STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, July 2026, pp 103–115.
Published online in International Academic Press (www.IAPress.org)

A Robust Cubic Spline Approach for Hierarchical Regression Models with Outliers

Sarah Sabah Akram [1], Hayder Raaid Talib [2,*], and Noor Abdul-Kareem Fayadh [2]

[1] *College of Administration and Economics, Wasit University, Wasit, Iraq*
[2] *College of Administration and Economics, Sumer University, Thi-Qar, Iraq*

Abstract This study is based on parameter estimation for hierarchical regression models in the presence of outliers. The robust cubic spline and the maximum likelihood estimation are compared. In order to test the efficiency of the two methods, Monte Carlo simulation experiments were conducted on sample sizes 15, 25, 40, and 60 at three different percentages of outliers 1%, 3%, and 5%. The assessment of performance was done by the absolute deviation error (ADE) and the coefficient of determination (R^2). The outcome of this study showed that the robust cubic spline performed better than maximum likelihood estimation.

Keywords hierarchical; Robust Cubic Spline; outlier values; multilevel models; M-estimation.

DOI: 10.19139/soic-2310-5070-3633

1. Introduction

1.1. General

Hierarchical/multilevel modeling is extensively used in statistics in the analysis of datasets where nesting or overlap exists, e.g., regional data or repeated measures in longitudinal study designs. The primary issue that arises in the use of such multilevel data is the presence of outliers since their inclusion renders estimation highly inaccurate. The study examines the impact of outliers on the estimation process within hierarchical regression models. Ordinary Least Squares and Maximum Likelihood Estimation (MLE) methods are commonly used yet susceptible to the effects of outliers, which can yield biased or unreliable results. So, it's important to develop better, more flexible estimation methods that can reduce the impact of outliers.

The intention of this study is to determine the influence of outliers in estimating parameters for hierarchical regression models, as well as to point out the shortcomings of standard estimation procedures when dealing with such cases. The research also focuses on evaluating the extent to which robust statistics could help in overcoming the mentioned challenges.

1.2. Hierarchical Regression Model

Hierarchical models are used to analyze data with nested or multilevel structures. They are especially useful when observations are grouped into higher-level units, which lets the researcher examining relationships both within and between groups. Hierarchical models are widely used in science, engineering, economics, and sociology because

*Correspondence to: Hayder Raaid Talib (Email: hayder.raaid@uos.edu.iq). College of Administration and Economics, Sumer University. P4F3+7H4, Rifai, Dhi Qar Governorate, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

they are so flexible. Theoretical, mathematical, or empirical methods can be used to create these models, as long as the necessary assumptions are met [1]. One of the best things about hierarchical models is that they can show how variables affect outcomes at different levels of a dataset. They also help clarify the connections between parts of complex data structures, which can improve understanding and decision-making. These models also provide parameter estimates at each level of the hierarchy, which can make the analysis more reliable [2].
Hierarchical regression has been utilized across various research domains due to its growing significance. Richardson et al. [3] employed this methodology to estimate parameters in a study examining mortality across various cohorts. Their approach facilitated the assessment of distinct regression models that were applied for various outcome categories within a suitable analytical framework. The findings indicated that hierarchical regression yielded more precise estimates; also, it was more resilient to the effects of outliers. Similarly, Bartholomew et al. [4] used hierarchical regression to examine the factors that contribute to low birth weight. Their analysis evaluated the impact of maternal weight, age, and height on neonatal weight; simultaneously, the influence of maternal education was evaluated.
In a study of Sipahioğlu et al.[5], they examined the influence of gender, academic performance, and the classroom emotional environment on school burnout among high school students. The researchers utilized a correlational design with random sampling to investigate the data via descriptive statistics, Pearson's correlation, and hierarchical regression. Their results showed that a positive emotional climate in the classroom was associated with lower burnout at school, whereas a negative emotional climate was associated with higher burnout. The hierarchical regression results further indicated that classroom emotional climate continued to be a significant predictor of school burnout, even after accounting for gender and academic achievement. The following is the standard form of the hierarchical regression model:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + e_{ij} \tag{1}$$

Here, Y_{ij} refers to the dependent variable for individual i in group j. Also, the parameter β_0 represents the model's intercept, while β_1 is the regression coefficient that is associated with the independent variable X_{ij}, which is the independent variable's value of the individual i in group j. The term u_j represents the effect specific to group j, and lastly, e_{ij} represents the random error for individual i in group j.

