Miscellanea

CENTRAL EUROPEAN REVIEW
OF ECONOMICS & FINANCE
Vol. 9, No. 3(2015), pp. 69-80

Zbigniew Śleszyński[1]

USING AKAIKE'S CRITERION AND BORDERED MATRICES FOR INITIAL VARIABLE SELECTION FOR ECONOMETRIC MODEL EVALUATED WITH LEAST SQUARES METHOD

Abstract

In this paper the usage of bordered matrices for selection of independent variables for econometric model evaluated with least squares method is shown. AIC was chosen as criterion of selection. Practical example of the method is also presented.

JEL Classification Code: **C010**, **C120**.

Keywords: Akaike's Information Criterion, least squares method, bordered matrices.

Introduction

Initial selection of explanatory variables for econometric model is really important issue. It often determines the final quality of the model. Therefore, it is essential to use proper variables. At the beginning researcher usually only know the response variable Y, which he tries to explain. Economic theory and intuition suggest what factors may affect the response variable. This way we obtain a set of potential explanatory variables. Often this set is too big and not all of the variables included should be put into the model – thus selection of chosen variables is required. This problem can be solved using numerous methods of variables selection for econometric models. Literature in this area is very extensive, ranging from the classic book by Draper Smith (1973) or Metody doboru zmiennych w modelach ekonometrycznych (Grabiński, Wydymus and Zeliaś 1982). Different model selection criterions are also

[1] Assistant Professor, Ph.D., K. Pulaski University of Technology and Humanities in Radom, Poland, Faculty of Economics, Department of International Business & Finance.

mentioned by Charemza and Deadman (1997). Big popularity in Poland has Helwig's method (Hellwig 1969).

In recent years there is a great interest in the selection of variables for econometric model based on Akaike's Information Criterion as well as comparing it with other criterions – Piłatowska (2011), Rosienkiewicz (2012). This paper focus on the practical aspects of applying and calculating AIC.

Basic considerations

At the beginning let's describe main formulas associated with AIC.

Let's consider a model:

$$Y = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \ldots + \alpha_k Z_k + \xi \tag{1}$$

Parameters are evaluated basing on n observations. Likelihood function for estimated model looks as follows:

$$L = \prod_{t=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{\left(y_t - \sum_{j=0}^{k} \alpha_j z_{jt}\right)^2}{2\sigma^2}} \tag{2}$$

Thus loglikelihood (see Dorosiewicz and others 1996):

$$\ln L = n \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{n} \left(y_t - \sum_{j=0}^{k} \alpha_j z_{jt}\right)^2 \tag{3}$$

For the model estimated with least squares method log likelihood function is equal to:

$$\ln \widehat{L} = -\frac{n}{2}\left(\ln(2\pi) + \ln\left(\frac{u^T u}{n}\right) + 1\right) \tag{4}$$

where u is model (1) residulas column vector. Equality (4) can be transformed into:

$$\ln \widehat{L} = -\frac{n}{2}\left(\ln(2\pi e) + \ln\left(\frac{u^T u}{n}\right)\right) \tag{5}$$

or

$$\ln \widehat{L} = -\frac{n}{2}\left(\ln\left(\frac{2\pi e}{n}\right) + \ln\left(u^T u\right)\right) \tag{6}$$

Let's note that one of measures of quality of the model is the convergence factor φ^2, which is calculated according to the formula:

$$\varphi^2 = \frac{\sum_{t=1}^{n}(y_t - y_t^*)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = \frac{u^T u}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{7}$$

Using (7), basing on (6) we obtain:

$$\ln \hat{L} = -\frac{n}{2}\left(\ln\left(\frac{2\pi e}{n}\right) + \ln\left(\varphi^2 \sum_{t=1}^{n}(y_t - \bar{y})^2\right)\right) \tag{8}$$

Relationship (8) we can rewrite as:

$$\ln \hat{L} = -\frac{n}{2}\left(\ln\left(2\pi e S_y^2\right) + \ln\left(\varphi^2\right)\right) \tag{9}$$

where:

$$S_y^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2 \tag{10}$$

is the variance of dependent variable Y empirical values.

