# Model Selection in Subset Polynomial Regression by Using Bootstrap Algorithm

*By* Suparman

International Journal of Advances in Intelligent Informatics
Vol. 4, No 2, July 2018, pp. xx-xx
ISSN: 2442-6571


# Model Selection in Subset Polynomial Regression by Using Bootstrap Algorithm

Suparman[a,1], Mohd Saifullah Rusiman[b,2]

[a] University of Ahmad Dahlan, Jalan Ringroad Selatan, Yogyakarta 55191, Indonesia
[b] Universiti Tun Hussein Onn Malaysia, Pagoh, Malaysia
[1] suparman@pmat.uad.ac.id*, [2] saifulah@uthm.edu.my
* corresponding author

ARTICLE INFO

*Article history*:
Received
Revised
Accepted

*Keywords*:
Bootstrap Algorithm
Subset Polynomial Regression
Model Selection

ABSTRACT

The subset polynomial regression model is a wider model than the polynomial regression model. In the subset polynomial regression model, the error is generally assumed to be normally distributed and the polynomial order is known. This study proposes an estimate of the parameters of the subset polynomial regression model in which the error is assumed to have any distribution and the polynomial order is unknown. The Bootstrap method is used to estimate the parameters of the subset polynomial regression model. Simulated data is used to test the performance of the Bootstrap method. The test results show that the bootstrap method can estimate well the parameters of the subset polynomial regression model. The bootstrap method is then implemented with real data.



## I. Introduction

The subset polynomial regression model is a polynomial regression in which some regression coefficients have a zero value. The advantage of this subset polynomial regression model is the user can select a regression model from all possible subsets of the polynomial regression model. The subset polynomial regression model has been studied by several authors. Jakabson and Lavendels [1] compared the formation of polynomial regression models using the subset selection approach and the adaptive basis function construction approach. In the subset selection approach, the least squares method is used. Overall the adaptive basis function construction approach was found to be superior to the subset selection approach. O'Neill et al. [2] used the method of a subset polynomial neural network to predict breast cancer. This method gives better results than the mammography method. Xie et al. [3] used the polynomial regression in medical image segmentaton. Suparman [4] proposed a subset polynomial regression model using error which has an exponential distribution. The Markov Chain Monte Carlo (MCMC) reversible jump method is used to estimate the parameters of the subset polynomial regression model. The subset polynomial regression model often assumes that the error has a normal distribution or exponential distribution. However, in everyday life it is often found that the distribution of the error is unknown.

The Bootstrap method developed in [5] is widely used in statistics and can be very useful in the context of regression [6]. A principle of the Bootstrap method is to try to get a good estimate based on minimal resources. In the case of statistical inference, minimal resources can be interpreted as little data, data that deviate from certain assumptions, and data that have no assumption about the distribution. Warton [6] used the bootstrap algorithm to estimate the parameters of a regression model. The Bootstrap method is applied in ecology. Garcia-Soidan et al. [7] used the Bootstrap method for spatial data. The estimator of the multivariate distribution function is used as the basis for the implementation of the Bootstrap method. Yazici et al. [8] used the Bootstrap method to obtain the empirical distribution of the parameters in the nonparametric regression of Conic Multivariate Adaptive Regression Splines (CMARS). The results showed that the bootstrap method provides an accurate parameter estimate. Beda et al. [9] used the Bootstrap method to calculate the confidence limits for spectral indices of heart-rate variability (HRV). Spectral indices are modeled


DOI:

