# Estimation in Mixed Poisson Regression Models

Remy Julius van de Ven

A thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy

School of Mathematics and Statistics
The University of Sydney

January 1996

## Statement of Authorship

This thesis contains no material which has been accepted as part of the requirements for any other degree or diploma at any university. To the best of my knowledge this thesis contains no material previously published, submitted for publication or written by any other individual other than myself, unless due reference to such is made in this thesis.

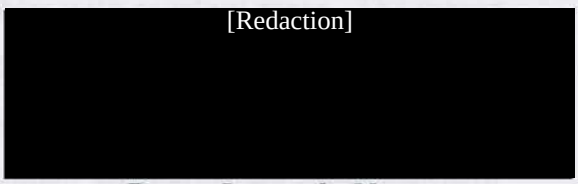
To date four papers out of this thesis have been, or are about to be, published.

van de Ven, R. (1993). Estimating the shape parameter for the negative binomial distribution. *J. Statist. Comput. Simul.* 46, 111-123.

van de Ven, R. and Weber, N.C. (1993). Bounds for the median of the negative binomial distribution. *Metrika* 40, 185-189.

van de Ven, R. and Weber, N.C. (1995). Log-linear models for mean and dispersion in mixed Poisson regression models. *Austral. J. Statist.* 37, 205-216.

van de Ven, R. and Weber, N.C. (1996). Percentile related bounds and estimator for the mean of the negative binomial distribution. *Statistics* 27, 345-352.




Remy J. van de Ven

Submitted: January 1996

Edited: April 1996

## Abstract

This thesis considers estimation of the parameters associated with models for count data displaying over-dispersion relative to the Poisson distribution where the over-dispersion is modelled using mixing. It is divided into seven chapters with Chapters Two to Five specific to the over-dispersed Poisson problem whilst Chapter Seven, which uses results from Chapter Six, is more general. The motivation for some of this work was the modelling of repeat counts of the number of fibres contained on microscopic slides as obtained by asbestos fibre counters and the subsequent estimation of mean fibre concentrations and counter variability.

Chapter One introduces the above mentioned asbestos fibre problem and follows this with an overview of the thesis. In Chapter Two a model for repeated measures count data over-dispersed relative to the Poisson distribution appropriate to the asbestos problem is given. To accommodate the over-dispersion a Poisson random variable is compounded with a positive random variable with mean equal one and variance linked linearly, via a log function, to a set of covariates. Maximum likelihood estimators of the parameters are obtained for the case where the compounding distribution is gamma and extended quasi-likelihood parameter estimators are obtained when the compounding distribution is unspecified. These two sets of parameter estimators are then shown to be comparable in certain circumstances.

In Chapter Three a special case of the general model in Chapter Two with a gamma compounding distribution is considered. Here repeat counts for a "subject" are taken as independent Poisson random variables with constant mean. The means are then modelled as independent observations from a gamma distribution. Two sets of moment estimators for the parameters of the model are obtained and generalized variances of the limiting distribution of the moment estimators are compared with the corresponding quantity for the maximum likelihood estimators. Also in this chapter we derive asymptotic results that explain some of the erratic behaviour of the moment estimators.

Chapter Four considers the estimation of the shape parameter of the negative

binomial distribution (NBD), this distribution being a special case of the model in Chapter Three. Here the results are given for a simulation study comparing four estimators for the shape parameter of the NBD distribution. Two criteria are used to compare the estimates obtained in the simulations, one being the traditional moment based criterion whilst the other is based on a new measure termed the "percentile measure". This measure, based on the difference between the percentiles of the true and estimated distribution function, is argued to be more appropriate in many cases.

In Chapter Five we continue studying the NBD and obtain some quantile related results. First we obtain bounds for the median in terms of the mean that are improvements on the bounds obtained by Payton, Young and Young (1989). Second we obtain percentile related bounds for the mean and use this to obtain a robust estimator for the mean of the NBD when the shape parameter is known.

The remaining two chapters are devoted to robust estimation in (generalized) linear mixed models. In Chapter Six a modification to the Fellner (1991) procedure for robustly estimating variance components in normal linear mixed models is proposed and studied. Also given is a robust moment based method. These robust methods are then applied in Chapter Seven to the generalized linear mixed model to obtain robust parameter estimators and the behaviour of these new estimators is studied via a simulation study.

From this simulation study in Chapter Seven it is concluded that the extension to generalized linear mixed models of the modification to Fellner's method has merit. There should though be scope for improvement in the method and this could be a subject for further research. In particular, a possible mechanism for achieving an improvement would be to have more robust starting values for the variance components in the iterative procedure proposed. One solution would be to develop quantile based variance components estimates in the generalized linear mixed model and to use these as starting values.

Another project for further research would be to obtain expressions for the variances of the fixed effects estimates for the linear mixed model obtained using the Fellner (1991) method. This would necessarily be an asymptotic result and

of interest in its own right. However, once this was available the modified Fellner method of Chapter Six and its extension to generalized linear mixed models given in Chapter Seven could be improved. This is the case as the modification to Fellner's method given in Chapter Six currently uses for these values the variances of the BLUP estimates of the fixed effects.

Finally, an alternative robust estimation procedure based on the results of Windham (1995) could be another subject for further research. That paper, which assumes the data are independent and identically distributed (*iid*), proposes an estimation procedure that weights datum according to the value of the estimated likelihood at that datum point. The procedure seems to have merit based on the examples considered in Windham's paper, which include some skewed distributions (e.g. gamma). Further research could consider extending the results to the non *iid* case, in particular data from generalized linear mixed model problems. It may be possible using such a procedure in generalized linear mixed model cases to reduce some of the bias that is inherent in procedures based on Winsorization.

## Acknowledgements

As with any large project it is rarely if ever the efforts of a single person and this thesis is no exception. Without the support and assistance in various forms from many people it is doubtful if this thesis would ever have been completed.

Foremost amongst these I would like to thank my supervisor Dr. Neville Weber whose contribution has been critical. Not only has Dr. Weber been a valuable source of inspiration and moral support, he has also patiently and carefully read through many drafts of the chapters in this thesis.

I would also like to thank the University of Sydney for providing me with employment throughout the duration of the project. This has been most beneficial as it has provided the ideal environment to undertake my research. Thanks also go to the staff of the School of Mathematics and Statistics who, without exception, have been only too willing to assist with problems, be they mathematical, computing or associated with library resources.

Finally, I would like to thank the unknown referees of papers that I have submitted for publication as their comments have contributed to improvements in the presentation of certain results.

# Table of contents

# Chapter 1

# Introduction

The purpose of this Introduction is twofold. Firstly we outline in some detail the problem that motivated the research project that lead to this thesis, that being the statistical analysis of counts of asbestos fibres. Secondly, we give an overview of the thesis together with some details of the chronological development of the work. This chronological perspective is important because some of the work in the early chapters has to a certain extent been superceded by the work in later chapters. Here the critical factor causing the separation was the publication of the paper by Schall (1991).

## 1.1 Asbestos fibre counting problem

The potential health consequences of asbestos fibres, primarily due to inhalation, have been long established and well documented; see for example Anon (1978), Anon (1984). Here the main consequence is cancer which may appear in a number of forms. One form is lung cancer which develops from asbestosis, a progressively disabling pulmonary fibrosis. Another is mesothelioma, a tumor made up of cells from the pleura or peritoneum. In addition there exists evidence to suggest links between asbestos exposure and other forms of cancer, e.g. cancer of the digestive tract.

As a consequence health risk authorities in a number of countries, including

Australia, have implemented procedures to monitor the asbestos fibre concentration in the workplace. These procedures, in short, require the sampling of air from the workplace environment, the extraction of the fibres from the sampled air and their deposition on a filter, the mounting of the filter on a microscopic slide and finally the determination of the asbestos fibre concentration on the filter. In order to make the procedure reproducible reference methods have been developed. These though do not completely eliminate the variability in the procedure, particularly in the sampling and counting stages.

For the counting stage the current reference method is based on the procedure documented in Anon (1988). Basically this method requires that the number of fibres on 100 randomly chosen fields of a filter be counted, counting as a half any fibre falling on the boundary of the field. Justification for the half modification can be based on the results of Hall (1985) wherein it is shown that under certain assumptions an estimate of the fibre loading (intensity) based on the half modification has smaller variance than an estimate based on the count of fibre centers (assuming the centers could be observed).

Now each randomly chosen field has an area less than 0.0001 of the total filter. Hence, if the fibre centers were uniformly distributed over the filter, the number of fibre centers on each field would be approximately independent and identically distributed Poisson random variables. However, counting as halves fibres on the boundary and rounding will give "under-dispersed Poisson" counts. This "under-dispersion" will be compensated for to some extent by the tendency for some clumping of fibres on the filter, a process leading to "over-dispersion" (see McCullagh and Nelder (1989) p. 198). Results from empirical studies (e.g. Ogden (1982)) tend to support the claim that repeat counts on a slide, obtained by the same counter, are approximately Poisson.

Counts from different counters however show significant "over-dispersion" relative to the Poisson distribution. This is due primarily to the between-counter variability which is not unexpected given that the counting rules are detailed and contain a degree of subjectivity. For example, according to Anon (1988), §8.5.1,

*A countable fibre is defined as any object having a maximum width less than 3 micrometres and a length greater than 5 micrometres and a length/width ratio greater than 3:1, and which does not appear to touch any particle with a maximum dimension greater than 3 micrometers.*

It is therefore not difficult to envisage situations where two counters would differ as to whether a particular object should be counted. The situation becomes more complicated when definitions are given for counting split fibres and for grouped fibres. However, with experience and participation in quality assurance programs (QAPs) the between-counter variabily decreases; see Ogden (1982).

