ROBUSTNESS OF ADAPTIVE METHODS FOR BALANCED NON-NORMAL DATA: SKEWED NORMAL DATA AS AN EXAMPLE

Reema Mousa Abushrida, Loai Mahmoud Al-Zoubi, Amer Ibrahim Al-Omari*
Al al-Bayt University, Faculty of Science, Department of Mathematics, Mafraq, Jordan.

ABSTRACT

Estimation of variances of the estimated regression coefficients and their estimators is based on fitting a linear regression model. One method for allowing for clustering in fitting a linear regression model is to use a linear mixed model with two levels. It is probably suitable to ignore clustering and use a single level model if the intra-class correlation estimate is close to zero.

In this paper, a two-stage survey is used to evaluate an adaptive strategy for estimating the variances of estimated regression coefficients. The strategy is based on testing the null hypothesis that random effect variance component is zero. If this hypothesis is accepted the estimated variances of estimated regression coefficients are extracted from the one-level linear model. Otherwise, the estimated variance is based on the linear mixed model, or, alternatively the Huber-White robust variance estimator is used. A simulation study is used to show that the adaptive approach provides reasonably correct inference in a simple case.

KEYWORDS: Skew normal distribution; Adaptive strategies; Huber-White.
MSC: 62J05

RESUMEN

La estimación de las varianzas de los coeficientes de regresión estimados y sus estimadores se basan en el ajuste del modelo de regresión lineal. Un método para permitir el hacer un clustering al hacer el ajuste es usar un modelo o nivel si el coeficiente de correlación intra-clase estimado esta cerca de cero.

En este trabajo una encuesta de dos etapas es usada para evaluar una estrategia adaptativa para estimar las varianzas de los estimados coeficientes de regresión. La estrategia se basa en hacer una prueba de hipótesis sobre que el efecto aleatorio de la componente de varianza es cero. Si esta hipótesis es aceptada la varianza estimada de los coeficientes de regresión estimados son extraídos de un modelo lineal de un nivel. En otro caso la varianza estimada se basa en un modelo lineal mixto o, alternativamente en el estimador robusto de la varianza de Huber-White. Un estudio de simulación se usa para mostrar que el enfoque adaptativo provee de inferencias razonablemente correctas en el caso simple.

1. INTRODUCTION

1.1. Cluster and Multistage Sampling

The basic idea in sampling is the inference about the population of interest based on the information contained in a sample. Good designs of sampling methods involve the use of probability methods, minimizing decision in the choice of survey units (Cochran, 1977).

Two-stage sampling is one of the sampling methods. It is used in many surveys of social, health, economic, and demographic topics (Kish, 1965).

In multistage sampling, the population is divided into groups called primary sampling units (PSUs). A random sample from each considered PSU is then selected. If all units within each considered PSU are selected then two-stage sampling is called cluster sampling (Cochran, 1977)

This work is a part of M. Sc. thesis of Reema M.A. Abushrida under the supervision of Dr. Loai M. Al-Zoubi and Dr. Amer I. Al-Omari, Department of Mathematics, Al al-Bayt University.
*Corresponding author. Email: alomari_amer@yahoo.com

For example, PSUs might be schools and units might be students in schools, or PSUs might be households and units might be people, or PSUs might be geographic areas and units might be households (for example, see Cochran, 1977; Kish, 1965).

Two-stage sampling is typically used because of the following:

- When the list of elements not available, but the list of clusters available or can be easily configured.
- The availability of time, effort and costs.
- Easy to draw and easy to analyze.
- Within-group correlations may be of interest. For example; the correlation between values for students in the same school might be of interest.

A disadvantage of the two-stage that it gives estimates less accurate than other sampling techniques. The intra-class correlation (ICC), ρ, is a measure of the relationship between the observations of the personnel of the same PSU. If the intra-class correlation is non-zero, the clustered nature of the design should be reflected in the analysis procedure. This is done by fitting a multilevel model (MLM) (Goldstein, 2003). In fact the intra-class correlation is often very small. Note that the intra-class correlation may take a negative value, but generally positive values. If all PSU in the population contains a number of units *M*, the smallest possible value of ρ is $\frac{-1}{M-1}$. This case occurs when the population is limited and non homogeneous within PSUs (Kish, 1965).

In the case of a similar number of observations in each PSU, ρ is mostly less than 0.1 when PSUs are geographical areas and the final units are families in these areas (Verma et al., 1980). But if PSUs are families and the final units are people within families, ρ are usually between 0 and 0.2 (Clark and Steel, 2002).

The design effect (*deff*) is used to measure the increase in variance that happened when two-stage sampling is used. Kish (1965) defined *deff* as the ratio of the variance of an estimator under a specific design to the variance of the estimator under simple random sampling without replacement.

For large number of PSUs with *M* units in each PSU and *m* sample size from each PSU, the *deff* for the sample mean is given by

$$deff = 1 + (m-1)\rho \, . \tag{1.1}$$

The *deff* cannot be expressed in terms of ρ when PSUs have unequal sample sizes. Hence, for a proposed design, Kish (1965) approximated the design effect by

$$deff = 1 + (\bar{m}-1)\rho \, , \tag{1.2}$$

where $\bar{m}$ stands for the average PSU sample size.

1.2 The Two-Level Linear Mixed Model

1.2.1 The Model

Multilevel models are a generalization of regression models, and as such can be used for a group of things, multilevel modeling is almost always an improvement, but to different degrees. Goldstein (2003) defined the two-level linear mixed model (LMM) as

$$\mathrm{y}_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + b_i + e_{ij},\ i=1, 2,\ldots, c,\ j = 1, 2,\ldots, m_i, \tag{1..3}$$

where y_{ij} is the dependent variable of interest, $\mathbf{x}_{ij}$ is a vector of covariates for unit j in the PSU i, c denotes the number of PSUs in the sample, m_i denotes the number of observations selected from PSU i, $\boldsymbol{\beta}$ is the vector of unknown regression coefficients, $b_i \sim N(0, \sigma_b^2)$, and e_{ij} is error term, distributed as $N(0, \sigma_e^2)$. Thus, $\mathrm{y}_{ij} \sim N(\boldsymbol{\beta}'\mathbf{x}_{ij}, \sigma_b^2 + \sigma_e^2)$, where $\sigma_b^2 + \sigma_e^2 = \sigma_y^2$ is the variance of y. We can estimate the variance of regression coefficient either by standard likelihood theory (West et al., 2007), or by using the robust Huber-White estimator (Huber, 1967; White, 1982). Also, we can use the maximum likelihood or restricted maximum likelihood methods to estimate the model parameters.

A simple special case of Model (1.2) is the intercept-only model, which includes just a grand mean parameter, and it is defined by equating x_{ij} to 1 for all i, j.

$$\mathrm{y}_{ij} = \beta + b_i + e_{ij}, \quad i=1, 2,\ldots, c, \quad j = 1, 2,\ldots, m_i. \tag{1.4}$$

Model (1.2) can be generalized as

$$\mathbf{Y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \tag{1.5}$$

where $\mathbf{X}$ is the $n \times p$ design matrix, $\mathbf{Y}= (y_1', \ldots, y_c')'$ is the complete set of $n=\sum_{i=1}^{c} m_i$ observations in the c groups, $y_i= (y_{i1}, \ldots, y_{im_i})'$ is the observed vector for the i^{th} PSU, and $\mathbf{V} = \mathrm{diag}\,(\mathbf{V}_i$, $i= 1,\ldots, c)$ as

$$\mathbf{V}_i = \sigma_b^2\mathbf{J}_{m_i} + \sigma_e^2\mathbf{I}_{m_i}, \tag{1.6}$$

where $\mathbf{J}_{m_i}$ is an $m_i \times m_i$ matrix with all entries equal to 1, and $\mathbf{I}_{m_i}$ is the $m_i \times m_i$ identity matrix. $\boldsymbol{\beta}$ is the vector of unknown regression coefficients (Sahai and Ojeda, 2005).

