RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS No. 91
Econometrics 28 2010

Konstancja Poradowska
Wrocław University of Economics

FORECASTING NEW EVENTS OCCURRENCE TIME ON THE EXAMPLE OF THE ANALYSIS OF FORESIGHT RESEARCH RESULTS

Abstract: The basic tools in the new events forecasting are the methods based on judgmental knowledge of the group of the qualified experts. The group of the qualified experts is usually not enough big and therefore it is groundless to use classical statistical tools. Instead it is common to utilize some subjective probability distribution for statistical interference. In previous papers to describe the subjective probability distribution I have applied triangular, two-sided-power, beta, Weibull's distributions and – in case of one person opinion analysis – trapezoidal and uniform distributions. In this paper I have additionally shown a method of new events forecasting for foresight research outputs analysis.

Key words: judgmental forecasting, subjective probability, Delphi method, foresight.

1. Introduction

Forecasts of new events occurrence time play an important role in managing the modern economy. The current forecast problems are for example the times when Poland will join Eurozone, when the share of biofuel in the transport fuel market will raise to 25%, and when new nuclear technology within atomic energy will be implemented. Undoubtedly, what raises the interest in forecasts in question in Poland is the growing popularity of foresight projects. The key task of foresight projects is to determine realization time of the given assumptions.

To forecast event occurrence time is not an easy task, especially when we are interested in new events which have not yet been observed. In such situations classical forecasting methods fail since they require empirical data from the past and they use opinions of experts to build forecasts.

This paper aims at presenting suggestions of how to gather experts opinions about the event occurrence time for the ultimate point and interval forecasts. As the tool for reasoning the notion of subjective probability has been applied. Forecasting on the basis of Delphic method, which is widely applied in Poland in foresight projects, has been chosen as the background for assumptions presented.

2. Delphi method in new events occurrence time forecasting

Delphi method is a tool widely used for forecasting and new events estimation. It was developed in Rand Corporation in the fifties of the last century. Its assumptions, however, at that time were used only to meet military needs [Dalkey 1969]. "Classical Delphi method" involves sending an anonymous questionnaire to a se-lected group of experts. The answers received are processed statistically and, on the condition that they converge sufficiently, they are aggregated for the final judgments. In order to achieve consensus, a questionnaire is usually distributed to the same group of respondents at least twice. In later rounds respondents receive, together with the same set of questions, the results of structure of answers from the earlier rounds. More information on Delphi method can be found in [Antoszkiewicz 1990].

In Delphic questionnaire a question about event realization time can be formulated in a variety of ways. To answer a question an expert might be requested to point to a particular moment of the event occurrence (e.g. a month, a year) while answers variants are not given or to choose one variant, usually in the form of five time periods, provided by the research organizers. An alternative question, however less frequently posed, asks in which time perspective can a particular event occur in relation to a moment defined in time. In each case obtaining results involves determining measures of location and variance in order to determine the convergence of expert opinion.

It should be emphasized that obtaining reliable results of the research calls for a large representative group of experts who ought to possess impressive knowledge and experience related with issues subject to research. However, it is frequently difficult to gather a lot of experts who are competent in a given field – especially when a new problem (for example latest advanced technologies development) are a subject to research. Therefore, in such a situation, it is more difficult to carry out a classical statistical analysis and to formulate adequate conclusions concerning future.

3. *Foresight* research related issues

In accordance with B. Martin's definition foresight is "the process involved in systematically attempting to look into the longer-term future of science, technology, the economy, the environment and society with the aim of identifying the emerging generic technologies and the underpinning areas of strategic research likely to yield the greatest economic and social benefits" [Martin 2001].

Therefore the most important objective of this type of research is to construct situation development scenarios that would cover a relatively long time perspective (usually 20-25 years) when extrapolation of the possessed knowledge about the past is not possible. That is why basic foresight tools include methods that are based on knowledge of experts, while Delphi questionnaire is most frequently applied tool to carry out the research (see e.g. [Kuwahara 2001; Miles, Keenan 2002; Oniszk--Popławska et al. 2004]).

