



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

University of Wollongong  
**Research Online**

---

Centre for Statistical & Survey Methodology  
Working Paper Series

Faculty of Engineering and Information Sciences

---

2009

# Adaptive Inference for Multi-Stage Survey Data

L. M. Al-Zou'bi

*University of Wollongong, lmaa515@uow.edu.au*

Robert Graham Clark

*University of Wollongong, rclark@uow.edu.au*

D. G. Steel

*University of Wollongong, dsteel@uow.edu.au*

---

## Recommended Citation

Al-Zou'bi, L. M.; Clark, Robert Graham; and Steel, D. G., Adaptive Inference for Multi-Stage Survey Data, Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 01-09, 2009, 34p.  
<http://ro.uow.edu.au/cssmwp/21>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:  
[research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)



***Centre for Statistical and Survey Methodology***

**The University of Wollongong**

**Working Paper**

01-09

Adaptive Inference for Multi-Stage Survey Data

Loai Mohamoud Al-Zou'bi, Robert Graham Clark and David G. Steel

*Copyright © 2008 by the Centre for Statistical & Survey Methodology, UOW. Work in progress, no part of this paper may be reproduced without permission from the Centre.*

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: [anica@uow.edu.au](mailto:anica@uow.edu.au)

# *Adaptive Inference for Multi-Stage Survey Data*

*Loai Mahmoud Al-Zou'bi, Robert Graham Clark and David G. Steel*

*Center for Statistical and Survey Methodology*

*School of Mathematics and Applied Statistics*

*University of Wollongong, NSW, 2522*

email: [lmaa515@uow.edu.au](mailto:lmaa515@uow.edu.au)

## **Abstract**

Two-stage sampling usually leads to higher variances for estimators of means and regression coefficients, because of intra-cluster homogeneity. One way of allowing for clustering in fitting a linear regression model is to use a linear mixed model with two levels. If the estimated intra-cluster correlation is close to zero, it may be acceptable to ignore clustering and use a single level model.

In this paper an adaptive strategy is evaluated 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.

**Key words:** Adaptive estimation, Variance components, cluster sampling, bias, multi-level models, Huber-White standard error

## **1 Introduction**

Two-stage sampling is used in many surveys of social, health, economic and demographic topics. Final population units are grouped into primary sampling units (PSUs). The first

stage of selection is a sample of PSUs and the second stage is a sample of units within selected PSUs. 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, see for example (Goldstein, 2003; Snijders and Bosker, 1999; Cochran, 1977; Kish, 1965).

Two-stage sampling is typically used because

- There is no sampling frame of final units, but a frame of PSUs (e.g. a list of suburbs) is available.
- Cost efficiency; for example it is much cheaper to draw a two-stage sample of 100 students from 10 schools than draw a simple random sample of 100 students, as those students might be dispersed over 100 schools (Snijders, 2001).
- Within-group correlations may be of interest in their own right. For instance, the correlation between values for students in the same school might be of interest.

A complication of two-stage sampling is that values of interest may tend to be more similar for units from the same PSU than for units from different PSUs. If so, this should be reflected in the analysis procedure. One way of doing this is by fitting a multilevel model.

Multilevel models are generalization of regression models. Let  $y_{ij}$  be a dependent variable of interest, and  $x_{ij}$  a vector of covariates for unit  $j$  in PSU  $i$ . The two-level linear mixed model (LMM) (Goldstein, 2003) is given by

$$y_{ij} = \beta'x_{ij} + b_i + e_{ij}, \quad i = 1, 2, \dots, c, \quad j = 1, 2, \dots, m_i \quad (1)$$

where  $c$  denotes the number of PSUs in the sample,  $m_i$  denotes the number of observations selected in PSU  $i$ ,  $\beta$  is the vector of unknown regression coefficients,  $b_i \sim N(0, \sigma_b^2)$  is a PSU specific random effect, and  $e_{ij}$  is assumed to be  $N(0, \sigma_e^2)$ . Therefore  $y_{ij} \sim N(\beta'x_{ij}, \sigma_b^2 + \sigma_e^2)$ , with variance  $\sigma_y^2 = \sigma_b^2 + \sigma_e^2$ . Variances of regression coefficient estimates can be estimated by either standard likelihood theory based on model (1) (West et al., 2007), or by using the robust Huber-White estimator (Huber, 1967; White, 1982). We have assumed that the sampling design is ignorable (Skinner and Marcel, 2004), so that a simple LMM can be applied to the sample. The issues associated with the effect of more complex sampling designs on used models is discussed by Pfeiffermann et al. (1998).

The intra-class correlation ( $\rho$ ) is a measure of the association between the regression residuals for members of the same PSU. It is the ratio of the population variance between PSUs and the total variance. Consequently, it is given by the formula  $\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$  (Kish, 1965).

In practice the intra-class correlation is often quite small. For example, if units within PSUs are no more homogenous than units over all PSUs, then the intra-class correlation is zero. On the other hand, if units from the same PSU have equal values then the intra-class correlation is 1. Generally the intra-class correlation is positive, but in case of equal number of observations in each PSU (Hox, 2002), it is usually less than 0.1 when PSUs are geographic areas and final units are households in these areas (Verma et al., 1980). When PSUs are households and final units are people in households it is between 0 and 0.2 (Clark and Steel, 2002). Values in the range 0.01-0.05 are possible.

There are number of possible approaches for estimating the regression coefficients and their variances when the intraclass correlation ( $\rho$ ) is thought to be small or has been estimated as a small value. One is to fit a linear mixed model regardless. Another is to fit a linear model assuming independent observations. However, if the sample design is relatively clustered, that is a large number of final units are selected from each PSU, the estimated variances resulting from a linear mixed model can be a lot larger those obtained from a linear model assuming independent observations, leading to wider confidence intervals. A third alternative is to use an adaptive strategy based on testing the null hypothesis that the random effect variance component,  $\sigma_b^2$ , is zero. If the null hypothesis is accepted we use the linear model for estimating the variances of the estimated regression coefficients  $\hat{\beta}$ . On the other hand, 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 or the Huber-White method.

This paper is divided into five sections. In Section 2 we will describe the linear mixed model including an outline of the standard likelihood theory estimator of  $\beta$  and  $var(\hat{\beta})$  and the Huber-White estimator of  $var(\hat{\beta})$ . In Section 3 an adaptive strategy will be described. In Section 4 a simulation study of the adaptive and other methods will be described. In Section 5 we will draw conclusions.

## 2 Fitting the Multilevel Model

### 2.1 The Model

Let  $\mathbf{X}$  be the  $m_i \times p$  design matrix, which is assumed to be of rank  $p$ ,  $\mathbf{Y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_c)$  is the complete set of  $n$  observations in the  $c$  PSUs, where  $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})'$  is the observed vector for the  $i^{\text{th}}$  PSU.

Model (1) can also be written as

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V}) \quad (2)$$

where  $\mathbf{V}$  is a block diagonal matrix,  $\mathbf{V} = \text{diag}(\mathbf{V}_i, i = 1, \dots, c)$ , and

$$\mathbf{V}_i = \sigma_b^2 \mathbf{J}_{m_i} + \sigma_e^2 \mathbf{I}_{m_i} \quad (3)$$

$\mathbf{J}_{m_i}$  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.

However, the variance components contained in  $\mathbf{V}_i$  are usually not known. Therefore, they are usually estimated by Restricted Maximum Likelihood (REML) giving the estimate  $\hat{\mathbf{V}}_i$ .

REML was first introduced by Patterson and Thompson (1971) 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  $\mathbf{y}$ , say  $\mathbf{k}'\mathbf{y}$ , where  $\mathbf{k}'$  is chosen so that  $\mathbf{k}'\mathbf{y}$  is free of fixed effects. One of the attractive aspects of REML is that it takes into account the degrees of freedom in estimation of the variance components by the use of the fixed effects part of the model, see for example Verbeke and Molenberghs (2000); Diggle et al. (1994); McCulloch and Searle (2001). There is also no loss of information about the variance components when the inference is derived from

$\mathbf{k}'\mathbf{y}$  rather than  $\mathbf{y}$ . Sahai and Ojeda (2005) presented the REML estimates of the variance components  $\hat{\sigma}_b^2$ , with  $\hat{\sigma}_b^2 \geq 0$ , and  $\hat{\sigma}_e^2$  to be the solutions of the following system of equations

$$\left. \begin{aligned} \frac{n-c}{\hat{\sigma}_e^2} + \sum_{i=1}^c \frac{\hat{\lambda}_i}{m_i} - \frac{\sum_{i=1}^c \frac{\hat{\lambda}_i^2}{m_i}}{\sum_{i=1}^c \hat{\lambda}_i} &= \frac{(n-c)MSE}{\hat{\sigma}_e^4} + \sum_{i=1}^c \frac{\hat{\lambda}_i^2}{m_i} (\bar{y}_i - \hat{\beta})^2 \\ \sum_{i=1}^c \hat{\lambda}_i - \frac{\sum_{i=1}^c \frac{\hat{\lambda}_i^2}{m_i}}{\sum_{i=1}^c \hat{\lambda}_i} &= \sum_{i=1}^c \hat{\lambda}_i^2 (\bar{y}_i - \hat{\beta})^2 \\ \hat{\beta} &= (\sum_{i=1}^c \mathbf{x}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i)^{-1} \sum_{i=1}^c \mathbf{x}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{y}_i \end{aligned} \right\} \quad (4)$$

where

$$\begin{aligned} MSA &= \frac{1}{(c-1)} \sum_{i=1}^c m_i (\bar{y}_i - \bar{y}_{..})^2 \\ MSE &= \frac{1}{n-c} \sum_{i=1}^c \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i.)^2 \\ \hat{\lambda}_i &= \frac{m_i}{\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2} \end{aligned}$$