Formula (9) describes log likelihood function maximum for the model estimated with least squares method.

Akaike's Information Criterion is defined by (Maddala 2008):

$$AIC = -2\ln \hat{L} + 2(k+1) \tag{11}$$

So for the model estimated with least squares method, basing on (9) and (11) we have:

$$AIC = n\left[\ln\left(2\pi e S_y^2\right) + \ln\left(\varphi^2\right)\right] + 2(k+1) \tag{12}$$

It is worth noting that AIC formulas described in the literature are a little different than (12) (for example Maddala 2008, Górecki 2010, Rosienkiewicz 2012). In most of cases the difference is caused by skipping the first sum component in equation (6) in the brackets.

It is worth adding that calculating AIC according to formula (12) leads to a result identical with that obtained using Gretl and in the case of other models it usually doesn't. Let's notice that if we consider different models with the same number of variables then according to AIC basing on (12), we choose the model with the lowest convergence coefficient φ^2.

Comparing different models for dependent variable Y, we choose the one with lowest AIC as the best. Therefore with the dependent variable Y and k potential explanatory variables to select the optimum set of variables, according to AIC, one would just estimate 2^k-1 indicators (similarly as in the case of Hellwig's method).

Let's rewrite formula (12), broken down into three components:

$$AIC = n \cdot \ln\left(2\pi e S_y^2\right) + n \cdot \ln\left(\varphi^2\right) + 2(k+1) \tag{13}$$

Let's note that if we have the dependent variable Y, k potential explanatory variables and the empirical data consisting of n observations – then by choosing variables for the model in accordance with the criterion of minimizing AIC given by (13), the first component of the sum occurring in the formula is constant. For various combinations of variables only two other components will change. To compare different subsets of variables - in practice we have to evaluate only convergence coefficients φ^2 for models with this subset used as exploratory variables.

Proposition of variable selection method using AIC

Before algorithm for choosing a satisfactory set of explanatory variables based on AIC will be shown, we will focus on finding the coefficient φ^2 occurring in the formula (13).

Let's denote by $(R(k)\ R_0(k))$ correlation couple describing model (1). As we know (Kolupa, Śleszyński 2010):

$$\varphi^2 = 1 - R_0^T R^{-1} R_0 \tag{14}$$

To compute φ^2 we can use bordered matrix P:

$$P = \left[\begin{array}{c|c} R & R_0 \\ \hline R_0^T & 1 \end{array}\right] \tag{15}$$

If on the matrix P given by (15) we perform elementary transformations which cause that:

– in place of matrix R we we obtain upper triangular matrix with diagonal elements equal to one R^*,
– in place of vector R_0^T we obtain zero vector,

then in place of number 1 we obtain φ^2. As a result matrix P will be transformed into P^*:

$$P^* = \left[\begin{array}{c|c} R^* & R_0{}^* \\ \hline 0,\ldots,0 & \varphi^2 \end{array}\right] \tag{16}$$

For model (1) matrix *P* given by (15) has degree (k+1). Let's notice that in practice, by transforming it to (16) we take care about each of *k* columns of correlation matrix *R*. Thus at the beginning we bring to zero all first columns elements below matrix R diagonal and the first component of vector R_0.

Then in the bottom right corner of the transformed matrix we will receive a number, equal to φ^2 of the model in which the only explanatory variable is the first variable of the model (1) – Z_1. According to formula (13) by substituting *k* with 1 we can calculate AIC for this model. If now we make similar operations on the second column of the matrix P then in the lower right corner of the transformed matrix we will receive a number, equal to the ratio φ^2 for the model in which the explanatory variables are the first and the second variable from model (1) – Z_1 and Z_2. According to formula (13) by substituting k this time with 2 we can calculate the value of the AIC for this model. Continuing the procedure until transforming the matrix *P* into P^* (16), as if „by the way" we determine the value of AIC for models in which sets of variables are respectively: $\{Z_1\}$, $\{Z_1, Z_2\}$, $\{Z_1, Z_2, Z_3\}$, …, $\{Z_1, Z_2, Z_3, …, Z_k\}$.

