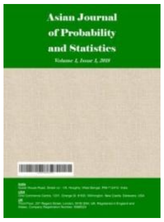


Asian Journal of Probability and Statistics

1(2): 1-21, 2018; Article no.AJPAS.41699

Modeling Count Data from Dependent Clusters with Poisson Mixed Models

K. A. N. K. Karunarathna[1,2*] and Pushpakanthie Wijekoon[3]

[1]*Department of Mathematics, Faculty of Science, Eastern University, Sri Lanka.*
[2]*Postgraduate Institute of Science, University of Peradeniya, Sri Lanka.*
[3]*Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka.*

Authors' contributions

This work was carried out in collaboration between both authors. Author PW designed and supervised the study. Author KANKK managed the literature survey, performed the statistical analysis and wrote the first draft of the manuscript. Author PW edited the manuscript. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2018/v1i224505
Editor(s):
(1) S. M. Aqil Burney, Department of Computer Science, University of Karachi, Pakistan.
Reviewers:
(1) Loc Nguyen, The University of Science, Vietnam.
(2) Bayo Lawal, Kwara State University, Nigeria.
Complete Peer review History: http://www.sciencedomain.org/review-history/25051

Received 18th March 2018
Accepted 27th May 2018
Published 8th June 2018

Original Research Article

ABSTRACT

Responses collected from dependent clusters are affected by the dependence among clusters and it should be taken into account in modeling such responses. In this study, a new approach was evaluated to incorporate cluster dependence in generalized linear Poisson mixed models for count responses from dependent clusters. Performance of this approach was evaluated by using a simulation process under three different designs and different covariates. The Marginal Generalized Quasi-likelihood (GQL) method was used for estimation of parameters with the cluster dependence. Monte Carlo likelihood (MCL) and Penalized Quasi-likelihood (PQL) estimates also were obtained for the purpose of comparison. Proposed approach was tested with a real data set also.

The proposed approach, with the incorporation of cluster dependence, gives better estimates for both fixed effects and variance of random effects with low standard errors with compared to the estimates obtained by ignoring the cluster dependence. Therefore, the proposed approach can be used for modeling count responses from a dependent cluster set up.

**Corresponding author: E-mail: nkkarunarathna@gmail.com;*

Keywords: Poisson mixed model; cluster counts; inter correlations; dependent clusters.

1. INTRODUCTION

Discrete data analysis is used in most of fields, and it consists of the analysis of count or binary data. Count data may be recorded for a cluster (panel, group or family). Depending on the situation, a cluster may be an individual or a set of individuals. A family in a village, a district in a province, a plot in a field, or even a location in an area may be considered as a cluster.

When a set of clusters is considered, they may be independent or dependent on others based on distances among clusters. When clusters have situated far away from each other, dependence among clusters is negligible, and clusters are assumed to be independent. But, this assumption is violated when clusters are closer to each other, because clusters tend to be dependent on other, especially on adjacent clusters. In this case, when responses from clusters are studied, effects come through dependence among clusters are also should be taken into consideration.

As a real example, assume that a researcher is interested to know effects of certain factors such as gender, educational level, and age on number of visits by family members to a family doctor (a measure for health status) within a year. In this example, families under study can be considered as clusters. It is obvious that health status of a family depends on the environment, especially on health status of adjacent families. Degree of dependence on health status of other families changes depending on distances among families. If families are far away from others, dependence of health status of a family on adjacent families may be negligible. When families are nearby, health status of a family is affected by its' adjacent families. This dependence can changes the effects of factors mentioned above. Therefore, in modeling effects of such factors on response, dependence among clusters also should be taken into account. Otherwise, effects of those factors may be incorrectly estimated.

A few authors have studied this problem for responses in continuous scale as well as in discrete scale.

Mariathas and Sutradhar [1] have studied this sort of dependence from locations for continuous responses with the linear mixed model by forming families of locations and relevant location specific random effects.

$$y_r = x_r'\beta + w_r'\boldsymbol{\gamma}_r^* + \varepsilon_r, \quad (1)$$

where $\boldsymbol{x}_r$ is a p-dimensional vector of covariates recorded for r th location, $\boldsymbol{\beta}$ is the regression covariate effects, $\mathbf{w}_r'$ is the weight vector corresponding to random effects vector $\boldsymbol{\gamma}_r^*$, and the random error ε_r is normally distributed with a zero mean and a constant variance. In defining covariance structure, they have considered decomposition of two families(say f_r and f_s) and number of common (n_{rs}^*) and uncommon($n_i^*; i = r, s$) locations. Then, covariance between y_r and y_s is defined in term of a distance based truncated equi-correlation for random effects of i th and j th locations, γ_i and γ_j as follows:

$$Corr(\gamma_i, \gamma_j) = \begin{cases} 1 & : \ if\ d = 0 \\ \phi & : \ if\ 0 < d \le d^* \\ 0 & : \ if\ d > d^* \end{cases} \quad (2)$$

where d is the euclidian distance between i th and j th locations, and d^* is the pre-specified distance defined by the researcher. Then, the first order pair-wise covariance between responses of r^{th} and s^{th} locations was shown as

$$Cov(y_r, y_s) = \frac{\sigma_\gamma^2}{\sqrt{n_r n_s}} \begin{bmatrix} \varphi\left\{n_{rs} + n_{rs}^*(n_r^* + n_s^*\right\} \\ + \left\{\varphi n_{rs}^*(n_{rs}^* - 1) + n_{rs}^*\right\} \end{bmatrix}, \quad (3)$$

where n_{rs} is the number of correlated pairs of locations among $n_r^* + n_s^*$ locations.

The moment- generalized least square hybrid method and maximum likelihood methods have been used to estimate β parameters in model (1), while the method of moments and maximum likelihood method have been used to estimate the variance of random effect, the variance of random error and the spatial correlation. A simulation study has been used to test performances of this method.

As in above study, many authors have studied different covariance structures in spatial setup for continuous data. Cressie [2], Jones and Vecchia [3], Berger et al. [4], Gelfand et al. [5], Cressie and Johannesson [6] have studied spatial correlation for continuous data. Correlation structures suggested by Cressie are exponential covariance function: ($C(h)=\sigma^2\exp(-a^2\|h\|^2)$, where $h\in\Re^d, a\in\Re$); Gaussian covariance function: ($C(h)=\dfrac{\sigma^2(a^2\|h\|/2)^v 2K_v(a^2\|h\|)}{\Gamma(v)}$) where $h\in\Re^d, a, v\in\Re, v>0$, K_v ; modified Bessel function of the second kind of order v); and the reciprocal covariance function ($C(h)=\sigma^2\left\{1+(\|h\|^2/b^2)\right\}^{-\beta}, h\in\Re^d, b\in\Re, \beta>0$), where $\|h\|$ is the euclidean norm of distance h, and σ^2 is the variance of the random covariate. All these correlation structures decay as the distance between two objects increases. With the spectral approach, Jones and Vecchia [3] have used the above three correlation structures for modeling spatial continuous data by using ARMA models. Based on Bayesian approaches, Berger et al. [4] and Gelfand et al. [5] have exploited some other covariance functions for continuous scale data.

A few studies are available for discrete responses also. Clayton and Kaldor [7] have used some nonlinear mixed models (including gamma and log-normal) to estimate relative risk $\theta_i=\dfrac{y_i}{E_i}$, where E_i is the expectation of the number of observed counts y_i in the i^{th} cluster, related to a covariate $\mathbf{x_i}$. They have assumed that conditional on relative risk θ_i , the count response y_i follows a Poisson distribution with parameter $\theta_i E_i$. Then the conditional expectation of y_i/θ_i has been given by

$$\begin{aligned} E(y_i/\theta_i) &= \theta_i E_i = \exp(\gamma_i)E_i \\ &= \exp(\mathbf{x_i'\beta}+\gamma_i) \end{aligned}, \quad (4)$$

where γ_i is the random effect belongs to i^{th} cluster. In this case, they assumed a log-normal distribution for θ_i , and the dependence of clusters has been incorporated via the random effects of adjacent clusters. Dependency among clusters has been modelled by applying conditional auto-regressive (CAR) model having the following properties.

$$E(\gamma_i/\gamma_j, i\neq j)=x_i'\beta+\rho\sum_j w_{ij}(\gamma_j-\mathbf{x_j'\beta}), \quad (5)$$

and $V(\gamma_i/\gamma_j, i\neq j)=\sigma^2$, (6)

where,

$$w_{ij}=\begin{cases}1: if\ i\ and\ j\ are\ adjacent\ clusters\\ 0: otherwise.\end{cases}$$

Finally, they identified the covariance structure of γ as $\sum=\sigma^2(I-\rho w)^{-1}$, where ρ is correlation index between two adjacent clusters.

Wijekoon et al. [8], have modeled count responses from locations where single response has been recorded from each location, by using Poisson generalized linear mixed model. Similar to Mariathas and Sutradhar [7], they also have constituted families of neighboring locations in a pre-specified distance and similar decomposition has been considered for families. When p-dimensional covariates $\mathbf{x}_s=\left[x_{s1},....,x_{sp}\right]'$ is specific to s^{th} location, they have used the following conditional Poisson model to model count responses.

$$y_s/\boldsymbol{\gamma}_s=\exp(\mathbf{x}_s'\boldsymbol{\beta}+\mathbf{a}'\boldsymbol{\gamma}_s)+e \quad (7)$$

where $\boldsymbol{\beta}$ is the effect of fixed covariates on y_s and ***a*** is the vector of weights for vector of random effects, $\boldsymbol{\gamma}_s$. Term e is the zero mean error component of the model. It has been assumed that random effect vector $\boldsymbol{\gamma}_s$ follow n_s-dimensional normal distribution with 0 mean vector and covariance matrix Φ_s . Here Φ_s is defined as $\Phi_s=\sigma_\gamma^2 C_{n_s n_s}(\phi)=\sigma_\gamma^2\left[\phi 1_{n_s}1_{n_s}'+(1-\phi)I_{n_s}\right]$, where $\mathbf{1}_{n_s}$ and $\mathbf{I}_{n_s}$ are n_s dimensional unit vector and matrix respectively, based on the correlation structure for random effects as given in (2).

