RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS No. 91

Econometrics 28 2010

Jan Zawadzki, Mateusz Goc

West Pomeranian Technological University, Szczecin

FORECASTING DISTRIBUTIONS OF VARIABLES WITH SEASONAL FLUCTUATIONS

Abstract: The following study presents the example of modelling and forecasting homogeneous distributions of the variable exhibiting seasonal fluctuations on the example of empirical distributions of the monthly unemployment rates by districts and cities with district rights in Poland. The study indicates that the process of creating the *ex ante* forecasts of distributions should be preceded by the *ex post* analysis of the accuracy of the forecasts of parameters and compliance of distributions.

Key words: homogeneous distributions, forecasting, seasonal data.

1. Theoretical introduction

Numerous examples of the description of empirical distributions of economic variables by means of theoretical distributions of random variables, usually continuous, may be found in a statistical and econometric literature. They concern various spheres of economy, starting from the pay distributions, through insurance, banking, demography and ending with the capital markets. However, far fewer studies have been devoted to forecasting the distributions. The studies by J. Kordos are the most extensive in this respect [Kordos 1970a, b, 1973].

Theoretical considerations concerning the formal foundations of forecasting the distributions have been conducted by B. Guzik [1991].

While starting the approximation of the empirical distribution of economic variable by means of the theoretical distribution of the random variable, it should be decided whether it is characterized by unimodality or multimodality as well as the direction and strength of asymmetry should be determined. Due to the fact that right-sided asymmetric distributions prevail in economics, empirical distributions will most frequently be approximated by means of the functions that allow the appearance of the asymmetry of this kind. If we model the distributions in time, it is crucial to investigate whether these will be the same or different distributions in the respective periods of time.

In case the empirical distribution of variable is characterized by its multimodality, the mixes of homogeneous or heterogeneous distributions are used for modelling, depending on whether each of separate intervals will be described by means of the

same type of distribution or different types. Mixing weights are usually established at the level of the share of the size of the given interval in the total number of cases.

One of the most difficult issues that must be handled while initiating the modelling, irrespective of the fact if the modelling concerns a unimodal distribution or the decomposition of a multimodal distribution, is establishing the analytical form of approximating distribution.

The introduction of the notion of an admissible distribution may be helpful in this respect. It is proposed to recognize as the admissible distribution such a theoretical distribution of a random variable, with relation to which there is no basis in at least one of the tests to reject the hypothesis on the compliance of empirical and theoretical distributions. This means that the empirical distribution may be described by means of many admissible distributions. The question arises: which of these distributions may be used in *ex ante* forecasting? At least a partial answer to this question may be obtained by conducting the *ex post* analysis of the compliance of empirical, theoretical and forecast distributions.

We will address the issue of modelling and forecasting of unimodal homogeneous distributions further in this study.

If the function domain is determined, then each of homogeneous admissible distributions may have its forecasts established on the basis of parameter forecasts. There are various propositions concerning the distribution forecasting methods in the literature, however, most of them refer to the situation in which time series of parameters are short. In this respect, the average mobile increment method can be mentioned.

If we have a greater number of observations at our disposal, we can use trend models or exponential fitting models without seasonal fluctuations for the purpose of creating parameter forecasts. However, if these are monthly or quarterly distributions of economic variables that are subject to modelling, then the predicators based on time series models with polynomial trends and constant seasonality or exponential trends with polynomials in an exponent and with relatively constant seasonality in the form [Zawadzki 1995]:

$$\hat{Y}_{iT} = \hat{f}(T) + \sum_{k=1}^{m} \hat{d}_{0ik} Q_{kT} \ , \ T = n+1, \ldots, n+\tau, \tag{1}$$

$$\hat{Y}_{iT} = e^{\hat{f}(T) + \sum_{k=1}^{m} \hat{\delta}_{0ik} Q_{kT}} \ , \ T = n+1, \ldots, n+\tau. \tag{2}$$

Holt-Winters models may be used for this purpose too.

2. The results of modelling and forecasting the monthly distributions of the unemployment rates

In the study, empirical distributions of the monthly unemployment rates, encompassing 323 districts and cities with district rights in the years 2005-2008, with the year

2008 constituting the period of the empirical verification of distributions, are subject to modelling and, subsequently, forecasting.

