Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: www.tandfonline.com/journals/gscs20

# A mixed-effects least square support vector regression model for three-level count data

Mohammad Moqaddasi Amiri, Leili Tapak & Javad Faradmal

**To cite this article:** Mohammad Moqaddasi Amiri, Leili Tapak & Javad Faradmal (2019) A mixed-effects least square support vector regression model for three-level count data, Journal of Statistical Computation and Simulation, 89:15, 2801-2812, DOI: 10.1080/00949655.2019.1636991

**To link to this article:** https://doi.org/10.1080/00949655.2019.1636991

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
2019, VOL. 89, NO. 15, 2801–2812
https://doi.org/10.1080/00949655.2019.1636991

Taylor & Francis
Taylor & Francis Group

Check for updates

# A mixed-effects least square support vector regression model for three-level count data

Mohammad Moqaddasi Amiri[a], Leili Tapak[b] and Javad Faradmal[c]

[a]Department of Biostatistics and Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran; [b]Department of Biostatistics, School of Public Health, Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran; [c]Department of Biostatistics and Epidemiology, School of Public Health, Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

**ABSTRACT**
Hierarchical study design often occurs in many areas such as epidemiology, psychology, sociology, public health, engineering, and agriculture. This imposes correlation in data structure that needs to be account for in modelling process. In this study, a three-level mixed-effects least squares support vector regression (MLS-SVR) model is proposed to extend the standard least squares support vector regression (LS-SVR) model for handling cluster correlated data. The MLS-SVR model incorporates multiple random effects which allow handling unequal number of observations for each case at non-fixed time points (a very unbalanced situation) and correlation between subjects simultaneously. The methodology consists of a regression modelling step that is performed straightforwardly by solving a linear system. The proposed model is illustrated through numerical studies on simulated data sets and a real data example on human Brucellosis frequency. The generalization performance of the proposed MLS-SVR is evaluated by comparing to ordinary LS-SVR and some other parametric models.

**ARTICLE HISTORY**
Received 23 April 2019
Accepted 24 June 2019

**KEYWORDS**
Kernel functions; least square support vector machine; count data; three level; random effect; machine learning; Brucellosis; longitudinal analysis; hierarchical; nonlinear regression

## 1. Introduction

Clustered data are frequently found in many study areas such as medical and public health sciences. For example patients are nested within hospitals; hospitals are nested within cities or children are nested within families. When a hierarchical sampling structure is considered to collect the data or when the data has a cluster nature, an intraclass correlation (within multiple levels) is introduced in the data. In this setting, hierarchical models should be used to analyse the data in which the effect of levels or clusters is introduced in the model using random effect terms. If the clustering structure of the data is ignored, the residuals will not be distributed independently. Consequently, the type-I error rate increases in testing regression coefficients for statistical significance [1].

**CONTACT** Javad Faradmal javad.faradmal@umsha.ac.ir Hamadan University of Medical Sciences, P.O. Box 65175-4171, Hamadan, Iran

One of the most commonly used methods for analysing clustered data are mixed-effects model. Based on the type of response variable, there are various types of mixed-effects models. The linear mixed-effects model (LMM) is used for continues outcomes [2,3]. Generalized linear mixed-effects models (GLMM) have been introduced for non-Gaussian outcomes (binomial, counts, ...). For example, Poisson mixed-effects model (PMM) can be used for clustered count data [2]. A Zero-inflated Poisson mixed model (ZPMM) is used in the presence of excess zeroes for a count response [4]. Alternative models such as (zero-inflated) negative binomial regression can also be useful in the presence of over-dispersion [5] and its variants for correlated data are (Zero-inflated) Negative Binomial mixed-effects Model (NBMM) and ZNBMM [6,7]. The (classical) parametric models consider the linear relationship between features and response variable and are not able to deal with complex and nonlinear effects of predictors as well as their interactions especially when there are a large number of covariates.

Support vector machines (SVMs) proposed by Vapnik [8] is a kernel machine learning approach that has been widely used for nonlinear classification and function estimation problems. Unlike the parametric approaches such as LMM, the SVM considers the nonlinear complex relations among features by using kernel functions. It is interesting in a sense that it solves a nonlinear problem by mapping the feature space into a higher dimension space where the problem may be solved linearly [9]. Moreover, unlike parametric methods SVMs can handle many features [10]. Although the SVM has shown a good performance in classification and prediction problems, it solves a quadratic optimization problem and requires numerical optimization algorithms which makes it computationally intensive. Least squares support sector machine (LS-SVM) was introduced by Suykens [11] as a least square version of SVM. In LS-SVM, the nonlinear modelling problem is solved through convex optimization techniques without suffering from several local minima. Therefore a unique solution is obtained. The optimization problem in the LS-SVM is in a linear equation form which is simple to solve and time-saving computationally. In addition, the LS-SVM has been shown to have a good generalization performance on an unseen test data [12]. Least squares support vector regression (LS-SVR) is the regression or function estimation form of LS-SVM.

Several studies have been conducted on clustered or longitudinal data analysis using kernel learning or SVM methods. A mixed-effects LS-SVM was introduced to classify longitudinal profiles into groups [13]. However, the outcomes in this study do not change with time. A multiple kernel learning method with random effects was used for predicting the class of longitudinal outcomes [14]. Particularly, kernel-based random effects methods such as mixed-effects LS-SVR are proposed to predict continues longitudinal outcomes [12,15]. To the best of our knowledge, no study exists that have used the SVM approaches in the three-level data.