They have derived mean, variance, covariance and correlation structure under two situations. In the first situation, independent random effects with equal weights have been considered. Under this situation, expected value and variance of response are $E(y_s) = \mu_s = \exp\left(\mathbf{x}_s'\boldsymbol{\beta} + \frac{1}{2n_s}\sigma_\gamma^2\right)$, and $V(y_s) = \mu_s + \mu_s^2 \exp\left(\frac{\sigma_\gamma^2}{n_s} - 1\right)$, respectively. Covariance between two locations y_s and y_w is $Cov(y_s, y_w) = \mu_s \mu_w \left(\exp(r_{sw}\sigma_\gamma^2) - 1\right)$.

The case of equi-correlated random effects has been considered in the second situation.

While the expected value of y_s has been derived as $E(y_s) = \mu_s = \exp\left(\mathbf{x}_s'\boldsymbol{\beta} + \frac{\sigma_\gamma^2}{2}\{1 + \varphi(n_s - 1)\}\right)$, the variance of y_s is given by $V(y_s) = \mu_s + \mu_s^2\left[\exp\{\sigma_\gamma^2(1 + \varphi(n_s - 1))\} - 1\right]$. Under the set up of equi-correlated random effects, the covariance between two locations y_s and y_w is given in the form

$$Cov(Y_s, Y_w) = \mu_s \mu_w \left(\exp\left(\frac{\varphi\sigma_\gamma^2}{2}(n_s + n_w - 2)\right)\right)\left[\exp\begin{pmatrix}\left(\frac{\sigma_\gamma^2 n_{sw}}{\sqrt{n_s n_w}}\right) \\ \{1 + \varphi(n_s + n_w - n_{sw} - 1)\} + \frac{\sigma_\gamma^2 \overline{n}_{sw}}{\sqrt{n_s n_w}}\end{pmatrix} - 1\right]. \quad (8)$$

where σ_γ^2 is the variance of random effect and r_{sw} is defined as $r_{sw} = \frac{n_{sw}}{\sqrt{n_s n_w}}$. The term $\overline{n}_{sw}$ is number of pairs of correlated locations in uncommon locations. Exact joint generalized quasi-likelihood estimation method has been used for parameter estimation. Performances have been tested by mean of a simulation study. In their study, same correlation has been used for all pairs of locations in the formed families. Not only that, they have assumed equal weights for all locations within the families. In real situation, such cases can be rarely found. At the same time, when family size increases, these assumptions can be highly violated. Dependence or correlation between two locations changes with the distance or gap (lag) between the selected locations. Dependence among adjacent clusters may arise through factors or covariates common to respective clusters and correlations among model errors. They also should be taken into consideration, when dependence among clusters is incorporated.

The aim of this study was to find out a different approach to incorporate cluster dependence in modeling count responses from dependent clusters. In this study, the cluster dependence of adjacent clusters was incorporated at two stages through covariates of adjacent clusters by means of a hierarchical model for cluster specific random effects and correlation structure of model random errors. A correlation index that changes with the lag or distance among clusters, was used for random errors.

Hence both fixed effects and random effects were to be incorporated in the model, generalized linear Poisson mixed model was used to model the count responses from clusters. Performances of the proposed approach were evaluated based on a simulation study with a linear set up of clusters. Proposed approach was tested for three different designs with different covariates. Parameters associated with the model were estimated by using the marginal generalized quasi-likelihood (GQL) method. Estimates of new approach were compared with the estimates of Penalized Quasi likelihood (PQL) and Monte Carlo likelihood (ML) approaches.

Details about the mean, variance, and covariance structure are given in section 2, while details about simulation process, design, and covariates used for the simulation are discussed in section 2.3. Section 2.4 presents the estimation methods. The results and applications are discussed in sections 3.1 and 3.2 respectively, while the conclusions are given in section 4.

2. METHODOLOGY

Let y_i be a single count response at i^{th} cluster (i =1,2,..,k) and γ_i be the corresponding random effect which is assumed to be identically and normally distributed with mean μ_γ and a common variance σ_γ^2 ,($\gamma_i \sim N(\mu_\gamma, \sigma_\gamma^2)$. It is assumed that random effects γ_i's are correlated. Model for the count response of i^{th} cluster y_i, depending on the random effect γ_i, is given by

$$y_i / \gamma_i = \exp(\mathbf{x}_i'\beta + \gamma_i) + \varepsilon_i, \quad (9)$$

where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})'$ is the vector of p-covariates for the i^{th} cluster, ε_i is identically and independently distributed pure white noise. Here, a normal distribution with zero mean and a common variance σ_ε^2, (i.e., $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is assumed for ε_i. Cluster dependency is incorporated through a hierarchical linear model for random effect. Marginal properties of the model were derived under two situations (cases) and they are discussed below.

2.1 Case 1: with Non-zero Mean for Random Effect $E(\gamma) = \mu_\gamma \neq 0$

Assume that k number of clusters are in a line and they are denoted by 1, 2,...k. At the first stage of incorporation of cluster dependence, random effect of i^{th} cluster, γ_i, was modeled by using covariates of adjacent clusters $(i\text{-}1)^{th}$ and $(i\text{+}1)^{th}$ as follow.

Let $$\gamma_i = \boldsymbol{z}_i'\boldsymbol{\theta} + \xi_i, \quad (10)$$

where $z_i' = x_{i-1}' + x_{i+1}'$ and $corr(\xi_i, \xi_j) \neq 0, i \neq j$, $\xi_i \sim N(0, \sigma_\xi^2)$ with $E(\gamma_i) = \mathbf{z}_i'\boldsymbol{\theta} \neq 0$ and $V(\gamma_i) = V(\xi_i)$. Coefficient $\boldsymbol{\theta}$ is the effect of $\mathbf{z}_i'$ on γ_i.

Hence, $\xi_i \sim N(0, \sigma_\xi^2)$, moment generating function of ξ_i is ,

$$M_{\xi_i}(t)_{t=1} = E(\exp(t\xi_i))_{t=1} = \exp\left(\frac{1}{2}\sigma_\xi^2\right) \quad (11)$$

Then, for any integer n(>0),

$$M_{\xi_i}(nt)_{t=1} = E(\exp(nt\xi_i)_{t=1} = \exp\left(\frac{n^2}{2}\sigma_\xi^2\right). \quad (12)$$

Hence,

$$V\left[\exp(t\xi_i)\right]_{t=1} = Var(\exp(\xi_i)) = E\left[\exp(\xi_i)^2\right] - \left[E(\exp(\xi_i))\right]^2 = \exp(\sigma_\xi^2)\left[\exp(\sigma_\xi^2) - 1\right] \quad (13)$$

The expected value of response y_i can be derived as,

$$\begin{aligned} E(y_i) &= E\left(E(y_i/\gamma_i)\right) = E\left(E\left[\exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\theta} + \xi_i\right)\right]\right) \\ &= E\left(\exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\theta} + \frac{1}{2}\sigma_\xi^2\right)\right) \\ &= \exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\boldsymbol{\theta} + \frac{1}{2}\sigma_\xi^2\right). \end{aligned} \quad (14)$$

The expected value of squared response y_i^2 is given as

$$\begin{aligned} E(y_i^2) &= E\left(E(y_i^2/\gamma_i)\right) \\ &= E\left[E\left(\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \gamma_i) + \varepsilon_i\right)^2\right] \\ &= E\left[E\left(\exp(2\boldsymbol{x}_i'\boldsymbol{\beta} + 2\gamma_i) + 2\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \gamma_i)\varepsilon_i + \varepsilon_i^2\right)\right] \\ &= \exp\left(2\boldsymbol{x}_i'\boldsymbol{\beta} + 2\boldsymbol{z}_i'\boldsymbol{\theta} + 2\sigma_\xi^2\right) + \sigma_\varepsilon^2, \end{aligned} \quad (15)$$

and the variance of response y_i is given by

$$\begin{aligned} V(y_i) &= E(y_i^2) - E(y_i)^2 \\ &= \exp\left(2\boldsymbol{x}_i'\boldsymbol{\beta} + 2\boldsymbol{z}_i'\boldsymbol{\theta} + \sigma_\xi^2\right)\left[\exp(\sigma_\xi^2) - 1\right] + \sigma_\varepsilon^2. \end{aligned} \quad (16)$$

At the second stage, cluster dependence was incorporated through the following correlation structure for random errors ξ_i and ξ_j corresponding to responses y_i and y_j in a specified distance d called as lag,

$$Corr(\xi_i, \xi_j) = \begin{cases} 1 & : \ if\ |i-j| = 0 \\ \phi^{|i-j|} & : \ if\ 0 < |i-j| \leq d \\ 0 & : \ if\ |i-j| > d. \end{cases} \quad (17)$$

Now, define $w=\mathbf{a}'\xi$, where $\mathbf{a}'=(1,1)$, and $\boldsymbol{\xi}'=(\xi_i,\xi_j)$. This *w* has a normal distribution ($N(0,\sigma_w^2)$) because it is a linear function of ξ_i and ξ_j. Since $Var(\xi)=\sigma_\xi^2$ and the correlation matrix of $\boldsymbol{\xi}$ is $Corr(\boldsymbol{\xi})=\begin{pmatrix}1 & \varphi^{|i-j|}\\ \varphi^{|i-j|} & 1\end{pmatrix}$, the variance covariance matrix of $\boldsymbol{\xi}$ can be written as

$$Cov(\boldsymbol{\xi})=\Sigma_{\boldsymbol{\xi}}=\begin{pmatrix}\sigma_\xi^2 & \varphi^{|i-j|}\sigma_\xi^2\\ \varphi^{|i-j|}\sigma_\xi^2 & \sigma_\xi^2\end{pmatrix}. \quad (18)$$

Then, the variance of w is derived as

$$\sigma_w^2=Cov(z)=\boldsymbol{a}'\,Cov(\boldsymbol{\xi})\,\boldsymbol{a}=2(1+\phi^{|i-j|})\sigma_\xi^2. \quad (19)$$

Therefore, according to (19), the moment generating function of w is,

$$\begin{aligned}M_w(t)_{t=1}&=E(\exp(t'w)_{t=1}=E(\exp(\xi_i+\xi_j))\\ &=\exp\left(\frac{1}{2}\sigma_w^2\right)=\exp\left[(1+\varphi^{|i-j|})\sigma_\xi^2\right]\end{aligned} \quad (20)$$

