Progress in Nutrition 2021; Vol. 23, Supplement 2: e2021291 DOI 10.23751/pn.v23iS2.12179

Original article

Jackknife Resampling Method for Estimation of Fuzzy Regression Parameters and Revised Tanaka Method

Dervis Topuz

Niğde Zübeyde Hanım Vocational School of Health Services, Niğde Ömer Halisdemir University, Niğde, Turkey.

Abstract. *Objective*: In this paper, hierarchical ways of constructing a fuzzy regression model using the jackknife resampling technique were the basis of the study. Fuzzy techniques are based on blurring the coefficients, while the jackknife technique is based on the delete one and delete-d observations. *Methods:* This study aimed to estimate the deviation, standard error, and confidence interval of the regression coefficients calculated by the jackknife Revised Tanaka (JFLR) technique, and to compare the performance of the jackknife least squares (JOLSR) technique with the relevant estimates. *Results*: The calculation of estimates is presented with a clinical numerical example. The deviation of the Revised Tanaka FLR model, standard errors, and confidence intervals of regression coefficients were found to be significantly smaller than the estimated JOLSR standard errors. The value of MSE calculated by fuzzy jackknife regression based on Revised Tanaka FLR technique was found to be bigger than MSE calculated by JOLSR technique. *Conclusion:* Jackknife OLS and jackknife FLR regression methods can be used effectively for parameter estimation, and the jackknife revised Tanaka regression method gives more reliable and valid results.

Key words: Jackknife OLS regression, Jackknife FLS regression, Anthropometric, Revised Tanaka method

Introduction

Statistical regression analysis is considered a useful estimation technique used in different scientific fields. Regression analysis, including statistical regression analysis and fuzzy regression analysis, aims to determine the best-fit model for describing the functional relationship between dependent variables and independent variables by exploiting the knowledge from the data pairs obtained as a result of the measurement. It is assumed that some differences may arise between observed values and the estimated values due to measurement errors and modeling errors. As the source of some differences, in many real-world problems, observations are usually described as approximate values instead of exact values due to a lack of information or inexact knowledge (1). In these cases, for linear regression with normal random errors ε_j having a constant variance, the least squares theory of regression estimation and inference provides clean, exact, and optimal methods for analysis. But for generalizations to non-normal errors and non-constant variance, exact methods rarely exist, and we are faced with approximate methods based on linear approximations to estimators and central limit theorems. Ordinary least squares linear regression, wherein ideal conditions resampling essentially reproduces the exact theoretical analysis, also offers the potential to deal with non-ideal circumstances such as non-constant variance. The probability distributions for the observations either cannot be found or can be done so only with great difficultly for such a data fuzzy set. The fuzzy set theory introduced by Zadeh (2) has caused to do valid and reliable applications in many areas of studies. The theory of fuzzy set is preferred because it handles uncertainty and vagueness.

The fuzzy regression analysis is a new statistical regression analysis technique that combines the ordinary least squares regression technique with the theory of fuzzy logic. The approach is one of the most widely used statistical techniques for evaluating the functional relationship between dependent and independent variables in uncertain situations. In fuzzy regression analysis, the relationship between dependent variables and independent variables is not as precise as in ordinary least squares regression analysis (3-5). In these cases, the jackknife resampling technique has the potential to provide more accurate analysis. The use of the jackknife resampling technique in linear regression and fuzzy linear regression is important. Regression analysis is one of the areas in which fuzzy set theory is used frequently since Tanaka (6) initiated research on fuzzy linear regression (FLR) analysis. This area is widely developed, and a wide variety of methods are proposed. In general, there are two approaches in the analysis of fuzzy regression models: the possibilistic approach and the fuzzy least squares regression (FLSR) analysis method, which is based on linear programming (LP). One approach to deal with FLSR is Linear Programming (LP). The fuzzy least squares (FLS) technique, an extension of the least squares technique to fuzzy set theory, was first introduced by Tanaka and developed by others. The different aspects of the method were investigated by Celmins (7), Diamond (8), Diamond and Körner (9), Savic and Pedrycz (10), Chang and Lee (11). The technique has been introduced to minimize the fuzziness of the analyzed data and the total spread of the output (see, for example (5, 12, 13)). The approach is based on blurring the coefficients. Blurring can be done in two ways. It is possible by 1) blurring the model coefficients estimated by the ordinary least squares technique at a specified "h level", or 2) estimating the coefficients as fuzzy numbers (14). Most of these fuzzy regression models are considered with fuzzy outputs and fuzzy parameters but non-fuzzy (crisp) inputs. The estimation process in fuzzy regression analysis is carried out based on both dependent and independent variables that take numeric values. For constructing models, a least-squares technique is usually adopted to find the regression coefficients using the collected observations of independent and dependent variables. The determination of the regression coefficients is important since they are used to describe the contribution of the corresponding dependent variable. Here, the basic idea is to minimize the fuzziness of the model by minimizing the total uncertainty of the fuzzy coefficients, which were calculated using all data.

Jackknife resampling method is the process of taking repeated samples from the main sample data set of n volumes selected by chance sampling method in order to obtain a new sample. The logic of the jackknife method is "leave someone out." In other words, in order to create new Jackknife samples from the main sample data set of n capacity, one observation is excluded each time and n pieces jackknife subsamples, each consisting of n-1 observations, are created. The observation excluded is returned (by putting it in place) and the process continues until all other observations in the sample have been excluded. Each Jackknife sample produced represents the property of the main sample's data. The jackknife or "leave one out" logic is a cross-validation technique first developed by Quenouille (1949, 1956) to estimate the bias and variance estimation from estimators (15, 16). The technique is a process of estimating the unknown parameters through Jackknife samples obtained by resampling the original sample (17). First, the parameters are estimated from the whole sample. Then, each element is, in turn, dropped from the sample and the parameter of interest is estimated from this smaller sample. A pseudo-value is then computed as the difference between the whole sample estimate and the partial estimate. These pseudo-values reduce the bias of the partial estimate. It is also used to obtain improved estimates and confidence intervals for complicated statistics (18). It is resampled artificially based on the original sample, and the estimated value of each parameter is obtained by the Jackknife sample as the observation value. The real purpose of the Jackknife is a logic that is used to obtain an unbiased prediction, reduce the random effects, and minimize higher risks (19, 20). The most important feature of the technique is that it can obtain an unbiased or small-bias estimate of the parameter values of the original sample by using only one sample by sampling (21).

