ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 152, 2000

*Krystyna Pruska**

BOOTSTRAP DISTRIBUTION OF OLS-ESTIMATORS FOR LINEAR REGRESSION MODELS

Abstract. The bootstraps methods can be widely applied in statistical research. In the paper the bootstraps OLS-estimators for linear models are considered. The results of simulation experiments for linear models with errors which have normal, Student and uniform distribution are presented. The estimates and the histograms for OLS-estimators are determined.

1. INTRODUCTION

The bootstrap methods give possibility of an approximation of unknown distribution of the random variable. They can be used in the estimation of parameters and function form of distribution and in verification of statistical hypotheses. In the paper the application of bootstrap methods is presented for approximation of OLS-estimator distribution in the case of linear models and small sample.

When random error of model is normally distributed and other classical assumption are fulfilled, those estimators have normal distribution, too. It is important to know distribution of OLS-estimators in the case of non-normal errors, because then we can construct confidence intervals for model parameters and verify the hypotheses about them. The bootstrap methods can be used for it. In the paper the models with errors, which have normal, Student or uniform distribution, are considered.

* University of Łódź, Chair of Statistical Methods.

2. BOOTSTRAP ESTIMATION OF LINEAR MODEL PARAMETERS ON THE BASIS OF OLS-ESTIMATION

OLS-method is very often used in statistical and econometrical investigations but it is difficult to determine the distribution of OLS-estimators for small sample and non-normal errors in linear models. The bootstrap methods give such possibility.

Let a model be given, which is of the form:

$$y_i = g(x_i, \beta) + \varepsilon_i \quad \text{for} \quad i = 1, \ldots, n \tag{1}$$

where

y_i – value of explained variable Y,

x_i – vector of values of explanatory variables (determined values),

β – vector of model parameters,

ε_i – value of random error,

g – linear function.

We assume that $(y_1, x_1), \ldots, (y_n, x_n)$ are obtained as a result of independent drawing.

Let $\hat{\beta}$ be OLS-estimator of parameter β, i.e. $\hat{\beta}$ is vector for which the function:

$$G(\tilde{\beta}) = \sum_{i=1}^{n} [y_i - g(x_i, \tilde{\beta})]^2 \tag{2}$$

has its minimum.

The construction of bootstrap distribution for OLS-estimator will be presented now (see Efron 1979).

We assume that

$$\hat{\varepsilon}_i = y_i - g(x_i, \hat{\beta}) \quad \text{for} \quad i = 1, \ldots, n \tag{3}$$

The following distribution

$$P(E_B = \hat{\varepsilon}_i) = \frac{1}{n} \quad \text{for} \quad i = 1, \ldots, n \tag{4}$$

is called the probability distribution from sample for $\hat{\varepsilon}_i$, $i = 1, \ldots, n$.

We generate n values from distribution (4). Let $\varepsilon_1^*, \ldots \varepsilon_n^*$ be them. Then we calculate values $y_1^*, \ldots, y_n^*$ in accordance with the formula:

$$y_i^* = g(x_i, \hat{\beta}) + \varepsilon_i^* \quad \text{for} \quad i = 1, \ldots, n \tag{5}$$

Next we determine OLS-estimate for β on the basis of the following model:

$$y_i^* = g(x_i, \beta) + \tilde{\varepsilon}, \quad \text{for} \quad i = 1, \ldots, n \tag{6}$$

where $\tilde{\varepsilon}_i$ is value of unknown random errors.

Generating values $\varepsilon_1^*, \ldots, \varepsilon_n^*$, calculating $y_1^*, \ldots, y_n^*$ and determining estimates of β for model (6) are repeated N times (for example $N = 1000$). Finally, we obtain N OLS-estimates for the model parameter. We denote them by $\beta_1^*, \ldots, \beta_N^*$. We may determine the histogram of their distribution, which can be called bootstrap distribution of OLS-estimator of β. The mean of N values of OLS-estimates for the model (6) is called bootstrap estimate of parameter β.

It can be shown that (see Efron, 1979; Efron, Tibshirani, 1993, p. 111–112):

$$E(\hat{\beta}^*) = \hat{\beta} \tag{7}$$

and

$$\operatorname{cov}(\hat{\beta}^*) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \tag{8}$$

where $\hat{\beta}^*$ denotes bootstrap estimator of β determined on the basis OLS--estimation, $\hat{\beta}$ is OLS-estimate of β, $\mathbf{X}' = (x_1, \ldots, x_n)$ and

$$\hat{\sigma}^2 = \frac{1}{2} \sum_{i=1}^{n} [y_i - g(x_i, \hat{\beta})]^2 \tag{9}$$

The bootstrap distributions of OLS-estimators are the approximation of OLS-estimator distribution. They can be applied for construction of confidence intervals and for verification of hypotheses about model parameters.

