Statistica & Applicazioni - Vol. XVIII, n. 2, 2020, pp. 137-160

ROBUST ESTIMATION IN JOINT MODELLING FOR HUMAN INTELLIGENCE

Thanigaivasan Gokul*
Mamandur Rangaswamy Srinivasan**
Michele Gallo***

SUMMARY

Joint models under generalized linear mixed model framework have received lot of attention among researchers in the field of psychology to analyse data with more than one response variable. The presence of aberrant observations in the data may influence the estimation of parameters in the existing method of estimation such as maximum likelihood, quasi-likelihood, etc. Hence, there exists a need for robust method of estimation under joint modelling to reduce the effect of influential data points. In this paper, two methods of robust estimation namely robust Maximum Likelihood method and robust Monte Carlo Newton-Raphson for joint longitudinal model has been compared with the usual maximum likelihood method to examine the association between the outcome variables of Spearman's G and S factors of human intelligence along with other covariates based on school lunch intervention data. In addition, a parametric bootstrap study is adopted to find the sensitivity and efficiency of the robust method in resampling techniques with varying sample sizes.

Keywords: Generalized Linear Mixed Model, Joint Model, Influential Observations, Robust Estimation.
DOI: 10.26350/999999_000035
ISSN: 18246672 (print) 2283-6659 (digital)

1. INTRODUCTION

An ongoing research in the field of cognitive psychology has sparked several statistical approaches in reaping the benefits of both psychological and statistical disciplines. In particular, the evolution of methods for measuring the human intelligence such as Spearman's G and S factors has provided a new scope in studying the mental abilities of school children. Until now, research on cognitive behaviour is restricted to modelling the data with a single response variable. However, joint modelling provides ample scope for the scientists who are interested in studying the behaviour of association with more than one response outcomes.

* Department of Statistics - University of Madras - CHENNAIM - India
(e-mai: gokultvasan@gmail.com).
** School of Mathematics and Statistics - University of Hyderabad - TELANGANA - India
(e-mail: ✉ mrsrin8@gmail.com).
*** Dipartimento di Scienze Umane e Sociali - Università di Napoli 'L'Orientale' - NAPOLI - Italia (e-mail: mgallo@unior.it).

Research in the earlier studies has shown that significant association exist between malnutrition and cognitive behavioural development of children. In this context, nutrition supplement plays a vital role in improving the child's intelligence and hence school lunch programs are introduced in order to improve the health of the school children (Pope, Roche, Morgan and Kolodinsky, 2018). Children suffering from mild-to-moderate malnutrition, show significant deficits in intellectual and behavioural functioning such as perceptual-spatial functioning, poorer school grades, reduced attentiveness and unresponsive play behaviour, as compared to their adequately nourished peers. Generally, cognitive measures are categorized into two factors namely; General (G) and Specific (S). Goharpey, Crewther D.P. and Crewther S.G. (2013) suggested that the Raven's Coloured Progressive Matrices (RCPM) tests are the valid measures of G factor for assessing the intelligence as a problem solving ability in children. The S factor includes Verbal Meaning (VM) test, similar to the Peabody Picture Vocabulary Test wherein the child has to select the picture matching a verbal label from a set of four pictures. Arithmetic Score (AS) deals with logical reasoning and some math work and Digit Span (DS) Test requires the child to repeat a series of digits after the experimenter has said them (Coyle, Elpers, Gonzalez, Freeman and Baggio, 2018; Muniz, Gomes and Pasian, 2016). The nature of the data is count or ordinal in most of the psychological studies. Thus, Poisson predictive model by Bilker, Hansen, Brensinger, Richard, Raquel and Gur (2012) shows that there exists a better correlation in the reduced model as compared to a 30-item model. Gra§hoff, Holling and Schwabe (2016) developed optimal design for the Rasch Poisson-Gamma model and the random effects parameter follows an underlying Gamma distribution.

Joint modelling under generalized linear mixed model (GLMM) framework has a wider scope in studying the association among the outcome variables (Gueorguieva, 2001; Iddi and Molenberghs, 2012). Hickey, Philipson, Jorgensen and Kolamunnage-Dona (2018) compared different hazard functions, latent association structure between the sub models, estimation approach and software implementation using joint modelling for longitudinal data. Loeys, Rosseel and Baten (2011) proposed a joint model for reaction time and accuracy in Psycholinguistic experiments. Palestro, Bahg, Sederberg, Lu, Steyvers and Turner (2018) studied Bayesian joint model for neural and behavioural measures of cognition.

The presence of outliers in the data may cause a serious problem in the estimation process. Hence, there is a need for a robust method in estimating the parameters. Literature is abundant on robust likelihood estimation methods (Qin, Bai and Zhu 2012; Sinha, 2004). McCulloch (1997) developed a Monte Carlo Newton-Raphson (MCNR) algorithm for approximating the ML estimates in GLMM's approach or improved the use of simulated ML methods. The MCNR estimates were compared to the exact ML likelihood estimates for simple models, and it was found that MCNR inherits the properties of the exact ML estimates. Yau and Kuk (2002) proposed robust methods for maximum quasi-likelihood and residual maximum quasi-likelihood estimation to limit the influence of outlying observations in GLMMs. The use of a re-descending ψ function effectively deletes large residuals rather than down-weighting them. In addition, Sinha (2004) proposed a modification of McCulloch (1997) as robust MCNR (RMCNR) algorithm for fitting Poisson mixed models. The method improved the use of importance sampling as well as the deterministic robust ML (RML) method in down-weighting the influential data points when estimating the parameters. Sutradhar and Bari (2011) proposed a robust generalized quasi likelihood estimation (RGQL) approach in order to obtain the consistent estimates for the parameters in the GLMMs with one or more outliers. Markatou, Afendras and Agostinelli (2018) discussed the robust cross validation technique to find the unusual observations in the data based on the weighted maximum likelihood methodology and compared their approach with the classical cross validation techniques for the selection of linear models. Lin, Fu, Qin and Zhu (2017) proposed a doubly robust estimation method for longitudinal data with dropouts. Wang, Jiang and Qiu (2015) proposed a robustified generalized likelihood ratio for longitudinal data using M- estimation method for estimating the coefficient functions.

In this paper, two methods of robust estimation namely robust maximum likelihood and robust Monte Carlo Newton-Raphson algorithm for joint longitudinal model has been compared with the classical maximum likelihood method of estimation to examine the association between the outcome variables of Spearman's G and S factors of human intelligence along with other covariates based on school lunch intervention study. The efficiency of the illustrated robust method of estimation for joint model is analysed by resampling technique with varying sample sizes. The paper is organized as follows: Section 2 follows the introduction in Section 1 and presents the methodology of robust estimation for joint mixed model approach for count data followed by bootstrapping methodology in Section 3. Section 4 presents the analysis based on joint modelling for the school lunch intervention dataset and results of parametric bootstrap technique. Section 5 of the paper provides the conclusion based on the application of the proposed method in cognitive studies.

2. METHODS OF ESTIMATION FOR JOINT MODELS

Consider a generalized linear mixed model with response $y_i = (y_1, ..., y_k)^t$, from subject i conditional on a vector of random effects u_i, are independent and follow a distribution $f_{y_{ij}|u_i}$ in the exponential family

$$f_{y_{ij}|u_i}(y_{ij}|u_i, \beta, \phi) = \exp\left\{\frac{y_{ij}\eta_{ij} - b(\eta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right\} \tag{1}$$

for some function a, b and c, with $i = 1, ..., K$ and $j = 1, ..., n_i$. The random effects $u_1, ..., u_k$ are assumed to have an independent multivariate normal density $f_{u_i}(u_i|\theta)$ with the mean vector 0 and covariance matrix $\sum(\theta)$, depending on the vector of unknown variance components $\theta = (\theta_1, ..., \theta_q)^t$. The variance-covariance matrix $\sum(\theta)$ of the i^{th} random effects vector u_i is not diagonal, in general, as the random effects within subjects can be correlated by nature.

The conditional mean response $\mu_{ij} = E(y_{ij}|u_i)$ depends on a set of covariates through the canonical link function $g(\mu_{ij}) = \eta_{ij} = x_{ij}^t\beta + z_{ij}^t u_i$ where x_{ij} and z_{ij} are covariates associated with the fixed and random effects, respectively. In this background (Sinha 2019), the conditional mean and variance functions are given by $\mu_{ij} = b'(\eta_{ij})$ and $\text{var}(y_{ij}|u_i) = a(\phi)b''(\eta_{ij})$, respectively, where $b'(\eta_{ij})$ and $b''(\eta_{ij})$ are the first and second derivatives of $b(\eta_{ij})$with respect to η_{ij}.

