

SELECTED ECONOMETRIC METHODS OF MODELLING THE WORLD'S POPULATION

Witold Rzymowski, Agnieszka Surowiec

Lublin University of Technology, Lublin, Poland

e-mails: w.rzymowski@pollub.pl; a.surowiec@pollub.pl

© 2018 Witold Rzymowski, Agnieszka Surowiec

This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivs license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>)

DOI: 10.15611/eada.2018.2.03

JEL Classification: C51

Abstract: Selected econometric methods of modelling the world's population size based on historical data are presented in the paper. Periodical variables were used in the models proposed in the paper. Moreover, a logistic-type function was used in modelling. The purpose of the paper was to obtain a model describing the world's population with the lowest possible maximal relative error and possibly the longest period of durability. In this work, 13,244 models from three families models were analyzed. Only a small part of such a large number of models satisfies the conditions of stability. The method of modelling the world's population size allows to obtain models with maximal relative errors not exceeding 0.5%. Selected models were used to prediction of the world's population up to 2050. The obtained results were compared with data published by the Organisation for Economic Co-operation and Development.

Keywords: nonlinear models, estimation, maximal relative error, population, forecast.

1. Introduction

For many years, scientists have been trying to predict the world's population size. The mathematical modelling of natural phenomena has been performed by biologists, mathematicians, physicists, economists. Nowadays, internet portals inform us how many people will be living in the world in 2050 or even in 2100.

The oldest model describing dynamics of a single population, introduced by British economist and demographer Thomas Malthus in 1798 is known in two versions: a continuous one and a discrete one. The continuous version is known in the form of a linear differential equation and the discrete in the form of a linear difference equation. The solutions of those equations are exponential and power functions, respectively. Thomas Malthus noted in his paper entitled "An Essay on the Principle of Population" [Malthus 1798] that the growth of the human population is

too quick compared with the growth of food supplies, which in his opinion must result inevitably in disaster.

The next researcher involved in modelling of the dynamics of the population was Belgian mathematician Pierre Verhulst. The model proposed by him around 1840 [Verhulst 1838] is probably the first known nonlinear model in population biology [Foryś, Poleszczuk 2011].

The Verhulst model is a logistic one. It was created as a result of a debate of a group of scientists over the Malthus model. Most participants of the debate agreed about the existence of the natural processes that decelerate the unlimited growth of population postulated by Malthus.

The logistic function was also used by biologist R. Pearl [Pearl, Reed 1924] in his research, and the Verhulst equation is sometimes referred to as the Verhulst-Pearl equation. Unfortunately, predictions based on the exponential or logistic model turned out to be false [Holzer 2003; Murray 1989]. Pearl and Reed [Pearl, Reed 1924] announced that the maximal size of the world's population would be 2 billion whereas this value had been already exceed in 1930. Another known model describing the world's population size is the Doomsday model [von Foerster et al. 1960; Smith 1977; Austin, Brewer 1971; Serrin 1975]. In recent years there have been attempts at using the arctangent function for purposes of the prediction of the world's population size, and that of a continent or a country [Smolik 1996; Hyb, Kaleta 2004]. Unfortunately, the above models, whereas describing historical data of the population size correctly, were not useful for the purposes of prediction. The predicted value had a too large an error. The authors agree with Jerzy Holzer's opinion that finding the parameters of a model for some periods of time is possible, but conclusions about the world's future population size may be very illusive [Holzer 2003]. In attempting to find the best possible model describing the world's population, in [Rzymowski, Surowiec 2012] a quasi-logistic model was proposed and a method of the estimation of the parameters of that model was described. Recently, in [Rzymowski, Surowiec 2017] a method of construction of a model describing the world's population size based on the 1950-2013 data available from [<http://stats.oecd.org>, 04 Mar 2016] using difference equations was presented. The obtained model of the world's population could be used for prediction and the relative error of prediction should not exceed 0.55% over 10 years.

In this paper we present some other ways of modelling the world's population size. The methods presented in the paper use periodical variables and a logistic-type function. The purpose of the paper is to obtain such models describing the world's population which would result in the best possible prediction throughout the longest possible period. In the paper the methods of obtaining the models were presented and the performance of the best models selected out of the proposed types are presented.