The above-mentioned studies have been done in a two-level setting. The intracluster correlation of the response is ignored and so the standard error of the mean of the response is estimated incorrectly when a simple random sample setting is used instead of the two-level cluster form [16]. As there can be found several situations where the data have a three level structure (e.g. the data on individuals are collected from centres in the second level within centres in the third level), additional correlation is introduced in the data. Ignoring this intracluster correlation in modelling procedure can lead to estimate the standard errors incorrectly [2]. In this study, we propose a three-level mixed-effects LS-SVR (MLS-SVR) model for predicting continues outcomes. In MLS-SVR, we use the random effect terms in LS-SVR as cluster effects. Then we compare the generalization performance of MLS-SVR with LS-SVR and parametric models in a real data example and a simulation study based on the real example with several scenarios.

The rest of the paper is organized as follows. First, we introduce the ordinary LS-SVR and propose the new MLS-SVR for three-level clustered data in Section 2. In Section 3, the numerical studies are performed through a real data example and a simulation study with several scenarios. Section 4 presents the discussion and conclusions.

## 2. Methods

In this section, we illustrate the standard LS-SVR along with the estimation technique. Then we propose the three-level MLS-SVR with estimation procedure for three-level cluster data. Also, we modify a generalized cross-validation (GCV) function presented in [12] to obtain the optimal regularization parameters.

### 2.1. Least square support vector regression

In this section, we describe the standard LS-SVR. Let the training data set be denoted by $G = \{(x_i, y_i)\}_{i=1}^{n}$ with covariate vector $x_i \in \mathbb{R}^d$ and the response$y_i \in \mathbb{R}$. In regression setting, the relationship between response variable and predictors is shown as follows

$$y(\boldsymbol{x}) = \omega^T \varphi(\boldsymbol{x}) + b \tag{1}$$

where the term $b$ is the bias term. The feature mapping function, $\varphi(\cdot)$, maps the input space into a higher dimensional feature space. The optimization problem of the nonlinear LS-SVR is considered as follows by using a regularization parameter $\lambda$

$$\min \quad J = \frac{1}{2}||\omega||^2 + \frac{\lambda}{2}\sum_{i=1}^{N} e_i^2 \tag{2}$$

over $(\omega,b,e)$ subject to equality constraints of

$$y_i = \omega^T \varphi(\boldsymbol{x}_i) + b + e_i, \quad i = 1, 2, \ldots, n \tag{3}$$

with weight vector $\omega \in \mathbb{R}^d$ and error variables $e_i \in \mathbb{R}$.

The primal Lagrange function can be constructed as

$$L = J - \sum_{i=1}^{N} \alpha_i(\omega^T \varphi(x_i) + b + e_i - y_i) \tag{4}$$

where $\alpha_i$ are the Lagrange multipliers. The conditions for optimality are given by

$$\begin{cases} \frac{\delta L}{\delta \omega} = 0 & \rightarrow \quad \omega = \sum_{i=1}^{N} \alpha_i \varphi(\boldsymbol{x}_i) \\ \frac{\delta L}{\delta b} = 0 & \rightarrow \quad \sum_{i=1}^{N} \alpha_i = 0 \\ \frac{\delta L}{\delta e_i} = 0 & \rightarrow \quad \alpha_i = \lambda e_i \\ \frac{\delta L}{\delta \alpha_i} = 0 & \rightarrow \quad \omega^T \varphi(\boldsymbol{x}_i) + b + e_i - y_i = 0 \end{cases} \tag{5}$$

After eliminating $e_i$ and $\omega$, we have the optimal values of $\alpha_i$ and $b$ by solving the linear equation as follows:

$$\begin{pmatrix} 0 & 1_n^T \\ 1_n & \Omega + I/\lambda \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \tag{6}$$

where $y = [y_1; y_2; \ldots; y_n]$, $\alpha = [\alpha_1; \ldots; \alpha_n]$, $1_n$ is the vector of ones of dimension $n$, $I_n$ is the identity matrix of dimension n and $\Omega$ is the $n \times n$ matrix with elements $\varphi(x_i)^T \varphi(x_j), i, j = 1, \ldots, n$.

The optimal values of bias ,$\hat{b}$, and Lagrange multipliers, $\hat{\alpha}_i$, are obtained by solving the linear equation. Then the optimal regression function for the given $x_0$ is obtained as:

$$y(x_0) = \sum_{i=1}^{N} \hat{\alpha}_i K(x_0, x_i) + \hat{b} \tag{7}$$

## *2.2. Three-level mixed effects LS-SVR*

Now we consider the mixed effects LS-SVR in three level cluster data setting. Given the training data set$G = \{(x_{ijk}, y_{ijk})\}_{i,j,k=1}^{N,n_i,n_{ij}}$, that $y_{ijk}$ is the $k$th response variable of the $j$th centre of the level-2in $i$th centre of the level-3corresponding to $p \times 1$ fixed effects covariates $x_{ijk}$. We assume that $y_{ijk}$ is related to $x_{ijk}$ in the regression form as

$$y_{ijk} = \omega^T \varphi(x_{ijk}) + b + \gamma_{0i} + \nu_{0ij} + e_{ijk} \quad i = 1, \ldots, N, j = 1, \ldots, n_i, k = 1, \ldots, n_{ij}. \tag{8}$$

where $\varphi(x_{ijk})$ is a nonlinear feature mapping function, $b$ is the bias term, $\nu_{0ij}$ is a random effect parameter for level-2 from $\mathrm{N}(0, \sigma_\nu^2)$, $\gamma_{0i}$ is a random effect parameter for level-3 from $\mathrm{N}(0, \sigma_\gamma^2)$ and $e_{ijk}$ is the residuals of the model from $\mathrm{N}(0, \sigma_e^2)$.