The density function $f_{u_{1i}|y_{1i}}(u_{1i}|y_{1i},\gamma)f_{u_{2i}|y_{2i}}(u_{2i}|y_{2i},\gamma)$ represents the conditional distribution of the subject random effects $u_{1i} \cdot u_{2i}$ given the observed response vector $y_{1i} \cdot y_{2i}$ from subject i. The maximum likelihood (ML) of γ may be obtained from the Newton-Raphson iterative equations $\gamma^{(m+1)} = \gamma^{(m)} - \{U'(\gamma^{(m)})\}^{-1}U(\gamma^{(m)})$, for $m = 0, 1, 2, ...$, where $U'(\gamma^{(m)})$ is the derivative of the score function $U(\gamma)$ with respect to γ evaluated at $\gamma = \gamma^{(m)}$ and $A(\gamma, u_i)$ actuality the "complete data" score vector for the i^{th} subject . McCulloch (1997) obtained the marginal likelihood of the parameters $\gamma = (\beta^t, \phi, \theta^t)^t$. The derivative $U'(\gamma)$ may be obtained as

$$\begin{aligned} U'(\gamma) = & \sum_{i=1}^{k} \iint \frac{\partial A(\gamma, u_{1i}, u_{2i})}{\partial \gamma} f_{u_{1i}|y_{1i}}(u_{1i}|y_{1i},\gamma)f_{u_{2i}|y_{2i}}(u_{2i}|y_{2i},\gamma)du_{1i}du_{2i} \\ & + \sum_{i=1}^{k} \iint A(\gamma, u_{1i}, u_{2i})A^t(\gamma, u_{1i}, u_{2i})f_{u_{1i}|y_{1i}}(u_{1i}|y_{1i},\gamma)f_{u_{2i}|y_{2i}}(u_{2i}|y_{2i},\gamma)du_{1i}du_{2i} \\ & - \sum_{i=1}^{k} \iint A(\gamma, u_{1i}, u_{2i})A^t(\gamma, u_{1i}, u_{2i})f_{u_{1i}|y_{1i}}(u_{1i}|y_{1i},\gamma)f_{u_{2i}|y_{2i}}(u_{2i}|y_{2i},\gamma)du_{1i}du_{2i} \\ & \times \iint A^t(\gamma, u_{1i}, u_{2i})f_{u_i|y_i}(u_{1i}|y_{1i},\gamma)(u_{2i}|y_{2i},\gamma)du_{1i}du_{2i}. \end{aligned}$$

Here the observed Fisher information is given by $I(\gamma) = -U'(\gamma)$. The asymptotic variance covariance matrix of the ML estimators $\tilde{\gamma}$ may be obtained from the Fisher information as $V(\tilde{\gamma}) = -I^{-1}(\gamma)$. The calculation of the likelihood function $L(\gamma)$, score function $U(\gamma)$ and Fisher information $I(\gamma)$ involves integrations with respect to the random effects u_i. For multidimensional random effects, these integrations require intensive computation, especially when there is a large number of subjects under study. McCulloch (1997) discussed some approximate algorithms to reduce the computational burden in this matter. When the underlying response distribution is correctly specified, the ML estimators $\tilde{\gamma}$ are typically asymptotically unbiased and follow an approximate normal distribution with the mean vector γ and variance-covariance matrix $V(\tilde{\gamma})$. These estimators, however, are generally sensitive to outliers or small departures from underlying distributional assumptions (Sinha, 2004). To bound the influence of outliers, we adopt a robust approach and given weights for influence points as described in Section 2.2.

The traditional method of estimating the parameters for mixed models is not resistant to the outliers or influential observations. In general, the influential observations or outliers are removed for any inferential analysis. However, there could be a chance of removing the important observations as outlier in the data which may lead to invalid conclusions (Li, Elashoff and Li, 2009). Yau and Kuk (2002) and Sinha (2004) considered the robust methods to limit the influential observations or outliers in estimating the parameters by maximizing a robust log-likelihood. Thus, the present study considers two methods of robust estimation namely robust maximum likelihood and robust Monte Carlo Newton-Raphson for joint longitudinal model for join modelling using generalized linear mixed model approach.

2.1 *Robust Estimation in Generalized Linear Mixed Models*

Standard statistical methods could seriously distort the estimates of regression parameters in the presence of outliers. Thus, a robust method was considered by Richardson and Welsh (1995) and Yau and Kuk (2002) to limit the influence of outlying observations on the estimate by maximizing a robust log-likelihood. Following the developments in linear mixed models (Fellner, 1986), the robust estimation procedure substitute pseudo-observations into the modified normal equation:

$$\begin{pmatrix} X'\Gamma^{-1}X & X'\Gamma^{-1}Z \\ Z'X\Gamma^{-1}X & X'\Gamma^{-1}Z + A^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} X'\Gamma^{-1}\tilde{y} \\ Z'\Gamma^{-1}\tilde{y} + A^{-1}\tilde{o} \end{pmatrix}, \tag{2}$$

where pseudo-observations $\tilde{y}$ and $\tilde{o}$ are defined by $\tilde{y} = X\tilde{\beta} + Z\tilde{u} + \Gamma^{1/2}\psi(\tilde{\xi})$, $\tilde{o} = \tilde{u} - A^{1/2}\psi(\tilde{\zeta})$ where $\tilde{\xi} = \Gamma^{1/2}\tilde{e}$, $\tilde{\zeta} = A^{-1/2}\tilde{u}$. Each of the components u_i has dimension v_i and distributed as $N(0, \theta_i A_i)$ and is independent of the other u-components. Here $\psi(\cdot)$ is the first derivative of the p-function in the robust log-likelihood. For the choice of $p(z) = z^2/2$, the estimation is equivalent to that in the previous section. One of the most commonly used p-function is the Huber p-function,

$$p(z) = \begin{cases} \dfrac{1}{2}z^2, & |z| \le K \\ K|z| - \dfrac{1}{2}K^2, & \text{otherwise.} \end{cases}$$

Subsequent applications assume the use of the Huber p-function with $K = 2$. Such a choice of K results in an efficiency of about 90% at the normal distribution in the one-component model (Yau and Kuk, 2002). For the Huber p-function, the derivate ψ is bounded but not re-descending, i.e. it does not satisfy the condition $\psi(z) \to 0$ as $|z| \to \infty$. The use of a re-descending ψ-function effectively deletes large residuals rather than down-weighting them. Though Richardson and Welsh (1995) used the Huber p-function, the subsequent development does not preclude the use of other common p-function for which the derivative ψ is bounded and re-descending, such as the Tukey bi-weight or Hampel piecewise linear ψ-functions. Richardson and Welsh (1995), expressed the constants $K_1 = E\{\psi(z)z\}$ and $K_2 = E\{\psi(z)\psi(z)\}$, where $z \sim N(0,1)$. Robust REML estimates of variance components are given by

$$\hat{\theta}_{(REMLI)} = \frac{\theta\psi(\tilde{\xi})'\tilde{\xi}}{K_1(n-p-v+r)}, \ \hat{\theta}_{i(REMLI)} = \frac{\theta_{1i}\psi(\tilde{\zeta}_{1i})'\tilde{\zeta}_{1i}}{K_1(v_{1i}+r_{1i})}, \frac{\theta_{2i}\psi(\tilde{\zeta}_{2i})'\tilde{\zeta}_{2i}}{K_1(v_{2i}+r_{2i})}, \tag{3}$$

where $\tilde{\zeta}_i = \theta_i^{-1/2} A_i^{-1/2} \tilde{u}_i$. Similarly, Yau and Kuk (2002) proposed a symmetric version of robust estimates of the variance components,

$$\hat{\theta}_{(REML)} = \frac{\theta\psi(\tilde{\xi})'\tilde{\xi}}{K_1(n-p-v+r)}, \hat{\theta}_{i(REML)} = \frac{\theta_{1i}\psi(\tilde{\zeta}_{1i})'\tilde{\zeta}_{1i}}{K_1(v_{1i}+r_{1i})}, \frac{\theta_{2i}\psi(\tilde{\zeta}_{2i})'\tilde{\zeta}_{2i}}{K_1(v_{2i}+r_{2i})} \tag{4}$$

Richardson and Welsh (1995) introduced two sets of variance component estimates termed respectively robust proposal REML I and II. Robust ML estimates of variance components can be obtained by replacing $n-p$ by n, r by $r^*$and r_i by r_i^*.

Here we describe how the robust estimation in linear mixed model can be extended to GLMM. Let us consider two outcome models from Ivanova, Molenberghs, Verbeke (2016), the joint density $f(Y_{1i}, Y_{2i})$ of the vectors Y_{1i} and Y_{2i} in *ith*-subject on n observations, $y_{1i} = (y_{11}, ..., y_{1n})', y_{2i} = (y_{21}, ..., y_{2n})'$, which can be written as $y_{1i} = \mu_1 + e_{1i},\ y_{2i} = \mu_2 + e_{2i}$ where $e_i = e_{1i}, e_{2i}$ is a vector of random errors, $e_{1i} = (e_{11}, ..., e_{1n})', e_{2i} = (e_{21}, ..., e_{2n})'$, with $E(e) = 0$ and $\text{var}(e_1) = \theta V(\mu_1)$, $\text{var}(e_2) = \theta V(\mu_2)$. Here θ is a dispersion parameter and $V(\mu)$ is a diagonal matrix of a function of $\mu = \mu_1 + \mu_2$. The expectation of $y_i = y_{1i}, y_{2i}$ for fixed $\mu = \mu_1 + \mu_2$, which relates to the linear predictor $\eta = \eta_{1i} + \eta_{2i}$ through a link function g, so $E(y_{1i}|u_1) = \mu_1, E(y_{2i}|u_2) = \mu_2,\ g(u_1) = \eta_1 = X\beta_1 + Zu_1,\ g(u_2) = \eta_2 = X\beta_2 + Zu_2$. Robust GLMM borrows from robust linear mixed models by an adjustment in the dependent variable and the variance of the error component. An adjusted dependent variable $y_{1i}^* = g(\mu_1) + g'(\mu_1)(y_{1i} - \mu_1), y_{2i}^* = g(\mu_2) + g'(\mu_2)(y_{2i} - \mu_2)$ (Schall, 1991) evaluated at the current parameter estimates. It follows that

$$y_{1i}^* \cong X\beta + Zu_1 + e_{1i}^*, y_{2i}^* \cong X\beta + Zu_2 + e_{2i}^*,$$

where $e_{1i}^* = (e_{11}^*, ..., e_{1n}^*)', e_{2i}^* = (e_{21}^*, ..., e_{2n}^*)';\ e_{1i}^* = e_{1i}\left(\frac{\partial\eta_{1i}}{\partial\mu_{1i}}\right), e_{2i}^* = e_{2i}\left(\frac{\partial\eta_{2i}}{\partial\mu_{2i}}\right)$

is called the adjusted error term. Its variance is $\text{var}(e_{1i}^*, e_{2i}^*) = \theta B^{-1}$, where θ is the dispersion parameter and $B = diag\left(V(\mu_{1i}, \mu_{2i})^{-1}\left\{\frac{\partial g^{-1}(\eta_{1i})}{\partial\eta_{1i}}\right\}^2\left\{\frac{\partial g^{-1}(\eta_{2i})}{\partial\eta_{2i}}\right\}^2\right)$.

