DETERMINANTS METHOD OF EXPLANATORY VARIABLES SET SELECTION TO LINEAR MODEL

Witold Rzymowski[1], Agnieszka Surowiec[2]

[1] Department of Applied Mathematics, Lublin University of Technology, Nadbystrzycka 38, 20-718 Lublin, Poland, e-mail: w.rzymowski@pollub.pl

[2] Department of Quantitative Methods for Management, Department of Management, Lublin University of Technology, Nadbystrzycka 38, 20-718 Lublin, Poland,, email: a.surowiec@pollub.pl

ABSTRACT

The determinants method of explanatory variables set selection to the linear model is shown in this article. This method is very useful to find such a set of variables which satisfy small relative error of the linear model as well as small relative error of parameters estimation of this model. Knowledge of the values of the parameters of this model is not necessary. An example of the use of the determinants method for world's population model is also shown in this article. This method was tested for $2^{24} - 1$ models for a set of 23 potential explanatory variables. 5 world's population models with one, two, three, four and five explanatory variables were chosen and analysed.

Keywords: linear regression analysis, least square parameter estimation, relative error, Gram matrix.

INTRODUCTION

In this article the linear model of the form:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k + \varepsilon$$

is analyzed, where:
- Y is the dependent variable,
- $X_1, X_2, \ldots, X_k$ are explanatory variables.

In order to initially select a set of explanatory variables for a linear model one should follow the available knowledge, experience and intuition (see [1], §1.1). Then one should take into account (see, for example [2] p. 193) the following recommendations:

A) The number of explanatory variables should not be too large.
B) Selected explanatory variables should be strongly correlated with the dependent variable and weakly correlated between themselves.
C) An econometric model with selected set of explanatory variables should be well matched to the data.

D) The estimation parameters errors should be small.

With the exception of recommendation A, the above-mentioned recommendations are quantitative and there are a lot of methods, see [3], making possible to take into account at least some of these recommendations.

The main measures of the quality of the linear model are: the relative error of the model, the estimation parameters relative errors and the coefficient of determination R^2. The most widespread in the Polish literature and from mathematical point of view is a very elegant Hellwig method [4] which includes only recommendation (B). Besides, using the Hellwig method, we are not able to determine whether the free parameter α_0 is needed in the model or not. Moreover, the value of the coefficient of determination R^2, more specifically the number $100R^2$, is interpreted as a percentage of the dependent variable variability explained by the model. As for models without a free parameter, the inequality $R^2 > 1$ is possible, then the coefficient of deter-

mination is useful only if the model contains a free parameter.

Because small estimation parameters relative errors of the model usually proclaim on the lack of collinear constraints between explanatory variables, then it should be given major consideration to the recommendations made by the points (C) and (D). The error of the model (the standard deviation of the residual component) can be calculated using the determinants of two appropriate matrices, (see [3], p. 102, formula (3.76) and [5]). An interesting fact is that relative errors of estimation parameters can also be set using the determinants of two appropriate matrices and this can be done without knowledge on the values of these parameters, (see formula (14)). The formulae (12) and (14) are the basis of the determinants methods.

However, it should be pointed out that none of the methods guarantees the achievement of the right choice and the empirical verification is only the correct final evaluation of the quality of the model.

For many years the problem to create a demographic model to predict the world population in a fixed time range has been investigated extensively. Population models of living beings tend to have a form:

$$y_t = f(t) + \varepsilon_t, \ t = 1,2,...N$$

The two oldest models: an exponential model of Malthus [6] and logistic of Verhulst [7] are the most well-known models in case of the human population. Both of these models really do not work in longer periods of time (see [8], §1.1). The models described in the works [9, 10] give a better description of the human population. But all of these models are time dependent models which are not considered in this work.

In this work an example of the use of the determinants method for world's population model given by the formula:

$$y_t = \alpha_0 + \alpha_1 x_{1,t} + \alpha_2 x_{2,t} + \ldots$$
$$+ \alpha_{23} x_{23,t} + \varepsilon_t, \ t = 1,2,...40, \qquad (0)$$

where: y_t is the tth observation of the world population and $x_{1,t}, x_{2,t}...., x_{23,t}$ is the tth observation of the populations in chosen countries or group of countries in year $1949 + t$ is done.