1.3. Multi-Level Model

Hierarchical statistical models are designed to analyze data in which parameters vary across different levels of the hierarchy. This means that data analysis at both the class and individual student levels can be conducted for academic achievements. Hierarchical statistical models extend the linear regression approach by allowing both linear and nonlinear forms of analysis. Development in computing capabilities, as well as the availability of advanced statistical packages, has led to the increasing application of such models.
Multilevel analysis can greatly help in dealing with data that exhibit a hierarchical nature, in that each observation is grouped in higher-order clusters. Data in the first level of such a model may comprise individual observations although there could be cases where the first level may be characterized by longitudinal data. Multilevel models thus allow for longitudinal analysis in the identification of trends in growth. Additionally, multilevel models can replace Analysis of Covariance (ANCOVA) by adjusting for covariates before evaluating treatment effects. Multilevel models can analyze these experiments without the homogeneity-of-slopes assumption required by ANCOVA. Multilevel models can be applied to data with 2-level, although 2-level models are the most common, and the remainder of this article focuses exclusively on them [15]. The dependent variable should be analyzed at the lowest level of the hierarchy.

$$Y_{ijk} = B_0 + B_1 X_{ijk} + u_j + u_k + e_{ijk} \tag{2}$$

Y_{ijk}: The variable for item i in set j and at level $k.uj$: j-group effect. u_k: k-level effect.

1.4. Outlier

An outlier denotes a value that remarkably deviates from the other data points in a dataset, either by being exceedingly high or low or by existing at the extremes of an ordered series. Statistically, it denotes data that may originate from a population different from the one under investigation. An outlier is defined as a data point having a standardized residual (RI) remarkably greater than the residuals of other observations in the collection [6, 7]. In multivariate settings, anomalies may involve several variables simultaneously. These anomalous observations can be expressed either as a full vector or as part of a sub-vector.

2. Estimation methods for the hierarchical regression model

2.1. Robust Cubic Spline

This method is robust for managing extreme values and random effects, improving the accuracy of results by yielding reliable estimation [8]. The general form of the hierarchical model is as follows:

$$Y_{ijk} = S(X_{ijk}) + u_{jk} + u_k + e_{ijk} \tag{3}$$

Y_{ijk} is the response value of data i in group j at level $k.S(X_{ijk})$: The regression function was estimated using a robust cubic spline. u_{jk}: random effect of group j at level $k.u_k$: random effect of level $k.e_{ijk}$: independent random error distributed $e_{ijk} \sim N(0, \sigma^2)$

We start with an initial estimate of the regression function $\hat{S}^{(\circ)}$ via the Cubic Spline method without the robust method, where we impose initial values for $\hat{u}_k^{(\circ)}$, $\hat{u}_{jk}^{(\circ)}$, with default values equal to zero.

Then the remainders are calculated $r_{ijk}^{(t)}$:

$$r_{ijk}^{(t)} = Y_{ijk} - \left(\hat{S}^{(t)}(X_{ijk}) + \hat{u}_{jk}^{(t)} + \hat{u}_k^{(t)}\right) \tag{4}$$

Based on the above residual equation, the weights w_{ijk} are calculated as follows:

To enhance the robustness of spline estimation, this study employs a weighting function based on the Huber-type M-estimation approach. The threshold parameter (δ) defines the point at which large residuals start to lose their influence. Here, (δ) is set to the widely recommended value of 1.345 to ensure high efficiency under normal conditions. Additionally, the scale parameter (s) (serves as a robust measure of residual dispersion) was recalculated at each iteration using the Median Absolute Deviation (MAD) with the formula:

$$s = 1.4826 \times median(|r_{ijk} - median(r_{ijk})|)$$

This technique reduces the influence of outliers by using smaller weights for outliers when calculating their contribution during the analysis. Huber's criterion was specifically selected since it utilizes the efficiency of least squares when dealing with small errors while mitigating the influence of large errors [20].

After determining the values of (δ) and s, the robust weights were applied to an iterative approach. In this approach, the updating of the coefficients of the spline and random effects continues until convergence has been achieved. The point of convergence is reached when the differences between successive parameter updates are less than 10^{-6}. An iterative technique forms a dependable foundation for fitting cubic splines in hierarchical regression.

$$w_{ijk}^{(t+1)} = \begin{cases} 1 & /r_{ijk}^{(t)}/ \leq \delta \\ \frac{1}{r_{ijk}^{(t)}} & /r_{ijk}^{(t)}/ > \delta \end{cases} \tag{5}$$

By re-estimating the regression function $\hat{S}^{(X)}$ using the system of equations based on weights $(w_{ijk}^{(t+1)})$:

$$\begin{aligned}\hat{S}^{(t+1)}(x) = \hat{\beta}_0^{(t+1)} + \hat{\beta}_1^{(t+1)} + \hat{\beta}_2^{(t+1)} x^2 + \\ \sum m = 1^m \hat{Y}_m^{(t+1)} \left(X_{ijk-t_m}\right)^3 +\end{aligned} \tag{6}$$

Where $\hat{\beta}_0$ is a fixed intercept in the model, $\hat{\beta}_1$ is the linear term coefficient, X_{ijk} is the slope of the relationship between the independent variable and the response variable, $\hat{\beta}_2$ is the quadratic term coefficient X_{ijk}^2, which adds a nonlinear effect to the relationship between Y_{ijk} and X_{ijk}, and t_m is the knots that specify the location of the relationship change. The term $[(X_{ijk-tm})^3+]$ (the truncated cubic basis function) can be defined as follows:

$$\left(X_{ijk-t_m}\right)_+^3 = \begin{cases} \left(X_{ijk-t_m}\right)^3 & X_{ijk} > t_m \\ 0 & X_{ijk} \le t_m \end{cases} \tag{7}$$

The formula in Equation (7) gives the basis of the truncated power function used in deriving the cubic spline. In this equation, $(X_{ijk-t_m})_+^3$ represents a cubic function that will be active only when the predictor value is greater than the knot t_m. Specifically, this function equals $(X_{ijk-t_m})^3$ when $X_{ijk} > t_m$ and is zero otherwise. This design lets the spline add localized flexibility at each knot, which lets the model change its curvature while keeping the first and second derivatives smooth. The truncated power basis is a common way to represent cubic splines. It can effectively capture nonlinear patterns in data [16, 21].

