach 1 i 2 przedstawiono graficznie wyniki modelowania rozkładów z czerwca 2004 r. oraz z grudnia 2007 r. W tabeli 3 przedstawiono wartości parametrów aproksymowanych funkcji.

Tabela 1. Wyniki przyporządkowania rozkładów stopy bezrobocia do krzywych Pearsona

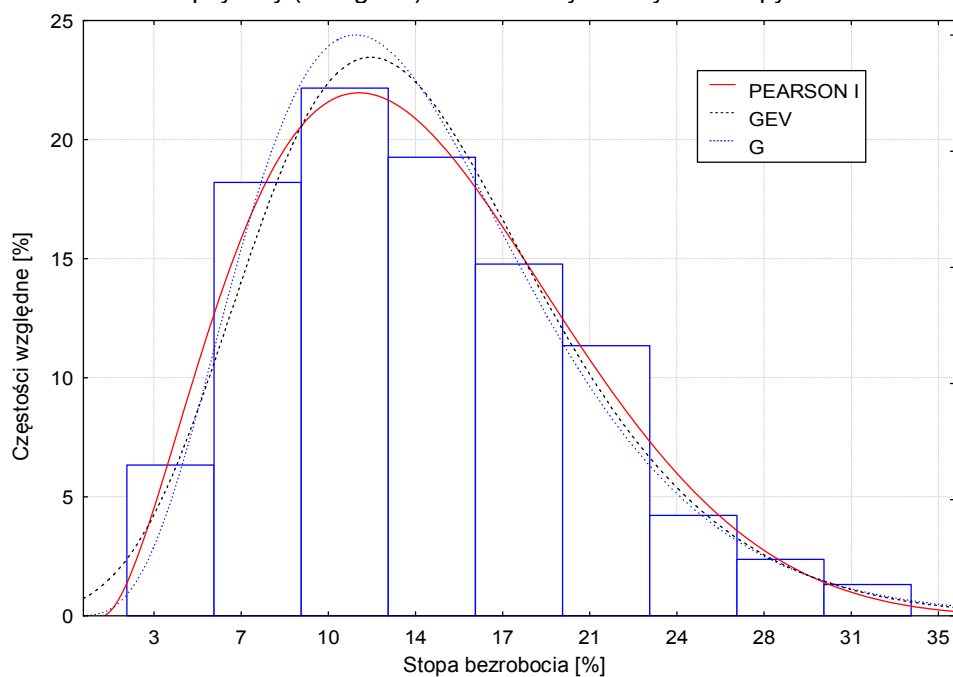
Rok	Miesiąc	κ	β_1	β_2	Typ krzywej Pearsona
2004	styczeń	-0,0626	0,1053	2,5012	I
	luty	-0,0610	0,1037	2,4925	I
	marzec	-0,0638	0,1061	2,5096	I
	kwiecień	-0,0647	0,1078	2,5107	I
	maj	-0,0738	0,1208	2,5407	I
	czerwiec	-0,0670	0,1122	2,5138	I
	lipiec	-0,0753	0,1242	2,5399	I
	sierpień	-0,0780	0,1291	2,5437	I
	wrzesień	-0,0894	0,1429	2,5855	I
	październik	-0,0947	0,1532	2,5920	I
	listopad	-0,0974	0,1574	2,5980	I
	grudzień	-0,0994	0,1589	2,6074	I
2005	styczeń	-0,0954	0,1545	2,5929	I
	luty	-0,0904	0,1482	2,5762	I
	marzec	-0,0881	0,1469	2,5633	I
	kwiecień	-0,0857	0,1448	2,5515	I
	maj	-0,0892	0,1523	2,5540	I
	czerwiec	-0,0908	0,1579	2,5484	I
	lipiec	-0,1037	0,1737	2,5961	I
	sierpień	-0,1111	0,1815	2,6230	I
	wrzesień	-0,1199	0,1936	2,6475	I
	październik	-0,1276	0,2028	2,6701	I
	listopad	-0,1347	0,2080	2,6957	I
	grudzień	-0,1269	0,2018	2,6682	I
2006	styczeń	-0,1022	0,1675	2,6023	I
	luty	-0,0946	0,1587	2,5749	I
	marzec	-0,0953	0,1571	2,5845	I
	kwiecień	-0,0905	0,1487	2,5753	I
	maj	-0,0843	0,1420	2,5496	I
	czerwiec	-0,0768	0,1309	2,5261	I
	lipiec	-0,0850	0,1404	2,5605	I
	sierpień	-0,0971	0,1587	2,5920	I
	wrzesień	-0,1054	0,1693	2,6178	I
	październik	-0,1210	0,1923	2,6558	I
	listopad	-0,1186	0,1916	2,6443	I
	grudzień	-0,1205	0,1991	2,6391	I
2007	styczeń	-0,1269	0,1970	2,6775	I
	luty	-0,1248	0,1935	2,6734	I
	marzec	-0,1231	0,1925	2,6665	I
	kwiecień	-0,1199	0,1854	2,6641	I
	maj	-0,1268	0,1974	2,6762	I
	czerwiec	-0,1557	0,2169	2,7696	I
	lipiec	-0,1443	0,2079	2,7377	I
	sierpień	-0,1493	0,2122	2,7514	I
	wrzesień	-0,1732	0,2434	2,8007	I
	październik	-0,2197	0,2820	2,9037	I
	listopad	-0,2170	0,2786	2,8990	I
	grudzień	-0,2346	0,2904	2,9341	I