In December 2006 in Poland National *Foresight* Programme "Poland 2020" was launched by the Minister of Science and Higher Education. The programme encompasses three research areas:

1) sustainable development of Poland,
2) information and telecommunications technologies,
3) security [www.mg.gov.pl].

National *Foresight* Programme in Poland should result in directing research and technologies development to those areas which guarantee dynamic economic development, and rationalizing spending realized from public resources. The key role in the research is played by defining realization time of a particular region development determinants which were previously indicated. However, when new advanced technology implementation is one of the determinants, it is frequently difficult to gather a group of competent enough experts. This, in turn, can arise doubts about applying classical techniques of Delphi questionnaires analysis, which are based on statistical measures of location.

In the further section of the paper an alternative to obtaining expert judgments about new events occurrence time shall be presented. The approach in question is based on the notion of subjective probability and can be implemented in case of a very small number of opinions. In case of extreme situations, it is possible to analyse an opinion provided by one expert only.

4. The idea of subjective probability

Subjective probability demonstrates a personal measure of the probability that some event will take place – it depends on our experiences, opinions heard, personal beliefs and even prejudices. Randomness is here a result of "lack of knowledge" about conditions influencing the event. According to B. de Finetti, regarded one of "fathers" of subjective probability, it can be deducted on the basis of observation of actions taken by other people, for instance bets taken in horse racing. The relation between the notion in question with expert opinions can be regarded as obvious one. Its implementation in this area is not new (see e.g. [Dittmann 2008; Orzeł 2005; Peter 1986; Poradowska 2008; Press 2003]).

When there is no empirical data which could be used to deduct distribution of a variable of interest (the variable can be a given event time occurrence), it is possible to assume *a priori* that there is a determined form of the probability distribution – simultaneously basing on different forms of information (particularly including experts' opinions and some subjective conviction that the opinions are right).

To describe subjective probability such density functions of the distributions are most frequently used which are explicitly definite by three parameters: minimal (a), the values of the variable, most probable values of the variable (w) and maximal values of the variable (b), like for instance a triangle distribution or beta[1]. Intuitively obvious interpretation of those parameters is significant in expert opinions analysis. Examples of selected density functions are presented in Figure 1.