$\hat{Y}_m$ is the coefficients of the cubic terms associated with each node tm¬ determine the extent of the influence of the cubic term at those knots by estimating the random effect $\hat{u}_{jk}$ as in the following formula:

$$\hat{u}_{jk}^{(t+1)} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\in^2} \cdot \frac{1}{n_{jk}} \sum_{j=1}^{n_{jk}} \left(Y_{ijk} - \hat{S}^{(t+1)}(X_{ijk} - u_k)\right) \tag{8}$$

Where σ_u^2 is the group random effect variance, $\sigma_\in^2$ the random error variance, n_{jk} the number of data in group j at level k. By estimation, the random effect for level k as a whole is in the following formula:

$$\hat{u}_k^{(t+1)} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\in^2} \cdot \frac{1}{n_k} \sum \left(\bar{y}_{jk} - \hat{S}^{(t+1)}(\bar{x}_{jk}) - \hat{u}_{jk}^{(t+1)}\right) \tag{9}$$

σ_v^2 random effect level variance. n_k number of groups at level k. $\bar{Y}_{jk}$ the mean of the values y_{ijk} within the group j at level k, and $\bar{X}_{jk}$ is the mean of the values $\bar{X}_{ijk}$ within the group j at level k [9].

When the conditions of stability are met:

$$\left\|\hat{S}^{(t+1)}(x) - \hat{S}^{(t)}(x)\right\| < \in \tag{10}$$

$$\left\|\hat{u}_{jk}^{(t+1)} - \hat{u}_{jk}^{(t)}\right\| < \in \tag{11}$$

Where $\in$ is a very small predetermined limit 10-6 at which the process stops. After completing the iterative process and the model is stable, the final formula for the estimate will be as follows:

$$\hat{y}_{ijk} = \hat{S}(x_{ijk}) + \hat{u}_{jk} + \hat{u}_k \tag{12}$$

Algorithm 1. Robust Cubic Spline (RCS) Estimation Procedure

The estimation of the Robust Cubic Spline (RCS) model is carried out through an iterative weighted procedure inspired by robust M-estimation principles. The complete algorithm can be summarized in the following steps[18]:

1. Initialization:

Initial spline coefficients are obtained using the classical cubic spline fit, while all random effects are initially set to zero. The initial residuals r_{ijk} are then computed.

2. Robust scale estimation:

A robust estimate of scale is calculated using the Median Absolute Deviation (MAD):

$s = 1.4826 \times median(|r_{ijk} - median(r_{ijk})|)$.

ensuring that the scale is not influenced by extreme residuals.

3. Weight computation using the Huber function:

The Huber weights are computed as

$w_{ijk} = \varphi(\frac{r_{ijk}}{s}, \delta), \delta = 1.345,$

where small residuals receive full weight and large residuals are down-weighted to reduce the influence of outliers.

4. Update of spline coefficients and random effects:

The process of updating the spline coefficients makes use of the weighted least squares technique, after which the random effects and variance components are updated using the weighted hierarchical model.

5. Convergence assessment:

The algorithm continues to run Steps 2 through 4 until it reaches convergence, defined as parameter updates less than 10^{-6}. Failure to converge implies that the algorithm must be rerun. This process generates reliable estimates for the spline coefficients and the random effect parameters, hence improving the effectiveness of the RCS method in the presence of outliers in hierarchical data structures [19].

2.2. Nonpolynomial Spline

Polynomial splines, including the cubic spline used in this paper, have been appreciated due to their ability to capture nonlinearities with ease. However, their effectiveness may be hindered by sudden function transitions and small sample sizes. This drawback may be addressed using non-polynomial splines utilizing basis functions with exponential, trigonometric, and hyperbolic terms. The advantage of these splines is that they have been shown to retain the general structure of the data and handle boundary issues [14, 15].

The use of nonpolynomial splines, along with iterative weighting procedures, aims to reduce the effect of outliers and ensure smooth function approximation. Such an approach would be especially useful for hierarchically structured regression equations, wherein there is a need to model nonlinear dependencies and incorporate random effects nesting. The adjustment of the spline basis to the curvature of the underlying data helps reduce the problem of overfitting, which commonly affects high-degree polynomials [1].