There is no explicit solution for this system of equations in general. While in the balanced data case ( $m_i = m$  for all  $i$ ) (Sahai and Ojeda, 2004), these estimates reduced to

$$\left. \begin{aligned} \hat{\sigma}_e^2 &= \min(MSE, \frac{n-c}{n-1} MSE + \frac{c-1}{n-1} MSA) \\ \hat{\sigma}_b^2 &= \frac{1}{m} \max(MSA - MSE, 0) \end{aligned} \right\} \quad (5)$$

The simple special case of model specified by 1 where the model includes just a grand mean parameter will be used in the simulation study in Section 4. This model is given by defining  $x_{ij}$  to be 1 for all  $i, j$

$$y_{ij} = \beta + b_i + e_{ij}, \quad i = 1, 2, \dots, c \quad ; \quad j = 1, 2, \dots, m_i \quad (6)$$

In this case,  $\hat{\beta}$  becomes

$$\hat{\beta} = \frac{\sum_{i=1}^c \hat{\lambda}_i \bar{y}_i}{\sum_{i=1}^c \hat{\lambda}_i} \quad (7)$$

This case is important in practice since often two stage surveys focus on estimation of means.

In the balanced case this reduces to

$$\hat{\beta} = \bar{y}_{..} \quad (8)$$



## 2.2 Likelihood Theory Estimator of $\text{var}(\hat{\beta})$

In this section, we will discuss the variances of the estimated regression coefficients and their estimators. The estimated variance of  $\hat{\beta}$  given in equation (4) is given by

$$\widehat{\text{var}}(\hat{\beta}) = \left\{ \sum_{i=1}^c \mathbf{x}_i \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i \right\}^{-1} \quad (9)$$

where  $\hat{\mathbf{V}}_i = \hat{\sigma}_b^2 \mathbf{J}_{m_i} + \hat{\sigma}_e^2 \mathbf{I}_{m_i}$ . For the special case (6) of an intercept-only model, this simplifies to

$$\widehat{\text{var}}(\hat{\beta}) = \left\{ \sum_{i=1}^c \frac{m_i}{\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2} \right\}^{-1} \quad (10)$$

In the balanced data case, the variance estimator simplifies further to

$$\widehat{\text{var}}(\hat{\beta}) = \frac{1}{c} \left[ \hat{\sigma}_b^2 + \frac{\hat{\sigma}_e^2}{m} \right] \quad (11)$$

The  $(1 - \alpha)100\%CI$  confidence interval for  $\hat{\beta}$  is given by

$$\hat{\beta} \pm t_{(df, 1-\frac{\alpha}{2})} \hat{se}(\hat{\beta}) \quad (12)$$

where  $\hat{se}(\hat{\beta}) = (\widehat{\text{var}}(\hat{\beta}))^{\frac{1}{2}}$  and the degrees of freedom (df) defined by Faes et al. (2004) as

$$df_{LMM} = n_e - 1$$

where  $n_e$  is the effective sample size, which is defined by Kish (1965) as the ratio of the sample size and the design effect ( $def f(\hat{\beta})$ ). Kish (1965) also defined the design effect for  $\hat{\beta} = \bar{y}$  by  $def f(\hat{\beta}) = 1 + (m - 1)\rho$ , where  $m$  is the average number of observations per PSU and  $\rho$  is the intraclass correlation. There is some debate over the appropriate degrees of freedom in (12). The degrees of freedom defined here are not exact, other approaches have

been suggested by Ruppert et al. (2003). For large samples this is a minor concern as the degrees of freedom will be large.

### 2.3 Huber-White Estimator

The estimator  $\widehat{var}(\hat{\beta})$  in (9) will be approximately unbiased provided the variance model (3) is correct. If this is not the case,  $\widehat{var}(\hat{\beta})$  will be biased and inference will be incorrect. An alternative to ML or REML estimates of  $var(\hat{\beta})$  is the robust variance estimate approach described by Liang and Zeger (1986), in the context of modeling longitudinal data using generalized estimating equations (GEE). This approach can be applied to the analysis of data collected using PSUs, where observations within PSUs might be correlated and the observations in different PSUs are independent.

This approach can be referred to as robust or Huber-White variance estimation (Freedman, 2006). This approach will be used as an alternative approach to estimating  $var(\hat{\beta})$  in this paper. The method yields asymptotically consistent covariance matrix estimates even if the variances and covariances assumed in model specified by 2 are incorrect. It is still necessary to assume that the observations from different PSUs are independent.

Equation 4 defines  $\hat{\beta}$  for the model specified in 2 and 3, the variance of this estimator is given by equation 9. An alternative estimator of  $\mathbf{V}_i$  is  $\hat{\mathbf{V}}_i^{Hub} = \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i'$ , where  $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{x}_i \hat{\beta}$ .  $\hat{\mathbf{V}}_i^{Hub}$  is approximately unbiased for  $\mathbf{V}_i$  even if (3) does not apply.

$$\begin{aligned}
 E(\hat{\mathbf{V}}_i^{Hub}) &= E(\hat{\mathbf{e}}_i \hat{\mathbf{e}}_i') \\
 &\approx E[(\mathbf{y}_i - \mathbf{x}_i \beta)(\mathbf{y}_i - \mathbf{x}_i \beta)'] \\
 &= \mathbf{V}_i
 \end{aligned} \tag{13}$$

Note that

$$\begin{aligned} \text{var}(\hat{\beta}) &= \text{var}\left(\left(\sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i\right)^{-1} \sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{y}_i\right) \\ &\approx \left(\sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{V}_i^{-1} \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i\right)^{-1} \end{aligned} \quad (14)$$

One way to make  $\widehat{\text{var}}(\hat{\beta})$  robust to misspecification of the variance model is to substitute the robust estimator  $\hat{\mathbf{V}}_i^{\text{Hub}}$  in (14) as follows

$$\widehat{\widehat{\text{var}}}(\hat{\beta}) = \left(\sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} (\hat{\mathbf{V}}_i^{\text{Hub}})^{-1} \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i\right) \left(\sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i\right)^{-1} \quad (15)$$

When there is only an intercept in the model ( $x_i = 1$ ), (15) becomes

$$\widehat{\widehat{\text{var}}}(\hat{\beta}) = \frac{\sum_{i=1}^c \hat{\lambda}_i^2 (\bar{y}_i - \hat{\beta})^2}{\left(\sum_{i=1}^c \hat{\lambda}_i\right)^2} \quad (16)$$

### Derivation of 16 from 15

As  $\hat{\mathbf{V}}_i = \hat{\sigma}_e^2 \mathbf{I}_{m_i} + \hat{\sigma}_b^2 \mathbf{J}_{m_i}$ , therefore

$$\begin{aligned} \sum_{i=1}^c \mathbf{x}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{x}_i &= \sum_{i=1}^c \mathbf{1}'_{m_i} \left[ \frac{1}{\hat{\sigma}_e^2} (\mathbf{I}_{m_i}) - \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} \mathbf{J}_{m_i} \right] \mathbf{1}_{m_i} \\ &= \sum_{i=1}^c \mathbf{1}'_{m_i} \left[ \frac{1}{\hat{\sigma}_e^2} (\mathbf{I}_{m_i}) - \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} (\mathbf{1}_{m_i} \mathbf{1}'_{m_i}) \right] \mathbf{1}_{m_i} \\ &= \sum_{i=1}^c \left[ \frac{1}{\hat{\sigma}_e^2} (\mathbf{1}'_{m_i} \mathbf{I}_{m_i} \mathbf{1}_{m_i}) - \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} \mathbf{1}'_{m_i} (\mathbf{1}_{m_i} \mathbf{1}'_{m_i}) \mathbf{1}_{m_i} \right] \end{aligned}$$