It was therefore decided that the National Association of Testing Authorities, Australia (NATA) would establish a National Asbestos Program within Australia to monitor the performance of Australian asbestos fibre counters. For this program, established in the late 1980s, NATA obtained a library of slides containing various asbestos fibre types and a range of nominal levels of asbestos fibre concentrations. Some of these were real environmental slides whilst others were artifically prepared. These latter slides were prepared in batches of 14 and were supposed to have nominally identical fibre concentrations. The slides were then circulated among the counters, masking a slide's identity each time it was sent to a counter.

One of objectives of the trial was thus to monitor the performance of the counters. To this end it was necessary to obtain estimates of the "true" fibre loadings of slides and estimates of the between-counter variabilty. Here the between-counter variabilty may depend on the a number of covariates such as counter experience. Estimates of these parameters could then be used to establish criteria to identify those counters not meeting the standards required.

The primary aim of this thesis is to consider this estimation problem.

## 1.2 Overview

Chapter 2 of this thesis considers a possible model for repeated measures count data over-dispersed relative to the Poisson distribution and appropriate to the asbestos

fibre counting problem. To accomodate the over-dispersion a Poisson random variable is compounded with a positive random variable with mean equal one and variance linked linearly, via a log function, to a set of covariates. Maximum likelihood estimates of the parameters are obtained for the case where the compounding distribution is gamma and extended quasi-likelihood parameter estimates are obtained when the compounding distribution is unspecified. These two sets of parameters estimates are then shown to be comparable in certain circumstances. An example to illustrate the procedure is given based on data taken from the National Asbestos Program. This work has been published in van de Ven and Weber (1995).

In Chapter 3 a special case of the general model in Chapter 2 with a gamma compounding distribution is considered. Here repeat counts for a "subject" are taken as independent Poisson random variables with constant mean. The means of the "subjects" are then modelled as independent observations from a two parameter gamma distribution with unit mean. In relation to the asbestos fibre counting problem this model is suitable for repeat counts by counters on slides from a batch of nominally identical slides, assuming the counters are a homogeous group. Two sets of moment estimates for the parameters of the model are given and generalized variances of the limiting distribution of the moment estimates are compared with the corresponding quantity for the maximum likelihood estimates. Also included in this chapter is an alternative asymptotic theory that explains some of the erratic behaviour of the moment estimates.

From the results of Chapter 3 we show that based on asymptotic theory the moment estimates for the model compare favourably with the maximum likelihood estimates. This then begs the question as to how the estimators compare for finite samples. This is the subject of Chapter 4. Herein a very special case of the model in Chapter 3 is studied, this being when each "subject" reports a single result. In this case the set of results correspond to a sample from a negative binomial distribution (NBD). Chapter 4 gives the results of a simulation study comparing the moment estimate, the maximum likelihood estimate and two other estimates for estimating the shape parameter of the NBD distribution for different parameter values and

different sample sizes. Here the results differ from previous studies in two ways. Firstly, all of the four estimates considered can be compared simultaneously based on our study. This was not possible with earlier studies as they considered different sub-groups of the estimates and used different parameter settings, different sample sizes and different criteria for dealing with infinite and negative estimates of the scale parameter, thus making comparisons across studies difficult. Secondly, a new measure is used to compare the estimates, termed the "percentile measure". This measure, based on the difference between the percentiles of the true and estimated distribution function, is argued to be more appropriate in many cases, particular so in the asbestos fibre counting problem. The results in this chapter have been published in van de Ven (1993).

A difficulty remains with the results given in Chapters 2, 3 and 4, this being that it is assumed that there are no outliers in the data. When outliers are present the estimation procedures in these chapters are not robust. In the remaining chapters we therefore consider robust procedures, beginning in Chapter 5 by considering the NBD. Here we first obtain bounds for the median in terms of the mean. These results have appeared in van de Ven and Weber (1993). We then obtain percentile related bounds for the mean and use these bounds to robustly estimate the mean when the shape parameter is known. These latter results have been published in van de Ven and Weber (1996).

The remaining two chapters are devoted to robust estimation in (generalised) linear mixed models with the results much more general than those of the earlier chapters. Here the basis of the work is a paper by Schall (1991), a paper that came to our attention after much of the work of the previous chapters was completed. Here we robustify Schall's procedure by extending some of the ideas from Fellner (1986), a paper considering linear mixed models. In this extension process it was observed that Fellner's method could be improved. Our modification to Fellner's procedure is the subject of Chapter 6, wherein Fellner's method is compared with our modified Fellner method and with a third robust method, one we have proposed based on moment methods. This modified Fellner method is then incorporated into

the extension to robust estimation in generalized linear mixed models (GLMMs), the subject of Chapter7. An example in Chapter 7 illustrates the application of this robust procedure to data from the National Asbestos Program.

## 1.3 Comment

Throughout the chapters that follow the material will be presented in a general context and not specifically related to the asbestos fibre counting problem. This is because many of the results obtained are appropriate to more general problems. However, when examples are used to illustrate theory we will use data relating to the specific problem where possible.

# Chapter 2

# Modelling mean and dispersion in mixed Poisson regression models.

To model independent count data it is not unusual to initially assume that the data are from Poisson distributions with means related, via a link function, to a linear function of a set of covariates for each datum value. The model, once the link function has been specified, can then be fitted using the techniques of generalized linear models as developed by Nelder and Wedderburn (1972) and which are comprehensively covered in the text by McCullagh and Nelder (1989).

In practice the data do not always fit this basic model and often show degrees of overdispersion relative to the Poisson model. One mechanism for accommodating this overdispersion is to fit the regression model using the quasi-likelihood method introduced in Wedderburn (1974). This method allows the variance function to include an additional parameter to account for overdispersion with this parameter often taken as a simple scaling factor. An overview of this method with reference to count data is given in McNeney and Petkau (1994). For more general models where the overdispersion parameter is related to some linear function of a set of covariates the extended quasi-likelihood function of Nelder and Pregibon (1987) offers a possible procedure. Smyth (1989) also offers a solution in this more general case utilising the Pearson residuals and method of quasi-likelihood.

An alternative to the quasi-likelihood based methods for accommodating overdis-

persion is the fully parametric approach of assuming that, conditional on the means, the data are from Poisson distributions but that the underlying means are themselves independent random variables from some distribution specified up to a set of parameters. Here the most commonly used distributions for modelling the means are the gamma (e.g. Lawless (1987a), Stein (1988)), lognormal (e.g. Hinde (1982)) and the inverse Gaussian (e.g. Stein (1988), Dean et al. (1989)).

If we now consider the related situation but with dependent count data there exist fewer available exact modelling distributions and to overcome this a number of approximate procedures have been developed.

When interest focuses on the regression model for the marginal mean and the covariances are known functions of the mean, we can use the quasi-likelihood method for dependent data, as given in McCullagh (1983), with the method allowing scope for overdispersion in the model. In a related situation but where the correlation matrix of the data has block diagonal form ($K$ blocks), dependent on a vector parameter $\zeta$, we can use the generalized estimating equations (GEEs) discussed in Liang and Zeger (1986) and Zeger and Liang (1986). These GEEs give $K^{\frac{1}{2}}$ consistent, asymptotically Gaussian estimators for the regression parameters for the mean under mild regularity conditions, even when the correlation structure is misspecified, provided $\zeta$ can be consistently estimated. Paik (1992) has extended the results of Liang and Zeger (1986) to also include modelling of the overdispersion as a function of a set of covariates.

Another method of modelling dependent count data, which is particularly useful for data exhibiting overdispersion, is to use a mixed model. Here, as for the independent case, we assume that conditional on the given mean values the data are independent random variables. The means are then modelled as random observations (not all independent) from other distributions. Using this modelling procedure, and under certain assumptions, approximate methods have been developed by Schall (1991), Breslow and Clayton (1993) and McGilchrist (1994). In each case the approximation derived is an extension of the best linear unbiased prediction (BLUP) method of Henderson (1963). Assumptions common to all three papers are that the

models are hierachical and that the random error components entering the linear predictor are realizations of Gaussian random variables.

An alternative to the above approximate procedures in the case of dependent data is to use exact procedures under the mixed model assumptions. This procedure however has the disadvantage that often, though not always, it requires computer intensive computation procedures for calculating high-dimensional integrals. Fortunately, for some models, the integrals to obtain the marginal distributions can be solved explicitly. Such a case is for conditionally distributed Poisson data when the means are gamma random variables. This method has been employed to model Poisson process data by, for example, Lawless (1987b) and Cooil (1991). van Duijn (1992) has also used this method for count data available in contingency table form with the row effects random and column effects fixed.

In this chapter we will use the mixed model approach to model dependent count data that is overdispersed relative to the Poisson distribution. Here we mix the Poisson distribution with, firstly a gamma to obtain exact results, and secondly a more general distribution and obtain approximate results using the extended quasi-likelihood of Nelder and Pregibon (1987). In both cases the data will be assumed to be in the form of repeated measures with the overdispersion introduced by the different "subjects". Also, the extent of the overdispersion introduced by each subject will be modelled using a log-linear model which will allow the inclusion of continuous covariates. In Section 2.1 we formulate the model we aim to consider in its full generality. Section 2.2 outlines briefly the derivation of the maximum likelihood estimating equations for the regression parameters of the model when the mixing distribution is gamma. This is followed by Section 2.3 wherein possible starting values are proposed. In Section 2.4 the more general mixing distribution is considered and estimates of the parameters are obtained using the extended quasi-likelihood estimator of Nelder and Pregibon (1987). The estimates we obtain here will be shown to be approximations to the MLEs obtained in Section 2.2. Section 2.5 contains as an example data from the National Asbestos Program discussed in Chapter 1. We conclude with a discussion in Section 2.6.