In practice, to end up with sensible set of explanatory variables, according to AIC and described method of proceeding, it is recommended to apply following rules:

1) explanatory variables in model (1) pre-sort by not growing absolute values of these variables correlation with the dependent variable Y,
2) if as the result of computations for s transformed matrix P column AIC value is bigger than the previous one – then in the (s-1) transformed matrix we remove row and column s and continue with the process. In practice it means that ZS won't be in the model
3) the process is finished when all the columns of matrix P are transformed. It is worth noting that proposed method doesn't guarantee that resulting variables combination will have minimum AIC, but it should be close to minimum.

In the next part of the paper numerical example of described process will be shown.

Example

Let's consider dependent variable Y and set of potential explanatory variables

$$A(6)=\{Z_1,Z_2,Z_3,Z_4,Z_5,Z_6\} \tag{17}$$

where:

Y – GDP in Poland at current prices in millions of $,
Z_1 – consumption at current prices in millions of $,
Z_2 – export at current prices in millions of $,
Z_3 – year,
Z_4 – inflation rate in %,
Z_5 – average annual exchange rate of US dollar in zł,
Z_6 – population in Poland in millions.

In the model empirical data from statistical yearbooks of Central Statistical Office from years 1995 – 2014 will be used, it means number of observations n=20. Empirical data is shown in table 1.

Table 1. Values of model variables

Y	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
142 172,50	111 006,43	32 643,54	1 995	27,80	2,42	38 284,00
159 919,15	127 265,34	35 333,58	1 996	19,90	2,70	38 294,00
159 045,05	127 320,17	37 151,91	1 997	14,90	3,28	38 290,00
173 474,83	137 938,29	45 072,27	1 998	11,80	3,49	38 277,00
169 700,57	136 284,56	40 939,38	1 999	7,30	3,97	38 263,00
171 873,73	140 422,42	46 799,42	2 000	10,10	4,35	38 254,00
190 521,26	158 242,26	51 878,65	2 001	5,50	4,09	38 242,20
198 704,99	168 824,86	57 143,77	2 002	1,90	4,08	38 218,50
217 524,24	182 335,11	72 634,16	2 003	0,80	3,89	38 190,60
253 778,33	208 556,38	87 912,97	2 004	3,50	3,65	38 173,80
304 476,01	247 576,67	106 336,40	2 005	2,10	3,23	38 157,10
343 338,92	275 939,73	131 156,81	2 006	1,00	3,10	38 125,50
428 948,93	336 738,71	166 535,95	2 007	2,50	2,77	38 115,60
530 185,12	427 117,30	203 158,31	2 008	4,20	2,41	38 135,90
437 022,66	351 997,30	164 216,35	2 009	3,50	3,12	38 167,30
476 624,66	385 195,48	192 890,54	2 010	2,60	3,02	38 529,90
524 256,60	416 947,43	226 142,61	2 011	4,30	2,96	38 538,40
496 129,57	394 852,63	223 509,67	2 012	3,70	3,26	38 533,30
526 030,75	415 671,03	242 639,21	2 013	0,90	3,16	38 495,70
547 899,15	429 004,63	257 078,54	2 014	0,00	3,16	38 484,00

Source: statistical yearbooks of Central Statistical Office and own calculations.