Then, expectation of product of responses y_i and y_j is given by

$$\begin{aligned}E(y_iy_j)&=E\left[E((y_i/\gamma_i)(y_j/\gamma_j))\right]\\ &=E\left[E\left(\exp((\mathbf{x}_i'+\mathbf{x}_j')\boldsymbol{\beta}+(\mathbf{z}_i'+\mathbf{z}_i')\boldsymbol{\theta}+(\xi_i+\xi_j))\right)\right]\\ &=E\left[\exp\left((\mathbf{x}_i'+\mathbf{x}_j')\boldsymbol{\beta}+(\mathbf{z}_i'+\mathbf{z}_j')\boldsymbol{\theta}\right)E\left(\exp(\xi_i+\xi_j)\right)\right]\\ &=E\left[\exp\left((\mathbf{x}_i'+\mathbf{x}_j')\boldsymbol{\beta}+(\mathbf{z}_i'+\mathbf{z}_j')\boldsymbol{\theta}\right)\exp\left((1+\phi^{|i-j|})\sigma_\xi^2\right)\right]\\ &=\exp\left((\mathbf{x}_i'+\mathbf{x}_j')\boldsymbol{\beta}+(\mathbf{z}_i'+\mathbf{z}_j')\boldsymbol{\theta}+(1+\phi^{|i-j|})\sigma_\xi^2\right)\end{aligned} \quad (21)$$

Covariance between responses y_i and y_j can be formed as

$$\begin{aligned}Cov(y_iy_j)&=E(y_iy_j)-E(y_i)E(y_j)\\ &=\exp\left((\boldsymbol{x}_i'+\boldsymbol{x}_j')\boldsymbol{\beta}+(\boldsymbol{z}_i'+\boldsymbol{z}_j')\boldsymbol{\theta}+\sigma_\xi^2\right)\left[\exp(\phi^{|i-j|}\sigma_\xi^2)-1\right]\end{aligned} \quad (22)$$

Therefore, correlation between responses y_i and y_j can be obtained by

$$Corr(y_i,y_j)=\frac{Cov(y_i,y_j)}{\sqrt{V(y_i)V(y_j)}} \quad (23)$$

Expected value of y_i^4 is

$$\begin{aligned}E(y_i^4)&=E\left(E(y_i^4/\gamma_i)\right)\\ &=E\left[E\left(\exp(\boldsymbol{x}_i'\boldsymbol{\beta}+\gamma_i)+\varepsilon_i\right)^4\right]\\ &=E\left[E\left(\exp(4\boldsymbol{x}_i'\boldsymbol{\beta}+4\gamma_i)+6\exp(2\boldsymbol{x}_i'\boldsymbol{\beta}+2\gamma_i)\varepsilon_i^2+\varepsilon_i^4\right)\right]\\ &=\exp\left(4\boldsymbol{x}_i'\boldsymbol{\beta}+4\boldsymbol{z}_i'\boldsymbol{\theta}+8\sigma_\xi^2\right)+6\exp\left(2\boldsymbol{x}_i'\boldsymbol{\beta}+2\boldsymbol{z}_i'\boldsymbol{\theta}+2\sigma_\xi^2\right)\sigma_\varepsilon^2+\sigma_\varepsilon^4\end{aligned} \quad (24)$$

By equation (15), $E(y_i^2)=\exp\left(2\mathbf{x}_i'\boldsymbol{\beta}+2\mathbf{z}_i'\boldsymbol{\theta}+2\sigma_\xi^2\right)+\sigma_\varepsilon^2$.

Variance of y_i^2 is,

$$\begin{aligned}V(y_i^2)&=E(y_i^4)-E(y_i^2)^2\\ &=\exp\left(4\boldsymbol{x}_i'\boldsymbol{\beta}+4\boldsymbol{z}_i'\boldsymbol{\theta}+4\sigma_\xi^2\right)\left[\exp(4\sigma_\xi^2)-1\right]+4\exp\left(2\boldsymbol{x}_i'\boldsymbol{\beta}+2\boldsymbol{z}_i'\boldsymbol{\theta}+2\sigma_\xi^2\right)\sigma_\varepsilon^2+\sigma_\varepsilon^4.\end{aligned} \quad (25)$$

Expectation of product of squared responses can be found as

$$E(y_i^2y_j^2)=E\left[E((y_i^2/\gamma_i)(y_j^2/\gamma_j))\right]$$

$$= E\left[E\left((\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \gamma_i) + \varepsilon_i)^2 \left(\exp(\boldsymbol{x}_j'\boldsymbol{\beta} + \gamma_j) + \varepsilon_j\right)^2\right)\right]$$

$$= E\left[E\left(\begin{array}{l}\left[\exp(2\boldsymbol{x}_i'\boldsymbol{\beta} + 2\gamma_i) + 2\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \gamma_i)\varepsilon_i + \varepsilon_i^2\right] * \\ \qquad\qquad \left[\exp(2\boldsymbol{x}_j'\boldsymbol{\beta} + 2\gamma_j) + 2\exp(\boldsymbol{x}_j'\boldsymbol{\beta} + \gamma_j)\varepsilon_j + \varepsilon_j^2\right]\end{array}\right)\right]$$

$$= E\left[\exp\left(2(\boldsymbol{x}_i' + \boldsymbol{x}_j')\boldsymbol{\beta} + 2(\boldsymbol{z}_i' + \boldsymbol{z}_j')\boldsymbol{\theta} + 2(\xi_i + \xi_j)\right)\right] + E\left[2\boldsymbol{x}_i'\boldsymbol{\beta} + 2\boldsymbol{z}_i'\boldsymbol{\theta} + 2\xi_i\right]E\left(\varepsilon_j^2\right)$$
$$+ E\left[2\boldsymbol{x}_j'\boldsymbol{\beta} + 2\boldsymbol{z}_j'\boldsymbol{\theta} + 2\xi_j\right]E\left(\varepsilon_i^2\right) + E\left(\varepsilon_i^2\varepsilon_j^2\right)$$

$$= \exp\left(2(\boldsymbol{x}_i' + \boldsymbol{x}_j')\boldsymbol{\beta} + 2(\boldsymbol{z}_i' + \boldsymbol{z}_j')\boldsymbol{\theta}\right)\exp\left(4(1 + \varphi^{|i-j|})\sigma_\xi^2\right) + \exp\left[2\boldsymbol{x}_i'\boldsymbol{\beta} + 2\boldsymbol{z}_i'\boldsymbol{\theta} + 2\sigma_\xi^2\right]\sigma_\varepsilon^2$$
$$+ \exp\left[2\boldsymbol{x}_j'\boldsymbol{\beta} + 2\boldsymbol{z}_j'\boldsymbol{\theta} + 2\sigma_\xi^2\right]\sigma_\varepsilon^2 + \sigma_\varepsilon^4$$

$$= \exp\left(2(\boldsymbol{x}_i' + \boldsymbol{x}_j')\boldsymbol{\beta} + 2(\boldsymbol{z}_i' + \boldsymbol{z}_j')\boldsymbol{\theta} + 4(1 + \varphi^{|i-j|})\sigma_\xi^2\right) + \exp\left[2\boldsymbol{x}_i'\boldsymbol{\beta} + 2\boldsymbol{z}_i'\boldsymbol{\theta} + 2\sigma_\xi^2\right]\sigma_\varepsilon^2$$
$$+ \exp\left[2\boldsymbol{x}_j'\boldsymbol{\beta} + 2\boldsymbol{z}_j'\boldsymbol{\theta} + 2\sigma_\xi^2\right]\sigma_\varepsilon^2 + \sigma_\varepsilon^4 \qquad (26)$$

Therefore covariance of squared responses is

$$Cov(y_i^2 y_j^2) = E(y_i^2 y_j^2) - E(y_i^2)E(y_j^2)$$
$$= \exp\left(2(\boldsymbol{x}_i' + \boldsymbol{x}_j')\boldsymbol{\beta} + 2(\boldsymbol{z}_i' + \boldsymbol{z}_j')\boldsymbol{\theta} + 4\sigma_\xi^2\right)\left[\exp\left(4\varphi^{|i-j|}\sigma_\xi^2\right) - 1\right]. \qquad (27)$$

Hence, correlation of squared responses can be obtained by

$$Corr(y_i^2, y_j^2) = \frac{Cov(y_i^2, y_j^2)}{\sqrt{V(y_i^2)V(y_j^2)}}. \qquad (28)$$

In case ξ_i are uncorrelated, random effects γ 's also be independent. Under this situation, the same expected value and the variance can be obtained. But, covariance of first order responses and second order responses become zero, implying that responses are independent. *i.e.* $Cov(y_i y_j) = Cov(y_i^2 y_j^2) = 0$

2.2 Case 2: Random Effects with Zero Mean ($E(\gamma) = \mu_\gamma = 0$)

Since $E(\gamma_i) = \mu_\gamma = z_i' B = 0$, then equation (10) becomes as $\gamma_i = \xi_i$, and hence $\sigma_\gamma^2 = \sigma_\xi^2$.
Because, γ_i is normally distributed, $\gamma_i \sim \mathrm{N}(0, \sigma_\gamma^2)$, as in case 1, it can be shown that,

$$M_{\gamma_i}(t)_{t=1} = \exp\left(\frac{1}{2}\sigma_\gamma^2\right). \qquad (29)$$

Then, the expectation and variance of count response of the i^{th} cluster are given respectively by,

$$E(y_i) = E\left[E(y_i / \gamma_i)\right] = E\left[\exp(\mathbf{x}_i'\boldsymbol{\beta} + \frac{1}{2}\sigma_\gamma^2)\right] = \exp\left(\mathbf{x}_i'\boldsymbol{\beta} + \frac{1}{2}\sigma_\gamma^2\right) \qquad (30)$$

and

$$Var(y_i) = E(y_i^2) - E(y_i)^2$$
$$= \exp(2\boldsymbol{x}_i^{'}\boldsymbol{\beta} + \sigma_\gamma^2)\left[\exp(\sigma_\gamma^2) - 1\right] + \sigma_\gamma^2. \quad (31)$$

Under the following correlation structure for γ_i and γ_j,

$$Corr(\gamma_i, \gamma_j) = \begin{cases} 1 & : \; if\,|i-j| = 0 \\ \varphi^{|i-j|} & : \; if\; 0 < |i-j| \le d \\ 0 & : \; if\,|i-j| > d, \end{cases} \quad (32)$$

covariance between y_i and y_j can be obtained as,

$$Cov(y_i, y_j) = E(y_i y_j) - E(y_i)E(y_j) = \exp\left[(\boldsymbol{x}_i' + \boldsymbol{x}_j')\boldsymbol{\beta} + \sigma_\gamma^2\right]\left[\exp(\phi^{|i-j|}\sigma_\gamma^2) - 1\right]. \quad (33)$$

Further, the covariance between y_i^2 and y_j^2 is given by

$$Cov(y_i^2, y_j^2) = \exp\left[2(\boldsymbol{x}_i' + \boldsymbol{x}_j')\boldsymbol{\beta} + 4\sigma_\gamma^2\right]\left[\exp(4\phi^{|i-j|}\sigma_\gamma^2) - 1\right]. \quad (34)$$

2.3 Simulation Study

The performance of proposed approach was tested by using a simulation study, as done in the literature, based on a sequence of linearly connected 100 clusters as shown in Fig. 1.