## 2.1 General model.

Let $(y_{ij}, \mathbf{x}_{ij}, \mathbf{z}_i)$ denote a sequence of observations, $i = 1, \cdots, K$ and $j = 1, \cdots, n_i$ where $y_{ij}$ is the $(i,j)^{th}$ response variable with associated vectors of covariates $\mathbf{x}_{ij}$ and $\mathbf{z}_i$ of dimensions $p$ and $q$ respectively. Here $\mathbf{x}_{ij}$ and $\mathbf{z}_i$ may have elements in common. The model we consider in this chapter is that the data $(y_{ij})$ are conditionally independent given proportionality constants $\gamma_i$ with

$$y_{ij} \mid \gamma_i \sim \text{Poisson}(\gamma_i \, \mu_{ij}) \quad ; \quad i = 1, \cdots, K$$
$$j = 1, \cdots, n_i \ .$$

We then take $\gamma_i$ $(i = 1, \cdots K)$ to be independent realisations of positive random variables having distribution functions $G(x; \alpha_i)$, expectations equal 1 and variances equal $\alpha_i^{-1}$. Finally, we let

$$\ln \mu_{ij} = \mathbf{x}'_{ij} \, \boldsymbol{\beta} \ ,$$
$$\ln \alpha_i = \mathbf{z}'_i \, \boldsymbol{\eta} \ ,$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are parameter vectors of dimension p and q respectively.

It is worth noting that the constraint that the mean of $G(x; \alpha_i)$ is 1 is not unnecessarily restrictive. Provided the mean is independent of $j$, $G(x; \alpha_i)$ can be rescaled to have unit mean and the rescaling factor incorporated into existing, or additional, parameters in the linear regressions of $\ln \mu_{ij}$ and $\ln \alpha_i$.

Recently various versions of this model have been studied. Lawless (1987a) and Dean et al. (1989) have considered the case $n_i = 1$ (i.e. independent data) for $G(x; \alpha_i)$ gamma and inverse-Gaussian and $\alpha_i = \alpha \ \forall \ i$. The Power Function Model of Stein (1988) is a special case of the above, again with independent data. Also to have considered the independent case for $G(x; \alpha_i)$ gamma with $\alpha_i = \alpha \ \forall \ i$ have been Collings and Margolin (1985) and Barnwal and Paul (1988). Their interest however focused on testing hypotheses rather than estimation, with the former testing $H_0$ : $\alpha = \infty$ and the latter $H_0 : \mu_1 = \mu_2 = \cdots = \mu_K$.

For the case with dependent data (i.e. not all $n_i = 1$) we are aware of no study that has considered this model in its general form. Some studies have considered

slight modifications under restricted conditions. For example, Collings and Margolin (1985) and Lawless (1987b) have considered the restricted case $\alpha_i = \alpha \; \forall \; i$ but their models differ from the above in that $\ln \mu_{ij}$ is proportional to $\mathbf{x}'_{ij} \, \boldsymbol{\beta}$ with the constant of proportionality known and dependent on $i$ and $j$. Thall (1988) on the other hand considers a similar situation but with $\ln \mu_{ij}$ differing from $\mathbf{x}'_{ij} \, \boldsymbol{\beta}$ by a known constant (offset), again depending on $i$ and $j$.

We now consider estimation of the regression parameters in the above model when $G(x; \alpha_i)$ is the distribution of a $\Gamma(\alpha_i, 1/\alpha_i)$ random variable as defined in Appendix A.

## 2.2 Maximum likelihood estimation under a gamma model.

If L denotes the likelihood under the model in the previous section when $G(x; \alpha_i)$ is a gamma distribution, then

$$L = \prod_{i=1}^{K} \left[ \frac{\Gamma(\alpha_i + y_{i\cdot})}{\Gamma(\alpha_i) \, (\prod_{j=1}^{n_i} y_{ij}!)} \left( \frac{\alpha_i}{\alpha_i + \mu_{i\cdot}} \right)^{\alpha_i} \prod_{j=1}^{n_i} \left( \frac{\mu_{ij}}{\alpha_i + \mu_{i\cdot}} \right)^{y_{ij}} \right]$$

where a dot in the subscript indicates summation over the replaced subscript.

To simplify the expression for what follows we now introduce some matrix notation. For $i = 1, \cdots, K$ let:

$$\begin{aligned}
\mathbf{y}_i &= (y_{i1}, \cdots, y_{in_i})' \\
\boldsymbol{\mu}_i &= (\mu_{i1}, \cdots, \mu_{in_i})' \\
\mathbf{S}_i &= \mathbf{y}_i - \boldsymbol{\mu}_i \\
\mathbf{l}_i &= \left( \frac{\partial \ln L}{\partial \mu_{i1}}, , \cdots, \frac{\partial \ln L}{\partial \mu_{in_i}} \right)' \\
\mathbf{A}_i &= \text{diag} \, (\mu_{i1}, \cdots, \mu_{in_i}) \quad \text{an } n_i \times n_i \text{ matrix} \\
\mathbf{X}'_i &= (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in_i}) \quad \text{a } p \times n_i \text{ matrix, and} \\
\mathbf{V}_i &= \mathbf{A}_i + \boldsymbol{\mu}_i \, \boldsymbol{\mu}'_i / \alpha_i \quad \text{which is the variance of } \mathbf{y}_i \; .
\end{aligned}$$

Further, let

$$\begin{aligned}
\mathbf{A}_{K+1} &= \text{diag}\,(\alpha_1, \cdots, \alpha_K); \\
\mathbf{Z}' &= (\mathbf{z}_1, \cdots, \mathbf{z}_K); \quad \text{a } q \times K \text{ matrix, and} \\
\mathbf{l}_{K+1} &= \left(\frac{\partial \ln L}{\partial \alpha_1}, \cdots, \frac{\partial \ln L}{\partial \alpha_K}\right)' \quad \text{where we have for } i = 1, \cdots, K \\
\frac{\partial \ln L}{\partial \alpha_i} &= I(y_{i\cdot} \geq 1) \sum_{r=1}^{y_{i\cdot}} \left(\frac{1}{\alpha_i + r - 1}\right) + \ln\left(\frac{\alpha_i}{\alpha_i + \mu_{i\cdot}}\right) + \left(\frac{\mu_{i\cdot} - y_{i\cdot}}{\mu_{i\cdot} + \alpha_i}\right) . \quad (2.1)
\end{aligned}$$

Note: To obtain this last expression we use the result that the digamma function $\psi(x) = d \ln \Gamma(x)/dx$ satisfies $\psi(n+x) - \psi(x) = \sum_{i=1}^{n} (x + i - 1)^{-1}$ for all $x > 0$ and integer $n \geq 1$ (see Abramowitz and Stegun (1968), page 258).

Using the result that $\mathbf{l}_i = \mathbf{V}_i^{-1}\mathbf{S}_i, \quad i = 1, \cdots, K$ the maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\eta}}$ say, can be shown to jointly satisfy the following equations:

$$\begin{aligned}
\sum_{i=1}^{K} \mathbf{X}_i' \mathbf{A}_i \mathbf{V}_i^{-1} \mathbf{S}_i &= \mathbf{0} \\
\mathbf{Z}' \mathbf{A}_{K+1} \mathbf{l}_{K+1} &= \mathbf{0}
\end{aligned}$$

Since the solutions for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\eta}}$ cannot be obtained explicitly from these two equations we propose use of Fisher's scoring method, for which we need the Fisher information matrix.

From the equations

$$\begin{aligned}
\frac{\partial^2 \ln L}{\partial \beta_r \partial \beta_{r'}} &= \sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ijr} x_{ijr'} \left(\frac{\partial \ln L}{\partial \mu_{ij}} \mu_{ij}^2 - y_{ij}\right) \\
&\qquad + \sum_{i=1}^{K} \left(\frac{y_{i\cdot} + \alpha_i}{(\mu_{i\cdot} + \alpha_i)^2} \sum_{j=1}^{n_i} \mu_{ij} x_{ijr} \sum_{j'=1}^{n_i} \mu_{ij'} x_{ij'r'}\right) \\
\frac{\partial^2 \ln L}{\partial \eta_s \partial \beta_r} &= \sum_{i=1}^{K} \left(\frac{y_{i\cdot} - \mu_{i\cdot}}{(\mu_{i\cdot} + \alpha_i)^2} \alpha_i z_{is} \sum_{j=1}^{n_i} \mu_{ij} x_{ijr}\right) \\
\frac{\partial^2 \ln L}{\partial \eta_s \partial \eta_{s'}} &= \sum_{i=1}^{K} \alpha_i^2 z_{is} z_{is'} \left(\frac{1}{\alpha_i} \frac{\partial \ln L}{\partial \alpha_i} + \frac{\mu_{i\cdot}^2 + \alpha_i y_{i\cdot}}{\alpha_i (\mu_{i\cdot} + \alpha_i)^2} - I(y_{i\cdot} \geq 1) \sum_{j=1}^{y_{i\cdot}} \frac{1}{(\alpha_i + j - 1)^2}\right) ,
\end{aligned}$$

where $\beta_r, \eta_s, x_{ijr}$ and $z_{is}$ are elements of $\boldsymbol{\beta}$, $\boldsymbol{\eta}$, $\mathbf{x}_{ij}$ and $\mathbf{z}_i$ respectively, we obtain the following block elements of Fisher's information matrix:

$$
\begin{aligned}
E\left(-\frac{\partial^2 \ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right) &= \sum_{i=1}^{K} \mathbf{X}_i'\mathbf{A}_i\mathbf{B}_i\mathbf{A}_i\mathbf{X}_i \\
E\left(-\frac{\partial^2 \ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\eta}'}\right) &= \mathbf{0} \\
E\left(-\frac{\partial^2 \ln L}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right) &= \mathbf{Z}'\mathbf{A}_{K+1}\mathbf{B}_{K+1}\mathbf{A}_{K+1}\mathbf{Z} = \sum_{i=1}^{K} \alpha_i^2 b_{K+1,i}\mathbf{z}_i\mathbf{z}_i' ,
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{B}_i = \mathbf{V}_i^{-1} &= \mathbf{A}_i^{-1} - (\mu_{i\cdot} + \alpha_i)^{-1}\mathbf{1}\mathbf{1}', \quad i = 1, \cdots, K \text{ with } \mathbf{1} \text{ a vector of 1s; and} \\
\mathbf{B}_{K+1} &= \operatorname{diag}(b_{K+1,1}, \cdots, b_{K+1,K}) \\
&\quad \text{where } b_{K+1,i} = \sum_{j=2}^{\infty} \frac{\Gamma(\alpha_i) j!}{j^2\,\Gamma(\alpha_i + j)} \left(\frac{\mu_{i\cdot}}{\mu_{i\cdot} + \alpha_i}\right)^j, \quad i = 1, \cdots, K.
\end{aligned}
$$