The aim of this research was to estimate the jackknife OLS and jackknife FLS regression coefficients with anthropometric measurement variables such as head circumference and abdominal circumference, height, chest circumference, and gender using the jackknife delete-one algorithm. It was also aimed to evaluate the relationships between anthropometric measurement variables such as head circumference with abdominal circumference, height, chest circumference, and gender by using in the clinical context.

Materials and Methods

Revised Tanaka regression (FLR) analysis

According to the FLSR approach, it is assumed that the deviations between the observed values and the predicted values are caused by the uncertainty of the system structure or the blurring of the regression coefficients, not from measurement and observation errors, contrary to the OLSR analysis method (22). Because of that, they introduced FLR model. He et al.(22) clarified that the hi ($0 \leq h \leq 1$) value, which is referred to as the degree of fit of the estimated fuzzy linear model to the given data, in Tanaka's FLR model depends not only on the estimated $\tilde{y}_i$'s spread but also on the distance between y_i's center and observed. Therefore, they proposed a new model by developing the objection function in Tanaka's FLR. It is seen in the proposed model that, the system of fuzziness decreases, and the average of estimated h_i values i.e $\bar{h}$ increases compared with Tanaka's model. So, the novelty of this article is to estimate the fuzzy regression parameter with the help of He et. al (23). So, He et. al's FLR is used in the parameter estimation of the fuzzy regression for the first time in this study. The title of our study is " Fuzzy Jackknife Regression Based on Revised Tanaka FLR". FLR which is mentioned in this paper revised Tanaka FLR. He et al. (23) denoted that their method gives better estimation than Tanaka FLR's.

That is, it assumes that the coefficients of the regression analysis model are related to its blur. For this purpose, the formula below is employed to estimate parameters of Fuzzy Regression Based on Revised Tanaka FLR ";

$$f = X \times \tilde{\beta} \rightarrow \tilde{Y}_i, \tilde{Y}_i = f(\tilde{\beta}, X) \qquad (1)$$

It is given by. Here; $\tilde{Y}_i$, x denotes the dependent variable in the symmetric triangular property structure estimated as a fuzzy number and is shown as $\tilde{Y}_i = (\tilde{y}_c, \tilde{e}_s)$. $\tilde{y}_c$, denotes the mean value (center) and $\tilde{e}_s$, denotes the propagation value.

In the case of fuzzy observations, consider a fuzzy linear regression for crisp explanatory and fuzzy response observations as follows:

$$\tilde{Y}_i = f(\tilde{\beta}, X) = \tilde{\beta}_0 + \tilde{\beta}_1 X_{i1} + \cdots + \tilde{\beta}_{p-1} X_{i(p-1)} = \tilde{\beta}_0 + \sum_{i=1}^{n} \tilde{\beta}_i X_i \qquad (2.a)$$

$$\tilde{Y}_i = \{c_0, s_0\} + \{c_1, s_1\} X_{i1} + \{c_2, s_2\} X_{i2} + \cdots + \{c_{p-1}, s_{p-1}\} X_{i(p-1)} \qquad (2.b)$$

It is defined by $(i = 1,2,3,\ldots,n)$. In the fuzzy LS regression model, the data of the dependent $\tilde{Y}_i$ variable can be real numbers or fuzzy numbers. It is generally assumed that the data for the dependent Y_i variable are symmetrical fuzzy numbers of interval type (24).

In which $\tilde{\beta}_j = [\tilde{\beta}_0 \text{ ve } \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3, \ldots \tilde{\beta}_j \ldots \ldots, \tilde{\beta}_{p-1}]^t$ is the coefficient values of the independent variables in the function and It is a set of dependent and independent variables formed in the form of $\{Y_i, X_{i1}, X_{i2}, X_{i3}, \ldots, X_{(p-1)n}\} = \{Y_i, X_{ij}\}$, and each dependent variable observation is expressed as $x \in X$ $(i = 1, \ldots, n,\ j = 1,2, \ldots, p-1)$. That is, they are crisp values of the explanatory variables.

$\tilde{\beta}_j = [\tilde{\beta}_0 \text{ ve } \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3, \ldots \tilde{\beta}_j \ldots \ldots, \tilde{\beta}_{p-1}]^t$ are fuzzy regression coefficients vectors with a symmetric triangular fuzzy number structure and they are fuzzy numbers in the form of $\tilde{\beta}_j = (c_j, s_j)$ $\tilde{\beta}_j$, (j: 0,1,2,3,….,p-1). c_j, is the $\mu_{\tilde{\beta}_i}(c_j) = 1$value representing the midpoint of the coefficients, that is, the center value, and has the form $c_j = [c_1, c_2 c_3, \ldots, c_n]^t$. s_j, shows the spread of the coefficients belonging to the fuzzy regression analysis model and is $s_j = [s_1, s_2, s_3, \ldots, s_n]^t$ shaped (24).

Each coefficient value $\tilde{\beta}_i = \{c_j, s_j\} = \{\tilde{\beta}_i : c_j - s_j \leq \tilde{\beta}_i \leq c_j + s_j\}$ has a symmetric triangular property structure and is, $\tilde{\beta}_i$ (j: 0,1,2,3,…,p-1) (6,25)

The $\tilde{\beta}_i = \{c_j, s_j\}$ value of the fuzzy coefficients was estimated by the minimum blur method proposed by Tanaka. The method is given in the following equation.