2.3. Maximum Likelihood Method (MLE)

Maximum likelihood estimation is a principal statistical method for parameter estimation. Proposed by R.A. Fisher in 1912 [10], it can be defined as the parameter values that maximize the logarithm of the likelihood function. In this study, the parameters of the hierarchical model are estimated using the maximum likelihood method based on the observed data y and the unknown coefficient θ [11, 17].

$$L(\theta; y) = P(y \mid \theta) \tag{13}$$

When formulating the maximum likelihood function, the probability distribution for each data y_{ijk} is assumed to follow the normal distribution.

$$y_{ijk} \sim N\left(S(x_{ijk}) + u_{jk} + u_k\,;\,\sigma_e^2\right) \tag{14}$$

Assuming the independence of data, the maximum likelihood function is as follows:

$$L(\theta, y) = \prod_{i=1}^{n}\prod_{j=1}^{j}\prod_{k=1}^{k}\phi\left(y_{ijk} \mid S(x_{ijk}) + u_{jk} + u_k, \sigma_e^2\right) \tag{15}$$

$\phi(.)$ is the probability density function of the normal distribution. By taking the natural logarithm of the likelihood function, we obtain the log-likelihood, which is often more convenient for analysis and estimation.

$$e(\theta; y) = -\frac{1}{2}\sum_{ijk}\left[\log\left(2\pi\sigma_e^2\right) + \frac{\left(y_{ijk} - S(x_{ijk}) - u_{jk} - u_k\right)^2}{\sigma_e^2}\right] \tag{16}$$

By estimation, the non-linear part $S(x)$ in the hierarchical model using the cubic spline method, the effect of the independent variable X_{ijk} must be separated from the rest of the parts. Therefore, $S(x)$ can be formulated using the cubic spline estimation method, as shown below.

$$S(x) = \beta_0 + \beta_1 X + \beta_2 X^2 + \sum_{m=1}^{m} y_m (X - K_m)^3 \tag{17}$$

The parameters to be estimated represent β_0, β_1, and β_2. After performing the previous step, we reduced the effect of the outliers during the estimation of $S(x)$ by reducing the error function as follows:

$$\min_{\beta, y}\sum_{ijk} p\left(y_{ijk} - S(x_{ijk})\right) \tag{18}$$

Estimation $\hat{S}(x)$ in the model is as follows:

$$Y_{ijk} - \hat{S}(x_{ijk}) = u_{jk} - u_k + e_{ijk} \tag{19}$$

After estimating $\hat{S}(x)$, we extracted the residuals to reduce the model:

$$e_{ijk} = Y_{ijk} - \hat{S}(x_{ijk}) \tag{20}$$

Equation 18 outlines the conditional likelihood of the observed responses, Y_{ijk}, given the fixed-effect coefficients and the random-effect structure of a three-level hierarchical model. This likelihood represents the probability density of level-1 observations, assuming normally distributed errors with a variance of σ_e^2. Because the model conditions the responses on both cluster-level and group-level random effects, Equation 18 serves as the foundational likelihood component necessary for constructing the maximum likelihood estimation of the hierarchical model [13].

e_{ijk} is the residual after removing the non-linear effect is represented by the model in the following form:

$$e_{ijk} = u_{jk} + u_k + e_{ijk} \tag{21}$$

Then the random effects are estimated using the remaining data (e_{ijk}) as follows:

$$e(\theta, e) = \frac{1}{2}\sum_{ijk}\left[\log\left(2\pi\sigma_e^2\right) + \frac{e_{ijk}^2}{\sigma_e^2}\right] \times \frac{1}{2}\left[\log\left(2\pi\sigma_u^2\right) + \frac{u_{jk}^2}{\sigma_u^2}\right] \times \frac{1}{2}\sum_{k}\left[\log\left(2\pi\sigma_v^2\right) + \frac{u_k^2}{\sigma_v^2}\right] \tag{22}$$

Equation (20) builds upon the conditional likelihood from Equation 18, expanding it into a complete joint log-likelihood expression by incorporating the distributions of the random effects at levels 2 and 3. Since the hierarchical model assumes that these random components are mutually independent and follow a normal distribution, the joint density is expressed as the product of the observation density and the densities of the random effects. Logarithmic transformation of this product turns the product into a sum of logs of likelihoods, hence capturing the additive nature of the equation as stated in Equation 20. This method is commonly used in hierarchical mixed effect models, and it forms the basis of maximum likelihood estimation of variance components [12].

By maximizing the likelihood function, we derive the maximum equations for each parameter as follows:

$$\frac{\partial e}{\partial u_{jk}} = -\frac{u_{jk}}{\sigma_u^2} + \sum_i \frac{e_{ijk}}{\sigma_e^2} = 0 \tag{23}$$

$$\hat{u}_{jk} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}\sum_i e_{ijk} \tag{24}$$

$$\frac{\partial e}{\partial u_k} = -\frac{u_{jk}}{\sigma_v^2} + \sum_i \frac{e_{ijk}}{\sigma_e^2} \tag{25}$$

$$\hat{u}_k = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\varepsilon^2}\sum_i e_{ijk} \tag{26}$$

Using each of the values of $\hat{u}_k$ and $\hat{u}_{jk}$ to estimation the variances σ_u^2, σ_e^2 and σ_ε^2 where: σ_u^2 variance between groups within and across classes, σ_e^2 variance between categories and σ_ε^2 random error variance.