Trends in the average unemployment rate in Poland have been presented in Figure 1. The figure indicates that the rates are subject to seasonal fluctuations.

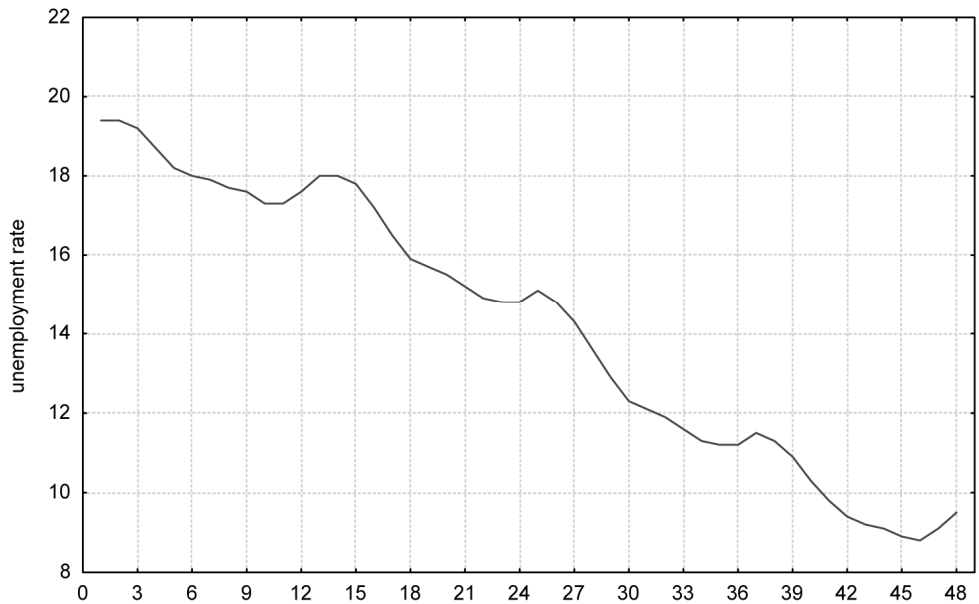


Figure 1. Actual values of the unemployment month rate in Poland rate in the years 2005-2008 by months

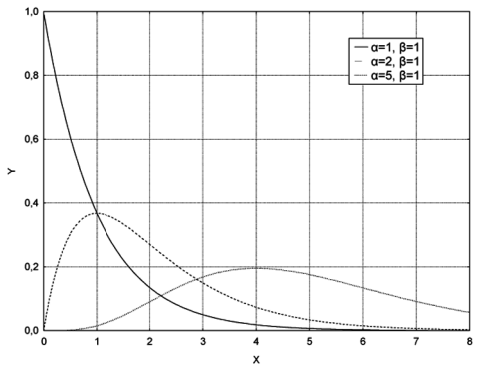
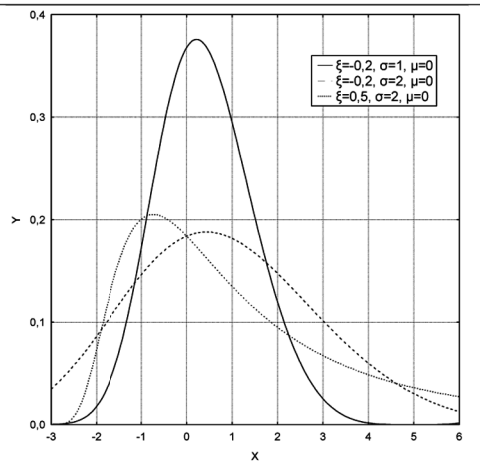
Source: own study.

Forecasting the distributions of the monthly unemployment rates will be preceded by modelling empirical distributions for each month in the years 2005-2007.

After an initial analysis, one two-parameter distribution (gamma-G) and one three-parameter distribution (of generalized extreme value – GEV) have been selected from over a dozen admissible distributions. The analytical forms of the density functions of these distributions ($f(x)$) and their cumulative distributions functions ($F(x)$) are specified in Table 1. This table also features a graphic presentation of the density function of both distributions for the selected values of parameters.