Of all the countries and groups of countries, available at [11], one selected only those whose population is greater than fifteen million. Table 1 presents a list of explanatory variables and dependent variable with alist of selected countries in

Table 1. The explanatory variables and dependent variable in world population model

X1	Australia	X13	Turkey
X2	Canada	X14	United Kingdom
X3	Chile	X15	United States
X4	France	X16	European Union (27 countries)
X5	Germany	X17	G7
X6	Italy	X18	Brazil
X7	Japan	X19	China
X8	Korea	X20	India
X9	Mexico	X21	Indonesia
X10	Netherlands	X22	Russian Federation
X11	Poland	X23	South Africa
X12	Spain	Y	World

world population model corresponding to them. The reason for choosing model (0) was the ease of getting a model with a potentially large number of explanatory variables. Note that the total number of set of potential explanatory variables in this case is equal to:

$$2^{24} - 1 = 16777215$$

The relevant calculations take into account data from years 1950–1989 while the effects of selected models was performed in full period of 1950–2012, what can be received as a makeshift for empirical verification.

In view of the large number of verified models it was necessary to write a computer program that performs appropriate calculations.

DETERMINANTS METHOD

Assume that $\{X_1, X_2,...X_k\}$ stands for the set of explanatory variables, where:

$$X_p = \begin{bmatrix} x_{p,1} \\ x_{p,2} \\ \vdots \\ x_{p,N} \end{bmatrix} \in R^N, \quad p = 1,2,...k$$

are linear independent and the condition

$$y_t > 0, \ t = 1,2,...,N$$

is true. Taking, if necessary,

$$x_{1,t} = J_t = 1, \ t = 1,2,...,N$$

we can consider the models with a free parameter. We will name „the variable" J the explanatory variable.

From the set of $\{X_1, X_2, \dots X_k\}$ let us choose a subset Z. In order to simplify and shorten the notations we assume that $Z = \{X_1, X_2, \dots, X_q\}$, where $1 \le q \le k$.

Then we consider the model:

$$y_t = \alpha_1 x_{1,t} + \alpha_2 x_{2,t} + \dots + \alpha_q x_{q,t} + \varepsilon_t,$$
$$t = 1, 2, \dots, N, \qquad (1)$$

whose parameters will be estimated by using Least Squares Method.

We will give the way to calculate the relative error $\mathrm{rel}S(Z, Y)$ of model (1) and relative errors $\mathrm{rel}S(Z, Y, \alpha_p)$ of its parameters estimation α_p, $p = 1, 2, \dots, q$, without determining the values of these parameters. For this purpose we will use the minors of the corresponding Gram matrix.

Least Squares Method – indications and standard formulae

Let:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}.$$

Of course:

$$Y, \varepsilon, X_p \in R^N, \quad p = 1, 2, \dots, q.$$

In these indications model (1) is such that:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_q X_q + \varepsilon.$$

For each:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix} \in R^q$$

we define:

$$Y(\beta) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_q X_q = \sum_{p=1}^{q} \beta_p X_p.$$

Note that:

$$Y(\beta) = \begin{bmatrix} y_1(\beta) \\ y_2(\beta) \\ \vdots \\ y_N(\beta) \end{bmatrix} = \beta_1 \begin{bmatrix} x_{1,1} \\ x_{1,2} \\ \vdots \\ x_{1,N} \end{bmatrix} + \beta_2 \begin{bmatrix} x_{2,1} \\ x_{2,2} \\ \vdots \\ x_{2,N} \end{bmatrix}_2 + \dots$$

$$\dots + \beta_q \begin{bmatrix} x_{q,1} \\ x_{q,2} \\ \vdots \\ x_{q,N} \end{bmatrix} = \mathbf{X}\beta, \qquad (2)$$

where:

$$\mathbf{X} \overset{\text{def}}{=} \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{q,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{q,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,N} & x_{2,N} & \cdots & x_{q,N} \end{bmatrix}$$

is a matrix, in which the columns are created by the coordinates of vectors $X_1, X_2, \dots, X_q$.