But  $\mathbf{1}_{m_i} \mathbf{I}_{m_i} = \mathbf{1}_{m_i}$ ,  $\mathbf{I}_{m_i} \mathbf{1}'_{m_i} = \mathbf{1}'_{m_i}$  and  $\mathbf{1}'_{m_i} \mathbf{1}_{m_i} = m_i$

$$\begin{aligned}
\therefore \sum_{i=1}^c \mathbf{x}'_i \hat{V}_i^{-1} \mathbf{x}_i &= \sum_{i=1}^c \left[ \frac{1}{\hat{\sigma}_e^2} (\mathbf{1}'_{m_i} \mathbf{1}_{m_i}) - \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} \mathbf{1}'_{m_i} \mathbf{1}_{m_i} \mathbf{1}'_{m_i} \mathbf{1}_{m_i} \right] \\
\sum_{i=1}^c \mathbf{x}'_i \hat{V}_i^{-1} \mathbf{x}_i &= \sum_{i=1}^c \left[ \frac{m_i}{\hat{\sigma}_e^2} - \frac{m_i^2 \hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} \right] \\
&\dots \\
&= \sum_{i=1}^c \left[ \frac{m_i}{\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2} \right] \\
&= \sum_{i=1}^c \hat{\lambda}_i \\
\therefore \left( \sum_{i=1}^c \mathbf{x}'_i \hat{V}_i^{-1} \mathbf{x}_i \right)^{-1} &= \left( \sum_{i=1}^c \hat{\lambda}_i \right)^{-1} \tag{17}
\end{aligned}$$

$$\begin{aligned}
\therefore \sum_{i=1}^c \mathbf{x}'_i \hat{V}_i^{-1} \hat{\mathbf{e}}_i \hat{\mathbf{e}}'_i \hat{V}_i^{-1} \mathbf{x}_i &= \sum_{i=1}^c \mathbf{1}'_{m_i} \left\{ \frac{1}{\hat{\sigma}_e^2} (\mathbf{I}_{m_i}) - \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} (\mathbf{J}_{m_i}) \right\} \hat{\mathbf{e}}_i \hat{\mathbf{e}}'_i \\
&\times \left\{ \frac{1}{\hat{\sigma}_e^2} (\mathbf{I}_{m_i}) - \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} (\mathbf{J}_{m_i}) \right\} \mathbf{1}_{m_i} \\
&= \sum_{i=1}^c \mathbf{1}'_{m_i} \left\{ \frac{1}{\hat{\sigma}_e^2} (\mathbf{I}_{m_i}) - \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} (\mathbf{1}_{m_i} \mathbf{1}'_{m_i}) \right\} \hat{\mathbf{e}}_i \hat{\mathbf{e}}'_i \\
&\times \left\{ \frac{1}{\hat{\sigma}_e^2} (\mathbf{I}_{m_i}) - \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} (\mathbf{1}_{m_i} \mathbf{1}'_{m_i}) \right\} \mathbf{1}_{m_i} \\
&\dots \\
&= \sum_{i=1}^c \left\{ \frac{1}{\hat{\sigma}_e^2} (\mathbf{1}'_{m_i}) - \frac{m_i \hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} (\mathbf{1}'_{m_i}) \right\} \hat{\mathbf{e}}_i \hat{\mathbf{e}}'_i \\
&\times \left\{ \frac{1}{\hat{\sigma}_e^2} (\mathbf{1}_{m_i}) - \frac{m_i \hat{\sigma}_b^2}{\hat{\sigma}_e^2 (\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2)} (\mathbf{1}_{m_i}) \right\} \\
&\dots \\
&= \sum_{i=1}^c \left\{ \frac{1}{\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2} \right\}^2 \mathbf{1}'_{m_i} \hat{\mathbf{e}}_i \hat{\mathbf{e}}'_i \mathbf{1}_{m_i} \tag{18}
\end{aligned}$$

But

$$\begin{aligned}
\mathbf{1}'_{m_i} \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \mathbf{1}_{m_i} &= \mathbf{1}'_{m_i} (\mathbf{y}_i - \mathbf{x}_i \hat{\beta}) (\mathbf{y}_i - \mathbf{x}_i \hat{\beta})' \mathbf{1}_{m_i} \\
&= (\mathbf{1}'_{m_i} \mathbf{y}_i - \mathbf{1}'_{m_i} \mathbf{1}_{m_i} \hat{\beta}) (\mathbf{y}_i' \mathbf{1}_{m_i} - \mathbf{1}'_{m_i} \mathbf{1}_{m_i} \hat{\beta}) \\
&\quad \dots \\
&= (m_i \bar{y}_i - m_i \hat{\beta})^2 \\
&= m_i^2 (\bar{y}_i - \hat{\beta})^2
\end{aligned} \tag{19}$$

Therefore

$$\begin{aligned}
\left[ \sum_{i=1}^c \mathbf{x}_i' \hat{V}_i^{-1} \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \hat{V}_i^{-1} \mathbf{x}_i \right] &= \sum_{i=1}^c \left\{ \frac{1}{\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2} \right\}^2 m_i^2 (y_{ij} - \hat{\beta})^2 \\
&= \sum_{i=1}^c \left\{ \frac{m_i}{\hat{\sigma}_e^2 + m_i \hat{\sigma}_b^2} \right\}^2 (y_{ij} - \hat{\beta})^2 \\
&= \sum_{i=1}^c \hat{\lambda}_i^2 (y_{ij} - \hat{\beta})^2
\end{aligned} \tag{20}$$

Therefore, from (20) and (17) the estimated Huber-White variance of  $(\hat{\beta})$  is given by

$$\begin{aligned}
\widehat{\text{var}}(\hat{\beta}) &= \left( \sum_{i=1}^c \hat{\lambda}_i \right)^{-2} \sum_{i=1}^c \hat{\lambda}_i^2 (y_{ij} - \hat{\beta})^2 \\
&= \frac{\sum_{i=1}^c \hat{\lambda}_i^2 (y_{ij} - \hat{\beta})^2}{\left( \sum_{i=1}^c \hat{\lambda}_i \right)^2}
\end{aligned} \tag{21}$$

where

$$\hat{\beta} = \frac{\sum_{i=1}^c \hat{\lambda}_i \bar{y}_i}{\sum_{i=1}^c \hat{\lambda}_i} \tag{22}$$

In the balanced data case, (i.e.  $m_i = m$ ), from equation (8) and since  $\hat{\lambda}_i$  is constant this estimator becomes

$$\widehat{\widehat{var}}(\hat{\beta}) = \frac{1}{c(c-1)} \sum_{i=1}^c (\bar{y}_i - \bar{y}_{..})^2 \quad (23)$$

### Derivation of 23 from 21

In this case  $\lambda_i = \lambda$  for all  $i$ , therefore

$$\begin{aligned} \widehat{\widehat{var}}(\hat{\beta}) &= \widehat{\widehat{var}}(\bar{y}_{..}) \\ &= \frac{1}{c(c-1)} \sum_{i=1}^c (\bar{y}_i - \bar{y}_{..})^2 \end{aligned} \quad (24)$$

## 2.4 REML Likelihood Ratio Test (RLRT) For Testing $H_0 : \sigma_b^2 = 0$

Suppose we want to test  $H_0 : \sigma_b^2 = 0$  vs.  $H_0 : \sigma_b^2 > 0$ , the REML estimators can be used to derive the likelihood ratio test (LRT) statistic for this test.

The problem of testing  $H_0 : \sigma_b^2 = 0$  using the likelihood ratio test for the large-sample is discussed by Self and Liang (1987) and Stram and Lee (1994). Under  $H_0$  the true parameter value is on the boundary of the parameter space, therefore the likelihood ratio test statistic has a distribution that is essentially the mixtures of  $\chi^2$  distributions under nonstandard conditions assuming that response variables are *iid* (Self and Liang, 1987). This assumption does not generally hold in linear mixed models. Stram and Lee (1994) used Self and Liang (1987) results to prove that the asymptotic distribution of the likelihood ratio test for testing  $H_0 : \sigma_b^2 = 0$  has an asymptotic 50:50 mixture of  $\chi^2$  with 0 and 1 degrees of freedom under  $H_0$  rather than the classical single  $\chi^2$  if the data are *iid* under the null and alternative hypotheses.

The asymptotic distribution of LRT under the null hypothesis is a 50:50 mixture of  $\chi^2$  (Freedman, 2006; Stram and Lee, 1994). As we are usually dealing with the unbalanced nested design then the asymptotic distribution of the LRT statistic under the null when testing a single variance component, i.e. when testing  $H_0 : \sigma_b^2 = 0$  vs  $H_0 : \sigma_b^2 > 0$ , is  $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$  (Chernoff, 1954). The regression parameters were estimated over the parameter space  $-\infty < \beta < \infty$ ,  $0 \leq \sigma_e^2 < \infty$  and  $0 \leq \sigma_b^2 < \infty$ , but  $\hat{\sigma}_b^2$  may equal to zero.

In our case Visscher (2006) showed that the REML-based likelihood ratio test (RLRT) is given by

$$-2\log(LRT) = (n-1)\log\left(\frac{n-c}{n-1} + \frac{c-1}{n-1}F\right) - (c-1)\log(F) \quad (25)$$

where  $F = \frac{MSA}{MSE}$  and  $MSA$  and  $MSE$  are defined in (5). Formula (25) is working for all values of  $F$  and that the value of  $-2\log(LRT)$  is zero if  $F \leq 1$ . Equation (25) has been used to test  $H_0 : \sigma_b^2 = 0$  against  $H_A : \sigma_b^2 > 0$  as well, which is used to define the adaptive strategies.