The λ-Kolmogorov and χ^2 compliance tests have been used for the purpose of testing empirical distributions [Domański 1979]. The estimate of empirical statistics (tests) for the following distributions: extreme values (GEV) and gamma (G) have been specified in Table 2. Critical values for both tests at the significance level $\alpha = 0.05$ equal respectively $\gamma = 1.36$ and $\chi^2_{Gew} = 11.07$ and $\chi^2_G = 12.59$.

Table 1. Analytical forms of gamma and extreme values distributions and graphs for examples values of parameters

Distribution	Density and cumulative distribution function	Charts for the examples of values of parameters
Gamma (G)	$f(x)=\frac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}\exp\left(\frac{-x}{\beta}\right)$ $F(x)=\frac{\Gamma_{x/\beta}}{\Gamma(\alpha)}$	
Generalized Extreme Values (GEV)	$f(x)=\begin{cases}\frac{1}{\sigma}\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}-1}\exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right) & -\infty<x\le\mu-\frac{\sigma}{\xi} \text{ for } \xi<0\\ & \mu-\frac{\sigma}{\xi}\le x<\infty \text{ for } \xi>0\\ \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right)\exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) & -\infty<x<\infty \text{ for } \xi=0\end{cases}$ $F(x)=\begin{cases}\exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right) & -\infty<x\le\mu-\frac{\sigma}{\xi} \text{ for } \xi<0\\ & \mu-\frac{\sigma}{\xi}\le x<\infty \text{ for } \xi>0\\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) & -\infty<x<\infty \text{ for } \xi=0\end{cases}$	

Source: [Johnson, Kotz, Balakrishman 1994; Kotz, Nadarajah 2000].

Table 2. Estimates of empirical statistics of GEV and G distributions by months in the years 2005-2007

Year	Month	Distribution			
		GEV		G	
		λ_{emp}	χ^2_{emp}	λ_{emp}	χ^2_{emp}
2005	January	0.4840	5.42	0.6130	5.17
	February	0.4688	5.00	0.6010	3.45
	March	0.5141	4.80	0.6605	4.52
	April	0.5420	6.32	0.6740	4.66
	May	0.6183	4.00	0.7626	3.99
	June	0.5618	5.31	0.6991	3.77
	July	0.5696	6.67	0.6827	4.70
	August	0.5669	4.49	0.6742	6.98
	September	0.6995	5.00	0.8048	4.27
	October	0.6350	3.40	0.7334	2.76
	November	0.6528	6.08	0.7493	6.21
	December	0.7279	6.92	0.8200	6.94
2006	January	0.6403	2.77	0.7569	5.49
	February	0.6837	2.72	0.7914	3.59
	March	0.6479	3.25	0.7777	4.09
	April	0.5544	3.71	0.6956	6.05
	May	0.4976	3.69	0.6551	4.04
	June	0.5626	2.51	0.7404	4.35
	July	0.5716	7.56	0.7396	7.98
	August	0.5344	9.65	0.6913	6.90
	September	0.5655	4.58	0.7373	7.68
	October	0.6674	6.37	0.8132	5.91
	November	0.8239	9.32	0.9722	6.33
	December	0.6181	5.61	0.7626	9.20
2007	January	0.6144	6.96	0.7700	5.54
	February	0.6130	6.70	0.7661	6.81
	March	0.6452	6.36	0.7986	7.25
	April	0.6385	4.54	0.7885	8.38
	May	0.5293	2.90	0.7172	7.18
	June	0.5897	4.32	0.7524	6.23
	July	0.5476	4.37	0.7119	8.00
	August	0.5430	2.37	0.7587	7.14
	September	0.5689	3.47	0.7398	4.27
	October	0.5215	3.49	0.6654	5.52
	November	0.6056	6.15	0.7491	4.23
	December	0.5369	4.00	0.6619	4.60

Source: own studies.

The information presented in the table indicates that the estimates of empirical statistics for both distributions are lower than the critical values in all investigated months. Thus, in the light of the aforesaid definition, they constitute admissible and at the same time homogeneous distributions.