The number:

$$\left\| Y - Y(\beta) \right\|^2 = \left\| Y - \sum_{p=1}^{q} \beta_p X_p \right\|^2 \overset{\text{def}}{=} \sum_{t=1}^{N} \left(y_t - \sum_{p=1}^{q} \beta_p x_{p,t} \right)^2$$

is a square of the length of the vector $Y - Y(\beta) \in R^N$. Estimating the parameters of model (1) using the Least Squares Method we find a vector:

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_q \end{bmatrix} \in R^q,$$

such that:

$$\left\| Y - Y(\alpha) \right\|^2 = \min_{\beta \in R^q} \left\| Y - Y(\beta) \right\|^2. \qquad (3)$$

Since:

$$\bigcup_{\beta \in R^q} Y(\beta) \subset R^N$$

is a linear subspace of space R^N, then exists a vector $\alpha \in R^q$ that satisfies the condition (3) and condition

$$\langle Y - Y(\alpha), Y(\beta) \rangle = 0, \quad \beta \in R^q, \qquad (4)$$

where:

$$\langle Y - Y(\alpha), Y(\beta) \rangle \overset{\text{def}}{=} \sum_{t=1}^{N} (y_t - y_t(\alpha)) y_t(\beta)$$

is Euclidean scalar product of vectors $Y - Y(\alpha)$, $Y(\beta) \in R^N$. Taking into account the formulae (2) and (4) we can obtain the condition:

$$\langle \mathbf{X}^T Y - \mathbf{X}^T \mathbf{X}\alpha, \beta \rangle = 0, \quad \beta \in R^q,$$

which shows that the equation:

$$\mathbf{X}^T \mathbf{X}\alpha = \mathbf{X}^T Y \qquad (5)$$

has a solution $\alpha \in R^q$. Since $X_1, X_2, \dots, X_q$ are linearly independent, the following formula holds

true $\det\!\left(\mathbf{X}^T\mathbf{X}\right) > 0$ and consequently:

$$\alpha = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T Y \qquad (6)$$

is the only solution of the equation (5), therefore, it is the only solution of the problem (3).

At the end of this chapter we write down the well-known formulae for relative error $\mathrm{rel}S(Z,Y)$ of model (1) and relative errors $\mathrm{rel}S(Z,Y,\alpha_p)$ of its parameters estimation α_p, $p = 1,2,...,q$.

In order to shorten the notations certain symbols are introduced. Put:

$$\sum Y = \sum_{t=1}^{N} y_t \; ,$$

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{q1} & b_{q2} & \cdots & b_{qq} \end{bmatrix}. \qquad (7)$$

If $\alpha \in R^q$ is done by formula (6), therefore,

$$S(Z,Y) = \sqrt{\frac{1}{N-q}\sum_{t=1}^{N}\left(y_t - \sum_{p=1}^{q}\alpha_p x_{p,t}\right)^2} = \frac{\|Y - \mathbf{X}\alpha\|}{\sqrt{N-q}} \qquad (8)$$

$$\mathrm{rel}S(Z,Y) = \frac{100N}{\sum Y}S(Z,Y) = \frac{100N}{\sum Y}\frac{\|Y - \mathbf{X}\alpha\|}{\sqrt{N-q}}, \qquad (9)$$

$$\mathrm{rel}S(Z,Y,\alpha_p) = \frac{100}{|\alpha_p|}S(Z,Y)\sqrt{b_{pp}} = \frac{100}{|\alpha_p|}\frac{\|Y - \mathbf{X}\alpha\|}{\sqrt{N-q}}\sqrt{b_{pp}} \qquad (10)$$

Determinants method. Gram matrix $\mathbf{G}(Z,Y)$

In this chapter the corresponding formulae without the values of parameters α_p for formulae (9) and (10) will be presented.