In Possibilistic Regression Analysis proposed by Tanaka and lshibuchi (1992) (26), the linear programming (LP) formulation considers triangular membership functions (not necessarily symmetric). The parameters of fuzzy regression can be estimated by using the objective function and constraints shown The LP formulation is follow (3):

$$\min Z(x)\ s^{*}|X_i| = SF + \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (S_0 + \sum_{j=0}^{m} s_j |X_{ij}|) + \sum_{i=1}^{n} d_i \quad (3)$$

$$0 \leq h_i = 1 - d_i / (S_0 + \sum_{j=1}^{m} s_j |X_{ij}|)$$

$$d_i = \left| y_i - (c_0 + \sum_{j=1}^{m} c_j |X_{ij}| \right.$$

and,

$$\min_{c,s} J = s_1, s_2, \ldots, s_n, s_j \geq 0\ \forall j;\ j = 0,1,2,\ldots,m,$$

$$s_j \geq 0, c_j \in R, X_{i0} = 1 (0 \leq h \leq 1, \forall_j = 0,1,2,\ldots,n)$$

Here, Z(x): function shows the total blur in the model. SF is the system fuzziness. m: is the number of observations regarding the dependent variable j: The number of arguments. x_{ij}: is the ith observation value of the jth independent variable. For each predicted $\tilde{Y}_i$ observation value, the constraint number must be 2x*n* (27). In order to minimize the total spread, the level h, $\tilde{Y}_i$, the predictor of each observation value Y_i, is assumed to have a turbidity tolerance $\mu_{\tilde{Y}_i}(Y_i) \geq h$ i=1,2,..,m. (28) In Equation 3, the objective function is weighted with the absolute values of the measurements of the distributions of the independent variables. The application of Jackknife resampling technique in fuzzy least squares regression analysis is given below.

Fuzzy Jackknife Regression Based on Revised Tanaka FLR Algorithm.

In this section, algorithm for fuzzy jackknife regression based on revised Tanaka (JFLSR) models based on the resampling observations was given. These approaches are usually applied when the regression models built from data have fixed explanatory variables. There are two cases of jackknife resampling. The first of them is based on deleting a single case from the original sample (delete one jackknife), and the second is based on deleting multiple cases from the original sample (delete d jackknife) sequentially (29-31). In the study, an algorithm based on deleting a single case (delete a jackknife) from the original sample was applied. The jackknife fuzzy regression analysis procedure is as follows.

Steps of The Algorithms for Delete-One Jackknife Revised Tanaka Regression.

To describe the resampling methods we start with an n sized sample $w_i = (Y_i, X_{ji})'$ and assume that w_i's are drawn independently and identically from a distribution of F, where $Y_i = (y_1, y_2, \ldots, y_n)'$ contains the responses, $X_{ji} = (x_{j1}, x_{j2}, \ldots, x_{jn})'$ is a matrix of dimension nxk, where j = 1,2,...k, i = 1,2,3,...,n. Let the px1 vector $w_i = (Y_i, X_{ji})'$, (i = 1,2,…,n) denote the values associated with ith observation. In this case, the set of observations are the vectors $(w_1, w_2, w_3, \ldots, w_n)$. The jackknife procedure based on delete-one (do) is as follows (32).

1$^{(do)}$. Draw n sized sample from population randomly and label the elements of the vector $w_i = (Y_i, X_{ji})'$ as the vector $Y_i = (y_1, y_2, \ldots, y_n)'$ and the matrix $X_{ji} = (x_{j1}, x_{j2}, \ldots, x_{jn})'$ where j = 1,2,...k, i = 1,2,3,...,n.

2$^{(do)}$. Omit the first row of the vector $w_i = (Y_i, X_{ji})'$ and label remaining n-1 sized observation sets $Y_i^{(J)} = (y_2^{(J)}, y_3^{(J)}, y_4^{(J)}, \ldots, y_n^{(J)})'$ and $X_{ji}^{(J)} = (x_{j2}^{(J)}, x_{j3}^{(J)}, x_{j4}^{(J)}, \ldots, x_{jn}^{(J)})'$ as delete -one Jackknife sample $(w_i^{(J)})$ and estimate the FLS regression coefficients $\tilde{\beta}^{(J_1)}$ from $(w_i^{(J)})$. Then, omit second row of the vector $w_i = (Y_i, X_{ji})'$ and label remaining n-1 sized observation sets $Y_i^{(J)} = (y_2^{(J)}, y_3^{(J)}, y_4^{(J)}, \ldots, y_n^{(J)})'$ and $X_{ji}^{(J)} = (x_{j2}^{(J)}, x_{j3}^{(J)}, x_{j4}^{(J)}, \ldots, x_{jn}^{(J)})'$ as $w_2^{(J)}$ and estimate the FLS regression coefficients $\tilde{\beta}^{(J_2)}$. Similarly, omit each one of the n observation sets and estimate the regression coefficients as $\tilde{\beta}^{(J_i)}$ alternately, where $\tilde{\beta}^{(J_i)}$ is Jackknife regression coefficient vector estimated after deleting of ith observation set from w_i.

3$^{(do)}$.Obtain the probability distribution $(F(\tilde{\beta}^{(J)}))$ of Jackknife estimates $\tilde{\beta}^{(J_1)}, \tilde{\beta}^{(J_2)}, \tilde{\beta}^{(J_3)}, \ldots, \tilde{\beta}^{(J_n)}$

4$^{(do)}$. Calculate the jackknife regression coefficient estimate which is the mean of the $(F(\tilde{\beta}^{(J)}))$ distribution (33) as;

$$\tilde{\beta}^{(J)} = \sum_{i=1}^{n} \tilde{\beta}^{(J_i)} / n = \bar{\tilde{\beta}}^{(J_i)} \quad (4)$$

5$^{(do)}$. Thus, the delete-one Jackknife regression equation is

$$\bar{Y}_i = f(\bar{\beta}, x) = \bar{\beta}_0^{(J)} + \bar{\beta}_1^{(J)} x_{j1}^{(J)} + \bar{\beta}_2^{(J)} x_{j2}^{(J)} + \cdots + \bar{\beta}_k^{(J)} x_{jn}^{(J)} \quad (5)$$

An illustrative study that shows how the delete-one Revised Tanaka Regression jackknife regression parameters are estimated was given in Table 2.

Revised Tanaka Jackknife bias, variance, confidence, and percentile interval.

