IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Model Selection Process in Time Series Analysis of Production System with Random Output

To cite this article: S. O. Ongbali *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **413** 012057

Model Selection Process in Time Series Analysis of Production System with Random Output

Ongbali S. O. [1], Igboanugo A. C. [2], Afolalu S. A. [1], Udo M. O. [1], Okokpujie I. P. [1]

[1]College of Engineering, Mechanical Engineering Department, Covenant University, Ota, Ogun State, Nigeria.

[2]Production Engineering Dept. faculty of Engineering, University of Benin, Edo state, Nigeria.

E-mail: samson.ongbali@covenantuniversity.edu.ng

Abstract. In time series investigation of characteristics of production system, different competing models are generally obtained particularly in production settings with stochastic output attributable to bottleneck problems. Consequently, selecting the best model that describes a production system becomes challenging and critical because some models that fit observed data most accurately may not predict future values correctly on account to model complexities. This research desires to demonstrate the procedure for model selection in production system with random output via the use of Adjusted Coefficient of Determination $(\bar{R}^2)$, Akaike and Schwarz criteria tools. Production output measurements obtained serve as input data to Autocorrelation Function and Partial Autocorrelation Function to obtain the order of Autoregressive, Autoregressive Moving Average and Autoregressive Integrated Moving Average models. The model parameters were estimated and used for predictions and compared with original and transformed data to obtain Sum of Squared Error (SSE). Afterward, the models were subjected to adequacy evaluation and subsequently tested with Akaike and Schwarz criteria. Among the competing models, ARIMA (3, 1, 1) model explain 66% variance of the dataset and wielded the lowest Akaike and Schwarz values of 534.41m and 534.34m respectively and thus selected as the model that represents the production system under investigation. The approach establishes that Adjusted Coefficient of Determination in conjunction with Akaike and Schwarz criteria are adequate tools for model selection in time series investigation particularly in stochastic situation
Keywords: Model selection, Time series, Random production output

1. Introduction

The objective of model selection is to discover a model that optimizes a process because exact model that describes a system is unknown to man. The primary objective of model selection is to compare competing models and select the best that describes a system. Model selection is a critical stage in time series investigation because we are always confronted with competing models particularly in production system with stochastic output attributable to bottleneck problems [12]. Inappropriate process of model selection results in choice of poor model with consequence. Modeling is approximation of reality, thus, model selection is to reject a model far from reality and select that which is close to reality [5]. [2] asserted that primary reason of model selection is to evaluate the performance of different models and choose the best among the models for specific dataset. Hence, overlooking model selection procedures result in worthless conclusion in statistical reasoning [29].

Researchers in the topic area over the decades have proposed different methods for model selection, for example, [2] argued that model performance is a function of its prediction ability and its selection is exceptionally imperative because it guides the choice of the quality of a selected model. Improved model performance obtained via model selection provide dependable future forecast of a system [1]. [3] claimed that aside model adequacy testing, objective of model selection include finding a good predictor that describes a system and Akaike Information Criterion (AIC) is a principal model selection method [4] and [14]. Akaike Information Criterion (AIC), Bayesian information criterion (BIC) and Structural Risk Minimization (SRM) where methods proposed by [2] for model selection. [5] derived Regularization Information Criterion (RIC) from Kullback-Leibler information number which is an extension of Takeuchi's Information Criterion (TIC) and AIC and then used it for model selection. [7] emphasized that Expert System (ES) can guide model selection for forecasting.

Similarly, [8] presented procedure for predictive testing of predicted residuals and noted that only parameter consistency is required for model selection. There are arguments whether to use model testing or model selection procedures in deciding the model that best fits a specific dataset but Arbitrage Pricing Theory (APT) model resolve the argument [9] and [20] asserted. [11] proposed the use final prediction Error (FPE) for model selection which is the expected variance of prediction error when Autoregressive is fitted. [10] used different methods for model selection including hypothesis testing, diagnostic tests, goodness-of-fit methods, Bayesian approaches and forecast evaluation methods. [13] asserted that cross-validation and Bayesian approach can be used for model selection. [15] used SURE-Autometrtcs Algorithm for Model Selection and asserted that the method performed well. [16], [18] and [24] proposed hypothesis tests and Selection Criteria using final prediction error (FPE) for model selection.