Assume that b_i is uncorrelated with e_{ij}, and that b_i and $b_{i'}$ for $i \neq i'$ are uncorrelated. Therefore,

$$\mathrm{Var}\,(\mathrm{y}_{ij}) = \mathrm{Var}\,(b_i) + \mathrm{Var}\,(e_{ij}) = \sigma_b^2 + \sigma_e^2,\ \ \mathrm{Cov}\,(\mathrm{y}_{ij}, \mathrm{y}_{ij'}) = \mathrm{Var}\,(b_i) = \sigma_b^2 \quad \text{for} \quad j \neq j', \tag{1.7}$$

and

$$\mathrm{Cov}\,(\mathrm{y}_{ij}, \mathrm{y}_{i'j}) = 0 \qquad \text{for } i \neq i', \text{ (Rao, 1997).}$$

Assuming balanced data design, with $i = 1, \ldots, c$ and $(j \neq j') = 1,\ldots, m$, Rao (1997) defined the intra-class correlation as

$$\rho = \frac{\mathrm{Cov}\,(\mathrm{y}_{ij}, \mathrm{y}_{ij'})}{\sqrt{\mathrm{Var}\,(\mathrm{y}_{ij})\mathrm{Var}\,(\mathrm{y}_{ij'})}}. \tag{1.8}$$

Therefore, substituting (1.7) into (1.8), we obtain

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}. \tag{1.9}$$

Notice that, under Model (1.4) the intra-class correlation is always greater than or equal to 0. Given estimates $\hat{\sigma}_b^2$ and $\hat{\sigma}_e^2$ of σ_b^2 and σ_e^2, respectively, an estimator for ρ is

$$\hat{\rho} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_e^2}. \tag{1.10}$$

1.2.2. Likelihood Theory Estimation of Model Parameters

The variance components σ_b^2 and σ_e^2 are generally unknown, and are usually estimated by Restricted Maximum Likelihood (REML), giving estimates $\widehat{\mathbf{V}}_i$ of $\mathbf{V}_i$.

Patterson and Thompson (1971) is first introduced REML as a modification of maximum likelihood. The REML method is often presented as a technique based on maximization of the likelihood of a set of linear combinations of the elements of the response variable $\boldsymbol{y}$, say $\boldsymbol{k'y}$, where $\boldsymbol{k}$ is chosen so that $\boldsymbol{k'y}$ is free of fixed effects. One of the attractive aspects of REML is that it takes into account the degrees of freedom used up by the estimation of the fixed effects (Diggle et al., 1994). The restricted log-likelihood function is given by West et al. (2007, p.28) by the equation

$$\ell_R = -\frac{1}{2}\,[(n\text{-}1)\,log\,(2\pi) + log\,|\mathbf{V}| + log\,\left|\mathbf{X}'\,\mathbf{V}^{-1}\mathbf{X}\right| + \mathbf{Y}'\mathbf{V}^{-1}\,\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\,\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\}\mathbf{V}^{-1}\mathbf{Y}], \tag{1.11}$$

where $\mathbf{V} = diag\,(\mathbf{V}_i)$ and $\mathbf{V}_i$ are given by (1.6). Maximizing (1.11) with respect to σ_b^2 and σ_e^2 gives the REML estimates of these parameters. The REML estimate of $\widehat{\boldsymbol{\beta}}$ is given by

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\,\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\,\mathbf{X}'\,\widehat{\mathbf{V}}^{-1}\mathbf{Y} = \left(\sum_{i=1}^{c}\mathbf{x}_i'\,\widehat{\mathbf{V}}_i^{-1}\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{c}\mathbf{x}_i'\,\widehat{\mathbf{V}}_i^{-1}\mathbf{y}_i\right). \tag{1.12}$$

The REML estimates, in the intercept-only model are defined by the following system of equations:

$$\frac{n-c}{\hat{\sigma}_e^2} + \sum_{i=1}^{c}\frac{\hat{\lambda}_i}{m_i} - \frac{\sum_{i=1}^{c}\hat{\lambda}_i^2}{\sum_{i=1}^{c}\hat{\lambda}_i} = \frac{(n-c)MSE}{\hat{\sigma}_e^4} + \sum_{i=1}^{c}\frac{\hat{\lambda}_i^2}{m_i}\,(\bar{y}_{i.} - \hat{\beta})^2,$$

$$\sum_{i=1}^{c}\hat{\lambda}_i - \frac{\sum_{i=1}^{c}\hat{\lambda}_i^2}{\sum_{i=1}^{c}\hat{\lambda}_i} = \sum_{i=1}^{c}\hat{\lambda}_i^2\,(\bar{y}_{i.} - \hat{\beta})^2, \tag{1.13}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{c}\hat{\lambda}_i\bar{y}_{i.}}{\sum_{i=1}^{c}\hat{\lambda}_i},\ \frac{\hat{\lambda}_i^2}{m_i}$$

(Sahai and Ojeda, 2005, p.106), where $\bar{y}_{i.}$ is the mean of PSU i and $MSE = \frac{1}{n-c}\sum_{i=1}^{c}\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_{i.})^2$, and $\lambda_i = \frac{m_i}{\sigma_e^2 + m_i\sigma_b^2} = [var((\bar{y}_{i.})]^{-1}$, is the variance reciprocal of the mean of PSU i, and $\hat{\lambda}_i = \frac{m_i}{\hat{\sigma}_e^2 + m_i\hat{\sigma}_b^2}$, is the estimate of λ_i. The system of equations in (1.13) must be solved numerically with respect to $\hat{\sigma}_b^2$ and $\hat{\sigma}_e^2$. In the balanced data case ($m_i = m$ for all i), the REML estimates have a simpler form. Let $MSA = \frac{m}{c-1}\sum_{i=1}^{c}(\bar{y}_{i.} - y_{..})2$, the system of Equations (1.13) becomes:

$$\hat{\sigma}_e^2 = min\left(MSE\,,\frac{n-c}{n-1}\,MSE + \frac{c-1}{n-1}\,MSA\right),\ \hat{\sigma}_b^2 = \frac{1}{m}\,max\,(MSA - MSE,\,0),\ \hat{\beta} = \bar{y}_{..}.$$

(Sahai and Ojeda, 2005, p.40).

1.2.3 Likelihood Theory Estimation of *var* ($\widehat{\boldsymbol{\beta}}$)

The estimated variance of the REML $\hat{\beta}$ is given by

$$\widehat{var}\,(\hat{\beta}) = (\mathbf{X}'\,\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1} = \left(\sum_{i=1}^{c}\mathbf{x}_i'\,\widehat{\mathbf{V}}_i^{-1}\mathbf{x}_i\right)^{-1}, \tag{1.14}$$

where

$$\widehat{\mathbf{V}}_i = \hat{\sigma}_b^2\mathbf{J}_{m_i} + \hat{\sigma}_e^2\mathbf{I}_{m_i}. \tag{1.15}$$

In the balanced data case, where $m_i = m$, the variance estimator becomes

$$\widehat{var}\,(\hat{\beta}) = \frac{1}{c}\left[\hat{\sigma}_b^2 + \frac{\hat{\sigma}_e^2}{m}\right]. \tag{1.16}$$

A confidence interval for β could be constructed using the equation

$$(1-\alpha)\,100\%\ CI\,(\beta) = \hat{\beta} \pm t_{\left(df\,,1-\frac{\alpha}{2}\right)}\sqrt{\widehat{var}\,(\hat{\beta})}\,. \tag{1.17}$$

Faes et al. (2009) suggested the following approximate confidence interval for the mixed models based on a scaled t-distribution:

$$(1-\alpha)\ 100\%\ CI\ (\beta) = \hat{\beta}\ \pm \delta^{-1}\ t_{\left(df\ ,1-\frac{\alpha}{2}\right)} \sqrt{\widehat{var}\ (\hat{\beta}\ )}\ , \tag{1.18}$$

where

$$\delta= \sqrt{\frac{\upsilon}{(\upsilon-2)\hat{V}(T)}},\ \upsilon= \sum_{i=1}^{c} \frac{m_i}{1+(m_i-1)\hat{\rho}} - 1\ ,\ \hat{V}(T) = 1+ \left(\frac{\hat{\beta}^2}{4\left(\widehat{var}\ (\hat{\beta}\ )\right)^3}\ \widehat{var}\left[\widehat{var}\ (\hat{\beta}\ )\right]\right), \tag{1.19}$$

with $\widehat{var}\ (\hat{\beta}\ )$ defined in (1.14), $T = \frac{\hat{\beta}}{\sqrt{\widehat{var}\ (\hat{\beta}\ )}}$ and the scale factor δ was chosen so that the first two moments of δt agreed with the moments of $t_{\upsilon-1}$.

Faes et al. (2009) did not declare how to estimate $V(T)$ or $\widehat{var}\left(\widehat{var}(\hat{\beta})\right)$; we use the parametric bootstrap to estimate $\widehat{var}\left(\widehat{var}(\hat{\beta})\right)$. Other approaches have been suggested; see for example Satterthwaite (1941) and Kenward and Roger (1997). The method of Faes et al. (2009) has the advantage that it extends naturally to non-Gaussian model, unlike the other approaches.