The presented considerations may be developed for the non-linear models.

3. BOOTSTRAP DISTRIBUTION OF OLS-ESTIMATORS FOR LINEAR MODEL WITH NORMAL, STUDENT AND UNIFORM ERRORS

In this section the bootstrap distributions of OLS-estimators are presented on the simulation examples. In this way we can investigate some properties of OLS-estimators and bootstrap OLS-estimators.

We consider the following model:

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_2 x_{3t} + \varepsilon_t \tag{10}$$

for $t = 1, \ldots, n$.

The experiments were conducted for $n = 20$, $\alpha_0 = 1$, $\alpha_1 = 3$, $\alpha_2 = 2$, $\alpha_3 = 1$. The values $x_{11}, ..., x_{1n}$ were generated from normal distribution $N(5; 1)$, $x_{21}, ..., x_{2n}$ – from distribution $N(3; 2)$ and $x_{31}, ..., x_{3n}$ – from distribution $N(10; 3)$. Then three groups of experiments were constituted.

In the first group the values of random error were generated from normal distribution. These experiments are denoted by $N1, ..., N10$. In the second group random error has Student distribution. The experiments have the following denotations $S1, ..., S5$. In the third group the values of error were generated from uniform distribution. These experiments are denoted by $U1, ..., U10$. The assumed values of standard deviation of error distribution are given in Tab. 1. The error expectation is equal to zero.

Table 1

Standard deviation of distributions in simulation experiments

Experiment number	Standard deviation
N1, S1, U1	1.732
N2, S2, U2	1.414
N3, S3, U3	1.291
N4, S4, U4	1.225
N5, S5, U5	1.183
N6, U6	0.100
N7, U7	0.300
N8, U8	0.500
N9, U9	1.000
N10, U10	2.000

Source: Author's assumptions.

In experiments S1, ..., S5 we have the Student distribution with 3, 4, 5, 6, 7 d.f. correspondingly.

On the basis of such data the values $y_1, ..., y_n$ were determined in each experiment according with formula (10).

Next the OLS-estimation and bootstrap OLS-estimation were carried out according with the procedure presented in Section 2. In bootstrap method the repetition number is equal to 1000. The results of all experiments are presented in Tab. 2–4.

Table 2

Results of experiments N1–N10

Experiment number (standard deviation of random error)	Parameter	True value of parameter	OLS-estimate of parameter (standard error)	Bootstrap OLS-estimate of parameter	Standard deviation of bootstrap distribution of parameter	Value of χ^2 test statistic
N1 (1.732)	α_0	1	5.689 (3.103)	5.570	2.664	4.649
	α_1	3	2.857 (0.406)	2.874	0.350	6.083
	α_2	2	1.538 (0.176)	1.536	0.155	4.039
	α_3	1	0.795 (0.148)	0.799	0.130	1.142
N2 (1.414)	α_0	1	4.829 (2.534)	4.731	2.175	4.649
	α_1	3	2.883 (0.331)	2.897	0.286	6.083
	α_2	2	1.623 (0.144)	1.622	0.128	3.978
	α_3	1	0.833 (0.121)	0.836	0.107	1.142
N3 (1.291)	α_0	1	4.495 (2.313)	4.406	1.986	4.649
	α_1	3	2.893 (0.303)	2.906	0.261	6.241

Table 2 (contd.)