The optimization problem of the nonlinear three-level mixed effects LS-SVR can be define as

$$\min \quad J = \frac{1}{2}||\omega||^2 + \frac{\lambda_1}{2} \sum_{i=1}^{N} \gamma_{0i}^2 + \frac{\lambda_2}{2} \sum_{i=1}^{N} \sum_{j}^{n_i} \nu_{0ij}^2 + \frac{\lambda_3}{2} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} e_{ijk}^2 \tag{9}$$

Subject to equality constraints

$$y_{ijk} = \omega^T \varphi(x_{ijk}) + b + \gamma_{0i} + \nu_{0ij} + e_{ijk} \quad i = 1, \ldots, N, j = 1, \ldots, n_i, k = 1, \ldots, n_{ij}. \tag{10}$$

Here $\lambda_1, \lambda_2,\lambda_3$ are regularization parameters. The Lagrange function for this problem can be constructed as

$$L = J - \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \alpha_{ijk}(\omega^T\varphi(x_{ijk}) + b + \gamma_{0i} + \nu_{0ij} + e_{ijk} - y_{ijk}) \tag{11}$$

where $\alpha_{ijk}$ are Lagrange multipliers. The conditions for optimality are given by

$$\begin{cases} \frac{\delta L}{\delta\omega} = 0 & \rightarrow \quad \omega = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \alpha_{ijk}\varphi(x_{ijk}) \\ \frac{\delta L}{\delta b} = 0 & \rightarrow \quad \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \alpha_{ijk} = 0 \\ \frac{\delta L}{\delta\gamma_{0i}} = 0 & \rightarrow \quad \lambda_1\gamma_{0i} - \sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \alpha_{ijk} = 0 \\ \frac{\delta L}{\delta\nu_{ij}} = 0 & \rightarrow \quad \lambda_2\nu_{0ij} - \sum_{k=1}^{n_{ij}} \alpha_{ijk} = 0 \\ \frac{\delta L}{\delta e_{ijk}} = 0 & \rightarrow \quad \lambda_3 e_{ijk} - \alpha_{ijk} = 0 \\ \frac{\delta L}{\delta\alpha_{ijk}} = 0 & \rightarrow \quad \omega^T\varphi(x_{ijk}) + b + \gamma_{0i} + \nu_{0ij} + e_{ijk} - y_{ijk} = 0 \end{cases} \tag{12}$$

After eliminating $\nu_{0ij}, \gamma_{0i}, \omega, e_{ijk}$ we can find the optimal values of $\alpha_{ijk}$ and b by solving the following linear system equation as

$$\begin{pmatrix} 0 & 1_{N_t}^T \\ 1_{N_t} & K + \frac{1}{\lambda_1}\tilde{\Gamma} + \frac{1}{\lambda_2}\tilde{V} + \frac{1}{\lambda_1}I \end{pmatrix}\begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \tag{13}$$

where $N_t = \sum_{i=1}^{n_i}\sum_{j=1}^{n_{ij}} n_{ij}$,$0_{N_t}$ and $1_{N_t}$ are the (Nt×1) vector of zeros and ones respectively, $K$ is the kernel function in the form $K(x_{ijk}, x_0) = \varphi'(x_{ijk})\varphi(x_0)$, $\tilde{V} = diag(V_1, \ldots, V_N)$

where $V_s$ is a square matrix of one by $n_{ij}$ dimension and $m$ is the total number of centres in level-2,$\alpha = (\alpha_1', \ldots, \alpha_N')$ where $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{in_i})'$ and $\alpha_{ij} = (\alpha_{ij1}, \ldots, \alpha_{ijn_{ij}})'$, $y = (y_1', \ldots, y_N')$ where $y_i = (y_{i1}, \ldots, y_{in_i})'$ and $y_{ij} = (y_{ij1}, \ldots, y_{ijn_{ij}})'$, $\tilde{\Gamma} = diag(\Gamma_1, \ldots, \Gamma_N)$where $\Gamma_i$ is a square matrix of one by $n_i$ dimension.

The optimal regression function for the given $x_0$ is as follow

$$y(x_0) = \hat{b} + \sum_{i=1}^{N}\sum_{j=1}^{n_i}\sum_{k=1}^{n_{ij}} \hat{\alpha}_{ijk}K(x_{ijk}, x_0) + \hat{\gamma}_{0i} + \hat{\nu}_{0ij} \tag{14}$$

### 2.2.1. Regularization parameters

The functional structure of the three-level mixed effects LS-SVR is characterized by hyper-parameters, the regularization parameters $\lambda_1, \lambda_2, \lambda_3$ and kernel function parameters. To obtain optimal values of these hyperparameters, we use the GCV function. The inverse of the leftmost matrix in (13) can be partitioned into follow submatrices as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{15}$$

where $S_{11}$ is a scalar, $S_{12}$ is a $(1 \times N_t)$ vector, $S_{21}$ is a $(N_t \times 1)$ vector and $S_{22}$ is a $(N_t \times N_t)$ matrix.

Then $\hat{y}$ can be expressed as $\hat{y} = Sy$, where $S = 1S_{12} + (K + (1/\lambda_1)\tilde{\Gamma} + (1/\lambda_2)\tilde{V} + (1/\lambda_1)I)S_{22}$. The GCV function can be obtained by applying the leave-one-out method and the first order Taylor expansion as follows

$$\mathrm{GCV}(\theta) = \frac{N_t y'(I - S)'(I - S)y}{(N_t - tr(S))^2} \tag{16}$$

where $\theta$ is a set of hyperparameters ($\lambda_1, \lambda_2, \lambda_3$, and kernel parameters). The optimal values of hyperparameters are that minimize the GCV function.

## 3. Numerical studies

In this section, we investigate the performance of the proposed MLS-SVR through a real example data set analysis related to the human Brucellosis frequency and a simulation study with several scenarios. In these numerical studies, we compared the prediction performance of the MLS-SVR with LS-SVR, LMM, PMM, NBMM, ZPMM, and ZNBMM.