2. Data and world population models

Let L_t represent the world's population size in the year $1949 + t$ for $t = 1, 2, \dots, 64$. The data used for the model construction are the world's population sizes which are presented in Table 1. The analyzed time interval covers the period 1950-2013.

Table 1. The world's population in 1950-2013 in thousands

t	$L_t/1000$	T	$L_t/1000$	t	$L_t/1000$	T	$L_t/1000$
1	2 525 779	17	3 397 475	33	4 608 963	49	5 975 304
2	2 572 851	18	3 468 522	34	4 691 560	50	6 051 478
3	2 619 292	19	3 541 675	35	4 776 393	51	6 127 700
4	2 665 865	20	3 616 109	36	4 863 602	52	6 204 147
5	2 713 172	21	3 691 173	37	4 953 377	53	6 280 854
6	2 761 651	22	3 766 754	38	5 045 316	54	6 357 992
7	2 811 572	23	3 842 874	39	5 138 215	55	6 435 705
8	2 863 043	24	3 919 182	40	5 230 452	56	6 514 095
9	2 916 030	25	3 995 305	41	5 320 817	57	6 593 228
10	2 970 396	26	4 071 021	42	5 408 909	58	6 673 106
11	3 026 003	27	4 146 136	43	5 494 900	59	6 753 649
12	3 082 830	28	4 220 817	44	5 578 865	60	6 834 722
13	3 141 072	29	4 295 665	45	5 661 086	61	6 916 184
14	3 201 178	30	4 371 528	46	5 741 822	62	6 997 999
15	3 263 739	31	4 449 049	47	5 821 017	63	7 080 072
16	3 329 123	32	4 528 234	48	5 898 689	64	7 162 120

Source: own elaboration based on <http://stats.oecd.org>, 4 March 2016.

The method presented in this work is based on the following observation. If

$$f(L_t) = g(t), \quad t = 1, 2, \dots, N, \quad (1)$$

and f is an injective function then

$$L_t = f^{-1}(g(t)), \quad t = 1, 2, \dots, N. \quad (2)$$

Three types of models are used to model the world's population size L_t . In each of them a trigonometric model in the following form

$$g_{[k,\tau]}(t) = a_{[k,\tau]} \cos\left(\frac{2\pi}{\omega_k} t\right) + b_{[k,\tau]} \sin\left(\frac{2\pi}{\omega_k} t\right) + c_{[k,\tau]} + \varepsilon_t, \quad t = 1, 2, \dots, \tau \quad (3)$$

plays the main role. We take

$$\omega_k = 199 + k, \quad k = 1, 2, \dots, 301,$$

where k defines periods ω_k of function $g_{[k,\tau]}$. τ is a parameter that represents the end of the parameter estimation range of the model. The model of an output variable is then, due to equations (1, 2, 3), given by the formula:

$$\hat{L}_t(k, \tau) = f^{-1} \left(a_{[k,\tau]} \cos \left(\frac{2\pi}{\omega_k} t \right) + b_{[k,\tau]} \sin \left(\frac{2\pi}{\omega_k} t \right) + c_{[k,\tau]} \right), \quad t = 1, 2, \dots, N. \quad (4)$$

In this work the following functions

$$f^A(L_t) = L_t \quad (5)$$

$$f^B(L_t) = \frac{\ln L_t}{23,9 - \ln L_t} \quad (6)$$

$$f^C(L_t) = \frac{\sqrt{L_t}}{390000 - \sqrt{L_t}}, \quad (7)$$

were used for $t = 1, 2, \dots, 64$. The Least Squares Method [Nowak 2006; Rao 1982] is used for the estimation of the parameters of the linear model (3) for all the cases A, B, C , for each $k = 1, 2, \dots, 301$ and $\tau = 20, 21, \dots, 63$. In this way, by using the equations (2, 3, 4, 5, 6, 7) one can obtain three families of models M^A, M^B, M^C and each of them contains

