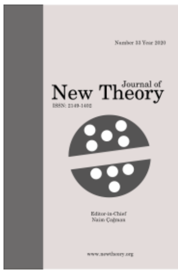
Journal of New Theory
ISSN: 2149-1402

33 (2020) 76-84
Journal of New Theory
http://www.newtheory.org
Open Access



Model Selection in Beta Regression Analysis Using Several Information Criteria and Heuristic Optimization

Emre DÜNDER[1], Mehmet Ali CENGİZ[2]

Article History
Received: 07.12.2020
Accepted: 17.12.2020
Published: 31.12.2020
Original Article

Abstract − In the context of generalized linear modeling (GLM), the beta regression analysis is used to estimate regression models when the dependent variable lies between (0,1). In this paper, we carried out a model selection process using several information criteria with heuristic optimization. We employed the differential evolution algorithm as a heuristic optimization method to select the best model for beta regression analysis. The results show that the alternative-type information criteria provide competitive results during the model selection process in beta regression analysis.

Keywords *− Beta regression, differential evolution algorithm, information criteria, model selection*

1. Introduction

In the regression analysis, the models are estimated according to the distribution of the dependent variable. Therefore, the distribution of the dependent variable should be appropriate, and the alternative choices can be used in the violation of normality. The generalized linear modelling (GLM) approach is employed to construct the regression models for several distribution-types such as Poisson, Gamma, Binomial, etc. Also, the selection of the optimal model including the best explanatory variable is very crucial. There are lots of attempts in the literature for the model selection in regression modeling. Sakate et al. [1] proposed a stepwise selection approach for Poisson regression analysis. Calcagno and Mazancourt [2] developed a software package to apply model selection for GLM. Örkcü [3] used and hybrid simulated annealing approach in regression analysis for model selection. Unler ve Murat [4] proposed a model selection approach based on the particle swarm optimization for binary dependent variables. The researchers also utilized the information criteria during the model selection [5-9].

Overall, we can see that the model selection process in GLM can be handled with three components:

1) GLM with an appropriate distribution,

2) The information criteria,

3) An appropriate optimization technique.

[1] emre.dunder@omu.edu.tr; [2] macengiz@omu.edu.tr
[1,2] Department of Statistics, Faculty of Arts and Science, Ondokuz Mayıs University, Samsun, Turkey

When the dependent variable is ranging between $(0,1)$ beta regression analysis can be used instead of the normal linear regression analysis [10]. By inspiring the current model approaches, we performed model selection in beta regression analysis using several information criteria.

To the best of our knowledge, this study is the first attempt to apply model selection under the alternative Bayesian and Information Complexity (ICOMP) type criteria for beta regression analysis. Also, we adopted the binary type of differential evolution algorithm to choose the optimal subset of the explanatory variables [11].

The article is organized as follows: In Section 2, we introduced the beta regression analysis and parameter estimates. In Section 3 we described the working mechanism of the binary differential evolution algorithm. In Section 4, we presented the information criteria which are used in the article. In Section 5, we introduced the numerical examples including the simulation studies and real data set applications. Finally, we presented the conclusion and discussion part in Section 6.

2. Beta regression analysis

In the regression analysis, beta regression analysis is used when the dependent variable includes ratio or percentage values in the interval $(0,1)$. The main reason for using beta regression analysis is that the ratio or percentage values of the data in the interval $(0,1)$ are suitable for the beta distribution due to the nature of the data [12]. The probability function of the beta distribution is expressed as follows:

$$f\ (y|\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma\big((1-\mu)\phi\big)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1},\qquad 0<y<1 \tag{1}$$

where μ shows the location parameter and ϕ shows the dispersion parameter. The parameter of μ is between $(0,1)$ and $\phi>0$. The expected value of the variable in beta distribution is $E(y)=\mu$ and the variance is $Var(y)=\mu\,(1-\mu)/(1+\phi)$ [13].