using an autoregressive model. Hall and Maiti [10] used the Bootstrap method to construct a mean error estimator and to calculate the predicted region. The Bootstrap technique can be applied to non-normal models. Colugnati et al. [11] used the Bootstrap method to obtain interval estimates for percentiles on the diagnosis of obesity and overweight in children and adolescents. Kant et al. [12] used a bootstrap-based neural network model for flood estimates. The results show that the bootstrap-based neural network model is a stable model. Ren et al. [13] used the Bootstrap method to determine the confidence interval for multihop distances. The use of Bootstrap method can eliminate the risk of small sample size and unknown distribution. Kleiner et al. [14] used Bootstrap for massive data. Jacek et al. [15] used the Bootstrap approach to estimate the uncertainty of surface response models. Chen et al. [16] used a bootstrap analysis to measure individual and regional differences in relative concentrations of gamma-aminobutyric acid in the human brain. Dongping [17] used the Bootstrap method to determine predictive point and prediction intervals to reduce the risk of misleading decisions in maintenance in prognostic devices. Liang et al. [18] used the Bootstrap Metropolis-Hasting algorithm for model selection and optimization. Mei et al. [19] used the Residual-Based Bootstrap Test to detect the constant coefficients in the Weighted Geographic Regression model. Mikshowsky et al. [20] used bootstrap aggregation sampling to improve the reliability of genomic predictions for Jersey sires. Olaniran et al. [21] used Bootstrap techniques to improve the selection and classification of Bayesian features. Zhen [22] used Bootstrap resampling to detect wideband signal number. Boubaka et al. [23] used the Bootstrap method to identify parameters for the dependent data. In this paper, the Bootstrap method will be used to determine the parameter estimator in the polynomial subset regression.

Suppose that $(y_t, x_t)$ is the pairing of the dependent variable and the independent variable, as well as $z_t$ is error and t = 1, 2, ....n where n is the number of observation. Let $k_{max}$ be a maximum order. The subset polynomial regression model which has an order k (k = 0, 1, ...., $k_{max}$) can be written as:

$$y_t = \beta_0 + \beta_{n_1} x_t^{n_1} + \beta_{n_2} x_t^{n_2} + \ldots + \beta_{n_k} x_t^{n_k} + z_t \qquad (1)$$

Here $\{n_1, n_2, ..., n_k\}$ is the subset of $\{1, 2, ..., k\}$ and $\beta = (\beta_0, \beta_{n_1}, \ldots, \beta_{n_k})'$ is the coefficient vector. The $z_t$ (t = 1, 2, 3, ..., n ) is an error with mean 0 and variance $\sigma^2$ that is identical but its distribution is unknown. Based on the data $(y_t, x_t)$ for t = 1, 2, ..., n, the parameters $\beta$, $\sigma^2$ and the polynomial regression subset models are estimated. This paper aims to estimate the parameters of the subset polynomial regression model using the Bootstrap method.

## II. Method

The method used to estimate the parameters of the subset polynomial regression model is as follows:

### *A. The Least Squares Estimate*

Equation (1) is a short form for a set of the following n simultaneous equations:

$$\begin{aligned} y_1 &= \beta_0 + \beta_{n_1} x_1^{n_1} + \beta_{n_2} x_1^{n_2} + \ldots + \beta_{n_k} x_1^{n_k} + z_1 \\ y_2 &= \beta_0 + \beta_{n_1} x_2^{n_1} + \beta_{n_2} x_2^{n_2} + \ldots + \beta_{n_k} x_2^{n_k} + z_2 \\ &\ldots \\ y_n &= \beta_0 + \beta_{n_1} x_n^{n_1} + \beta_{n_2} x_n^{n_2} + \ldots + \beta_{n_k} x_n^{n_k} + z_n \end{aligned} \qquad (2)$$

In matrix form, equation (2) can be written as:

$$Y = X\beta + Z \qquad (3)$$

where

$$Y=\begin{bmatrix}Y_1\\Y_2\\\cdots\\Y_n\end{bmatrix},\ X=\begin{bmatrix}1 & X_1^{n_1} & X_1^{n_2} & \cdots & X_1^{n_k}\\1 & X_2^{n_1} & X_2^{n_2} & \cdots & X_2^{n_k}\\\cdots & \cdots & \cdots & & \cdots\\1 & X_n^{n_1} & X_n^{n_2} & \cdots & X_n^{n_k}\end{bmatrix},\ \beta=\begin{bmatrix}\beta_0\\\beta_{n_1}\\\cdots\\\beta_{n_k}\end{bmatrix},\text{ and } Z=\begin{bmatrix}Z_1\\Z_2\\\cdots\\Z_n\end{bmatrix}.$$