In addition, [17] proposed mean squared error of prediction as criterion for selecting model and claimed that it is better than using residual sum of squares. [19] recommended the use of single statistic as the sum of squared residuals for model selection. [21] offered unbiased Akaike Information Criterion (AICu) for model selection and claimed that it outperforms biased correction Akaike Information Criterion (AICc) but [28] claim that AICc outperform other forms of model selection tools. [23] projected penalized profile likelihood method for model selection and claimed it outperform the adaptive methods.

Furthermore, [25] develop a Cp statistic with a known distribution for model selection. [31] emphasize advantage of AIC for model selection in bootstrapping. [27] proposed AIC and BIC, Predictive Least Squares (PLS) and Sequentially Normalized Least Squares (SNLS) for model selection. [29] proposed model selection method based on multiple hypothesis testing including AIC, BIC, FPE and Minimum Description Length (MDL). [30] asserted that there are four model selection methods namely, parametric vision, data generating process (DGP), evaluation based on fit, and ignoring model uncertainty on inference and subsequently recommended semiparametric method for model selection.

The forgoing presented different methods of model selection procedures but with application to specific situations different from area we are exploring. It is evidence also that there is no unified method for model selection but some of the procedures recommended in literature are complex, laborious, time consuming and more theoretical than empirical but our approach is simple and easy to apply in practice. Section one reviews literature in the topic area, section two presents methodology, and section three depicts results and discussions and finally section draw conclusion of the study.

2. Methodology

The key variables required for model selection using Adjusted Coefficient of Determination $(\overline{R}^2)$ Akaike and Schwarz are the number of model parameters and the Sum of Squared Errors (SSE). Production output measurement obtained from a manufacturing company serve as input data to Autocorrelation function and Partial autocorrelation function computations. The order AR(p), ARMA (p, q), ARI(p, d) and ARIMA(p, d, q) models were identified and their values were estimated using OLS technique in conjunction Matlab pseudocode. The parameter values were used for prediction and compared with the original and the transformed data to obtain Sum of Squared Error (SSE). The identified order of parameters (p, d, q) along with SSEs serve as input variables to the Adjusted Coefficient of Determination $(\overline{R}^2)$, Akaike and Schwarz criteria test to select the model that represents the system under investigation

A Brief Framework of Adjusted Coefficient of Determination, Akaike and Schwarz Criteria

IOP Conf. Series: Materials Science and Engineering **413** (2018) 012057 doi:10.1088/1757-899X/413/1/012057

Adequacy of a model is evaluated by using Adjusted Coefficient of Determination $\overline{R}^2$ which gives idea of how many data points fall within the regression line to examine the relationship within a dataset. Adjusted coefficient of determination $\overline{R}^2$ shows the percentage of variation explained by the independent variable that affects the dependent variable.

Equation for Adjusted Coefficient of Determination is given by,

$$\overline{R}^2 = 1 - \frac{\sum_1^n (y_t - \hat{y})^2 / (n-k)}{\sum_1^n (y_t - \overline{y})^2 / (n-1)} \quad (1)$$

Were, k = number of parameters

n = number of dependent variables.

Equation for Akaike and Schwarz criteria are given respectively by,

Akaike (AIC) = $n \log(SSE) + 2k$ (2)

Schwarz $BIC = n \log(SSE) + k \log n$ (3)

Where, k = number of parameters that are fitted in the model

Log = natural logarithm

n = number of observations in the series

SSE = sum of squared errors, given as, $SSE = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$

3. Results

Four models – AR (4), ARMA (4, 3) ARI (3, 1) and ARIMA (3, 1, 1) were obtained from analysis and were subjected to model adequacy evaluation and subsequently model selection.