Based on the probability distribution, the log-likelihood function for the beta regression is written as:

$$L(\mu,\phi)=\log\Gamma(\phi)-\log\Gamma(\mu\phi)-\log\Gamma\big((1-\mu)\phi\big)+(\mu\phi-1)\log y+\big((1-\mu)\phi-1\big)\log(1-y) \tag{2}$$

By following the GLM approach, the beta regression can be written as the following form:

$$g(\mu)=\beta_0+\sum_{k=1}^{p}\beta_{ik}\,X_{ik},i=1,2,\dots,n \tag{3}$$

where $g(.)$ denotes the link function and $\beta=(\beta_1,\beta_2,\dots,\beta_p)$ vector represents the regression coefficients, and they are estimated with the numerical methods such as Newton-Rapson or Fisher scoring on the log-likelihood function. The detailed information can be found in [13].

Table 1. The link functions for beta regression

Link function	Formula
Logit	log(μ (1 − μ))
Log-log	-log(-log(μ))
Complementary log-log	log(-log((1 − μ)))
Probit	Φ^{-1}(μ)
Cauchy	tan(π(μ-0.5))
Log	log(μ)

The beta regression model can be estimated with various link functions. Table 1 shows the possible link functions that can be used within beta regression framework. In Table 1, $\log(.)$ shows the natural logarithm, $\Phi(.)$ denotes the cumulative link function of the standard normal distribution and $\tan(.)$ indicates the trigonometric tangent function [10,14].

3. Differential evolution algorithm for binary search

Heuristic techniques are very common to optimize a target function. Heuristic methods are implemented effectively inside the statistical modeling. Especially, the model selection is carried out with a proper heuristic optimization technique for determining the best variable subset. To carry out the model selection, each variable is coded as binary and the optimal set is chosen such as optimizing a target function.

The mentioned approach is employed in regression modeling and we can perform the model selection in beta regression analysis. Differential evolution is a very effective optimization method [15] and we used this method as a binary version. In our approach, the target functions are the information criteria that evaluates the several beta regressions models.

Differential evolution algorithm consists of four main steps:

1) Generating the initial population
2) Mutation
3) Crossover
4) Selection

As it occurs in the similar algorithms, the initial population is generated at the starting point [11]. The initial population is generated as the following way:

$$x_i^G = x_{i\,(L)} + rand_i\,[0,1]\,(x_{i\,(H)} - x_{i\,(L)}) \tag{4}$$

where $x = (x_1, x_2 \dots x_p)$ is the vector of the parameters, $x_{i\,(H)}$ and $x_{i\,(L)}$ are the bounds and $rand$ is the generated parameters as uniformly distributed. Since we employ a binary search, the bounds are limited between [0,1] and the parameters are rounded. Figure 1 represents the coding scheme for model selection.

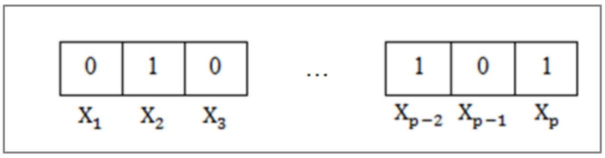


Fig. 1: The coding of the explanatory variables

The mutation process is applied to make the search process more robust and durable. This process also allows new regions in the search space to be discovered. For this purpose, vectors called trial parameters are created. Trial parameters are formed by adding the weighted difference between two units onto a third unit [11]. The parameters re-derived in this way are evaluated within the objective function (i.e. information criteria) according to the values of the previous units. If the value of the objective function consisting of the re-derived vectors is better than the previous value, it is replaced with the more appropriate parameter vectors. This process for each size proceeds as follows:

$$v^{G+1} = x_{r3}^{\ G} + \mathrm{F}\,(x_{r1}^{\ G} - x_{r2}^{\ G}) \tag{5}$$

where $\mathrm{F} \in [0,1]$ is a scaling factor for $r_1 \neq r_2 \neq r_3$.

After the mutation stage, a crossover process is applied over the parameter vectors. Crossing is used to strengthen the success level of the parameter vectors obtained in the mutation stage and to define new vectors by acting from existing vectors. In the crossover stage, a crossover constant defined as $CR \in [0,1]$ is processed as follows:

$$x^{G+1} = \begin{cases} u^{G+1}, & f(u^{G+1}) \leq f(x^G) \\ x^G, & f(u^{G+1}) > f(x^G) \end{cases} \quad (6)$$

where f(.) denotes the target function. This process is iteratively employed until reaching the optimum.

4. Information criteria and model selection

Information criterion is a measure that shows the performance of a statistical model. Mainly, information criteria attempt to penalize the bias and they are widely used within the scope of regression models, especially when choosing the most appropriate model.