It is desired in many problems to test the hypothesis that the parameter lies in part of the parameter space  $\Omega$  of dimension  $p$ , say  $\Omega_0$ , versus the alternative that this parameter lies in the complement of  $\Omega_0$ , say  $\Omega_1$ . Suppose  $\Omega_0 \subset \Omega$  is of dimension  $r$  and the parameter is an interior point of  $\Omega$  but it lies on the boundary of  $\Omega_0$  and  $\Omega_1$ , therefore subject to regularity conditions  $-2\ln LRT$  follows the  $\chi^2$  distribution with  $p-r$  degrees of freedom. Chernoff (1954) derived the distribution of the likelihood ratio test when  $\Omega_0$  and  $\Omega_1$  have the same dimension as  $\Omega$ , with the parameter is an interior point in  $\Omega$  and lies on the boundary of  $\Omega_0$  and  $\Omega_1$ . Unfortunately these regularity conditions are not satisfied in our case as our

parameter lies on the boundary of the parameter space.

### 3 An Adaptive Strategy

In this paper, we consider two adaptive strategies. Both of them rely on the idea of testing the variance component  $\sigma_b^2$  in model (1). If we reject  $H_0 : \sigma_b^2 = 0$ , we use the first adaptive strategy which is utilizing the LMM-REML estimators of  $var(\hat{\beta})$  defined in equations (10) and (11) in unbalanced and balanced data cases, respectively. On the other hand, if we accept  $H_0$  the LM-REML estimator defined in equations (10) and (11) is employed but  $\sigma_b^2 = 0$  in this case which is equivalent to the standard linear model with independent errors.

This strategy is explained in Figure 1 below.

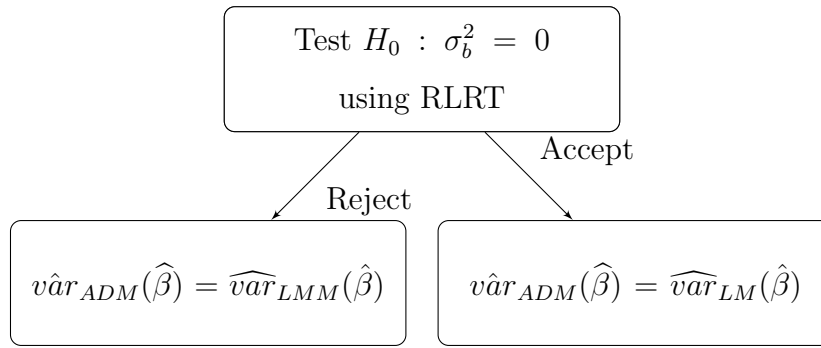


Figure 1: Flowchart explaining the adaptive procedure using the estimated variance extracted from the LMM

where  $\widehat{var}_{LM}(\hat{\beta})$  is the estimator of  $var_{LM}(\hat{\beta})$  using the LM strategy,  $\widehat{var}_{LMM}(\hat{\beta})$  is the estimator of  $var_{LMM}(\hat{\beta})$  using the LMM strategy and  $\widehat{var}_{ADM}(\hat{\beta})$  is the adaptive estimator.

The second adaptive strategy, explained in Figure 2, is identical, except that  $\widehat{var}_{Hub}(\hat{\beta})$  is used instead of  $\widehat{var}_{LMM}(\hat{\beta})$  when  $H_0$  is rejected.



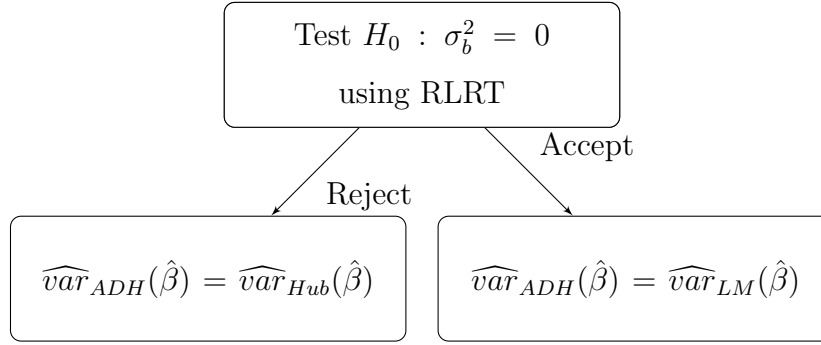


Figure 2: Flowchart explaining the adaptive procedure using Huber-White estimator

where  $\widehat{var}_{Hub}(\hat{\beta})$  is the estimator of  $var_{Hub}(\hat{\beta})$  using the Huber-White strategy.

The benefit of the adaptive strategy is that we use the simple linear model to derive variance estimators, unless there is strong evidence that  $H_0 : \sigma_b^2 > 0$ . This has benefit of simplifying the model and may also give tighter confidence intervals. However, it is not clear whether the adaptive approaches will give valid confidence intervals for  $\beta$ , because the confidence intervals assume non-adaptive procedures.

## 4 Simulation Study

A simulation study was conducted to compare the adaptive and non-adaptive methods for estimating  $var(\hat{\beta})$ . Data were generated from model specified by 6, with  $m_i = m$  and an intercept only model, the values of  $\rho$ ,  $m$  and  $c$  were varied. 1000 samples were generated in each case.

This study is divided into three parts. The first was the estimation of the regression coefficients  $\beta$  and the random effects variance component  $\sigma_b^2$  as well as  $var(\hat{\beta})$ . The estimated

regression coefficients  $\hat{\beta}$  and the estimators of  $var(\hat{\beta})$  were calculated for the LMM and LM models using the *lme* and *lm* packages (Pinheiro and Bates, 2000) in the R statistical environment (R Development Core Team, 2007).

The hypothesis  $H_0 : \sigma_b^2 = 0$  was tested as described in Section 2.4. The two adaptive strategies (ADM) and (ADH) are defined as

$$\widehat{var}_{ADM}(\hat{\beta}) = \begin{cases} \widehat{var}_{LMM}(\hat{\beta}) & \text{if } H_0 \text{ is not retained} \\ \widehat{var}_{LM}(\hat{\beta}) & \text{if } H_0 \text{ is retained} \end{cases} \quad (26)$$

$$\widehat{var}_{ADH}(\hat{\beta}) = \begin{cases} \widehat{var}_{Hub}(\hat{\beta}) & \text{if } H_0 \text{ is not retained} \\ \widehat{var}_{LM}(\hat{\beta}) & \text{if } H_0 \text{ is retained} \end{cases} \quad (27)$$

90% confidence intervals for  $\hat{\beta}$  are given by

$$(1 - \alpha)100\%CI = \hat{\beta} \pm t_{(df, 1 - \frac{\alpha}{2})} \hat{SE}(\hat{\beta}) \quad (28)$$

where  $\alpha = 0.1$  and the degrees of freedom (df) are defined to be:

$$df = \begin{cases} n - 1 & , \text{ using LM Est.} \\ n_e - 1 & , \text{ using LMM Est.} \\ c - 1 & , \text{ using Huber-White Est.} \end{cases} \quad (29)$$

Degrees of freedom for the first and second adaptive strategies (ADM) and (ADH) are defined as

$$df_{ADM} = \begin{cases} n - 1 & , \text{ if } H_0 \text{ accepted} \\ n_e - 1 & , \text{ if } H_0 \text{ rejected} \end{cases} \quad (30)$$

$$df_{ADH} = \begin{cases} c - 1 & , \text{ if } H_0 \text{ accepted} \\ n_e - 1 & , \text{ if } H_0 \text{ rejected} \end{cases} \quad (31)$$

Tables 1 - 4 show the ratio of the mean estimated variance of  $\hat{\beta}$  using the four strategies of estimation (ADM, ADH, LMM and Huber) to the true variance with  $\rho$  values 0, 0.025,

0.05 and 0.1. In all tables we used  $\beta = 0$  and  $\alpha = 0.1$ . They include the non-coverage probabilities for testing  $H_0 : \beta = 0$  and the lengths of the confidence intervals of  $\beta$  as well as the probability that  $H_0 : \sigma_b^2 = 0$  is rejected. Results on non-coverage and confidence interval length are shown in graphical form in Figures 3 - 12. In these graphs we also include the LM strategy of estimation so that the effect of completely ignoring the clustered nature of the data can be examined.

The variance estimators are approximately unbiased as all ratios are approximately 1. The only exception is the variance estimator using the LMM strategy, it tends to be biased when we have 2 PSUs with all numbers of observations per PSU for all values of  $\rho$ . It, also tends to be biased when we have 5 PSUs with 10 or fewer observations per PSU in case of  $\rho = 0, 0.025$  and  $0.05$ . In case of  $\rho = 0.1$  it tends to be biased at  $m=2$  and  $5$  and  $c = 5$ .