To obtain the values of $b_{K+1,i}$ for $i = 1, \cdots, K$ we use the result that if y has a negative binomial distribution with mean $\mu$ and variance $\mu + \mu^2/\alpha$ then

$$
E\left(I(y \geq 1) \sum_{j=1}^{y} \frac{1}{(\alpha + j - 1)^2}\right) = \sum_{j=1}^{\infty} \frac{\Gamma(\alpha)\; j!}{j^2\;\Gamma(\alpha + j)} \left(\frac{\mu}{\mu + \alpha}\right)^j \tag{2.2}
$$

This result was given in Fisher (1941) and has subsequently been shown by Henderson (1992).

If we have starting values $\hat{\boldsymbol{\beta}}_{(0)}$ and $\hat{\boldsymbol{\eta}}_{(0)}$ we can then, from Fisher's scoring method, progressively iterate between the following two equations

$$
\hat{\boldsymbol{\beta}}_{(S+1)} = \hat{\boldsymbol{\beta}}_{(S)} + \left(\sum_{i=1}^{K} \mathbf{X}_i'\mathbf{A}_i\mathbf{B}_i\mathbf{A}_i\mathbf{X}_i\right)^{-1} \sum_{i=1}^{K} \mathbf{X}_i'\mathbf{A}_i\mathbf{V}_i^{-1}\mathbf{S}_i \tag{2.3}
$$

$$
\hat{\boldsymbol{\eta}}_{(S+1)} = \hat{\boldsymbol{\eta}}_{(S)} + (\mathbf{Z}'\mathbf{A}_{K+1}\mathbf{B}_{K+1}\mathbf{A}_{K+1}\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{A}_{K+1}\mathbf{l}_{K+1} \tag{2.4}
$$

where the second term on the right hand side of (2.3) is calculated at $\hat{\boldsymbol{\beta}}_{(S)}$ and $\hat{\boldsymbol{\eta}}_{(S)}$, and the second term on right hand side of (2.4) is calculated at $\hat{\boldsymbol{\beta}}_{(S+1)}$ and $\hat{\boldsymbol{\eta}}_{(S)}$. This process is repeated until convergence to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\eta}}$.

Then, as $K \to \infty$, we have $[K^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), K^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})]'$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix given by

$$
\lim_{K\to\infty} K \begin{pmatrix} \left(\sum_{i=1}^{K} \mathbf{X}_i'\mathbf{A}_i\mathbf{B}_i\mathbf{A}_i\mathbf{X}_i\right)^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{Z}'\mathbf{A}_{K+1}\mathbf{B}_{K+1}\mathbf{A}_{K+1}\mathbf{Z})^{-1} \end{pmatrix}
$$

provided this covariance matrix has a positive definite limit. Here the result follows as the typical "regularity assumptions" of weak conditions on derivatives of the log-likelihood and the existence of moments of the data are met. These conditions for the general case with nonhomogeneous observations, and in particular for independent multivariate, nonhomogeneous observations as we have here, are reviewed in a paper by Fahrmeir (1987).

**Note:** It must be stressed here that this result is a limiting distribution result and that the limiting variances of the estimates, in particular $\hat{\eta}$, do not equal those of the limiting distribution. This follows as the problem here is a generalization of the parameter estimation problem for the negative binomial distribution for which the estimate of the shape parameter does have not a proper distribution. See, for example, Anscombe (1950).

## 2.3 Starting values.

As starting values for $\hat{\beta}$ and $\hat{\eta}$ one suggestion, provided both regression models have been configured to contain constant terms, is to use $\hat{\beta}_{(0)} = (\hat{\beta}_{1(0)}, 0, \cdots, 0)'$ and $\hat{\eta}_{(0)} = (\hat{\eta}_{1(0)}, 0, \cdots, 0)'$ with

$$\begin{aligned} \exp(\hat{\beta}_{1(0)}) &= \tfrac{1}{K}\textstyle\sum_{i=1}^{K} \bar{y}_{i\cdot} \quad \text{where} \quad \bar{y}_{i\cdot} = \tfrac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \;\; ; \text{ and} \\ \exp(\hat{\eta}_{1(0)}) &= \frac{(\sum_{i=1}^{K} n_i\hat{\mu})^2}{\sum_{i=1}^{K} n_i^2(\bar{y}_{i\cdot} - \hat{\mu})^2 - \sum_{i=1}^{K} n_i\hat{\mu}} \quad \text{where } \hat{\mu} = \exp(\hat{\beta}_{1(0)}). \end{aligned}$$

Here $\exp(\hat{\beta}_{1(0)})$ and $\exp(\hat{\eta}_{1(0)})$ are moment estimators of $\mu_{ij}$ and $\alpha_i$ respectively when $\mu_{ij} = \mu$ and $\alpha_i = \alpha$ for all $i, j$. (For details see Chapter 3.)

As an alternative to the above starting value $\hat{\beta}_{(0)}$ we can use an estimate obtained from the following method which is based on the method of Breslow (1984), but extended to dependent data. This method, in addition to giving a better starting value, has the advantage that it does not require that the regression model for $\ln(\mu_{ij})$ contains a constant term.

Let $T_{ij} = \ln(y_{ij})$, or $\ln(y_{ij} + \frac{1}{2})$ if $y_{ij} = 0$. Using a first order Taylor approxi-

mation we have $T_{ij} \approx \ln(\mu_{ij}) + \frac{1}{\mu_{ij}}(y_{ij} - \mu_{ij})$ and hence

$$\begin{aligned} E(T_{ij}) &\approx \ln(\mu_{ij}) = \mathbf{x}'_{ij}\,\boldsymbol{\beta} \\ V(T_{ij}) &\approx \frac{1}{\alpha_i} + \frac{1}{\mu_{ij}} \\ Cov(T_{ij}, T_{ij'}) &\approx \frac{1}{\alpha_i} \text{ for } j \neq j'. \end{aligned}$$

Letting $\mathbf{T}_i = (T_{i1}, \cdots, T_{in_i})'$ we therefore have $V(\mathbf{T}_i) \approx \frac{1}{\alpha_i}\mathbf{1}\mathbf{1}' + \mathbf{A}_i^{-1}$ where $\mathbf{A}_i^{-1}$ is defined above. Denoting $V(\mathbf{T}_i)$ by $\mathbf{W}_i$ we could, if we knew $\mathbf{W}_i$, use a w.l.s. regression to obtain the following initial estimate of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_{(0)} = \left(\sum_{i=1}^{K} \mathbf{X}_i'\mathbf{W}_i^{-1}\mathbf{X}_i\right)^{-1} \sum_{i=1}^{K} \mathbf{X}_i'\mathbf{W}_i^{-1}\mathbf{T}_i$$

$$\text{where } \mathbf{W}_i^{-1} = \mathbf{A}_i - \frac{1}{\alpha_i + \mu_{i\cdot}}\boldsymbol{\mu}_i\boldsymbol{\mu}_i'$$

using the same notation as in Section 2.2.

However, as $\mathbf{W}_i$ is unknown we need to replace it by an estimate $\hat{\mathbf{W}}_i$. One solution is to let $\hat{\mathbf{W}}_i$ equal $\mathbf{W}_i$ calculated at $\alpha_i^{-1} = 0$ and $\mu_{ij} = \max(y_{ij}, \frac{1}{2})$. In this case $\hat{\mathbf{W}}_i^{-1}$ is a diagonal matrix with diagonal elements $\max(y_{ij}, \frac{1}{2})$. This will not be the optimal starting value but should prove better than the previous starting value. For a slightly better initial estimate we recommend using $\alpha_i = \exp(\hat{\eta}_{1(0)})$, as given above, when calculating $\hat{\mathbf{W}}_i$, instead of $\alpha_i^{-1} = 0$.

This estimate of $\hat{\boldsymbol{\beta}}_{(0)}$ can now be used to upgrade the $\hat{\boldsymbol{\eta}}_{(0)}$ estimate, again assuming that $\alpha_i = \alpha$ for all $i$, by replacing $\exp(\hat{\eta}_{1(0)})$ above by

$$\exp(\hat{\eta}_{1(0)}) = \frac{(\sum_{i=1}^{K} \hat{\mu}_{i\cdot})^2}{\sum_{i=1}^{K}(y_{i\cdot} - \hat{\mu}_{i\cdot})^2 - \sum_{i=1}^{K} \hat{\mu}_{i\cdot}}$$

where $\ln(\hat{\mu}_{i\cdot}) = \sum_{j=1}^{n_i} \mathbf{x}'_{ij}\,\hat{\boldsymbol{\beta}}_{(0)} = \mathbf{1}'\mathbf{X}_i\,\hat{\boldsymbol{\beta}}_{(0)}$.