Tabela 2. Zgodność rozkładów empirycznych stopy bezrobocia z rozkładami PEARSON I, GEV i G

Rok	Miesiąc	Rozkład					
		PEARSON I		GEV		G	
		λ_{emp}	χ^2_{emp}	λ_{emp}	χ^2_{emp}	λ_{emp}	χ^2_{emp}
2004	styczeń	0,2793	7,54	0,5145	4,23	0,6966	4,69
	luty	0,2936	7,56	0,5120	5,96	0,6785	6,99
	marzec	0,3106	7,55	0,5005	6,04	0,6576	5,38
	kwiecień	0,3461	6,07	0,4666	5,20	0,6290	6,78
	maj	0,5509	4,00	0,4396	4,46	0,5885	3,11
	czerwiec	0,2158	2,79	0,5521	5,84	0,7367	4,86
	lipiec	0,2032	3,03	0,5079	5,33	0,6152	6,09
	sierpień	0,1624	1,95	0,5422	6,09	0,5712	4,68
	wrzesień	0,2293	2,50	0,4513	1,20	0,6078	3,37
	październik	0,2603	4,75	0,5007	2,02	0,6374	3,18
	listopad	0,2630	4,97	0,5046	4,76	0,6576	5,15
	grudzień	0,3236	7,81	0,5628	4,23	0,6917	7,65
2005	styczeń	0,2723	7,25	0,4840	5,42	0,6130	5,17
	luty	0,3108	6,33	0,4688	5,00	0,6010	3,45
	marzec	0,2961	7,78	0,5141	4,80	0,6605	4,52
	kwiecień	0,4449	7,75	0,5420	6,32	0,6740	4,66
	maj	0,2636	7,41	0,6183	4,00	0,7626	3,99
	czerwiec	0,2559	4,53	0,5618	5,31	0,6991	3,77
	lipiec	0,2412	6,06	0,5696	6,67	0,6827	4,70
	sierpień	0,2038	4,32	0,5669	4,49	0,6742	6,98
	wrzesień	0,2511	4,76	0,6995	5,00	0,8048	4,27
	październik	0,2432	5,25	0,6350	3,40	0,7334	2,76
	listopad	0,2478	2,65	0,6528	6,08	0,7493	6,21
	grudzień	0,2866	5,23	0,7279	6,92	0,8200	6,94
2006	styczeń	0,3517	7,77	0,6403	2,77	0,7569	5,49
	luty	0,6001	13,12	0,6837	2,72	0,7914	3,59
	marzec	0,2995	8,15	0,6479	3,25	0,7777	4,09
	kwiecień	0,3075	4,90	0,5544	3,71	0,6956	6,05
	maj	0,2900	4,29	0,4976	3,69	0,6551	4,04
	czerwiec	0,4382	3,73	0,5626	2,51	0,7404	4,35
	lipiec	0,2933	1,65	0,5716	7,56	0,7396	7,98
	sierpień	0,2349	2,60	0,5344	9,65	0,6913	6,90
	wrzesień	0,3275	3,48	0,5655	4,58	0,7373	7,68
	październik	0,2965	3,56	0,6674	6,37	0,8132	5,91
	listopad	0,3134	5,91	0,8239	9,32	0,9722	6,33
	grudzień	0,3451	6,07	0,6181	5,61	0,7626	9,20
2007	styczeń	0,5923	10,40	0,6144	6,96	0,7700	5,54
	luty	0,3639	5,10	0,6130	6,70	0,7661	6,81
	marzec	0,3672	8,06	0,6452	6,36	0,7986	7,25
	kwiecień	0,2956	3,89	0,6385	4,54	0,7885	8,38
	maj	0,2703	4,08	0,5293	2,90	0,7172	7,18
	czerwiec	0,2566	2,30	0,5897	4,32	0,7524	6,23
	lipiec	0,2648	2,10	0,5476	4,37	0,7119	8,00
	sierpień	0,3892	5,20	0,5430	2,37	0,7587	7,14
	wrzesień	0,3185	4,11	0,5689	3,47	0,7398	4,27
	październik	0,2678	3,43	0,5215	3,49	0,6654	5,52
	listopad	0,5160	5,12	0,6056	6,15	0,7491	4,23
	grudzień	0,1992	3,49	0,5369	4,00	0,6619	4,60