Non-coverage for  $\beta$  was close to the nominal rate of 10% when  $\rho = 0$  for all methods.

For  $\rho \neq 0$ , Huber non-coverage was close to 10% in all cases. The LMM non-coverage was close to 10% in most cases. However, when  $\rho$  was large (0.05 or 0.1) and  $m$  was large (10 or more) and  $c$  was small (2 or 5), the LMM non-coverage was much larger. This may be because of the difficulty in determining the appropriate degrees of freedom.

Table 1: Variance ratios, length and non-coverage of the 90% confidence intervals for  $\beta$ , and power of testing  $H_0 : \sigma_b^2 = 0$  with  $\rho=0$ .

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$						Non-Coverage of CI for $\beta$				Pr(reject $H_0$ )		Confidence Interval Length			
		ADM	ADH	LMH	Hub	TADM	TADH	TLMH	THub	Lrt	CADM	CADH	CLMM	CHub			
2	2	1.148	1.148	1.365	0.990	8.4	8.6	7.0	10.4	10.2	2.722	2.837	3.290	5.014			
2	5	1.268	1.268	1.509	1.044	8.9	9.1	7.3	10.2	6.2	1.280	1.540	1.431	3.161			
2	10	1.205	1.205	1.463	1.013	9.5	9.5	8.1	9.3	5.5	0.830	1.045	0.923	2.266			
2	25	1.371	1.371	1.652	1.177	8.5	8.7	7.6	8.4	6.5	0.522	0.714	0.574	1.505			
2	50	1.298	1.298	1.570	1.070	8.7	8.7	7.3	9.2	4.9	0.356	0.462	0.390	1.013			
5	2	1.073	1.073	1.183	0.981	10.0	10.0	8.5	9.8	10.3	1.183	1.192	1.259	1.266			
5	5	1.098	1.098	1.211	1.001	7.5	7.4	6.7	8.0	8.0	0.708	0.723	0.746	0.805			
5	10	1.179	1.179	1.298	1.063	8.2	8.3	7.6	9.5	7.8	0.496	0.510	0.521	0.572			
5	25	1.052	1.052	1.168	0.943	10.0	9.9	8.7	11.2	7.3	0.310	0.319	0.326	0.360			
5	50	1.048	1.048	1.161	0.924	9.6	9.5	8.3	10.6	5.8	0.217	0.222	0.227	0.250			
10	2	1.038	1.038	1.115	0.984	8.3	8.2	7.3	9.1	8.7	0.779	0.781	0.810	0.793			
10	5	1.145	1.145	1.206	1.087	7.8	7.8	7.5	8.4	8.8	0.487	0.492	0.500	0.509			
10	10	1.142	1.142	1.209	1.048	9.8	9.7	9.0	8.8	8.7	0.344	0.347	0.353	0.356			
10	25	0.995	0.995	1.046	0.904	10.3	10.2	9.8	10.2	6.7	0.215	0.217	0.220	0.222			
10	50	1.038	1.038	1.089	0.932	10.0	9.8	9.4	11.2	7.0	0.152	0.153	0.155	0.156			
25	2	1.085	1.085	1.133	1.047	9.2	9.0	8.6	9.5	9.2	0.479	0.479	0.490	0.478			
25	5	1.052	1.052	1.061	1.015	9.9	9.9	9.7	11.3	9.9	0.303	0.304	0.304	0.305			
25	10	0.995	0.995	1.003	0.938	10.3	10.3	10.3	10.9	7.8	0.212	0.213	0.213	0.212			
25	25	1.012	1.012	1.021	0.941	10.1	10.0	10.1	10.4	7.5	0.134	0.135	0.135	0.133			
25	50	1.039	1.039	1.066	0.968	8.9	8.8	8.5	9.9	7.9	0.095	0.095	0.096	0.095			

Table 2: Variance ratios, length and non-coverage of the 90% confidence intervals for  $\beta$ , and power of testing  $H_0 : \sigma_b^2 = 0$  with  $\rho=0.01$ .

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$						Non-Coverage of CI for $\beta$						Pr(reject $H_0$ )		Confidence Interval Length			
		ADM	ADH	LMH	Hub	TADM	TADH	TLMH	THub	Lrt	CADM	CADH	CLMM	CHub					
2	2	1.141	1.141	1.353	0.987	8.3	8.2	7.2	10.4	10.3	2.752	2.870	3.316	5.050					
2	5	1.240	1.240	1.478	1.041	10.0	9.9	7.7	8.9	7.0	1.307	1.605	1.464	3.250					
2	10	1.163	1.163	1.394	1.000	10.1	10.1	8.5	9.2	6.9	0.852	1.126	0.942	2.348					
2	25	1.273	1.273	1.528	1.179	11.2	11.2	9.1	8.6	8.2	0.546	0.808	0.605	1.663					
2	50	1.107	1.107	1.339	1.081	13.8	13.8	11.4	9.3	10.3	0.392	0.634	0.436	1.265					
5	2	1.063	1.063	1.172	0.974	9.4	9.6	8.5	10.4	10.9	1.191	1.201	1.268	1.275					
5	5	1.090	1.090	1.197	1.010	9.5	9.4	8.6	8.9	9.4	0.717	0.735	0.755	0.823					
5	10	1.147	1.147	1.277	1.086	9.7	9.7	8.3	9.5	10.9	0.509	0.529	0.538	0.603					
5	25	0.953	0.953	1.075	0.946	12.7	12.5	10.7	10.4	13.2	0.326	0.344	0.347	0.402					
5	50	0.911	0.911	1.021	0.940	14.4	13.9	12.2	11.1	21.0	0.241	0.262	0.257	0.308					
10	2	1.037	1.037	1.115	0.991	8.5	8.5	8.0	9.2	9.5	0.785	0.788	0.817	0.803					
10	5	1.088	1.088	1.163	1.067	9.0	8.9	8.0	9.5	9.3	0.491	0.496	0.508	0.522					
10	10	1.071	1.071	1.143	1.034	9.8	9.6	9.4	9.5	11.6	0.350	0.355	0.362	0.373					
10	25	0.923	0.923	0.995	0.939	11.6	11.0	10.4	10.9	18.8	0.228	0.234	0.237	0.252					
10	50	0.839	0.839	0.903	0.877	15.3	15.1	13.4	11.6	29.0	0.170	0.177	0.177	0.192					
25	2	1.070	1.070	1.120	1.040	9.7	9.7	8.3	8.6	9.4	0.481	0.482	0.493	0.482					
25	5	1.028	1.028	1.044	1.017	10.4	10.4	10.2	10.3	13.5	0.307	0.308	0.309	0.313					
25	10	0.927	0.927	0.941	0.922	11.5	11.5	11.5	11.4	14.5	0.217	0.219	0.219	0.223					
25	25	0.916	0.916	0.933	0.947	12.5	12.3	12.4	10.7	24.8	0.142	0.143	0.143	0.149					
25	50	0.964	0.964	1.000	1.010	11.1	10.6	10.3	8.9	48.7	0.109	0.111	0.111	0.116					

Table 3: Variance ratios, length and non-coverage of the 90% confidence intervals for  $\beta$ , and power of testing  $H_0 : \sigma_b^2 = 0$  with  $\rho=0.025$ .

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\beta)$						Non-Coverage of CI for $\beta$				Pr(reject $H_0$ )		Confidence Interval Length			
		ADM	ADH	LMH	Hub	TADM	TADH	TLMH	THub	Lrt	CADM	CADH	CLMH	CHub			
2	2	1.128	1.128	1.336	0.985	8.4	8.6	7.1	9.2	10.8	2.799	2.928	3.356	5.127			
2	5	1.187	1.188	1.429	1.035	10.9	10.9	8.8	8.3	7.6	1.335	1.666	1.507	3.370			
2	10	1.080	1.080	1.319	0.987	12.6	12.6	10.5	9.5	7.8	0.876	1.199	0.982	2.488			
2	25	1.181	1.181	1.399	1.158	15.4	15.5	12.2	8.7	12.5	0.591	0.998	0.655	1.867			
2	50	1.024	1.024	1.204	1.062	20.5	20.5	16.5	10.4	18.8	0.454	0.921	0.507	1.563			
5	2	1.048	1.048	1.160	0.972	8.6	8.8	7.5	11.1	11.2	1.201	1.211	1.281	1.293			
5	5	1.056	1.056	1.174	1.013	10.6	10.4	8.7	9.6	10.3	0.727	0.748	0.771	0.850			
5	10	1.091	1.091	1.238	1.095	10.8	10.7	8.7	9.9	14.8	0.526	0.554	0.563	0.644			
5	25	0.919	0.919	1.021	0.957	15.6	15.3	13.3	11.4	25.8	0.360	0.397	0.383	0.461			
5	50	0.892	0.892	0.965	0.942	17.9	16.0	15.6	10.3	42.6	0.288	0.335	0.304	0.380			
10	2	1.031	1.031	1.111	0.996	9.3	9.4	8.3	9.2	10.4	0.793	0.796	0.827	0.815			
10	5	1.042	1.042	1.123	1.052	9.7	9.5	9.1	8.9	12.3	0.500	0.506	0.520	0.540			
10	10	1.017	1.017	1.089	1.028	11.0	10.4	10.0	9.3	19.2	0.364	0.373	0.378	0.399			
10	25	0.905	0.905	0.970	0.958	12.9	12.2	11.8	10.9	38.5	0.254	0.267	0.265	0.289			
10	50	0.837	0.837	0.870	0.869	16.6	15.2	14.9	12.1	62.0	0.208	0.225	0.214	0.236			
25	2	1.057	1.057	1.108	1.037	9.7	9.7	8.3	9.1	11.1	0.486	0.487	0.498	0.489			
25	5	1.009	1.009	1.025	1.020	10.7	10.6	10.2	9.7	20.4	0.315	0.317	0.317	0.325			
25	10	0.891	0.891	0.905	0.917	12.4	12.1	11.9	11.1	29.2	0.228	0.230	0.230	0.239			
25	25	0.921	0.921	0.936	0.957	13.1	12.4	13.0	11.2	61.3	0.161	0.166	0.163	0.171			
25	50	1.022	1.023	1.029	1.032	10.4	9.2	9.9	8.4	91.1	0.137	0.142	0.138	0.143			