1.2.4. Huber-White Estimator of *var* ($\hat{\boldsymbol{\beta}}$)

The generalized estimation equation (GEE) approach to linear modeling of clustered data can use either ordinary least squares (OLS) or generalized least squares (*GLS*).The *OLS* estimator for β is defined by

$$\hat{\boldsymbol{\beta}}_{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \tag{1.20}$$

and

$$var\ (\hat{\boldsymbol{\beta}}_{ols}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}\ (\mathbf{X}'\mathbf{X})^{-1}. \tag{1.21}$$

In general, $\mathbf{V}$ is not known and it can be estimated by $\hat{\mathbf{V}}$, therefore the estimated variance for $\hat{\boldsymbol{\beta}}_{ols}$ is defined by

$$\widehat{var}(\hat{\boldsymbol{\beta}}_{ols}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}\ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{1.22}$$

The estimator $\widehat{var}(\hat{\boldsymbol{\beta}}\ )$ in (1.14) is approximately unbiased provided that the variance model (1.6) is correct. Otherwise, $\widehat{var}(\hat{\boldsymbol{\beta}}\ )$ is biased and the inference will be incorrect. An alternative to ML or REML estimates of $var\ (\hat{\boldsymbol{\beta}}\ )$ is the robust variance estimate approach described by Liang and Zeger (1986), this approach can be applied to the analysis of data collected using PSUs.

This approach can be referred to as robust or Huber-White variance estimation (Huber, 1967; White, 1982). The method yields asymptotically consistent covariance matrix estimates even if the variances and covariance's assumed in Model (1.1) are incorrect. It is necessary to assume that observations from different PSUs are independent. In Equation (1.12) the variance of $\hat{\boldsymbol{\beta}}$ is estimated by substituting REML estimates of σ_b^2 and σ_e^2 into $\mathbf{V}_i$. An alternative estimator of $\mathbf{V}_i$ is $\hat{\mathbf{V}}_i^{Hub} = \hat{e}_i\hat{e}_i'$, where $\hat{e}_i = y_i - \mathrm{x}_i'\hat{\beta}$, $\hat{\mathbf{V}}_i^{Hub}$ is approximately unbiased for $\mathbf{V}_i$ even if (1.4) does not apply.

Take the expectation of $\hat{\mathbf{V}}_i^{Hub}$ to find that

$$E\left(\hat{\mathbf{V}}_i^{Hub}\right) = E\left(\hat{e}_i\hat{e}_i'\right) \approx E\left[(y_i - \mathrm{x}_i'\beta)((y_i - \mathrm{x}_i'\beta)'\right] = \mathbf{V}_i\ . \tag{1.23}$$

Note that

$$\begin{aligned} var\ (\hat{\boldsymbol{\beta}}\ ) &= var\left(\left(\sum_{i=1}^{c} \mathbf{x}_i'\ \hat{\mathbf{V}}_i^{-1}\mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{c} \mathbf{x}_i'\ \hat{\mathbf{V}}_i^{-1}\mathbf{y}_i\right)\right) \\ &\approx \left(\sum_{i=1}^{c} \mathbf{x}_i'\ \hat{\mathbf{V}}_i^{-1}\mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{c} \mathbf{x}_i'\ \hat{\mathbf{V}}_i^{-1}\mathbf{V}_i\ \hat{\mathbf{V}}_i^{-1}\ \mathbf{x}_i\right) \left(\sum_{i=1}^{c} \mathbf{x}_i'\ \hat{\mathbf{V}}_i^{-1}\mathbf{x}_i\right)^{-1}. \end{aligned} \tag{1.24}$$

One way to construct a robust estimator of $var\ (\hat{\boldsymbol{\beta}})$ is to substitute the robust estimator $\hat{\mathbf{V}}_i^{Hub}$ in (1.24) as follows (Liang and Zeger, 1986):

$$\widehat{var}_{Hub}(\hat{\beta}) = \left(\sum_{i=1}^{c} \mathbf{x}_i' \ \hat{\mathbf{V}}_i^{-1}\mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{c} \mathbf{x}_i' \ \hat{\mathbf{V}}_i^{-1}\hat{\mathbf{V}}_i^{Hub}\hat{\mathbf{V}}_i^{-1}\ \mathbf{x}_i\right) \left(\sum_{i=1}^{c} \mathbf{x}_i' \ \hat{\mathbf{V}}_i^{-1}\mathbf{x}_i\right)^{-1}. \quad (1.25)$$

Using the intercept only model ($\mathbf{x}_{ij}$=1), (1.25) becomes:

$$\widehat{var}_{Hub}(\hat{\beta}) = \frac{\sum_{i=1}^{c} \hat{\lambda}_i^2 \left(\bar{y}_{i.} - \hat{\beta}\right)^2}{\left(\sum_{i=1}^{c} \hat{\lambda}_i\right)^2}. \quad (1.26)$$

Since $\hat{\lambda}$ is constant, (1.26) reduces in the balanced data case (m_i= m), to:

$$\widehat{var}_{Hub}(\hat{\beta}) = \frac{1}{c(c-1)} \sum_{i=1}^{c}(\bar{y}_{i.} - \bar{y}_{..})^2. \quad (1.27)$$

1.4. Testing $H_0 : \sigma_b^2 = 0$ in the Linear Mixed Model Using RLRT

The problem of testing H_0: σ_b^2= 0 using the likelihood ratio test (LRT) is discussed by Self and Liang (1987) using ML estimators for the variance components. A best choice is to use REML estimators to derive the LRT statistic for testing H_0: σ_b^2= 0. Self and Liang (1987) assumed that the true parameter values are on the boundary of the parameter space, and showed that the large sample distribution of the likelihood ratio test is a mixture of χ^2 distributions under nonstandard conditions provided that the response variables are *iid*.

Stram and Lee (1994) used the results of Self and Liang (1987) to prove that the asymptotic distribution of the LRT for testing H_0: σ_b^2= 0 has an asymptotic 50:50 mixture of χ^2 with 0 and 1 degrees of freedom under H_0 rather than the classical single χ^2 if the data are *iid* under the null and alternative hypotheses. From (1.11), the restricted likelihood ratio test is given by

$$\Lambda = -2\ log\ (\text{RLRT}) = 2 \ \underset{H_1}{MAX}\ \ell_R(\beta,\sigma_b^2,\sigma_e^2) - 2\ \underset{H_0}{MAX}\ \ell_R(\beta,\sigma_b^2,\sigma_e^2). \quad (1.28)$$

In the intercept-only model case (1.4) assuming balanced data, Visscher (2006) introduced the REML-based likelihood ratio test as

$$\Lambda=\begin{cases}(n-1)\log\left(\frac{n-c}{n-1}+\frac{c-1}{n-1}\,F\right)-\ (c-1)\log(F) & if \quad F>1\\ 0 & if \quad F\ \le\ 1,\end{cases} \quad (1.29)$$

where $F = \frac{MSA}{MSE}$.

1.5. Adaptive Procedures

Adaptive sampling provides a fit solution to the problem estimate. One method is to fit mixed linear model in any case. Another method is to fit a linear model assuming independent observations, i.e. ρ= 0. But, if a large number of final units are selected from each PSU, variations resulting from the estimated linear mixed model can be much larger than those obtained from the linear model with independent observations, which would lead to broader confidence intervals; also linear mixed model is more complex than the simple linear model.

Here, we will provide third alternative method for estimation, which depends on testing the null hypothesis, as σ_b^2 = 0. If the null hypothesis is not rejected we use the linear model for estimating the variances of the estimated regression coefficient $\hat{\beta}$. Moreover, if the null hypothesis is rejected we use the estimated variance for $\hat{\beta}$, either using the standard likelihood theory variance estimator for the LMM ($\widehat{var}_{LMM}(\hat{\beta})$) or the Huber-White method ($\widehat{var}_{Hub}(\hat{\beta})$).

1.6. Skew Normal Distributions and Their Properties

Normal distribution is the most popular distribution because of its many attractive properties and moreover there are two main reasons for its popularity: the first, is the effect of the central limit theorem, in most cases the distribution of observations is at least approximately normal; the second, the normal distribution and its sampling distribution are easily tractable.

The idea of skew-normal distribution was first introduced by Azzalini (1985). It is an extension of the normal distribution through the shape parameter γ. It is ended up with standard normal random variable for γ = 0 and to half-normal when γ approaches ∞.

We know that the probability density function (pdf) and cumulative distribution function (*cdf*) of the standard normal random variable are given as follows

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \qquad \Phi(x) = \int_{-\infty}^{x} \phi(u)du\,, \qquad \phi(x) = \phi(-x)\,, \qquad \text{and} \;\; \Phi(x) + \Phi(-x) = 1.$$

where the product of $\phi(x)$ and $\Phi(x)$, gives another attractive class of random variables. It is called the skew-normal distributions. (Azzalini (1985)), with skewness parameter γ.

The study of the skew normal distribution explores an approach for statistical analysis without the symmetry assumption for the underlying distribution.

Let *Y* be a continuous random variable. Let ϕ and Φ denote the standard normal density and corresponding distribution function, respectively. Then *Y* is said to have a skew-normal distribution with the parameter γ, $-\infty < \gamma < \infty$ if the density of Y is

$$f(\mathrm{y}; \gamma) = 2\, \phi(\mathrm{y})\, \Phi\,(\gamma y), \qquad -\infty < \gamma\,,\, \mathrm{y} < \infty, \tag{1.30}$$

where ϕ and Φ denote the standard normal density and corresponding distribution function, respectively, i.e., Y ~ *SN* (γ). (Azzalini, 1985).