### 3.1. Human brucellosis frequency example

The brucellosis data set contains monthly number of patients diagnosed with human brucellosis from cities of five provinces in the west of Iran (Hamadan, Kermanshah, Lorestan, Ilam, and Kordestan) in 2016–2017. We use the count of brucellosis as the response variable. Some climatic variables such as monthly sum of sunshine hours, monthly averages of temperature, humidity, rainfall and wind speed are considered as covariates. Moreover, we use the ratio of rural to urban population, livestock (cattle, sheep, and goat) population (frequency) and area of croplands, forest, and grassland (hectare) as covariates. These covariates have been extracted from Statistical Yearbook of Iran in 2016–2017 and IRAN Meteorological Organization site [17,18].

We have 12 observations (the observations related to 12 months for each city; level-1) in each city (level-2) and 54 cities nested in five provinces (level-3). So the total observations are 648. The data partitioning technique is used for making the train and test data set as follows [12]. For the training data, we randomly take the two-thirds of 12 observations for each city (eight observations) and add them together for all cities. The rest of them are used as the testing data. So the size of the training and testing data sets was 432 and 216 for brucellosis data respectively. According to the type of response variable (count) and the existence of excess zero, we fit MLS-SVR and LS-SVR with radial basis function (RBF) as the kernel. The optimal values of the hyperparameters were obtained from GCV function. Also some other compatible parametric mixed-effects models for three-level data such as LMM, PMM, NBMM, ZPMM, and ZNBMM were applied on training data and the

**Table 1.** Performance comparison of proposed MLS-SVR and other models for brucellosis example (standard error in parenthesis).

| Measure | MLS-SVR | LS-SVR | LMM | PMM | NBMM | ZPMM | ZNBMM |
|---|---|---|---|---|---|---|---|
| MSE | **27.97** | 60.12 | 31.54 | 36.84 | 35.27 | 36.52 | 35.27 |
| | (0.58) | (0.99) | (0.62) | (3.18) | (2.71) | (3.17) | (2.71) |
| MAE | 3.29 | 5.07 | 3.43 | 3.22 | **3.21** | **3.21** | **3.21** |
| | (0.02) | (0.02) | (0.02) | (0.03) | (0.03) | (0.03) | (0.03) |

**Table 2.** Performance comparison of proposed MLS-SVR and other models for brucellosis example with 12-fold cross-validation.

| Measure | MLS-SVR | LS-SVR | LMM | PMM | NBMM | ZPMM | ZNBMM |
|---|---|---|---|---|---|---|---|
| MSE | **29.03** | 60.97 | 32.94 | 48.32 | 47.43 | 48.32 | 47.43 |
| MAE | **3.31** | 5.14 | 3.57 | 3.68 | 3.68 | 3.68 | 3.68 |

parameters of each model were estimated. Then we used the testing data set to investigate the generalization performance of each model. This procedure was repeated 100 times. We use mean square error (MSE) and mean absolute error (MAE) for comparison of fitted models generalization. The results were given in Table 1.

According to Table 1, the proposed MLS-SVR yielded the smallest mean of MSEs. The means of MAEs for NBMM, ZPMM, and ZNBMM are smaller than proposed MLS-SVR, although the standard error of the MLS-SVR was smaller than that of the parametric models.

In order to compare the observed and estimated (predicted) values of the human Brucellosis frequency, we used a twelve-fold cross-validation technique. So, the observations of each month were placed in one fold. Therefore, each time point has one prediction value. We fit the MLS-SVR, LS-SVR, LMM, PMM, NBMM, ZPMM, and ZNBMM and compare the generalization performance with MSE and MAE measures.

The results of the twelve-fold cross-validation were reported in Table 2. Table 2 describes the mean of MSEs and MAEs for proposed MLS-SVR which shows that it is smaller than other methods.

Also, we draw the box-plot of predictions for each method in Figure 1. These box-plots were drawn in four actual observed values (0, 1, 5, and 10). Figure 1 shows that the prediction accuracy of the proposed MLS-SVR for five and 10 values is better than zero and one values.

### 3.2. Simulation example

A Monte Carlo simulation study was conducted to assess the performance of different approaches in prediction of level1 observations.

Suppose $y_{ijk}$ be the response of $k$th subject within $j$th centre of the level-2within $i$th centre of the level-3 related to covariates $x_{ijk}^{(p)}$. We generated data from zero-inflated negative binomial distribution with mean $\mu_{ijk} = \exp\{\,(Beta * \tau_{ijk}) + b + \gamma_{0i} + \nu_{0ij}$ and set the dispersion parameter equal to 5.5 (dispersion parameter obtained from the brucellosis data) where $\tau_{ijk} = (x_{ijk}^{(1)}, (\sin((\pi/4)x_{ijk}^{(p)})/(\pi/4)x_{ijk}^{(p)})), \quad p = 2, \ldots, 13, i = 1, \ldots, N, j = 1, \ldots, n_i, k = 1, \ldots, n_{ij}, x_{ijk}^{(1)}$ is the time variable (month $= 1, \ldots, 12$), $b = 1.2$, $Beta = (0.001,$

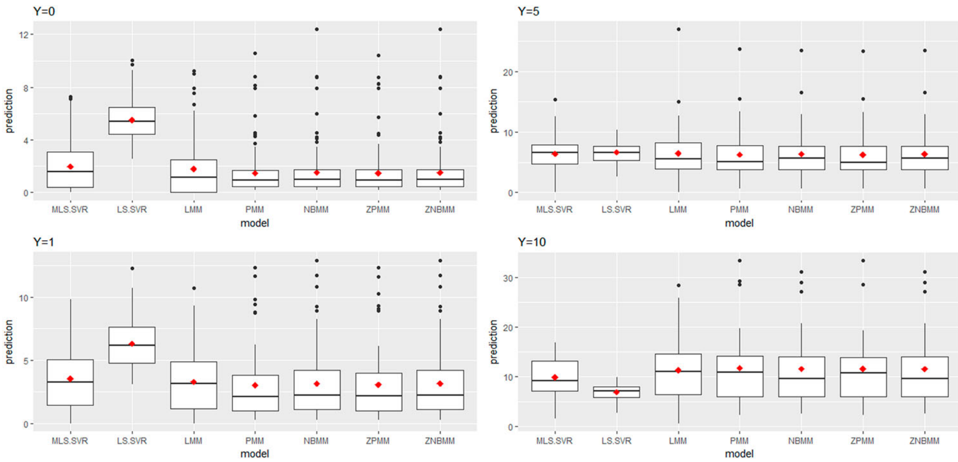


**Figure 1.** Box-plots of predictions in 12-fold cross-validation (red points are the mean of predictions).

0.04,−0.08,0.09,0.17,0.04,−0.03,−0.37,0.3,−0.21, 0.05,−0.22, −0.2), the covariates $x_{ijk}^{(p)}$, $p = 2, \ldots, 13$ generated from normal distribution with zero mean and unit variance. The correlation between the first five covariates is