Table 4: Variance ratios, length and non-coverage of the 90% confidence intervals for  $\beta$ , and power of testing  $H_0 : \sigma_b^2 = 0$  with  $\rho=0.05$ .

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\hat{\beta})$						Non-Coverage of CI for $\beta$						Pr(reject $H_0$ )		Confidence Interval Length			
		ADM	ADH	LMH	Hub	TADM	TADH	TLMH	THub	Lrt	CADM	CADH	CLMM	CHub					
2	2	1.103	1.103	1.309	0.982	8.5	9.4	7.3	9.5	11.3	2.864	3.002	3.426	5.256					
2	5	1.102	1.102	1.355	1.026	12.5	12.3	9.6	8.6	8.0	1.373	1.731	1.577	3.561					
2	10	1.008	1.008	1.227	0.971	15.9	15.9	13.0	9.9	10.7	0.935	1.392	1.052	2.716					
2	25	1.127	1.127	1.285	1.132	19.8	19.6	16.8	10.0	19.8	0.679	1.347	0.744	2.179					
2	50	1.016	1.016	1.117	1.044	25.8	25.3	20.7	10.0	31.2	0.569	1.406	0.616	1.967					
5	2	1.030	1.030	1.142	0.970	9.2	9.2	7.8	10.4	12.3	1.221	1.232	1.304	1.323					
5	5	1.007	1.007	1.141	1.014	12.1	12.0	9.9	10.1	12.7	0.748	0.773	0.801	0.897					
5	10	1.059	1.060	1.191	1.101	13.6	13.3	10.6	9.3	23.8	0.566	0.612	0.606	0.709					
5	25	0.913	0.913	0.996	0.969	18.9	17.9	15.5	11.0	42.0	0.419	0.483	0.445	0.549					
5	50	0.907	0.907	0.947	0.942	20.0	16.3	17.8	10.4	63.8	0.365	0.444	0.379	0.480					
10	2	1.022	1.022	1.103	1.001	9.4	9.5	8.3	9.4	12.4	0.808	0.812	0.843	0.837					
10	5	0.995	0.995	1.084	1.035	11.4	11.2	10.2	10.1	18.9	0.520	0.529	0.543	0.571					
10	10	0.972	0.972	1.045	1.024	13.0	12.5	10.6	9.1	31.2	0.391	0.406	0.408	0.440					
10	25	0.940	0.940	0.971	0.972	14.8	13.4	13.5	10.8	66.7	0.306	0.330	0.313	0.345					
10	50	0.858	0.858	0.868	0.868	16.5	13.3	15.7	11.9	87.2	0.267	0.294	0.270	0.299					
25	2	1.040	1.040	1.093	1.034	9.9	10.0	8.9	8.9	14.2	0.495	0.496	0.508	0.501					
25	5	0.987	0.987	1.007	1.024	10.7	10.4	10.3	9.2	31.0	0.329	0.332	0.332	0.344					
25	10	0.875	0.876	0.891	0.915	12.7	12.4	12.4	10.2	51.5	0.249	0.254	0.251	0.264					
25	25	0.958	0.958	0.961	0.966	12.5	11.8	12.3	11.2	91.6	0.196	0.202	0.196	0.204					
25	50	1.046	1.046	1.046	1.046	8.9	8.4	8.9	8.4	99.9	0.174	0.181	0.174	0.181					

Table 5: Variance ratios, length and non-coverage of the 90% confidence intervals for  $\beta$ , and power of testing  $H_0 : \sigma_b^2 = 0$  with  $\rho=0.1$ .

PSUs	Obs	$E(\widehat{var}(\hat{\beta}))/var(\beta)$						Non-Coverage of CI for $\beta$				Pr(reject $H_0$ )		Confidence Interval Length			
		ADM	ADH	LMH	Hub	TADM	TADH	TLMH	THub	Lrt	CADM	CADH	CLMM	CHub			
2	2	1.048	1.048	1.262	0.977	9.6	10.7	7.8	8.9	11.0	2.946	3.082	3.585	5.517			
2	5	1.000	1.000	1.249	1.010	15.3	15.4	12.4	8.8	10.8	1.494	1.984	1.732	3.949			
2	10	0.929	0.929	1.112	0.953	20.4	20.3	15.9	10.1	15.9	1.069	1.790	1.204	3.184			
2	25	1.062	1.062	1.180	1.103	26.6	26.2	21.5	9.6	27.9	0.847	1.895	0.924	2.747			
2	50	1.014	1.015	1.058	1.027	28.9	27.2	24.6	10.5	45.2	0.787	2.181	0.825	2.627			
5	2	0.998	0.998	1.114	0.970	10.5	10.7	8.8	10.9	14.8	1.264	1.278	1.354	1.388			
5	5	0.974	0.974	1.092	1.014	13.3	13.1	10.8	9.2	21.7	0.810	0.855	0.868	0.990			
5	10	1.048	1.048	1.142	1.103	16.0	15.6	13.4	8.8	39.8	0.658	0.738	0.696	0.833			
5	25	0.958	0.958	0.989	0.983	18.9	15.0	16.4	8.9	67.4	0.550	0.662	0.565	0.703			
5	50	0.930	0.931	0.941	0.940	19.2	13.7	18.1	11.4	85.3	0.505	0.629	0.511	0.644			
10	2	1.004	1.004	1.087	1.009	10.3	10.2	8.6	8.9	16.2	0.839	0.844	0.878	0.880			
10	5	0.958	0.958	1.030	1.014	12.9	12.3	12.0	9.9	33.6	0.568	0.586	0.592	0.632			
10	10	0.973	0.973	1.021	1.021	13.4	12.2	11.2	8.6	56.4	0.460	0.487	0.474	0.518			
10	25	0.971	0.972	0.982	0.982	14.5	12.0	13.4	10.4	88.9	0.398	0.435	0.402	0.442			
10	50	0.872	0.872	0.872	0.873	14.9	11.6	14.9	11.6	98.0	0.365	0.401	0.365	0.402			
25	2	1.013	1.013	1.070	1.031	10.9	10.9	10.1	9.9	20.4	0.514	0.516	0.529	0.527			
25	5	0.992	0.992	1.007	1.030	11.0	10.9	10.8	9.8	56.7	0.364	0.370	0.367	0.382			
25	10	0.903	0.903	0.907	0.916	12.1	11.3	11.8	10.3	86.0	0.298	0.307	0.299	0.311			
25	25	0.976	0.976	0.976	0.976	11.9	10.8	11.9	10.8	99.9	0.253	0.261	0.253	0.261			
25	50	1.056	1.056	1.056	1.056	9.9	9.0	9.9	9.0	100.0	0.235	0.243	0.235	0.243			



Figure 3: confidence interval non-coverage using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0$

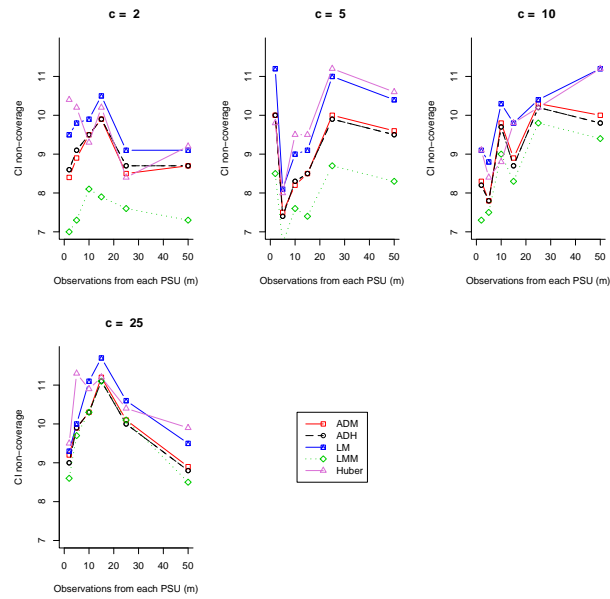


Figure 4: confidence interval non-coverage using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0.01$