The general structure of an information criterion is defined as follows:

$$\text{Information criterion} = -\,2\,LL + \text{Penalty} \quad (7)$$

where LL is the log-likelihood of the model.

The penalty of the criterion has a huge impact on the selected models in regression analysis. The formulations of the most common criteria such as AIC and BIC are given as follows:

$$AIC(k) = -2LL + 2k \quad (8)$$
$$BIC(k) = -2LL + k\log(n) \quad (9)$$

where k shows the number of free parameters and n shows the sample size. In regression models, k represents the number of explanatory variables.

Moreover, the different terms can be added as the penalty inside the information criteria. For example, ICOMP-type criteria include the covariance matrix of the statistical models with a complexity function [16,17]. There are also different criteria called as the information matrix-based information criterion (IBIC) and scaled unit information prior Bayesian information criterion (SPBIC) [18].

Table 2. Penalty terms of the seven information criteria

Information criteria	Penalty
ICOMPifim	2C(F^{-1})
ICOMPpeu	k+2C(F^{-1})
ICOMPpeuln	k+log(n)C(F^{-1})
ICOMPperf	n+2C(F^{-1})
CICOMP	k(1+log(n)) + 2C(F^{-1})
IBIC	klog(n/2π) + log $\lvert(\hat{\Sigma}_{model})^{-1}\rvert$
SPBIC	p (1 - log(k/ ($\beta^T(\hat{\Sigma}_{model})^{-1}\beta$)))

Table 2 demonstrates the penalty terms of five ICOMP-type and two Bayesian-type information criteria. In ICOMP-type criteria, $C(.)$ denotes the complexity function and F is the Fisher information matrix of the statistical model. We considered the C_{1F} complexity, shown as:

$$C_{1F}(.) = \frac{1}{4\lambda_m{}^2}\sum_{i=1}^{p}(\lambda_i - \lambda_m)^2 \quad (10)$$

where λ_m is the arithmetic mean of the eigenvalues which obtained from the inverse of Fisher information matrix [19]. In our study, we consider the Fisher information matrix of the beta regression model. In IBIC and SPBIC, $\hat{\Sigma}_{model}$ is the variance-covariance matrix of the beta regression models.

5. Numerical examples

In this part we performed a simulation study and real data applications on model selection for beta regression analysis. Mainly, we assessed the performance of the seven criteria in terms of the model selection capability.

We formulated the simulation settings from the following equation [20]:

$$\log\left(\frac{\mu}{1-\mu}\right) = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \tag{11}$$

The expected value, μ is generated by considering the logit link function. After generating the expected values, we simulated the response variable from the beta distribution with a fixed dispersion parameter ϕ.

In simulation part, we obtained the ratio of correct selected variables (C) and incorrectly selected variables (I) for each correct and incorrect variable set. The number of runs is 100 for the simulation settings.

In real data analysis part, we evaluated the prediction errors and the number of the significant-insignificant variables. All the implementations were performed betareg and DEoptim packages, existing in R software [21, 22].

5.1 Simulation study-1

The simulation part includes a design which has the multicollinearity among the predictors. The explanatory variables are generated as the following way:

$$x_1 = 10 + \varepsilon_1 \tag{12}$$
$$x_2 = 10 + 0.3\varepsilon_1 + \alpha\varepsilon_2 \tag{13}$$
$$x_3 = 10 + 0.3\varepsilon_1 + 0.5604\alpha\varepsilon_2 + 0.8282\alpha\varepsilon_3 \tag{14}$$
$$x_4 = -8 + x_1 + 0.5x_2 + 0.3x_3 + 0.5\varepsilon_4 \tag{15}$$
$$x_5 = -5 + 0.5x_1 + x_2 + 0.5\varepsilon_5 \tag{16}$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5 \sim N(0,1)$. We fixed $\alpha = 0.3$ for this setting and it causes the multicollinearity [23]. The true model includes x_1, x_2, x_3 and the generation process is conducted as follows:

$$L = 5 + 1.75x_1 + 1.25x_2 + 0.5x_3 + e \tag{17}$$
$$M = 1/(1 + exp(-L)) \tag{18}$$
$$Y \sim Beta(M\phi, (1 - M)\phi) \tag{19}$$

As it is seen from above, the correct model includes $\{x_1, x_2, x_3\}$. We considered $\phi = 5, 50$ with an error term $e \sim N(0,1)$ and added five irrelevant variables as $d\ x\ R_d \sim U(0,1)\ d=1,2,3,4,5$. The sample size was chosen as $n = 50, 100, 300, 500$.