Robust estimation in a GLMM is achieved by an iterative algorithm which replaces $\tilde{y}_i = \tilde{y}_{1i}, \tilde{y}_{2i}$ by $\tilde{y}_i^* = \tilde{y}_{1i}^*, \tilde{y}_{2i}^*$, $\tilde{e}_i = \tilde{e}_{1i}, \tilde{e}_{2i}$ by $\tilde{e}_i^* = \tilde{e}_{1i}^*, \tilde{e}_{2i}^*$ and Γ by θB^{-1} in (2) and then uses the robust estimating equations (3) and (4) until convergence is reached. Explicitly, the robust version of the modified normal equation is

$$\begin{pmatrix} X'(\theta^{-1}B)X & X'(\theta^{-1}B)Z \\ Z'(\theta^{-1}B)X & X'(\theta^{-1}B)Z + A^{-1} \end{pmatrix}\begin{pmatrix} \tilde{\beta} \\ \tilde{u}_1 \end{pmatrix} = \begin{pmatrix} X'(\theta^{-1}B)\tilde{y}_{1i}^* \\ Z'(\theta^{-1}B)\tilde{y}_{1i}^* + A^{-1}\tilde{o} \end{pmatrix},$$

$$\begin{pmatrix} X'(\theta^{-1}B)X & X'(\theta^{-1}B)Z \\ Z'(\theta^{-1}B)X & X'(\theta^{-1}B)Z + A^{-1} \end{pmatrix}\begin{pmatrix} \tilde{\beta} \\ \tilde{u}_2 \end{pmatrix} = \begin{pmatrix} X'(\theta^{-1}B)\tilde{y}_{2i}^* \\ Z'(\theta^{-1}B)\tilde{y}_{2i}^* + A^{-1}\tilde{o} \end{pmatrix} \tag{5}$$

where $\tilde{y}_{1i}^* = X\tilde{\beta} + Z\tilde{u}_1 + \theta^{1/2}\psi(\tilde{\xi}),\ \tilde{y}_{2i}^* = X\tilde{\beta} + Z\tilde{u}_2 + \theta^{1/2}\psi(\tilde{\xi}),\ \tilde{o} = \tilde{u} - A^{1/2}\psi(\tilde{\zeta})$, $\tilde{\xi} = \theta^{-1/2}B^{1/2}\tilde{e}_{1i}^*$, $\tilde{\xi} = \theta^{-1/2}B^{1/2}\tilde{e}_{2i}^*$ and $\tilde{\zeta} = A^{-1/2}\tilde{u}_1$, $\tilde{\zeta} = A^{-1/2}\tilde{u}_2$.

Robust variance component estimates are obtained as in (3) and (4). Since the matrices A and B are positive definite, their square roots are well defined. The approach can distinguish between outlying observations and outlying clusters via an examination of $\tilde{\xi}$ and $\tilde{\zeta}$ respectively (Schall, 1991). Yau and Kuk (2002) illustrated with an example of poisson model by using the same ψ-function in the expressions $\tilde{y}^*_{1i}, \tilde{y}^*_{2i}$ and $\tilde{o}$. Further, a possible extension is to use different ψ-function or tuning constants in these two expressions.

2.2 *Robust Estimation using Newton-Raphson Algorithm*

For simplicity, we consider the dispersion parameter $\phi = 1$, as it is the case for both binary and count data. Based on ML method in Section 2, to estimate the model parameters $\gamma = (\beta^t, \theta^t)^t$, we adopt the robust approach of Sinha (2004) with some modifications to estimate the variance components. Consider two longitudinal outcomes, say, Y_{1ij} and Y_{2ij} to denote the j^{th} measurement on the i^{th} subject respectively for the outcomes either continuous, ordinal or count in nature ($i = 1, \ldots, K$ $j = 1, \ldots, n_{1i}$ and $j = 1, \ldots, n_{2i}$). A joint model is built by describing the joint density $f(Y_{1\mathrm{i}}; Y_{2\mathrm{i}})$ of the first outcome vector Y_{1i} and second outcome vector Y_{2i}. Random effect $u_i = (u_{1i}, u_{2i})$ will be changed into $v_{1i} = \sum^{-1/2} u_{1i}, v_{2i} = \sum^{-1/2} u_{2i}$ normally distributed, given by

$$L(\beta, \theta) = \prod_{i=1}^{K} \int\int \prod_{j=1}^{n_{1i}} \prod_{j=1}^{n_{2i}} f_{y_{1ij}|v_{1i}}(y_{1ij}|v_{1i}, \beta, \phi) f_{v_{1i}}(v_{1i}|\theta) \times f_{y_{2ij}|v_{2i}}(y_{2ij}|v_{2i}, \beta, \phi) f_{v_{2i}}(v_{2i}|\theta) dv_{1i} dv_{2i} \tag{6}$$

In this situation, the canonical parameter is given by $\eta_{1ij} = x^t_{1ij}\beta + z^t_{1ij}\sum^{1/2} v_{1i}$, $\eta_{2ij} = x^t_{2ij}\beta + z^t_{2ij}\sum^{1/2} v_{2i}$. From (6), the ML estimating equations for γ may be obtained as

$$\sum_{i=1}^{k} \sum_{j=1}^{n_{1i}} \sum_{j=1}^{n_{2i}} E\Big[(\partial/\partial\gamma)\{\log f_{y_{1ij}|v_{1i}}(y_{1ij}|v_{1i}, \beta, \phi) \cdot f_{y_{2ij}|v_{2i}}(y_{2ij}|v_{2i}, \beta, \phi)\}|y_{1i}, y_{2i}\Big] = 0 \tag{7}$$

where E denotes the conditional expectation with respect to the random effects v_{1i}, v_{2i} given the observed data y_{1i}, y_{2i}. From (7), the ML estimating equations for β take the form

$$\sum_{i=1}^{k} \sum_{j=1}^{n_{1i}} \sum_{j=1}^{n_{2i}} E\big[\{y_{1ij} - \mu_{1ij}(\beta, \phi, v_{1i}) \cdot y_{2ij} - \mu_{2ij}(\beta, \phi, v_{2i})\} x_{2ij}|y_{1i}, y_{2i}\big] = 0 \tag{8}$$

where $\mu_{1ij}(\beta, \phi, v_{1i}) = E(y_{1ij}|v_{1i}), \mu_{2ij}(\beta, \phi, v_{2i}) = E(y_{2ij}|v_{2i})$. Clearly, the score function in the above equation is proportional to both (y_{1ij}, y_{2ij}) and (x_{1ij}, x_{2ij}), and hence is unbounded for outliers in both $y-$ and $x-$directions. To bound the influence of outliers, Sinha (2004) suggested robust estimators of the regression coefficients by solving the equations

$$
\begin{aligned}
&\sum_{i=1}^{k}\sum_{j=1}^{n_{1i}} E\left[\{\psi_{1c}(r_{1ij}) - E(\psi_{1c}(r_{1ij})|v_{1i})\}\sigma_{1ij}w(x_{1ij})x_{1ij}|y_{1i}\right] = 0,\\
&\sum_{i=1}^{k}\sum_{j=1}^{n_{2i}} E\left[\{\psi_{2c}(r_{2ij}) - E(\psi_{2c}(r_{2ij})|v_{2i})\}\sigma_{2ij}w(x_{2ij})x_{2ij}|y_{2i}\right] = 0
\end{aligned}
\tag{9}
$$

with respect to β, where

$$
r_{1ij} = (y_{1ij} - \mu_{1ij})/\sigma_{1ij},\; r_{2ij} = (y_{2ij} - \mu_{2ij})/\sigma_{2ij}
$$

with

$$
\mu_{1ij} = \mu_{1ij}(\beta, \theta, v_{1i}) = E(y_{1ij}|v_{1i}),\; \mu_{2ij} = \mu_{2ij}(\beta, \theta, v_{2i}) = E(y_{2ij}|v_{2i})
$$

and

$$
\sigma^2_{1ij} = \sigma^2_{1ij}(\beta, \theta, v_{1i}) = var(y_{1ij}|v_{1i}) = b''(\eta_{2ij}), \psi_{1c},
$$

$$
\sigma^2_{2ij} = \sigma^2_{2ij}(\beta, \theta, v_{2i}) = var(y_{2ij}|v_{2i}) = b''(\eta_{2ij}), \psi_{2c}
$$

where Huber's ψ-function is given by $(\psi_{1c}(r) \cdot \psi_{2c}(r)) = \max(-c, \min(r, c))$ and $w(x)$ is a weight function that is used to downweight any outliers in the covariates x. The tuning constant c in $\psi_{1c}(r) \cdot \psi_{2c}(r)$ is used to obtain a certain level of efficiency by the robust estimators at the underlying distribution. It is, however, difficult to find an optimal value of the tuning constant analytically.

