SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

ISSN 1301-4048 e-ISSN 2147-835X Period Bimonthly Founded 1997 Publisher Sakarya University
http://www.saujs.sakarya.edu.tr/

Title: Investigation of Parametric, Non-Parametric and Semiparametric Methods in Regression Analysis Structures

Authors: Esra YAVUZ, Mustafa ŞAHİN

Recieved: 2022-07-22 00:00:00

Accepted: 2022-09-22 00:00:00

Article Type: Research Article

Volume: 26
Issue: 6
Month: December
Year: 2022
Pages: 1111-1116

How to cite
Esra YAVUZ, Mustafa ŞAHİN; (2022), Investigation of Parametric, Non-Parametric and Semiparametric Methods in Regression Analysis Structures. Sakarya University Journal of Science, 26(6), 1111-1116, DOI: 10.16984/saufenbilder.1147135
Access link
https://dergipark.org.tr/en/pub/saufenbilder/issue/74051/1147135

Sakarya University Journal of Science 26(6), 1111-1116, 2022

Investigation of Parametric, Non-Parametric and Semiparametric Methods in Regression Analysis

Esra YAVUZ*[1], Mustafa ŞAHİN[2]

Abstract

Regression analysis is known as statistical methods applied to model and analyze the relationship between variables. Regression method can be examined as parametric, non-parametric and semiparametric regression methods. The parametric regression method assumes that the dependent variable is in a linear relationship with the independent variables and that the shape of the relationship is known. If these assumptions are not met, non-parametric regression methods are applied. However, these methods cause difficulties especially in the interpretation part due to the problem of multidimensionality when there is more than one independent variable. Thus, when there is more than one independent variable, some of the independent variables may be in a linear relationship with the dependent variable, while the other part may be in a nonlinear relationship. Thus, in order to model these relationships, semiparametric regression methods, which are the additive combination of parametric and non-parametric regression methods, are used. In this study, parametric regression method, definition of non-parametric regression method and assumption conditions are given. It has been shown that the semiparametric regression method can be applied in cases where these assumptions are not met. Thus, in the study, regression methods were examined in three different parts, and parametric, non-parametric and semiparametric regression methods were examined theoretically.

Keywords: Parametric, non-parametric, semiparametric, regression

1. INTRODUCTION

Regression analysis, one of the most important subjects of statistics, is an analysis method applied to measure the relationship between two or more variables. Regression analysis gives information about the existence of a relationship between dependent and independent variables and, if there is, about the type and strength of that relationship [1]. In general, regression analysis investigates whether there is an effect of another variable on a variable.

* Corresponding author: yavuz7346@gmail.com
[1] Şırnak University
ORCID: https://orcid.org/0000-0002-5589-297X
[2] Kahramanmaraş Sütçü İmam University
E-mail: ms66@ksu.edu.tr
ORCID: https://orcid.org/0000-0003-3622-4543

Here, while the dependent variable is one, the variables that affect it (independent variables) can be one or more. A regression model with one independent variable is called simple linear regression, and a regression model with more than one independent variable is called multiple linear regression analysis. Regression analysis is based on the assumptions that the independent variables affect the dependent variable linearly, that the relationship between dependent and independent variables is known, and that the dependent variable has a normal distribution.

However, in many estimation problems, some of the independent variables do not affect the dependent variable linearly. Thus, the need to examine regression models that include nonlinear or more complex variables arises. For this reason, regression analysis is examined in three different groups as parametric, non-parametric and semiparametric regression [2].

In the parametric regression model, it is assumed that the shape of the regression function is known beforehand. At the same time, it is assumed that the error variances are constant for all independent variables, there is no autocorrelation between the error terms, the variables are normally distributed, and there is no multicollinearity problem between the independent variables [3]. If these conditions are not met, the estimations made for the regression function lead to incorrect results and interpretations. In cases where the assumptions are not met, these estimations will not be a good estimation result, so a non-parametric regression model that can stretch the linearity assumption in the parametric regression model is needed. While strong assumptions are required in the parametric regression model, these assumptions are not required in the non-parametric regression model [4].

In the non-parametric regression model, the shape of the function is not predetermined and there are no assumptions provided in the parametric regression model. The only part of the assumption requirement is that the mean of the error terms is zero and the variance is finite. Thus, flexibility is created in the emergence of the relationship between the variables. In non-parametric regression, it is known for the ability of the data to express itself [5]. In addition to its advantages in non-parametric regression, it also has some difficulties. One of these challenges is the need for a large amount of data. Another challenge is the complexity of operations and the need for intensive computer use. In cases where the number of independent variables is more than two, alternative methods are needed because it is difficult to apply the analyzes with non-parametric regression method and to interpret the graphs [6]. Thus, semiparametric regression model can be applied in cases where parametric and non-parametric regression methods are not used. In semiparametric regression models, time dependent and chance variables are included in the model with non-parametric methods, while continuous independent variables are included in the model with parameric methods [7].