Basing on data from table 1 we evaluate correlation pair ($R(6)$ $R_0(6)$) describing the model, where Y is dependent variable and $A(6)$ given by (17) are explanatory variables.

$$R = \begin{bmatrix} 1 & 0{,}98944 & 0{,}95704 & -0{,}61549 & -0{,}45475 & 0{,}44239 \\ 0{,}98944 & 1 & 0{,}96481 & -0{,}59626 & -0{,}42943 & 0{,}52192 \\ 0{,}95704 & 0{,}96481 & 1 & -0{,}76070 & -0{,}22645 & 0{,}45566 \\ -0{,}61549 & -0{,}59626 & -0{,}76070 & 1 & -0{,}27107 & -0{,}03352 \\ -0{,}45475 & -0{,}42943 & -0{,}22645 & -0{,}27107 & 1 & -0{,}11313 \\ 0{,}44239 & 0{,}52192 & 0{,}45566 & -0{,}03352 & -0{,}11313 & 1 \end{bmatrix}$$

$$R_0(6)=\begin{bmatrix}0{,}99937\\0{,}99198\\0{,}95359\\-0{,}59935\\-0{,}46778\\0{,}45228\end{bmatrix}$$

It is worth noting that the potential variables of the model have been ordered so that correlation vector components $R_0(6)$ meet recommended first condition (are ordered by decreasing absolute value of the correlation coefficient with the dependent variable Y). We can then build bordered matrix P given by formula (15).

$$P=\left[\begin{array}{cccccc|c}1 & 0{,}98944 & 0{,}95704 & -0{,}61549 & -0{,}45475 & 0{,}44239 & 0{,}9993681\\0{,}98944 & 1 & 0{,}96481 & -0{,}59626 & -0{,}42943 & 0{,}52192 & 0{,}991981\\0{,}95704 & 0{,}96481 & 1 & -0{,}76070 & -0{,}22645 & 0{,}45566 & 0{,}953594\\-0{,}61549 & -0{,}59626 & -0{,}76070 & 1 & -0{,}27107 & -0{,}03352 & -0{,}599355\\-0{,}45475 & -0{,}42943 & -0{,}22645 & -0{,}27107 & 1 & -0{,}11313 & -0{,}46778\\0{,}44239 & 0{,}52192 & 0{,}45566 & -0{,}03352 & -0{,}11313 & 1 & 0{,}45228\\\hline 0{,}9993681 & 0{,}991981 & 0{,}953594 & -0{,}599355 & -0{,}46778 & 0{,}45228 & 1\end{array}\right] \qquad (18)$$

To compute φ^2 for the model with one explanatory variable Z_1 – we perform elementary transformations on matrix P given by (18) – bringing first column elements below diagonal to zeros. As a result:

$$P_{(Z_1)}=\left[\begin{array}{cccccc|c}1 & 0{,}9894 & 0{,}9570 & -0{,}6155 & -0{,}4547 & 0{,}4424 & 0{,}99937\\0 & 0{,}0210 & 0{,}0179 & 0{,}0127 & 0{,}0205 & 0{,}0842 & 0{,}00317\\0 & 0{,}0179 & 0{,}0841 & -0{,}1717 & 0{,}2088 & 0{,}0323 & -0{,}00284\\0 & 0{,}0127 & -0{,}1717 & 0{,}6212 & -0{,}5510 & 0{,}2388 & 0{,}015744\\0 & 0{,}0205 & 0{,}2088 & -0{,}5510 & 0{,}7932 & 0{,}0880 & -0{,}01331\\0 & 0{,}0842 & 0{,}0323 & 0{,}2388 & 0{,}0880 & 0{,}8043 & 0{,}010170\\\hline 0 & 0{,}0032 & -0{,}0028 & 0{,}0157 & -0{,}0133 & 0{,}0102 & 0{,}00126339\end{array}\right] \qquad (19)$$

Using the first column from table 1 we compute:

$$S_y^2=\frac{1}{20}\sum_{t=1}^{20}\left(y_t-\bar{y}\right)^2=22878724746 \qquad (20)$$

Thus:

$$20\cdot\ln\left(2\pi e S_y^2\right)=533{,}827007 \qquad (21)$$

Basing on (19) and (21), using (13) we have:

$$AIC_{(Z_1)} = 20 \cdot \ln\left(2\pi e S_y^2\right) + 20 \cdot \ln(0{,}00126339) + 2(1+1) = \\ 533{,}827007 + (-133{,}47909) + 4 = 404{,}347917 \tag{22}$$