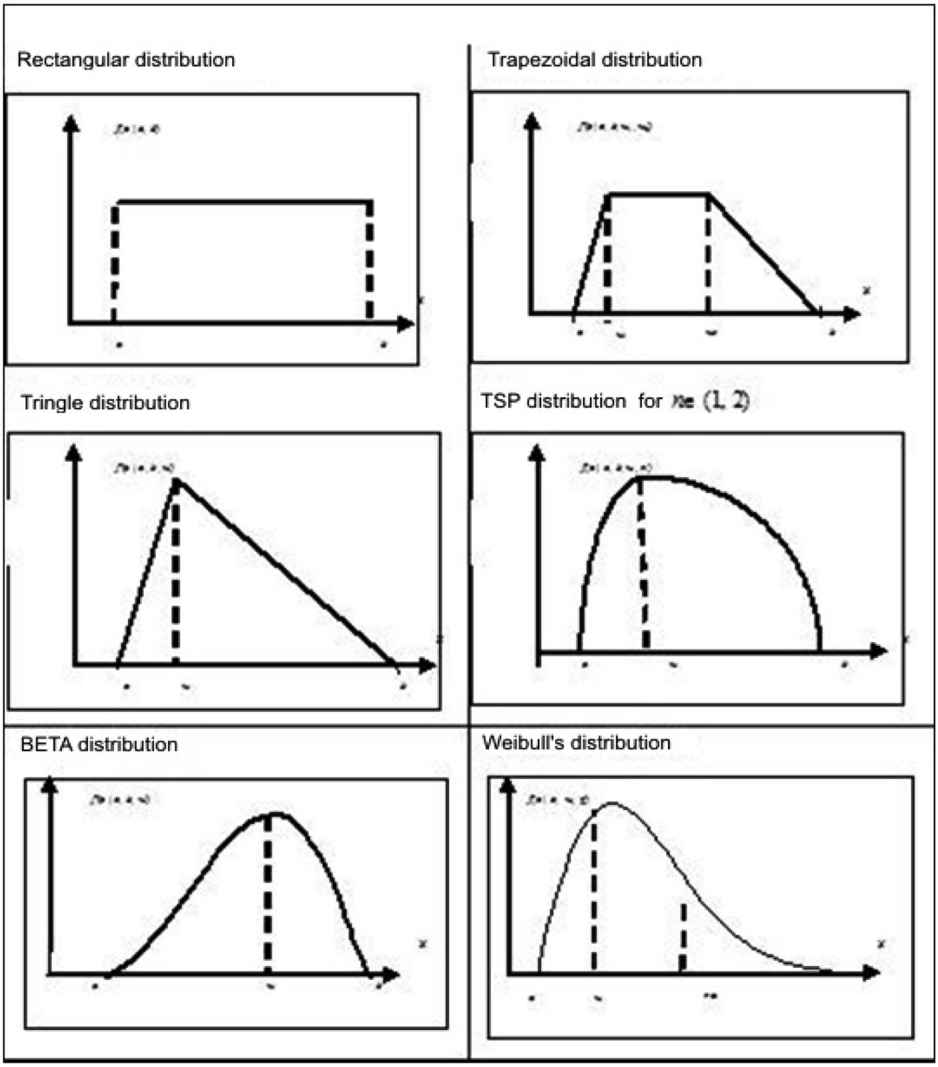

Figure 1. Examples of density functions of selected probability distributions

Source: own elaboration.

[1] Distribution in question is commonly used to estimate event duration time in PERT method that was frequently employed while assessing risk of investment projects.

Adequate formulas which allow to determine the expected value and standard deviation of individual distributions are included in Table 1.

In an extreme situation when opinion is provided by only one expert, an expert is asked to determine an interval $[a, b]$ in which, in his opinion, real value of the variable will be placed for 100%. Afterwards an expert is asked to point to one value from the interval in question (w) which he estimates as the one with the highest probability of realization. Assumption about probability density function shape between indicated values should be made on the basis of all information about the problem studied which is available to experts and researchers.

When a large group of experts takes part in the research, distribution of their opinions can be treated as a kind of approximation of real variable distribution. Extremes answers are then taken for values a and b, while value w is determined on the level of a modal value which is found by employing interpolation formula [Poradowska 2008; Poradowska, Wójciak 2009]. Moreover, it is assumed that event occurrence time has such a constant probability distribution whose density function shape resembles expert answers distribution.

Generalisation of a triangle distribution, i.e. the TSP distribution (two-sided power distribution) should additionally be listed here as a less known form of the distribution [Poradowska 2008; Van Dorp, Kotz 2002]. The distribution in question includes an additional shape parameter n. When $n = 1$, the distribution comes down to a uniform distribution of the $[a, b]$ interval. When $n = 2$, TSP distribution becomes a triangle distribution. When $1 < n < 2$, the density function diagram looks like a "triangle of convex sides". However, if $n > 2$, the triangle in question has "concave sides". By choosing n parameter in a right way, TSP distribution may help describe numerous known shapes of probability distributions including unimodal, U-shaped and J-shaped ones.

When the open time interval is one of the variants presented to experts, Weibull's distribution may be deemed most suitable. Its density function is limited exclusively from the bottom with the value a. To determine parameters of distribution in question, it is necessary to determine on the basis of expert opinions a selected upper percentile of the distribution, the value q_α. It is expected that from possible values of the variable $(1 - \alpha)$% are bigger than it.

An important "advantage" of using unimodal probability distributions in event occurrence time approximation is the fact that for distribution in question the shortest of intervals can be determined $[x_1, x_2]$. The interval in question with a given probability p will cover the real value of variable[2]. For a triangle distribution and its generalization it can be done analytically. Interval limits are as follows:

$$x_1 = a + (w - a)(1 - p)^{\frac{1}{n}}, \qquad x_2 = b - (b - w)(1 - p)^{\frac{1}{n}}. \tag{1}$$

[2] This way of proceeding corresponds with a principle of interval forecast [Pawłowski 1973].