$$301 \times 44 = 13,244$$

models of the output variable L_t in the following form

$$\hat{L}_t^A(k, \tau) = g_{[k,\tau]}(t), \quad t = 1, 2, \dots, N, \quad (8)$$

$$\hat{L}_t^B(k, \tau) = \exp \frac{23.9 g_{[k,\tau]}(t)}{1 + g_{[k,\tau]}(t)}, \quad t = 1, 2, \dots, N \text{ and} \quad (9)$$

$$\hat{L}_t^C(k, \tau) = \left(\frac{390000 g_{[k,\tau]}(t)}{1 + g_{[k,\tau]}(t)} \right)^2, \quad t = 1, 2, \dots, N. \quad (10)$$

Let consider that the period of durability of analyzed models \hat{L}_t^A , \hat{L}_t^B and \hat{L}_t^C is long. In the case of \hat{L}_t^A one can create the forecasts for $t = 1, 2, \dots, \infty$, in the case of \hat{L}_t^B the model will no longer be usable if $23.9 - \ln L_t = 0$, that is if $L_t = \exp(23.9) \approx 2.4 \cdot 10^{10}$, and the \hat{L}_t^C model will not work if $390000 - \sqrt{L_t} = 0$, that is if $L_t = 390000^2 = 1.5 \cdot 10^{11}$. The parameters 23.9 and 390 000 respectively in models (2.9, 2.10) are chosen arbitrarily. Moreover, one can see that the models (9, 10) are not sensitive to the change of these parameters.

3. Model quality assessment

The best model selection criterion that is used in this work is the criterion of minimizing the maximum relative error [Chen et al. 2010; Rzymowski, Surowiec 2012]. This criterion is used for all the families of models M^A, M^B, M^C given by the equations (8, 9, 10) respectively.

The selection procedure is the following.

First, we calculate:

- The relative errors for each models $\hat{L}_t(k, \tau)$ (see equations (8, 9, 10))

$$\delta_t(k, \tau) = \frac{|L_t - \hat{L}_t(k, \tau)|}{L_t} 100 [\%], t = 1, 2, \dots, 64. \quad (11)$$

- The maximum relative errors

– In the domain of parameter estimation

$$\delta_L(k, \tau) = \max_{t=1, 2, \dots, \tau} \delta_t(k, \tau). \quad (12)$$

– In the domain of “forecast” when $\tau < 64$,

$$\delta_R(k, \tau) = \max_{t=\tau+1, \tau+2, \dots, 64} \delta_t(k, \tau). \quad (13)$$

– In the domain $t = 1, 2, \dots, 64$

$$\delta_{\max}(k, \tau) = \max(\delta_L(k, \tau), \delta_R(k, \tau)). \quad (14)$$

Next, we choose:

- The model with the lowest maximum relative error:

$$\delta_{\min}(k^*, \tau^*) = \min_{k, \tau} \delta_{\max}(k, \tau). \quad (15)$$

- The models that satisfy the following conditions:

$$(I) \quad \delta_{\max}(\hat{k}, \hat{\tau}) \leq 0.5\% \quad (16)$$

and

$$(II) \quad \delta_R(\hat{k}, \hat{\tau}) \leq \delta_L(\hat{k}, \hat{\tau}). \quad (17)$$

The condition (I) (inequality 16) was established on the basis of previous studies [Rzymowski, Surowiec 2017]. All the models that satisfy the conditions (I) and (II) appear to be stable.

Finally, we find the world’s population models with the smallest number of initial data used for the estimation of the structural parameters in equation (3) that satisfies the conditions (I) and (II) (see equations 16, 17). This smallest number of initial data is denoted by $\hat{\tau}_-$. \hat{K} is the range that contains the parameters \hat{k} corresponding to the parameter $\hat{\tau}_-$.

4. Results

4.1. Models and their characteristics

The results presented in this chapter pertain to the resulting models of world population, as given by equations (8, 9, 10), where 13,244 models were studied for each of the respective model families M^A , M^B , M^C . Only a small proportion of this vast number of models meets the (I) condition and conditions (I) and (II) simultaneously.