The jackknife bias, variance, and confidence intervals are estimated by using the following equations from $\left(F(\tilde{\beta}^{(J)})\right)$ distribution (19, 34, 35). The jackknife bias equals,

$$\text{bias}_J(\tilde{\beta}) = (n-1)(\tilde{\beta}^{(J)} - \tilde{\beta}) \qquad (6)$$

$$\text{var}(\tilde{\beta}^{(J)}) = \frac{(n-1)}{n}\sum_{i=1}^{n}\left[(\tilde{\beta}^{(Ji)} - \tilde{\beta}^{(J)})(\tilde{\beta}^{(Ji)} - \tilde{\beta}^{(J)})'\right] \qquad (7)$$

where $\left(\tilde{\beta}_i^{(Ji)}\right)$ is the estimate produced from the replicate with i[th] observation set or j[th] group deleted (36).

Jackknife (1-α) 100 % confidence interval equals (37).

$$\left(\tilde{\beta}^{(J)} - t_{n-p,\frac{\alpha}{2}} * S_e(\tilde{\beta}^{(J)}) < \beta < \tilde{\beta}^{(J)} + t_{n-p,\frac{\alpha}{2}} * S_e(\tilde{\beta}^{(J)})\right) \qquad (8)$$

where $t_{n-p,\frac{\alpha}{2}}$ is the critical value of t with probability α/2 the right for n-p degrees of freedom; and $S_e(\tilde{\beta}^{(J)})$ is the standard error of the $\tilde{\beta}^{(J)}$.

The jackknife percentile Interval can be constructed from the quantiles of the jackknife sampling distribution of $\tilde{\beta}^{(J)}$. The (α/2) % and (1-α/2) % percentile interval is

$$\tilde{\beta}^{(J)}_{(lower)} < \beta < \tilde{\beta}^{(J)}_{(upper)} \qquad (9)$$

where $\tilde{\beta}^{(J)}$ is the ordered jackknife estimates of fuzzy least squares regression coefficient from Equation 11 or 13, lower = (α/2) n, and upper = (1-α/2) n (38).

Goodness of Fit Test Criteria for Fuzzy Regression Models

In the study, in order to determine to what extent, the real observation values and the estimated fuzzy results are compatible with each other, the goodness of fit test criteria.

- MSE: Mean square error,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2 \qquad (10)$$

- Coefficient of determination (R^2),

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2}{\sum_{i=1}^{n}(\tilde{Y}_i)^2}\right) \qquad (11)$$

Here; : $\tilde{Y}_i$ The nx1 dimensional predicted values show the vector. n: number of observations, p: number of input variables, Y_i: Shows the observed values (39).

Results

Head circumference measurement is an important indicator in predicting brain development, especially in early childhood (40-41). Abnormal values in the head circumference can be an early sign of many diseases, including developmental retardation, as well as inherited disorders. It is known that most of the change in the head circumference is due to genetic factors (42). In many studies, it has been shown that in addition to genetic and age factors, parental head circumference has an important effect on determining the reasons for the change in head size (43). In the study conducted to determine whether gender affects the head circumference, it was determined that the mean head circumference of the newborn boys is approximately 0.7-0.8 cm. larger than that of the newborn girls. Anthropometric measurement data of 50 healthy newborns (50% male, 50% female) constituted the study sample. Anthropometric measurements of the variables such as head circumference, abdominal circumference, body length, chest circumference, and gender of the newborns were taken retrospectively from specially created patient files and those variables are used in the JOLS and JFLS regression models that are summarized in previous sections. This example focuses onthe illustration and application of jackknife techniques in JOLS and JFLS regression analysis. The data pairs $w_i = (Y_i, X_{ji})'$ of Table 1 population, (i = 1, …, 50) are used to demonstrate the proposed procedure in a case where the crisp input X and crisp output Y_i.

The jackknife procedure based on deletion (do) is applied to the data in Table 1 as follows.

1[(do)]. First, the ordinary least squares regression (OLSR) model was fitted to data given in and the results of the ordinary least squares regression were summarized in Table 2.

All regressions in Table 2 are significant (P<0.01) and the determination of coefficient $R^2 = 0.558$, respectively. The regression of BÇ on B and C is significant

Table 1. n = 50 volume original data set

$w_i^{(I)}$	BÇ(Y)	KÇ(X_1)	B(X_2)	GÇ(X_3)	C
1	35	34.5	51	38	2
2	36	35	51.5	35.5	2
3	33	30	43	30	2
4	34	29	46	31	2
-	-	-	-	-	-
47	35	30	48	32	1
48	29	28	46	29.5	1
49	35.5	32	47	33	1
50	33	32	48	32.5	1

BÇ: Head circumference (cm); KÇ: Abdominal circumference (cm); B: body length (cm); GÇ: Chest circumference (cm), C: gender

as a result of variance analysis (P<0.01**). According to the t-tests for significance of regression coefficients, KÇ and GÇ of the regression coefficients are significant (P<0.01).

2^(do)^. 3^(do)^. 4^(do)^. 5^(do)^. The jackknife (jackknife samples, each of size n-1=50-1=49) regression procedure, from the data given in Table 1, calculating the OLS jackknife and FLR jackknife estimates of the OLS jackknife and FLR jackknife regression parameters for each sample are shown in Table 3.