Since the mean square of the estimation values u_{nj}^k at the level of all groups, it is as follows:

$$\hat{\sigma}_u^2 = \frac{1}{J_k}\sum j = 1^j = \sum_{k=1}^{K} \hat{u}_{jk}^2 \tag{27}$$

After calculations, the mean squares of the estimation values $\hat{u}_K$ at the category level

$$\sigma_u^2 = \frac{1}{K}\sum K = 1^k \hat{u}_k^2 \tag{28}$$

The mean squared error remaining after removing the effect of u_k, u_{jk}.

$$\sigma_e^2 = \frac{1}{I_{JK}}\sum i = 1^I \sum_{j=1}^{J}\sum_{k=1}^{K}\left(e_{ijk} - \hat{u}_{jk} - \hat{u}_k\right)^2 \tag{29}$$

Through an iterative process that updates the coefficients to enhance estimation performance, the likelihood function is maximized repeatedly until the results stabilize, ultimately yielding the final estimation formula [10]:

$$\hat{y}_{ijk} = \hat{S}(X_{ijk} + \hat{u}_{jk} + \hat{u}_k + e_{ijk}) \tag{30}$$

3. Data generation

Numerous techniques exist for generating random data that adhere to a normal distribution, such as the Box-Muller method and the approximate asymptotic method. The Box-Muller method is favored due to its simplicity and precision. To generate data according to a multivariate normal distribution with a mean vector μ and a variance matrix $\sum$, denoted as $X \sim N(\mu, \sum)$, the following algorithm can be used employed:

* Generating Z_i values using the (Box-Muller) method

$$Z_i \sim N(0,1), \qquad i = 1,2,\ldots P$$

$$Z \sim MNp(0,1)$$

Since Z is an $n \times p$ matrix.

* Decompose the variance matrix $\sum$ using the Cholesky decomposition method to derive C. $C'C = \sum$, where C is a lower triangular matrix of order $P \times P$ and represents the square root of the matrix $\sum$. The segmentation process occurs as follows:

a) The d_i vector is calculated by the following formula:

$$d_i = a_{ij} - \sum_{k=1}^{j-1} \ell_{ik}^2 - d_k$$
$$\ell_{ik} = [(a_{ij} - \sum_{k=1}^{j-1} \ell_{ik}\ell_{ik} - d_k)]d_j^{-1}, j,k = 1,2,\ldots,P$$

a) The C matrix is extracted according to the following formula:

$$C = [LD^{(\frac{1}{2})}] \text{ where } D = diag(d_i) \text{ and } L = [\ell_{ij}]$$

* In order to compute the matrix X, which is characterized by the multivariate normal distribution with a mean vector $\sum$ and a covariance matrix $\sum'$, the below formula is applied:

$$X' = \mu + CZ' \text{ and } X \sim MNp(\mu, \sum)$$

4. Generating Regression Model Data

In addition, the X matrix, corresponding to the explanatory variables, is created. A random error term, normally distributed with a mean of 0 and a given variance σ_e^2 is also generated. The response variable values are subsequently determined using the linear relationship below:

$$Y = X\beta + e$$

The values of β are designated in accordance with the model under examination and correspond to the characteristics of the phenomenon being analyzed, grounded in its theoretical framework. Nonparametric functions are approximated by Cubic B-splines, with parameters or penalties established through cross-validation. The quantity of knots for each combination varies from 1 to 6. This method includes the previously developed explanatory variables, random errors, and a total of eight variables. The model parameters are initialized using the subsequent default values:

$$\beta = 3.281, -4.891, -3.891, 6.782, 2.666, 1.176, -3.546, 4.763$$

The study examined four sample sizes: 15, 25, 40, and 60, corresponding to small, medium, and large datasets, and introduced outliers at rates of 1%, 3%, and 5%. Each experiment was conducted 1,000 times to assess algorithm efficiency. The Maximum Likelihood Estimation and the Robust Cubic Spline were evaluated using Absolute Deviation Error (ADE), the coefficient of determination R^2, and the Robust Cubic Spline as performance indicators [18].

5. Analyzing the results of simulation experiments

Table 1 below gives the estimation procedures, Maximum Likelihood Estimation and Robust Cubic Spline, alongside the 1% outlier level and the sample sizes. The table highlights the impact of having outliers on each of the estimation procedures.