Lag of one (that is $d^* = 1$) was considered in setting dependence among clusters. In this setup, the first cluster is dependent on the second cluster only, while the last (100th) cluster depends only on the 99th cluster. All other clusters are dependent on adjacent clusters in left and right sides. For example, the i^{th} cluster is dependent on the (i-1)th and (i+1)th clusters.

Simulation study was performed under three different designs explained below with both discrete and continuous covariates for clusters. Under the first design, only two covariates (X_1, X_2) were considered. Covariate X_1 was discrete-binary generated from Bernoulli distribution with probability 0.7, and X_2 was a continuous covariate generated from Uniform distribution within the interval (1, 3). Values 0.3 and 0.8 were used as true values of covariates effects (β_1, β_2) under this design. In the second design, only two covariates (X_1, X_2) were used. The first covariate takes value 1 for all the clusters, while the other was generated from the Uniform distribution as in the first design. In this design, as the true values of covariates effects (β_1, β_2), values 0.3 and 0.8 were used. Three covariates $(X_1, X_2, and\ X_3)$ were used under the third design. The first covariate, X_1 is having a value 1 for all clusters and the other two covariates (X_2 and X_3) were from Bernoulli distribution. Values 0.3, 0.5 and 0.8 were used as the true values of covariates effects (β_1, β_2, β_3) in the third design.

With a given variance σ_γ^2, random effect γ_i for the i^{th} cluster was generated as explained below. First, a random number was generated from the standard normal distribution (normal distribution with zero mean and variance one), say number is g_i. Then, the value obtained through the product of the number g_i and the square root of variance of random effect σ_γ^2, as $\gamma_i = g_i \times \sqrt{\sigma_\gamma^2}$, used as the value of γ_i. Similarly, random effects for other clusters also were generated.

Then, cluster dependence was included through the random effect component of the mixed model. The random effect for the i^{th} cluster (i=2, 3,...., 99) after incorporating cluster dependence (say γ_i^*) was obtained by

$$\gamma_i^* = (\phi\gamma_{i-1} + \gamma_i + \phi\gamma_{i+1})(1/sqrt(1+2\phi)) \quad . \quad (35)$$

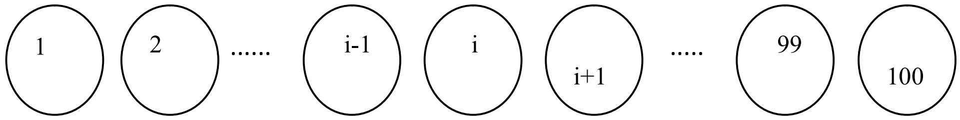


Fig. 1. A sequence of linearly located 100 clusters

Random effects of adjacent clusters have been weighted in term of clusters correlation φ. But, for the first and last clusters, random effects with cluster dependence were obtained by using $\gamma_1^* = (\gamma_1 + \varphi\gamma_2)(1/sqrt(1+\varphi))$ and $\gamma_{100}^* = (\gamma_{100} + \varphi\gamma_{99})(1/sqrt(1+\varphi))$ respectively.

Count responses were generated based on the assumption that conditional on random effects, count responses follow Poisson probability distribution, that is $y_i / \gamma_i \sim Poisson(\mu_i)$, where $\mu_i = \exp(x_i'\beta + \gamma_i^*)$.

Then, z_i variables were formed by using $z_i' = x_{i-1}' + x_{i+1}'$. Cluster dependence incorporated random effects were modeled by using the z_i variables as $\gamma_i^* = \mathbf{z_i'\theta} + \xi_i$, where θ is the effect of z_i on γ_i^*, which was estimated by the least squared method. Assume $\hat{\theta}$ is the estimate of θ. Here, the variance of ξ_i, σ_ξ^2 was estimated by using the estimated random errors obtained through the equation $\hat{\xi}_i = \gamma_i^* - z_i'\hat{\theta}$. Under the second case of the first approach, there was not such a need for usage of z variable.

Simulation was performed for several values (0.1, 0.2, 0.3, 0.4,) of correlation index φ, and four values (0.3, 0.5, 0.8, and 1.2) of the variance of random effects σ_γ^2 with 500 iterations. Generalized quasi likelihood estimates were obtained with the incorporation of cluster dependence. Further, for the purpose of comparison, Penalized Quassi likelihood and Monte Carlo likelihood estimates also were obtained. The performance of estimates was compared by using the simulated mean (SM: average of converged values of estimates of parameters) and the simulated standard error (SSE). R software was used for simulation process. Penalized Quassi likelihood estimates were obtained by using glmmPQL function in glmmPQL package and Monte Carlo likelihood estimates were obtained by using glmm function in the glmm package.

2.4 Estimation of Parameters

2.4.1 Generalized quasi-likelihood

Marginal generalized quasi-likelihood (GQL) estimating equations of Sutradhar [9] were used to estimate covariates effects and variance of random effects, based on Gauss-Newton iterative procedure. In estimating those parameters, it is assumed that value of φ and σ_ξ^2 are known. The first order responses are used to construct the marginal GQL estimation equation to estimate β . Then the GQL estimation equation for β is

$$\frac{\partial \mu'}{\partial \beta} \Sigma^{-1}(y-\mu) = 0, \tag{36}$$

and the solution of this equation could be obtained by using the Gauss-Newton iterative equation

$$\hat{\beta}_{GQL}(r+1) = \hat{\beta}_{GQL}(r) + \left[\frac{\partial \mu'}{\partial \beta}\Sigma^{-1}\frac{\partial \mu}{\partial \beta'}\right]_r^{-1}\left[\frac{\partial \mu'}{\partial \beta}\Sigma^{-1}(y-\mu)\right]_r, \tag{37}$$

where $[\;]_r$ denotes that the expression within the square bracket is evaluated at $\hat{\beta} = \hat{\beta}_{GQL}(r)$, the estimate obtained for the r[th] iteration, μ and Σ are given by equation (14), and equation (22), respectively.

The second order response based GQL estimation was used to obtain the marginal estimation of σ_ξ^2 as follows. Let $\mathbf{U} = (y_1^2, ..., y_s^2, ..., y_k^2)$, be the vector of second-order responses of all k clusters. Also, let $\boldsymbol{\lambda} = E(\mathbf{U}) = (\lambda_{11}, ..., \lambda_{ss}, ..., \lambda_{kk})'$, where

$\lambda_{ss} = E(y_s^2)$ for all s=1,...,k , as given by equation (15). Now define covariance matrix of U, $Cov(\mathbf{U}) = \mathbf{\Omega}$, and elements of this covariance matrix are giving by equation (27) and estimated σ_ξ^2 . Then GQL estimating equation for σ_ξ^2 is

$$\frac{\partial \lambda'}{\partial \sigma_\gamma^2} \boldsymbol{\Omega}^{-1}(\boldsymbol{U} - \boldsymbol{\lambda}) = 0. \quad (38)$$

The solution of this was also obtained by using the iterative equation

$$\hat{\sigma}^2_{\gamma\ GQL}(r+1) = \hat{\sigma}^2_{\gamma\ GQL}(r) + \left[\frac{\partial \lambda'}{\partial \sigma_\gamma^2} \boldsymbol{\Omega}^{-1} \frac{\partial \lambda}{\partial \sigma_\gamma^2}\right]_r^{-1} \left[\frac{\partial \lambda'}{\partial \sigma_\gamma^2} \boldsymbol{\Omega}^{-1}(\boldsymbol{U} - \boldsymbol{\lambda})\right]_r \quad (39)$$

where $[\ ]_r$ denotes that the expression within the square brackets, and is evaluated at $\hat{\sigma}_\gamma^2 = \hat{\sigma}^2_{\gamma\ GQL}(r)$ for the r[th] iteration.

2.4.2 <u>Penalized Quasi-likelihood</u>

The PQL approach assumes that the random effects γ_i are fixed effects parameters and estimate them along with β , before estimating σ_γ^2 by assuming that σ_γ^2 is known. For β and γ_i respectively, the estimating equations are obtained by maximizing the penalized quasi-likelihood function

$$h(\gamma_i) = -\sum_{j=1}^{n_i} y_{ij}(x_{ij}^{/}\beta + \gamma_i) + \sum_{j=1}^{n_i} \exp(x_{ij}^{/}\beta + \gamma_i) + \frac{\gamma_i^2}{2\sigma_\gamma^2} \quad (40)$$

with respect to β and γ_i , and the corresponding estimating equations are given by

$$g_1^*(\beta, \gamma_i) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} [y_{ij} - \exp(x_{ij}^{/}\beta + \gamma_i)] x_{ij} = 0 \quad (41)$$

and

$$g_2^*(\beta, \gamma_i, \sigma_\gamma^2) = \sum_{j=1}^{n_i} [y_{ij} - \exp(x_{ij}^{/}\beta + \gamma_i)] - \frac{\gamma_i}{\sigma_\gamma^2} = 0 \quad . (42)$$

To obtain an estimate (restricted maximum likelihood estimate) for the variance parameter, σ_γ^2 , profile quasi-likelihood based score equation is obtained by differentiating the profile quasi-likelihood function with respect to σ_γ^2 as,

$$g_3^*(\hat{\beta}_{PQL}, \sigma_\gamma^2, \hat{\gamma}_{PQL}) = \frac{\partial ql(\hat{\beta}_{PQL,}\ \sigma_\gamma^2, \hat{\gamma}_{PQL})}{\partial \sigma_\gamma^2} \quad (43)$$