$$\begin{bmatrix} 1 & -0.6 & 0.0 & 0.4 & -0.6 \\ & 1 & -0.6 & -0.9 & 0.9 \\ & & 1 & 0.7 & -0.6 \\ & & & 1 & -0.8 \\ & & & & 1 \end{bmatrix} \qquad (17)$$

### *3.2.1. Simulation scenarios*

We considered two different values (5 and 10) for the number of centres at level-2 and level-3. The sample size of level-1 was 12 (number of the month) for each centre of level-2 in all scenarios. The variance of the random effects was considered in two levels as (0.02, 0.9) and (0.02, 0.5) for level-2 and level-3, respectively. We considered the probability of the zero in zero-inflated negative binomial distribution in two levels (0 and 0.1). There were 32 settings in this simulation study. In all conditions, we used data partitioning technique for splitting data to training and testing sets, by taking one-third of the 12 observations in each centre of level-2 (4 observations) and adding them together for the testing data set. The rest of them were used as the training data set. We fitted the MLS-SVR, LS-SVR, LMM, PMM, NBMM, ZPMM, and ZNBMM on the training data and estimated the parameters of each model. The RBF was used as the kernel function in the MLS-SVR and LS-SVR methods. Then, we evaluated the generalization performance for each model using two measures of MSE and MAE on the test data set. This procedure was repeated 100 times in each scenario.

The results of the simulation study were given in Table 3. The mean of MSEs for the proposed MLS-SVR model was smaller than that of the other methods in all simulation scenarios. Also, according to the mean of MAEs, the generalization performance of the proposed MLS-SVR was better than that of the other models except in the limited number of simulation scenarios.

## 4. Discussion and conclusions

This paper proposed a new least square support vector regression (LS-SVR) method for cluster-correlated data situation. We used random effect terms as cluster effects in ordinary LS-SVR. An important strength of our proposed model is its flexibility in modelling (non)linear and complex relationships between predictors and response, while it takes into account the hierarchical structure of the data and it is computationally efficient due to the need to solve only a linear system.

Through areal data example, we observed that the proposed MLS-SVR provided a better result in terms of prediction performance based on the MSE measure. According to the MAE measure, the proposed MLS-SVR had better generalization performance than the LS-SVR and LMM, but the performance of the PMM, NBMM, ZPMM, and ZNBMM was better than the MLS-SVR. Nevertheless, the later methods had a greater standard error compared the proposed method in Table 1. The reason for this result can be found in the definition and properties of MSE and MAE measures. Unlike the MAE measure, the MSE is sensitive to the outliers and large errors. In Figure 1, the box-plots of the PMM, NBMM, ZPMM, and ZNBMM showed that there were several outlier values in the obtained predictions using these methods. The effects of these outliers and large errors were shown in MSE and were ignored in MAE.

According to Figure 1, the proposed model was more accurate in the prediction of large values (5 and 10) than the small values (zero and one). Also, the PMM, NBMM, ZPMM, and ZNBMM resulted in more outlier values in predictions than the proposed MLS-SVR.

In our simulation studies, the generalization performance of the proposed MLS-SVR model was better than other methods based on MSE measure in all scenarios. Moreover, it outperformed other methods in terms of the MAE measure in lots of scenarios. The reason that in some scenarios, the proposed MLS-SVR did not yield the smallest mean of the MAEs can be found in the properties of the MSE and MAE measures. The existence of outliers and large errors affect the MSE greatly.

Although the sample size of level-1 was fixed in the real data example, the proposed method can be used for the unbalanced three-level clustered data. Like other SVM methods, the proposed MLS-SVR can be used in the high dimensional data.

In this paper, we used RBF as the kernel function. The other kernel functions like polynomial can be used in the proposed MLS-SVR. Also, the amount of dispersion parameter was fixed in all simulation scenarios. Simulation with other dispersion parameters can be considered in future studies.

If we introduce a symmetric and positive definite block-diagonal matrix like $W$ as a weight in Equation (2) and solve the related optimization problem, then this new approach can be considered as the weighted method of LS-SVR [19] and be used in clustered data sets. Indeed, the $^1/_\lambda$ term in Equation (6) replaced by $W^{-1}$and the $\tilde{\Gamma}$, $\tilde{V}$, and I in Equation (13) can be extracted from $W^{-1}$.