Table 3. The ratio of the correctly selected variables beta regression models for $\phi = 5$

Criteria	Sample size				Average
	$n = 50$	$n = 100$	$n = 300$	$n = 500$	
AIC	77.333	90.667	99.333	100.000	91.833
BIC	67.333	83.333	98.000	98.639	86.827
IBIC	60.000	74.000	90.667	96.599	80.316
SPBIC	68.667	86.667	98.667	99.320	88.330
CICOMP	50.000	66.000	88.667	96.599	75.316
ICOMPifim	68.667	86.667	99.333	100.000	88.667
ICOMPpeu	68.667	86.667	99.333	100.000	88.667
ICOMPpeuln	68.667	86.667	99.333	100.000	88.667
ICOMPperf	68.667	86.667	99.333	100.000	88.667

Table 4. The ratio of the incorrectly selected variables beta regression models for $\phi = 5$

Criteria	Sample size				Average
	$n = 50$	$n = 100$	$n = 300$	$n = 500$	
AIC	19.200	13.200	16.800	14.286	15.871
BIC	3.600	2.800	2.000	0.816	2.304
IBIC	0.400	0.000	0.000	0.000	0.100
SPBIC	8.800	5.600	4.000	3.265	5.416
CICOMP	1.200	0.000	0.000	0.000	0.300
ICOMPifim	8.400	5.600	4.400	4.898	5.824
ICOMPpeu	8.400	5.600	4.400	4.898	5.824
ICOMPpeuln	8.400	5.600	4.400	4.898	5.824
ICOMPperf	8.400	5.600	4.400	4.898	5.824

Table 5. The ratio of the correctly selected variables beta regression models for $\phi = 50$

Criteria	Sample size				Average
	$n = 50$	$n = 100$	$n = 300$	$n = 500$	
AIC	99.333	100.000	100.000	100.000	99.833
BIC	99.333	100.000	100.000	100.000	99.833
IBIC	95.333	100.000	100.000	100.000	98.833
SPBIC	98.667	100.000	100.000	100.000	99.667
CICOMP	94.000	100.000	100.000	100.000	98.500
ICOMPifim	97.333	100.000	100.000	100.000	99.333
ICOMPpeu	97.333	100.000	100.000	100.000	99.333
ICOMPpeuln	97.333	100.000	100.000	100.000	99.333
ICOMPperf	97.333	100.000	100.000	100.000	99.333

Table 6. The ratio of the incorrectly selected variables beta regression models for $\phi = 50$

Criteria	Sample size				Average
	$n = 50$	$n = 100$	$n = 300$	$n = 500$	
AIC	18.400	19.600	16.800	12.400	16.800
BIC	7.200	5.200	3.200	1.200	4.200
IBIC	0.000	0.000	0.000	0.000	0.000
SPBIC	4.000	4.000	2.000	0.800	2.700
CICOMP	0.000	0.000	0.000	0.000	0.000
ICOMPifim	1.600	3.200	2.400	1.200	2.100
ICOMPpeu	1.600	3.200	2.400	1.200	2.100
ICOMPpeuln	1.600	3.200	2.400	1.200	2.100
ICOMPperf	1.600	3.200	2.400	1.200	2.100

The simulation results are shown in Table 3-6. When checking the results, we see that AIC was able to select the correct variables in beta regression models. However, covariance-based information criteria give competitive results while determining the true models. ICOMP-type and BIC-type alternative criteria become superior to the classical ones while excluding the incorrect variables. The ratio of the incorrectly selected variables is lower than AIC and BIC. As the sample size increases, the model selection capability of ICOMP-type and BIC-type criteria become much obvious.

5.2 Real data analysis

In this part, we conducted real data analysis applications on two data sets. We considered two real benchmark data sets, Bodyfat and Vitamin. Bodyfat [24] consists $p = 17$ and Vitamin [25] consists $p = 11$ explanatory variables. In each data sets the range of the response variable is (0,1).