Table 2. Number of models fulfilling the (I) condition and conditions (I) and (II) simultaneously for the M^A , M^B , M^C families of models

Families of models	(I)*	(I) and(II)**
M^A	36	13
M^B	141	88
M^C	14	5

* (I) – models that satisfy condition (I) given by equation (16).

** (I) and(II) – the models that satisfy conditions (I) and (II) given by equations (16) and (17).

Source: own elaboration.

The number of models meeting these conditions for each of the respective model families M^A , M^B , M^C is listed in Table 2. The largest number of models fulfilling the condition (I) and conditions (I) and (II) concurrently are found in M^B (see equation 2.9), and the lowest in M^C (see equation 2.10) family.

Table 3 presents the lowest quantity of initial data $\hat{\tau}_-$ included during the estimation of the remaining parameters of world's population models that met conditions (I) and (II) simultaneously and the k values that represent that $\hat{\tau}_-$ value for each of the respective M^A , M^B , M^C model families. In the case of the M^B family the \hat{K} set contains a single element, and thus the whole M^B family has just one model with a relative forecast error that does not exceed 0.5%, whereas in the M^A family there are seven models with that level of relative forecast model with three further ones in the M^C family. On the basis of data presented in Table 3, the longest – over a

Table 3. The lowest quantity of initial data included during estimation of the remaining parameters of world population models that met conditions (I) and (II) simultaneously $\hat{\tau}_-$ and the related \hat{K} set for M^A , M^B , M^C families of models

Families of models	$\hat{\tau}_-$	\hat{K}
M^A	54	35÷41
M^B	42	74
M^C	54	43÷45

Source: own elaboration.

twenty-year long forecast period - was achieved for the M^B family models, whereas in the case of the M^A and M^C families, the relative error for only a ten year forecast did not exceed 0.5%.

Table 4 includes a list of τ^* and k^* values (see equation (15)) and the corresponding values of maximum relative errors in the domain of parameter estimation (see equation (12)) and in the domain of the forecast (see equations (13)) for the M^A , M^B , M^C families of models. Similar values of the τ^* parameters were attained in the work [Rzymowski, Surowiec 2017] for models of the world's population based on the difference equation.

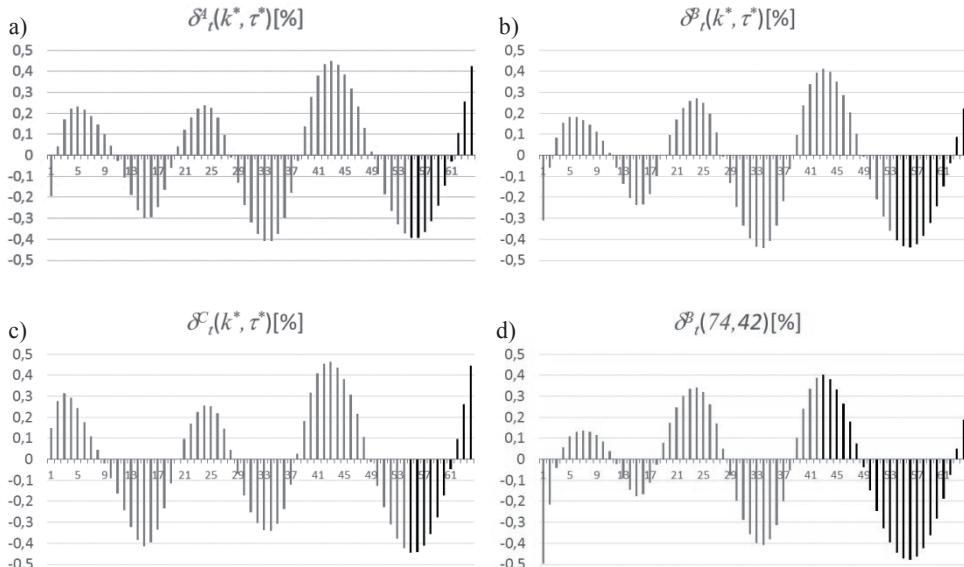
Table 4. Maximum relative errors in parameter estimation and in forecasting that correspond to the models with the lowest relative errors for the M^A , M^B , M^C families of models and values of parameter τ^* and k^*

Families of models	τ^*	k^*	$\delta_L(k^*, \tau^*)$	$\delta_R(k^*, \tau^*)$
M^A	54	35	0.45	0.42
M^B	53	78	0.44	0.44
M^C	54	43	0.46	0.44

$\delta_L(k^*, \tau^*)$ – maximum relative error defined by equation (12),

$\delta_R(k^*, \tau^*)$ – maximum relative error defined by equation (13).