The weight function $w(x)$ is given by $w(x) = \min(1, (\delta/d_x)^{\gamma/2})$ with $\gamma \geq 1$, where d_x is the Mahalanobis distance $d_x = (x - c_x)^t S_x^{-1}(x - c_x)$, with c_x and S_x being some robust estimators of the mean and variance of x, and δ is given by the 95th percentile of the χ^2_p distribution with $p = \dim(x)$. Note that for estimating the variance components θ, given a set of robust estimators of β, Sinha (2004) considers solving the score equations $E[(\partial/\partial\theta)\log f(u_i, \theta)|y_i] = 0$ with respect to θ, where E denotes the expectation with respect to the conditional distribution $f_{u|y}(u|y)$ of the vector of random effects $u_i = (u_1^t, ..., u_k^t)^t$ given the response vector $y_i = (y_1^t, ..., y_k^t)^t$. Further, when the data contains large outliers, the resulting estimators of θ can still give considerable bias. Thus, to improve the robustness properties of the estimators, Sinha (2019) proposed to solve the equations

$$
\begin{aligned}
\phi_l(\theta) &\equiv \sum_{i=1}^{k}\sum_{j=1}^{n_{1i}} E\left[\{\psi_{1c}(r_{1ij}) - E(\psi_{1c}(r_{1ij})|v_{1i})\}\sigma_{1ij}w(z_{1ij})z^t_{1ij}\Delta_l v_{1i}|y_{1i}\right] = 0,\\
\phi_l(\theta) &\equiv \sum_{i=1}^{k}\sum_{j=1}^{n_{2i}} E\left[\{\psi_{2c}(r_{2ij}) - E(\psi_{2c}(r_{2ij})|v_{2i}))\}\sigma_{2ij}w(z_{2ij})z^t_{2ij}\Delta_l v_{2i}|y_{2i}\right] = 0
\end{aligned}
\tag{10}
$$

with respect to θ_l, the l^{th}component of θ, for $l = 1, ..., q$ and $\Delta_l = \partial\sum^{1/2}/\partial\theta_l$. As before, here the expectations are taken with respect to the conditional distribution of the random effects v_{1i}, v_{2i} given the observed response vector y_{1i}, y_{2i}. The expectations, however, cannot be obtained in closed form and numerical methods are needed to compute these expectations. Equations (9) and (10) are solved simultaneously using a numerical method for the (RML) estimators $\hat{\beta}$ and $\hat{\theta}$. Here we describe the iterative Newton-Raphson method for solving these equations. For this, Sinha (2004) rewrite (9) in the matrix form

$$\sum_{i=1}^{k} E[X_{1i}^t W_{1i} d_{1i}, X_{2i}^t W_{2i} d_{2i} | y_{1i}, y_{2i}] = 0, \tag{11}$$

where X_{1i}, X_{2i} denotes the design matrix for subject *i*, W_{1i}, W_{2i} is a diagonal matrix with the *jth* diagonal element $w_{1ij} = \sigma_{1ij} w(x_{1ij}), w_{2ij} = \sigma_{2ij} w(x_{2ij})$, and the vector d_{1i}, d_{2i} has its *jth* element $d_{1ij} = \psi_{1c}(r_{1ij}) - E(\psi_{1c}(r_{1ij})|v_{1i})$, $d_{2ij} = \psi_{2c}(r_{2ij}) - E(\psi_{2c}(r_{2ij})|v_{2i})$.

Expanding the left side of (11) using a first-order Taylor series approximation and also using the Fisher scoring technique, the iterative equations for $\hat{\beta}$ may be obtained as

$$\begin{aligned}\beta^{(m+1)} = \beta^{(m)} + \left[\sum_{i=1}^{k} E(X_{1i}^t W_{1i} D_{1i} X_{1i}, X_{2i}^t W_{2i} D_{2i} X_{2i} | y_{1i}, y_{2i})\right]^{-1} \\ \sum_{i=1}^{k} E(X_{1i}^t W_{1i} d_{1i}, X_{2i}^t W_{2i} d_{2i} | y_{1i}, y_{2i}),\end{aligned} \tag{12}$$

where the second term on the right side is evaluated at $\beta = \beta^{(m)}$ for $m = 0, 1, 2, ...$, and D_{1i}, D_{2i} is a diagonal matrix with the *jth* diagonal element $D_{1ij} = -(\partial/\partial\eta_{1ij}) d_{1ij}. D_{2ij} = -(\partial/\partial\eta_{2ij}) d_{2ij}$. Similarly, from (10) the iterative equations for $\hat{\theta}$ may be obtained as

$$\begin{aligned}\theta^{(m+1)} = \theta^{(m)} + \left[\sum_{i=1}^{k} E(H_{1i}^t W_{1i}^* D_{1i} X_{1i}, H_{2i}^t W_{2i}^* D_{2i} X_{2i} | y_{1i}, y_{2i})\right]^{-1} \\ \times \sum_{i=1}^{k} E(H_{1i}^t W_{1i}^* d_{1i}, H_{2i}^t W_{2i}^* d_{2i} | y_{1i}, y_{2i}),\end{aligned} \tag{13}$$

where H_{1i}, H_{2i} is an $n_{1i} \times q, n_{2i} \times q$ with its *jth* row vector $h_{1ij}^t = (h_{1ij1}, ..., h_{1ijq})$, $h_{2ij}^t = (h_{2ij1}, ..., h_{2ijq})$ with the *lth* element $h_{1ijl} = z_{1ij}^t \Delta_l v_{1i}$, $h_{2ijl} = z_{2ij}^t \Delta_l v_{2i}$ and $W_{1i}^*, W_{2i}^*$is a diagonal matrix with the *jth* diagonal element $w_{1ij}^* = \sigma_{1ij} w(z_{1ij})$, $w_{2ij}^* = \sigma_{2ij} w(z_{2ij})$.

The Metropolis step is then incorporated into the Newton-Raphson iterative and second term on the right side of (13) is evaluated at $\theta = \theta^{(m)}$for $m = 0, 1, 2, ...$ There is a close connection between the adopted robust MCNR and the classical ML estimators introduced in Section 2. When $\psi_c(r) = r$ and $w(x) = 1$, the robust estimators are equivalent to ML estimators. The complete algorithm for estimating $\gamma = (\beta^t, \theta^t)^t$ may be described as follows:

Step 1: Choose initial values $\beta^{(0)}$ and $\theta^{(0)}$. Set $m = 0$. These initial estimates can be chosen as the ordinary MCNR estimates.

Step 2: (a) Calculate $\beta^{(m+1)}$ and $\theta^{(m+1)}$ from the iterative equations (12) and (13), respectively.

(b) Set $m = m + 1$.

Step 3: Continue step 2 until convergence is achieved. Declare the estimators at convergence to be the RMCNR estimators $\hat{\gamma} = (\hat{\beta}^t, \hat{\theta}^t)^t$ and here we take the joint log-likelihood as $l(\hat{\theta}^t|y_{1i}), l(\hat{\theta}^t|y_{2i})$. This study adopted the log-likelihoods Sakamoto (2018) and ideas from Yau and Kuk (2002) to obtain log-likelihood using Newton-Raphson algorithm.

The choice of a good set of initial estimates $\beta^{(0)}$ and $\theta^{(0)}$ is often necessary to improve the speed of convergence. The ordinary ML estimates of β and θ are an ideal choice of initial values, which may be readily obtained from standard software such as SAS or R. However, these initial values may not guarantee a convergence in iterations, and sometimes a trial and error method may be useful for finding a suitable set of initial values.

3. BOOTSTRAP FOR JOINT MODELS

For the case of classical estimators with no outliers, some authors use non-parametric bootstrap methods by generating bootstrap samples of residuals from the original dataset. These methods, however, are not suitable for data contaminated with outliers, as the bootstrap samples are influenced by outliers in the original dataset. Salibian-Barrera and Van Aelst (2008) noted that the proportion of outliers in bootstrap samples may be much higher than those in the original data, which may lead to poor performance from the non-parametric bootstrap method. Sinha (2019) proposed a parametric bootstrap method to approximate the mean squared prediction errors of the small area estimators and here we adopt the same method for joint modelling.