Let:

$$\mathbf{G}(Z) = \mathbf{X}^T\mathbf{X} = \begin{bmatrix} \sum X_1^2 & \sum X_1 X_2 & \cdots & \sum X_1 X_q \\ \sum X_2 X_1 & \sum X_2^2 & \cdots & \sum X_2 X_q \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_q X_1 & \sum X_q X_2 & \cdots & \sum X_q^2 \end{bmatrix}$$

and

$$\mathbf{G}(Z,Y) = \begin{bmatrix} \sum X_1^2 & \sum X_1 X_2 & \cdots & \sum X_1 X_q & \sum X_1 Y \\ \sum X_2 X_1 & \sum X_2^2 & \cdots & \sum X_2 X_q & \sum X_2 Y \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum X_q X_1 & \sum X_q X_2 & \cdots & \sum X_q^2 & \sum X_q Y \\ \sum X_1 Y & \sum X_2 Y & \cdots & \sum X_q Y & \sum Y^2 \end{bmatrix}$$

represent Gram matrices, where:

$$\sum X_p X_{p'} = \sum_{t=1}^{N} x_{p,t} x_{p',t}, \; p, p' \in \{1,2,...,q\},$$

$$\sum X_p Y = \sum_{t=1}^{N} x_{p,t} y_t, \; p \in \{1,2,...,q\}.$$

For every $p = 1,2,...,q$, the notation $\mathbf{G}_p(Z)$ represents the matrix that was created from matrix $\mathbf{G}(Z)$ by removing the pth row and the pth column, the notation $\mathbf{G}_p(Z,Y)$ represents the matrix that was created from the matrix $\mathbf{G}(Z,Y)$ by removing the last $(q+1)^{\text{th}}$ row and the pth column, the notation $\hat{\mathbf{G}}_p(Z,Y)$ represents the matrix that was created from the matrix $\mathbf{G}(Z)$ by replacement the pth column:

$$\begin{bmatrix} \sum X_1 X_p \\ \sum X_2 X_p \\ \vdots \\ \sum X_q X_p \end{bmatrix}$$

by the column:

$$\begin{bmatrix} \sum X_1 Y \\ \sum X_2 Y \\ \vdots \\ \sum X_q Y \end{bmatrix}$$

Remark 1. In case of $q = 1$ (a model with one explanatory variable X) we have:

$$y_t = \alpha x_t + \varepsilon_t, \; t = 1,2,...,N \; ,$$

$$\mathbf{G}(Z) = \left[\sum_{t=1}^{N} x_t^2\right],$$

$$\mathbf{G}(Z,Y) = \begin{bmatrix} \sum_{t=1}^{N} x_t^2 & \sum_{t=1}^{N} x_t y_t \\ \sum_{t=1}^{N} y_t x_t & \sum_{t=1}^{N} y_t^2 \end{bmatrix},$$

$$\mathbf{G}_1(Z,Y) = \hat{\mathbf{G}}_1(Z,Y) = \left[\sum_{t=1}^{N} x_t y_t\right]$$

and we assume $\mathbf{G}_1(Z) \overset{\text{def}}{=} [1]$.

The model errors

Let us define the sets:

$$\mathsf{R}(Z) = \left\{\sum_{p=1}^{q}\beta_p X_p : \; 0 \le \beta_p \le 1, \; p = 1,2,...,q\right\}$$

$$\mathsf{R}(Z,Y) = \left\{\sum_{p=1}^{q}\beta_p X_p + \beta_{q+1}Y : \; 0 \le \beta_p \le 1, \; p = 1,2,...,q+1\right\}$$

Number $\sqrt{\det \mathbf{G}(Z)}$ is the q-dimensional volume (q-dimensional Hausdorff measure) of the parallelepiped $\mathsf{R}(Z)$, and number $\sqrt{\det \mathbf{G}(Z,Y)}$ is the $q+1$-dimensional volume ($q+1$-dimensional Hausdorff measure) of the parallelepiped $\mathsf{R}(Z,Y)$. Besides the parallelepiped $\mathsf{R}(Z)$ is the base of parallelepiped $\mathsf{R}(Z,Y)$, and number $\|Y - \mathbf{X}\alpha\|$ is the height of parallelepiped $\mathsf{R}(Z,Y)$. Therefore, we have:

$$\sqrt{\det \mathbf{G}(Z,Y)} = \|Y - \mathbf{X}\alpha\| \sqrt{\det \mathbf{G}(Z)} .$$

Taking into consideration the formulae (8) and (9) we get:

$$S(Z,Y) = \sqrt{\frac{\det \mathbf{G}(Z,Y)}{(N-q)\det \mathbf{G}(Z)}} , \quad (11)$$

$$\mathrm{rel}S(Z,Y) = \frac{100N}{\sum Y} \sqrt{\frac{\det \mathbf{G}(Z,Y)}{(N-q)\det \mathbf{G}(Z)}} . \quad (12)$$