## 2.4 Estimation under a general model using extended quasi-likelihood.

In the situation where one is unwilling to assume that the mixing distributions $G(x; \alpha_i)$ are gamma and instead wishes to leave them unspecified, except possibly

for the existence of cumulants up to a certain order, then there currently exist very few methods for modelling the variances ($\alpha_i^{-1}$) of these mixing distributions. In this section we derive estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\eta}}$ say, using the GEEs of Liang and Zeger (1986) coupled with the extended quasi-likelihood method of Nelder and Pregibon (1987).

The extended quasi-likelihood ($Q^+$) as defined by Nelder and Pregibon (1987) for a single datum value $y$ is given by

$$Q^+ = -\tfrac{1}{2}\left(\ln\{2\pi\phi V_\theta(y)\} + \phi^{-1}D_\theta(y;\mu)\right)$$

where $E(y) = \mu, \text{Var}\,(y) = \phi V_\theta(\mu)$ and

$$D_\theta(y;\mu) = -2\int_y^\mu \frac{y-t}{V_\theta(t)}\,dt\ .$$

Since, for the model in Section 2.1 we have $y_{i\cdot}$ $(i = 1,\cdots,K)$ independent with means $\mu_{i\cdot}$ and variances $\mu_{i\cdot} + \mu_{i\cdot}^2/\alpha_i$ the extended quasi-likelihood for the totals $y_{1\cdot}, \cdots, y_{K\cdot}$ is

$$\begin{aligned} Q^+ &= \sum_{i=1}^K \left(-\frac{1}{2}\ln\{2\pi(y_{i\cdot} + y_{i\cdot}^2/\alpha_i)\} + \int_{y_{i\cdot}}^{\mu_{i\cdot}} \frac{y_{i\cdot}-t}{t(1+t/\alpha_i)}\,dt\right) \\ &= \sum_{i=1}^K \left(-\frac{1}{2}\ln(2\pi y_{i\cdot}) - \frac{1}{2}\ln(1+\frac{y_{i\cdot}}{\alpha_i}) + y_{i\cdot}\ln(\frac{\mu_{i\cdot}}{y_{i\cdot}}) + (y_{i\cdot}+\alpha_i)\ln(\frac{\alpha_i+y_{i\cdot}}{\alpha_i+\mu_{i\cdot}})\right). \end{aligned}$$

Two points need to be made now. First, since we are using $Q^+$ to model discrete data we have a problem when $y_{i\cdot} = 0$ for any $i$ as then $V_{\theta_i}(y_i) = 0$ and hence $Q^+ = \infty$. To overcome this we recommend that $Q^+$ be calculated at $y_{i\cdot} + c$ for some small $c$ rather than $y_{i\cdot}$ if $y_{i\cdot} = 0$.

Secondly, as the $y_{1\cdot}, \cdots, y_{K\cdot}$ are sufficient for $\alpha_1, \cdots, \alpha_K$, and therefore for $\hat{\boldsymbol{\eta}}$, it is not necessary to extend the definition of the extended quasi-likelihood to dependent data in order to model the full data set.

Now, if we knew $\boldsymbol{\eta}$, then from the GEEs of Liang and Zeger (1986) we again have as our estimate of $\boldsymbol{\beta}$ the solution to $\sum_{i=1}^K \mathbf{X}_i'\mathbf{A}_i\mathbf{V}_i^{-1}\mathbf{S}_i = \mathbf{0}$. On the otherhand, if $\boldsymbol{\beta}$ were known then the maximum extended quasi-likelihood estimate of $\boldsymbol{\eta}$ would satisfy $\mathbf{Z}'\mathbf{A}_{K+1}\mathbf{l}_{K+1}^+ = \mathbf{0}$ where the $i^{th}$ element of $\mathbf{l}_{K+1}^+$, $i = 1,\cdots,K$ is given by

$$\frac{\partial Q^+}{\partial \alpha_i} = \ln\left(\frac{\alpha_i+y_{i\cdot}}{\alpha_i}\right) + \left(\frac{y_{i\cdot}}{2\alpha_i(\alpha_i+y_{i\cdot})}\right) + \ln\left(\frac{\alpha_i}{\alpha_i+\mu_{i\cdot}}\right) + \left(\frac{\mu_{i\cdot}-y_{i\cdot}}{\mu_{i\cdot}+\alpha_i}\right) \quad (2.5)$$

To solve for $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\eta}}$, starting with $\tilde{\boldsymbol{\beta}}_{\mathbf{0}}$ and $\tilde{\boldsymbol{\eta}}_{\mathbf{0}}$, we suggest iteration through the following combined Fisher Scoring, Newton-Raphson method

$$\tilde{\boldsymbol{\beta}}_{(S+1)} = \tilde{\boldsymbol{\beta}}_{(S)} + \left(\sum_{i=1}^{K} \mathbf{X}_i' \mathbf{A}_i \mathbf{B}_i \mathbf{A}_i \mathbf{X}_i\right)^{-1} \sum_{i=1}^{K} \mathbf{X}_i' \mathbf{A}_i \mathbf{V}_i^{-1} \mathbf{S}_i \tag{2.6}$$

$$\tilde{\boldsymbol{\eta}}_{(S+1)} = \tilde{\boldsymbol{\eta}}_{(S)} + \left(\mathbf{Z}' \mathbf{A}_{K+1} \mathbf{B}_{K+1}^{+} \mathbf{A}_{K+1} \mathbf{Z}\right)^{-1} \mathbf{Z}' \mathbf{A}_{K+1} \mathbf{l}_{K+1}^{+} \tag{2.7}$$

where $B_{K+1}^{+} = \text{diag}\,(b_{K+1,1}^{+}, \cdots, b_{K+1,K}^{+})$ with

$$b_{K+1,i}^{+} = -\left(\frac{1}{\alpha_i}\frac{\partial Q^{+}}{\partial \alpha_i} + \frac{\partial^2 Q^{+}}{\partial \alpha_i^2}\right), \text{ and}$$

$$\frac{\partial^2 Q^{+}}{\partial \alpha_i^2} = \left(\frac{\mu_{i\cdot}^2 + \alpha_i y_{i\cdot}}{\alpha_i(\mu_{i\cdot} + \alpha_i)^2}\right) - \left(\frac{y_{i\cdot}}{\alpha_i(\alpha_i + y_{i\cdot})} + \frac{y_{i\cdot}(2\alpha_i + y_{i\cdot})}{2\alpha_i^2(\alpha_i + y_{i\cdot})^2}\right)$$

using the same approach as in §2.

An interesting observation to be made here is that the estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ obtained from (2.6) and (2.7) can be considered as approximations to the estimates obtained from (2.3) and (2.4). To see this note that the two sets of score equations differ only in that $\mathbf{l}_{K+1}$ for the MLE is repaced by $\mathbf{l}_{K+1}^{+}$. Comparing (2.1) and (2.5), these two vectors differ only in that $I(y_{i\cdot} \geq 1)\sum_{r=1}^{y_{i\cdot}}(\alpha_i + r - 1)^{-1}$ is replaced by an approximation to it given by $\ln(\frac{\alpha_i + y_{i\cdot}}{\alpha_i}) + (\frac{y_{i\cdot}}{2\alpha_i(\alpha_i + y_{i\cdot})})$.

To appreciate this approximation note that for $y_{i\cdot} \geq 1$

$$\int_1^{y_{i\cdot}+1} \frac{1}{\alpha_i + x - 1}\, dx \leq \sum_{r=1}^{y_{i\cdot}} \left(\frac{1}{\alpha_i + r - 1}\right) \leq \frac{1}{\alpha_i} + \int_2^{y_{i\cdot}+1} \frac{1}{\alpha_i + x - 2}\, dx$$

giving

$$\ln\left(\frac{\alpha_i + y_{i\cdot}}{\alpha_i}\right) \leq \sum_{r=1}^{y_{i\cdot}} \left(\frac{1}{\alpha_i + r - 1}\right) \leq \ln\left(\frac{\alpha_i + y_{i\cdot}}{\alpha_i}\right) + \frac{y_{i\cdot}}{\alpha_i(\alpha_i + y_{i\cdot})}\,.$$

As this latter inequality holds when $y_{i\cdot} = 0$, provided we interpret the middle term as zero if $y_{i\cdot} = 0$, we have on averaging the bounds

$$I(y_{i\cdot} \geq 1)\sum_{r=1}^{y_{i\cdot}} \frac{1}{\alpha_i + r - 1} \approx \ln\left(\frac{\alpha_i + y_{i\cdot}}{\alpha_i}\right) + \left(\frac{y_{i\cdot}}{2\alpha_i(\alpha_i + y_{i\cdot})}\right)\,.$$

This approximation is satisfactory for $\alpha_i$ not too small (e.g. $\alpha_i \geq 0.5$) and improves as $\alpha_i$ increases.