Where $\hat{\beta}_{PQL}$ and $\hat{\gamma}_{i\ PQL}$ (i = 1, . . . ,K) are the Penalized quasi-likelihood estimators of β and $\tilde{\gamma}_i$, Then the following equation can be derived and estimate of σ_γ^2 denoted by $\hat{\sigma}^2_{r,\ PQL}$ can be obtained by solving

$$\sum_{i=1}^{K} \hat{\gamma}_{i,PQL}^2 - \sigma_\gamma^4 \sum_{i=1}^{K} \frac{\sum_{j=1}^{n_i} \exp(x_{ij}^{/} \hat{\beta}_{PQL} + \hat{\gamma}_{i,PQL})}{1 + \sigma_\gamma^2 \sum_{j=1}^{n_i} \exp(x_{ij}^{/} \hat{\beta}_{PQL} + \hat{\gamma}_{i,PQL})} = 0 \quad (44)$$

2.4.3 <u>Montecarlo likelihood</u>

In this approach, estimates are obtained by approximating the following likelihood,

$$L(\beta, \sigma_\gamma^2) = \int \prod_{i=1}^{K} \prod_{j=1}^{n_i} f(y_{ij} / \gamma_i, \beta) \left(2\prod \sigma_\gamma^2\right)^{-1/2} \exp(-\gamma_i^2 / 2\sigma_\gamma^2) d\gamma_i, \quad (45)$$

Where

$$f(y_{ij} / \gamma_i, \beta) = \exp\left[\begin{bmatrix} y_{ij}(x_{ij}'\beta + \gamma_i) \\ -\exp(x_{ij}'\beta + \gamma_i) \end{bmatrix} + b(y_{ij})\right]. \quad (46)$$

Estimates for β and σ_γ^2 are iteratively obtained based on McCulloch's EM approach [10] by using

$$\frac{1}{N} \sum_{w=1}^{N} \log\left(f(y / \gamma_i^{(w)}, \beta)\right), \quad (47)$$

And $\frac{1}{N} \sum_{w=1}^{N} \log\left(g(\gamma_i^{(w)} / \sigma_\gamma^2)\right)$ respectively, with random effects generated from metropolis algorithm [11]. In this study, n_i was 1 for all $i = 1, 2, .., K$ and $N = K$.

2.4.3.1 McCulloch's EM algorithm

This algorithm is as follows.

1. Choose starting values for $\beta^{(0)}$ and $\sigma_\gamma^{2(0)}$, $r=0$ where r denotes iteration.
2. Generate N number of values of the random effect by using $f(\gamma_i^- \mid y, \beta^{(r)}, \sigma_\gamma^{2(r)})$ and

(a) Choose $\beta^{(r+1)}$ to maximize a Monte Carlo estimate of $E\left[\log f(y \mid \gamma_i, \beta)\right]$

That is, maximize $\left[\frac{1}{N}\right]\sum_{w=1}^{N}\log f(y \mid \gamma_i^{(w)}, \beta).$

(b) Choose $\sigma_\gamma^{2(r+1)}$ to maximize $\left[\frac{1}{N}\right]\sum_{w=1}^{N}\log g(\gamma_i^{(w)} \mid \sigma_\gamma^2),$

where
$g(\gamma_i^{(w)} \mid \sigma_\gamma^2) = (2\pi\sigma_\gamma^2)^{-1/2}\exp\{-\gamma_i^2 / 2\sigma_\gamma^2\}$.

(c) Set $r=r+1$.

3. If convergence is achieved, then declare $\beta^{(r+1)}$ and $\sigma_\gamma^{2(r+1)}$ to be the maximum likelihood estimates; otherwise go back to step 2.

3. RESULTS AND DISCUSSION

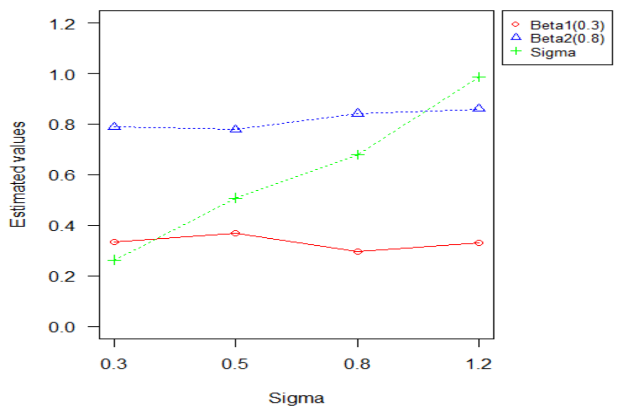
3.1 Results of Simulation

The proposed approach gave accurate estimates for all parameters β_1, β_2, β_3 and σ_γ^2 under the low values of correlation φ, up to 0.4. The simulated results with 500 iterations for $\varphi = 0.1$, $\varphi = 0.2$, $\varphi = 0.3$, and $\varphi = 0.4$ under all designs are given in Table 1, Table 2, Table 3, and Table 4 (in appendix) respectively. Estimates obtained with the incorporation of cluster dependence at both stages, are given in the column named as CDBS and estimates obtained by incorporating cluster dependence at the first stage only are given in the column CDFS. Estimates obtained without incorporation of cluster dependence by using Monte Carlo likelihood and Penalized Quasi likelihood methods are given in MCL and PQL columns respectively.

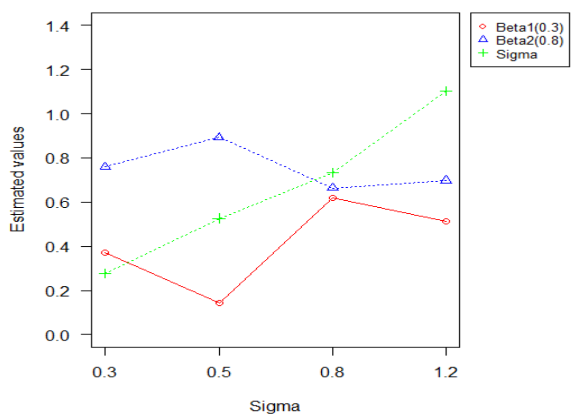
Proposed approach has given better estimates for both regression effects and variance of random effects. With compared to estimates obtained by ignoring cluster dependence (estimates in columns MCL and PQL), most of estimates obtained with the incorporation of cluster dependence (estimates in columns CDBS, CDFS) were closer to the actual values of the parameters. For example, consider estimates of regression effects (β_1, β_2) in Table 1 (in appendix) under the first design. PQL estimates for ($\beta_1 = 0.3, \beta_2 = 0.8$) were 0.3478 and 0.7703 respectively, while the corresponding MCL estimates were 0.3717 and 0.7745. Estimates with incorporation of cluster dependence at the first stage were 0.3661 and 0.7576, while the estimates obtained with incorporation of cluster dependence at both stages, were 0.3343 and 0.7829 respectively. In the same case, the variance of random effects $\sigma_\gamma^2 = 0.3$ was estimated as 1.1169 and 0.0013 by PQL and MCL respectively. GQL estimates with the incorporation of cluster dependence at the first stage and both stages were 0.2630 and 0.2607 respectively. In most of cases, it could be seen that GQL estimates obtained with the incorporation of cluster dependence at the both stages were closer than the GQL estimates obtained by incorporating dependence at the first stage only. But differences between corresponding values in columns CDBS and CDFS were small. That may be due to the low lag length (d=1) which was used in this study. When value of σ_γ^2, design and covariates changes, any pattern of estimates could not be observed. Further, it can be seen that standard error of the estimates were very low.

Fig. 2 shows how estimates were closer to their actual values and hence, the accuracy of the estimates.

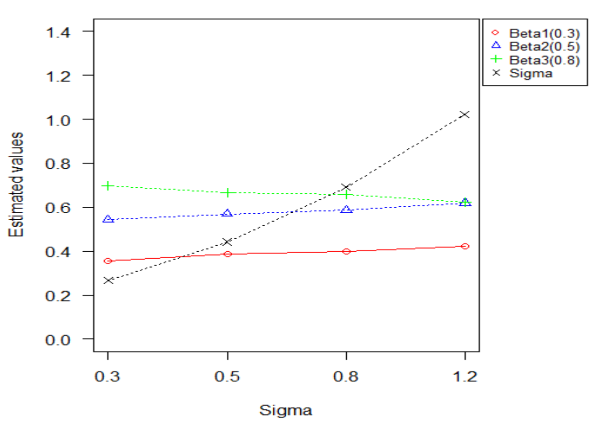
A plot of generated responses and estimated responses by using the fitted model with estimates obtained using proposed approach, are given in Fig. 3. Lines of both generated and estimated responses take the same pattern with low deviations in values. These lines also confirm suitability of estimates obtained from the proposed approach.