Experiment number (standard deviation of random error)	Parameter	True value of parameter	OLS-estimate of parameter (standard error)	Bootstrap OLS-estimate of parameter	Standard deviation of bootstrap distribution of parameter	Value of χ^2 test statistic
N3 (1.291)	α_2	2	1.655 (0.131)	1.654	0.115	3.978
	α_3	1	0.847 (0.110)	0.850	0.097	1.142
N4 (1.225)	α_0	1	4.316 (2.194)	4.231	1.884	4.649
	α_1	3	2.897 (0.287)	2.911	0.247	6.083
	α_2	2	1.673 (0.125)	1.672	0.109	4.039
	α_3	1	0.855 (0.105)	0.858	0.092	0.765
N5 (1.183)	α_0	1	4.203 (2.120)	4.122	1.820	4.649
	α_1	3	2.902 (0.277)	2.914	0.239	6.241
	α_2	2	1.684 (0.120)	1.683	0.106	3.978
	α_3	1	0.861 (0.101)	0.863	0.089	0.0765

N6 (0.1)	α_0	1	1.271 (0.179)	1.264	0.154	4.649
	α_1	3	2.992 (0.023)	2.993	0.200	6.241
	α_2	2	1.973 (0.010)	1.973	0.009	4.039
	α_3	1	0.988 (0.009)	0.988	0.008	0.765
N7 (0.3)	α_0	1	1.812 (0.537)	1.792	0.461	4.649
	α_1	3	2.975 (0.070)	2.978	0.061	6.241
	α_2	2	1.919 (0.031)	1.920	0.027	3.978
	α_3	1	0.965 (0.026)	0.965	0.023	0.765
N8 (0.5)	α_0	1	2.354 (0.896)	2.319	0.769	4.042
	α_1	3	2.959 (0.117)	2.964	0.101	6.241
	α_2	2	1.867 (0.051)	1.867	0.045	3.978
	α_3	1	0.941 (0.043)	0.942	0.038	0.7653

Table 2 (contd.)

Experiment number (standard deviation of random error)	Parameter	True value of parameter	OLS-estimate of parameter (standard error)	Bootstrap OLS-estimate of parameter	Standard deviation of bootstrap distribution of parameter	Value of χ^2 test statistic
N9 (1.0)	α_0	1	3.707 (1.792)	3.638	1.538	4.649
	α_1	3	2.917 (0.234)	2.927	0.202	6.083
	α_2	2	1.733 (0.102)	1.732	0.089	4.039
	α_3	1	0.882 (0.085)	0.884	0.075	0.765
N10 (2.0)	α_0	1	6.415 (3.583)	6.276	3.076	4.649
	α_1	3	2.835 (0.469)	2.855	0.404	6.241
	α_2	2	1.466 (0.204)	1.465	0.179	4.039
	α_3	1	0.763 (0.171)	0.768	0.151	0.765

Table 3

Results of experiments S1–S5

Experiment number (standard deviation of random error)	Parameter	True value of parameter	OLS-estimate of parameter (standard error)	Bootstrap OLS-estimate of parameter	Standard deviation of bootstrap distribution of parameter	Value of χ^2 test statistic
S1 (1.732)	α_0	1	4.271 (1.674)	4.300	1.484	10.749
	α_1	3	2.555 (0.219)	2.549	0.193	17.037
	α_2	2	2.101 (0.095)	2.103	0.081	8.627
	α_3	1	0.867 (0.080)	0.867	0.070	4.148
S2 (1.414)	α_0	1	2.621 (3.612)	2.587	3.175	6.921
	α_1	3	2.903 (0.473)	2.907	0.401	7.246
	α_2	2	1.825 (0.205)	1.828	0.191	14.599
	α_3	1	0.910 (0.172)	0.910	0.152	2.626
S3 (1.291)	α_0	1	1.081 (2.556)	1.042	2.326	4.516
	α_1	3	2.962 (0.334)	2.964	0.301	24.712

Table 3 (contd.)

Experiment number (standard deviation of random error)	Parameter	True value of parameter	OLS-estimate of parameter (standard error)	Bootstrap OLS-estimate of parameter	Standard deviation of bootstrap distribution of parameter	Value of χ^2 test statistic
S3 (1.291)	α_2	2	1.890 (0.145)	1.891	0.137	7.268
	α_3	1	1.032 (0.122)	1.033	0.113	16.325
S4 (1.225)	α_0	1	1.909 (1.998)	1.913	1.782	2.141
	α_1	3	2.881 (0.260)	2.884	0.232	2.545
	α_2	2	1.969 (0.113)	1.965	0.099	7.467
	α_3	1	0.982 (0.095)	0.982	0.087	3.907
S5 (1.183)	α_0	1	−3.377 (1.084)	−3.408	0.931	1.656
	α_1	3	3.470 (0.142)	3.470	0.126	16.117
	α_2	2	2.229 (0.062)	2.230	0.054	7.522
	α_3	1	1.161 (0.052)	1.163	0.044	5.869