Table 7. Kolmogorov-Smirnov test for the beta distribution of the response variables

Data set	$M\phi$	$M(1-\phi)$	Statistic	p
Bodyfat	4.783	20.372	0.061	0.302
Vitamin	8.341	4.310	0.136	0.127

Table 7 shows the goodness of fit tests testing the fitness of the beta distribution for the response variables. The test results reveal that the response variables follow the beta distribution.

To test the predictive performance of the criteria in beta regression models, we used the standardized absolute errors [26] as follows:

$$SAE = \sum_{i=1}^{n} \frac{\varepsilon_i - E(\varepsilon)}{\sigma_\varepsilon} \tag{20}$$

where $\varepsilon = Y - \hat{Y}$ is the prediction error and $E(\varepsilon)$, σ_ε show the standart deviation of the expected values of the errors, respectively.

Table 8. The performance results for Bodyfat data set

Criteria	Significant (V)	Insignificant (V)	SAE
AIC	6	1	189.585
IBIC, SPBIC, CICOMP	3	0	183.348
BIC, ICOMPifim, ICOMPpeu, ICOMPpeuln, ICOMPperf	4	0	178.944

Table 9. The performance results for Vitamin data set

Criteria	Significant (V)	Insignificant (V)	SAE
AIC	7	2	2.08E-02
BIC, SPBIC	7	0	2.09E-02
ICOMPifim, ICOMPpeu, ICOMPpeuln, ICOMPperf	6	0	2.22E-02
IBIC, CICOMP	4	0	2.11E-02

Table 9 shows the real data analysis performance results of the information criteria for the selected beta regression models. The number of selected significant and insignificant variables are represented as Significant (V) and Insignificant (V), respectively.

The models selected by AIC are not satisfactory because of including insignificant variables and relatively higher errors. AIC seems to cause overfitting in the selected beta regression models. Most of the ICOMP-type criteria give promising results with low errors and a high number of significant variables. BIC-type alternative criteria also excluded the irrelevant variables, and the predictive errors are competitive. Although alternative ICOMP-type and BIC-type chose the different beta regression models, they tend to select only the significant variables and provide satisfactory prediction results.

6. Conclusion and Discussion

Model selection is one of the most important fields in regression analysis. The information criteria are much useful to carry out the model selection process. Within the scope of many regression models included in GLM, model selection results differ depending on the information criteria and selection mechanism. In this study, we focused on the model selection task beta regression analysis using several information criteria and a heuristic optimization approach. Firstly, we assessed the alternative information criteria on beta regression analysis for the model selection task.

We implemented the numerical examples with the simulation and real data set analysis. The simulation studies demonstrate the alternative information criteria provide better results since they can exclude the wrong variables in the selected beta regression models. Also, they tend to select correct variables as the sample size increases. Especially, we should emphasize that the alternative criteria provide satisfactory results in the presence of multicollinearity. The alternative ICOMP-type and BIC-type criteria are not affected by the multicollinearity and exclude the irrelevant variables. The increment of the sample size demonstrates this fact more obviously. The real data set examples also provide the model selection skills of the alternative criteria in both estimation and prediction results.

The different type of model selection strategies can be tried such as forward, backward and stepwise selection procedures for beta regression modeling. But it is well-known that the heuristic methods produce more efficient results. When using the information criteria inside the model selection process, the goal is to find the optimum value in the criteria so differential evolution algorithm is quite successful on the optimization process.

The information criteria that we used in the article are favorable on the model selection for beta regression. The classical criteria, especially for high multicollinearity, tend to select redundant variables. Overall, we can conclude that when the variance-covariance matrix exists as a penalty in the information criteria, it can improve the model selection results in beta regression analysis.

Acknowledgement

This study is a part of the PhD dissertation of Dünder [20] and derived from this thesis.

References

[1] D. M. Sakate, D. N. Kashid, D.T. Shirke, *Subset Selection in Poisson Regression*. Journal of Statistical Theory and Practice, 5(2) (2011) 207-219.

[2] V. Calcagno, and C. de Mazancourt, *glmulti: An R Package for Easy Automated Model Selection with (generalized) Linear Models*, Journal of Statistical Software, 34(12) (2010) 1-29.

[3] H. H. Örkcü, *Subset Selection in Multiple Linear Regression Models: A Hybrid of Genetic and Simulated Annealing Algorithms*, Applied Mathematics and Computation, 219(23) (2013) 11018-11028.