For beta or Weibull's distribution – numeric procedure remains [Dittmann, Poradowska 2008; Poradowska, Wójciak 2009].

Deciding to apply trapezoidal distribution, instead of using one most probable value (w) it is necessary to determine interval $[w_1, w_2] \subset [a, b]$, which should include the most values of variable regarded as most probable by experts. The approach in question may, however, be deemed controversial since the interval $[w_1, w_2]$ long in relation to $[a, b]$, may prove low convergence of experts. This, in turn, is a setback for formulating judgments on the expert opinions. Answers distribution similar to rectangular also indicates the lack of convergence in a group of experts. It is worth, however, to notice that both trapezoidal and rectangular distribution may be implemented do analyze an opinion of one expert. Instead of determining one value it is usually easier for one expert to determine interval (here time interval) in which the real value of variable studied should be put. When a researcher does not have any additional information about the problem forecasted, apart from the interval limits, rectangular distribution seems to be the most suitable.

5. Subjective probability in the analysis of *foresight* research results

It was attempted to apply subjective probability in the analysis of results of *foresight* project which concerned new energy technologies.

In Delphic 125 surveys that were created for the research the most important role was played by a question concerning the realization time. In their answers experts pointed to one of five variants: till 2010, in the years of 2011-2020, 2021-2030, after 2030 and never. Typical distributions of answers provided are presented below (see Figure 2).

It is assumed that a particular thesis will not be completed when share of "never" answers is the largest out of all variants. Such theses should not be subject to any further analysis. However, in results discussed such situations did not appear. In remaining cases the survey answers other than "never" were defined. On the basis of answers three values of a forecast variable were set: a – minimal, w – the most probable and b – maximal or $q\alpha$ – α rank quantile. The year of 2008 was assumed to be the minimal value of the forecast variable because the research was conducted that year. The most probable value was found on the basis of answers provided by experts. Interpolation formula of a modal value was employed. A modal value could not be set in 40 of the studied theses (see Example 2, Figure 2). This made their further analysis impossible.

If a set of possible values of variable could be deemed, upper bound that is no expert would point to the last open time interval, year 2030 should be understood as the maximum value (b) and triangle, beta, PERT and TSP distributions would be considered. When the minimal value (b) was assumed to be 2030 and to compare results triangle distributions, beta PERT and TSP were considered. In case of the

TSP distribution the n parameter may be estimated by means of the highest probability method [Van Dorp, Kotz 2002]. The statististical value χ^2 may be regarded as one of the criteria of an adequate distribution choice. In questions from the field of advanced technologies a group of experts always pointed to the "after 2030" variant. Therefore it was decided to apply Weibull's distribution. The $q\alpha$ value was assumed to be 2031 and α probability – percentage of answers that pointed to three first (closed) time intervals.