Source: own elaboration.



The gray bar on the graph corresponds to the value of relative error for the estimation period, the black bar for that in the forecast period.

Fig. 1. Graph of relative errors for $t = 1, 2, \dots, 64$ for the models of M^A (a), M^B (b), M^C (c), families with the lowest maximum relative errors and for the M^B family model with the longest forecast period (d)

Source: own elaboration.

In the case of the M^A and M^C families of models, the models with the lowest relative error are also the models with the longest possible forecast in those respective families (see Table 3), the M^B family model is the model with the lowest maximum relative error of all the models analyzed in this work. The relative error for an eleven-year forecast is below 0.45%.

Figure 1 presents the graphs of relative errors (see equation 11) for $t=1,2,\dots,64$ for the models of M^A (Figure 1a), M^B (Figure 1b), M^C (Figure 1c) families with the lowest maximum relative errors. Figure 1d presents a graph of relative errors for the M^B family model with the longest forecast period. When we compare the graphs of Figure 1 it seems that the most stable one is the model with the lowest relative error belonging to the M^C family.

4.2. Prediction

On the basis of the models characterized by their lowest maximum relative error (see equation 3.5) in each of the M^A , M^B , M^C families of models (see Table 4), in this work we performed a prediction of the world's population up to 2050. The resulting values are presented in Table 5 and Figures 2 and 3. Table 5 compares the results with data published on [<http://stats.oecd.org>, 04 Mar 2016] and [<http://wiadomosci.onet.pl/ciekawostki/ludnosc-swiata-infographic/bqqt3y>, 20 Jun 2017] websites for the years 2015 and 2050. The 2050 prediction on the basis of each of the models suggested is lower than both the OECD¹ and Onet forecasts. The 2015 prediction on the basis of each of the suggested models meets the OECD forecast.

Table 5. Comparison of world population (in billions of people) forecasts obtained with the use of the M^A , M^B , M^C families models with the lowest relative errors with OECD and Onet data.

Year/Model	M^A	M^B	M^C	OECD	Onet
2015	7,3	7,3	7,3	7,3	7,2
2050	8,5	9	8,3	9,6	9,5

OECD – forecast from [<http://stats.oecd.org>, 4 March 2016].

Onet – forecast from [<http://wiadomosci.onet.pl/ciekawostki/ludnosc-swiata-infographic/bqqt3y>, 20 June 2017].

Source: own elaboration based on [<http://stats.oecd.org>, 4 March 2016] and [<http://wiadomosci.onet.pl/ciekawostki/ludnosc-swiata-infographic/bqqt3y>, 20 June 2017].

Figure 2 presents the resulting forecasts for 2050 and their comparison with data published by OECD [<http://stats.oecd.org>, 4ch Mar 2016].

The M^C family model, even if it seemed to be the most stable model, results in the most underestimated forecast when we compare it to the values published by the

¹ OECD – Organisation for Economic Co-operation and Development.