Semiparametric regression method is also known as partial linear regression method due to the additive combination of parametric and non-parametric regression. In the semiparametric regression method, if the independent variables are unrelated, the coefficients of the parametric variables are estimated by the least squares method, and the non-parametric variables are estimated by methods such as spline [8]. The fact that the semiparametric regression method is less restrictive than parametric and non-parametric methods has led to the use of this method more. In addition, since the sample size is less important in this method, especially in applications consisting of small samples, the researcher has the opportunity to continue the

practice even if he or she has no knowledge about the parameter and variable. One of the most important features that distinguishes the semiparametric method from other methods is that the research continues even if the assumptions are not fulfilled [9].

Recently, in addition to parametric and non-parametric regression methods, the need for semiparametric regression method has increased. The aim of this study is to use semiparametric regression method in addition to these methods in cases where parametric and non-parametric regression methods are insufficient [10].

2. METHODS

2.1. Regression Methods

The regression method is a method applied to measure the relationship between two or more variables that have a cause-effect relationship between them. In the regression, the existence of the relationship between the variables, and if there is a relationship, estimations and predictions about the strength of the relationship are made. In the regression method, one of the variables is applied as the dependent variable and the others as the independent variable [11].

where x is the independent variable and y is the relationship between the dependent variable,

$$y = f(x) \tag{1}$$

shown in the form.

The purpose of the regression method is to examine the existence of the relationship between the independent variables on the dependent variable and to estimate the values that the dependent variable can take [12]. The regression method is shown in three different groups as parametric, non-parametric and semiparametric regression [13].

2.1.1. Parametric regression method

The method of examining the mean relationship between dependent and independent variables with a mathematical function and determining the parameters in this function clearly is called parametric regression [14]. Parameters are measures such as mean, variance and ratio of the population. In the parametric regression method, the $x_1, x_2, \dots, x_p$ arguments are assumed to be linear. $X = (x_1, x_2, \dots, x_p)$represents the mean of y dependent variable versus the change in independent variables [15],

$$y = \boldsymbol{X}\beta + \varepsilon_{\mathrm{i}} \tag{2}$$

shown in the form.

The parametric regression method is based on assumptions. It assumes that the functional shape between the independent variables, $\boldsymbol{X}\beta$, is considered linear and that the β parameters are finite [15].

2.1.2. Nonparametric regression method

Non-parametric regression model created with the dependent variable (y) and the independent variable (x), whose relationship with the dependent variable is unknown,

$$y_i = f(x_i) + \varepsilon_i\,, \quad i = 1,2, \dots, n \tag{3}$$

shown in the form. In the non-parametric regression method, besides estimating the parameters, it aims to estimate the unknown mean function $f(x_i)$.

Although there are no limiting assumptions when estimating in non-parametric regression, some features must be provided. When the number of independent variables is large, it

becomes increasingly difficult to make an estimation. Thus, the interpretation of the resulting graphics becomes more complicated [16]. Thus, it causes the problem of dimensionality. In addition to these situations, it is difficult to apply discrete independent variables in the non-parametric regression method and to interpret the effects of the y variable with the increase in the number of independent variables. As a result of these difficulties, the application of the semiparametric regression method provides an advantage [17].

2.1.3. Semiparametric regression method

Semiparametric regression method, in which some of the variables in the model are considered as parametric and the other part as non-parametric regression,

$$y_i = \alpha + x_1\beta_1 + \ldots + x_j\beta_j + f_1(x_{j+1}) + \cdots + f_k(x_k) + \varepsilon \quad (4)$$

is expressed as. In this model, j variables have a linear effect on the dependent variable y and represent the parametric part of the model. It creates a non-linear effect on the dependent variable y of other variables and represents the non-parametric part of the model. Since more than one variable is considered in the non-parametric regression part, the problems that occur in the non-parametric regression method are also valid for these models. In order to eliminate the problems that occur, the variables in the non-parametric part of the model are included in the model additively, and a new model is created. In addition, there may be dummy variables in the parametric part of the model [18].

At the stage of model creation, first of all, the variables are determined. Then, the functional or mathematical shape of the model should be determined. While creating the mathematical figure, first of all, the graphics should be interpreted. By interpreting the graphs created with the dependent variable and the independent variables, information is given about the structure of the relationship between the variables. After examining the graphs of the variables, a semiparametric regression model is applied in which some of the independent variables are suitable for parametric and the other part is suitable for non-parametric relationships. The parametric part of the semiparametric regression method creates the linear relationship, and the non-parametric part creates the nonlinear relationship. For this reason, the semiparametric regression method is also called semi-linear methods.

Iterative algorithms are generally applied in the estimation of semiparametric regression methods. These algorithms are many algorithms used for estimation of methods and these algorithms are analyzed in many computer programs. The most preferred programs in the analysis of these algorithms are R software, SAS software and Stata software [15].