To compute φ^2 for the model with Z_1 and Z_2 as explanatory variables – on matrix $P_{(Z1)}$ given by (19) we perform elementary transformations bringing second columns elements below diagonal to zeros. This time we receive:

$$P_{(Z_1,Z_2)} = \left[\begin{array}{cccccc|c} 1 & 0{,}9894 & 0{,}9570 & -0{,}6155 & -0{,}4547 & 0{,}4424 & 0{,}99937 \\ 0 & 1 & 0{,}8509 & 0{,}6055 & 0{,}9763 & 4{,}0076 & 0{,}150775 \\ 0 & 0 & 0{,}0689 & -0{,}1825 & 0{,}1913 & -0{,}0394 & -0{,}00554 \\ 0 & 0 & -0{,}1825 & 0{,}6135 & -0{,}5634 & 0{,}1878 & 0{,}013826 \\ 0 & 0 & 0{,}1913 & -0{,}5634 & 0{,}7732 & 0{,}0058 & -0{,}016406 \\ 0 & 0 & -0{,}0394 & 0{,}1878 & 0{,}0058 & 0{,}4668 & -0{,}002526 \\ \hline 0 & 0 & -0{,}0055 & 0{,}0138 & -0{,}0164 & 0{,}0025 & 0{,}00078572 \end{array}\right] \tag{23}$$

Basing on (23), using (22) we have:

$$AIC_{(Z_1,Z_2,Z_3)} = 20 \cdot \ln\left(2\pi e S_y^2\right) + 20 \cdot \ln(0{,}00034001) + 2(3+1) = \\ 533{,}827007 + (-159{,}73062) + 8 = 382{,}096384 \tag{24}$$

As AIC value has lowered – we add one another explanatory variable Z_3. After transforming the third column of matrix (23) we have:

$$P_{(Z_1,Z_2,Z_3)} = \left[\begin{array}{cccccc|c} 1 & 0{,}9894 & 0{,}9570 & -0{,}6155 & -0{,}4547 & 0{,}4424 & 0{,}99937 \\ 0 & 1 & 0{,}8509 & 0{,}6055 & 0{,}9763 & 4{,}0076 & 0{,}150775 \\ 0 & 0 & 1 & -2{,}6502 & 2{,}7785 & -0{,}5719 & -0{,}080456 \\ 0 & 0 & 0 & 0{,}1299 & -0{,}0564 & 0{,}0834 & -0{,}000856 \\ 0 & 0 & 0 & -0{,}0564 & 0{,}2416 & 0{,}1152 & -0{,}0010144 \\ 0 & 0 & 0 & 0{,}0834 & 0{,}1152 & 0{,}4443 & -0{,}0056945 \\ \hline 0 & 0 & 0 & -0{,}0009 & -0{,}0010 & -0{,}0057 & 0{,}00034001 \end{array}\right] \tag{25}$$

Thus:

$$AIC_{(Z_1,Z_2,Z_3)} = 20 \cdot \ln\left(2\pi e S_y^2\right) + 20 \cdot \ln(0{,}00034001) + 2(3+1) = \\ 533{,}827007 + (-159{,}73062) + 8 = 382{,}096384 \tag{26}$$

Also this time AIC value has lowered – we add one another explanatory variable Z_4. By transforming 4-th column of matrix (25) we have:

$$P_{(Z_1,Z_2,Z_3,Z_4)} = \left[\begin{array}{cccccc|c} 1 & 0,9894 & 0,9570 & -0,6155 & -0,4547 & 0,4424 & 0,99937 \\ 0 & 1 & 0,8509 & 0,6055 & 0,9763 & 4,0076 & 0,150775 \\ 0 & 0 & 1 & -2,6502 & 2,7785 & -0,5719 & -0,080456 \\ 0 & 0 & 0 & 1 & -0,4341 & 0,6423 & -0,006592 \\ 0 & 0 & 0 & 0 & 0,2172 & 0,1515 & -0,001386 \\ 0 & 0 & 0 & 0 & 0,1515 & 0,3907 & -0,005145 \\ \hline 0 & 0 & 0 & 0 & -0,0014 & -0,0051 & 0,00033437 \end{array}\right] \quad (27)$$