Let y_{1ij}, y_{2ij} denote a count response from the *jth* measure in the *ith* subject, consider a Poisson mixed model with an intercept random effect and Following Ivanova, Molenberghs and Verbeke (2016), a joint model is built by describing the joint density $f(Y_{1i}, Y_{2i})$ of the vectors Y_{1i} and Y_{2i}, random effect will be assumed that $u_i = (u_{1i}, u_{2i})$ respectively

$$\begin{aligned} &y_{1ij}|u_{1i} \sim \text{ind.Poisson } (\mu_{1ij}), \\ &y_{2ij}|u_{2i} \sim \text{ind.Poisson } (\mu_{2ij}), \quad i = 1, ..., k, \ j = 1, ..., n_{1i}, j = 1, ..., n_{2i} \\ &\mu_{1ij} = \exp(\beta x_{1ij} + u_{1i}), \mu_{2ij} = \exp(\beta x_{2ij} + u_{2i}), \\ &\mu_{1i}, \mu_{2i} \sim \text{ind. } N(0, \sigma_u^2) \end{aligned} \tag{14}$$

The classical estimates of β and σ_u can be obtained by solving the preceding equation numerically. McCulloch (1997) developed MCNR algorithm for solving these estimating equations, and obtained approximate ML estimates of the parameters. For given robust estimates $\hat{\beta}$ and $\hat{\sigma}_u$, generate subject-specific random effects u_{1i}^*, u_{2i}^* from $N(0, \sum(\hat{\sigma}_u))$ to create a bootstrap population $y_{1i}^* = \{y_{1ij}^*; j = 1, ..., N_{1i}\}$,

$y^*_{2i} = \{y^*_{2ij}; j = 1, ..., N_{2i}\}$ (with auxiliary variables x_{1ij}, x_{2ij} and z_{1ij}, z_{2ij}) from the distribution $f_{y_{1ij}|u_{1i}}(y_{1ij}|u_{1i}, \beta), f_{y_{2ij}|u_{2i}}(y_{2ij}|u_{2i}, \beta)$ using the linear predictor

$$\eta^*_{1ij} = g(\mu^*_{1ij}) = x^t_{1ij}\hat{\beta} + z^t_{1ij}u^*_{1i}, \eta^*_{2ij} = g(\mu^*_{2ij}) = x^t_{2ij}\hat{\beta} + z^t_{2ij}u^*_{2i}, \; i = 1, ..., k. \quad (15)$$

Compute the corresponding bootstrap population mean $\bar{Y}^*_{1i}, \bar{Y}^*_{2i}$.

Step 1. Generate a bootstrap random sample $\{y^*_{1ij}, y^*_{2ij};\; j \in s_{1i},\; j \in s_{2i},\; i = 1, ..., k\}$ from the above bootstrap population and compute the robust bootstrap estimators $(\hat{\beta}^*, \hat{\sigma}^*_u)$ and bootstrap predictors $\hat{u}^*_{1i}, \hat{u}^*_{2i}$, and hence robust predictors $t_{1i}t_{2i}(\hat{\beta}^*, \hat{\sigma}^*_u, \hat{u}^*)$ of the bootstrap subject mean $\bar{Y}^*_{1i}, \bar{Y}^*_{2i}$.

Step 2. Repeat Steps 1 and 2 to generate B bootstrap populations $\{y^{*(1)}_{1i}, ..., y^{*(b)}_{1i}, ..., y^{*(B)}_{1i}\}$, $\{y^{*(1)}_{2i}, ..., y^{*(b)}_{2i}, ..., y^{*(B)}_{2i}\}$ and a corresponding bootstrap sample from each bootstrap population. For the sample drawn from the b^{th} bootstrap population $\{y^{*(b)}_{1i}, y^{*(b)}_{2i}; i = 1, ..., k\}$, compute the corresponding bootstrap estimators $t_{1i}t_{2i}(\hat{\beta}^{*(b)}, \hat{\sigma}^*_u(b), \hat{u}^{*(b)})$ of the population subject means $\bar{Y}^*_{1i}(b), \bar{Y}^*_{2i}(b)$.

Obtain an estimate of mean squared prediction error and compute $\hat{\sigma}_u$, as well as the log-likelihood Sakamoto (2018) proposed marginal AICs for mixed-effects models based on a Monte Carlo method. $l(\hat{\sigma}^{(b)}_u|y_{1i}),\; l(\hat{\sigma}^{(b)}_u|y_{2i})$ of the original y_{1i}, y_{2i} data evaluated at $\hat{\sigma}_u{}^{(b)}$, for each bootstrap replicate.

The conditional mean and variance of the response (y_{1ij}, y_{2ij}) are given by $E(y_{1ij}|u_{1i}) = \exp(\beta x_{1ij} + u_{1i}), E(y_{2ij}|u_{2i}) = \exp(\beta x_{2ij} + u_{2i})$. Here the parameters of interest are the slope coefficient β and variance component σ^2_u. From (12), for given σ_u the iterative equation for $\hat{\beta}$ takes the form

$$\beta^{(m+1)} = \beta^{(m)} + \left[\sum_{i=1}^{k} E(X^t_{1i}W_{1i}D_{1i}X_{1i}, X^t_{2i}W_{2i}D_{2i}X_{2i}|y_{1i}, y_{2i})\right]^{-1}$$
$$\times \sum_{i=1}^{k} E(X^t_{1i}W_{1i}d_{1i}, X^t_{2i}W_{2i}d_{2i}|y_{1i}, y_{2i})|_{\beta=\beta^{(m)}}, \quad (16)$$

where the design matrix $X_{1i} = (x_{1i1}, ..., x_{1i1})^t, X_{2i} = (x_{2i1}, ..., x_{2i1})^t$, the weight matrix W_{1i}, W_{2i} has its *jth* diagonal element $w_{1ij} = \sqrt{\mu_{1ij}}w(x_{1ij}),\; w_{2ij} = \sqrt{\mu_{2ij}}w(x_{2ij})$ with $\mu_{1ij} = \mu_{1ij}(\beta, \sigma_u, v_{1i}),\; \mu_{2ij} = \mu_{2ij}(\beta, \sigma_u, v_{2i})$ and $v_{1i} = u_{1i}/\sigma_u, v_{2i} = u_{2i}/\sigma_u\tilde{N}(0, 1)$, the vector d_{1i}, d_{2i} has the *jth* element

$$d_{1ij} = \psi_{1c}(r_{1ij}) - E(\psi_{1c}(r_{1ij})|v_{1i}),$$
$$d_{2ij} = \psi_{2c}(r_{2ij}) - E(\psi_{2c}(r_{2ij})|v_{2i})$$

with $r_{1ij} = (y_{1ij} - \mu_{1ij})/\sqrt{\mu_{1ij}},\; r_{2ij} = (y_{2ij} - \mu_{2ij})/\sqrt{\mu_{2ij}}$, and the diagonal matrix D_{1i}, D_{2i} has the *jth* element

$$D_{1ij} = -(\partial/\partial\eta_{1ij})d_{1ij},$$
$$D_{2ij} = -(\partial/\partial\eta_{2ij})d_{2ij}$$

with $\eta_{1ij} = \beta x_{1ij} + \sigma_u v_{1i}$, $\eta_{2ij} = \beta x_{2ij} + \sigma_u v_{2i}$.

After some algebra, Sinha (2019) showed that

$$D_{1ij} = \sqrt{\mu_{1ij}}[\psi'_{1c}(r_{1ij}) - E\{\psi'_{1c}(r_{1ij})|v_{1i}\}] + (1/2)[r_{1ij}\psi'_{1c}(r_{1ij}) - E\{r_{1ij}\psi'_{1c}(r_{1ij})|v_{1i}\}] + \sqrt{\mu_{1ij}}E\{r_{1ij}\psi'_{1c}(r_{1ij})|v_{1i}\}$$

$$D_{2ij} = \sqrt{\mu_{2ij}}[\psi'_{2c}(r_{2ij}) - E\{\psi'_{2c}(r_{2ij})|v_{2i}\}] + (1/2)[r_{2ij}\psi'_{2c}(r_{2ij}) - E\{r_{2ij}\psi'_{2c}(r_{2ij})|v_{2i}\}] + \sqrt{\mu_{2ij}}E\{r_{2ij}\psi'_{2c}(r_{2ij})|v_{2i}\}.$$

From (13), for given β the iterative equation for $\hat{\sigma}_u$ takes the form

$$\sigma_u^{(m+1)} = \sigma_u^{(m)} + \left[\sum_{i=1}^{k} E(H_{1i}^t W_{1i}^* D_{1i} X_{1i}, H_{2i}^t W_{2i}^* D_{2i} X_{2i}|y_{1i}, y_{2i})|_{\sigma_u=\sigma_u^{(m)}}\right]^{-1} \times \sum_{i=1}^{k} E(H_{1i}^t W_{1i}^* d_{1i}, H_{2i}^t W_{2i}^* d_{2i}|y_{1i}, y_{2i})|_{\sigma_u=\sigma_u^{(m)}}, \quad (17)$$

where the design matrix H_{1i}, H_{2i} has the *jth* row $h_{1ij} = v_{1i}, h_{2ij} = v_{2i}$ and the diagonal weight matrix W_{1i}^*, W_{2i}^* has the *jth* diagonal element $w_{1ij}^* = \sigma_{1ij} = \sqrt{\mu_{1ij}}$, $w_{1ij}^* = \sigma_{1ij} = \sqrt{\mu_{1ij}}$. Sinha (2019) proposed bootstrap samples from the underlying distribution instead of a contaminated distribution for outliers. A similar behavior was observed in the bootstrap procedure of Sinha and Rao (2009) and Sinha (2019). This robust estimation is considered for the joint modelling for G factor with every S factor individually to examine the association between the response variables in the school lunch intervention study.