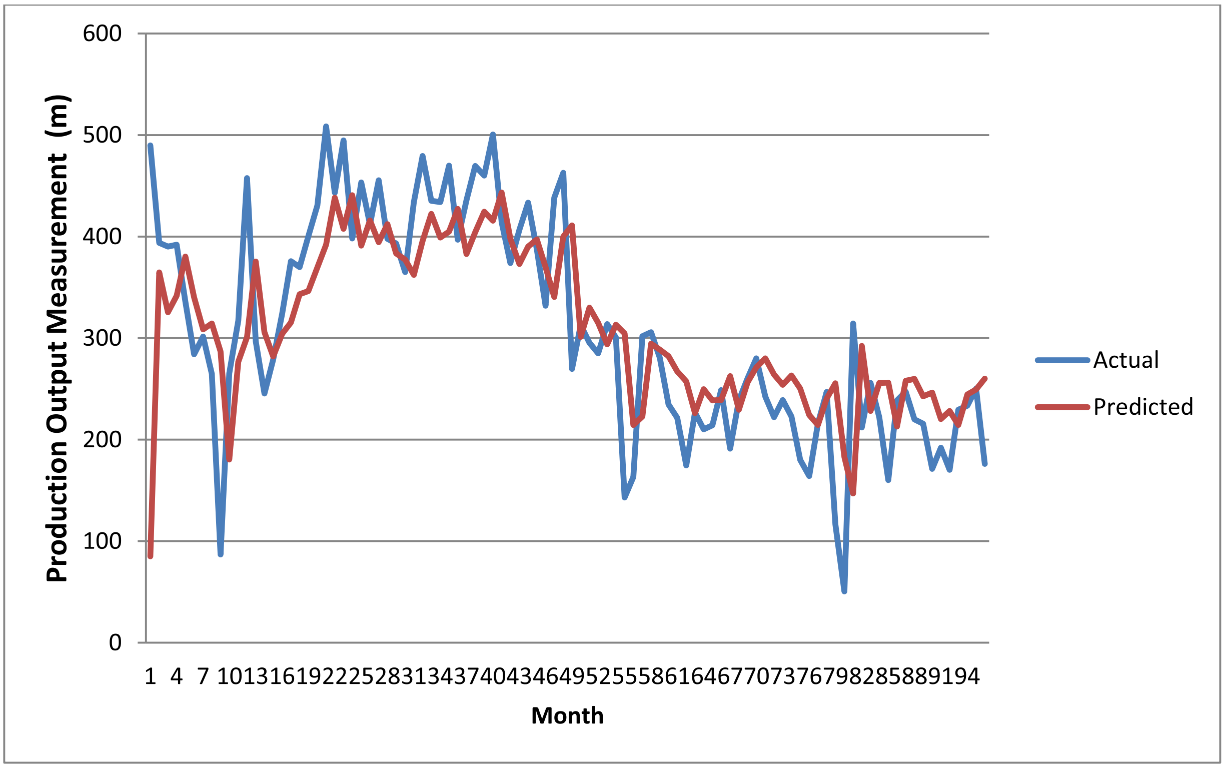

Figure 1. AR (4) model prediction superimpose on actual data with SSE = 551610.804m.

IOP Conf. Series: Materials Science and Engineering **413** (2018) 012057 doi:10.1088/1757-899X/413/1/012057