Subsequently, on the basis of the monthly time series for three parameters of GEV distribution and two parameters of G distribution time series models have been estimated with the 1st order, 2nd order and 3rd order polynomial trends and constant seasonality – they have respectively been marked as: *l*, *kw* and *3st*. Exponential models with polynomials of the same order in exponent and relatively constant seasonality have also been estimated. They have been marked with the following symbols: *w*, *w2st.* and *w3st.* The estimates of the parameters of the stochastic structure: coefficients of determination (R^2) and coefficients of random volatility (V_S) for the evaluated equations have been specified in the third and fourth columns of the table. The trends in these estimates indicate that the evaluated equations describe the trends of parameters in time well or very well. They inform about high estimates of the coefficients of determination and low values of the coefficients of random volatility. This observation is confirmed by the statistical values (to be found in the next two columns) of significant parameters of trend and parameters accompanying 0-1 variables. For the majority of equations the number of the latter ones is not lower than 5.

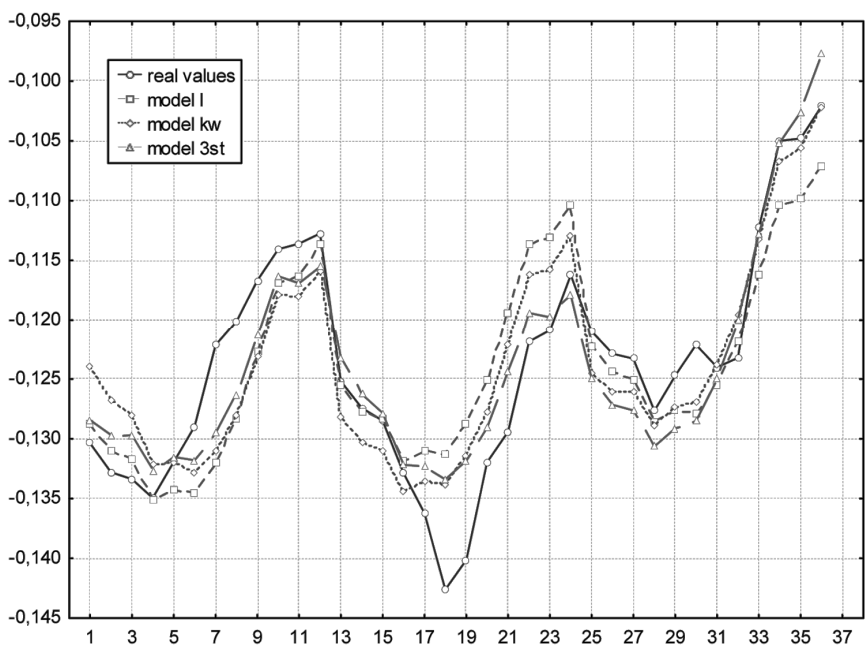


Figure 2. Actual and fitted values of ξ parameter of the GEV distribution

Source: own studies.

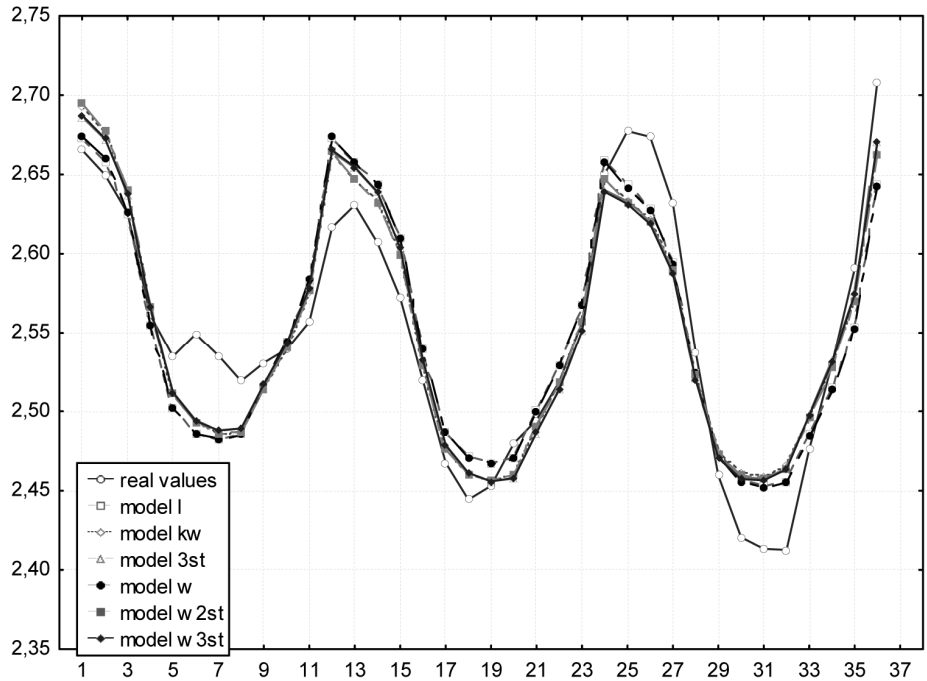


Figure 3. Actual and fitted values of β parameter of the G distribution

Source: own studies.