To obtain the least squares estimate of $\beta$, first write the sample subset polynomial regression:

$$y_t=\hat{\beta}_0+\hat{\beta}_{n_1}x_t^{n_1}+\hat{\beta}_{n_2}x_t^{n_2}+\ldots+\hat{\beta}_{n_k}x_t^{n_k}+z_t \quad (4)$$

for t = 1, 2, 3, ..., n, which can be written briefly in matrix notation as:

$$Y=X\hat{\beta}+e \quad (5)$$

where

$$Y=\begin{bmatrix}Y_1\\Y_2\\\cdots\\Y_n\end{bmatrix},\ X=\begin{bmatrix}1 & X_1^{n_1} & X_1^{n_2} & \cdots & X_1^{n_k}\\1 & X_2^{n_1} & X_2^{n_2} & \cdots & X_2^{n_k}\\\cdots & \cdots & \cdots & & \cdots\\1 & X_n^{n_1} & X_n^{n_2} & \cdots & X_n^{n_k}\end{bmatrix},\ \hat{\beta}=\begin{bmatrix}\hat{\beta}_0\\\hat{\beta}_{n_1}\\\cdots\\\hat{\beta}_{n_k}\end{bmatrix},\text{ and } e=\begin{bmatrix}e_1\\e_2\\\cdots\\e_n\end{bmatrix}.$$

Here, $\hat{\beta}$ is a column vector of the least squares estimator of the subset polynomial regression coefficient and e is a column vector of the residual n. According to the least squares method, the least squares estimator is obtained by minimizing

$$\sum_{t=1}^{n}e_t^2=\sum_{t=1}^{n}(y_t-\hat{\beta}_0-\hat{\beta}_{n_1}X_t^{n_1}-\ldots-\hat{\beta}_{n_k}X_t^{n_k})^2 \quad (6)$$

This is achieved by partially differentiating (6) to $\beta_0,\beta_{n_1},\ldots,\beta_{n_k}$ and the result obtained is equal to zero. This process produces k + 1 simultaneous equations in k + 1 unknown variables.

$$n\hat{\beta}_0+\hat{\beta}_{n_1}\sum_{t=1}^{n}X_t^{n_1}+\hat{\beta}_{n_2}\sum_{t=1}^{n}X_t^{n_2}+\ldots+\hat{\beta}_{n_k}\sum_{t=1}^{n}X_t^{n_k}=\sum_{t=1}^{n}y_t$$

$$\hat{\beta}_0\sum_{t=1}^{n}X_t^{n_1}+\hat{\beta}_{n_1}\sum_{t=1}^{n}X_t^{2n_1}+\hat{\beta}_{n_2}\sum_{t=1}^{n}X_t^{n_1}X_t^{n_2}+\ldots+\hat{\beta}_{n_k}\sum_{t=1}^{n}X_t^{n_1}X_t^{n_k}=\sum_{t=1}^{n}x_t^{n_1}y_t$$

$$\hat{\beta}_0\sum_{t=1}^{n}X_t^{n_2}+\beta_{n_1}\sum_{t=1}^{n}X_t^{n_2}X_t^{n_1}+\hat{\beta}_{n_2}\sum_{t=1}^{n}X_t^{2n_2}+\ldots+\hat{\beta}_{n_k}\sum_{t=1}^{n}X_t^{n_2}X_t^{n_k}=\sum_{t=1}^{n}x_t^{n_2}y_t \quad (7)$$

...

$$\hat{\beta}_0\sum_{t=1}^{n}X_t^{n_k}+\beta_{n_1}\sum_{t=1}^{n}X_t^{n_k}X_t^{n_1}+\hat{\beta}_{n_2}\sum_{t=1}^{n}X_t^{n_k}X_t^{n_2}+\ldots+\hat{\beta}_{n_k}\sum_{t=1}^{n}X_t^{2n_k}=\sum_{t=1}^{n}x_t^{n_k}y_t$$

In matrix form, equation (7) can be presented as:

$$(X'X)\hat{\beta}=X'Y \quad (8)$$

If the inverse of (X'X) exists, say $(X'X)^{-1}$, then by multiplying in both sides of (8) by this inverse, the result is as follows.

$$(X'X)^{-1}(X'X)\hat{\beta}=(X'X)^{-1}X'Y$$

or

$$\hat{\beta}=(X'X)^{-1}X'Y$$

The least squares estimator for $\beta = (\beta_0, \beta_{n_1}, \ldots, \beta_{n_k})'$ is

$$\hat{\beta} = (X'X)^{-1}X'Y$$

And the least squares estimator for $\sigma^2$ is:

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\beta}'X'Y}{n-k-1}$$

*B. Statistical Criteria*

The $C_k$ statistical criteria [5] is used to select the best polynomial subset regression model. The best subset polynomial regression model selected is a subset polynomial regression model that has the smallest $C_k$ value. The $C_k$ value is calculated using the following equation:

$$C_k = \frac{\sum_{t=1}^{n}(y_t - \hat{\beta}_0 - \hat{\beta}_{n_1}x_t^{n_1} - \ldots - \hat{\beta}_{n_k}x_t^{n_k})^2}{n} - \frac{2k\hat{\sigma}^2}{n}$$

*C. Bootstrap Method*

The Bootstrap method developed in [5] is a simulation method based on data that can be applied to statistical inference problems. A basic principle of bootstrapping is resampling i.e. resampling / artificial observation of $z_1, z_2, \ldots, z_n$ that already exists.

Let $\hat{F}$ be an empirical distribution taken with probability 1/n at each observed value $z_1, z_2, \ldots, z_n$. Let B be a number of the resampling. The Bootstrap sample is defined as a random sample of size n composed of $\hat{F}$, e.g. the b[th] Bootstrap sample (b = 1, 2, ... , B) is denoted by $z_1^b, z_2^b, \ldots, z_n^b$. The Bootstrap sample $z_1^b, z_2^b, \ldots, z_n^b$ is a random sample of size n taken with the return of population $z_1, z_2, \ldots, z_n$. Members of the bootstrap sample $z_1^b, z_2^b, \ldots, z_n^b$ comprising the original samples $z_1, z_2, \ldots, z_n$, appear once, appear twice, appear more than twice, or do not appear in the original sampling process. The computational steps to determine the 100(1-α)% confidence interval for $\hat{y}_{t+1}$ are as follows:

- Calculate $\hat{\beta}$ and $\hat{\sigma}^2$ from the original data.
- Calculate $\hat{z}_t$ using the equation

$$\hat{z}_t = \hat{y}_t - \hat{\beta}_0 - \hat{\beta}_{n_1}x_t^{n_1} - \ldots - \hat{\beta}_{n_k}x_t^{n_k}.$$

- For b = 1, 2, ...., B:
  - Resampling $\hat{z}_t^{(b)}$.
  - Compute $\hat{y}_t^{(b)}$ with the equation

$$y_t^{(b)} = \beta_0 + \beta_{n_1}x_t^{n_1} + \ldots + \beta_{n_k}x_t^{n_k} + \hat{z}_t^{(b)}.$$

  - Compute $\hat{\beta}^{(b)}$, $\hat{\sigma}^{2(b)}$ and $\hat{y}_{t+1}^{(b)}$.
- Compute $\hat{\beta}_{boot}$, $\hat{\sigma}_{boot}^2$ and $\hat{y}_{(t+1)(boot)}^{(b)}$.
- Calculate the 100(1-α)% confidence interval for $\hat{y}_{t+1}$

## III. Results and Discussion

As an illustration, we apply the Bootstrap algorithm to determine the prediction interval in simulated data (simulation study) and real data (case study). A simulation study was undertaken to confirm the performance of the bootstrap algorithm whether it works properly. Case studies are given to provide examples of the application of research in solving problems in everyday life. Here resampling is done as much as $B = 2000$ and $\alpha = 0.05$.

### A. *Simulated Data*

Figure 1 shows a graph of 1000 synthesis data of the subset polynomial regression model with order 2. The value of x is determined but the value of y is made using equation (1). The values of regression coefficients and the variance of error are $\beta_0 = 1$ and $\beta_2 = 0.5$ and $\sigma^2 = 9$.

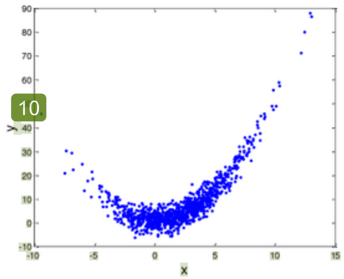

Figure 1: Simulated data

The simulated data in Figure 1 are matched against the subset polynomial regression model. Here $k_{max}$ = 2. The Bootstrap algorithm is used to estimate the best subset polynomial regression model, the subset polynomial regression coefficient and the variance $\sigma^2$. Estimation of the subset polynomial regression model is done by looking at $C_k$ statistical value for the three regression models of the subset polynomial. The $C_k$ statistical value for the three regression models of the subset polynomial can be seen in Table 1.