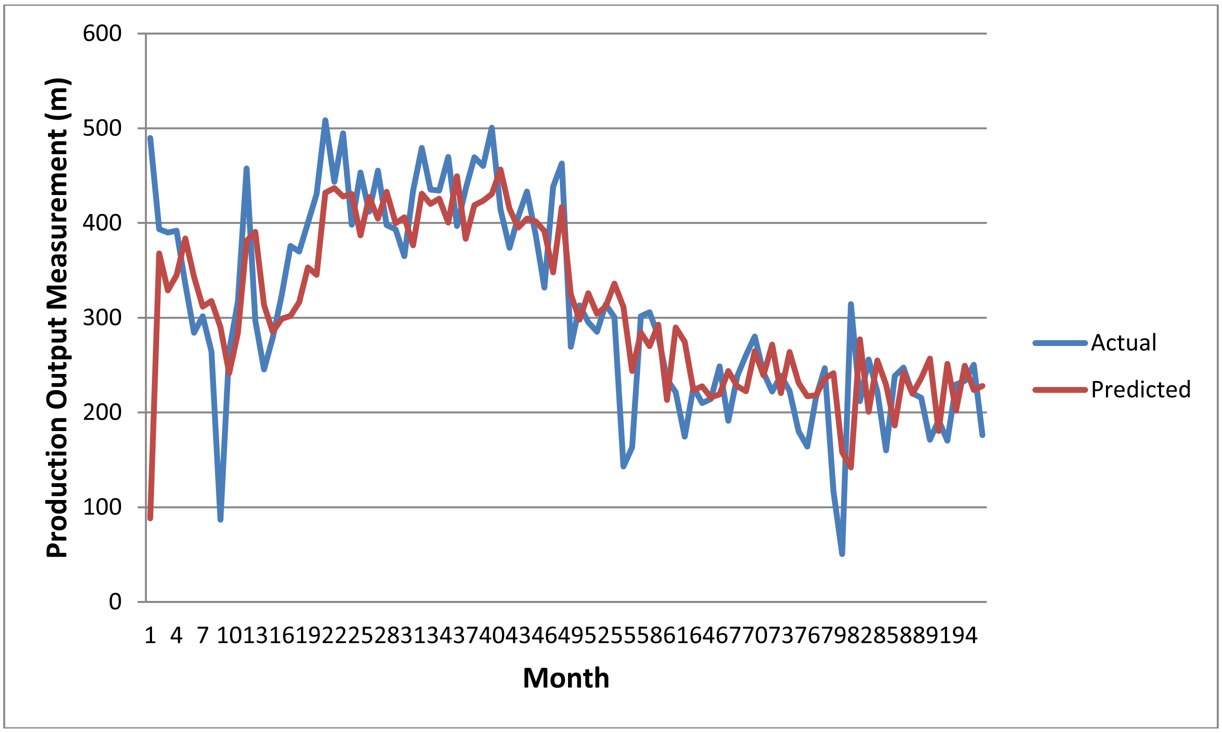

Figure 2. ARMA (4, 3) model prediction superimpose on actual data with SSE = 487623.122m