Table 4

Results of experiments U1–U10

Experiment number (standard deviation of random error)	Parameter	True value of parameter	OLS-estimate of parameter (standard error)	Bootstrap OLS-estimate of parameter	Standard deviation of bootstrap distribution of parameter	Value of χ^2 test statistic
U1 (1.732)	α_0	1	3.054 (2.944)	3.115	2.754	4.673
	α_1	3	2.793 (0.385)	2.792	0.363	12.626
	α_2	2	1.789 (0.167)	1.780	0.149	5.510
	α_3	1	1.062 (0.140)	1.061	0.127	3.086
U2 (1.414)	α_0	1	2.677 (2.403)	2.727	2.248	6.045
	α_1	3	2.831 (0.314)	2.830	0.296	13.644
	α_2	2	1.828 (0.137)	1.821	0.122	5.510
	α_3	1	1.050 (0.114)	1.050	0.104	3.086
U3 (1.291)	α_0	1	2.531 (2.194)	2.577	2.053	4.673
	α_1	3	2.846 (0.287)	2.845	0.270	13.644

Table 4 (contd.)

Experiment number (standard deviation of random error)	Parameter	True value of parameter	OLS-estimate of parameter (standard error)	Bootstrap OLS-estimate of parameter	Standard deviation of bootstrap distribution of parameter	Value of χ^2 test statistic
U3 (1.291)	α_2	2	1.843 (0.125)	1.836	0.111	5.095
	α_3	1	1.046 (0.105)	1.046	0.097	2.702
U4 (1.225)	α_0	1	2.453 (2.082)	2.496	1.948	6.045
	α_1	3	2.854 (0.272)	2.853	0.257	13.644
	α_2	2	1.851 (0.118)	1.845	0.106	5.510
	α_3	1	1.044 (0.099)	1.043	0.090	3.086
U5 (1.183)	α_0	1	2.403 (2.011)	2.445	1.881	6.045
	α_1	3	2.859 (0.263)	2.858	0.248	13.644
	α_2	2	1.856 (0.114)	1.850	0.102	5.510
	α_3	1	1.042 (0.096)	1.042	0.087	3.086

U6 (0.1)	α_0	1	1.119 (0.170)	1.122	0.159	4.673
	α_1	3	2.988 (0.022)	2.989	0.021	13.644
	α_2	2	1.988 (0.010)	1.987	0.009	5.510
	α_3	1	1.004 (0.008)	1.004	0.007	2.702
U7 (0.3)	α_0	1	1.356 (0.510)	1.366	0.477	4.673
	α_1	3	2.964 (0.067)	2.964	0.063	13.644
	α_2	2	1.963 (0.029)	1.962	0.026	5.510
	α_3	1	1.011 (0.024)	1.011	0.022	3.086
U8 (0.5)	α_0	1	1.593 (0.850)	1.611	0.795	4.673
	α_1	3	2.940 (0.111)	2.940	0.105	13.644
	α_2	2	1.939 (0.048)	1.937	0.043	5.095
	α_3	1	1.018 (0.040)	1.018	0.037	3.086

Table 4 (contd.)

Experiment number (standard deviation of random error)	Parameter	True value of parameter	OLS-estimate of parameter (standard error)	Bootstrap OLS-estimate of parameter	Standard deviation of bootstrap distribution of parameter	Value of χ^2 test statistic
U9 (1.0)	α_0	1	2.186 (1.700)	2.221	1.590	6.045
	α_1	3	2.880 (0.222)	2.880	0.209	13.644
	α_2	2	1.878 (0.097)	1.873	0.086	5.510
	α_3	1	1.036 (0.081)	1.035	0.073	3.086
U10 (2.0)	α_0	1	3.372 (3.399)	3.442	3.180	4.673
	α_1	3	2.761 (0.447)	3.760	0.419	13.644
	α_2	2	1.756 (0.493)	1.747	0.172	5.095
	α_3	1	1.071 (0.162)	1.071	0.147	2.702