Table 1: The $C_k$ statistical value

| Subset Polynomial Regression Model with Order 2 | $C_k$ Statistical Value |
|---|---|
| $y = \beta_0 + \beta_1 x$ | 62.5997 |
| $y = \beta_0 + \beta_2 x^2$ | 9.1938 |
| $y = \beta_0 + \beta_1 x + \beta_2 x^2$ | 9.2170 |

From Table 1 it can be seen that the smallest $C_k$ statistical value is achieved by the second subset polynomial regression model. Thus, the second regression is the best subset polynomial regression model. Based on the regression of the best subset polynomial, then the parameters of the corresponding subset polynomial regression model are estimated using the least squares method. The results are $\hat{\beta}_0 = 0.9323$ and $\hat{\beta}_2 = 0.5070$ and $\hat{\sigma}^2 = 9.1756$. If the parameter values and estimator values of both regression and variance coefficients are compared then it appears that the Bootstrap algorithm can work well in estimating parameters based on synthesis data. Prediction for the value of $y_{1000}$ if x = 16.4176 is 9.2569 and the corresponding 95% confidence interval is (9.0984, 9.4117).

### B. *Real Data*

Table 2 shows the business tendency index (y) and the consumer tendency index (x) in the second quarter of 2000 to the fourth quarter of 2009.

Table 2 : The business tendency index (BTI) and the consumer tendency index (CTI).
Source: http://www.bps.go.id

| **Year** | **Quarter** | **BTI** | **CTI** |
|---|---|---|---|
| 2000 | II | 122.50 | 113.29 |
| | III | 117.44 | 108.04 |
| | IV | 116.06 | 114.23 |

| | | | |
|---|---|---|---|
| 2001 | I | 107.73 | 110.52 |
| | II | 111.75 | 104.10 |
| | III | 105.36 | 119.21 |
| | IV | 101.03 | 125.19 |
| 2002 | I | 100.03 | 113.75 |
| | II | 113.38 | 116.65 |
| | III | 108.77 | 119.96 |
| | IV | 102.37 | 120.28 |
| 2003 | I | 95.78 | 105.87 |
| | II | 105.16 | 117.28 |
| | III | 111.41 | 114.17 |
| | IV | 114.13 | 121.73 |
| 2004 | I | 104.35 | 115.20 |
| | II | 113.74 | 112.30 |
| | III | 114.12 | 120.22 |
| | IV | 115.03 | 109.96 |
| 2005 | I | 98.93 | 96.72 |
| | II | 106.31 | 98.68 |
| | III | 105.7 | 93.20 |
| | IV | 98.45 | 94.43 |
| 2006 | I | 95.12 | 96.01 |
| | II | 108.5 | 109.77 |
| | III | 108.72 | 109.16 |
| | IV | 107.43 | 106.96 |
| 2007 | I | 100.19 | 106.93 |
| | II | 110.96 | 105.78 |
| | III | 112.58 | 109.48 |
| | IV | 112.25 | 106.10 |
| 2008 | I | 104.41 | 95.01 |
| | II | 111.72 | 93.84 |
| | III | 111.12 | 102.78 |
| | IV | 102.19 | 100.93 |
| 2009 | I | 96.91 | 102.15 |
| | II | 110.43 | 106.42 |
| | III | 112.86 | 107.79 |
| | IV | 108.45 | 108.76 |

The data in Table 2 are matched against the subset polynomial regression model. Here $k_{max} = 3$. The bootstrap algorithm was used to obtain the subset polynomial regression model, the regression model parameters and the variance $\sigma^2$. Estimation of subset polynomial regression model is done by looking at the statistical value of $C_k$ for 7 models.

From Table 3 it can be seen that the smallest $C_k$ statistical value is achieved by the 4th polynomial subset regression model. Thus, the fourth subset polynomial regression model is the best subset polynomial regression model.