When the independent variables in the semiparametric regression model are unrelated, the coefficients of the variables in the parametric part of the model are analyzed with the least squares method, and the non-parametric variables are analyzed with the estimations of non-parametric methods such as spline. However, in the semiparametric regression method, the parametric and non-parametric parts of the model may be related to each other [13]. Algorithms are needed in which the relationships between the variables are taken into account when relevant. The most preferred among these algorithms are the backward fit algorithm, the Newton-Raphdon algorithm, the Speckman approach and the partial spline approaches [16].

3. CONCLUSIONS AND DISCUSSION

When there is more than one independent variable in the regression analysis, some of the independent variables may be in a linear relationship with the dependent variable, while the other part may be in a nonlinear relationship. In addition, in cases where the number of independent variables is more than two, it becomes difficult to apply non-parametric regression analyzes and interpret graphs. In order to model such relationships, semiparametric regression methods, which are a combination of parametric and non-parametric regression methods, are used.

In this study, first of all, parametric regression method, definition of non-parametric regression method and assumption conditions are given. It has been shown that the semiparametric regression method can be applied in cases where these assumptions are not met. Thus, in the study, regression methods were examined in three different parts, and parametric, non-parametric and semiparametric regression methods were examined theoretically.

Acknowledgments
This study was presented at the "IFSCOM2022 8th International IFS and Contemporary Mathematics Conference".

Funding
The author (s) has no received any financial support for the research, authorship or publication of this study.

Authors' Contribution
The authors contributed equally to the study.

The Declaration of Conflict of Interest/ Common Interest
No conflict of interest or common interest has been declared by the authors.

The Declaration of Ethics Committee Approval
This study does not require ethics committee permission or any special permission.

The Declaration of Research and Publication Ethics
The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] G. Aneiros-Pérez, P. Vieu, “Nonparametric time series prediction: A semi-functional partial linear modeling”, Journal of Multivariate Analysis, vol. 99, no 5, pp. 834-857, 2008.

[2] M. Aytaç, “Applied non-parametric statistical tests”, Uludag University Press, Bursa, Turkey, 1991.

[3] J. Begun, W. Hall, W. Huang, J. Wellner, “Information and asymptotic efficiency in parametric-nonparametric models”, The Annals of Statistics, vol. 11, pp. 432-452, 1983.

[4] M. Çitil, F. Tuğrul, “Some New Equalities On the Intutionistic Fuzzy Modal Operators”, Sakarya University Journal of Science, vol. 22, no 6, pp. 1524-1531 2018.

[5] J, Harezlak, D. Ruppert, M. P. Wand, "Semiparametric regression with R", New York: Springer, 2018.

[6] L. Keele, "Semiparametric Regression For The Social Sciences", Chichester: John Wiley & Sons, 2008.

[7] R. Li, H. Liang, "Variable selection in semiparametric regression modeling", Annals of statistics, vol. 36, no 1, p. 261, 2008.

[8] D. Y. Lin, L. J. Wei, I. Yang, Z. Ying, "Semiparametric regression for the mean and rate functions of recurrent events", Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 62, no 4, pp. 711-730, 2000.

[9] J. Liu, R. Zhang, W. Zhao, "A robust and efficient estimation method for single index models". Journal of Multivariate Analysis, vol. 122, pp. 226-238, 2013.

[10] M. Mammadov, A. F. Yüzer, D. Aydın, "Splayn correction regression and correction parameter selection", 4th Statistics Congress proceedings and poster abstracts book, Belek-Antalya, September 25-28, pp: 148-149, 2005.

[11] X. Shi, " Applications of nonparametric and semiparametric methods in economics and finance". PhD Thesis, Economics in the Graduate School of Binghamton University, New York, 2009.

[12] D. Ruppert, M. P. Wand, R. J. Carroll, " Semiparametric regression" (No. 12). Cambridge university press, 2003.

[13] X. Shi, " Applications of nonparametric and semiparametric methods in economics and finance". PhD Thesis, Economics in the Graduate School of Binghamton University, New York, 2009.

[14] N. Tezcan, "Non-parametric regression analysis". Atatürk University Journal of Economics and Administrative Sciences, vol. 25, pp. 341-352, 2011.

[15] S. Toprak, "Semi-parametric regression models with measurement errors". PhD Thesis, Dicle University, Institute of Science, Department of Mathematics, Diyarbakır, Turkey, pp: 98, 2015.

[16] A. Yatchew, "Semiparametric regression for the applied econometrician". Cambridge University Pres, Cambridge, UK, pp: 213, 2003.

[17] E. Yavuz, M. Şahin, " Semiparametric Regression Models and Applicability in Agriculture". Black Sea Journal of Agriculture, 9-10, 2022.

[18] M. Turanı, S. Bağdatlı, "Semiparametric Regression". Suggestion Journal, vol. 9, no 35, pp. 207-213, 2011.