4. DATA ANALYSIS

A controlled Kenya school children feeding intervention study consists of 546 individuals involving the cognitive records among their intake of nutritional supplements (Neumann, Bwibo, Murphy, Sigman, Whaley, Allen, Guthrie, Weiss and Demment, 2003) is considered. Each nutrition group was comprised of 9 out of 12 schools with children aged 6-14. The school lunch intervention began at time t = 0 by adding the supplements: meat, milk and oil added as calories to determine the effects of human intelligence outcome measures. Responses are collected on Raven's Coloured Progressive Matrices test and analysed as a measure of G factor. Data are measured at five different points of time as indicated. Round 1 data is baseline data collected before the onset of intervention or as pre-intervention. Round 2 was taken as soon as the intervention started, while rounds 3, 4, and 5 were during the second, fourth, and sixth month after intervention started as indicating post intervention scores. At the end of study, only 374 individuals had a full-sequence data resulting from the fact that 172 individuals are missing after the first, second and fifth rounds of school lunch intervention. Hence, 374 subjects were considered in this intervention study out of them 189 are boys and 185 are girl children. 96 children were given calorie supplement, 126 children were given meat supplement, 80 were given milk and 72 were considered as control group in this study. Table 1 gives the descriptive summaries of treatment and outcomes observed at 5 time points.

TABLE 1. - *Descriptive summary*

		Minimum	**Maximum**	**Mean**	**Std. Deviation**
G Factor	Raven's coloured progressive matrices	0	31	5.49	3.09
S Factor	Arithmetic score	0	17	5.69	4.02
	Verbal meaning	0	40	8.53	8.09
	Digit span total	0	16	4.02	3.58
Covariates	Age	5	12	7.03	1.13
	Height	101.10	134.95	115.65	5.82
	Weight	14.30	50.15	20.23	3.38
	Head Circumstance	45.40	56.60	50.62	1.42
	Socio economic status	36.00	167.00	84.03	21.36
	Read test	0	12	6.91	5.10
	Write test	0	11	5.31	4.90

We have built a robust joint model for General Intelligence factor as measured by RCPM test and other three S-factors assessing VM, AS and DS with a view to study the association between outcomes, and how it evolves over time. There is also baseline covariate information on each subject, including gender, age, time, socio economic status, height, and weight. The study consists of developing suitable joint longitudinal model for the response variables (G and S factors) with set of covariates under consideration. In this context, the study on improvement of mental skills is correlated with the cognitive ability and this association can only be studied if both outcomes are modelled jointly using a suitable GLMM. The data analysis comprised of two steps. Firstly, the analysis starts with identifying the potential influential observations through model residuals in the data set. Finally, the small adjustment has been made to those influential observations and fitted the joint model to compare the performance between ML and Robust ML, Robust MCNR methods of estimation.

Gumedze and Chatora (2014) introduced the variance shift outlier model and studied the effect of outliers in over dispersed count data using generalized linear mixed model. Rakhmawati, Molenberghs, Verbeke and Faes (2015) considered count, binary and time to event data using GLMM and identified the influential observations and performed three methods of analysis to accommodate overdispersion in the data. Following Gumedze and Chatora (2014), the outliers are detected using GLMM. Residuals from the fitted GLMM for the response RCPM, AS, VM, and DS with the explanatory variables namely, milk, meat, calories, and control group reveal that the data contains few influential observations which may distort the estimates of the parameter. The observations 66, 245, 266, 270, 410, 548, 638, 652, 670, 693, 879, 985, 1012, 1260, 1341, 1392, and 1395 are identified as influential observation in RCPM among four nutritional supplements. Similarly, AM yields many observations as influential, whereas VM and DS contains sufficiently less influential observations. The diagrammatic representation of the influential observations through boxplot is presented in Figures 1 and 2.

FIGURE 1. - *Boxplots of RCPM and AS with respect to nutritional supplements*

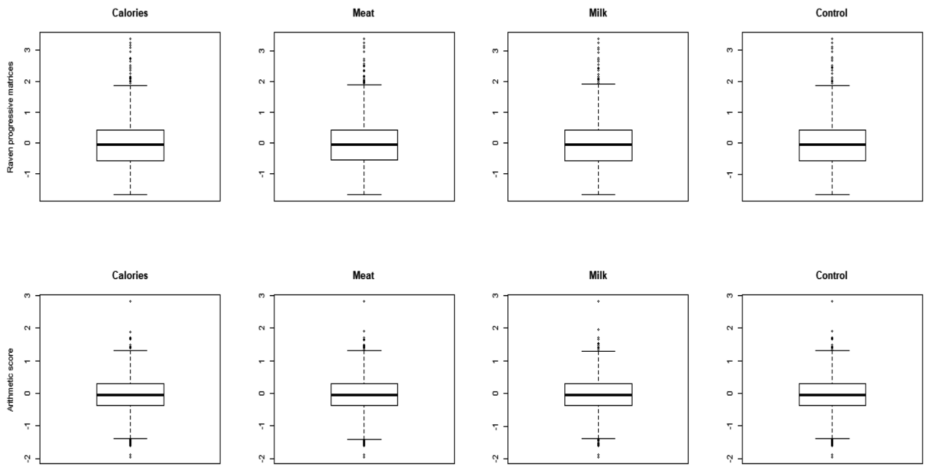
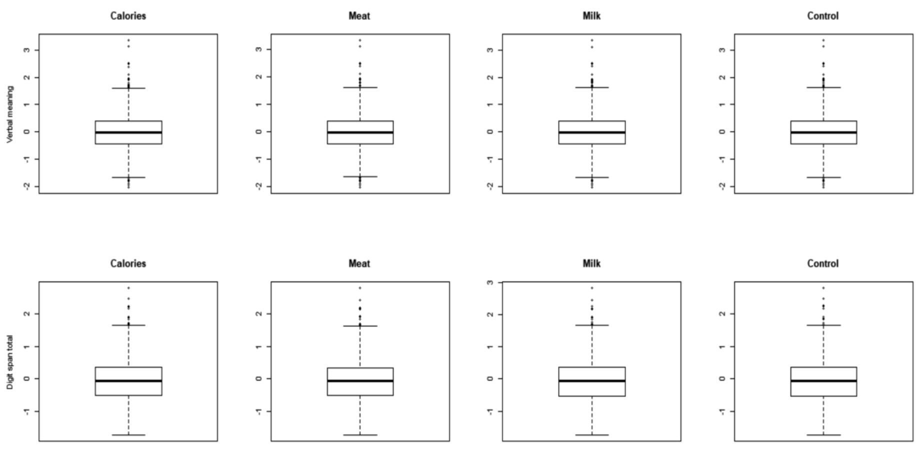


FIGURE 2. - *Boxplots of VM and DS with respect to nutritional supplements*



Once the influential observations are identified, an adjustment is included to those observations and the revised data set is analysed by fitting robust joint model between the G and S factors. Following Sinha (2004) and Yau and Kuk (2002), the idea of robust MCNR, robust ML are utilised by fitting joint model for human intelligence factors and the performance of the robust methods are compared with ML method. Cavanaugh and Neath (2019) discussed the properties and theoretical practices of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) towards the model selection process. Hence, we considered measures such as AIC, BIC and log-likelihood criteria to explore the objective of the study.

Joint modelling are included for RCPM, Y_{1ij} first outcome based on (14) and for AS/VM/DS, Y_{2ij} second outcome based on (14). Appendix A and presents the results of joint modelling with AIC, BIC, and log likelihood values of the cognitive ability assessed by RCPM as the G factor and VM, AS, DS as S factors comparing ML and RMCNR, RML methods.

For the RMCNR algorithm, the choice of the tuning constant c in Huber's psi function $\psi_{1c}(r) \cdot \psi_{2c}(r)$ is an important aspect. Following Sinha (2019) a suitable value of c based on a simulation study, was obtained using a sequence of values of c in the interval $[1, 2.5]$. The value of $c = 1.2$ produced estimates with approximately 90% efficiency at the underlying distributions. For robust ML estimation, $\psi(\cdot)$ is the first derivative of the Huber p-function in the robust log-likelihood used as suitable weights. In the robust estimation method, the identified influential observations are adjusted by giving weights, which leads to the absence of influential points in the joint model.

Based on the criteria values, meat supplements turnout to be the best choice in developing the cognitive abilities among the school children. From Appendix A it is proved and observed that

(i) the AIC, BIC and log-likelihood for the comparison of models on analytical ability by RCPM in association with numerical ability by AS are closer to each other for all the nutritional supplements but meat supplement yields the least criteria values, resulting in improvement of intelligence as compared to other supplements namely, milk, calories and control. Similarly, when comparing the analytical ability by RCPM in association with linguistic ability by Verbal Meaning and immediate memory by Digit Span total, the results revealed that meat supplement produces smaller criteria values indicating that the supplement of meat helps in improving the intelligence on children;

(ii) the analysis revealed that the robust MCNR performs well than other methods. The results revealed that the meat supplement is the best among the other nutritional supplements in developing the cognitive abilities and performance of school children;

(iii) lastly, in studying the association between the G - S factor, RCPM - DS showed better association than compared to other G - S factors by considering the outcomes as independent and the association between the outcome vectors are captured completely by the association between the random effects considered.

Sakamoto (2019) showed that the model with variance components possess boundary issues and the information criteria such as AIC, AICC, BIC, etc., gives more or less the same results and suggested that the bootstrap approach may be helpful in drawing a valid inference. Thus, the following section explores the bootstrap technique to evaluate the robust GLMM towards intervention study on cognitive behaviour.