Table 3: The $C_k$ statistical value

| Subset Polynomial Regression Model with Order 3 | $C_k$ Statistical Value |
|---|---|
| $y = \beta_0 + \beta_1 x$ | 37.8193 |
| $y = \beta_0 + \beta_2 x^2$ | 38.3861 |
| $y = \beta_0 + \beta_3 x^3$ | 38.4738 |
| $y = \beta_0 + \beta_1 x + \beta_2 x^2$ | 35.9786 |

| | |
|---|---|
| $y=\beta_0+\beta_1x+\beta_3x^3$ | 36.9467 |
| $y=\beta_0+\beta_2x^2+\beta_3x^3$ | 36.8494 |
| $y=\beta_0+\beta_1x+\beta_2x^2+\beta_3x^3$ | 37.6857 |

Based on this subset best polynomial regression model, then the parameters of the corresponding subset polynomial model are estimated. The results are $\hat{\beta}_0=-189.1774$, $\hat{\beta}_1=5.2858$, $\hat{\beta}_2=-0.0234$ and $\hat{\sigma}^2=32.6954$. The prediction for $y_{41}$ if x = 108.76 is 108.7878 and the 95% confidence interval for $y_{41}$ if x = 108.76 is (106.8255, 110.7612).

## IV. Conclusion

The paper showed how the Bootstrap algorithm can be used to generate parameter estimates in polynomial subset regression model and determine the prediction interval for the dependent variable in the polynomial subset regression model if the error has any distribution. The simulation results showed that the Bootstrap algorithm could estimate well the parameters and determine the prediction interval.

The data of business tendency index (y) and consumer tendency index (x) in the second quarter of 2000 up to the fourth quarter of 2009 were matched against a subset polynomial regression model. The Bootstrap algorithm obtained the subset polynomial regression model as follows:

$$y=\hat{\beta}_0+\hat{\beta}_1x+\hat{\beta}_2x^2$$

where $\hat{\beta}_0=-189.1774$, $\hat{\beta}_1=5.2858$, and $\hat{\beta}_2=-0.0234$. This subset polynomial regression model is very useful for decision making, for example to predict the value or calculate the prediction interval of variable y in the future. This Bootstrap algorithm can be applied also to other models of econometrics.

## Acknowledgment

The authors would like to thank Assoc. Prof. Allan Leslie White, Ph.D, University of Western Sydney, for his suggestions to improve this paper.

## References

[1] G. Jekabsons and J. Lavendels, "A Comparison of Subset Selection and Adaptive Basis Function Construction for Polynomial Regression Model Building," *Computer Sciences*, vol. 38, no. 38, pp. 187-197, 2009.

[2] T. O'Neill, J.Penm and J.S. Penm, "A Subset Polynomial Neural Networks Approach for Breast Cancer Diagnosis," *International Journal of Electronic Healtheare*, vol. 3, no. 3, pp. 293-302. 2007.

[3] C.H. Xie, Y.J. Liu and J.Y. Chang, "Medical Image Segmentation Using Rough Set and Local Polynomial Regression," *Multimedia Tools and Applications*, vol. 74, no. 6, pp. 1885-1914, 2015.

[4] Suparman, "Selection in Subset Polynomial Regression by Using Reversible Jump MCMC," *Journal of Environmetal Science, Computer Science and Engineering & Technology*, vol. 5, no. 3, pp. 214-219, 2016.

[5] B. Efron and R. Tibshirani, An Introduction to the Bootstrap. Chapman & Hall : New York, 1993.

[6] D. Warton, L. Thibaut and Y.A. Wang, "The PIT-trap-A "model-tree" Bootstrap Procedure for Inference about Regression Models with Discrete," Multivariate Responses, *PloS ONE*, vol. 12, no. 7, pp. 1-18, 2017.

[7] P. Carcia-Soidan, R. Menezes and O. Rubinos, "Bootstrap Approaches for Spatial Data," *Stochastic Environmental Research & Risk Assessment*, vol. 28, no. 5, pp. 1207-1219, 2014.

[8] C. Yazici, I. Batnaz and F. Yerlikaya-Ozkurt, "A Computational Approach to Nonparametric Regression : Bootstrapping CMARS Method," *Machine Learning*, vol. 101, no. 1, pp. 211-230, 2015.

[9] A. Beda, D.M. Simpson and L. Faes, "Estimation of Confidence Limits for Descriptive Indexes Derived from Autoregressive Analysis of Time Series : Methods and Application to Heart Rate Variability," *PLoS ONE*, vol. 12, no. 10, pp. 1-22, 2017.