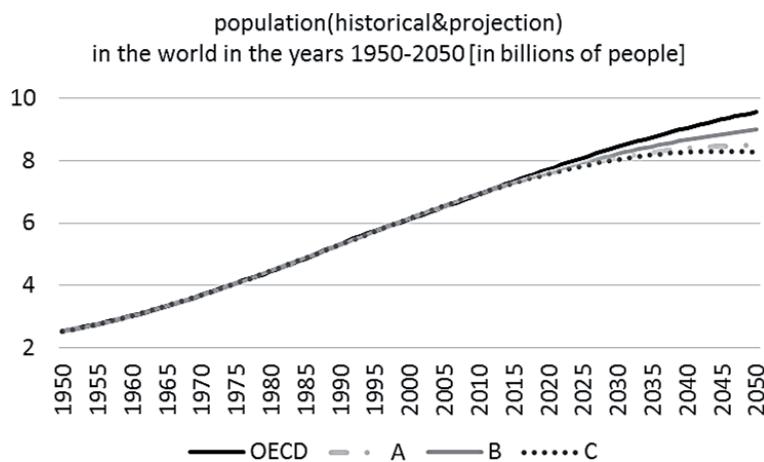


Fig. 2. World population forecasts up to 2050 according to the OECD and the models the lowest maximum relative errors of the M^A , M^B , M^C families

Source: own elaboration.

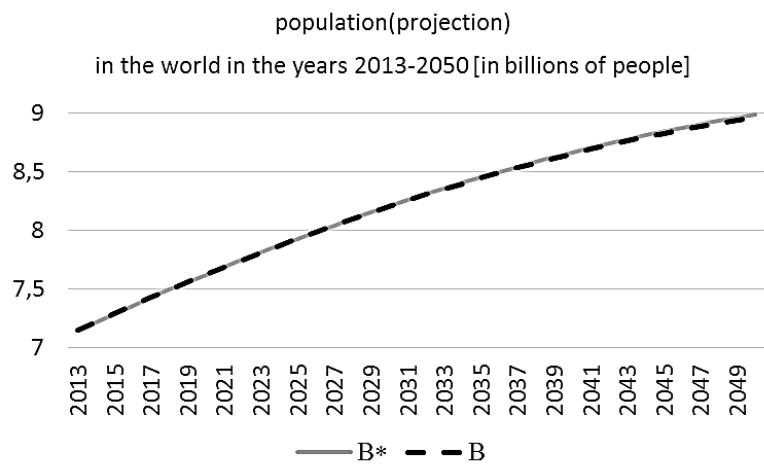


Fig. 3. The world's population forecast from 2013 to 2050 obtained for the models from the M^B family
Source: own elaboration.

OECD. The M^B family model with the lowest possible relative error achieved in this work results in a forecast that is closest to the OECD published data. Furthermore, results of forecasts for the M^B family of models with the lowest relative error (B^*) and longest forecast ($B_$) are comparable, as presented in Figure 3.

5. Final remarks

Despite the statement that population forecasting by fitting mathematical curves is notably unreliable because it ignores so many important factors of demography [Dorn 1962], the models of population obtained by using the method of combining two types of models: the trigonometric one and the logistic one provide a remarkably good fit with nearly all available data. The method presented in this work gives surprisingly good results in modelling the world's population. The models with the maximum relative error not exceeding 0.5% in the period 1950-2013 can be obtained. The maximum relative error in [Rzymowski, Surowiec 2017] does not exceed 0.6 % for the world's population. The result presented in this work is better than the one obtained in [Rzymowski, Surowiec 2017] but the curves that represent the forecasts are very similar. The forecasts obtained on the basis of the models with the lowest maximum relative error for each model family M^A , M^B , M^C analyzed in this work, are reflected to a better degree reflect in these statements: "There must be an upper limit on the earth's life support capabilities, and therefore the population cannot grow without bound" [Robertson et al. 1961] and "A model predicts that the world's population will stop growing in 2050"² than the forecast of the OECD.

Maybe the model from the M^B family with the lowest maximum relative error is the one that will provide the most accurate forecast for the longest period.