Resampling in longitudinal repeated measures data has been investigated scarcely since 1990. Bootstrap is one of the most widely used resampling techniques in literature, which does not require any distributional assumptions of the data. Thai, Mentré, Holford, Veyrat-Follet and Comets (2013) compared bootstrap methods, such as parametric and nonparametric bootstrap that resample both random effects and residuals in linear mixed-effects models, to better take into account the hierarchical structure of multi-level and longitudinal data. Moreover, Sinha (2019) considered the Monte Carlo technique for the robust ML methods using parametric bootstrap method and the same method of resampling technique is adopted in this paper. It is also important to note that the proposed bootstrap method has been found to be useful for providing good approximations to the mean squared prediction errors of the small area estimators. Each simulation run was based on 1000 replicates of data sets. From (5), Schall (1991) proposed the use of iterative reweighted least squares estimation to obtain β, σ_u estimates by solving the overdetermined equations

$$\begin{pmatrix} (\theta^{-1/2}B^{1/2})X & (\theta^{-1/2}B^{1/2})Z \\ 0 & A^{-1/2} \end{pmatrix} \begin{pmatrix} \beta \\ \sigma_u \end{pmatrix} = \begin{pmatrix} (\theta^{-1/2}B^{1/2})y^*_{1i} \\ 0 \end{pmatrix} \begin{pmatrix} (\theta^{-1/2}B^{1/2})y^*_{2i} \\ 0 \end{pmatrix} \quad (18)$$

Yau and Kuk (2002) examined the performance of robust ML method and similarly, we used robust Monte Carlo NR algorithm for parametric bootstrap technique. For the Poisson mixed model in (14), the parameters were fixed at $\beta = 1$ and $\sigma_u = 0.25$. The values of the auxiliary variable x_{ij} were generated from the normal distribution $N(1, 1)$ for all population units $(i = 1, ..., k, j = 1, ..., N_{1i}, j = 1, ..., N_{2i})$. Initially, the estimators were studied for the case when no outliers were considered in the data. Then data were contaminated with 2% and 6% outliers and the results are presented in Appendix B.

To create outliers in the Poisson mixed model, we randomly replaced a certain proportion of y_{1i}, y_{2i} values by those obtained from the mixture of Poisson distributions: $0.5 \times \text{Poisson}(\lambda_1) + 0.5 \times \text{Poisson}(\lambda_2)$ with $\lambda_1 = \exp(0) = 1$ and $\lambda_2 = \exp(4) = 75.3$. These types of ''model deviations'' generally produce ''mean-shift outliers'' in the data, and here our goal is to assess whether the adopted method is robust against these outliers. The performance of the bootstrap samples towards the robust model is assessed by the model information criteria AIC, BIC and the log-likelihood.

The conclusions based on the bootstrap samples of varied outlier percentage (0%, 2%, and 6%) are as follows:

(i) as the number of outliers increases, the AIC, BIC, and log-likelihood criteria show that the joint model fit becomes worse irrespective of the nutritional supplements;

(ii) the numerical values of AIC, BIC and log-likelihood from bootstrap study shows that the nutritional supplement meat consistently performs well irrespective of the contamination of outliers;
(iii) based on the criteria values, the robust joint model fitted for factor RCPM DS yield better performance than the other factors.

It is clear from the empirical results that RMCNR, RML and ML methods perform almost equally well when there are no outliers in the data. The RMCNR method, however, appears to be more efficient than the RML, ML method when data are contaminated with outliers. So the robust method may be explored in practice even when there is no clear indication of outliers in the data.

5. CONCLUSIONS

Longitudinal studies are common in many psychometric studies particularly on cognitive ability of school children. Literature is abundant on studying the association between the G and S factors of human intelligence. However, many researchers are often interested to analyse each response separately. Gokul, Srinivasan and Swaminathan (2022) considered Concordance Correlation Coefficient as a measure of association and interchangeability among treatment groups. Here, we consider joint models with correlated random effects to capture the association between response variables and their covariates. The presence of aberrant or influential observations in the data may distort the inference of the data. Hence, there exists a need for robust estimation method, which improves the parameter estimates by giving due weights to the influential observations.

In this paper, we have illustrated two robust methods of estimation for joint modelling by imputing some adjustments to the outlying observations using GLMM approach. The sensitivity and efficiency of the robust joint model are examined through a bootstrap study with varying outlier percentages from the complete data. The Kenya intervention study suggests that the supplementation of meat has shown gradual improvement on cognitive performance of school children compared to other supplements. Based on the AIC, BIC, and log-likelihood value of the joint mixed model, the influence of ‘‘G’’ factor (Raven’s coloured progressive matrices) is quite likely to be felt more on attention (Digit Span) immediate memory than the other two ‘‘S’’ factors namely, Arithmetic Score and Verbal meaning involving reasoning abilities on linguistic and numerical domain considered to be of higher order cognitive functions.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referees for their constructive comments and suggestions, which helped to vastly improve the manuscript.

APPENDIX A

A.1 - *Joint model for RCPM versus three S factors AS, VM, and DS*

Nut.	Sup	RCPM-DS			RCPM - VM			RCPM - AS		
	Criteria	**ML**	**RML**	**RMCNR**	**ML**	**RML**	**RMCNR**	**ML**	**RML**	**RNCNR**
Calories	**AIC**	14623.41	14588.51	14242.72	16712.89	16541.17	16195.32	15719.72	15234.22	14888.42
	BIC	14789.33	14755.13	14376.87	16878.21	16707.71	16329.47	15884.27	15401.78	15022.97
	Log-Lik	7276.98	7143.66	6931.07	8316.25	8048.55	7835.96	7827.22	7245.42	7032.83
Milk	**AIC**	14615.31	14583.37	14237.52	16705.52	16533.67	16187.82	15713.67	15227.12	14881.92
	BIC	14780.49	14751.23	14372.97	16871.23	16701.12	16322.87	15395.77	15395.77	15017.47
	Log-Lik	7268.87	7140.51	6927.92	8307.24	8038.78	7826.19	7819.99	7239.91	7027.32
Meat	**AIC**	14624.42	14575.71	14229.92	16711.91	16520.27	16174.42	15721.52	15218.73	14872.92
	BIC	14790.17	14742.93	14364.67	16877.57	16689.93	16311.67	15884.97	15385.58	15007.27
	Log-Lik	7276.01	7131.96	6909.37	8314.35	8028.38	7805.79	7826.89	7229.18	7011.59
Control	**AIC**	14626.90	14600.78	14254.92	16716.48	16554.77	16208.92	15724.57	15244.53	14898.72
	BIC	14791.25	14768.22	14389.97	16880.33	16723.73	16345.47	15888.41	15.411.48	15033.17
	Log-Lik	7277.17	7156.54	6943.95	8317.04	8064.62	7852.03	7829.83	7264.21	7042.16

APPENDIX B

B.1 - *Average AIC, BIC and Likelihood obtained from Bootstrap Results* for RCPM versus AS

Nut Sup	Model	RCPM - AS								
	Methods	**ML**	**RML**	**RMCNR**	**ML**	**RML**	**RMCNR**	**ML**	**RML**	**RMCNR**
	Cont. Criteria	**6%**	**6%**	**6%**	**2%**	**2%**	**2%**	**0%**	**0%**	**0%**
Calories	**AIC**	22829.91	19375.87	18759.37	21341.07	18851.43	18319.16	20668.23	18521.47	18038.81
	BIC	23625.61	19760.61	18131.03	21645.43	19076.49	18531.9	20914.32	18720.28	18227.27
	LogLik	16487.31	13224.73	12619.9	15259.28	12777.42	12255.2	14723.5	12500.5	12031.96
Milk	**AIC**	18868.53	18051.79	17591.51	18084.72	17744.91	17365.27	17659.87	17488.81	17157.59
	BIC	18760.12	18297.89	17826.78	18135.94	17794.81	17403.64	17703.38	17523.39	17181.17
	LogLik	13788.32	12.167.51	11717.98	13334.36	11871.65	11504.77	13046.31	11623.86	11306.5
Meat	**AIC**	21204.42	15537.47	15134.6	20401.07	15236.26	14908.73	19902.65	15011.78	14731.71
	BIC	21892.1	15502.02	15086.09	20489.2	15252.56	14914.07	19955.8	15024.29	14733.84
	LogLik	15440.06	11629.19	11246.47	14681.47	11422.96	11109.15	14205.63	11255.05	10989.13
Control	**AIC**	24709.79	21394.06	20723.09	22923.46	20783.88	20196.24	22100.33	20387.8	19853.27
	BIC	24827.91	21150.18	20467.53	23123.47	20578	19980.34	22151.92	20120.18	19575.53
	LogLik	17669.18	14182.63	13524.72	16143.98	13658.88	13083.17	15341.83	13268.86	12748.94

B.2 - *Average AIC, BIC and Likelihood obtained from Bootstrap Results for RCPM versus VM*