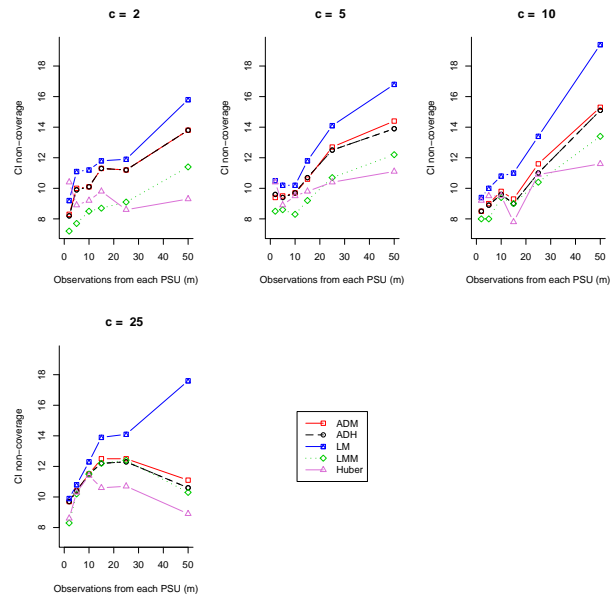


Figure 5: confidence interval non-coverage using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0.025$

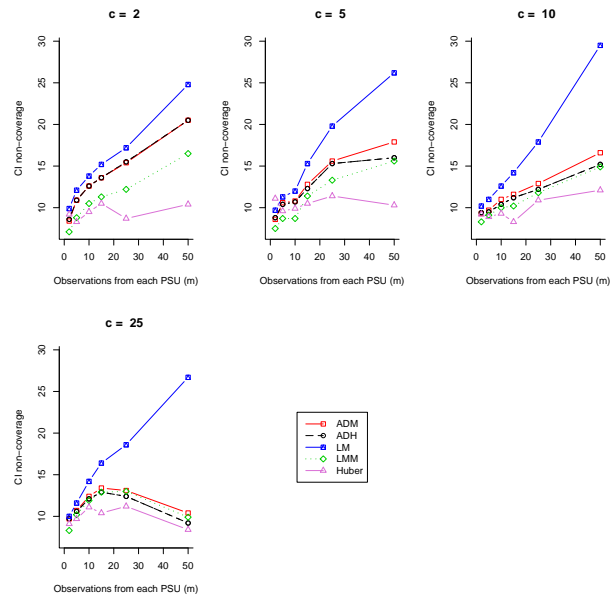


Figure 6: confidence interval non-coverage using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0.05$

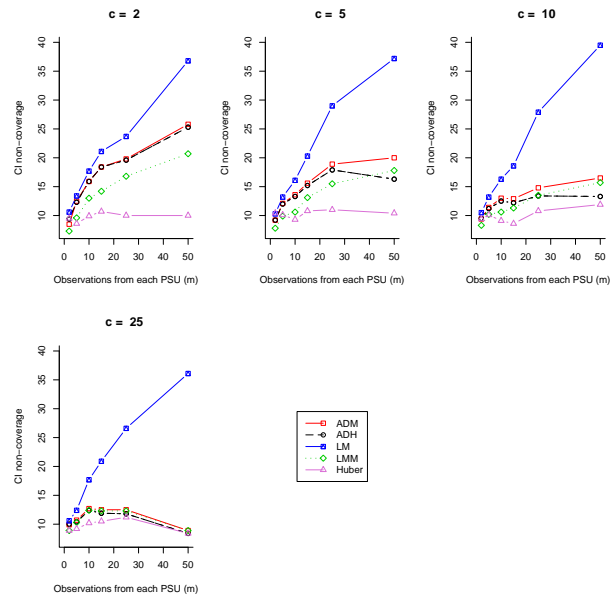


Figure 7: confidence interval non-coverage using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0.1$

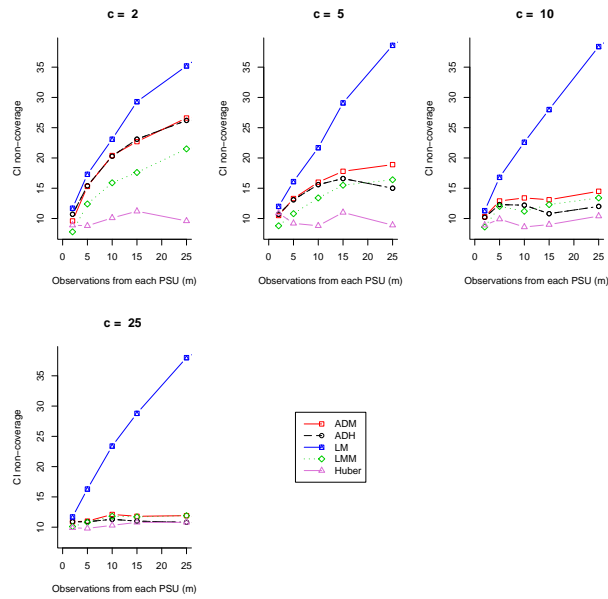


Figure 8: confidence interval lengths using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0$

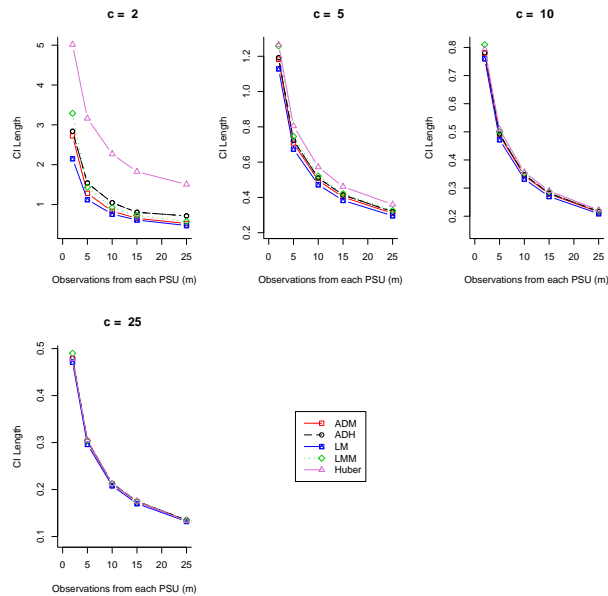


Figure 9: confidence interval lengths using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0.01$

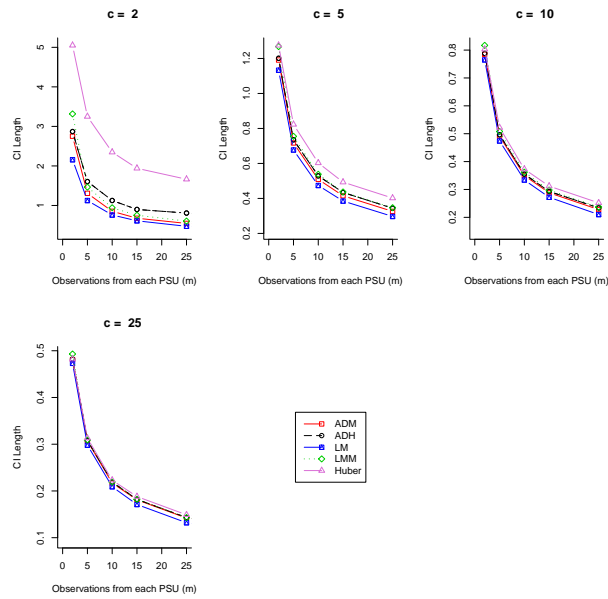


Figure 10: confidence interval lengths using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0.025$

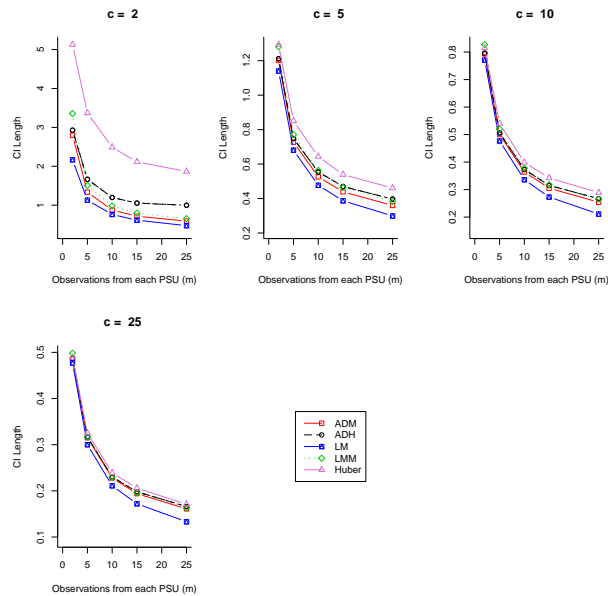


Figure 11: confidence interval lengths using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0.05$