The objective function in equation (3) is calculated as follows:

$$\min Z(x) = SF + \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} \left(S_0 + \sum_{j=0}^{m} s_j |X_{ij}|\right) + \sum_{i=1}^{n} d_i$$

$$Z = 2x\left[50\, a_0^S + a_1^S \sum_{i=1}^{50} x_{1i} + a_2^S \sum_{i=1}^{50} x_{2i} + a_3^S \sum_{i=1}^{50} x_{3i} + a_4^S \sum_{i=1}^{50} x_{4i}\right] + \sum_{i=1}^{50} d_i \qquad (12)$$

$$Z = 2x\,[50 x\, a_0^S + 1580x\, a_1^S + 2435x\, a_2^S + 1692.50x\, a_3^S + 73.0x\, a_4^S] + d_1 + d_2 + .. + d_{50}$$

$$Z = 2x\,[50 x 0.605 + 1580 x - 0.022 + 2435 x - 0.012 + 1692.50 x 0.015 + 73.0x0.035] = 5.8$$

Above objective functions' constraints are shown in equation (13) and equation (14) for the first and last data. For the first data;

$$d_1/(s_0 + 34.5s_1 + 51s_2 + 38s_3 + 2.0s_4) \leq 0.5$$

$$d_1 = |35.0 - c_o - |(c_0 + 34.5c_1 + 51c_2 + 38c_3 + 2.0c_4) \leq 0.5 \qquad (13)$$

For the last data;

$$d_{50}/(s_0 + 32.0s_1 + 48.0s_2 + 32.5s_3 + 1.0s_4) \leq 0.5$$

$$d_{50} = |33.0 - c_o - |(c_0 + 32.0c_1 + 48.0c_2 + 32.5c_3 + 1.0 \leq 0.5 \qquad (14)$$

The coefficients obtained from Lingo 16. Software are as follows:

The summaries of some OLS jackknife and FLS jackknife values of regression coefficients are presented in Table 4.

Jackknife samples are generated omitting each one of the n observation sets and estimated the regression coefficients as $\tilde{\beta}_1^{(I)}$. The histograms of the jackknife estimates conform quite atypical to the limiting normal distribution for all regression coefficients.

The combined display of statistics calculated by jackknife OLS and jackknife Revised Tanaka FLR regression analysis methods applied for the estimation of head circumference of newborns is as in Table 5.

According to the JOLS and JFLR regression analysis methods, it was determined that there was no statistically significant difference between the observed head circumference values (cm) of the newborns and the mean measurements of the estimated head circumference values (cm). Based on these results, it was also seen in this study that the approaches were calculated with minimum deviation as a result of the study

Table 2. The summary statistics of regression coefficients for OLS regression

Variables	$\hat{\beta}$	S.E.($\hat{\beta}$)	t	Sig	95% Confidence Interval
Constant	11.467	4.051	2,831	.007	(3.309) –(19.626)
KÇ(X_1)	-0.014	0.056	-0.260	.796	(-0.127) – (0.098)
B(X_2)	0.273	0.100	2.730	.009	(0.072) - (0.474)
GÇ(X_3)	0.267	0.116	2.309	.026	(0.034) - (0.500)
C(X_4)	0.890	0.295	3.017	.004	(0.296) - (1.484)
R^2=0.558, N = 50, SSE = 0.994, F = 14.216**					

BÇ: Head circumference (cm); KÇ: Abdominal circumference (cm); B: body length (cm); GÇ: Chest circumference (cm), C: gender, SSE: sum of squares of error.

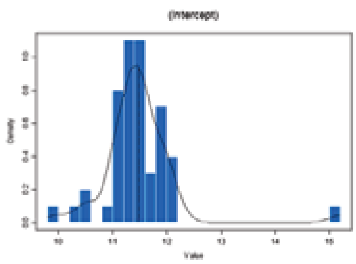
Table 3. The of the jackknife (jackknife samples, each of size n-1=50-1=49) regression procedure from the data given, calculating the Jackknife OLS ve jackknife FLR regresyon estimates of the regression parameters for each sample newborn's head cırcumference model.

		Observation sets					**JOLSR**					**JFLR**				
$W_i^{(J)}$	**Variables**	1	2	3	-	50	$\bar{\beta}_0^{(J)}$	$\bar{\beta}_1^{(J)}$	$\bar{\beta}_2^{(J)}$	$\bar{\beta}_3^{(J)}$	$\bar{\beta}_4^{(J)}$	$\bar{\beta}_0^{(J)}$	$\bar{\beta}_1^{(J)}$	$\bar{\beta}_2^{(J)}$	$\bar{\beta}_3^{(J)}$	$\bar{\beta}_4^{(J)}$
1	BÇ(Y)	omitted	36	33	-	33	9.952	-0.018	0.272	0.316	0.935	-6.471	0.094	0.412	0.496	1.058
	KÇ(X_1)		35	30	-	32										
	B(X_2)		51.5	43	-	48										
	GÇ(X_3)		35.5	30	-	32.5										
	C(X_4)		2	2	-	1										
2	BÇ(Y)	35	omitted	33	-	24.50	11.134	-0.011	0.281	0.262	0.901	-10.750	0.690	0.661	-0.295	0.718
	KÇ(X_1)	34.5		30	-	25.10										
	B(X_2)	51		43	-	24										
	GÇ(X_3)	38		30	-	19.07										
	C(X_4)	2		2		1										
3	BÇ(Y)	35	36	omitted	-	33	10.379	-0.011	0.292	0.270	0.850	-11.161	0.490	0.724	-0.151	0.746
	KÇ(X_1)	34.5	35		-	32										
	B(X_2)	51	51.5		-	48										
	GÇ(X_3)	38	35.5		-	32.5										
	C(X_4)	2	2			1										
-	- -	- -	- -	- -	- -	- -	- -	- -	- -			- -	- -	- -	-	-
50	BÇ(Y)	35	36	33	-	omitted	11.545	-0.011	0.276	0.259	0.866	-8.331	0.594	0.586	-0.163	0.602
	KÇ(X_1)	34.5	35	30	-											
	B(X_2)	51	51.5	43												
	GÇ(X_3)	38	35.5	30	-											
	C(X_4)	2	2	2	-											
	$\beta^{(J_0)} = \sum_{i=1}^{n} \beta^{(J_i)}/50$, $\tilde{\beta}^{(J_0)} = \sum_{i=1}^{n} \tilde{\beta}^{(J_i)}/50$						11.479	-0.010	0.271	0.264	0.890	-7.726	0.572	0.574	-0.148	0.637

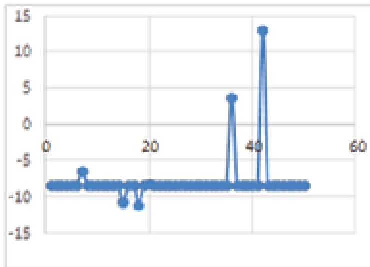
Table 4. The summary statistics of the regression coefficients for OLS jackknife and FLR Revised Tanaka jackknife regression (n=50)