Bibliography

- Austin A.L., Brewer J.W., 1971-1972, *World population growth and related technical problems*, Technological Forecasting and Social Change, 3, pp. 23-49.
- Chen K., Guo P., Lin Y., Ying Z., 2010, *Least absolute relative error estimation*, Journal of American Statistical Association, vol. 105, no. 491, pp. 1104-1112.
- Dorn H. F., 1962, *World population growth: an international dilemma*, Science, 135, pp. 283-290.
- Foryś U., Poleszczuk J., 2011, *Modelowanie matematyczne w biologii i medycynie*, Uniwersytet Warszawski, Warszawa, <http://mst.mimuw.edu.pl/wyklady/mbm/wyklad.pdf>.
- Holzer J.Z., 2003, *Demografia*, PWE, Warszawa.
- Hyb W., Kaleta J., 2004, *Porównanie metod wyznaczania współczynników modelu matematycznego na przykładzie prognozy liczby ludności świata*, Przegląd Naukowy Inżynieria i Kształtowanie Środowiska, 2(29), pp. 94-99.
- Malthus T.R., 1798, *An essay on the Principal of Population*. J. Johnson, in St. Paul's Churchyard: London.
- Murray J.D., 1989, *Mathematical Biology 19 of Biomathematics Texts*, Berlin Springer.
- Nowak E., 2006, *Zarys metod ekonometrii. Zbiór zadań*, PWN, Warszawa.
- Pearl R., Reed L.J., 1924, *The growth of human population*, Studies in Human Biology: 584-637, Pearl R. (ed.), William and Wilkins, Baltimore.
- Rao C.R., 1982, *Modele liniowe statystyki matematycznej*, PWN, Warszawa.
- Robertson J.S., Bond V.P., Cronkite E.P., Hutton W.E., Howland W.E., Shinbrot M., von Foerster H., Mora P.M., Amiot L.W., 1961, *Doomsday*, Science, 133, pp. 936-946.

² <http://www.sciencedaily.com/releases/2013/04/130404072923.htm>, 2013.

- Rzymowski W., Surowiec A., 2012, *Method of Parameters estimation of Pseudologistic Model*, [in:] *Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne 2*, Zieliński Z.E. (ed.) WSH, Kielce, pp. 256-265.
- Rzymowski W., Surowiec A., 2017, *Modelling population growth with difference equation method*, *Przegląd Statystyczny*, 64 (3), pp. 339-351.
- Serrin J., 1975, *Is 'Doomsday' on target? (Letter)*, *Science*, 189, pp. 86-88.
- Smith D.A., 1977, *Human Population Growth: Stability or Explosion?* *Mathematics Magazine*, 50 (4), pp. 186-197.
- Smolik S., 1996, *Long-term Projection of Numerical Growth of Population*, II Krajowa Konferencja Zastosowań Matematyki w Biologii i Medycynie.
- Verhulst P.F., 1838, *Notice sur la loi que la population suit dans son d'accroissement*, *Correspondance Mathématique et Physique Publiée par A. Quetelet*, 10, pp. 113-121.
- von Foerster H., Mora P.M., Amiot L.W., 1960, *Doomsday. Friday, 13 November, A.D. 2026*, *Science*, vol. 132, no 3436, pp. 1291-1295.
- <http://stats.oecd.org>, Data extracted on 04 Mar 2016 13:39 UTC (GMT) from OECD.Stat.
- <http://www.sciencedaily.com/releases/2013/04/130404072923.htm>, 4 April 2013. Plataforma SINC. ScienceDaily, *A model predicts that the world's populations will stop growing in 2050*.
- <http://wiadomosci.onet.pl/ciekawostki/ludnosc-swiata-infographic/bqqt3y>.

WYBRANE METODY MODELOWANIA LICZBY LUDNOŚCI ŚWIATA

Streszczenie: W pracy przedstawiono wybrane metody ekonometryczne modelowania liczby ludności świata na podstawie danych historycznych. W zaproponowanych w pracy modelach użyto zmiennych okresowych. Ponadto w modelowaniu wykorzystano funkcję typu logistycznego. Celem pracy było uzyskanie modelu opisującego liczbę ludności świata o jak najmniejszym maksymalnym błędzie względnym i możliwie długim okresie żywotności. Zaproponowana w pracy metoda modelowania liczby ludności świata pozwala uzyskać modele o maksymalnym błędzie względnym nieprzekraczającym 0,5%. Wybrane modele wykorzystano do predykcji liczby ludności świata do roku 2050.

Slowa kluczowe: modele nieliniowe, estymacja parametrów strukturalnych, maksymalny błąd względny modelu, populacja, predykcja.