Nut Sup	Model	RCPM - AS								
	Methods	ML	RML	RMCNR	ML	RML	RMCNR	ML	RML	RMCNR
	Cont. Criteria	6%	6%	6%	2%	2%	2%	0%	0%	0%
Calories	AIC	27754.31	23141.33	22419.38	24774.1	22184.75	21534.44	23472.5	21586.81	20992.11
	BIC	28252.07	23298.63	22563.01	25076.69	22296.19	21633.69	23527.06	21593.83	20988.01
	LogLik	20687.12	16445.29	15723.25	18042.35	15564.87	14921.8	17080.62	15102.28	14521.08
Milk	AIC	21856.22	21245.67	20676.32	20638.54	20526.87	20029.66	19936.72	20147.02	19704.63
	BIC	22839.17	21437.71	20857.72	20784.85	20751.65	20243.63	19986.64	20203.98	19750.69
	LogLik	17031.68	14950.25	14394.55	15895.92	14479.67	13995.88	15275.76	14199.12	13769.85
Meat	AIC	25299.75	18265.59	17754.42	23164.75	17850.19	17407.01	22384.69	17504.8	17114.59
	BIC	25760.88	18634.83	18112.94	23622.85	17969.81	17514.69	22439.8	17564.23	17162.1
	LogLik	18425.11	13439.92	12937.76	17165.45	13041.48	12610.85	16551.01	12729.6	12352.45
Control	AIC	30078.16	25188.35	24398.75	26628.73	24115.55	23406.91	24733.14	23263.89	22617.08
	BIC	31152.72	25308.75	24508.83	26927.14	23996.62	23277.79	24797.22	23049.14	22390.22
	LogLik	21423.25	17330.99	16553.21	18982	16525	15389.47	17781.82	15973.25	15339.4

B.3 - *Average AIC, BIC and Likelihood obtained from Bootstrap Results for RCPM versus DS*

Nut Sup	Model	RCPM - AS								
	Methods	**ML**	**RML**	**RMCNR**	**ML**	**RML**	**RMCNR**	**ML**	**RML**	**RMCNR**
	Cont. Criteria	**6%**	**6%**	**6%**	**2%**	**2%**	**2%**	**0%**	**0%**	**0%**
Calories	**AIC**	15857.01	14468.64	13955.58	14583.5	13613.69	13177.32	14158.99	13200.15	12823.57
	BIC	16316.95	14710.54	14184.36	14722.73	13656.86	13210.11	14191.32	13153.89	12767.27
	LogLik	14769.56	12281.68	11780.14	13536.84	11450.8	11026.81	13125.93	11047.69	10684.39
Milk	**AIC**	13756.11	12881.92	12475.98	12741.01	12164.73	11831.53	12402.66	11813.71	11538.6
	BIC	14115.89	12963.5	12541.98	12853.31	12209.08	11865.21	12432.45	11842.97	11557.59
	LogLik	12501.41	12127.81	11729.88	11540.37	11510.59	11191.26	11220.02	11206.93	10946.43
Meat	**AIC**	14432.73	12002.57	11652.8	13378.18	11309.09	11032.47	13026.66	10976.13	10.750.15
	BIC	14538.28	12245.16	11881.21	13430.85	11400.43	11109.63	13061.7	10994.56	10758.91
	LogLik	13749.73	10626.48	10285.97	12868.35	9966.18	9698.67	12574.56	9642.42	9431.71
Control	**AIC**	18308.44	16957.34	16388.11	16210.83	15595.4	15106.24	15511.63	14952.22	14523.57
	BIC	19568.56	17517.33	16936.09	16554.5	15593.93	15094.17	15549.82	14696.57	14256.93
	LogLik	16250.12	13695.44	13137.89	14448.24	12512.17	12036.74	13847.61	11951.01	11536.22

REFERENCES

Bilker W.B., Hansen J.A., Brensinger C.M., Richard J., Raquel E., Gur R.C. (2012). Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *HHS Public Access*, **19**(3), 354-369.

Cavanaugh J.E., Neath A.A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews - Computational Statistics*, **1460**, 1-11.

Coyle T.R., Elpers K.E., Gonzalez M.C., Freeman J., Baggio J.A. (2018). General Intelligence (g), ACT Scores, and Theory of Mind: (ACT)g Predicts Limited Variance Among Theory of Mind Tests. *Intelligence*, **71**, 85-91.

Fellner W.H. (1986). Robust estimation of variance components. *Technometrics*, **28**, 51-60.

Goharpey N., Crewther D.P., Crewther S.G. (2013). Research in Developmental Disabilities Problem solving ability in children with intellectual disability as measured by the Raven's Colored Progressive Matrices. *Research in Developmental Disabilities*, **34**(12), 4366-4374.

Gokul T., Srinivasan M.R., Swaminathan V.D. (2022). Measuring interchangeability in school lunch intervention data using concordance correlation coefficient. *Statistics and Application*, In Press.

Graßhoff U., Holling H., Schwabe R. (2016). Optimal Design for the Rasch Poisson-Gamma Model. In J. Kunert, C.H. Mller, A.C. Atkinson (Eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis* (pp. 133-141). Springer, Geneva, Switzerland.

Gueorguieva R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*, **1**, 177-193.

Gumedze F.N., Chatora T.D. (2014). Detection of outliers in longitudinal count data via overdispersion. *Computational Statistics and Data Analysis*, **79**, 192-202.

Hickey G.L., Philipson P., Jorgensen A., Kolamunnage-Dona R. (2018). A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *Journal of the Royal Statistical Society. Series A-Statistics in Society*, **181**(4), 1105-1123.

Iddi S., Molenberghs G. (2012). A joint marginalized multilevel model for continuous and binary longitudinal outcomes. *Journal of Applied Statistics*, **39**(11), 2413-2430.

Ivanova A., Molenberghs G., Verbeke G. (2016). Mixed models approaches for joint modelling of different types of responses. *Journal of Biopharmaceutical Statistics*, **26**(4), 601-618.

Li N., Elashoff R.M., Li G. (2009). Robust Joint Modeling of Longitudinal Measurements and Competing Risks Failure Time Data. *Biometrical Journal*, **51**(1), 19-30.

Lin H., Fu B., Qin G., Zhu Z. (2017). Doubly robust estimation of generalized partial linear models for longitudinal data with dropouts. *Biometrics*, **73**(4), 1132-1139.

Loeys T., Rosseel Y., Baten K. (2011). A Joint Modeling Approach for Reaction Time and Accuracy. *Psychometrika*, **76**(3), 487-503.

Markatou M., Afendras G., Agostinelli C. (2018). Weighted Cross validation in model selection. *Wiley Interdisciplinary Reviews: Computational Statistics*, **1439**, 1-12.

McCulloch C.E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, **92**, 162-170.

Muniz M., Gomes C.M.A., Pasian S.R. (2016). Factor structure of Raven's Coloured Progressive Matrices. *Psico-USF*, **21**(2), 259-272.

Neumann G.C., Bwibo N.O., Murphy S.P., Sigman M., Whaley S., Allen L.H., Guthrie D., Weiss R.E., Demment M.W. (2003). Animal Source Foods Improve Dietary Quality, Micronutrient Status, Growth and Cognitive Function in Kenyan School Children: Background, Study Design and Baseline Findings. *The Journal of nutrition*, **133**, 3941S-3949S.

Palestro J.J., Bahg G., Sederberg P.B., Lu Z.L., Steyvers M., Turner B.M. (2018). A tutorial on joint models of neural and behavioral measures of cognition. *Journal of Mathematical Psychology*, **84**, 20-48.

Pope L., Roche E., Morgan C.B., Kolodinsky J. (2018). Sampling tomorrow lunch today: Examining the effect of sampling a vegetable-focused entre on school lunch participation, a pilot study. *Preventive Medicine Reports*, **12**, 152-157.

Qin G., Bai Y., Zhu Z. (2012). Robust empirical likelihood inference for generalized partial linear models with longitudinal data. *Journal of Multivariate Analysis*, **105**(1), 32-44.

Rakhmawati T., Molenberghs G., Verbeke G., Faes C. (2015). Local influence diagnostics for generalized linear mixed models with overdispersion. *Journal of Applied Statistics*, **44**(4), 620-641.

Richardson A.M., Welsh A.H. (1995). Robust restricted maximum likelihood in mixed linear models. *Biometrics*, **51**, 1429-1439.

Sakamoto W. (2018). Bias-reduced marginal Akaike information criteria based on a Monte Carlo method for linear mixed-effects models Bias-reduced mAIC for LME models. *Scandinavian Journal of Statistics*, **46**(1), 1-29.

Sakamoto W. (2019). Inference on variance components near boundary in linear mixed effect models. *WIREs Computational Statistics*, **1466**, 1-10.

Salibian-Barrera M., Van Aelst S. (2008). Robust model selection using fast and robust bootstrap. *Computational Statistics and Data Analysis*, **52**, 5121-5135.

Schall R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719-727.

Sinha S.K. (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association*, **99**(466), 45-460.

Sinha S.K. (2019). Robust small area estimation in generalized linear mixed models. *Metron*, **77**, 201-225.

Sinha S.K., Rao J.N.K. (2009). Robust small area estimation. *Canadian Journal of Statistics*, **37**, 381-399.

Sutradhar B.C., Bari W. (2011). Robust quasi-likelihood inference in generalized linear mixed models with outliers. *Journal of Statistical Computation and Simulation*, **81**(2), 233-258.

Thai H.T., Mentr F., Holford N.H., Veyrat-Follet C., Comets E. (2013). A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. *Pharmaceutical Statistics*, **12**(3), 129-140.

Wang Z., Jiang J., Qiu Q. (2015). Robust Inference for Time-Varying Coefficient Models with Longitudinal Data. *Open Journal of Statistics*, **5**(7), 702-713.

Yau K.K.W., Kuk A.Y.C. (2002). Robust estimation in generalized linear mixed models. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **64**(1), 101-117.