In the proposed MLS-SVR, we used only the random intercept effect. We will consider the random intercept and trend effects in a future study as soon. Also, We are

**Table 3.** Performance comparison of proposed MLS-SVR and other models for simulation study (standard error in parenthesis).

| | | | MLS-SVR | | LS-SVR | | LMM | | PMM | | NBMM | | ZPMM | | ZNBMM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Level-3 centres (variance) | Number of Level-2 centres (variance) | Probability of Excess Zero | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 5 | 5 | 0.1 | **34.54** | **2.66** | 68.68 | 4.29 | 37.25 | 2.96 | 47.54 | 2.88 | 43.94 | 2.80 | 40.49 | 2.76 | 37.32 | 2.72 |
| (0.8) | (0.5) | | (7.00) | (0.14) | (12.1) | (0.29) | (7.70) | (0.16) | (11.5) | (0.16) | (12.5) | (0.16) | (8.22) | (0.13) | (7.43) | (0.13) |
| | | 0 | **17.09** | **2.18** | 55.79 | 3.98 | 18.60 | 2.36 | 25.55 | 2.31 | 20.19 | 2.26 | 25.11 | 2.32 | 20.34 | 2.26 |
| | | | (2.47) | (0.10) | (12.8) | (0.27) | (2.86) | (0.11) | (7.15) | (0.11) | (3.48) | (0.10) | (6.69) | (0.11) | (3.58) | (0.10) |
| 10 | 5 | 0.1 | **25.90** | 2.49 | 63.56 | 4.15 | 26.70 | 2.62 | 28.98 | 2.55 | 27.69 | 2.51 | 27.90 | 2.50 | 26.86 | **2.48** |
| (0.8) | (0.5) | | (2.81) | (0.07) | (7.98) | (0.16) | (2.82) | (0.08) | (3.22) | (0.08) | (2.94) | (0.07) | (3.13) | (0.07) | (2.86) | (0.07) |
| | | 0 | **15.76** | **2.14** | 61.45 | 3.96 | 16.19 | 2.24 | 17.88 | 2.20 | 17.18 | 2.17 | 17.79 | 2.20 | 17.20 | 2.18 |
| | | | (1.96) | (0.06) | (14.2) | (0.19) | (1.91) | (0.07) | (2.13) | (0.06) | (1.97) | (0.06) | (2.12) | (0.06) | (1.97) | (0.06) |
| 5 | 10 | 0.1 | **25.84** | 2.51 | 73.17 | 4.13 | 27.21 | 2.64 | 27.88 | 2.56 | 26.87 | 2.53 | 26.76 | 2.52 | 26.24 | **2.50** |
| (0.8) | (0.5) | | (3.63) | (0.11) | (19.8) | (0.27) | (4.15) | (0.12) | (4.03) | (0.11) | (3.48) | (0.10) | (3.71) | (0.10) | (3.46) | (0.10) |
| | | 0 | **20.30** | **2.29** | 58.38 | 4.01 | 20.91 | 2.39 | 21.59 | 2.33 | 21.32 | 2.30 | 21.60 | 2.33 | 21.32 | 2.31 |
| | | | (3.30) | (0.09) | (9.71) | (0.23) | (3.32) | (0.10) | (3.19) | (0.09) | (3.36) | (0.09) | (3.21) | (0.09) | (3.36) | (0.09) |
| 10 | 10 | 0.1 | **22.34** | 2.41 | 54.94 | 3.95 | 22.62 | 2.47 | 24.02 | 2.43 | 22.88 | 2.41 | 23.03 | 2.40 | 22.38 | **2.39** |
| (0.8) | (0.5) | | (1.93) | (0.07) | (6.50) | (0.15) | (1.96) | (0.07) | (2.18) | (0.07) | (2.00) | (0.07) | (1.98) | (0.07) | (1.86) | (0.07) |
| | | 0 | **16.10** | 2.20 | 53.56 | 4.06 | 16.25 | 2.24 | 16.92 | 2.21 | 16.70 | **2.19** | 16.94 | 2.21 | 16.71 | 2.20 |
| | | | (1.31) | (0.06) | (6.10) | (0.17) | (1.31) | (0.06) | (1.38) | (0.06) | (1.35) | (0.06) | (1.38) | (0.06) | (1.35) | (0.06) |
| 5 | 5 | 0.1 | **7.15** | **1.83** | 10.90 | 2.29 | 7.77 | 1.94 | 8.39 | 1.93 | 8.08 | 1.91 | 8.00 | 1.91 | 7.89 | 1.90 |
| (0.04) | (0.5) | | (0.29) | (0.03) | (0.60) | (0.05) | (0.33) | (0.03) | (0.40) | (0.03) | (0.37) | (0.03) | (0.34) | (0.03) | (0.34) | (0.03) |
| | | 0 | **7.28** | **1.78** | 14.38 | 2.47 | 7.78 | 1.88 | 8.13 | 1.85 | 8.03 | 1.84 | 8.11 | 1.85 | 8.02 | 1.84 |
| | | | (0.52) | (0.03) | (1.36) | (0.06) | (0.52) | (0.03) | (0.53) | (0.03) | (0.53) | (0.03) | (0.55) | (0.03) | (0.53) | (0.03) |
| 10 | 5 | 0.1 | **8.23** | **1.90** | 13.33 | 2.43 | 8.52 | 1.95 | 8.82 | 1.94 | 8.72 | 1.92 | 8.64 | 1.93 | 8.57 | 1.92 |
| (0.04) | (0.5) | | (0.36) | (0.02) | (0.78) | (0.04) | (0.37) | (0.02) | (0.41) | (0.02) | (0.42) | (0.02) | (0.40) | (0.02) | (0.39) | (0.02) |
| | | 0 | **6.80** | **1.78** | 12.31 | 2.37 | 7.01 | 1.82 | 7.28 | 1.81 | 7.16 | 1.80 | 7.26 | 1.81 | 7.16 | 1.80 |
| | | | (0.23) | (0.02) | (0.59) | (0.04) | (0.23) | (0.02) | (0.26) | (0.02) | (0.25) | (0.02) | (0.26) | (0.02) | (0.25) | (0.02) |
| 5 | 10 | 0.1 | **8.04** | **1.88** | 13.99 | 2.40 | 8.26 | 1.92 | 8.61 | 1.92 | 8.53 | 1.90 | 8.41 | 1.90 | 8.31 | 1.89 |
| (0.04) | (0.5) | | (0.43) | (0.03) | (1.39) | (0.05) | (0.43) | (0.03) | (0.49) | (0.03) | (0.49) | (0.03) | (0.50) | (0.03) | (0.46) | (0.03) |
| | | 0 | **6.56** | **1.78** | 11.67 | 2.37 | 6.80 | 1.82 | 6.96 | 1.81 | 6.92 | 1.80 | 6.96 | 1.82 | 6.92 | 1.80 |
| | | | (0.20) | (0.02) | (0.49) | (0.04) | (0.21) | (0.02) | (0.23) | (0.02) | (0.23) | (0.02) | (0.22) | (0.02) | (0.23) | (0.02) |
| 10 | 10 | 0.1 | **7.60** | **1.85** | 12.30 | 2.36 | 7.70 | 1.87 | 7.79 | 1.87 | 7.75 | **1.85** | 7.61 | **1.85** | **7.60** | **1.85** |
| (0.04) | (0.5) | | (0.21) | (0.02) | (0.51) | (0.03) | (0.22) | (0.02) | (0.20) | (0.02) | (0.22) | (0.02) | (0.20) | (0.02) | (0.21) | (0.02) |
| | | 0 | **6.62** | **1.77** | 12.08 | 2.38 | 6.73 | 1.79 | 6.84 | 1.79 | 6.77 | **1.77** | 6.84 | 1.78 | 6.77 | **1.77** |
| | | | (0.15) | (0.01) | (0.39) | (0.03) | (0.15) | (0.02) | (0.16) | (0.02) | (0.15) | (0.02) | (0.16) | (0.02) | (0.15) | (0.02) |