Table 1. The estimation methods (Maximum Likelihood and Robust Cubic Spline), along with the 1% outlier ratio and the corresponding sample sizes.

sample size	Parameters	Maximum Likelihood		Robust cubic spline	
		β	95% C.I	β	95% C.I
15	β_1	2.027	1.672,2.383	3.028	2.674,3.383
	β_2	-3.674	-3.116,-4.234	-4.100	-3.488,-4.714
	β_3	-2.904	-2.227,-3.583	-3.679	-3.182,-4.178
	β_4	5.772	5.189,6.357	6.173	5.893,6.455
	β_5	2.002	1.671,2.335	2.378	2.028,2.731
	β_6	1.326	1.044,1.609	1.641	1.477,1.807
	β_7	-3.018	-2.561,-3.477	-3.304	-2.899,-3.711
	β_8	4.013	3.364,4.663	4.281	3.892,4.681
25	β_1	2.192	1.562-2.824	2.563	2.172,2.955
	β_2	-3.109	-2.672,-3.547	-4.011	-3.673,-4.351
	β_3	-2.566	-2.100,-3.033	-3.177	-2.783,-3.573
	β_4	5.782	5.199,6.367	6.166	5.673,6.661
	β_5	1.894	1.270,2.520	2.300	2.104,2.498
	β_6	1.037	0.903,1.173	1.327	1.055,1.601
	β_7	-3.186	-2.894,-3.481	-3.728	-3.034,-4.424
	β_8	4.237	3.977,4.450	4.314	3.871,4.759
40	β_1	3.045	2.894,3.198	3.385	2.996,3.776
	β_2	-3.022	-2.673,-3.372	-4.005	-3.783,-4.228
	β_3	-2.259	-2.021,-2.499	-3.007	-2.566,-3.450
	β_4	6.023	5.639,6.409	6.217	5.757,6.678
	β_5	2.326	2.025,2.629	2.793	2.256,3.332
	β_6	1.099	0.936,1.264	1.673	1.117,2.231
	β_7	-2.466	-2.177,-2.757	-3.105	-2.670,-3.542
	β_8	4.084	3.573,4.596	4.288	3.901,4.677
60	β_1	3.051	2.477,3.627	3.471	3.014,3.929
	β_2	-3.209	-2.888,-3.550	-3.447	-3.003,-3.893
	β_3	-3.011	-2.560,-3.463	-3.752	-3.115,-4.391
	β_4	6.033	5.377,6.691	6.701	6.103,7.301
	β_5	2.277	2.001,2.556	2.609	2.100,3.121
	β_6	1.300	1.071,1.531	1.507	1.215,1.801
	β_7	-3.211	-2.609,-3.815	-3.448	-3.022,-3.876
	β_8	4.401	3.793,5.011	4.707	4.032,5.384

The same estimation methods, along with the 3% outlier ratio and the corresponding sample sizes, are shown in Table 2.

Similarly, Table 3 represents the above-mentioned methods with 5

Table 2. Estimation methods and sample sizes with 3% outlier ratio.

sample size	Parameters	Maximum Likelihood		Robust cubic spline	
		β	**95% C.I**	β	**95% C.I**
15	β_1	2.027	1.672,2.383	3.028	2.674,3.383
	β_2	-3.674	-3.116,-4.234	-4.100	-3.488,-4.714
	β_3	-2.904	-2.227,-3.583	-3.679	-3.182,-4.178
	β_4	5.772	5.189,6.357	6.173	5.893,6.455
	β_5	2.002	1.671,2.335	2.378	2.028,2.731
	β_6	1.326	1.044,1.609	1.641	1.477,1.807
	β_7	-3.018	-2.561,-3.477	-3.304	-2.899,-3.711
	β_8	4.013	3.364,4.663	4.281	3.892,4.681
25	β_1	2.192	1.562-2.824	2.563	2.172,2.955
	β_2	-3.109	-2.672,-3.547	-4.011	-3.673,-4.351
	β_3	-2.566	-2.100,-3.033	-3.177	-2.783,-3.573
	β_4	5.782	5.199,6.367	6.166	5.673,6.661
	β_5	1.894	1.270,2.520	2.300	2.104,2.498
	β_6	1.037	0.903,1.173	1.327	1.055,1.601
	β_7	-3.186	-2.894,-3.481	-3.728	-3.034,-4.424
	β_8	4.237	3.977,4.450	4.314	3.871,4.759
40	β_1	3.045	2.894,3.198	3.385	2.996,3.776
	β_2	-3.022	-2.673,-3.372	-4.005	-3.783,-4.228
	β_3	-2.259	-2.021,-2.499	-3.007	-2.566,-3.450
	β_4	6.023	5.639,6.409	6.217	5.757,6.678
	β_5	2.326	2.025,2.629	2.793	2.256,3.332
	β_6	1.099	0.936,1.264	1.673	1.117,2.231
	β_7	-2.466	-2.177,-2.757	-3.105	-2.670,-3.542
	β_8	4.084	3.573,4.596	4.288	3.901,4.677
60	β_1	3.051	2.477,3.627	3.471	3.014,3.929
	β_2	-3.209	-2.888,-3.550	-3.447	-3.003,-3.893
	β_3	-3.011	-2.560,-3.463	-3.752	-3.115,-4.391
	β_4	6.033	5.377,6.691	6.701	6.103,7.301
	β_5	2.277	2.001,2.556	2.609	2.100,3.121
	β_6	1.300	1.071,1.531	1.507	1.215,1.801
	β_7	-3.211	-2.609,-3.815	-3.448	-3.022,-3.876
	β_8	4.401	3.793,5.011	4.707	4.032,5.384