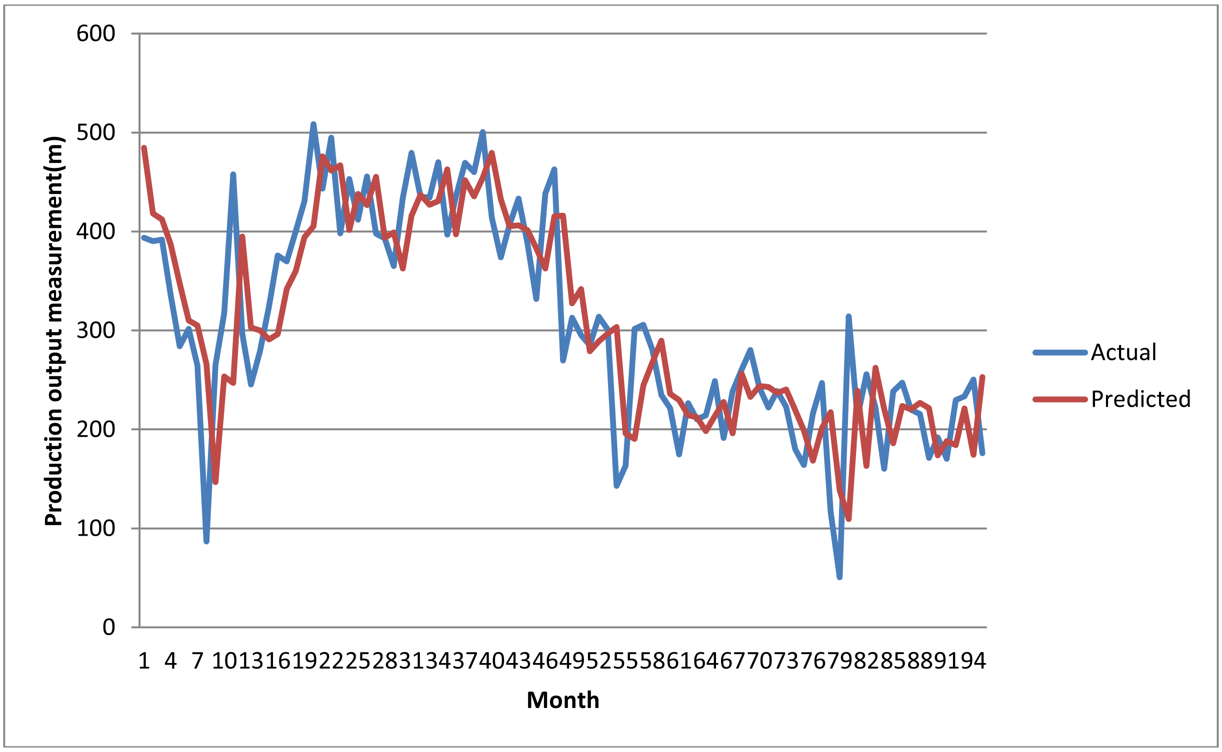

Figure 3. ARI (3, 1) Model prediction superimpose on transformed data

Predicted values in figure 3 tracked the transformed data closely at some points with SSE = 385822.75m