Parameters estimation and estimation errors

In view of the formula (6):

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_q \end{bmatrix} = \alpha = (\mathbf{G}(Z))^{-1} \mathbf{X}^T Y$$

we obtain:

$$\alpha_p = \frac{\det \hat{\mathbf{G}}_p(Z,Y)}{\det \mathbf{G}(Z)} = (-1)^{p-1} \frac{\det \mathbf{G}_p(Z,Y)}{\det \mathbf{G}(Z)},$$

$$p = 1,2,...,q. \quad (13)$$

Remark 2. In case of model of the form $y_t = \alpha x_t$, $t = 1,2,...,N$, (see Remark 1) we get:

$$\alpha = (-1)^{1-1} \frac{\det \mathbf{G}_1(Z,Y)}{\det \mathbf{G}(Z)} = \frac{\det \sum XY}{\det \sum X^2} = \frac{\sum XY}{\sum X^2} .$$

Forasmuch (see (7)):

$$b_p = \frac{\det \mathbf{G}_p(Z)}{\det \mathbf{G}(Z)}, \quad p = 1,2,...,q$$

then (see (11)):

$$S(Z,Y,\alpha_p) = S(Z,Y)\sqrt{b_p} =$$

$$= \sqrt{\frac{\det \mathbf{G}(Z,Y)\det \mathbf{G}_p(Z)}{(N-q)\det^2 \mathbf{G}(Z)}} .$$

Now, taking into account the formulae (10) and (13), for every $p = 1,2,...,q$, we get

$$\mathrm{rel}S(Z,Y,\alpha_p) = 100\sqrt{\frac{\det \mathbf{G}(Z,Y)}{(N-q)\det \mathbf{G}_p(Z,Y)}} . \quad (14)$$

WORLD POPULATION MODELS

Initial set of explanatory variables

The potential explanatory variables X_p for the model of world population Y can be the numbers of people in chosen countries or in groups of countries. We chose X_p presented in Table 1. Therefore the considered model is of the form (1) and the relative errors $\mathrm{rel}S(Z,Y)$ and $\mathrm{rel}S(Z,Y,\alpha_p)$ for all non-empty sets:

$$Z \subset \{J, X_1, X_2,..., X_{23}\}$$

were determined on the basis of the data from the years 1950–1989.

The next stage, the verification, was performed for only those models with one, two, three, four and five explanatory variables out of these presented in Table 1, for which none of the relative errors of parameters estimation exceeds 5%. The set of such designated models are divided into two groups: the first - the models without the free parameter and the second – the models with the free parameter. Then, in each of these parts 5 models with the smallest errors $\mathrm{rel}S(Z,Y)$ with one, two, three, four and five explanatory variables were chosen and analysed. Finally we analyse 25 models without a free parameter and 16 models with a free parameter. (In the second group there is one model with four explanatory variables and there is no model with five explanatory variables which satisfy the criterion 5% error).

For each selected model we calculate the theoretical values of world population $\hat{y}_t$, $t = 1,2,...,63$ (a full range of data from the years 1950–2012), the maximum relative errors:

$$\mathrm{rel}\delta_{40}(Z,Y) = \max_{t=1,2,...40} \frac{|y_t - \hat{y}_t|}{y_t} 100\% \text{ and}$$

$$\mathrm{rel}\delta_{63}(Z,Y) = \max_{t=1,2,...63} \frac{|y_t - \hat{y}_t|}{y_t} 100\%$$

Tables 2 and 3 show the obtained results. Table 2 presents the results for models without the free parameters, the Table 3 presents the similar results for models with the free parameter. The numbers of explanatory variables and the

Table 2. The explanatory variables and the relative errors in the world population linear model without the free parameter. The cell shaded in each errors' column corresponds to the smallest error