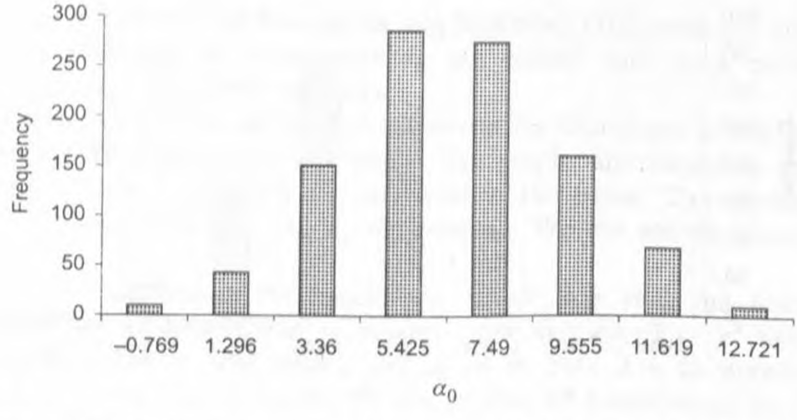

Fig. 1. Histogram of bootstrap distribution of OLS-estimator for parameter a_0 in experiment N_1

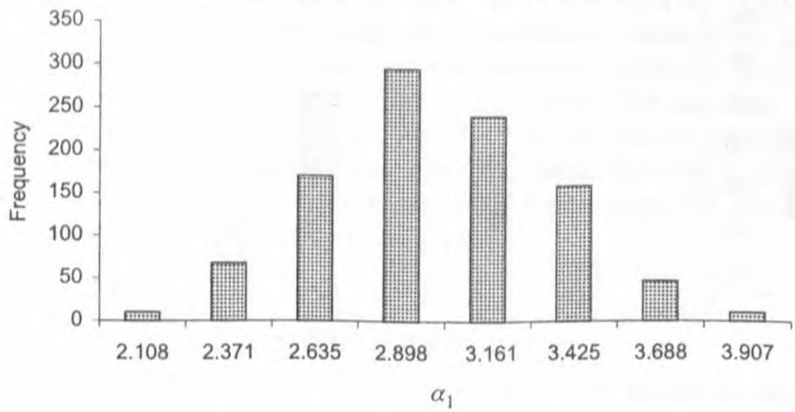

Fig. 2. Histogram of bootstrap distribution of OLS-estimator for parameter a_1 in experiment N_1

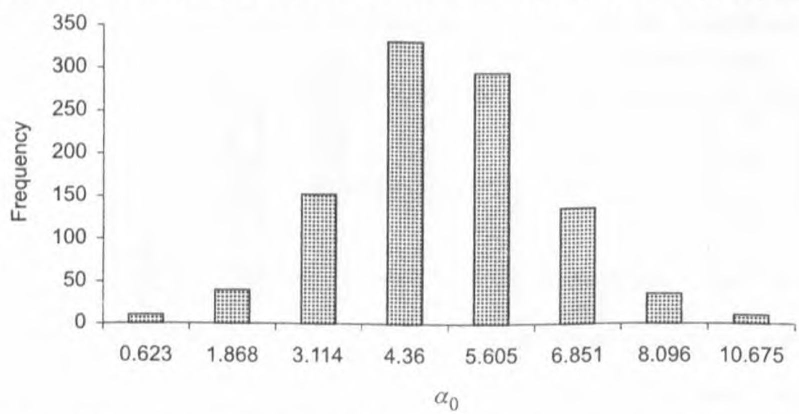

Fig. 3. Histogram of bootstrap distribution of OLS-estimator for parameter a_0 in experiment S_1

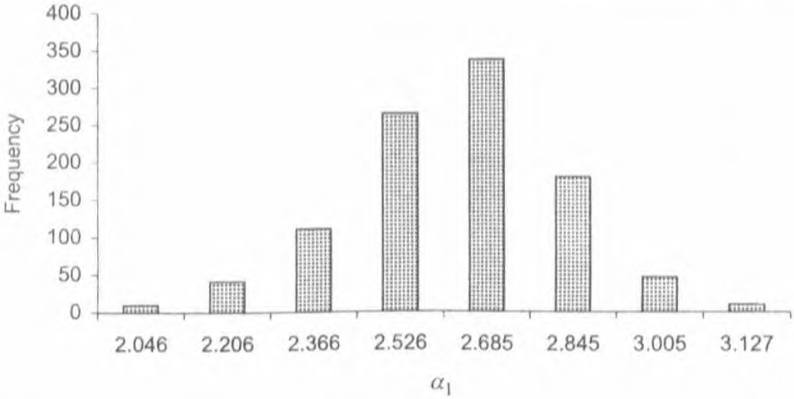

Fig. 4. Histogram of bootstrap distribution of OLS-estimator for parameter a_1 in experiment S_1