Examples of results of modelling one parameter of GEV(ξ) and one of G(β) distributions have been presented in Figures 2 and 3.

The comparison of the estimates of the parameters of stochastic structure and the number of significant structural parameters indicates that the evaluated models have at least good predictive properties. Therefore, the forecasts of parameters for the following 12 months of 2008 have been constructed together with their subsequent empirical verification on the basis of all evaluated equations. The mean absolute percentage error (MAPE) have been calculated for the forecast horizons equals: 3, 6, 9 and 12 months. The estimates of these parameters are to be found in four last columns of the analyzed table.

The models with the best predictive properties or the lowest evaluations of relative errors of *ex post* forecasts for the respective prospects of the forecast have been put in bold.

In most cases, the models with the most favourable predictive properties did not characterize by the lowest estimates of errors of relative *ex post* forecasts of parameters. This suggests the existence of a discrepancy between the predictive properties of equations and the accuracy of forecasts.

Table 3. Goodness of fit of time series models: number of significant parameters and the accuracy of *ex post* forecasts

Distribution	Distribution parameter	Model	Goodnes of fit		Number of significant parameters		MAPE [%]			
			R^2	V_s [%]	trend	seasonal	$h = 3$	$h = 6$	$h = 9$	$h = 12$
GEV	ξ	*l*	0.5245	5.35	2	2	5.97	11.44	> 20%	> 20%
		kw	0.6234	4.76	2	4	**1.05**	**3.77**	**12.18**	**18.37**
		3st	**0.7128**	**4.16**	**3**	**5**	12.81	14.28	15.60	> 20%
	σ	*l*	0.9418	2.09	2	7	**0.66**	**1.43**	**2.49**	**2.61**
		kw	**0.9872**	**0.98**	**3**	**9**	3.94	4.78	5.43	6.84
		3st	0.9867	1.00	1	9	3.56	4.24	4.67	5.85
		w	0.9185	2.54	2	3	0.65	1.94	3.43	3.80
		w2st	0.9837	1.14	2	9	4.44	4.46	4.37	5.30
		w3st	0.9829	1.16	1	9	4.31	4.29	4.14	5.01
	μ	*l*	0.9479	3.70	2	1	5.77	5.99	5.82	**4.92**
		kw	0.9944	1.21	3	9	4.97	8.88	13.37	18.79
		3st	**0.9966**	**0.94**	**3**	**11**	**0.12**	**1.59**	**3.00**	4.97
		w	0.9103	5.11	2	0	6.62	8.92	10.50	10.21
		w2st	0.9964	1.02	2	11	4.54	5.38	6.85	9.69
		w3st	0.9968	0.96	3	11	2.75	3.02	3.78	5.88
G	α	*l*	0.9155	4.10	2	0	9.48	10.82	10.90	9.24
		kw	0.9880	1.55	2	1	2.62	5.26	9.39	15.28
		3st	0.9907	1.36	4	5	2.62	2.04	2.06	**3.89**
		w	0.8827	5.97	2	0	11.11	13.15	14.15	13.57
		w2st	0.9931	1.23	3	5	**0.39**	1.53	3.52	6.66
		w3st	**0.9931**	**1.23**	**2**	**5**	0.89	**0.88**	**2.06**	4.52
	β	*l*	0.7685	1.52	1	8	5.23	5.30	5.11	5.00
		kw	**0.7879**	**1.45**	**2**	**8**	4.14	3.87	3.35	2.94
		3st	0.7811	1.48	1	8	**3.37**	**2.81**	**2.22**	**2.00**
		w	0.7664	1.52	2	8	5.32	5.37	5.17	5.06
		w2st	0.7841	1.47	2	8	4.23	3.98	3.48	3.06
		w3st	0.7776	1.49	1	8	3.40	2.86	2.27	2.09

Source: own calculations.