Table 3. Estimation methods and sample sizes with the 5 % outlier ratio.

sample size	Parameters	Maximum Likelihood		Robust cubic spline	
		β	**95% C.I**	β	**95% C.I**
15	β_1	2.367	2.064,2.673	3.105	2.771,3.439
	β_2	-4.011	-3.461,-4.562	-4.303	-3.711,-4.896
	β_3	-3.017	-2.583,-3.453	-3.471	-3.099,-3.845
	β_4	5.562	5.017,6.108	6.288	5.656,6.922
	β_5	2.109	1.700,2.520	2.501	2.225,2.779
	β_6	1.017	0.991,1.045	1.206	1.004,1.409
	β_7	-2.891	-2.372,-3.413	-3.405	-3.020,-3.792
	β_8	4.199	3.473,4.927	4.357	4.001,4.714
25	β_1	2.783	2.688-2.880	3.302	2.945,3.659
	β_2	-3.217	-3.009,-3.427	-4.287	-3.788,-4.788
	β_3	-2.823	-2.477,-3.171	-3.288	-2.904,-3.673
	β_4	5.461	5.003,5.921	6.364	5.989,6.741
	β_5	1.784	1.330,2.241	2.402	2.066,2.740
	β_6	1.107	0.899,1.318	1.204	1.001,1.409
	β_7	-3.003	-2.683,-3.325	-3.411	-3.162,-3.662
	β_8	4.104	3.652,4.558	4.422	4.046,4.799
40	β_1	3.108	2.977,3.241	3.200	2.878,3.524
	β_2	-3.134	-2.772,-3.497	-4.322	-3.982,-4.663
	β_3	-2.381	-2.116,-2.647	-3.689	-2.922,-4.457
	β_4	5.558	5.001,6.117	6.503	5.882,7.125
	β_5	2.105	1.887,2.325	2.500	2.337,2.664
	β_6	1.005	0.848,1.164	1.117	1.004,1.132
	β_7	-2.502	-2.336,-2.669	-3.337	-2.905,-3.771
	β_8	4.005	3.668,4.344	4.553	4.044,5.063
60	β_1	3.002	2.890,3.115	3.201	3.003,3.401
	β_2	-3.693	-3.211,-4.177	-4.506	-3.993,-5.021
	β_3	-3.187	-2.863,-3.513	-3.694	-3.099,-4.291
	β_4	5.937	5.276,6.600	6.602	6.209,6.997
	β_5	2.190	1.946,2.435	2.547	2.214,2.901
	β_6	1.267	1.100,1.436	1.169	1.087,1.253
	β_7	-3.186	-2.877,-3.497	-3.503	-3.307,-3.701
	β_8	4.271	3.855,4.689	4.588	4.173,5.005

The results of the Absolute Deviation Error and the coefficient of determination for the simulated cases, along with the 1% outlier ratio and the corresponding sample sizes, are shown in Table 4.

Table 4. Absolute Deviation Error and coefficient of determination for simulated cases with 1% outlier ratio

N	Methods	Absolute Deviation Error	Coefficient of determination
15	Maximum Likelihood	6.783	0.622
	Robust cubic spline	5.893	0.714
25	Maximum Likelihood	5.893	0.646
	Robust cubic spline	5.294	0.721
40	Maximum Likelihood	5.593	0.661
	Robust cubic spline	4.892	0.744
60	Maximum Likelihood	5.100	0.691
	Robust cubic spline	4.394	0.771

The results of the Absolute Deviation Error and the coefficient of determination for the simulated cases, together with the 3% outlier ratio and the corresponding sample sizes, are shown in Table 5.

Table 5. Absolute Deviation Error and coefficient of determination for simulated cases with 3% outlier ratio.

N	Methods	Absolute Deviation Error	Coefficient of determination
15	Maximum Likelihood	7.281	0.593
	Robust cubic spline	6.291	0.601
25	Maximum Likelihood	6.882	0.618
	Robust cubic spline	5.893	0.648
40	Maximum Likelihood	6.492	0.673
	Robust cubic spline	5.335	0.690
60	Maximum Likelihood	5.833	0.684
	Robust cubic spline	4.782	0.791

The results of the Absolute Deviation Error and the coefficient of determination for the simulated cases, along with the 5% outlier ratio and the corresponding sample sizes, are shown in Table 6.

Table 6. Absolute Deviation Error and coefficient of determination for simulated cases with 5% outlier ratio.