[10] P. Hall and T. Maiti, "On Parametric Bootstrap Methods for Small Area Prediction," *Journal of the Royal Statistical Society : Series B*, vol. 68, no. 2, pp. 221-238, 2006.

[11] F.A.B Colugnati, F. Louzada-Neto and J.A. Taddei, "An Application of Bootstrap Resampling Method to Obtain Confidence Interval for Percentile Fatness Cutoff Points in Childhood and Adolescence Overweight Diagnoses," *International Journal of Obesity*, vol. 29, no. 3, pp. 340-347, 2005.

[12] A. Kant, P.K. Suman, B.K. Giri, M.K. Tiwari, C. Chatterjee, P.C. Nayak and S. Kumar, "Comparison of Multi-Ogjective Evolutionary Neural Network, Adaptive Neuro-Fuzzy Inference System and Bootstrap-Based Neural Network for Flood Forecasting," *Neural Comput & Applic*, vol. 23, no. 1, pp. 231-246, 2013.

[13] Y. Ren, N. Yu, X. Wang, L. Li and J. Wan, "Nonparametric Bootstrap-Based Multihop Localization Algorithm for Large-Scale Wireless Sensor Networks in Complex Environments," *International Journal of Distributed Sensor Networks*, pp. 1-9, 2013.

[14] A. Kleiner, A. Talwalkar, P. Sarkar and M.I. Jordan, "A Scalable Bootstrap for Massive Data," *J.R. Statist. Soc. B*, vol. 76, no. 4, pp. 795-816, 2014.

[15] P. Jacek, R. Norbert, S. Malgorzata, G. Andrii and K. Maciej, "Bootstrap Identification of Confidence Intervals for the Non-Linear DoE Model," *Applied Mechanics and Materials*, vol. 712, pp. 11-16, 2015.

[16] M. Chen, C, Liao, S. Chen, Q. Ding, D. Zhu, H. Liu X. Yan and J. Zhong, "Uncertainty Assessment of Gamma-Aminobutyric Acid Concentration of Different Brain Regions in Individual and Group Using Residual Bootstrap Analysis," *Med Biol Eng Comput*, vol. 55, pp. 1051-1059, 2017.

[17] L. Dongping, "Failure Prognosis with Uncertain Estimation Based on Recursive Models Re-Sampling Bootstrap and ANFIS," *IAENG International Journal of Computer Science,* vol. 43, no. 2, pp. 253-262, 2016.

[18] F. Liang, J. Kim and Q. Song, "A Bootstrap Metropolis-Hastings Algoritm for Bayesian Analysis of Big Data," *Technometrics*, vol. 58, no. 3, pp. 304-318, 2016.

[19] C-L Mei, M. Xu and N. Wang, "A Bootstrap Test for Constant Coefficients in Geographically Weighted Regression Models," *International Journal of Geographical Information Science*, vol. 30, no. 8, pp. 1622-1643, 2016.

[20] A.A. Mikshowsky, D. Gianola and K.A. Weigel, "Improving Reliability of Genomic Predictions for Jersey Sires Using Bootstrap Aggregation Sampling," *J. Dairy Sci.*, vol. 99, pp. 3632-3645, 2016.

[21] O.R. Olaniran, S.F. Olaniran, W.B. Yahya, A.W. Banjoko, M.K. Garba, L.B. Amusa and N.F Gatta, "Improved Bayesian Feature Selection and Classification Methods Using Bootstrap Prior Techniques," Anale. Seria Informatica., vol. 14, no. 2, pp. 46-52, 2016.

[22] J. Zhen, "Detection of Wideband Signal Number Based on Bootstrap Resampling," *International Journal of Antennas and Propagation*, pp. 1-8, 2016.

[23] T. Boukaba, M.N.E. Korso, A.M. Zoubir and D. Berkani, "Bootstrap Based Sequential Detection in Non-Gaussian Correlated Clutter," *Progress in Electromagnetics Research*, vol. 81, pp. 125-140, 2018.