Before ending this section we observe that if $\tilde{\boldsymbol{\eta}}$ is $K^{1/2}$-consistent for $\boldsymbol{\eta}$ given $\boldsymbol{\beta}$ then from Liang and Zeger (1986), Theorem 2, we have $K^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically ($K \to \infty$) multivariate normal with mean $\mathbf{0}$ and covariance matrix given by $\lim_{K\to\infty} K(\sum_{i=1}^{K} \mathbf{X}_i'\mathbf{A}_i\mathbf{B}_i\mathbf{A}_i\mathbf{X}_i)^{-1}$. However, as was shown by Davidian and Carroll (1988) estimators based on the extended quasi-likelihood need not be consistent and hence we cannot claim asymptotic normality for $\tilde{\boldsymbol{\beta}}$. Davidian and Carroll (1988) do claim however that the bias for estimators based on the extended quasi-likelihood will in general not be significant if the true distribution is a member of the exponential family.

## 2.5 Example.

As an example we take data from the National Asbestos Program discussed in Chapter 1. The data are given in Table 2.1. These results are for six slides containing similar asbestos dust type and which have been counted by 113 different counters, each identified by a unique identification code (Id). In this table the column labelled y gives the counts whilst the column labelled CS gives the counter's status. Here a counter status of '1' or '2' identifies reference counters, a '3' accredited counters and '4' non-accredited counters. Further, the counters with status '1' correspond to two automatic image analysers.

Let $y_{ij}$ denote the $j^{th}$ count for the $i^{th}$ counter and $\mu_{ij}$ the unconditional expectation of $y_{ij}$. We use the model

$$\ln \mu_{ij} = \sum_{k=1}^{7} x_{ijk}\beta_k$$

where $x_{ijk} = 1$ if the $j^{th}$ count for the $i^{th}$ counter is for the $k^{th}$ slide, and 0 otherwise ($k = 1, \cdots, 6$) whilst $x_{ij7}$ indicates if the counter status is '1'. The inclusion of this latter covariate is to take note of the proportionately lower counts obtained by the image analysers in previous studies (see, for example, Crawford and Cowie (1984)). Fitting this model, assuming no difference between counters within each counter class, we obtain the parameter estimates

Table 2.1 Counts reported by different counters for a set of six slides.

| Slide 1 | | | Slide 2 | | | Slide 3 | | | Slide 4 | | | Slide 5 | | | Slide 6 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Id. | y | CS | Id. | y | CS | Id. | y | CS | Id. | y | CS | Id. | y | CS | Id. | y | CS |
| 1 | 20 | 1 | 1 | 13 | 1 | 4 | 28 | 2 | 4 | 18 | 2 | 4 | 28 | 2 | 4 | 24 | 2 |
| 2 | 19 | 1 | 3 | 26 | 2 | 4 | 26 | 2 | 4 | 32 | 2 | 4 | 36 | 2 | 4 | 29 | 2 |
| 3 | 38 | 2 | 4 | 35 | 2 | 4 | 26 | 2 | 12 | 27 | 3 | 4 | 23 | 2 | 4 | 20 | 2 |
| 4 | 26 | 2 | 4 | 48 | 2 | 4 | 30 | 2 | 17 | 20 | 3 | 7 | 18 | 3 | 8 | 30 | 3 |
| 4 | 34 | 2 | 4 | 29 | 2 | 4 | 12 | 2 | 25 | 11 | 3 | 10 | 22 | 3 | 9 | 19 | 3 |
| 4 | 32 | 2 | 4 | 32 | 2 | 4 | 21 | 2 | 38 | 28 | 4 | 15 | 21 | 3 | 11 | 22 | 3 |
| 4 | 32 | 2 | 4 | 26 | 2 | 4 | 24 | 2 | 41 | 20 | 4 | 19 | 22 | 3 | 14 | 20 | 3 |
| 4 | 32 | 2 | 4 | 28 | 2 | 4 | 26 | 2 | 46 | 29 | 4 | 30 | 14 | 3 | 22 | 22 | 3 |
| 4 | 30 | 2 | 4 | 24 | 2 | 16 | 35 | 3 | 48 | 14 | 4 | 31 | 20 | 3 | 23 | 14 | 3 |
| 4 | 22 | 2 | 4 | 30 | 2 | 21 | 29 | 3 | 60 | 22 | 4 | 32 | 22 | 4 | 24 | 19 | 3 |
| 4 | 22 | 2 | 4 | 24 | 2 | 26 | 28 | 3 | 65 | 26 | 4 | 34 | 16 | 4 | 28 | 14 | 3 |
| 5 | 30 | 2 | 5 | 50 | 2 | 47 | 46 | 4 | 69 | 26 | 4 | 37 | 15 | 4 | 29 | 40 | 3 |
| 6 | 30 | 2 | 20 | 36 | 3 | 49 | 25 | 4 | 73 | 29 | 4 | 44 | 28 | 4 | 35 | 16 | 4 |
| 13 | 17 | 3 | 33 | 29 | 4 | 51 | 16 | 4 | 101 | 38 | 4 | 54 | 8 | 4 | 36 | 12 | 4 |
| 18 | 21 | 3 | 39 | 14 | 4 | 55 | 28 | 4 | | | | 57 | 24 | 4 | 40 | 33 | 4 |
| 27 | 19 | 3 | 46 | 34 | 4 | 58 | 38 | 4 | | | | 60 | 6 | 4 | 42 | 20 | 4 |
| 43 | 24 | 4 | 50 | 14 | 4 | 62 | 27 | 4 | | | | 65 | 35 | 4 | 56 | 18 | 4 |
| 45 | 12 | 4 | 52 | 18 | 4 | 72 | 19 | 4 | | | | 66 | 28 | 4 | 59 | 22 | 4 |
| 46 | 15 | 4 | 58 | 35 | 4 | 76 | 26 | 4 | | | | 73 | 28 | 4 | 81 | 15 | 4 |
| 47 | 38 | 4 | 61 | 20 | 4 | 78 | 35 | 4 | | | | 74 | 21 | 4 | 83 | 22 | 4 |
| 53 | 47 | 4 | 64 | 31 | 4 | 85 | 26 | 4 | | | | 82 | 26 | 4 | 93 | 24 | 4 |
| 58 | 19 | 4 | 67 | 32 | 4 | 97 | 26 | 4 | | | | 92 | 16 | 4 | 94 | 15 | 4 |
| 63 | 38 | 4 | 71 | 26 | 4 | 98 | 34 | 4 | | | | 104 | 20 | 4 | | | |
| 68 | 44 | 4 | 77 | 22 | 4 | 99 | 34 | 4 | | | | 108 | 11 | 4 | | | |
| 68 | 58 | 4 | 84 | 38 | 4 | | | | | | | | | | | | |
| 70 | 25 | 4 | 86 | 23 | 4 | | | | | | | | | | | | |
| 75 | 31 | 4 | 87 | 34 | 4 | | | | | | | | | | | | |
| 79 | 23 | 4 | 88 | 41 | 4 | | | | | | | | | | | | |
| 80 | 10 | 4 | 90 | 50 | 4 | | | | | | | | | | | | |
| 84 | 10 | 4 | 91 | 24 | 4 | | | | | | | | | | | | |
| 89 | 10 | 4 | 95 | 28 | 4 | | | | | | | | | | | | |
| 96 | 26 | 4 | 103 | 43 | 4 | | | | | | | | | | | | |
| 100 | 16 | 4 | 105 | 25 | 4 | | | | | | | | | | | | |
| 102 | 14 | 4 | 107 | 47 | 4 | | | | | | | | | | | | |
| 102 | 17 | 4 | 110 | 26 | 4 | | | | | | | | | | | | |
| 106 | 10 | 4 | 112 | 30 | 4 | | | | | | | | | | | | |
| 109 | 29 | 4 | 113 | 19 | 4 | | | | | | | | | | | | |
| 111 | 17 | 4 | | | | | | | | | | | | | | | |
| 112 | 30 | 4 | | | | | | | | | | | | | | | |

| | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ |
|---|---|---|---|---|---|---|---|
| Estimate | 3.25 | 3.41 | 3.32 | 3.19 | 3.05 | 3.06 | -0.45 |

with a deviance of 457.66 on 153 degrees of freedom. It is therefore clear that the model assuming independence does not fit the data. By including a counter effect in the model we can see that a significant proportion of the overdispersion is due to the variability of the counter bias.

To model the extra Poisson variation introduced by the individual counters we introduce a proportionality constant $\gamma_i$ with variance $\alpha_i^{-1}$ and modelled by

$$\ln \alpha_i = \eta_1 + z_{i2}\eta_2$$

where $z_{i2}$ indicates if the $i^{th}$ counter has status '4'. This allows the variance of the non-accredited counters to differ from the other more experienced counters. Here it would be preferable to also allow the variance of the automatic image analysers to differ from the others as their results are obtained somewhat differently from the other counters. However, as we have only three counts here it was felt more appropriate to include them with the experienced counters.

If we now assume that the $\gamma_i$ are independently drawn from a gamma distribution with unit mean and variance $\alpha_i^{-1}$, estimates for the parameters in the model, together with estimates of the associated standard errors are:

| | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | $\eta_1$ | $\eta_2$ |
|---|---|---|---|---|---|---|---|---|---|
| Estimate | 3.20 | 3.39 | 3.24 | 3.17 | 3.08 | 3.07 | -0.39 | 3.39 | -1.01 |
| Std. Error | 0.05 | 0.05 | 0.06 | 0.08 | 0.06 | 0.06 | 0.20 | 0.54 | 0.58 |

We therefore observe, for example, that the results for the automatic image analysers are on average estimated to be approximately $\frac{2}{3}$ those for the other counters. Also, the variability of the counts for the non-accredited counters exceeds (not necessarily significantly) those for the others, as would be expected.

Using these estimates of the parameters we can obtain estimates of prediction intervals for the assessment of future counts on these slides. Here we use the fact that the unconditional distribution of a count for a given slide has a negative binomial

distribution with mean $\mu_{ij}$ and shape parameter $\alpha_i$. Thus, for example, an estimate of an approximate 95% prediction interval for a future count by an accredited counter on slide 6 with estimated mean 21.54 would be [11, 34].