Basing on matrix (27) we can see that convergence rate for model with $\{Z_1, Z_2, Z_3, Z_4\}$ as explanatory variables has lowered (it is now 0,00033437), but AIC is equal to:

$$\begin{aligned} &AIC_{(Z_1,Z_2,Z_3,Z_4)} = 20 \cdot \ln\left(2\pi e S_y^2\right) + 20 \cdot \ln(0{,}00033437) + 2(4+1) = \\ &533{,}827007 + (-160{,}06529) + 10 = 383{,}761716 \end{aligned} \quad (28)$$

so is bigger than for the previous model (compare with (26)). It means that Z_4 shouldn't be taken into the model. We replace the variable Z_4 with another explanatory variable Z_5. In practice this means that we get back to bordered matrix given by (25), remove 4-th row and column and make elementary transformations of fifth column. As a result we have:

$$P_{(Z_1,Z_2,Z_3,Z_5)} = \left[\begin{array}{cccccc|c} 1 & 0,9894 & 0,9570 & -0,6155 & -0,4547 & 0,4424 & 0,99937 \\ 0 & 1 & 0,8509 & 0,6055 & 0,9763 & 4,0076 & 0,150775 \\ 0 & 0 & 1 & -2,6502 & 2,7785 & -0,5719 & -0,080456 \\ 0 & 0 & 0 & 0,1299 & -0,0564 & 0,0834 & -0,000856 \\ 0 & 0 & 0 & -0,0564 & 1 & 0,4770 & -0,0041981 \\ 0 & 0 & 0 & 0,0834 & 0 & 0,3893 & -0,0052106 \\ \hline 0 & 0 & 0 & -0,0009 & 0 & -0,0052 & 0,00033575 \end{array}\right] \quad (29)$$

Received in the lower right corner of the matrix (29) convergence coefficient is greater than the value obtained in the matrix (27) and the models have the same number of variables – thus we can already say that the AIC value will be higher than the value of the formula (28), and so than the (26):

$$\begin{aligned} &AIC_{(Z_1,Z_2,Z_3,Z_5)} = 20 \cdot \ln\left(2\pi e S_y^2\right) + 20 \cdot \ln(0{,}00033575) + 2(4+1) = \\ &533{,}827007 + (-159{,}98269) + 10 = 383{,}844312 \end{aligned} \quad (30)$$

It means that also variable Z_5 shouldn't be taken into the model. This time in matrix (25) we remove 4-th and 5-th row and column and run elementary transformations on 6-th column of the matrix. As a result:

$$P_{(Z_1,Z_2,Z_3,Z_6)} = \left[\begin{array}{cccccc|c} 1 & 0{,}9894 & 0{,}9570 & -0{,}6155 & -0{,}4547 & 0{,}4424 & 0{,}99937 \\ 0 & 1 & 0{,}8509 & 0{,}6055 & 0{,}9763 & 4{,}0076 & 0{,}150775 \\ 0 & 0 & 1 & -2{,}6502 & 2{,}7785 & -0{,}5719 & -0{,}080456 \\ 0 & 0 & 0 & 0{,}1299 & -0{,}0564 & 0{,}0834 & -0{,}000856 \\ 0 & 0 & 0 & -0{,}0564 & 0{,}2416 & 0{,}1152 & -0{,}0010144 \\ 0 & 0 & 0 & 0{,}0834 & 0{,}1152 & 1 & -0{,}01281066 \\ \hline 0 & 0 & 0 & -0{,}0009 & -0{,}0010 & 0 & 0{,}00026703 \end{array}\right] \quad (31)$$

$$AIC_{(Z_1,Z_2,Z_3,Z_6)} = 20 \cdot \ln\left(2\pi e S_y^2\right) + 20 \cdot \ln(0{,}00026703) + 2(4+1) = \\ 533{,}827007 + (-164{,}56317) + 10 = 379{,}263837 \quad (32)$$