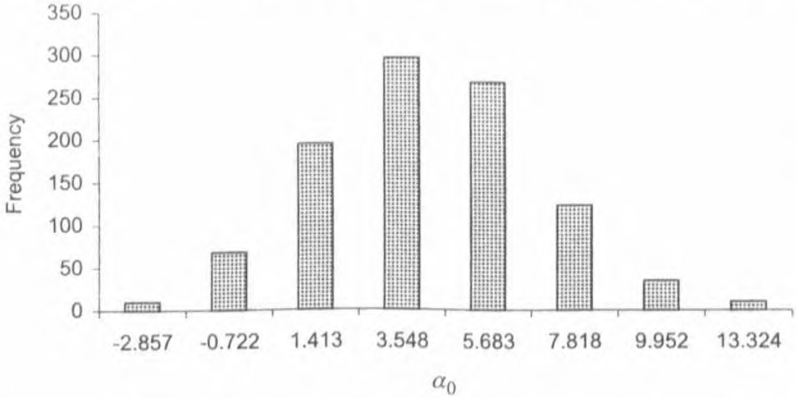

Fig. 5. Histogram of botstrap distribution of OLS-estimator for parameter a_0 in experiment U_1

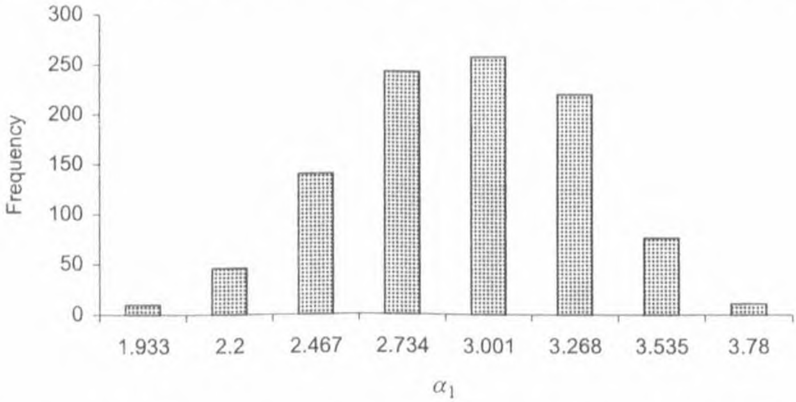

Fig. 6. Histogram of botstrap distribution of OLS-estimator for parameter a_1 in experiment U_1

We can notice that OLS-estimates and bootstrap OLS-estimates are near but the evaluations of parameters α_0 are biased with large error for a relatively large standard deviation.

For each experiment and for each parameter the histograms of distribution of bootstrap OLS-estimator were made. The graphs are numerous and for this reason a few histograms are presented in the paper. The experiments N1, S1, U1 and parameters α_0, α_1 were chosen. We can see the histograms in Fig. 1–6.

For each experiment the hypothesis, which says that the bootstrap distribution of OLS-estimator is normal, was verified. The χ^2 test was applied and values of test statistic are given in Tab. 2–4. In experiments N1–N10 the results are obvious. We know, that OLS-estimators for linear models with normal errors are normally distributed. In other experiments the verified hypothesis is rejected in 15 cases out of 100 cases for significance level 0.05. It is consistent with properties of estimators obtained by pseudo maximum likelihood method which are asymptotically normally distributed (see Gourieroux, Monfort, Trognion, 1984). We can treat OLS--estimators in experiment S1–S5 and U1–U10 as estimators obtained by this method. In our experiments the sample is small (its size is equal to 20), but we can observe that distribution of OLS-estimators for our model often has approximately normal distribution.

4. FINAL REMARKS

The aim of the paper is to show wide possibilities of bootstrap methods on the example of estimation of linear model. They enable determining estimates of unknown parameters and approximation of unknown probability distribution. They can also find application in estimation procedures and in verification of statistical hypotheses, especially in those cases, when determining the distribution of estimators and test statistics is difficult.

REFERENCES

Efron B. (1979), *Bootstrap Methods: Another Look at the Jackknife*, "The Annals of Statistics", 7, 1–26.

Efron B., Tibshirani R. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Gourieroux G., Monfort A., Trognion A. (1984), *Pseudo Maximum Likelihood Methods: Theory*, "Econometrica", **52**, 681–700.

Pruska K. (1996), *Metody regresji przełącznikowej i ich zastosowanie*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.