Rys. 1. Rozkład empiryczny (histogram) oraz rozkłady teoretyczne stopy bezrobocia w czerwcu 2004 r.



Rys. 2. Rozkład empiryczny (histogram) oraz rozkłady teoretyczne stopy bezrobocia w grudniu 2007 r.

Tabela 3. Zgodność rozkładów empirycznych stopy bezrobocia z rozkładami PEARSON I, GEV i G

Miesiąc	Rozkład	Parametry
Czerwiec 2004	PEARSON I	$K = 2,00 \times 10^{-7} \quad a_1 = 49,22 \quad a_2 = 6,78 \quad m_1 = 3,02 \quad m_2 = 1,41$
	GEV	$\xi = -0,1522 \quad \sigma = 7,01 \quad \mu = 19,56$
	G	$\alpha = 9,03 \quad \beta = 2,51$
Grudzień 2007	PEARSON I	$K = 1,54 \times 10^{-12} \quad a_1 = 46,8 \quad a_2 = 1,44 \quad m_1 = 6,06 \quad m_2 = 1,77$
	GEV	$\xi = -0,1021 \quad \sigma = 5,50 \quad \mu = 11,57$
	G	$\alpha = 5,26 \quad \beta = 2,71$

WNIOSKI

Z przeprowadzonych w pracy rozważań wynika, że oceny statystyk empirycznych λ -Kolmogorowa dla rozkładu typu I Pearsona we wszystkich przypadkach są niższe od oceny statystyk dla rozkładów wartości ekstremalnych (GEV) oraz gamma (G). Oznacza to, że rozkład ten najlepiej aproksymuje badaną zmienną (a ściślej charakteryzuje się mniejszymi różnicami maksymalnych różnic funkcji gęstości). Otrzymanie dla wszystkich miesięcy rozkładów homogenicznych pozwoli na zbudowanie rozkładów prognoz za pomocą metody polegającej na budowie prognoz parametrów. Prognozy parametrów wyznacza się na podstawie predyktorów opartych na równaniach szeregu czasowego z wahaniami sezonowymi lub modelach Holta-Wintersa.

PIŚMIENNICTWO

- Goc M.** 2009. Zastosowanie metod ekonometrycznych w modelowaniu i prognozowaniu rozkładów wybranych zjawisk ekonomicznych. Praca doktorska. Szczecin, AR (maszynopis).
- Kordos J.** 1973 a. Metody analizy i prognozowania rozkładów płac i dochodów ludności. Warszawa, PWE.
- Kordos J., Stroińska Z.** 1973 b. Zastosowanie krzywych Pearsona do badania rozkładów płac i dochodów ludności. Warszawa, GUS.
- Kotz S., Nadarajah S.** 2001. Extreme value distributions. Theory and applications. London, Imperial College Press.
- Johnson N.L., Kotz S., Balakrishnan N.** 1994. Continuous Univariate Distributions, vol. 1. New York, Wiley.
- Rose C., Smith M.D.** 2002. Mathematical Statistics with Mathematica. New York, Springer.
- Stuart A., Ord K.** 1994. Kendall's advanced theory of statistics, vol. I. Distribution theory. New York, Hodder Arnold.