# Model Selection in Subset Polynomial Regression by Using Bootstrap Algorithm

ORIGINALITY REPORT

26%

SIMILARITY INDEX

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | www.bps.go.id<br>Internet | 218 words — 7% |
| 2 | ijain.org<br>Internet | 148 words — 5% |
| 3 | himayatullah.weebly.com<br>Internet | 88 words — 3% |
| 4 | "Table of contents", 2016 12th International Conference on Mathematics, Statistics, and Their Applications (ICMSA), 2016<br>Crossref | 39 words — 1% |
| 5 | pdfs.semanticscholar.org<br>Internet | 35 words — 1% |
| 6 | T. Olsson, A. Koptioug. "Statistical analysis of antenna robustness", IEEE Transactions on Antennas and Propagation, 2005<br>Crossref | 27 words — 1% |
| 7 | eibn.org<br>Internet | 18 words — 1% |
| 8 | ubt.opus.hbz-nrw.de<br>Internet | 18 words — 1% |
| 9 | lib.dr.iastate.edu<br>Internet | 16 words — 1% |

10 1slon.ru
Internet
15 words — < 1%

11 Don-Hee Han. "Facial Dimensions and Predictors of Fit for Half-Mask Respirators in Koreans", AIHA Journal, 1/2003
Crossref
13 words — < 1%

12 animsci.agrenv.mcgill.ca
Internet
13 words — < 1%

13 Jēkabsons, Gints, and Jurijs Lavendels. "A comparison of subset selection and adaptive basis function construction for polynomial regression model building", Scientific Journal of Riga Technical University Computer Sciences, 2009.
Crossref
13 words — < 1%

14 Kant, Amal, Pranmohan K. Suman, Brijesh K. Giri, Mukesh K. Tiwari, Chandranath Chatterjee, Purna C. Nayak, and Sawan Kumar. "Comparison of multi-objective evolutionary neural network, adaptive neuro-fuzzy inference system and bootstrap-based neural network for flood forecasting", Neural Computing and Applications, 2013.
Crossref
12 words — < 1%

15 figshare.com
Internet
11 words — < 1%

16 academic.odysci.com
Internet
11 words — < 1%

17 journals.plos.org
Internet
11 words — < 1%

18 "Hydrologic Modeling", Springer Nature, 2018
Crossref
9 words — < 1%

19 Borchartt, Tiago B., Roger Resmini, Leonardo S. Motta, Esteban W.G. Clua, Aura Conci, Mariana J.A. Viana, Ladjane C. Santos, Rita C.F. Lima, and Angel Sanchez.
9 words — < 1%

"Combining approaches for early diagnosis of breast diseases using thermal imaging", International Journal of Innovative Computing and Applications, 2012.
Crossref

20 www.dice.ucl.ac.be
Internet
9 words — < 1%

21 Tae Wan Kim. "Parameter optimization using a regression model and fitness function in laser welding of aluminum alloys for car bodies", International Journal of Precision Engineering and Manufacturing, 04/2011
Crossref
8 words — < 1%

22 www.profileofpcnayak.org
Internet
8 words — < 1%

23 Yazıcı, Ceyda, Fatma Yerlikaya-Özkurt, and İnci Batmaz. "A computational approach to nonparametric regression: bootstrapping CMARS method", Machine Learning, 2015.
Crossref
8 words — < 1%

24 Dengkui Li, Changlin Mei. "A two-stage estimation method with bootstrap inference for semi-parametric geographically weighted generalized linear models", International Journal of Geographical Information Science, 2018
Crossref
8 words — < 1%

25 "Proceedings of the Pacific Rim Statistical Conference for Production Engineering", Springer Nature, 2018
Crossref
8 words — < 1%

26 users.jyu.fi
Internet
8 words — < 1%

27 Arda Yunianta, Omar Mohammed Barukab, Norazah Yusof, Nataniel Dengen, Haviluddin Haviluddin, Mohd Shahizan Othman. "Semantic data mapping technology to solve semantic data problem on heterogeneity aspect", International Journal of Advances in Intelligent
8 words — < 1%

Informatics, 2017
Crossref

28 Marchetti, S.. "Non-parametric bootstrap mean squared error estimation for M-quantile estimators of small area averages, quantiles and poverty indicators", Computational Statistics and Data Analysis, 201210 — 7 words — < 1%
Crossref

EXCLUDE QUOTES OFF
EXCLUDE BIBLIOGRAPHY ON
EXCLUDE MATCHES OFF