The component γ is called the shape parameter because it regulates the shape of the density function. As γ increases (in absolute value), the skewness of the distribution increases.

The mean and the variance at the skew normal random variable *Y* are, respectively E(*Y*) = $\sqrt{\frac{2}{\pi}}\,\delta$, and Var(Y) = $1-\left(\frac{2}{\pi}\right)\delta^2$, where $\delta = \frac{\gamma}{\sqrt{1+\gamma^2}}$.

In practice, to fit real data we work with an affine transformation Z = μ +σY, with $\mu \in \mathbb{R}$ and $\sigma > 0$. Then the density of Z can be written as:

$$\mathrm{g}\,(\mathrm{z};\mu, \sigma, \gamma\,) = \frac{2}{\sigma}\;\phi\left(\frac{2-\mu}{\sigma}\right)\Phi\left(\frac{2-\mu}{\sigma}\,\gamma\right). \tag{1.31}$$

It can be written in abbreviation as Z ~ SN (μ, σ^2, γ).

Azzalini (1985) showed the following properties:

- As γ tends to ∞, (1.29) becomes f(y) = ϕ(y), $0 \le \mathrm{y} < \infty$ which is the half-normal **(**folded normal) probability density function.
- If Y~ N (0; 1) and *X* ~SN (γ), then both |Y| and |X| have the same pdf.
- If Y ~SN (γ), then $Y^2 \sim\; \chi^2_1$.

- If Y ~SN (γ), then $-Y$ ~SN ($-\gamma$).

In this section, the methods based on fitting a linear mixed model are applied to data having skew-normal distribution.

The skew-normal distribution refers to a parametric class of probability distributions which includes the standard normal as a special case.

2. ADAPTIVE STRATEGIES

Two adaptive strategies are considered here. Both of them depend on the idea of testing the variance component σ_b^2 in Model (1.2). If we reject H_0: $\sigma_b^2 = 0$ in the first, the LMM estimators of $var\,(\hat{\beta})$ defined in Equation (1.14) will be used. The standard linear model with independent errors will be used if H_0 is not rejected, as in this case σ_b^2 will be assumed to be zero.

The robust Huber White estimator $\widehat{var}_{Hub}(\hat{\beta})$ will be used instead of $\widehat{var}_{LMM}(\hat{\beta})$ if H_0 is rejected, as the second adaptive strategy.

The two adaptive strategies (ADM) and (ADH) are defined as

$$\widehat{var}_{ADM}(\hat{\beta}) = \begin{cases} \widehat{var}_{LMM}(\hat{\beta}) & \text{if} \quad H_0 \ \text{ is not rejected} \\ \widehat{var}_{LM}(\hat{\beta}) & \text{if} \quad H_0 \ \text{ is rejected,} \end{cases} \tag{2.1}$$

and

$$\widehat{var}_{ADH}(\hat{\beta}) = \begin{cases} \widehat{var}_{Hub}(\hat{\beta}) & \text{if} \quad H_0 \ \text{ is not rejected} \\ \widehat{var}_{LM}(\hat{\beta}) & \text{if} \quad H_0 \ \text{ is rejected.} \end{cases} \tag{2.2}$$

The Huber-White variance estimator is approximately an unbiased. For the intercept-only model, it is easy to show that

$$\frac{E\left(\widehat{var}_{Hub}(\hat{\beta})\right)}{var\,(\hat{\beta})} = \frac{\left(\sum_{i=1}^{c}\lambda_i\right)^2 - \sum_{i=1}^{c}{\lambda_i}^2}{\left(\sum_{i=1}^{c}\lambda_i\right)^2}, \tag{2.3}$$

where $\hat{\beta}$ and $var\,(\hat{\beta})$ are given by (1.13) and (1.26), respectively. Hence, a bias-adjusted estimator is given by dividing (1.26) by the right hand side of (2.4), yields

$$\widehat{var}_{Hub}(\hat{\beta}) = \frac{1}{\left(\sum_{i=1}^{c}\hat{\lambda}_i\right)^2 - \sum_{i=1}^{c}\hat{\lambda}_i^2} \sum_{i=1}^{c}\hat{\lambda}_i^2\left(\bar{y}_{i.} - \hat{\beta}\right)^2. \tag{2.4}$$

The LMM 90% confidence intervals for β are given by

$$(1-\alpha)\ 100\%\ CI = \hat{\beta} \pm \delta^{-1}\, t_{\left(df\,,1-\frac{\alpha}{2}\right)} \sqrt{\widehat{var}\ (\hat{\beta}\,)}\,, \tag{2.5}$$

where $\delta = \sqrt{\frac{v}{(v-2)\hat{V}(T)}}$, α = 0.1 and the degrees of freedom (*df*) are defined to be:

$$df = \begin{cases} n-1, & \text{using } LM\ Est. \\ v-1, & \text{using } LMM\ Est. \\ c-1, & \text{using } Huber-White\ Est. \end{cases} \tag{2.6}$$

Degrees of freedom for adaptive strategies ADM and ADH are defined as

$$df_{ADM} = \begin{cases} n-1 & \text{if} \quad H_0 \quad \text{is not rejected} \\ v-1 & \text{if} \quad H_0 \quad \text{is rejected,} \end{cases} \tag{2.7}$$

and

$$df_{ADH} = \begin{cases} n-1 & \text{if} \quad H_0 \quad \text{is not rejected} \\ c-1 & \text{if} \quad H_0 \quad \text{is rejected,} \end{cases} \tag{2.8}$$

where υ represents the effective sample size, with $\hat{\upsilon} = \frac{n}{\widehat{deff}\ (\hat{\beta}\ )}$. The effective sample size is the ratio of the sample size to the design effect of the $\widehat{\beta}$. The degrees of freedom for the linear mixed model are only an approximation (Faes et al., 2009). However, the degrees of freedom of the linear model and Huber-White are exact (MacKinnon and White, 1985).

The adaptive confidence intervals may not have the correct coverage rates as they might not incorporate the model selection uncertainty. The extent of this problem will be evaluated by simulation. An alternative approach would be to fit both the LM and LMM and base estimates and inference on model averaging of these two models (see for example Hoeting et al., 1999; Yuan and Yang, 2005). The adaptive method was tested by AL-Zoubi *et at*., 2010 for the normal data and by AL-Zoubi (2012) for exponential data.

3 SIMULATION STUDY

In this section, a simulation study is conducted to compare the adaptive and non-adaptive methods for estimating *var* ($\widehat{\beta}$). Data are generated from the skew normal distribution, with m_i= *m* and an intercept only Model (1.4). The values of ρ, *m* and *c* are varied. The value of γ is fixed to be 1. 1000 samples are generated in each case. The values of σ_b^2 and σ_e^2 are set to $\frac{\rho}{1-\rho}\ \gamma$ and 1, respectively, to ensure that the intra-class correlation is ρ.

For each sample the estimated regression coefficient $\widehat{\beta}$ and the estimators of *var* ($\widehat{\beta}$) are calculated for the LMM and LM models using the *lme4* and *lm* packages (Pinheiro and Bates, 2000) in the R statistical environment (R Development Core Team, 2007). The true variance of $\widehat{\beta}$ is determined by calculating the variance over all 1000 simulations.

The two adaptive strategies ADM and ADH are defined by (2.2) and (2.3) and 90% confidence intervals are calculated for the LMM method using the method of Faes et al. (2009). Huber-White confidence intervals and the adaptive confidence intervals are calculated as discussed in Section 1.2.4. The hope is that the adaptive procedures give shorter confidence intervals as they will use the LM when H_0 is not rejected and for small sample sizes these cases still have $\hat{\rho}$ away from zero. As the sample size increases, H_0 will only be not rejected when $\hat{\rho}$ is close to zero.

The restricted likelihood ratio test described in Section 1.4 is evaluated for testing H_0: $\sigma_b^2 = 0$ The parameter ρ is varied over a range of values of 0, 0.01, 0.025, 0.05 and 0.1; *c* is varied over 2, 5, 10 and 25; and *m* is varied over 2, 5, 10, 15, 25 and 50.

3.1. Simulation Results on Adaptive Confidence Intervals for $\boldsymbol{\beta}$ for Balanced Data

A simulation study based on equal sized PSUs, m_i= *m*, and an intercept only model is conducted to compare the adaptive and non-adaptive methods for estimating *var* ($\widehat{\beta}$). In this study we used the parametric bootstrap to estimate *V* (*T*) because the scale parameter δ relies on *V* (*T*) (see Equation 1.19) and Faes et al. (2009) did not specify how *V* (*T*) can be estimated.