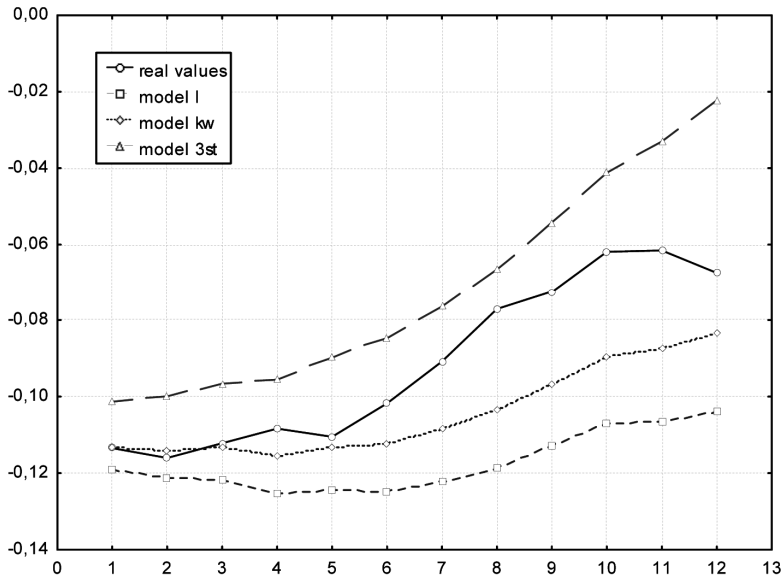


Figure 4. *Ex post* forecasts of parameter ξ of GEV distribution

Source: own calculations.

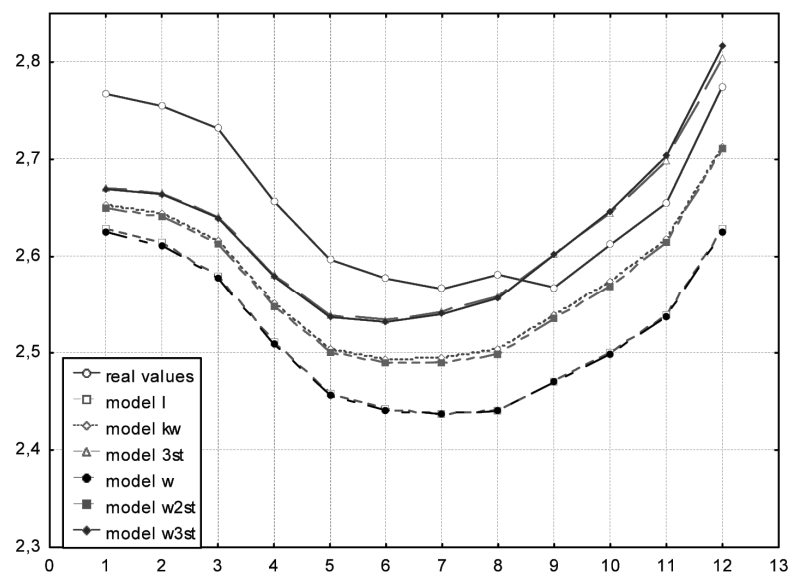


Figure 5. *Ex post* forecasts of parameter β of G distribution

Source: own calculations.

The results of forecasting the same parameters, for which the modelling results have been presented in Figures 2 and 3, are to be found in Figures 5 and 6.

The comparison of the pairs of figures referring to the modelling and forecasting ξ and β parameters and the estimates presented in Table 3 indicate that despite similar estimates of the coefficients of random volatility for the respective equations, the estimates of the errors of *ex post* forecasts are much more diverse.