(*continued*).

**Table 3.** Continued.

| | | | MLS-SVR | | LS-SVR | | LMM | | PMM | | NBMM | | ZPMM | | ZNBMM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of Level-3 centres (variance) | Number of Level-2 centres (variance) | Probability of Excess Zero | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| 5 | 5 | 0.1 | **11.79** | **2.04** | 24.49 | 2.93 | 12.17 | 2.12 | 13.50 | 2.11 | 12.78 | 2.07 | 12.72 | 2.08 | 12.48 | 2.06 |
| (0.8) | (0.04) | | (1.50) | (0.08) | (5.76) | (0.19) | (1.59) | (0.09) | (2.01) | (0.08) | (1.74) | (0.08) | (1.72) | (0.08) | (1.65) | (0.08) |
| | | 0 | **10.75** | **1.98** | 23.75 | 3.08 | 11.17 | 2.07 | 11.79 | 2.02 | 11.53 | 2.01 | 11.76 | 2.02 | 11.53 | 2.02 |
| | | | (1.09) | (0.07) | (3.02) | (0.17) | (1.13) | (0.08) | (1.22) | (0.07) | (1.18) | (0.07) | (1.21) | (0.07) | (1.18) | (0.07) |
| 10 | 5 | 0.1 | **10.96** | 2.01 | 21.46 | 2.88 | 11.15 | 2.04 | 12.03 | 2.04 | 11.42 | 2.01 | 11.29 | 2.00 | 11.13 | **1.99** |
| (0.8) | (0.04) | | (0.91) | (0.05) | (2.69) | (0.11) | (1.02) | (0.06) | (1.22) | (0.06) | (1.03) | (0.05) | (1.00) | (0.05) | (0.99) | (0.05) |
| | | 0 | **9.76** | **1.91** | 22.24 | 2.93 | 9.81 | 1.94 | 10.19 | 1.93 | 9.96 | **1.91** | 10.12 | 1.92 | 9.97 | **1.91** |
| | | | (1.19) | (0.05) | (3.52) | (0.13) | (1.18) | (0.05) | (1.18) | (0.05) | (1.20) | (0.05) | (1.13) | (0.05) | (1.20) | (0.05) |
| 5 | 10 | 0.1 | **11.15** | **2.03** | 21.02 | 2.95 | 11.52 | 2.10 | 12.22 | 2.10 | 11.60 | 2.06 | 11.67 | 2.06 | 11.41 | 2.04 |
| (0.8) | (0.04) | | (0.94) | (0.07) | (2.30) | (0.14) | (0.96) | (0.07) | (1.05) | (0.07) | (0.97) | (0.07) | (1.00) | (0.07) | (0.95) | (0.07) |
| | | 0 | **11.41** | **2.01** | 28.36 | 3.23 | 11.83 | 2.08 | 13.12 | 2.06 | 12.30 | 2.04 | 13.11 | 2.06 | 12.31 | 2.04 |
| | | | (1.54) | (0.08) | (5.63) | (0.22) | (1.64) | (0.09) | (2.45) | (0.09) | (1.93) | (0.09) | (2.45) | (0.09) | (1.93) | (0.09) |
| 10 | 10 | 0.1 | **13.72** | 2.10 | 30.69 | 3.14 | 13.88 | 2.13 | 14.75 | 2.14 | 14.19 | 2.11 | 14.40 | 2.11 | 13.99 | **2.09** |
| (0.8) | (0.04) | | (1.85) | (0.06) | (6.47) | (0.17) | (1.85) | (0.07) | (2.10) | (0.07) | (2.00) | (0.07) | (2.08) | (0.07) | (1.94) | (0.06) |
| | | 0 | **10.15** | **1.89** | 27.85 | 3.04 | 10.31 | 1.92 | 11.25 | 1.92 | 10.45 | **1.89** | 11.22 | 1.92 | 10.44 | **1.89** |
| | | | (1.29) | (0.06) | (6.23) | (0.16) | (1.34) | (0.06) | (1.78) | (0.06) | (1.36) | (0.05) | (1.78) | (0.06) | (1.36) | (0.05) |
| 5 | 5 | 0.1 | **3.99** | **1.56** | 4.13 | 1.58 | 4.30 | 1.61 | 4.41 | 1.62 | 4.34 | 1.61 | 4.35 | 1.62 | 4.33 | 1.61 |
| (0.04) | (0.04) | | (0.09) | (0.02) | (1.00) | (0.02) | (0.10) | (0.02) | (0.10) | (0.02) | (0.10) | (0.02) | (0.10) | (0.02) | (0.10) | (0.02) |
| | | 0 | **3.76** | **1.49** | 4.01 | 1.54 | 4.00 | 1.54 | 4.03 | 1.54 | 4.02 | 1.54 | 4.06 | 1.55 | 4.03 | 1.54 |
| | | | (0.09) | (0.02) | (0.11) | (0.02) | (0.10) | (0.02) | (0.09) | (0.02) | (0.10) | (0.02) | (0.10) | (0.02) | (0.10) | (0.02) |
| 10 | 5 | 0.1 | **4.04** | **1.56** | 4.18 | 1.59 | 4.15 | 1.58 | 4.20 | 1.58 | 4.16 | 1.58 | 4.16 | 1.58 | 4.15 | 1.58 |
| (0.04) | (0.04) | | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) |
| | | 0 | **3.79** | **1.49** | 4.04 | 1.54 | 3.88 | 1.51 | 3.89 | 1.51 | 3.88 | 1.51 | 3.90 | 1.51 | 3.88 | 1.51 |
| | | | (0.07) | (0.01) | (0.08) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) |
| 5 | 10 | 0.1 | **4.12** | **1.58** | 4.29 | 1.61 | 4.22 | 1.60 | 4.29 | 1.60 | 4.22 | 1.60 | 4.26 | 1.60 | 4.23 | 1.60 |
| (0.04) | (0.04) | | (0.07) | (0.01) | (0.08) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) | (0.07) | (0.01) |
| | | 0 | **3.89** | **1.51** | 4.15 | 1.57 | 3.96 | 1.53 | 3.98 | 1.53 | 3.97 | 1.53 | 3.97 | 1.53 | 3.97 | 1.53 |
| | | | (0.08) | (0.01) | (0.10) | (0.02) | (0.08) | (0.01) | (0.08) | (0.01) | (0.08) | (0.01) | (0.08) | (0.01) | (0.08) | (0.01) |
| 10 | 10 | 0.1 | **3.97** | **1.55** | 4.16 | 1.58 | 4.01 | **1.55** | 4.07 | 1.56 | 4.01 | **1.55** | 4.02 | **1.55** | 4.01 | **1.55** |
| (0.04) | (0.04) | | (0.06) | (0.01) | (0.06) | (0.01) | (0.06) | (0.01) | (0.06) | (0.01) | (0.06) | (0.01) | (0.06) | (0.01) | (0.06) | (0.01) |
| | | 0 | **3.68** | **1.49** | 3.89 | 1.63 | 3.71 | **1.49** | 3.74 | **1.49** | 3.71 | **1.49** | 3.73 | **1.49** | 3.71 | **1.49** |
| | | | (0.05) | (0.01) | (0.05) | (0.01) | (0.05) | (0.01) | (0.05) | (0.01) | (0.05) | (0.01) | (0.05) | (0.01) | (0.05) | (0.01) |