To apply the parametric bootstrap method to estimate var (T), 100 samples are generated from the intercept-only model (1.4) with variances $\hat{\sigma}_b^2$ and $\hat{\sigma}_e^2$. For each sample, we estimated β and var ($\hat{\beta}$) to find the value of $T = \frac{\hat{\beta}}{\sqrt{\widehat{var}\ (\hat{\beta}\ )}}$. The variance of the 100 values of T was calculated and used to estimate V (T).

Another way to estimate var (T) is to estimate $\widehat{var}\left[\widehat{var}\ (\hat{\beta}\ )\right]$, and then substitute into (1.19), but Faes et al. (2009) also didn't specified how to estimate this parameter, therefore we have tried to do that using the parametric bootstrap. The same procedure above is used, but now we estimated var ($\hat{\beta}$) from the fitted model and then calculated the variance of the 100 estimated values of var ($\hat{\beta}$). Then $\widehat{var}$ (T) was calculated by coding Equation (1.19) in R. However the method of estimating V (T) by calculating the variance of the 100 estimated values of T performed better than the method uses $\widehat{var}\left[\widehat{var}\ (\hat{\beta}\ )\right]$, to estimate var (T).

The hypothesis H_0: $\sigma_b^2 = 0$ is tested as described in Section 1.2 using the restricted likelihood ratio test defined in Equation (1.29). The two adaptive strategies ADM and ADH are as defined in Section 2. 90% confidence intervals are calculated for the LMM method using the method of Faes et al. (2009). Huber-White confidence intervals are calculated, and the adaptive confidence intervals are calculated as discussed in Section 2.

Tables 1-5 show the ratio of the mean estimated variance of $\hat{\beta}$ and $E\left(\widehat{var}\ (\hat{\beta}\ )\right)/var\ (\hat{\beta}\ )$ using the four strategies of estimation (ADM, ADH, LMM and Huber) with values of ρ equals to 0, 0.01, 0.025, 0.05 and 0.1. In all tables we used $\beta = 0$ and significance level $\alpha = 0.1$ for testing $\sigma_b^2 = 0$. The tables show the non-coverage rates of 90% confidence intervals of β and the average lengths of these confidence intervals. The proportion of samples where H_0: $\sigma_b^2 = 0$ is rejected are also shown.

3.2. Simulation Study of Skew-Normal Data in a Balanced Two-Stage Design

A simulation study is conducted to compare the adaptive and non-adaptive methods for estimating var ($\hat{\beta}$) and associated confidence intervals where data are skew-normally distributed. This study is based on equal sample sizes within PSUs. Data are generated from the intercept only Model (1.4) assuming that b_i has a skew normal with variance $\sigma_b^2 = \frac{\rho}{1-\rho}\ \gamma$ and a skewness parameter $\gamma = 1$, and $e_{ij} \sim N\ (0, 1)$.

These strategies are the linear model strategy, the linear mixed model strategy, the robust Huber-White variance estimator strategy and the two adaptive strategies, the LMM based and the Huber based adaptive strategies. The parameter ρ was varied over a range of values of 0, 0.01, 0.025, 0.05 and 0.1. The number of PSUs, c, is varied over a range of values of 2, 5, 10 and 25 and the PSU sample size is varied over a range of values of 2, 5, 10, 15, 25 and 50.

The hypothesis H_0: $\sigma_b^2 = 0$ is tested as described in Section 1.2 using the restricted likelihood ratio test defined in Equation (1.29). The two adaptive strategies ADM and ADH are as defined in Section 2.2. 90% confidence intervals for β are calculated for the LMM method using the method of Faes et al. (2009). Huber-White confidence intervals for β are calculated, and the adaptive confidence intervals for β are calculated as discussed in Section 2.2.

For skew-normal distribution with one value of ($\gamma = 1$), the results are summarized in Tables 1-5. Here, we assumed that the PSUs have the same number of observations, that is $m_i = m$, for all i=1, 2, . . . , c.

As in Subsection 2, the ratio of the estimated variance to the true variance of $\widehat{\beta}$,$E\left(\widehat{var}\ (\widehat{\beta}\ )\right)/var\ (\widehat{\beta}\ )$ is calculated. The tables also include the non-coverage rates for β as well as the average lengths of the 90% confidence intervals for β. The restricted likelihood ratio test probabilities of rejecting $H_0: \sigma_b^2 = 0$ are included in these tables as well. Four strategies of estimation are included in the tables, ADM, ADH, LMM and Hub.

3.3.Variance Estimation

Based on Tables 1-5 we can conclude that:

1. The variance estimators are generally approximately unbiased, as all ratios are approximately 1. However, there are some exceptions for variance estimator using the LMM strategy, where , variance estimators are tended to be biased as

a) For ρ = 0 when there are 10 or less sample PSUs with all numbers of observations per PSU except when there are 10 sample PSUs with 5 observations per PSU.

b) For ρ = 0.01, when there are 2 and 5 sample PSUs with all numbers of observations per PSU and when there are 10 sample PSUs with 5,10,15 and 25 observations per PSU.

c) For ρ = 0.025, 0.0 when there are 2 and 5 sample PSUs with all numbers of observations per PSU except when there are 5 sample PSUs with 50 observations per PSU. It also tended to be biased when there are 10 sample PSUs with 2 and 50 for ρ = 0.025 and 5 for ρ = 0.05 observations per PSU.

d) For ρ = 0.1, when there are 2 sample PSUs with observations $\leq$ 25 per PSU , when there are 5 sample PSUs with 2 and 5 observations per PSU and when there are 10 sample PSUs with 2 and 25 observations per PSU.

2. The other exception where the ADM and the ADH variance estimators are tended to be biased in the cases:

a) For ρ = 0, when there are 2 and 5 sample PSUs with all numbers of observations per PSU except when there are:

- 5 sample PSUs with 2 and 5 observations per PSU.
- 10 sample PSUs with 15 and 50 observations per PSU.
- 25 sample PSUs with 2 and 5 observations per PSU.

b) For ρ = 0.01, when there are:

- 2 Sample PSUs with observations $\leq$ 25 per PSU.
- 5 sample PSUs with 5 observations per PSU.
- 10 sample PSUs with 5, 10 and 25 observations per PSU.

c) For ρ = 0.025, when there are 2 sample PSUs with 10 or less observations per PSU and when there are 5 sample PSUs with 5 observations per PSU.

d) For ρ = 0.05 when there are 2 sample PSUs with 2 and 5 observations per PSU and when there are 5 sample PSUs with 2 observations per PSU.

e) For ρ = 0.1 when there are 2 sample PSUs with 10 observations per PSU and when there are 5 sample PSUs with 5 observations per PSU.

3.4. Confidence Intervals

1. The tables show that the non-coverage rates of confidence intervals for β are almost around the nominal rate (α=10%) when ρ = 0, for all methods. Also, it close to the nominal rate when ρ = 0.01 and 0.025, when there are 5 or less sample PSUs with 5 or less units per PSU and when c =10 with m= 2. For ρ = 0.05, they are close to the nominal rate when there are 5 or less sample PSUs with 2 observations per PSU. But they are far from the nominal rate (10%) for other values of ρ.
2. The ADH average lengths of confidence intervals for β are almost always shorter than its counterpart based on the Huber–White variance estimation for all values of ρ. When there are 2 and 5 sample PSUs it is very obvious that the ADH average lengths of confidence intervals for β are much shorter than Huber average lengths of confidence intervals for β for all numbers of observations per PSU as follows:
   a) With orders 40-55% for ρ = 0 and 0.01 when there are 2 sample PSUs with all numbers of observations per PSU.
   b) With order of 7-12% for ρ = 0 when there are 5 sample PSUs with all numbers of observations per PSU.
   c) For ρ = 0.01 and 0.1 when there are 5 sample PSUs with orders 5-15%.
   d) For ρ = 0.01, 0.025, 0.05 and 0.1 when there are 10 sample PSUs with orders 2-8%.
   e) With ρ = 0.025 and 0.05 when there are 2 sample PSUs with orders 37-53%.
   f) When there are 5 sample PSUs with orders 7-16%.
   g) For ρ = 0.1 when there are 2 sample PSUs with orders 25-48%.
3. The ADM average lengths of confidence intervals for β are almost always shorter than the LMM average lengths of confidence intervals for β for all values of ρ. When there are 2 and 5 sample PSUs it is very obvious with all numbers of observations per PSU that the ADM average lengths of confidence intervals for β are much shorter than LMM average lengths of confidence intervals for β as:
   a) With orders 7-10% for ρ = 0 when there are 2 sample PSUs with all numbers of observations per PSU.
   b) For ρ = 0 and 0.01, the ADM average lengths are shorter for ρ = 0.01, 0.025, 0.05 and 0.1 when there are 2 sample PSUs with orders 7-12%.
   c) For ρ = 0.025, 0.05 and 0.1 and when there are 5 sample PSUs with orders 2-7%.