At the last stage, on the basis of the equations characterized by the lowest estimates of the errors of relative *ex post* forecasts of parameters, the forecasts of distributions have been created, and subsequently, testing the p compliance of the distributions acquired in this manner with the empirical distributions took place. Table 4 specifies the estimates of the empirical statistics of the tests of compliance of empiri-

Table 4. The comparison of the compliance of theoretical and forecast distributions of the unemployment rate with the empirical distributions in the year 2008

Distribution	Month	The values of statistics for theoretical distributions		The values of statistics for forecast distributions	
		λ_{emp}	χ^2_{emp}	λ^*_{emp}	χ^{2*}_{emp}
GEV	January	0.5891	2.27	0.5938	2.56
	February	0.5046	2.98	0.5124	2.96
	March	0.5360	1.36	0.5474	1.67
	April	0.5862	5.34	0.4604	3.09
	May	0.7067	4.20	0.8058	3.34
	June	0.9189	3.00	0.7744	5.46
	July	0.9670	7.87	0.9164	9.94
	August	0.8377	10.43	1.0094	13.49
	September	0.7230	6.27	1.0842	14.06
	October	0.7365	9.08	1.0746	9.13
	November	0.7055	8.53	0.6282	8.40
	December	0.6403	6.95	0.8486	10.04
G	January	0.6687	3.83	1.1761	9.82
	February	0.6247	3.74	1.2062	10.55
	March	0.6800	5.89	1.2592	9.78
	April	0.6623	4.22	1.0857	7.22
	May	0.7653	3.84	1.2829	9.04
	June	0.8032	5.70	1.1685	8.60
	July	0.8494	6.90	1.3258	15.17
	August	0.7554	6.17	1.4397	9.52
	September	0.6822	10.12	1.4994	16.64
	October	0.7548	7.52	1.5518	22.66
	November	0.7394	11.68	2.0101	33.28
	December	0.6500	5.53	2.4586	36.62

Source: own calculations.

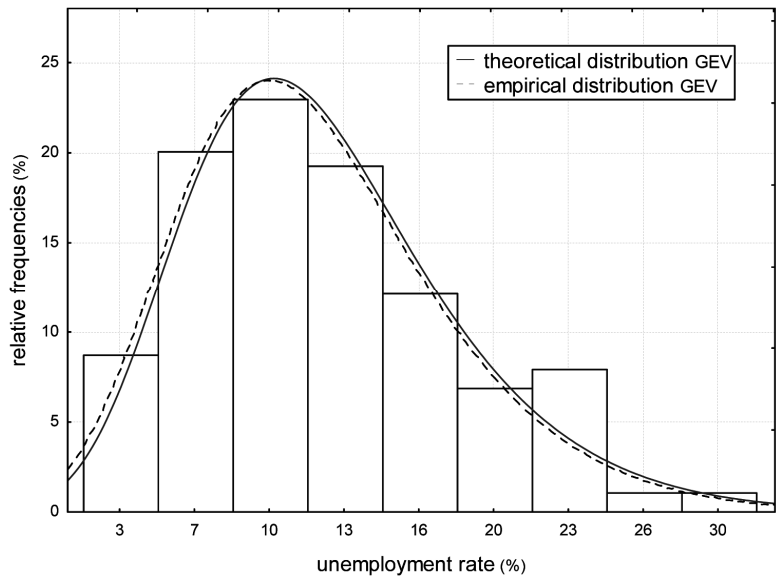


Figure 6. Empirical, theoretical and forecasting distributions of unemployment rate in December 2008

Source: own studies.

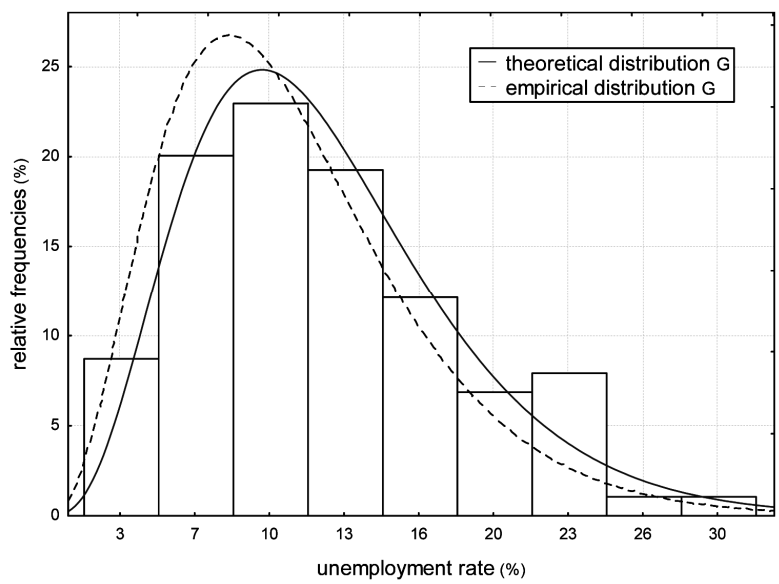


Figure 7. Empirical, theoretical and forecasting distributions of unemployment rate in December 2008

Source: own studies.

cal distributions: with theoretical distributions (λ_{emp} and χ^2_{emp}) and forecast distributions (λ^*_{emp} and χ^{2*}_{emp}). The information to be found in the table indicates that the estimates of empirical statistics obtained in all months of 2008 were lower than the critical values for theoretical approximates for both distributions. However, such a correlation does not take place in case of forecast distributions.