(a) Design 1

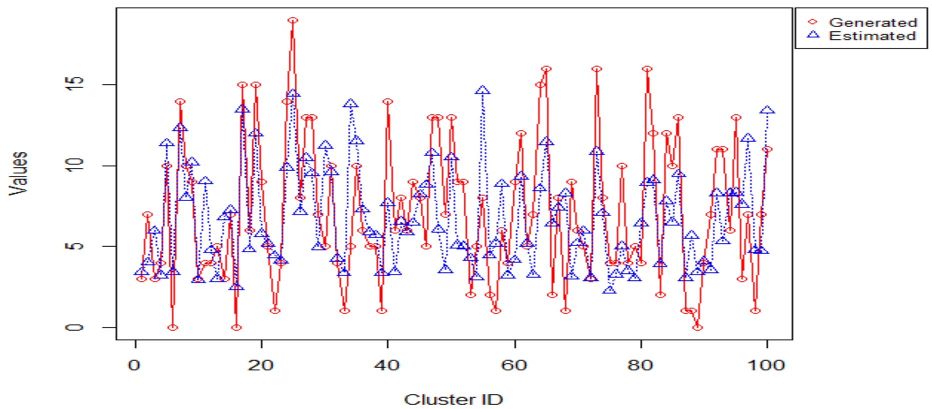


(b) Design-2

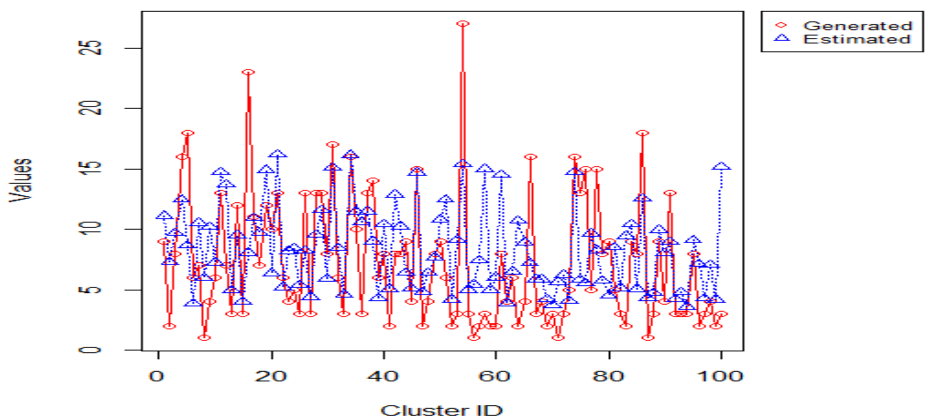


(c) Design 3

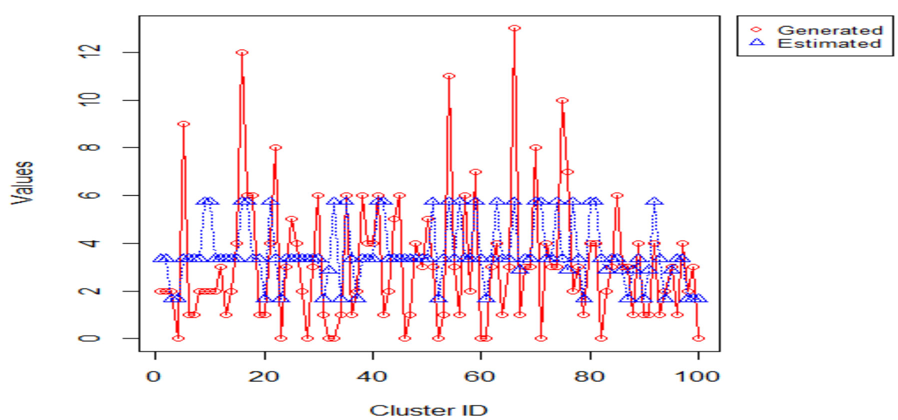
Fig. 2. Plot of estimated values for design-1 with $\phi = 0.1$



(a) Design 1



(b) Design 2



(c) Design 3

Fig. 3. Plot of generated and estimated responses for design-1 with $\phi = 0.1$

3.2 An Application

Performance of proposed approach was tested by using Scottish lip cancer data set (Table 5 in appendix 2) which has been used for several studies. Data have been collected over a period of five years by covering 56 districts (counties) in Scotland. Data set contains observed and expected counts of lip cancer, the latitude and longitude of each county. The expected counts have been found based on population, its age and sex distribution in each county. Further the percentage of population engaged in agriculture, fishing or forestry (PAFF) is available as a covariate. Adjacent districts for each county also have been given.

In this study, while PAFF was used as the covariate, available observed counts were used as the responses to obtain expected counts under the proposed approach. The covariance of responses was found based on the given adjacent districts. It was assumed that all adjacent districts have the same dependence with the respective district. Under the approach, estimates of β and σ_γ^2 were obtained for different values of φ and σ_γ^2. Estimated values of β, σ_γ^2 and φ when the estimate of σ_γ^2 was closer to its' true value were taken as the most suitable estimates of β, σ_γ^2 and φ. All most all the procedures were same as explained in the simulation study. Random effects were estimated by applying PQL method and used those estimated random effects as the random effect of each cluster.

Estimates obtained for β, σ_γ^2 and φ for this data set under proposed approach are given in Table 6 below. This says that correlation among adjacent districts is about 0.2.

Table 6. Estimates of parameters under each approach

Estimates of parameters		
$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\phi}$
0.10	1.49	0.20

Based on the estimates of parameters β and σ^2 in Table 6, counts were estimated. Accuracy of estimated counts were found by using mean error (ME), mean absolute deviation (MAD), and mean absolute percentage error (MAPE) and they are given in Table 7. ME, MAD, and MAPE obtained by using observed and expected values in the data set are listed under APP-0, while the results for proposed approach are listed under APP-1. This approach has given closer values for ME, MAD and MAPE.

Table 7. Accuracy of estimated counts under proposed approach

Approaches	Accuracy measures		
	ME	MAD	MAPE
App-0	0.0035	6.38	84.33
App-1	4.40	7.45	67.39

4. CONCLUSION

Results obtained for the simulated data and real data, confirm that proposed approach give better estimates for the parameters in the model with low standard deviation for moderate cluster dependence. Therefore, the proposed approach can be used to incorporate cluster dependencies in modeling count data from dependent clusters by using Poisson mixed models when there is weak or moderate dependence among adjacent clusters.

Under the simulation study, this approach was evaluated for a linear set up of clusters. As a lag length, one (d=1) was used. , In that, dependence of two nearest clusters in left and right side to the given cluster was considered. By using a larger value for d, accuracy of estimates could be improved. In the application of proposed approach for the cancer data, dependence of more than two clusters (counties) was considered because clusters (counties) in the cancer data sets are not having a linear setup and each cluster is dependent on more than two clusters. This confirms that the proposed approach is applicable even for non linear set up of clusters. According to Sutradhar [9], GQL estimation approach always gives consistent estimates. Therefore, in this study, GQL estimation method was used for parameter estimation.

Method suggested by Wijekoon et al. [8], requires formation of spatial families of clusters (locations), decomposition of formed families, identification of common and uncommon locations to families, and identification of correlated random effects. Hence, it is a long procedure. With compared their method, proposed approach by this study is simpler and easy to perform. At the same time, the proposed approach gives lower standard errors for estimates with compared to method suggested by Wijekoon et al. [8].

This study can be extended for nominal or ordinal scale responses as further research. At the same time, case that random effects are having different variances also can be taken for further studies.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Mariathas HH, Sutradhar BC. Variable family size based spatial moving correlations model. Sankhya B: Indian J. Stat. 2016;1-38.
2. Cressie N. Statistics for Spatial Data. Wiley, New York; 1991.
3. Jones RH, Vecchia AV. Fitting continuous ARMA models to unequally spaced spatial data. J Am Stat Ass. 1993;88:947–954.
4. Berger JO, Oliveria VD, Sans OB. Objective Bayesian analysis of spatially correlated data. J Am Stat Ass. 2001;96: 1361-1374.
5. Gelfand AE, Kim HJ, Sirmans CF, Banerjee SS. Spatial modeling with spatially varying coefficient processes. J Am Stat Ass. 2003;98:387–396.
6. Cressie N, Johannesson G. Fixed rank Kriging for very large spatial data sets. J R Stat Soc Series B Stat Methodol, B. 2008;70:209–226.
7. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics. 1987;43(3):671-681.
8. Wijekoon P, Oyet A, Sutradhar BC. Pair-wise family based correlation model for spatial count data. Sankya B. Indian J. Stat. 2018;1-52.
9. Sutradhar BC. On exact quasi likelihood inference in generalized linear mixed models. Sankhya B. Indian J. Stat. 2004; 66:261-289.
10. McCulloch CE. Maximum likelihood algorithms for generalized linear mixed models. J. Amer. Statist. Assoc. 1997;92: 162 -170.
11. Tanner MA. Tools for Statistical Inference: Observed Data and Data Augmentation. Berlin: Springer-Verlag; 1993.

APPENDIX 1. SIMULATED RESULTS

Under each design of covariates, 0.0 was used as the initial value of all parameters: β_1; β_2 ; and β_3, while 0.1 was used as the initial value of σ_γ^2 .

Table 1. Simulated results for $\varphi = 0.1$

Design	σ^2	QTY	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\beta}_3$				$\hat{\sigma}^2$			
			CDBS	CDFS	MCL	PQL	CDBS	CDFS	MCL	PQL	CDBS	CDFS	MCL	PQL	CDBS	CDFS	MCL	PQL
First	0.3	SM	0.3343	0.3661	0.3717	0.3478	0.7879	0.7576	0.7745	0.7703	_	_	_	_	0.2630	0.2607	0.0013	1.1169
		SSE	0.0038	0.0038	0.0038	0.0037	0.0015	0.0015	0.0025	0.0015					0.0028	0.0027	0.0006	0.0165
(β_1, β_2)	0.5	SM	0.3700	0.3909	0.4123	0.3718	0.7782	0.7779	0.7633	0.7676	_	_	_	_	0.5079	0.5102	0.0113	1.1487
=		SSE	0.0037	0.0036	0.0041	0.0039	0.0014	0.0014	0.0035	0.0015					0.0037	0.0037	0.0021	0.028
$(0.3, 0.8)$	0.8	SM	0.2969	0.3046	0.3596	0.3280	0.8411	0.8446	0.9594	0.8582	_	_	_	_	0.6804	0.6795	0.0020	0.7602
		SSE	0.0032	0.0032	0.0092	0.0035	0.0015	0.0031	0.0039	0.002					0.0026	0.0026	0.0011	0.0397
	1.2	SM	0.3309	0.3413	0.5076	0.3845	0.8598	0.8564	1.0944	0.8789	_	_	_	_	0.9864	0.9835	0.0299	0.4019
		SSE	0.0030	0.0030	0.0075	0.0035	0.0016	0.0015	0.0057	0.0015					0.0032	0.0026	0.0031	0.0246
Second	0.3	SM	0.3705	0.3715	0.5249	0.4208	0.7590	0.7403	0.7514	0.7227	_	_	_	_	0.2778	0.276	2.08E-06	0.2836
		SSE	0.0068	0.0066	0.0162	0.0070	0.0029	0.0032	0.0076	0.0028					0.0016	0.0016	8.15E-07	0.024
(β_1, β_2)	0.5	SM	0.1442	0.1417	0.7583	0.3539	0.8924	0.8958	0.8272	0.8600	_	_	_	_	0.5265	0.5268	1.02E-05	0.0678
=		SSE	0.0047	0.0042	0.0080	0.0070	0.0029	0.0027	0.0042	0.0031					0.0016	0.0016	3.37E-07	0.0146
$(0.3, 0.8)$	0.8	SM	0.6187	0.6387	0.9260	1.0196	0.6640	0.6541	0.6060	0.5803	_	_	_	_	0.7337	0.7333	2.23E-05	2.5844
		SSE	0.0063	0.0063	0.0153	0.0112	0.0030	0.0029	0.0038	0.0031					0.0031	0.0031	8.84E-07	0.0601
	1.2	SM	0.5124	0.5179	0.9329	0.8003	0.6985	0.6962	0.6768	0.6468	_	_	_	_	1.1022	1.1002	4.37E-06	2.1060
		SSE	0.0058	0.0059	0.0154	0.0065	0.0025	0.0025	0.0078	0.0026					0.0024	0.0024	2.20E-06	0.0499
Third	0.3	SM	0.3560	0.3651	0.4175	0.3528	0.5436	0.5591	0.5735	0.5504	0.6986	0.6971	0.7093	0.6939	0.2666	0.2661	4.35E-06	0.5203
		SSE	0.0071	0.0069	0.0105	0.0073	0.0053	0.0052	0.0109	0.0052	0.0076	0.0076	0.0093	0..0071	0.0021	0.002	7.47E-07	0.0242
$(\beta_1, \beta_2, \beta_3)$	0.5	SM	0.3859	0.3888	0.5057	0.3521	0.5694	0.5748	0.6263	0.5732	0.6660	0.6635	0.6963	0.6716	0.4418	0.4415	7.30E-06	0.1732
=		SSE	0.0067	0.0071	0.0148	0.0082	0.0048	0.0047	0.0185	0.005	0.0068	0.0073	0.0149	0.0079	0.0024	0.0024	2.13E-06	0.0200
$(0.3, 0.5, 0.8)$	0.8	SM	0.3983	0.4046	0.5856	0.3484	0.5869	0.5862	0.6268	0.5881	0.6574	0.6536	0.6557	0.6359	0.6911	0.6863	7.12E-06	0.1279
		SSE	0.0066	0.0065	0.0084	0.0079	0.0048	0.0048	0.0146	0.0050	0.0071	0.0068	0.0100	0.0075	0.0030	0.0033	1.48E-06	0.0200
	1.2	SM	0.4206	0.4379	0.4238	0.3694	0.6191	0.6322	0.6480	0.6132	0.6243	0.6190	0.7740	0.6194	1.0222	1.0180	6.11E-06	0.0941
		SSE	0.0064	0.0061	0.0041	0.0080	0.0043	0.0044	0.0122	0.0051	0.0060	0.0061	0.1587	0.0070	0.0032	0.0035	1.88E-06	0.0192