4. CONCLUSIONS

Based on results obtained we may conclude the following:
For each value of ρ, the Huber-White variance estimators are unbiased. For all values of ρ, the length of the ADH average lengths of confidence intervals for β are shorter than the Huber-White confidence intervals for β.
When ρ = 0, non-coverage rates are approximately around the nominal rate (α=10%). LMM average lengths of confidence intervals for β are nearly wider than the ADM average lengths of confidence intervals for β regardless the value of ρ.
The ADM, adaptive based on LMM variance estimator for β as alternative, confidence intervals are shorter than the LMM confidence intervals in designs with 5 or less sample PSUs with all average numbers of observations per PSU for all values of intra-class correlation ρ. The ADM confidence intervals are a bit shorter for designs with 5 sample PSUs with all average numbers of observations per PSU for all values of intra-class correlation ρ. The ADM confidence intervals are shorter for designs with number of sample

PSUs, c =10 and m=2 and 50 when ρ = 0 and 0.01, respectively. Also, c = 10 with 10 or more observations per PSU when ρ = 0.025, and with 25 or less observations per PSU when ρ = 0.05 and 0.1. The ADM confidence intervals are shorter for these designs with 2-12%. Otherwise, ADM and LMM confidence intervals performance are approximately the same.

The ADH, adaptive based on Huber-White variance estimator for β as an alternative, confidence intervals are much shorter than the Huber-White confidence intervals in designs with 2 and 5 sample PSUs with, approximately all average numbers of observations per PSU for all values ρ. The ADH confidence intervals are shorter for designs with 10 sample PSUs with m ≤ 25 for ρ =0, 0.05 and 0.1, also with all average numbers of observations per PSU for ρ=0.01 and 0.025. The ADH confidence intervals are shorter for these designs with 5-55%. There were no significant differences, otherwise.

ADH non-coverage rates are smaller than Huber-White non-coverage rates in designs with all sample PSUs with all numbers of observations per PSU when ρ = 0 and ADH non-coverage rates are larger than Huber-White non-coverage rates except in designs with c = 2, 15 and 25 with $\overline{m}$= 3 and c = 5 and 10 with $\overline{m}$ = 3 and 10 when ρ=0.025. ADH non-coverage rates are larger than Huber-White non-coverage rates except in designs with c = 2 and 10 with $\overline{m}$ = 3 when ρ = 0.1

ADM non-coverage rates are larger than LMM non-coverage rates in designs with c = 2 with all numbers of observations per PSU for all values of ρ, with c = 5 with $\overline{m}$ = 3 when ρ = 0, with c = 5 with all numbers of observations per PSU and with $\overline{m}$ = 3 and 25 when ρ = 0.025 and 0.1, respectively. The ADM and ADH confidence intervals are shorter than LMM and Huber-White confidence intervals in designs with c ≤ 5 with all numbers of observations per PSU for all values of ρ. There are no relevant differences, otherwise.

Table 1: Simulation results of inferences about *β*, and testing H_0: $H_0 : \sigma_b^2 = 0$ using RLRT with *ρ* = 0, using skew- normal data with γ= 1 based on balanced samples

		$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$				Non- Coverage of CI for β (%)				Pr (Rej H_0) (%)	Confidence Interval Length			
c	*m*	ADM	ADH	LMM	Hub	ADM	ADH	LMM	Hub	RLRT	ADM	ADH	LMM	Hub
2	2	1.283	1.283	1.502	1.111	8.4	8.4	11.0	10.6	10.5	4.985	3.026	5.453	5.188
2	5	1.169	1.169	1.401	0.934	9.4	9.0	10.0	9.7	5.2	1.194	1.459	1.284	3.065
2	10	1.175	1.175	1.427	0.983	11.0	11.0	10.6	8.9	4.8	0.849	1.018	0.941	2.274
2	15	1.201	1.201	1.457	1.008	10.9	10.9	10.4	8.2	4.4	0.696	0.825	0.767	1.880
2	25	1.222	1.223	1.592	1.106	8.7	8.7	9.1	9.0	6.0	0.558	0.690	0.612	1.470
2	50	1.198	1.198	1.451	0.987	9.4	9.4	10.3	9.5	4.6	0.378	0.454	0.417	1.012
5	2	1.031	1.031	1.147	0.982	10.2	10.1	10.2	9.9	10.3	1.190	1.199	1.212	1.295
5	5	1.071	1.071	1.186	0.959	10.8	10.8	10.9	11.6	9.0	0.725	0.732	0.744	0.796
5	10	1.124	1.124	1.259	1.024	8.8	8.8	9.0	10.1	7.6	0.505	0.510	0.520	0.573
5	15	1.114	1.114	1.216	0.963	9.5	9.5	10.2	12.1	7.7	0.412	0.417	0.417	0.456
5	25	1.146	1.146	1.284	1.039	8.4	8.4	8.8	8.1	6.2	0.311	0.314	0.322	0.358
5	50	1.160	1.160	1.206	1.061	10.4	10.4	9.4	9.3	6.8	0.223	0.225	0.231	0.256
10	2	1.072	1.072	1.156	1.031	8.4	8.4	9.0	8.5	10.5	0.786	0.786	0.792	0.802
10	5	1.012	1.012	1.069	0.937	10.2	10.3	11.2	11.6	7.0	0.487	0.488	0.486	0.500
10	10	1.042	1.042	1.101	0.965	9.4	9.4	9.9	9.9	7.7	0.346	0.346	0.346	0.357
10	15	1.098	1.098	1.159	1.006	9.4	9.4	10.8	11.2	8.0	0.282	0.282	0.281	0.289
10	25	1.057	1.057	1.119	0.984	9.9	9.8	10.7	9.9	8.9	0.218	0.219	0.219	0.227
10	50	1.118	1.118	1.193	1.035	9.2	9.2	9.0	9.2	6.7	0.153	0.153	0.155	0.159
25	2	1.088	1.088	1.136	1.051	10.2	10.1	10.6	10.1	9.4	0.480	0.481	0.478	0.479
25	5	1.088	1.088	1.098	1.030	8.8	8.8	9.9	10.0	9.3	0.304	0.303	0.300	0.302
25	10	1.034	1.034	1.044	0.994	10.0	10.0	11.0	9.9	9.1	0.214	0.214	0.211	0.215
25	15	1.017	1.017	1.024	0.955	9.5	9.5	10.7	10.2	8.6	0.175	0.175	0.172	0.174
25	25	1.068	1.068	1.081	1.025	10.0	10.0	10.2	10.3	9.0	0.135	0.135	0.134	0.136
25	50	1.064	1.064	1.095	1.018	9.0	9.0	9.3	10.0	9.0	0.096	0.096	0.095	0.096

Table 2: Simulation results of inferences about β, and testing $H_0 : \sigma_b^2 = 0$ using RLRT with ρ = 0.01, using skew- normal data with γ= 1 based on balanced samples

		$E(\widehat{var}\ (\widehat{\boldsymbol{\beta}}\ ))/var\ (\widehat{\boldsymbol{\beta}}\ )$				Non- Coverage of CI for β (%)				Pr (Rej H_0) (%)	Confidence Interval Length			
c	*m*	ADM	ADH	LMM	Hub	ADM	ADH	LMM	Hub	RLRT	ADM	ADH	LMM	Hub
2	2	1.186	1.186	1.290	1.016	8.0	8.0	11.7	9.2	10.9	5.287	3.050	5.835	5.103
2	5	1.142	1.142	1.276	0.918	9.9	9.5	11.2	11.2	4.9	1.238	1.474	1.236	3.047
2	10	1.075	1.075	1.285	0.863	12.8	12.8	14.5	12.0	4.9	0.857	1.037	0.933	2.223
2	15	1.153	1.153	1.408	0.985	12.3	12.3	12.6	10.4	4.8	0.699	0.837	0.774	1.900
2	25	1.085	1.085	1.232	0.962	14.2	14.2	15.1	10.8	7.0	0.563	0.721	0.633	1.545
2	50	1.038	1.038	1.275	1.006	19.9	19.9	18.2	8.2	9.1	0.431	0.582	0.492	1.223
5	2	1.008	1.008	1.114	0.939	11.5	11.4	12.1	12.4	10.3	1.188	1.198	1.205	1.277
5	5	1.123	1.123	1.253	1.036	10.7	10.7	12.6	11.4	8.2	0.718	0.723	0.739	0.806
5	10	1.019	1.019	1.155	0.962	12.6	12.6	12.8	11.7	8.9	0.512	0.518	0.535	0.591
5	15	1.060	1.060	1.184	0.986	13.7	13.5	13.5	13.6	9.9	0.421	0.427	0.434	0.484
5	25	1.028	1.028	1.146	0.983	17.1	17.0	15.3	14.2	11.1	0.329	0.334	0.342	0.386
5	50	1.005	1.005	1.124	0.993	21.1	20.7	20.6	17.9	16.0	0.245	0.249	0.257	0.288
10	2	0.984	0.984	1.058	0.944	10.8	10.5	11.4	11.8	10.8	0.785	0.787	0.790	0.802
10	5	1.080	1.080	1.141	1.017	12.4	12.2	12.9	13.9	9.6	0.496	0.497	0.496	0.512
10	10	1.089	1.089	1.156	1.050	12.6	12.4	12.8	12.1	11.2	0.352	0.353	0.356	0.370
10	15	1.019	1.019	1.084	0.978	19.3	19.2	19.3	18.1	12.9	0.291	0.291	0.295	0.305
10	25	1.103	1.103	1.184	1.105	20.2	20.0	20.2	17.2	15.2	0.228	0.228	0.232	0.244
10	50	0.935	0.935	1.008	0.972	32.4	32.1	30.8	27.5	23.5	0.170	0.171	0.176	0.185
25	2	1.007	1.007	1.053	0.979	11.4	11.4	12.6	11.2	10.4	0.480	0.481	0.479	0481
25	5	1.098	1.098	1.106	1.050	16.0	15.9	16.8	16.6	12.7	0.306	0.306	0.302	0.305
25	10	1.033	1.033	1.047	1.021	21.2	21.2	22.0	20.9	15.0	0.219	0.218	0.217	0.222
25	15	0.984	0.984	0.998	0.993	27.3	27.2	28.1	25.8	14.6	0.178	0.178	0.177	0.183
25	25	0.998	0.998	1.014	1.016	37.6	37.6	37.6	36.3	23.7	0.143	0.142	0.142	0.147
25	50	0.982	0.982	1.023	1.030	55.7	56.1	54.1	52.4	39.2	0.107	0.106	0.109	0.111