[4] A. Unler, A. Murat, *A Discrete Particle Swarm Optimization Method for Feature Selection in Binary Classification Problems*, European Journal of Operational Research, 206(3) (2010) 528-539.

[5] H. Akaike, *Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models*, Biometrika, 60(2) (1973) 255-265.

[6] G. Schwarz, *Estimating the Dimension of a Model*, The Annals of Statistics, 6(2) (1978) 461–464.

[7] H. Bozdogan, *Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions*, Psychometrika, 52(3) (1987) 345-370.

[8] C. M Hurvich, C. L. Tsai, *Regression and Time Series Model Selection in Small Samples*, Biometrika, 76(2) (1989) 297-307.

[9] H. Bozdogan, *Statistical Data Mining and Knowledge Discovery*, CRC Press, 15-56 (2004) Boca Raton.

[10] F. Cribari-Neto, A. Zeileis, *Beta Regression in R*, Journal of Statistical Software, 34(2) (2010) 1-24.

[11] K. M. Mullen, D. Ardia, D. L. Gil, D. Windover and J. Cline, *DEoptim: An R Package for Global Optimization by Differential Evolution*. Journal of Statistical Software, 40(6) (2011) 1-26.

[12] S. Ferrari, F. Cribari-Neto, *Beta Regression for Modelling Rates and Proportions*. Journal of Applied Statistics, 31(7) (2004) 799-815.

[13] W. Zhao, R. Lv, Y. Zhang, J. Liu, *Variable Selection for Varying Dispersion Beta Regression Model*, Journal of Applied Statistics, 41(1) (2014) 95-108.

[14] T. C. M. Dias and C. A. R. Diniz, *The Use of Several Link Functions on a Beta Regression Model: a Bayesian Approach*, In AIP Conference Proceedings, 1073(1) (2008) 144-149, American Institute of Physics.

[15] D. Karaboğa, *Yapay Zeka Optimizasyon Algoritmaları*, Nobel Yayınevi, (2011) 201-221, İstanbul.

[16] H. Bozdogan, *Akaike's Information Criterion and Recent Developments in Information Complexity*, Journal of Mathematical Psychology, 44(1) (2000) 62-91.

[17] E. Pamukçu, H. Bozdogan, H. and S. Çalık, *A Novel Hybrid Dimension Reduction Technique for Undersized High Dimensional Gene Expression Data Sets Using Information Complexity Criterion for Cancer Classification*. Computational and Mathematical Methods in Medicine (2015) doi:10.1155/2015/370640.

[18] K. A. Bollen, S. Ray, J. Zavisca and J. J. Harden, *A Comparison of Bayes Factor Approximation Methods Including Two New Methods*, Sociological Methods & Research, 41(2) (2012) 294-324.

[19] H. Bozdogan, *A New Class of Information Complexity (ICOMP) Criteria with an Application to Customer Profiling and Segmentation*, Istanbul University Journal of the School of Business, 39(2) (2010) 370-398.

[20] E. Dünder, *Model Selection in Beta Regression Analysis Using Heuristic Optimization Algorithms*, PhD Dissertation, Ondokuz Mayıs University (2017) Samsun, Turkey.

[21] A. Zeileis, F. Cribari-Neto, B. Gruen, I. Kosmidis, A. B. Simas, A. V. Rocha, M. A. Zeileis, (2016). Package 'betareg'.

[22] D. Ardia, K.M. Mullen, B. G. Peterson and J. Ulrich, 'DEoptim': Differential Evolution in 'R'. (2016) version 2.2-4.

[23] E. K. Koc, H. Bozdogan, *Model Selection in Multivariate Adaptive Regression Splines (MARS) Using Information Complexity as the Fitness Function*, Machine Learning, 101(1-3) (2015) 35-58.

[24] J. Verzani, Using R: Data Sets, Etc. for the Text "Using R for Introductory Statistics", Second Edition. R package version 2.0-5, (2015) https://CRAN.R-project.org/package=UsingR

[25] P. Rossi, PERregress: Regression Functions and Datasets R Package Version 1.0-8, (2013) https://CRAN.R-project.org/package=PERregress.

[26] E. Deniz, O. Akbilgic, J. A. Howe, *Model Selection Using Information Criteria Under a New Estimation Method: Least Squares Ratio*, Journal of Applied Statistics, 38(9) (2011) 2043-2050.