	Variables	**Observed**	**Mean**	**S.E.**	**Bias**	**Confidence intervals**
						5%, 95% Persentile Interval
JOLSR	Constant	11.467	11.469	4.830	0.574	(10.556)-(12.115)
	KÇ(X_1)	-0.014	-0.010	0.253	0.218	(-0.023)-(-0.009)
	B(X_2)	0.272	0.271	0.123	-0.083	(0.252)-(0.291)
	GÇ(X_3)	0.267	0.264	0.175	-0.104	(0.233)-(0.297)
	C(X_4)	0.889	0.890	0.255	0.021	(0.836)-(0.946)
JFLR	Constant	11.467	-7.726	1.504	0.605	(10.411)-(12.181)
	KÇ(X_1)	-0.014	0.572	0.0	-0.022	(-0.026)-(-0.008)
	B(X_2)	0.272	0.574	0.0	-0.012	(0.246)-(0.293)
	GÇ(X_3)	0.267	-0.148	0.0	0.015	(0.196)-(0.309)
	C(X_4)	0.889	0.637	0.0	0.035	(0.833)-(0.953)

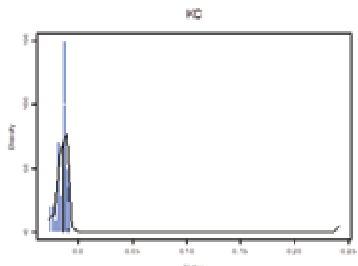
BÇ: Head circumference (cm); KÇ: Abdominal circumference; B: body length; GÇ: Chest circumference, C: gender



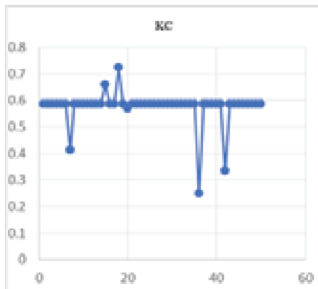
(a)



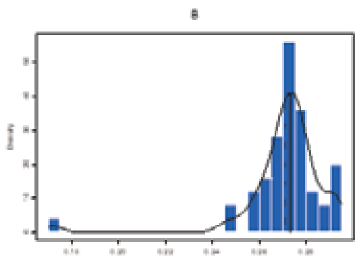
(a)



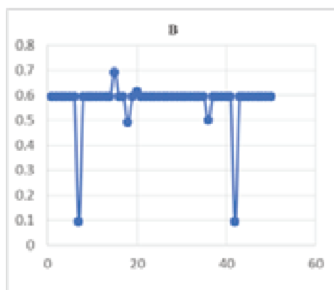
(b)



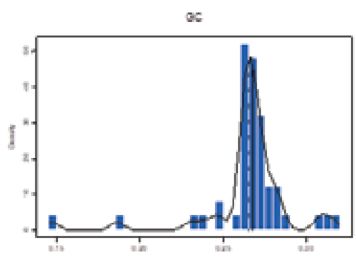
(b)



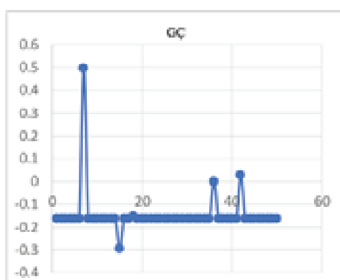
(c)



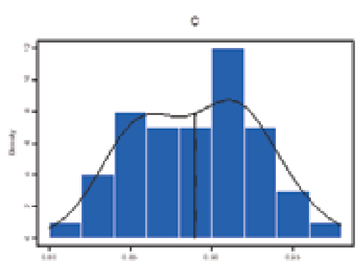
(c)



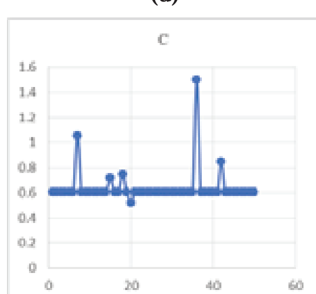
(d)



(d)



(e)



(e)

Figure 1. Histogram of jackknife OLS (n=50, (a), (b), (c), (d), (e)) and jackknife FLR (n=50, (a), (b), (c), (d), (e)) regression parameter estimates.

Table 5. Statistics of the estimated mean head circumference values of newborns with the JOLSR and JFLR analysis approach

$w_i^{(j)}$	Observed Head circumference values of newborns(cm))	Head circumference values estimated with the JOLSR approach (cm)	Head circumference values estimated with the JFLR approach (cm)
1	35	36.77	36.93
2	36	36.24	37.88
3	33	32.53	30.95
4	34	36.51	31.95
…	…	...	…
47	35	33.53	32.89
48	29	32.34	30.97
49	35.5	33.50	33.31
50	33	33.64	33.96
Mean	34.64	34.60	34.22
Standard error	1.434	33.64	21.97
Goodness of Fit Test Criteria		MSE(J)= 0.89	MSE(F)= 3.53
		$R^2_{observed/\ Jestimated}$= 0.999	$R^2_{observed/\ Festimated}$= 0.997

conducted for the application of JOLS and Revised Tanaka (JFLR) regression analysis methods in cases where uncertainties and outlier observation values were found in clinical studies. The mean head circumference observed and estimated by the JOLSR and JFLR regression methods were calculated as 34.60 cm and 34.22 cm, respectively.

According to the JOLSR and JFLR regression methods, among the anthropometric measurements of the newborns, there was a significant relationship between $r_{BC-C} = 0.433$, $r_{BC-B} = 0.579$, $r_{BC-GÇ} = 0.616$ (correlation is significant at 0.01 level(2-tailed) and the head circumference of the newborns, while no relationship was found between the abdominal circumference. According to the JOLSR and JFLR regression methods, it was determined that the height and gender of the newborn are the main factors affecting the head circumference of the newborn. In the analysis of the head circumference values of newborns, it was determined by both models that gender should be taken into account as well as genetic characteristics. In addition, it was concluded that the head circumference results calculated with the JFLR model represented the reality more strongly.