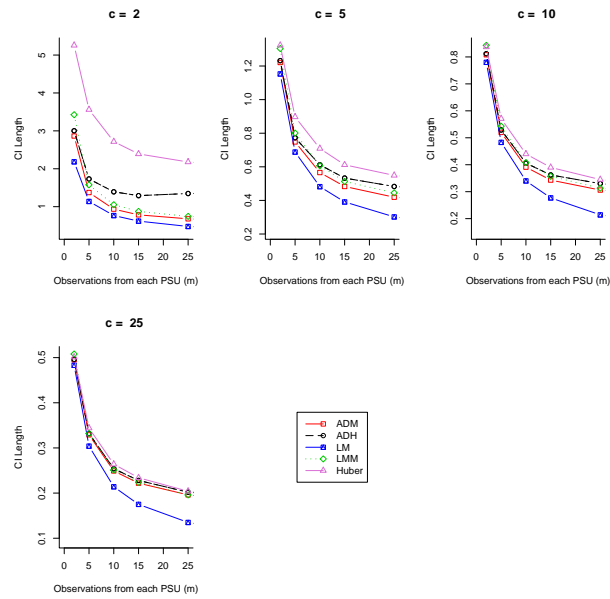
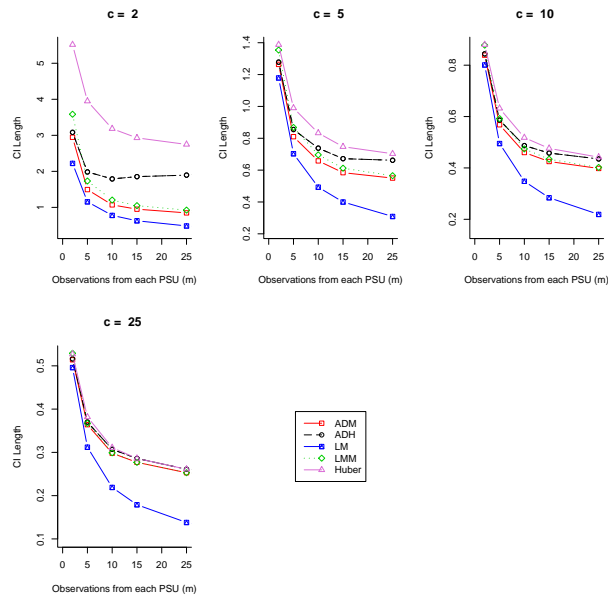


Figure 12: confidence interval lengths using different variance estimation methods and for various values of  $m$  and  $c$ ,  $\rho=0.1$



We used the effective sample size as suggested by Faes et al. (2004). We also tried using the sample size and the approach of Ruppert et al. (2003), but results were even worse.

Figure 3 shows that LM non-coverage was close to 10% when  $\rho = 0$ . It was very high otherwise as shown by Figures 4 - 7. Hence, use of LMM without at least checking  $H_0 : \sigma_b^2 = 0$  is not a strategy that should never be used.

Figure 8-12 show that confidence intervals using the LM strategy are the shortest, however this strategy is not viable because of its high non-coverage when  $\rho \neq 0$ . The Huber based approach gives the widest in general. The ADM and ADH confidence intervals are almost always shorter than the LMM and Huber ones, respectively. When there were 2 and 5 PSUs it is very clear that ADM and ADH are much shorter than LMM and Hub, respectively, for all values of  $\rho \neq 0$ . In case where there were 25 PSUs these lengths become closer. For example:

- for  $c = 2$  and  $m = 10$  with  $\rho = 0.05$  the ADM and ADH confidence intervals lengths are 0.935 and 1.392, respectively, while these lengths were 1.052 and 2.716 for LMM and Huber, respectively;
- in the case of  $c = 25$  and  $m = 5$  and  $\rho = 0.025$  ADM and LMM confidence intervals lengths are 0.315 and 0.317, respectively, while the ADH and Huber are 0.317 and 0.325, respectively.
- in the case of  $c = 10$  and  $m = 50$  and  $\rho = 0.1$  the ADM and ADH confidence intervals lengths are 0.365 and 0.401, respectively, while the LMM and Huber are 0.365 and 0.402, respectively.

## 5 Conclusion

1. Designs with few clusters and large sample sizes in each cluster appear to be non-robust to intra-cluster correlation. In these designs, even a small intraclass correlation will substantially inflate the variance of the mean, however the cluster-level variance component is unlikely to be significant even if the intraclass correlation is as high as 0.1. As a result, when the number of clusters ( $c$ ) is 2 or 5, and the number of observations per cluster ( $m$ ) is 25, both of the adaptive estimators have higher than desirable non-coverage, of the order of 15%. It appears that for these extreme designs, clustering must be allowed for in variance estimates, even if the clustering is not statistically significant.
2. In all other designs, the adaptive methods are reasonably reliable, with non-coverage fairly close to the nominal 10%.
3. In comparing the Linear Mixed Model (LMM) with the adaptive version (ADM), we find that:

The LMM tends to be too conservative with (non-coverage less than 10%) except for the extreme designs mentioned in 1. This is presumably due to the difficulty in defining the appropriate degrees of freedom for this method. In contrast, ADM has narrower confidence intervals and has non-coverage closer to the nominal 10%.

The ADM confidence intervals are noticeably narrower for  $c$  equal to 2 and 5, but there is not much to choose between ADM and LMM for  $c=25$ .

4. In comparing the robust Huber-White approach with the adaptive version (ADH), we find that:

Both the Huber and ADH approaches have non-coverage close to the nominal 10% except in the extreme designs mentioned in 1.

The Huber method gives wide confidence intervals when  $c$  is small (2 or 5) even though the non-coverage is close to the nominal 10%. This is because the degrees of freedom for this method is equal to  $(c-1)$ . ADH has much narrower confidence intervals, because its degrees of freedom are equal to  $(n-1)$  rather than  $(c-1)$  if the cluster-level variance component is not significant.

5. This leads to the following recommendations:

Designs with fewer than 10 clusters, and a large sample size in each cluster should be avoided, even if the intra-cluster correlation is believed to be low.

Provided this advice is followed, clustering can be ignored if the cluster-level variance effect is insignificant. This gives close to correct coverage, while giving shorter confidence intervals (at least slightly).

6. Future research will focus on whether different criteria (other than significance for the cluster-level variance component) give better adaptive confidence intervals, and on unbalanced designs and non-normal data.



## References

- Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25(3):573–578.
- Clark, R. G. and Steel, D. G. (2002). The effect of using household as a sampling unit. *International Statistical Review*, 70(2):289–314.
- Cochran, W. G. (1977). *Sampling Techniques*. John Wiley and Sons, Chichester.
- Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford University Press, Oxford.
- Faes, C., Molenberghs, H., Aerts, M., Verbeke, G., and Kenward, M. G. (2004). The effective sample size and a novel small sample degrees of freedom method. Technical Report , 0651, Interuniversity Attraction Pole.
- Freedman, D. A. (2006). On the so-called “Huber sandwich estimator” and “robust standard errors”. *The American Statistician*, 60:299–302.
- Goldstein, H. (2003). *Multilevel Statistical Models*. Kendall’s Library of Statistics 3. Arnold, London, London, third edition.
- Hox, J. (2002). *Multilevel Analysis: Techniques and Application*. Laurence Erlbaum Associates, Mahwah, New Jersey.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard

- conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California, Berkeley*, 11:221–233.
- Kish, L. (1965). *Survey Sampling*. John Wiley and Sons, Chichester.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.
- McCulloch, C. and Searle, S. (2001). *Generalized, Linear, and Mixed Models*. John Wiley and Sons, Chichester.
- Patterson, H. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554.
- Pfeffermann, D., Skinner, C. J., Holmes, D., Goldstein, H., and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of Royal Statistical Society B*, 60:23–56.
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects Models in S and S-plus*. Statistics and Computing. Springer, New York.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, New York.

- Sahai, H. and Ojeda, M. M. (2004). *Analysis of Variance for Random Models Volume 1: Balanced Data, Theory, Methods, Applications and Data analysis*. Birkhauser, Boston.
- Sahai, H. and Ojeda, M. M. (2005). *Analysis of Variance for Random Models Volume 2: Unbalanced Data, Theory, Methods, Applications and Data analysis*. Birkhauser, Boston.
- 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(398):605–610.
- Skinner, C. J. and Marcel, D. T. V. (2004). The impact of complex sampling designs in the analysis of longitudinal survey data. *CJ*, 7:2040–2045.
- Snijders, T. A. B. (2001). *Multilevel Modelling of Health Statistics*, chapter 11, pages 159–174. John Wiley and Sons, Chichester.
- Snijders, T. A. B. and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. SAGE publication.
- Stram, D. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50:1171–1177.
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- 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*, 143:431–473.

Visscher, P. M. (2006). A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Research and Human Genetics*, 9(4):490–495.

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/CRC, Boca Raton, Florida.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25.