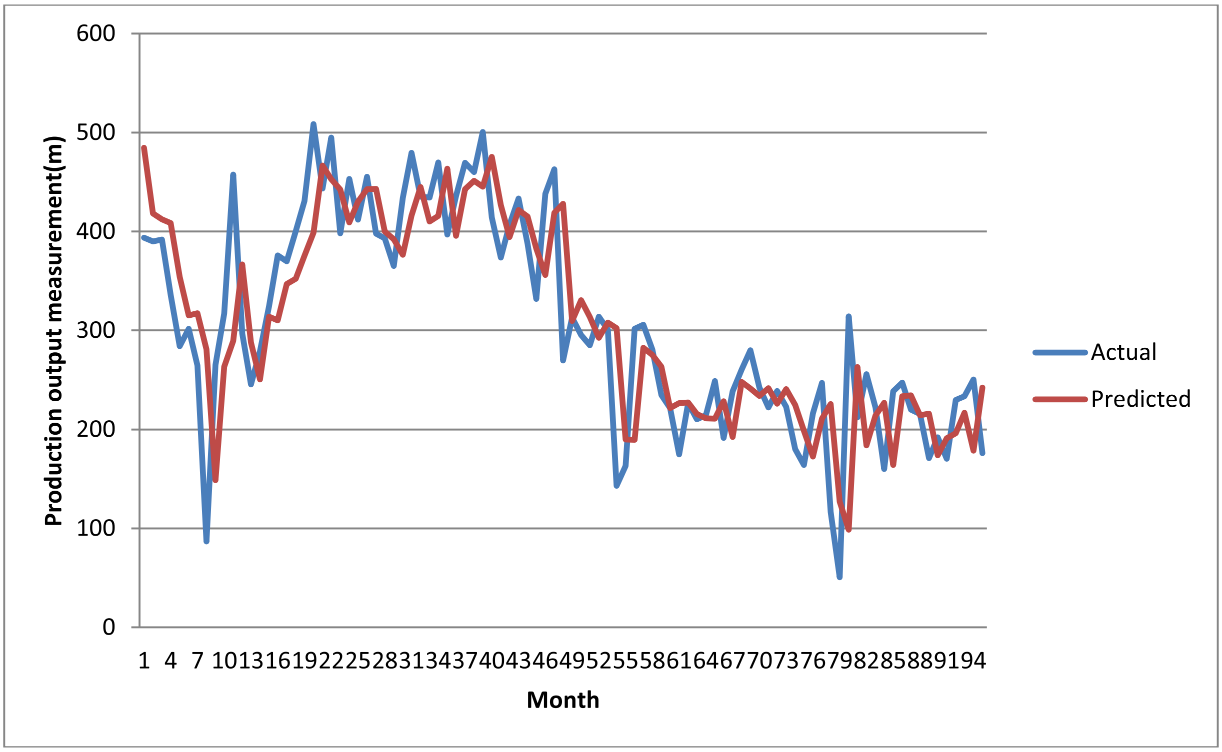

Figure 4 ARIMA (3, 1, 1) Model prediction superimpose on transform data

The predicted values in figure 4 tracked the transformed data closer than ARI (3, 1) model with SSE = 364910.024m

Table 1. Evaluation of Model Adequacy Using Adjusted Coefficient of Determination $(\bar{R}^2)$

Model	Adjusted Coefficient of Determination $(\bar{R}^2)$
AR (4)	52%
ARMA (4, 3)	54%
ARI (3, 1)	62%
ARIMA (3, 1, 1)	66%

ARMA(3, 1, 1) model in table 1 outperformed other competing models as indicated in table 1 by wielding $(\bar{R}^2)$ value of 66% signifying that the model explains 66% of the variance of the independent variable that affects the dependent variables

Table 2. Akaike and Schwarz Criteria Values

Model	Akaike (AIC)	Schwarz (BIC)
AR (4)	559.20	559.13
ARMA (4, 3)	554.056	553.985
ARI (3, 1)	536.71	536.64
ARIMA (3, 1, 1)	534.41	534.34

The AIC and BIC values in table 2 show that ARIMA (3, 1, 1) model surpasses other models by wielding the lowest Akaike and Schwarz values of 534.41 and 534.34 respectively. The Adjusted Coefficient of Determination, Akaike and Schwarz values support ARIMA (3, 1, 1) model to be the best model that described the production system under investigation. Thus, ARIMA (3, 1, 1) is selected and fitted to the time series data analysed. Consequently, ARIMA (3, 1, 1) model can be fitted to the production system and use for out-of-sample prediction to approximate future production output.

4. Conclusion

Selecting the right model leads to accurate prediction and guarantee sustainable production system. The foregoing demonstrates that Adjusted Coefficient of Determination in conjunction with Akaike and Schwarz criteria are powerful tools for model selection particularly in production setting with random output. Most times, Adjusted Coefficient of Determination points out the appropriate model to be fitted to a data set and usually supported by Akaike and Schwarz criteria for final selection as demonstrated in this study.

Acknowledgements

We acknowledge the financial support offered by Covenant University for actualization of this research work for publication.