Table 2. Simulated results for $\varphi = 0.2$

Design	σ^2	QTY	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\beta}_3$				$\hat{\sigma}^2$			
			CDBS	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**
First	0.3	SM	0.3490	0.3518	0.3582	0.3332	0.7625	0.7621	0.7772	0.7728	_	_	_	_	0.2698	0.2650	1.60E-03	1.1001
		SSE	0.0038	0.0038	0.0043	0.0038	0.0014	0.0014	0.0027	0.0014					0.0028	0.0028	8.00E-04	0.0162
(β_1, β_2)	0.5	SM	0.3895	0.3906	0.4138	0.3614	0.7586	0.7574	0.8056	0.7975	_	_	_	_	0.4680	0.4655	0.1033	0.9427
=		SSE	0.0036	0.0036	0.0060	0.0039	0.0015	0.0015	0.0046	0.0020					0.0046	0.0045	0.0020	0.0307
$(0.3, 0.8)$	0.8	SM	0.3913	0.3927	0.4359	0.3726	0.7853	0.7840	0.8880	0.8239	_	_	_	_	0.7971	0.8099	0.0103	0.9069
		SSE	0.0045	0.0046	0.0045	0.0050	0.0027	0.0025	0.0055	0.0018					0.0125	0.0137	0.0019	0.0426
	1.2	SM	0.4455	0.4382	0.4146	0.3121	0.7674	0.7708	0.7594	0.7805	_	_	_	_	1.1088	1.1090	0.0235	0.8215
		SSE	0.0073	0.0073	0.0071	0.0095	0.0037	0.0038	0.0074	0.0041					0.0050	0.0051	0.0025	0.0365
Second	0.3	SM	0.5022	0.5064	0.5156	0.4999	0.7016	0.6988	0.7089	0.6915	_	_	_	_	0.2311	0.2294	2.93E-06	0.6908
		SSE	0.0061	0.0195	0.0135	0.0085	0.0026	0.0031	0.0075	0.0032					0.0018	0.0019	2.14E-07	0.0297
(β_1, β_2)	0.5	SM	0.3127	0.3182	0.1912	0.3220	0.7916	0.7787	0.9455	0.7742	_	_	_	_	0.4304	0.4243	1.25E-05	0.2121
=		SSE	0.0068	0.0064	0.0080	0.0068	0.0029	0.0022	0.0041	0.0031					0.0022	0.0083	5.43E-07	0.0162
$(0.3, 0.8)$	0.8	SM	0.4413	0.4421	0.3288	0.5831	0.7220	0.7165	0.7148	0.6867	_	_	_	_	0.8043	0.8001	1.96E-06	1.1470
		SSE	0.0064	0.0068	0.0045	0.0077	0.0027	0.0027	0.0052	0.0030					0.0025	0.0036	5.46E-07	0.0495
	1.2	SM	0.3394	0.3249	0.5271	0.6830	0.8156	0.8305	0.8953	0.7309	_	_	_	_	0.9395	0.9372	7.92E-06	1.1019
		SSE	0.0075	0.0061	0.0253	0.0083	0.0045	0.0036	0.0118	0.0033					0.0086	0.0086	6.21E-07	0.0602
Third	0.3	SM	0.3269	0.3327	0.3838	0.3292	0.5508	0.5461	0.5777	0.5640	0.7257	0.7182	0.7068	0.7030	0.2893	0.2834	4.58E-06	0.5127
		SSE	0.0068	0.0065	0.0083	0.0073	0.0055	0.0049	0.0108	0.0050	0.0071	0.0071	0.0076	0.0072	0.0022	0.0021	7.04E-07	0.0238
$(\beta_1, \beta_2, \beta_3)$	0.5	SM	0.3784	0.3831	0.4510	0.3328	0.5621	0.5626	0.5907	0.5818	0.6715	0.6611	0.6902	0.6802	0.4593	0.4519	7.17E-06	0.2968
=		SSE	0.0069	0.0064	0.0098	0.0078	0.0049	0.0052	0.0096	0.0050	0.0068	0.0061	0.0109	0.0068	0.0030	0.0028	1.45E-06	0.0252
$(0.3, 0.5, 0.8)$	0.8	SM	0.3986	0.4090	0.5689	0.3414	0.5602	0.5620	0.5598	0.5757	0.6562	0.6533	0.6679	0.6517	0.7343	0.7261	4.18E-06	0.5210
		SSE	0.0063	0.0065	0.0118	0.0082	0.0051	0.0051	0.0151	0.0055	0.0063	0.0063	0.0132	0.0069	0.0035	0.0035	7.12E-07	0.3140
	1.2	SM	0.4155	0.4184	0.6990	0.3828	0.5623	0.5895	0.5920	0.5947	0.6400	0.6278	0.6300	0.6135	1.1241	1.1211	7.32E-06	0.2547
		SSE	0.0061	0.0060	0.0127	0.0084	0.0042	0.0041	0.0138	0.0049	0.0064	0.0062	0.0134	0.0071	0.0044	0.0044	1.07E-06	0.0281

Table 3. Simulated results for $\varphi = 0.3$

Design	σ^2	QTY	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\beta}_3$				$\hat{\sigma}^2$			
			CDBS	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**
First	0.3	SM	0.3480	0.3495	0.3605	0.3282	0.7621	0.7618	0.7808	0.7734	_	_	_	_	0.2902	0.2817	0.0009	1.1019
(β_1, β_2)		SSE	0.0038	0.0037	0.0081	0.0038	0.0014	0.0014	0.0035	0.0014					0.0032	0.0032	0.0006	0.0156
=	0.5	SM	0.3665	0.3691	0.4014	0.3471	0.7424	0.7434	0.7723	0.7695	_	_	_	_	0.5417	0.5495	0.0065	0.7695
(0.3,0.8)		SSE	0.0040	0.0038	0.0114	0.0039	0.0015	0.0015	0.0054	0.0015					0.0044	0.0045	0.0016	0.0015
	0.8	SM	0.3025	0.3025	0.2966	0.2019	0.8491	0.8549	0.9751	0.8381	_	_	_	_	0.5288	0.5484	0.0025	0.3159
		SSE	0.0031	0.0031	0.0031	0.0039	0.0018	0.0018	0.0029	0.0015					0.0054	0.0050	0.0007	0.0205
	1.2	SM	0.2478	0.2626	0.2443	0.2775	0.8429	0.8395	0.7969	0.8851	_	_	_	_	0.7817	0.7754	0.0053	1.5172
		SSE	0.0036	0.0035	0.0035	0.0038	0.0031	0.0015	0.0023	0.0019					0.0051	0.0050	0.0013	0.0370
Second	0.3	SM	0.3276	0.3299	0.4943	0.3766	0.7779	0.7720	0.8005	0.7441	_	_	_	_	0.2991	0.2965	1.73E-06	0.4185
(β_1, β_2)		SSE	0.0071	0.0068	0.0393	0.0079	0.0032	0.0031	0.0170	0.0032					0.0018	0.0019	5.20E-07	0.0277
=	0.5	SM	0.3665	0.3799	0.5165	0.4179	0.7566	0.7424	0.7655	0.7291	_	_	_	_	0.5112	0.5130	7.22E-06	0.6237
(0.3,0.8)		SSE	0.0062	0.0068	0.0177	0.0078	0.0029	0.0056	0.0082	0.0031					0.0023	0.0022	2.39E-06	0.0099
	0.8	SM	0.4022	0.4061	0.6778	0.5486	0.7329	0.7272	0.7699	0.7060	_	_	_	_	0.8675	0.8569	9.06E-07	1.3200
		SSE	0.0062	0.0063	0.0234	0.0075	0.0027	0.0035	0.0141	0.0028					0.0028	0.0028	1.31E-07	0.0485
	1.2	SM	0.4836	0.4861	0.9395	0.7652	0.7297	0.7463	0.6982	0.6543	_	_	_	_	0.9812	0.9795	2.42E-06	0.9281
		SSE	0.006	0.0058	0.0101	0.0067	0.0027	0.0026	0.0040	0.0030					0.0099	0.0098	4.19E-07	0.0641
Third	0.3	SM	0.4076	0.4025	0.8317	0.5174	0.5767	0.5840	0.5939	0.5206	0.6215	0.6322	0.4735	0.5353	0.2344	0.2216	3.43E-05	0.7937
$(\beta_1, \beta_2, \beta_3)$		SSE	0.0070	0.0075	0.0097	0.0071	0.0053	0.0050	0.0053	0.0052	0.0067	0.0073	0.0069	0.0069	0.0032	0.0034	1.32E-06	0.0246
=	0.5	SM	0.3636	0.3833	0.3996	0.5390	0.5475	0.5438	0.5518	0.5926	0.6912	0.6703	0.6497	0.6637	0.4962	0.7783	7.71E-01	7.8E-06
(0.3, 0.5, 0.8)		SSE	0.0066	0.0070	0.0064	0.0123	0.0051	0.0050	0.0045	0.0146	0.0067	0.0075	0.0067	0.0126	0.0029	0.0040	3.80E-03	1.1E-06
	0.8	SM	0.3833	0.3996	0.5390	0.3424	0.5438	0.5518	0.5926	0.5760	0.6703	0.6497	0.6637	0.6610	0.7783	0.7712	7.86E-06	0.3158
		SSE	0.0070	0.0064	0.0123	0.0082	0.0050	0.0045	0.0146	0.0049	0.0075	0.0067	0.0126	0.0071	0.0040	0.0038	1.14E-06	0.0302
	1.2	SM	0.3364	0.4097	0.6216	0.4259	0.5295	0.5874	0.5499	0.6165	0.6321	0.6164	0.6303	0.5904	1.2800	1.2757	7.36E-06	1.1689
		SSE	0.0058	0.0056	0.0094	0.0094	0.0047	0.0044	0.0045	0.0050	0.0060	0.0055	0.0089	0.0069	0.0053	0.0053	8.41E-07	0.0617