If we do not assume a gamma mixing distribution and instead estimate the parameters using the extended quasi-likelihood method given in Section 2.4 we obtain the same estimates for $\beta_1, \cdots, \beta_7$ but we now obtain $\eta_1 = 3.37$ and $\eta_2 = -1.01$, that is, only a small change in the $\eta_1$ estimate.

Finally, because the gamma and log-normal distributions can look reasonably similar for certain parameter values, and because in our example we have an hierarchical model, we model the data using the procedures of Schall (1991), Breslow and Clayton (1993) and McGilchrist (1994). Here the aim is to see how their methods, which agree for this example, compare with the exact method. The model, conditional on the random effect $\gamma_i$ which is assumed now to be log-normal, is that $\ln E\,(\boldsymbol{y}_i|\gamma_i) = \boldsymbol{X}_i\,\boldsymbol{\beta} + \mathbf{1}\,b_i$ where $\mathbf{1}$ is a vector of 1's and the $b_i$ are independent normal random varables with zero means and variances $\sigma_1^2$ for counters with status '1', '2' or '3' and $\sigma_2^2$ for counters with status '4'. The parameter estimates, using REML estimates for the variance components, are

| | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | $\sigma_1^2$ | $\sigma_2^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Estimate | 3.15 | 3.34 | 3.20 | 3.15 | 3.07 | 3.05 | -0.34 | 0.060 | 0.135 |

Recalling that if $\ln(V) \sim N(0, \sigma^2)$ the variance of $V$ is $\exp(\sigma^2)\,(\exp(\sigma^2) - 1)$, we obtain the following estimates of the variances of the proportionality constants: $\hat{\alpha}_1^{-1} = 0.066$ and $\hat{\alpha}_2^{-1} = 0.165$. This would thus correspond to the values $\eta_1 = 2.8$ and $\eta_2 = -0.8$ in the model $\ln \alpha_i = \eta_1 + z_{i2}\eta_2$. We thus see that the fixed effects estimates are not markedly different using this approximate inference procedure and the exact method but there are slight differences in the estimates for the parameters in the variance model.

## 2.6 Discussion.

Using the procedures developed in this paper it is possible to estimate the parameters of the mean model for repeated measures Poisson data with log link in the presence of heterogeous overdispersion. This is achieved by compounding the Poisson random variables with a second random variable, the variance of which is linearly related, again via a log link, to a set of covariates. One immediate advantage of this procedure is that it will result in more efficient estimates than when the heterogeneity is ignored. A second advantage is that when the degree of overdispersion is also of interest, as in the example given above, the overdispersion can be quantified. In this example a measure of the overdispersion is required in order to obtain estimates of prediction intervals.

To estimate the parameters in the models the iterative procedure that is given is the Fisher Scoring (FS) method in the case of a gamma compounding distribution and a combined Fisher Scoring, Newton-Raphson (FS-NR) method for the more general compounding distribution. In the two cases the FS method was used where possible as it is generally considered superior (converges more rapidly) to the NR method; see for example Kale (1962). However, to obtain the MLEs for the example given this is not the case. Beginning with a slightly modified starting value to that given in Section 2.3 and iterating until there is no further change in the third decimal place, the FS method requires 30 iterations to converge compared with 4 iterations using a combined FS-NR method equivalent to that developed for the extended quasi-likelihood solution. The FS method also requires 30 iterations when beginning at the originally proposed starting value.

The reason for the slightly modified starting value mentioned in the above comparison was because the combined FS-NR did not converge from the starting value proposed in Section 2.3, a situation that is always possible with iterative procedures. Unfortunately, the problem of non-convergence in this situation may go beyond the simple exercise of choosing a new starting value. Here there is always the possibility that one or more of the estimates of $\alpha_i$ will tend to zero, in which case one

or more of the parameter estimates in the variance regression model will necessarily tend in magnitude to infinity. In the simplest case of the model considered here, corresponding to the negative binomial distribution, such a case corresponds to the sample variance being less than the sample mean. For the more general case though no such simple characterization procedure is available. Therefore, in some cases of non-convergence of estimates, it must be borne in mind that no choice of starting value will necessarily overcome non-convergence. However, if there is overdispersion and the covariates in the log linear model for the variance are chosen judiciously, this should not be a problem.

# Chapter 3

# Moment estimators for a mixed gamma-Poisson model with repeated measures.

In Chapter 2 the maximum likelihoood estimates (MLE) were obtained for the regression parameters for the mixed gamma-Poisson model when both the unconditional mean of the Poisson and the variance of the gamma are log linear. To obtain these estimates an iterative procedure was given which utilised the Fisher scoring algorithm.

We now consider a special case of that model, that being when both log linear models each only contain a constant term. In this case we have in effect a simple extension of the negative binomial distribution. It will therefore be more appropriate in this case to estimate, not the parameters in the regression model, but rather the exponentiation of these values.

In Section 3.1 we give the model together with the maximum likelihood estimates of the two parameters. We then obtain, in the next section, two method of moment estimators based on first and second sample moments. Moment estimators based on these sample moments in general give rise to reasonably efficient estimators whilst giving the added advantage that they can be obtained explicitly. In Section 3.3 we obtain some results for the limiting distribution of these moment estimators

and then Section 3.4 provides a comparison of the asymptotic generalized variances for the two moment estimators and the MLE.

In Section 3.5 we present an alternative asymptotic theory wherein we allow the parameter values to vary with increasing sample size, in contrast to the previous asymptotic results where the parameters were kept constant. Here the parameters are allowed the vary in a manner not unlike situations giving rise to very variable estimates for $\alpha$ and we show that the limiting distribution of the two moment estimators of $\alpha$ can have very heavy tailed distributions with all moments infinite. We finish the chapter with some concluding remarks contained in Section 3.6.

## 3.1 Model and maximum likelihood estimates.

Let $Y_{ij}$ denote the $j^{th}$ result for the $i^{th}$ "subject" for $i = 1, \cdots, K$ and $j = 1, \cdots, n_i$, where the $n_i$ are uniformly bounded. Further comment relating to this condition on the $n_i$ will be given later but suffice to say here that this condition is sufficient to ensure later results.

Conditioning on a variable $\gamma_i$ we now assume that the $Y_{ij}$ are independent, Poisson random variables with mean $\gamma_i\ \mu$ whilst the $\gamma_i$ are assumed independent, gamma random variables having unit mean and variance $\alpha^{-1}$. In the special case $n_i = 1$ for all $i$, the $Y_{ij}$ are unconditionally independent negative binomial random variables.

For this model we have the following likelihood (L)

$$L = \prod_{i=1}^{K} \left[ \frac{\Gamma(\alpha + y_{i\cdot})}{\Gamma(\alpha) \prod_{j=1}^{n_i} (y_{ij}!)} \left( \frac{\alpha}{n_i\ \mu + \alpha} \right)^{\alpha} \left( \frac{\mu}{n_i\ \mu + \alpha} \right)^{y_{i\cdot}} \right].$$

Denoting $\ln L$ by $\mathcal{L}$ and differentiating with respect to the parameters we obtain

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{y_{\cdot\cdot}}{\mu} - \sum_{i=1}^{K} n_i \left( \frac{y_{i\cdot} + \alpha}{n_i\ \mu + \alpha} \right) \tag{3.1}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{K}\left(I(y_{i\cdot} \geq 1)\sum_{r=1}^{y_{i\cdot}}\frac{1}{\alpha+r-1}\right) + \sum_{i=1}^{K} ln\left(\frac{\alpha}{n_i\,\mu+\alpha}\right) + \sum_{i=1}^{K}\left(\frac{n_i\,\mu - y_{i\cdot}}{n_i\,\mu+\alpha}\right) \qquad (3.2)$$

where a dot as a subcript denotes summation over the corresponding values. Again, the leading term in (3.2) follows from the note given with (2.1).

To obtain the maximum likelihood estimators we solve $\frac{\partial \mathcal{L}}{\partial \mu} = 0$ and $\frac{\partial \mathcal{L}}{\partial \alpha} = 0$. The standard procedure for solving these two equations is Fisher's scoring method for which we need the expected values of the second derivatives of $\mathcal{L}$. Here the second derivatives are given by

$$\frac{\partial^2 \mathcal{L}}{\partial \mu^2} = \sum_{i=1}^{K}\frac{n_i^2(y_{i\cdot}+\alpha)}{(n_i\,\mu+\alpha)^2} - \frac{y_{\cdot\cdot}}{\mu^2}\,,$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \mu} = \sum_{i=1}^{K} n_i\left(\frac{n_i\,\mu - y_{i\cdot}}{(n_i\,\mu+\alpha)^2}\right)\,,$$

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = \sum_{i=1}^{K}\left(\frac{(n_i\,\mu)^2+\alpha y_{i\cdot}}{\alpha(n_i\,\mu+\alpha)^2}\right) - \sum_{i=1}^{K}\left(I(y_{i\cdot}\geq 1)\sum_{r=1}^{y_{i\cdot}}\frac{1}{(\alpha+r-1)^2}\right).$$

Letting $N = \sum_{i=1}^{K} n_i$ we have

$$E\left(-\frac{\partial^2 \mathcal{L}}{\partial \mu^2}\right) = \sum_{i=1}^{K}\frac{\alpha}{\mu^2}\left(\frac{n_i\mu}{n_i\mu+\alpha}\right)\,,$$

$$E\left(-\frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \mu}\right) = 0\,, \quad \text{and}$$

$$E\left(-\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right) = \sum_{i=1}^{K}\sum_{r=2}^{\infty}\frac{r!\,\Gamma(\alpha)}{r^2\,\Gamma(\alpha+r)}\left(\frac{n_i\,\mu}{n_i\mu+\alpha}\right)^r$$

where the last equation follows from (2.2).