Table 3: Simulation results of inferences about β, and testing $H_0 : \sigma_b^2 = 0$ using RLRT with ρ = 0.025, using skew- normal data with γ= 1 based on balanced samples

		$E(\widehat{var}\ (\widehat{\boldsymbol{\beta}}\ ))/var\ (\widehat{\boldsymbol{\beta}}\ )$				Non- Coverage of CI for β (%)				Pr (Rej H_0) (%)	Confidence Interval Length			
c	*m*	ADM	ADH	LMM	Hub	ADM	ADH	LMM	Hub	RLRT	ADM	ADH	LMM	Hub
2	2	1.235	1.235	1.462	1.074	7.5	7.5	11.7	10.1	11.2	5.578	3.083	6.097	5.137
2	5	1.134	1.134	1.274	0.996	11.6	11.2	11.2	9.1	7.1	1.223	1.638	1.439	3.324
2	10	1.148	1.148	1.418	1.019	14.4	14.4	14.4	11.0	6.5	0.891	1.129	1.002	2.400
2	15	1.019	1.019	1.226	0.892	16.5	16.5	16.4	12.5	7.3	0.753	0.972	0.831	2.003
2	25	0.998	0.998	1.204	0.946	21.6	21.6	19.4	11.2	9.9	0.625	0.858	0.703	1.702
2	50	1.019	1.019	1.220	1.027	26.7	26.7	22.0	10.6	13.4	0.502	0.733	0.569	1.275
5	2	1.089	1.089	1.206	1.025	10.2	10.1	10.4	10.0	8.8	1.192	1.200	1.220	1.294
5	5	1.060	1.060	1.180	1.018	14.4	14.4	13.4	11.4	11.6	0.751	0.757	0.774	0.857
5	10	1.058	1.058	1.185	1.024	14.9	14.8	16.0	13.2	12.3	0.531	0.539	0.551	0.619
5	15	0.996	0.996	1.135	0.989	21.2	20.9	19.5	16.6	10.9	0.430	0.435	0.452	0.511
5	25	0.964	0.964	1.096	1.014	26.1	25.8	23.4	17.6	19.0	0.360	0.368	0.384	0.438
5	50	0.909	0.909	1.010	0.964	31.6	31.0	29.1	22.5	31.8	0.290	0.298	0.309	0.346
10	2	1.016	1.016	1.088	0.966	11.7	11.7	12.6	12.5	12.0	0.791	0.793	0.796	0.803
10	5	0.961	0.961	1.018	0.933	16.3	16.1	17.1	15.7	11.0	0.501	0.503	0.503	0.526
10	10	0.927	0.927	0.993	0.929	22.7	22.5	23.1	19.9	16.7	0.365	0.366	0.372	0.390
10	15	0.980	0.980	1.062	1.015	27.9	27.7	26.9	23.7	18.0	0.302	0.302	0.311	0.328
10	25	0.888	0.889	0.951	0.930	35.9	35.4	34.5	30.0	30.0	0.250	0.251	0.257	0.272
10	50	1.066	1.066	1.127	1.123	43.7	44.1	40.3	38.2	50.9	0.205	0.206	0.212	0.220
25	2	0.948	0.948	0.991	0.925	17.1	17.3	17.8	17.9	11.6	0.483	0.483	0.481	0.484
25	5	0.966	0.966	0.978	0.960	26.2	26.0	27.8	25.4	13.9	0.310	0.309	0.307	0.315
25	10	0.954	0.954	0.978	0.984	37.7	38.0	38.4	36.2	21.4	0.225	0.224	0.225	0.233
25	15	1.009	1.009	1.013	1.051	47.9	48.3	48.7	46.6	28.8	0.189	0.188	0.190	0.196
25	25	0.990	0.990	1.004	1.031	59.5	60.3	59.5	58.6	47.5	0.158	0.156	0.158	0.162
25	50	0.907	0.907	0.921	0.927	73.3	74.5	73.0	73.9	78.8	0.128	0.128	0.129	0.130

Table 4: Simulation results of inferences about β, and testing $H_0: \sigma_b^2 = 0$ using RLRT with ρ = 0.05, using skew- normal data with γ= 1 based on balanced samples

		$E(\widehat{var}(\widehat{\beta}))/var(\widehat{\beta})$				Non- Coverage of CI for β (%)				Pr (Rej H_0) (%)	Confidence Interval Length			
c	*m*	ADM	ADH	LMM	Hub	ADM	ADH	LMM	Hub	RLRT	ADM	ADH	LMM	Hub
2	2	1.097	1.097	1.284	0.951	9.7	9.7	13.9	9.7	10.6	5.400	3.042	5.884	5.274
2	5	1.179	1.179	1.423	1.037	12.9	12.8	12.6	9.9	7.9	1.238	1.730	1.452	3.421
2	10	1.022	1.022	1.269	0.980	17.7	17.7	14.5	11.1	9.2	0.943	1.295	1.081	2.656
2	15	1.073	1.073	1.229	1.067	20.3	20.3	17.9	10.9	10.4	0.822	1.149	0.946	2.314
2	25	1.035	1.035	1.202	1.004	27.7	27.7	24.8	13.4	15.3	0.748	1.118	0.826	1.923
2	50	0.944	0.944	1.089	0.982	34.1	34.1	27.8	14.4	22.7	0.663	1.064	0.745	1.677
5	2	1.126	1.126	1.250	1.080	10.5	10.5	11.9	11.2	12.2	1.215	1.222	1.239	1.219
5	5	1.022	1.022	1.153	1.011	17.2	17.2	15.6	13.8	13.1	0.763	0.772	0.798	0.882
5	10	1.026	1.026	1.169	1.054	21.7	21.6	20.4	15.9	17.4	0.565	0.574	0.601	0.671
5	15	0.976	0.976	1.093	1.017	28.4	28.0	27.1	21.5	21.1	0.475	0.486	0.501	0.571
5	25	0.970	0.970	1.076	1.028	32.4	31.7	28.1	21.2	32.8	0.419	0.429	0.448	0.496
5	50	0.957	0.957	1.020	1.005	35.6	34.2	33.3	27.4	52.0	0.374	0.383	0.393	0.424
10	2	1.051	1.051	1.134	1.016	13.7	13.7	14.7	13.7	11.1	0.804	0.806	0.812	0.823
10	5	1.055	1.055	1.132	1.065	23.4	23.1	22.8	20.3	15.6	0.517	0.519	0.525	0.552
10	10	0.939	0.939	1.010	0.979	28.6	29.0	27.4	24.9	24.9	0.388	0.388	0.399	0.420
10	15	0.878	0.878	0.943	0.922	38.0	38.0	36.2	33.2	31.4	0.330	0.330	0.341	0.358
10	25	0.995	0.995	1.044	1.039	42.6	42.7	40.9	39.0	52.2	0.296	0.296	0.304	0.315
10	50	0.887	0.887	0.905	0.906	50.2	49.5	49.8	48.1	79.1	0.260	0.258	0.264	0.266
25	2	0.982	0.982	1.032	0.972	23.3	23.3	23.4	23.9	11.4	0.490	0.490	0.491	0.495
25	5	0.990	0.990	1.005	1.008	38.4	38.6	38.3	36.6	22.4	0.321	0.321	0.320	0.330
25	10	1.059	1.059	1.077	1.112	55.2	56.3	55.9	53.1	37.6	0.241	0.240	0.241	0.251
25	15	0.996	0.996	1.011	1.040	62.5	63.7	62.2	62.0	53.7	0.210	0.208	0.211	0.216
25	25	0.991	0.991	0.998	1.008	75.2	75.4	75.0	74.4	81.0	0.185	0.183	0.186	0.186
25	50	1.055	1.055	1.056	1.057	85.5	85.7	85.4	85.7	96.9	0.161	0.160	0.161	0.160