Table 4. Simulated results for $\varphi = 0.4$

Design	σ^2	QTY	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\beta}_3$				$\hat{\sigma}^2$			
			CDBS	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**	**CDBS**	**CDFS**	**MCL**	**PQL**
First	0.3	SM	0.3325	0.3332	0.3369	0.3107	0.7656	0.7643	0.7873	0.7765	_	_	_	_	0.2996	0.2871	0.0006	1.1127
		SSE	0.0038	0.0040	0.0043	0.0040	0.0014	0.0014	0.0050	0.0014					0.0028	0.0029	0.0003	0.0155
(β_1, β_2)	0.5	SM	0.3577	0.3500	0.3743	0.3339	0.7450	0.7415	0.7699	0.7756	_	_	_	_	0.5866	0.5699	0.0094	1.2659
$=$		SSE	0.0041	0.0036	0.0045	0.0041	0.0014	0.0015	0.0027	0.0016					0.0043	0.0043	0.0018	0.0272
$(0.3, 0.8)$	0.8	SM	0.5600	0.5919	0.7106	0.5592	0.7745	0.7633	0.9194	0.8628	_	_	_	_	0.6243	0.6243	0.0086	0.3413
		SSE	0.0055	0.0058	0.0079	0.0055	0.0019	0.0017	0.0052	0.0027					0.0094	0.0094	0.0022	0.0305
	1.2	SM	0.2688	0.2461	0.3159	0.3634	0.8708	0.8627	1.0565	0.5212	_	_	_	_	1.0170	0.9864	0.0399	0.7547
		SSE	0.0048	0.0065	0.0069	0.0051	0.0026	0.0031	0.0064	0.0015					0.0055	0.0060	0.0033	0.0334
Second	0.3	SM	0.3113	0.3413	0.3886	0.3355	0.7854	0.7814	0.7851	0.7612	_	_	_	_	0.3064	0.3051	6.81E-07	0.4145
		SSE	0.0067	0.0067	0.0127	0.0078	0.0030	0.0050	0.0069	0.0033					0.0019	0.0017	1.29E-07	0.0268
(β_1, β_2)	0.5	SM	0.3424	0.3432	0.4930	0.3799	0.7695	0.7677	0.7914	0.7460	_	_	_	_	0.5329	0.5374	3.06E-06	0.5294
$=$		SSE	0.0063	0.0066	0.0231	0.0074	0.0028	0.0029	0.0127	0.0030					0.0025	0.0025	1.43E-06	0.0354
$(0.3, 0.8)$	0.8	SM	0.3719	0.3835	0.5951	0.5104	0.7472	0.7425	0.7982	0.7250	_	_	_	_	0.8985	0.8993	1.79E-06	1.3593
		SSE	0.0059	0.0061	0.0209	0.0073	0.0026	0.0026	0.0152	0.0028					0.0032	0.0032	3.04E-07	0.0496
	1.2	SM	0.2563	0.2482	0.7212	0.9227	0.8236	0.8262	0.7850	0.6539	_	_	_	_	1.2665	1.2697	1.69E-05	1.8983
		SSE	0.0063	0.007	0.0088	0.0081	0.0033	0.0039	0.0028	0.0029					0.0036	0.0036	5.29E-07	0.0345
Third	0.3	SM	0.3117	0.3190	0.3939	0.3208	0.5572	0.5565	0.6173	0.5570	0.7386	0.7283	0.7715	0.7049	0.2997	0.2944	3.99E-06	0.5892
		SSE	0.0068	0.0066	0.0235	0.0074	0.0053	0.0052	0.0250	0.0053	0.0073	0.0071	0.0277	0.0072	0.0024	0.0025	7.11E-07	0.0252
$(\beta_1, \beta_2, \beta_3)$	0.5	SM	0.3579	0.3697	0.4463	0.3535	0.5453	0.5442	0.5553	0.5674	0.7060	0.6817	0.6794	0.6642	0.4898	0.4847	8.25E-06	0.3411
$=$		SSE	0.0067	0.0066	0.0096	0.0074	0.0052	0.0050	0.0052	0.0048	0.0076	0.0067	0.0507	0.0068	0.0030	0.0041	1.46E-06	0.0265
$(0.3, 0.5, 0.8)$	0.8	SM	0.3896	0.3241	0.9859	0.6758	0.5434	0.4633	0.5111	0.4941	0.5527	0.6752	0.5596	0.5874	1.0295	1.0152	1.80E-05	1.2735
		SSE	0.0070	0.0064	0.0091	0.0098	0.0049	0.0047	0.0063	0.0048	0.0070	0.0063	0.0073	0.0061	0.0047	0.0045	1.06E-06	0.0492
	1.2	SM	0.3417	0.3555	0.6043	0.4004	0.5176	0.5788	0.5436	0.5931	0.7071	0.6701	0.6832	0.6344	1.3202	1.314	9.55E-06	0.9256
		SSE	0.0063	0.0060	0.0233	0.0099	0.0049	0.0044	0.0145	0.0049	0.0064	0.0060	0.0154	0.0069	0.0047	0.0048	8.46E-07	0.0511

APPENDIX 2. DATA SET

Table 5. Scottish lip cancer data set

District	Observed(y)	Expected	%AFF(x)	Latitude	Longitude	Adjacent Districts
1	9	1.4	16	57.29	5.5	5,9,11,19
2	39	8.7	16	57.56	2.36	7,10
3	11	3	10	58.44	3.9	6,12
4	9	2.5	24	55.76	2.4	18,20,28
5	15	4.3	10	57.71	5.09	1,11,12,13,19
6	8	2.4	24	59.13	3.25	3,8
7	26	8.1	10	57.47	3.3	2,10,13,16,17
8	7	2.3	7	60.24	1.43	6
9	6	2	7	56.9	5.42	1,11,17,19,23,29
10	20	6.6	16	57.24	2.6	2,7,16,22
11	13	4.4	7	58.12	6.8	1,5,9,12
12	5	1.8	16	58.06	4.64	3,5,11
13	3	1.1	10	57.47	3.98	5,7,17,19
14	8	3.3	24	54.94	5	31,32,35
15	17	7.8	7	56.3	3.1	25,29,50
16	9	4.6	16	57	3	7,10,17,21,22,29
17	2	1.1	10	57.06	4.09	7,9,13,16,19,29
18	7	4.2	7	55.65	2.88	4,20,28,33,55,56
19	9	5.5	7	57.24	4.73	1,5,9,13,17
20	7	4.4	10	55.35	2.9	4,18,55
District	Observed(y)	Expected	%AFF(x)	Latitude	Longitude	Adjacent Districts
21	16	10.5	7	56.7	2.98	16,29,50
22	31	22.7	16	57.1	2.2	10,16
23	11	8.8	10	56.4	5.27	9,29,34,36,37,39
24	7	5.6	7	55.6	3.96	27,30,31,44,47,48, 55,56
25	19	15.5	1	56.2	3.3	15,26,29
26	15	12.5	1	56.1	3.6	25,29,42,43
27	7	6	7	55.2	4.09	24,31,32,55
28	10	9	7	55.9	2.8	4,18,33,45
29	16	14.4	10	56.6	4.09	9,15,16,17,21,23, 25,26,34,43,50
30	11	10.2	10	55.9	3.8	24,38,42,44,45,56
31	5	4.8	7	55.47	4.55	14,24,27,32,35,46,47
32	3	2.9	24	55	4.36	14,27,31,35
33	7	7	10	55.8	3.2	18, 28,45,56
34	8	8.5	7	56.3	4.73	23,29,39,40,42,43,51,52,54
35	11	12.3	7	55.2	4.98	14,31,32,37,46
36	9	10.1	0	55.9	4.95	23,37,39,41
37	11	12.7	10	55.7	5.02	23,35,36,41,46
38	8	9.4	1	55.9	4.18	30,42,44,49,51,54

District	Observed(y)	Expected	%AFF(x)	Latitude	Longitude	Adjacent Districts
39	6	7.2	16	56.1	4.99	23,34,36,40,41
40	4	5.3	0	56	4.91	34,39,41,49,52
41	10	18.8	1	55.8	4.82	36,37,39,40,46,49,53
42	8	15.8	16	56	4	26,30,34,38,43,51
43	2	4.3	16	56.1	3.96	26,29,34,42
44	6	14.6	0	55.8	4.09	24,30,38,48,49
45	19	50.7	1	55.9	3.4	28,30,33,56
46	3	8.2	7	55.6	4.75	31,35,37,41,47,53
47	2	5.6	1	55.7	4.45	24,31,46,48,49,53
48	3	9.3	1	55.7	4.27	24,44,47,49
49	28	88.7	0	55.9	4.55	38,40,41,44,47,48,52,53,54
50	6	19.6	1	56.4	3.2	15,21,29
51	1	3.4	1	56	4.27	34,38,42,54
52	1	3.6	0	56.1	4.64	34,40,49,54
53	1	5.7	1	55.7	4.7	41,46,47,49
54	1	7	1	55.9	4.45	34,38,49,51,52
55	0	4.2	16	55.6	3.38	18,20,24,27,56
56	0	1.8	10	55.1	3.4	18,24,30,33,45,55

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/25051