Discussion and Conclusion

Usually, researchers make recommendations by using methods that contain different assumptions, trying to show them as alternatives to each other. However, it should be kept in mind that jackknife OLS and jackknife FLS regression approaches cannot be substituted for each other. Because the applicability of each method is limited by various assumptions, their applications differ from each other. The jackknife OLS regression and jackknife FLR regression methods estimate the variation of a statistic from the variation of that statistic between subsamples, rather than from parametric assumptions, and may yield different results in many situations. So, they provide a way of decreasing bias and obtaining standard errors in situations where the standard methods might be expected to be inappropriate. But when the jackknife OLS regression method is used to estimate the standard error of a statistic, it gives very few different results when on the same data, whereas the jackknife FLR regression method gives the same result each time. In these cases, Heltshe and Forrester (1985) also reported that not only sample size but also the total number of individuals in the sample is important in improving the jackknife estimators (44, 45). There are a limited number of studies in the literature examining anthropometric measurements affecting the head circumference of the newborn. However, it was stated that under normal nutritional and health conditions, body size, gender, and head circumference are significantly related. However, a few studies have emphasized that the apparent general relationship between the newborn's abdominal circumference, body length, chest circumference, gender, and head circumference is misleading and causes misinterpretations (44, 46, 47). In order to be a solution to these problems, on the data set with a small sample, the Jackknife fuzzy regression method was preferred to obtain reliable and valid estimations. Hence, the jackknife bias, the standard errors, and confidence intervals of the KÇ, B, GÇ, and C coefficients based on the distribution $\left(F\left(\hat{\beta}^{(J)}\right)\right)$ are substantially larger than the jackknife FLS and estimated asymptotic OLS standard errors. The Jackknife OLS percentile ranges are also larger than the Jackknife FLR percentile ranges of the B and C coefficients.

As a result, Jackknife OLS and jackknife FLR regression methods can be used effectively for parameter estimation. Comparisons can be made between them. They are preferred to the nonlinear regression method due to some theoretical properties such as not having any distribution assumption on residues and therefore allow inference even if errors do not conform to normal distribution, the explanatory performance of the proposed model using the mean square of error is satisfactory. However, the jackknife OLS regression method is easier to apply to complex sampling logic than the jackknife FLR regression method. The application of both regression techniques depends on the development of computer technologies and could also be more frequently used if statistical computer packages featured these analyses. Jackknife resampling method works for descriptive and spread measures statistics, e.g. the estimators such as averages, bias, variances, but not for statistics, e.g. medians, maxima value, smallest value, etc. The estimator preferred should be appropriate to the model used or reliable and valid estimations can not be obtained. When we compare the methods according to the clinical results, the jackknife Revised Tanaka regression method is preferred to the Jackknife OLS regression method because it can calculate reliable and valid results that represent the reality better.

Funding: This study was completed by no means of financial support or funding.

References

1. Chen L, Hsueh C. Fuzzy regression models using the least-squares method based on the concept of distance, *IEEE Transactions on Fuzzy Systems* 2009; 17, 1259-1272. Doi: 10.1109/tfuzz.2009.2026891.
2. Zadeh LA. Fuzzy Sets, Information and Control 1965; 8, 338-353.
3. Yang MS, Liu HH. Fuzzy least-squares algorihms for interactive fuzzy linear regression models, Fuzzy Sets and Systems 2003; 135, 305-316. Doi: 10.1016/S0165-0114(02)00123-9
4. Danesh S. Fuzzy parameters estimation via hybrid methods, Hacettepe Journal of Mathematics and Statistics 2016; 47, 1605-1624. Doi: 10.15672/hjms.201614621831
5. Razzaghnia T. Regression parameters prediction in data set with outliers using neural network, Hacettepe Journal of Mathematics and Statistics, 2019; 48, 1170-1184. Doi: 10.15672/hujms.544496
6. Tanaka H, Uejima S, Asai K. Linear regression analysis with fuzzy model, IEEE Transactions on Systems, Man, and Cybernetics, 1982;12, 903 – 907. Doi: 10.1109/tsmc.1982.4308925
7. Celmins A. Least squares model fitting to fuzzy vector data, Fuzzy sets and Systems, 1987; 22, 245-269. Doi: 10.1016/0165-0114(87)90070-4
8. Diamond P. Least squares fitting of several fuzzy variables, Proc. of the Second IFSA Congress, Tokyo; July 20–25.1987; 1, 339-332
9. Diamond P. Fuzzy least squares, Information Science, 1988; 46, 141-157. Doi: 10.1016/0020-0255(88)90047-3
10. Savic D, Pdrycz W. Evaluation of fuzzy regression models, Fuzzy Sets and Systems, 1991; 391, 51-63. Doi: 10.1016/0165-0114(91)90065-X
11. Chang PT, Lee ES. Fuzzy least absolute deviations regression and the conflicting trends in fuzzy parameters, Computers and Mathematics with Applications, 1994; 285, 89-101. Doi: 10.1016/0898-1221(94)00143-X
12. Ming M, Friedman M, Kandel A. General fuzzy least squares, Fuzzy Sets and Systems, 1997; 88, 107-118. doi.10.1016/S0165-0114 (96)00051-6
13. Hong DH, Song JK, Young H. Fuzzy least-squares linear regression analysis using shape preserving operations, Information Sciences 2001; 138, 185-193.
14. D'Urso P, Gastaldi T. A least-squares approach to fuzzy linear regression analysis, Computational Statistics and Data Analysis, 2000; 34, 427-440. Doi: 10.1016/S0167-9473(99)00109-7
15. Quenouille MH. Approximate tests of correlation in time-series, Journal of the Royal Statistical Society. Series B, 1949; 11, 68–84. doi:10.1111/j.2517-6161.1949.tb00023.x.
16. Quenouille MH. Notes on bias in estimation, Biometrika, 1956; 43, 353-360. doi:10.1093/biomet/43.3-4.353