References

[1] Eija Ferreira. Model Selection in Time Series Machine Learning Applications [D]. Academic dissertation of Technology and Natural Sciences of the University of Oulu, Linnanmaa, 2015.
[2] Hastie T., Tibshirani R. and Friedman J. The Elements of Statistical Learning: Data Mining [J]. Inference and Prediction. Springer Series in Statistics. Springer-Verlag, New York, NY, 2nd edition, 2009.
[3] Hannan E.J., Krishanaiah P. R. and Rao M. M. Various Model Selection Techniques in Time Series Analysis[J]. Eds., Handbook of Statistics, 1985, Vol. 5,
[4] Clifford M. Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples [J]. Biometrika, 1989, 76, 2, pp. 297-307
[5] Shibata R. Statistical Aspects of Model Selection. Working Paper [C]. International Institute for Applied Systems Analysis, 1989, A-2361 Laxenburg, Austria
[6] Markus B., Wolfgang A., Peter F., Tim L., Silvia M., and Alexander R. Visual Analytics for Model Selection in Time Series Analysis [J]. IEEE Computer Society, 2013
[7] Delurgio Stephen A (1998). Forecasting Principles and Applications [T]. Irwin McGraw-Hill New York.
[8] Dufour J, Ghysels E. and Hall A. Generalized Predictive Tests and Structural Change Analysis in Econometrics [J]. International Economic Review, 1994, Vol. 35, No. 1, pp. 199-229
[9] Bharti M. I., Al-Shanfari H and Hossain M. Z. Econometric Analysis of Model Selection and Model Testing [J]. Journal of the American Statistical Association, 2002
[10] Dufour J-M. Model selection. [T] New Palgrave Dictionary of Economics, 2nd Edition, 2007.
[11] Akaike H. Fitting autoregressive models for prediction [J]. Annals of the Institute of Statistical Mathematics, 1969, 21, 243–247.
[12] Igboanugo A. C. and Ongbali S. O. A Factorial Study of Bottleneck Problems in Multi-Stage Production System: Results from Textile Industry [J]. University of Benin Journal of Science & Technology, 2012, Vol.1, No. 1
[13] Burnham K. P. and Anderson D. R. Model Selection and Multimodel Inference A Practical Information-Theoretic Approach [J]. Second Edition. (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, 2002, NY 10010, USA),
[14] Burnham K. P., Anderson D. R. and Huyvaert K. P. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons [J]. Behav Ecol Sociobiol, 2011, 65:23–35
[15] Norhayati Y. SURE-Autometrtcs Algorithm for Model Selection in Multiple Equations [D]. PhD Thesis, Universiti, Utara, Malaysia, 2016.

IOP Conf. Series: Materials Science and Engineering **413** (2018) 012057 doi:10.1088/1757-899X/413/1/012057

[16] Hannes L. and Benedikt M. P. The Model Selection [J]. Yale University, 2000.
[17] **Allen D. M.** Mean Square Error of Prediction as a Criterion for Selecting Variables [J]. Technometrics, 2000, Volume 13, 1971 - Issue 3
[18] Allen D. M. The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction [J]. Technometrics, 1974, Vol. 16, No. 1, pp. 125-127
[19] Amemiya T. Selection of Regressors [J]. International Economic Review, 1980, 21(2):331-54
[20] Bhatti M. I. and Al-Shanfari H. Economic Analysis of Model Selection and Model Testing [J]. 2006, Routledge
[21] McQuarrie A., Shumway R., and Tsai C-L. The model selection criterion AICu [J]. Statistics & Probability Letters, 1998, Vol. 34, Issue
[22] McQuarrie A. D. R. and Tsai C-L "Regression and Time Series Model Selection," [C] World Scientific, Singapore City, (2000).
[23] Chen Z., Tang M-L., Gao W and Shi N-Z. New Robust Variable Selection Methods for Linear Regression Models [J]. Scandinavian Journal of Statistics, 2014.
[24] Liebscher E A. Universal Selection Method in Linear Regression Model [J]. Open Journal of Statistics, 2012, Vol.2, No. 2
[25] Riani M. and Atkinson A. C. The Selection of Time Series Models Perhaps with Regressors [J]. London School of Economics, London WC2A 2AE, UK, 2010.
[26] Berkowitz J. and Kilian L. Recent Developments in Bootstrapping Time Series [J]. Econometrics Reviews, 2000, 19(1) 1 – 48
[27] Maatta J. Model Selection Methods for Linear Regression and Phylogenetic Reconstruction [D]. PhD Thesis, Series of Publications A, Report A-2016-3, 2016.
[28] Hurvich C. M and Tsai C-L. Regression and Time Series Model Selection in Small Samples [J] Biometrika, 1989, 76, 2, pp. 297-307
[29] Leeb h. and Pötscher B. M. Model Selection and Inference: Facts and Fiction [J]. Econometric Theory, 2005, 21, 21-59
[30] Hansen B. E. Challenges for Econometric Model Selection [J]. Econometric Theory, 21, 2005.