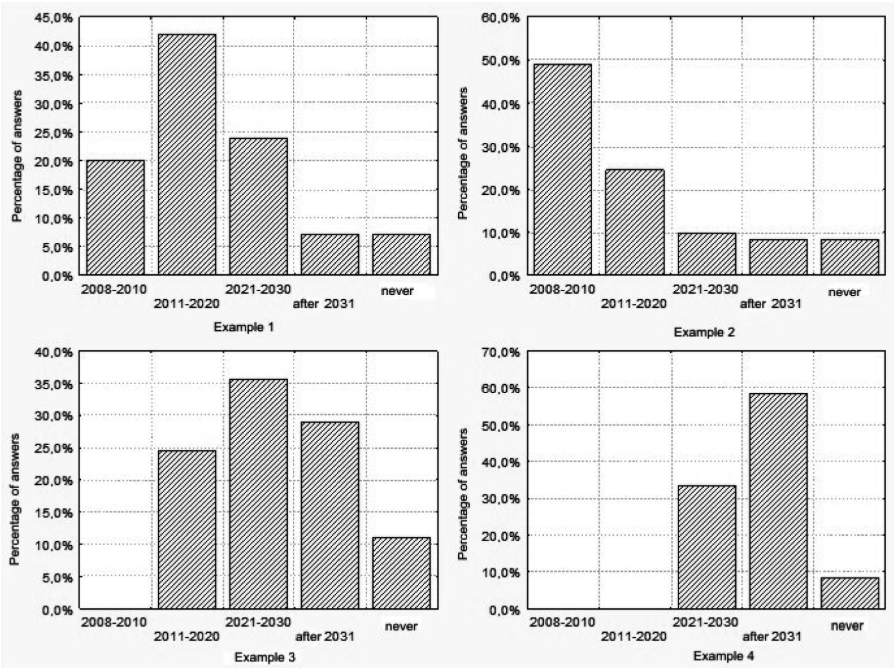

Figure 2. Exemplary distribution of the results of thesis completion time

Source: [Poradowska, Wójciak 2009].

Knowing α probability and q_α, the $q\alpha$ quantile the maximum value (b) was found by means of the equation: $\int_a^{q_\alpha} f(x)dx = \alpha$ where $f(x)$ is a density function of probability distribution.

Assuming a form of variable distribution shape between indicated values, interval forecasts of the theses realization time were built. Formulation of point forecasts was given up because of the very nature of the forecast variable pointing to a particular year was deemed unjustified in this case. Exemplary forecast of a particular event realization time built in accordance with a procedure described is presented below.

Table 1. Formulas of expected value and standard deviation of selected probability distributions

Distribution	Setpoints	Expected value	Standard deviation
Uniform	a, b	$\frac{a+b}{2}$	$\frac{b-a}{\sqrt{12}}$
Trapezoidal	a, b, w_1, w_2	$\frac{1}{3}\left(a+w_1+w_2+b-\frac{bw_2-aw_1}{b-a+w_2-w_1}\right)$	$\sqrt{\frac{1}{18}(b-w_1)^2+(w_2-a)^2+(w_1-a)(b-w_2)-\frac{2(b-a)(w_2-w_1)(w_2-a)(b-w_1)}{(b-a+w_2-w_1)^2}}$
Triangle	a, b, w	$\frac{a+w+b}{3}$	$\sqrt{\frac{(b-a)^2+(w-a)(b-w)}{18}}$
TSP	a, b, w, n	$\frac{a+(n-1)w+b}{n+1}$	$\sqrt{\frac{n(b-a)^2-2(n-1)(w-a)(b-w)}{(n+2)(n+1)^2}}$
Beta	a, b, w	$\frac{a+4w+b}{6}$	$\sqrt{\frac{(m-a)(b-m)}{7}}$
Weibull's*	a, w, q_α	$\gamma\cdot\Gamma\left(1+\frac{1}{\beta}\right)+a$	$\gamma\cdot\sqrt{\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)}$

* β and γ parameters in Weibull's distribution formulas are found by means of the following equations:

$$(1)\quad \left[-\ln(1-\alpha)\cdot\frac{\beta}{\beta-1}\right]^{\frac{1}{\beta}}=\frac{q_\alpha-a}{w-a};\quad (2)\quad \gamma=(w-a)\left(\frac{\beta}{\beta-1}\right)^{\frac{1}{\beta}}$$

Source: own elaboration.

6. Example

Results presented concern realization time of one of technology *foresight* theses. Table 2a presents distribution of answers of all 39 experts who took part in a thematic panel. Since the answers “never” constituted only 10% of all the answers it was assumed that a thesis will be realized. Subjective probability distribution for its realization time was found on the basis of answers other than “never” distribution (see Table 2b).

Table 2a. All answers distribution

Answer variant	Number of experts	Percentage of experts
Never	6	17
After 2030	3	9
2021-2030	8	23
2011-2020	14	40
Till 2010	4	11
In total	35	100

Table 2b. Distribution of answers other than „never”

Answer variant	Number of experts	Percentage of experts
After 2030	3	10
2021-2030	8	28
2011-2020	14	48
Till 2010	4	14
In total	29	100

Source: own elaboration on the basis of [Czaplicka-Kolorz K. (ed.) 2007b].