17. Lu X, Su L. Jackknife model averaging for quantile regressions, Journal of Econometrics 2015; 188, 40–58. Doi: 10.1016/j.jeconom.2014.11.005
18. Friedl H, Stampfer E. Jackknife resampling, Encyclopedia of Environmetrics, 2002; 2, 1089-1098. doi:10.1002/9780470057339.vaj001
19. Tukey J.W. Bias and confidence in not-Quite large sample, Annals of Mathematical Statistics, 1958; 29, 614-623. Doi: 10.1214/aoms/1177706647.
20. Abdi H, Williams LJ. Jackknife. Encyclopedia of research design, 2010; 654-661. Thousand Oaks, Sage.
21. Xu G. Jackknife estimates and its application, Journal of Shandong Normal University, 2000; 15, 454–455._
22. Chang YHO, Ayyub BM. Fuzyy regression methotds-a comparative assessment, Fuzzy Sets and Systems, 2001; 119,187-203. Doi: 10.1016/s0165-0114(99)00091-3
23. He YQ, Chan L, Kand ML. Balancing productivity and consumer satisfaction for profitability: statistical and fuzzy regression analysis, European Journal of Operational Resarch, 2007: 176(1), 252-263
24. Kim B, Bishu RR. Evaluation of fuzzy linear regression models by comparison membership functions, Fuzzy Sets and Systems, 1998; 100, 343–352. Doi: 10.1016/S0165-0114(97)00100-0
25. Wang HF, Tsaur RC. Insight of a fuzzy regression model, Fuzzy Sets and Systems, 2000; 112, 355-69. doi:org/10.1016/S0165-0114 (97)00375-8
26. Tanaka H, Ishibuchi H. Possibilistic regression analysis based on linear programing*;* 1992.
27. Moskowitz H, Kim K. On assesing the H value in fuzzy linear regression, Fuzzy Sets and Systems, 1993; 58, 303-27. Doi: 10.1016/0165-0114(93)90505-C
28. Hojati M, Bector C.R, Smimou KA. Simple method for computation of fuzzy linear regression. European Journal of Operational Research 2005; 166, 172-184. doi:10.1016/j.ejor.2004.01.039
29. Efron B, Gong G. A Leisurely look at the bootstrap, the jackknife, and cross-validation, The American Statistician 1983; 37, 36-48. Doi: 10.1080/00031305.1983.10483087
30. Wu CFJ. Jackknife, bootstrap and other resampling methods in regression analysis, Annals of Statistics, 1986; 14, 1261-1295. Doi: 10.1214/aos/1176350142
31. Shao J, Tu D. Applications to time series and other dependent data, The Jackknife and Bootstrap. Springer Series in Statistics. Springer, New York, NY, 1995; 386-415. Doi: 10.1007/978-1-4612-0795-5_9
32. Sahinler S, Topuz D. Bootstrap and Jackknife resampling algorithms for estimation of regression parameters, Journal of Applied Quantitative Methods, 2007; 2, 188-199.
33. Fox J. Applied regression analysis, linear models, and related methods*,* 1997; Sage Publications.
34. Miller RG. The Jackknife–a review. Biometrika 1974; *61*, 1–15. doi:10.1093/biomet/61.1.1
35. Efron B, The jackknife, the bootstrap and other resampling plans, Society for industrial and applied mathematics, 1982.
36. Friedl H, Stampfer E. Jackknife resampling. Encyclopedia of Environmetrics, 2, Eds.: A. El- Shaarawi, W. Piegorsch, Wiley: Chichester, 2002a; 1089-1098.
37. Tibshirani RJ, Efron B. An introduction to the bootstrap. Monographs on statistics and applied probability, 1993; 57, 1-436.
38. Friedl H, Stampfer E. Resampling methods. Encyclopedia of Environmetrics. 3, Eds. A. El- Shaarawi, W. Piegorsch, Wiley: Chichester, 2002b; 1768-1770.
39. Topuz D. Clinical Data Obtained for Prediction of the Weight of the Newborn Analysis with Classical and Fuzzy Linear Regression Models, Turkiye Klinikleri J Biostat, 2020; 12, 320-34. Doi: 10.5336/biostatic.2020-74850
40. Kliegman RM, Behrman RE, Jenson HB, Stanton BM. Nelson textbook of pediatrics e-book, Elsevier Health Sciences, 2007.
41. Daştan Gürler S, Boran P. Ebeveyn antropometrik ölçümlerinin çocuk baş çevresi üzerine etkisi, Çocuk Dergisi, 2018; 18, 113-120. Doi: 10.5222/j.child.2018.04880.
42. Weaver DD, Christian JC. Familial variation of head size and adjustment for parental head circumference, The Journal of Pediatrics, 1980; 96, 990-994. Doi: 10.1016/S0022-3476(80)80623-8.
43. Sawada A, Ikeda H, Kimura-Ohba, Matsuzawa, Awaya T, Shiotani, et al. Head growth evaluation in early childhood, from the Japan Children's study, Pediatrics International, 2010; 52, 343-346. doi.org/10.1111/j.1442-200X.2009.03002.x
44. Christiane S, Holle G, Michael H. The association between weight, height and head circumference reconsidered, Pediatric Research, 2017; 81, 825-830. doi:10.1038/pr.2017.3
45. Adhiambo S, Makumı N, Odhıambo R, Orwa G. On jackknife confidence ıntervals assocıated with nonparametrıc regressıon estımators for a finite population total, East African Journal Of Statistics,2017;5,41-48.
46. Ariani D, Nasution Y.N, Yuniarti D. Perbandingan metode bootstrap dan jackknife resampling dalam menentukan nilai estimasi dan ınterval kontingensi parameter regresi, Jurnal Eksponensial, 2017; 8, 43-50.
47. Fernandes AAR, et al. Metode Statistika Multivariat Pemodelan Persamaan Struktural (SEM) Pendekatan WarpPLS. Universitas Brawijaya Press, 2017.

Correspondence
Dervis Topuz
Niğde Zübeyde Hanım School of Health Services Vocation,
Niğde Ömer Halisdemir University, Niğde, Turkey
https://orcid.org/0000-0001-9415-1421
Email: topuz@ohu.edu.tr