going to extend the MLS-SVR method in the future studies for obtaining the robust estimations using a weighted method to decrease the effect of y-outliers in multilevel data sets.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This study was a part of PhD thesis of the first author in Biostatistics and funded by the Vice-Chancellor for Research and Technology of Hamadan University of Medical Sciences [grant number 9609286041].

## References

[1] McNeish DM, Stapleton LM. The effect of small sample size on two-level model estimates: a review and illustration. Educ Psychol Rev. 2016;28(2):295–314.
[2] Hedeker D, Gibbons RD. Longitudinal data analysis. Vol. 451. New Jersey: John Wiley & Sons; 2006.
[3] Verbeke G. Linear mixed models for longitudinal data. Linear mixed models in practice. New York: Springer; 1997. p. 63–153.
[4] Hall DB. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics. 2000;56(4):1030–1039.
[5] Hilbe JM. Negative binomial regression. New York: Cambridge University Press; 2011.
[6] Booth JG, Casella G, Friedl H, et al. Negative binomial loglinear mixed models. Stat Model. 2003;3(3):179–191.
[7] Yau KK, Wang K, Lee AH. Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. Biometrical J. 2003;45(4):437–452.
[8] Vapnik V. Statistical learning theory. Vol. 3. New York: Wiley; 1998.
[9] Rohwer JA, Abdallah CT. One-vs-one multiclass least squares support vector machines for direction of arrival estimation. Albuquerque (NM): Sandia National Labs; 2003.
[10] Joachims T. Text categorization with support vector machines: learning with many relevant features. Machine Learning: ECML-98. 1998:137–142.
[11] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett. 1999;9(3):293–300.
[12] Shim J, Sohn I, Hwang C. Kernel-based random effect time-varying coefficient model for longitudinal data. Neurocomputing. 2017;267:500–507.
[13] Luts J, Molenberghs G, Verbeke G, et al. A mixed effects least squares support vector machine model for classification of longitudinal data. Comput Stat Data Anal. 2012;56(3):611–628.
[14] Chen T, Zeng D, Wang Y. Multiple kernel learning with random effects for predicting longitudinal outcomes and data integration. Biometrics. 2015;71(4):918–928.
[15] Seok KH, Shim J, Cho D, et al. Semiparametric mixed-effect least squares support vector machine for analyzing pharmacokinetic and pharmacodynamic data. Neurocomputing. 2011;74(17):3412–3419.
[16] Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Anal. 2006;1(3):473–514.
[17] Statistical center of Iran, Statistical Yearbook of Iran in 2016–2017. Available from: https://www.amar.org.ir.
[18] IRAN Meteorological Organization. 2016–2017. Available from: http://irimo.ir.
[19] Suykens JAK, De Brabanter J, Lukas L, et al. Weighted least squares support vector machines: robustness and sparse approximation. Neurocomputing. 2002;48(1-4):85–105.