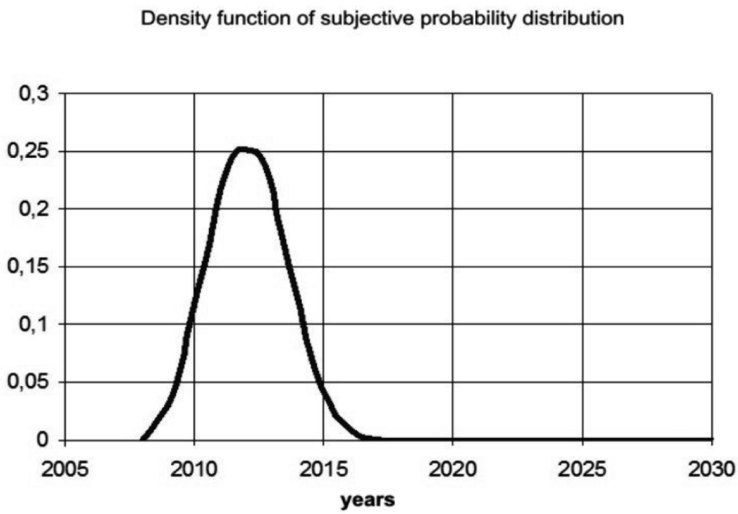

Figure 3. Density function of Weibull’s distribution determined on the basis of results included in Table 2b

Source: own elaboration.

Weibull's distribution parameters are as follows: a = 2008 (the year when the research was carried out); w = 2012 (the value found by employing interpolation formula of a modal value, having replaced numbers with density values); q_α = 2031 where probability α = 0.90 (90% of experts answered that the thesis will have been realized by 2031). Density function shape of such distribution is presented in Figure 3.

On the basis of formulas included in Table 1 the expected value and standard deviation were counted: m = 2012.07 (which may be interpreted as January 2012), s = 1.46. Therefore the rule of unloaded forecast, similarly to the rule of highest probability, indicates that the thesis will be realized in 2012. If results of both rules differ, according to the postulate of maximizing of forecast realization chances, point forecast shall be built on the level of a modal value. However, in the face of a long-term perspective of the forecast built as well as the character of the event forecasted, that is the lack of possibility to extrapolate the knowledge possessed, it seems to be more suitable to define time interval than to build point forecast. In the time interval in question it is possible to expect the realization of thesis assumptions[3].

Table 3 presents limits of intervals (x_1 and x_2) which were found in a numerical way as well as their Δ for selected probabilities p.

Table 3. Limits and lengths of forecast intervals for selected probabilities p

p	x_1	x_2	Δ
0.50	2011.0	2013.0	2.0
0.80	2010.1	2013.9	3.8
0.90	2009.6	2014.5	4.9

Source: own elaboration.

With 0.90 probability it is possible to expect that thesis will be realized in the interval from 2010 to 2015. Similar conclusions may be made for other assumed probabilities.

7. Conclusions

Subjective probability-based reasoning may provide a convenient tool to be used while forecasting new events occurrence time, especially in case of a small group of experts, i.e. when standard statistical tools cannot be used. Probability distributions that may be implemented to analyse experts' judgements include a triangle, TSP, beta and Weibull's distributions. In case of having one expert only also a rectangular and trapezoidal distribution can be implemented.

Only in case when a number of experts is sufficiently large (a few dozen) and their competencies are highly evaluated, a distribution of answers may be deemed a satisfactorily reliable estimation of the actual distribution of the variable subject to

[3] How to build such an interval was discussed in the works [Poradowska 2008, 2009].

research. An assumed probability distribution should be also influenced by researcher's knowledge, experience and intuition. "Correct" determination of a distribution form is of much importance because despite the same parameters determined on the basis of experts' judgements (e.g. a, w and b) different probability distributions may give different forecasts. Moreover, values lengths of forecast intervals for different probability distributions show remarkable differences. When a group of experts is small and the very nature of any phenomenon subject to research provides no foundations for reasoning that the probability distribution has a 'more sophisticated' shape, it is recommended to employ a triangle distribution, or even a rectangular one. However, forecasts built on rectangular distribution may demonstrate a higher level of uncertainty.

References

Antoszkiewicz J.D., *Metody heurystyczne. Twórcze rozwiązywanie problemów*, PWE, Warszawa 1990.