Number of variables in model	No	Explanatory variables	$relS_{40}$ (Z, Y)	$rel\delta_{40}$ (Z, Y)	$rel\delta_{63}$ (Z, Y)
1	1	19	1.439	2.738	11.309
	2	3	1.974	4.583	4.583
	3	1	2.405	4.059	7.095
	4	18	6.233	18.362	18.362
	5	23	7.083	14.623	17.343
2	1	17;23	0.288	0.596	1.757
	2	5;23	0.292	0.486	3.070
	3	4;23	0.302	0.7387	1.432
	4	16;23	0.35	0.6024	1.804
	5	9;12	0.355	0.6416	4.846
3	1	1;10;20	0.149	0.3995	2.414
	2	9;10;20	0.16	0.4138	0.930
	3	1;6;20	0.168	0.521	2.846
	4	1;16;20	0.172	0.4985	2.852
	5	1;4;20	0.173	0.4587	2.897
4	1	3;15;19;20	0.042	0.1148	0.893
	2	3;19;20;22	0.042	0.1428	2.370
	3	3;17;19;20	0.048	0.1531	0.891
	4	7;15;19;20	0.062	0.1615	1.231
	5	2;4;19;20	0.068	0.1482	**0.518**
5	1	7;8;19;21;22	**0.037**	0.1037	5.650
	2	8;14;19;20;21	0.044	**0.0696**	1.352
	3	7;14;17;19;20	0.06	0.2188	1.385
	4	5;7;17;19;20	0.064	0.2281	0.907
	5	3;8;11;19;21	0.069	0.1218	2.537

Table 3. The numbers of explanatory variables and the relative errors in the world population linear model with the free parameter. The cell shaded in each errors' column corresponds to the smallest error

| Number of variables in model	No	Explanatory variables	$relS_{40}$ (Z, Y)	$rel\delta_{40}$ (Z, Y)	$rel\delta_{63}$ (Z, Y)
1	1	0;13	0.681	1.774	1.860
	2	0;9	0.686	2.095	4.167
	3	0;21	0.717	3.083	3.083
	4	0;23	0.885	2.177	3.564
	5	0;18	1.025	2.455	4.621
2	1	0;1;20	0.202	0.558	3.187
	2	0;1;23	0.210	0.517	1.760
	3	0;16;20	0.263	0.475	1.421
	4	0;18;20	0.277	0.579	1.072
	5	0;3;20	0.305	0.650	2.986
3	1	0;3;19;20	0.086	0.286	1.034
	2	0;2;19;23	0.103	0.202	1.934
	3	0;2;19;20	0.110	0.276	**0.702**
	4	0;15;19;23	0.116	0.223	1.551
	5	0;3;12;16	0.267	1.157	6.133
4	1	0;11;14;19;20	**0.039**	**0.113**	1.622
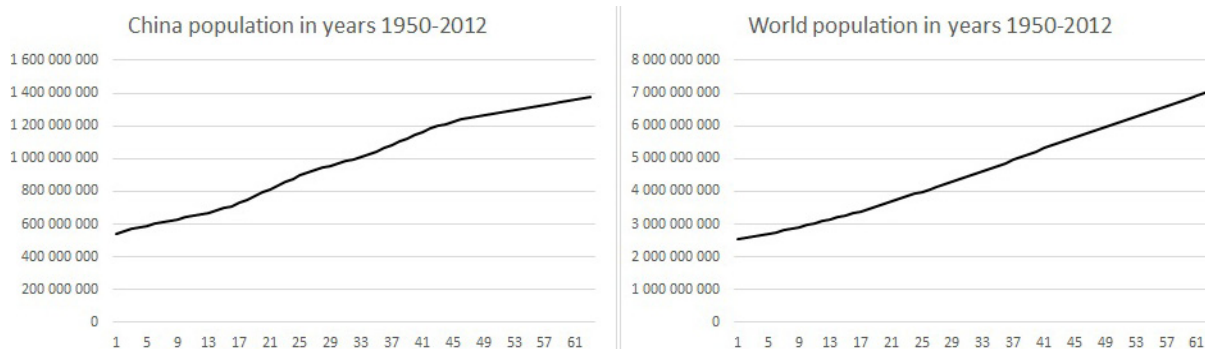
Fig. 1. Comparison the China population and the world population on the basis of the data
from the years 1950–2012 [11]

relative errors for each world population linear model are presented.