In case of the distribution of extreme values, the estimates of the empirical statistics of –λ-Kolmogorov test (λ^*_{emp}) were also lower than the critical values for all months. In case of the test of compliance χ^2 the values of statistics χ^{2*}_{emp} the estimates higher than the critical values were obtained for August and September.

The analysis of λ^*_{emp} and χ^{2*}_{emp} statistics obtained for G distribution indicates that they were lower than the critical values from January to June and additionally, respectively, in July and August. This means that in case of G distribution, the prospects of the forecast should not exceed 6-7 months.

The results of modelling and *ex post* forecasting of the unemployment rates for GEV and G distributions in December 2008 have been presented in the graphic form in Figures 6 and 7.

The analysis of the foregoing figures indicates that in case of GEV distribution the density curves: theoretical and forecast, are almost convergent, which testifies to a high compliance between the forecast distribution and the empirical distribution. The forecast for the gamma distribution exhibits significant discrepancies with the empirical distribution. The appearance of such a discrepancy is indicated by high values of λ^*_{emp} and χ^{2*}_{emp} statistics, which are significantly higher than λ_{emp} and χ^2_{emp} statistics. Thus, the conclusion that has been put forward before, and which refers to a considerably shorter forecast prospects for the G distribution, remains valid.

3. Summary

The observations included herein indicate that the *ex ante* forecasting process of homogeneous distributions should be preceded by the analysis of the accuracy of ex post forecasts of these parameters. The same type of procedure shall be conducted with reference to the distribution forecasts so that the best model was selected from all the admissible models for the purpose of *ex ante* distribution forecasting.

The study has also proved that the time series models with seasonal fluctuations may successfully be used in the modelling and forecasting the parameters of homogeneous distributions in the situation when the phenomenon examined, and thus the distribution parameters, characterize by the appearance of seasonal fluctuations.

References

Domański Cz., *Statystyczne testy nieparametryczne*, PWE, Warszawa 1979, pp. 88-89.

Guzik B., *Prognozowanie funkcji*, Prace Naukowe Instytutu Cybernetyki Ekonomicznej, Zeszyty Naukowe AE w Poznaniu, Zeszyt Nr 184, Poznań 1991, pp. 7-20.

Johnson N.L., Kotz S., Balakrishnan N., *Continuous Univariate Distributions*, Vol. 1, Wiley, New York 1994.
Kordos J., *Modele prognoz rozkładu płac według wysokości*, [in:] *Wybrane problemy prognoz statystycznych*, t. 11, GUS, Warszawa 1970a, pp. 230-243.
Kordos J., *Prognoza rozkładu płac za 1970 r. w gospodarce uspołecznionej*, Wiadomości Statystyczne 1970b, no. 5 (108), pp. 39-41.
Kordos J., *Metody analizy i prognozowania rozkładów płac i dochodów ludności*, PWE, Warszawa 1973.
Kotz S., Nadarajah S., *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London 2000.
Zawadzki J., *Ekonometryczne metody prognozowania procesów gospodarczych*, Wydawnictwo Akademii Rolniczej, Szczecin 1995.

O PROGNOZOWANIU ROZKŁADÓW ZMIENNYCH WYKAZUJĄCYCH WAHANIA SEZONOWE

Streszczenie: w pracy na przykładzie rozkładów empirycznych miesięcznych stóp bezrobocia według powiatów i miast na prawach powiatów przedstawiono przykład modelowania i prognozowania rozkładów homogenicznych zmiennej wykazującej wahania sezonowe. Wykazano w niej, że proces budowy prognoz rozkładów *ex ante* powinien być poprzedzony analizą *ex post* dokładności prognoz parametrów i zgodności rozkładów.