Czaplicka-Kolorz K. (ed.), *Scenariusze rozwoju technologicznego kompleksu paliwowo-energetycznego dla zapewnienia bezpieczeństwa energetycznego kraju – część 2. Scenariusze opracowane na podstawie foresightu energetycznego dla Polski na lata 2005-2030*, Główny Instytut Górnictwa, Katowice 2007b.

Dalkey N.C., *The Delphi Method. An Experimental Study of Group Opinion*, Rand, Santa Monica 1969.

Dittmann P., *Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie*, Oficyna Ekonomiczna, Kraków 2008.

Dittmann P., Poradowska K. (2008), *Experts' Opinions in Forecasting for Enterprises Management*, Faculty of Economics and Management Press, Zielona Góra 2008.

Kuwahara T., *Technology Foresight in Japan – The Potential and Implications of DELPHI Approach*, NISTEP Study Material 77, 2001, www.nistep.go.jp, pp. 125-141.

Martin B.R. (2001). *Technology Foresight in a Rapidly Globalizing Economy*, International Conference on "Technology Foresight for Central and Eastern Europe and the Newly Independent States", Vienna, Austria, 4-5 April 2001.

Miles I., Keenan M., *Practical Guide to Regional Foresight in the United Kingdom*, Unit "Science and Technology Foresight", European Commission/DG Research, 2002.

Oniszk-Popławska A., Bonacina M., Holst J.B., Verte D., Weinert T., *Metoda foresightu technologicznego zastosowana do oceny przyszłości energetycznej Europy w projekcie EurEnDel*, Projekt EurEnDel, January 2004, *www.izt.de.*

Orzeł J., *Rola metod heurystycznych, w tym grupowej oceny ekspertów, oraz prawdopodobieństwa subiektywnego w zarządzaniu ryzykiem operacyjnym*, Bank i Kredyt, May 2005.

Pawłowski Z., *Prognozy ekonometryczne*, PWN, Warszawa 1973.

Peter C.F., *The axioms of subjective probability*, Statistical Science 1986, Vol. 1, No 3.

Poradowska K., *Możliwości wykorzystania rozkładu trójkątnego do konstrukcji prognoz punktowych i przedziałowych*, Postępy Statystyki, Ekonometrii i Matematyki Stosowanej w Polsce Południowej, University of Economics Publishing House, Kraków 2008.

Poradowska K., Wójciak M., *Uogólniony rozkład trójkątny w analizie badania typu foresight*, Dynamiczne Modele Ekonometryczne, Prace Naukowe UMK, Zeszyt 389, Toruń 2009.

Press J.S., *Subjective and Objective Bayesian Statistics: Principles, Models and Applications*, Wiley & Sons, New Jersey 2003.

Van Dorp J.R., Kotz S., *A novel extension of the triangular distribution and its parameter estimation*, The Statistician 2002, Vol. 51, Part 1, pp. 63-79.

PROGNOZOWANIE CZASU ZAJŚCIA NOWYCH ZDARZEŃ NA PRZYKŁADZIE ANALIZY WYNIKÓW BADANIA *FORESIGHT*

Streszczenie: podstawowymi narzędziami w prognozowaniu czasu zajścia nowych zdarzeń są metody oparte na wiedzy ekspertów merytorycznych. Grupa takich ekspertów często bywa niewystarczająco liczna, aby przy formułowaniu ostatecznych sądów uzasadnione było korzystanie z klasycznych metod statystycznych. W takiej sytuacji można zaproponować wnioskowanie w oparciu o rozkład prawdopodobieństwa subiektywnego analizowanej zmiennej. Do opisu prawdopodobieństwa subiektywnego można wykorzystać następujące rozkłady: trójkątny oraz jego uogólnienie (rozkład TSP), beta, Weibulla, a w szczególnym przypadku korzystania z sądu tylko jednego eksperta – dodatkowo rozkład prostokątny oraz trapezowy. Stosując ideę prawdopodobieństwa subiektywnego, w referacie przedstawiono sposób konstrukcji prognoz czasu zajścia nowych zdarzeń na potrzeby analizy wyników badania *foresight*.