The resulting value of AIC given by (32) is smaller than all the values obtained previously and Z_5 is the last one variable from the set of possible explanatory variables $A(6) = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\}$. This means the end of our proceeding. The chosen set of variables, according to the illustrated method is a 4-piece set $\{Z_1, Z_2, Z_3, Z_6\}$. It should be noted that in practice, according to the algorithm of the method, we analyzed only 6 out of 63 possible combinations of the variables but it can be easily noticed that the resulting subset has the lowest AIC value of all 63 possible subsets of the set $A(6)$ given by (17). This does not mean that it must always be, but set of variables selected according to the proposed method should be satisfactory. It should be added that the resulting model in the following steps should be verified.

At the end, in order to confirm the correctness of the calculations presented in the example, we attach Table 2 – printout of the Gretl program, together with evaluated AIC values.

Table 2. Model 1: Least squares method, using observations 1995-2014 (T = 20), Dependent variable: Y

	Coefficient	Std. error	t-Student	p-value
const	5,12693e+06	953663	5,3760	0,00008
Z_1	1,03609	0,0463455	22,3557	<0,00001
Z_2	0,533522	0,0827506	6,4474	0,00001
Z_3	-2302,74	432,336	-5,3263	0,00008
Z_6	-13,6178	6,72551	-2,0248	0,06107
Dependent variable average	322581,4		Dependent variable std dev	155186,6
Sum of squared residuals	1,22e+08		Residues standard error	2854,062
R square	0,999733		Adjusted R square	0,999662
F(4, 15)	14039,74		P value for test F	1,34e-26
Log likelihood	184,6319		**AIC**	**379,2638**
Bayes. Schwarz Criterion	384,2425		Hannan-Quinn Criterion	380,2357
rho1	-0,088852		Durbin-Watson Criterion	2,136946

Source: output from Gretl basing on data from tab 1.

Summary

The problem of initial selection of variables into an econometric model is an important issue. Among the numerous methods AIC has high popularity. Author thinks that formula (13) is important as it allows to effectively determine the value of this criterion for the model estimated by least squares. For this purpose it is necessary to calculate the values of φ^2 for the analysed models. Application of boundary matrix in the presented algorithm definitely simplifies the process of calculations. It should be added that the necessary elementary transformations, even for a high degree of matrix are easy to perform using for example MS Excel spreadsheet. The empirical example illustrates this. It is worth noting that the selection procedure gave an optimal set of variables but this is not the rule and in the author's opinion it requires further simulation studies.

Literature

Charemza W. W., Deadman D. F. (1997), *Nowa ekonometria*. Warszawa: Polskie Wydawnictwo Ekonomiczne.

Dorosiewicz S., Gruszczyński M., Kołatkowski D., Kuszewski T., Podgórska M., Syczewska E. (1996), *Ekonometria*. Warszawa. Oficyna Wydawnicza SGH.

Draper N.R., Smith H. (1973), *Analiza regresji stosowana*. Warszawa PWN.

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

Górecki B. J. (2010), *Ekonometria podstawy teorii i praktyki*. Warszawa: Wydawnictwo Key Text.

Hellwig Z. (1969), *Problem optymalnego wyboru predykant*. Przegląd Statystyczny R XVI Zeszyt 3-4, 221-237.

Kolupa M., Śleszyński Z. (2010), *Metody ekonometryczne*. Radom: Wydawnictwo Politechniki Radomskiej.

Maddala G. S. (2008), *Ekonometria*. Warszawa. Wydawnictwo Naukowe PWN SA.

Piłatowska M. (2011), *Porównanie kryteriów informacyjnych i predykcyjnych w wyborze modelu*. Sopot. Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego, nr 4/8, 499-512.

Rosienkiewicz M. (2012), *Porównanie metod Akaike i Hellwiga w zakresie efektywności konstrukcji modelu regresyjnego*. Wiadomości Statystyczne, nr 10 (617), 27-42.