FINAL REMARKS

Remarks about the world population model

The maximum relative errors of all analysed models with a free parameter and one explanatory variable do not exceed 5% (see Table 3). Among the selected models with one explanatory variable, but without free parameter there is only one such model, while the maximum relative errors of each of all selected models with two explanatory variables and free parameter also does not exceed 5%.

The models with three parameters show the greatest stability. For models without a free parameter we have $\text{rel}\delta_{40} \leq 0.521$, $\text{rel}\delta_{63} \leq 2.887$ and for models with a free parameter we have $\text{rel}\delta_{40} \leq 0.65$, $\text{rel}\delta_{40} \leq 3.187$.

In addition, we found that among the listed models with one explanatory variable there are such whose maximum relative error in the full range of observation ($t = 1, 2,..., 63$) is equal to the maximum relative error in the range of adjustment ($t = 1, 2,..., 40$). These are the models with variable corresponding to the number of population in Chile and in Brazil (see Table 2) and in Indonesia (see Table 3).

It was found that in the all possible set of models ($2^{24} - 1$) there is not a model with six or more parameters and theirs estimation relative errors not exceeding 5%.

The model without the free parameter that has the smallest error $\text{rel}\delta_{63} \approx 0.5\%$, see Table 2, contains variables that represent the number of peo-

ple in Canada, France, China and India. In turn, the model with the free parameter that has the smallest error $\text{rel}\delta_{63} \approx 0.7\%$, see Table 3, contains variables that represent the number of people in the same countries, except France. The variables X_{19} and X_{20} that represent the number of people in China and India, are most common in both Tables 2 and 3. In addition, both of these variables in both Tables 2 and 3, are present in two models with the smallest errors $\text{rel}\delta_{63}$. It's no surprising such a result, because the total number of the population of these countries is close to 37% of the world's population. On the other hand, a comparison of the plots for China and world of population (see Fig. 1) makes this less obvious, at least in relation to variable X_{19}.

The occurrence rarity in Tables 2 and 3 of variables X_{16} and X_{17} that represent the number of people in EU and in G7 is thought-provoking.

Not every model that is selected in accordance with the recommendations of the theoretical procedure of the variables selection must be a good description of the studied phenomenon. An example is a model with variables $\{X_7, X_8, X_{19}, X_{21}, X_{22}\}$, presented in Table 2.

Remarks about the determinats method

Determinants method provides an easy programmable way to choose the explanatory variables to any linear model even if the potential number of explanatory variables is very large.

The determinants method takes into account the quality recommendation A.

In the determinants method procedure the models which satisfy the theoretical recommendations B to D can be easily found.

REFERENCES

1. Chow G.C., Econometrics. Mc-Graw-Hill Book Company, New York 1983.

2. Zeliaś A., Pawełek B., Wanat S., Prognozowanie ekonomiczne. Teoria. Przykłady. Zadania. PWN, Warszawa 2003.

3. Grabiński T., Wydymus S., Zeliaś A., Metody doboru zmiennych w modelach ekonometrycznych. PWN, Warszawa 1982.

4. Hellwig Z., Problem optymalnego wyboru predykant. Przegląd Statystyczny, nr 3-4, 1969.

5. Kolupa M., O pewnym sposobie obliczania współczynnika zbieżności. Przegląd Statystyczny, nr. 1, 1975.

6. Malthus T.R., An essay on the Principal of Population. Penguin Books, 1970. Originally published in 1798.

7. Verhulst P.F., Notice sur la loi que la population suit dans son d'accroissement. Corr. Math. et Phys., vol. 10, 1838, 113–121.

8. Murray J.D., Mathematical Biology. I: An Introduction, Springer-Verlag, Berlin Heidelberg, 1989. (wyd. polskie: Murray J.D., Wprowadzenie do biomatematyki. PWN, Warszawa 2006).

9. Hyb W., Kaleta J., Porównanie metod wyznaczania współczynników modelu matematycznego na przykładzie prognozy liczby ludności świata. Przegląd naukowy Inżynieria i Kształtowanie środowiska, vol. 2(29), 2004, 94–99.

10. Rzymowski W., Surowiec A., Method of Parameters estimation of Pseudologistic Model. [In:] Zieliński Z.E. (ed.), Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne. vol. 2, Kielce 2012, 256–265.

11. http://stats.oecd.org