N	Methods	Absolute Deviation Error	Coefficient of determination
15	Maximum Likelihood	7.919	0.502
	Robust cubic spline	6.744	0.539
25	Maximum Likelihood	7.483	0.558
	Robust cubic spline	6.693	0.594
40	Maximum Likelihood	6.977	0.603
	Robust cubic spline	5.935	0.659
60	Maximum Likelihood	6.382	0.664
	Robust cubic spline	5.199	0.702

6. Conclusion

From the results obtained, the Robust Cubic Spline technique performs relatively well compared to Maximum Likelihood Estimation in scenarios where outliers exist. Hierarchical regression is suitable in situations where the model needs to fit both linear and nonlinear relationship within complex data structures. It can be used in various disciplines such as health care, education, social sciences, finance, and economics. It was observed that the increase in the percentage of outliers caused higher ADE but lower R2 values while the sample size caused lower ADE values. As the percentage of outliers increased from 1% to 5%, ADE values increased by 15% for MLE but by 13% for RCS. The increase in sample size from 15 to 60 caused a drop in ADE values for both techniques. However, the decrease in ADE values for RCS is relatively higher than that of MLE at various sample sizes and percentages of outliers. In some instances, the width of confidence intervals for RCS was higher than MLE.

Acknowledgement

The authors like to thank their universities for partially supporting this work.

REFERENCES

1. Al-Rawi, M. H., & Talib, H. R., *Robust estimation using non-polynomial splines for nonlinear regression models with outliers.*, Iraqi Journal of Statistical Science, (2022)19(1), 45–59.
2. Bartholomew DC, Biu OE, Arimie CO, *Hierarchical Regression Modeling of Some Factors Affecting Weight of Child at Birth*, 2021;17(12):11-27.
3. Bickel,D.R., *Robust and efficient estimation of the mode of continuous data*, 2002 ”http://www.mcg.edu/research/biostat/bickel.html.
4. Campbell, N. A, *Robust procedures in Multivariate Analysis I: Robust covariance Estimation*, 2002 ”http://www.mcg.edu/research/biostat/bickel.html.
5. Chen MK, Chien HN, Liu RL, *An Empirical Study on the Learning Experiences and Outcomes of College Student Club Committee Members Using a Linear Hierarchical Regression Model*, Appl Syst Innis. 2023;6(1):1–14.
6. Daniel P. Normolle, *An algorithm for robust nonlinear analysis of radioimmunoassays and other bioassays*, 2003 Ann Arbor,MI 48109-2029 , 313-936-1013.
7. David, T. O , *Applied Robust Statistics*, 2004; Chapman & Hall, New York.
8. de Jong PF , *Hierarchical regression analysis in structural equation modeling*, 1999;6(2):198-211.
9. Hayder Raaid Talib, Hassan Hopoop Razaq, Sarah Adel Madhioom, *Comparison between classical and robust estimation methods for regression model parameters in the case of incomplete data*, Adv Theory Nonlinear Anal Appl. 2024;7(4):17-34.
10. Hmood MY, Talib HR , *A method of multi-dimensional variable selection for additive partial linear models*, Chil J Stat. 2024;15(2):126–46.
11. Jurckova J, Sen PK , *Robust Statistical Procedures: Asymptotics and Interrelations*, New York: John Wiley & Sons; 2005.
12. Liu, C. & Rubin, D. B, *Maximum likelihood estimation of factor analysis using the ECME algorithm with complete and incomplete data*, 2002; statistica sinica 8, p. 437-446.
13. Liu C, Rubin DB, *Maximum likelihood estimation of factor analysis using the ECME algorithm with complete and incomplete data*, Stat Sin. 1998;8(3):729–47.
14. Majeed A. Yousif , Ravi P. Agarwal , Pshtiwan Othman Mohammed , Alina Alb Lupas, Rashid Jan and Nejmeddine Chorfi, *Advanced Methods for Conformable Time-Fractional Differential Equations: Logarithmic Non-Polynomial Splines*, Splines” Axioms 2024, 13, 551. https://doi.org/10.3390/axioms13080551.
15. Majeed A. Yousif and Faraidun K. Hamasal, *Novel simulation of the time fractional Burgers–Fisher equations using a non-polynomial spline fractional continuity method*, AIP Advances 12, 115018 (2022); https://doi.org/10.1063/5.0128819.
16. Morrison, D. F., *Expectation and Variances of Maximum Likelihood Estimates of the Multivariate Normal Distribution Parameters with Missing Data*, JASA, 1971; vol. 53, P. 326- 339.
17. Peter, J. H, *Robust Statistics*, 1981; John Wiley & Sons, New York.
18. 18. Richardson DB, Hamra GB, Ma Clehose RF, Cole SR, Chu H., *Hierarchical Regression for Analyses of Multiple Outcomes*, Am J Epidemiol. 2015;182(5):459-67.
19. Roche. A., *EM Algorithm and Variants: An Informal Tutorial*, ftp://ftp.cea.fr/pub/dsv/madic/pubils/roche-em .
20. SİPAHİOĞLU M, MANAP A, DİLEKÇİ Ü, YILDIRIM M, *Classroom Emotional Climate and School Burnout in Secondary School Students: A Hierarchical Regression Model*, Curr Res Soc Sci. 2023;9(2):113–29.
21. Talib HR, Madhloom SA, Aazer KK., *Comparison between classical and robust methods for linear model parameters in the presence of outlier’s values*, Edelweiss Appl Sci Technol. 2024;8(6):8506–13.