Table 5: Simulation results of inferences about β, and testing $H_0: \sigma_b^2 = 0$ using RLRT with ρ = 0.1, using skew- normal data with γ= 1 based on balanced samples

		$E(\widehat{var}(\widehat{\beta}))/var(\widehat{\beta})$				Non- Coverage of CI for β (%)				Pr (Rej H_0) (%)	Confidence Interval Length			
c	*m*	ADM	ADH	LMM	Hub	ADM	ADH	LMM	Hub	RLRT	ADM	ADH	LMM	Hub
2	2	1.003	1.003	1.209	0.904	12.2	12.2	15.5	12.2	10.9	5.228	3.031	5.829	5.322
2	5	1.064	1.064	1.297	0.996	16.3	16.0	15.3	11.4	10.4	1.421	1.927	1.564	3.724
2	10	1.103	1.103	1.229	1.100	24.0	24.0	20.5	13.2	14.6	1.082	1.677	1.230	3.009
2	15	0.989	0.989	1.142	0.973	28.2	28.0	25.2	12.7	16.6	0.993	1.543	1.102	2.617
2	25	1.052	1.052	1.198	1.097	33.3	33.2	28.0	11.6	24.1	0.965	1.585	1.073	2.468
2	50	0.955	0.956	1.038	0.987	38.7	38.5	31.9	12.7	34.7	0.980	1.627	1.063	2.187
5	2	1.061	1.061	1.179	1.017	14.6	14.5	14.5	13.4	13.3	1.226	1.238	1.264	1.237
5	5	1.102	1.102	1.235	1.114	21.2	21.0	18.5	15.4	18.2	0.814	0.829	0.856	0.946
5	10	0.944	0.944	1.049	0.991	31.8	31.0	29.0	21.4	28.0	0.632	0.647	0.668	0.749
5	15	0.907	0.907	0.996	0.964	35.4	33.9	30.7	24.5	38.0	0.577	0.592	0.612	0.678
5	25	0.949	0.949	1.003	0.988	35.5	33.9	32.4	26.0	55.4	0.543	0.554	0.567	0.606
5	50	0.882	0.883	0.901	0.897	37.6	36.1	36.1	32.7	74.2	0.524	0.527	0.535	0.551
10	2	1.068	1.068	1.156	1.063	18.5	18.3	19.1	19.1	13.9	0.830	0.832	0.842	0.862
10	5	0.975	0.975	1.054	1.025	31.5	32.0	29.0	26.2	24.9	0.555	0.555	0.572	0.600
10	10	0.933	0.933	0.991	0.982	40.2	39.9	39.6	35.2	42.7	0.436	0.438	0.448	0.470
10	15	0.951	0.951	0.995	0.995	43.5	43.9	42.1	40.1	60.0	0.406	0.405	0.419	0.429
10	25	1.059	1.059	1.083	1.085	52.2	51.7	51.0	49.3	76.9	0.373	0.372	0.379	0.385
10	50	1.037	1.037	1.041	1.041	55.7	56.3	55.1	55.6	94.8	0.351	0.348	0.352	0.350
25	2	0.984	0.984	1.036	0.989	36.7	36.3	36.7	36.2	18.5	0.503	0.504	0.507	0.512
25	5	0.956	0.956	0.977	1.008	57.2	57.8	57.2	55.6	37.0	0.343	0.340	0.344	0.356
25	10	0.977	0.977	0.987	1.005	73.3	73.3	73.2	72.3	71.2	0.279	0.276	0.280	0.283
25	15	0.957	0.957	0.961	0.968	80.4	81.2	80.3	80.6	89.0	0.255	0.252	0.255	0.255
25	25	1.023	1.023	1.024	1.025	85.7	85.5	85.7	85.5	98.2	0.231	0.230	0.231	0.230
25	50	1.008	1.009	1.008	1.009	91.9	91.6	91.9	91.6	100.0	0.210	0.209	0.210	0.209

RECEIVED DECEMBER, 2013
REVISED FEBRUARY, 2014

REFERENCES

[1] AL-ZOU'BI, L.M. (2012): Adaptive inference for Multi-Stage unbalanced exponential survey data. **Electrical Journal of Applied Statistical Analy***sis*. Accepted for publication, 2012.
[2] AL-ZOU'BI, L.M., CLARK, R.G., and STEEL, D.G. (2010): Adaptive inference for multi-stage survey data. **Communications in Statistics-Simulation and Computation**, 39, 1334- 1350.
[3] AZZALINI, A. (1985): A class of distributions which includes the normal ones. **Scandinavian Journal of Statistics**, 12, 171-178.
[4] CLARK, R.G., and STEEL, D.G. (2002): The effect of using household as sampling unit. **International Statistical Review,** 70, 289-314.
[5] COCHRAN, W.G. (1977): **Sampling Techniques**. John Wiley and Sons, Inc, , N. York.
[6] DIGGLE , P.J., LIANG, K.Y., and ZEGER, S.L. (1994): **Analysis of Longitudinal Data**. Clarendon Press, New York..
[7] FAES, C., MOLENBERGHS, H., AERTS, M., VERBEKE, G., and KENWARD, M.G. (2009): The effective sample size and alternative small-sample degrees of freedom method. **The American Statistician**, 63, 389- 399.
[8] GOLDSTEIN, H. (2003): **Multilevel Statistical Models**. 3rd ed. Kendall's Library of Statistics 3, Arnold, .London.
[9] HUBER, P.J. (1967): The behavior of maximum likelihood estimates under non-standard conditions. **Proceedings of the Fifth Berekley Symposium on Mathematical Statistics and Probability**, University of California, Berekley, 11, 1967, 221-233.
[10] KISH, L. (1965): **Survey Sampling**. John Wiley and Sons, Inc, , N. York.
[11] LIANG, K.Y., and ZEGER,S.L. (1986): Longitudinal data analysis using generalized linear models. **Biometrika,** 73, 13-22.
[12] LONGFORD, N.T. (2008): An alternative analysis of variance. **Statistics and Operations Research Transactions,** 32, 77-92.
[13] MACKINNON, J.G., and WHITE, H. (1985): Some hetero skedasticity-consistent covariance matrix estimators with improved finite sample properties. **Journal of Econometrics**, 29, 305-325.
[14] PATTERSON, H.D., and THOMPSON, R. (1971): Recovery of inter-block information when block sizes are unequal. **Biometrika**, 58, 545- 554.
[15] PINHEIRO, J.C., and BATES, D.M. (2007): **Mixed-effects Models in S and S-plus**. Springer, New York.
[16] R DEVELOPMENT CORE TEAM. (2007): **R: A Language and Environment for Statistical Computing**.
[17] R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2007.
[18] RAO, P.S.R.S. (1997): **Variance Components Estimation, Mixed Models, Methodologies and Applications**. Chapman and Hall/CRC, Boca Raton, Florida.
[19] STRAM, D.O., and LEE, J.W. (1994): Variance components testing in the longitudinal mixed effects model. **Biometrics,** 50, 1171-1177.
[20] SELF, S.G., and LIANG, K.Y. (1987):Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. **Journal of the American Statistical Association**, 82, 605-610.
[21] VISSCHER, P.M. (2006): A note on the asymptotic distribution of likelihood ratio tests to test variance components. **Twin Research and Human Genetics**, 9, 490- 495.
[22] VERMA, V., SCOTT, C., and O'MUIRCHEARTAIGH, C. (1980): Sample design and sampling errors for the world fertility survey. **Journal of the Royal Statistical Society, Series A**, 143(4), 431–473.
[23] WEST, B.T., WELCH, K.B., and GALECKI, A.T.(2007): **Linear Mixed Model: A Practical Guide Using Statistical Software.** Chapman and Hall/CRC, Boca Raton, Florida.
[24] WHITE, H. (1982): Maximum likelihood estimation of misspecified models. **Econometrica,** 50, 1-25.
[25] YUAN, Z., and YANG, Y. (2005): Combining linear regression models: when and how? **Journal of the American Statistical Association**, 100, 1202- 1214.