The MLEs are then obtained by progressively iterating between the following two equations until the required degree of convergence is met.

$$\hat{\mu}_{(S+1)} = \hat{\mu}_{(S)} + \left[E\left(-\frac{\partial^2 \mathcal{L}}{\partial \mu^2}\right)\right]^{-1}_{\hat{\mu}_{(S)},\hat{\alpha}_{(S)}}\left[\frac{\partial \mathcal{L}}{\partial \mu}\right]_{\hat{\mu}_{(S)},\hat{\alpha}_{(S)}}\,,$$

$$\hat{\alpha}_{(S+1)} = \hat{\alpha}_{(S)} + \left[E\left(-\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right)\right]^{-1}_{\hat{\mu}_{(S+1)},\hat{\alpha}_{(S)}}\left[\frac{\partial \mathcal{L}}{\partial \alpha}\right]_{\hat{\mu}_{(S+1)},\hat{\alpha}_{(S)}}\,, \qquad S = 0, 1, \cdots$$

where $\partial \mathcal{L}/\partial \mu$ and $\partial \mathcal{L}/\partial \alpha$ are given in (3.1) and (3.2) respectively.

We then have that as $K \to \infty$, $[K^{1/2}(\hat{\mu} - \mu), K^{1/2}(\hat{\alpha} - \alpha)]$ has a limiting distribution that is bivariate normal with zero means and covariance matrix given by

$$\lim_{K\to\infty} K \begin{pmatrix} \left(\sum_{i=1}^{K} \frac{\alpha}{\mu^2}\left(\frac{n_i\mu}{n_i\mu+\alpha}\right)\right)^{-1} & 0 \\ 0 & \left(\sum_{i=1}^{K}\sum_{r=2}^{\infty} \frac{r!\,\Gamma(\alpha)}{r^2\,\Gamma(\alpha+r)}\left(\frac{n_i\mu}{n_i\mu+\alpha}\right)^r\right)^{-1} \end{pmatrix} . \quad (3.3)$$

Here there are no restrictions required on the $n_i$ other than that they are all greater than or equal one as the covariance matrix is positive definite for all sequences $n_i$ when $\mu$ and $\alpha$ are both interior points of $(0, \infty)$.

## 3.2 Method of moment estimates.

As an alternative to the maximum likelihood estimators for the parameters we now give two method of moment estimators. These two are equivalent for the case $n_i = n \; \forall \, i$ and when $n = 1$ correspond to the usual moment estimators for the mean $(\mu)$ and shape $(\alpha)$ parameters of the negative distribution distribution denoted by $NBD(\mu, \alpha)$.

Let a bar denote the averaging operator, so that $\bar{Y}_{i\cdot} = Y_{i\cdot}/n_i$ and $\bar{Y}_{\cdot\cdot} = Y_{\cdot\cdot}/N$ . Our two sets of method of moment estimators (MME) are:

**MME-1**

$$\begin{aligned} \hat{\mu}_1 &= \bar{Y}_{\cdot\cdot} \; , \\ \hat{\alpha}_1 &= \frac{(N - \sum_{i=1}^{K} n_i^2/N)\;\hat{\mu}_1^2}{S_1 - (K-1)\;\hat{\mu}_1} \; , \quad \text{where } S_1 = \sum_{i=1}^{K} n_i(\bar{Y}_{i\cdot} - \hat{\mu}_1)^2 \end{aligned}$$

**MME-2**

$$\begin{aligned} \hat{\mu}_2 &= \frac{1}{K}\sum_{i=1}^{K} \bar{Y}_{i\cdot} \; , \\ \hat{\alpha}_2 &= \frac{(K-1)\hat{\mu}_2^2}{S_2 - \frac{(K-1)}{K}\,(\sum_{i=1}^{K} n_i^{-1})\;\hat{\mu}_2} \; , \quad \text{where } S_2 = \sum_{i=1}^{K}(\bar{Y}_{i\cdot} - \hat{\mu}_2)^2 \; . \end{aligned}$$

Here the primary difference between the two sets of estimators is that the former gives equal weight to each $Y_{ij}$ whilst the latter gives equal weight to the $\bar{Y}_{i\cdot}$. The motivation behind the latter is to avoid giving too much weight to those "subjects" for which we happen to have many repeat measures.

To obtain the derivation of these two sets of moment estimators we first observe that the $Y_{i\cdot}$ $(i = 1, \cdots, K)$ are independent $NBD(n_i\mu, \alpha)$ random variables and hence both $\hat{\mu}_1$ and $\hat{\mu}_2$ are unbiased estimators of $\mu$. To simplify the derivation of the $\alpha$ estimators we let

$$Z_i = (\bar{Y}_{i\cdot} - \mu) \tag{3.4}$$

with $\mathbf{Z}' = (Z_1, \cdots, Z_K)$ and use the following two results.

**Lemma 3.1**

$$S_1 = \frac{1}{N}\mathbf{Z}'\mathbf{A}_1\mathbf{Z} ,$$

where $\mathbf{A}_1 = \left\{a_{ij}^{(1)}\right\}$ with $a_{ii}^{(1)} = n_i(N - n_i)$ and $a_{ij}^{(1)} = -n_i n_j$ for $i \neq j$.

**Proof.**

$$\begin{aligned} S_1 &= \sum_{i=1}^{K} n_i \left[\left(\frac{N - n_i}{N}\right)\bar{Y}_{i\cdot} - \sum_{j\neq i}\frac{n_j}{N}\bar{Y}_{j\cdot}\right]^2 \\ &= \frac{1}{N}\sum_{i=1}^{K}\frac{n_i}{N}\left[(N - n_i)Z_i - \sum_{j\neq i} n_j Z_j\right]^2 \\ &= \frac{1}{N}\sum_{i=1}^{K}\frac{n_i}{N}\mathbf{Z}'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z} = \frac{1}{N}\mathbf{Z}'\left(\sum_{i=1}^{K}\frac{n_i}{N}\mathbf{u}_i\mathbf{u}_i'\right)\mathbf{Z} , \end{aligned}$$

where

$$\mathbf{u}_i' = (u_{i1}, \cdots, u_{iK}) \quad \text{with} \quad \begin{aligned} u_{ij} &= N - n_i \quad \text{if } i = j , \\ &= -n_j \quad \text{if } i \neq j . \end{aligned}$$

The result then follows.

•

**Lemma 3.2**

$$S_2 = \frac{1}{N}\mathbf{Z}'\mathbf{A}_2\mathbf{Z} ,$$

where $\mathbf{A}_2 = \left\{a_{ij}^{(2)}\right\}$ with $a_{ii}^{(2)} = (K - 1)$ and $a_{ij}^{(2)} = -1$ for $i \neq j$.

**Proof.**

The proof here is similar to the proof of Lemma 3.1.

•

From Lemma 3.1, we have

$$E\,(S_1) = \frac{1}{N} Tr\,(\mathbf{A_1}\; E\,(\mathbf{ZZ}')) = \frac{1}{N}\sum_{i=1}^{K} n_i(N-n_i)\; V\,(\bar{Y}_{i\cdot})\;, \tag{3.5}$$

which, using the result $V\,(Y_{i\cdot}) = n_i\;\mu + (n_i\;\mu)^2/\alpha$, gives

$$E\,(S_1) \;=\; (K-1)\;\mu + (N - \sum_{i=1}^{K} n_i^2/N)\frac{\mu^2}{\alpha}\;. \tag{3.6}$$

We now obtain $\hat{\alpha}_1$ from (3.6), estimating $\mu$ by $\hat{\mu}_1$.

Similarily, we can show that

$$E\,(S_2) \;=\; (K-1)\left(\mu\sum_{i=1}^{K} n_i^{-1}/K + \mu^2/\alpha\right)$$

and from this we obtain $\hat{\alpha}_2$, estimating $\mu$ with $\hat{\mu}_2$.

## 3.3 Properties of the method of moment estimates.

We begin this section by giving the variances for the two moment estimators for $\mu$.

**Theorem 3.1**

$$V\,(\hat{\mu}_1) \;=\; \frac{\mu}{N} + \frac{\mu^2}{N^2\alpha}\sum_{i=1}^{K} n_i^2\;, \quad \text{and} \tag{3.7}$$

$$V\,(\hat{\mu}_2) \;=\; \frac{\mu}{K^2}\sum_{i=1}^{K} n_i^{-1} + \frac{\mu^2}{K\alpha}\;. \tag{3.8}$$

**Proof.**

Since $Y_{i\cdot}$ $(i = 1,\cdots,K)$ are independent $NBD(n_i\mu,\alpha)$ random variables we have that the mean and variance of $Y_{i\cdot}$ equal $n_i\mu$ and $n_i\mu + (n_i\mu)^2/\alpha$ respectively. The above results are thus readily obtained.

•

Unfortunately, in order to compare the two estimators of $\alpha$ ($\hat{\alpha}_1$ and $\hat{\alpha}_2$) we are unable to obtain results corresponding to Theorem 3.1, nor for covariances of $\hat{\mu}_i$ and $\hat{\alpha}_i$ , $i = 1,2$. This is because, in addition to the intractability of these values,

there is also the problem that the quantities may not exist. This would be the case if the denominator for $\hat{\alpha}_i$ , $i = 1, 2$ could take the value zero with positive probability. For example, for the negative binomial case (i.e. $n_i = 1 \; \forall \; i$) the denominator would be zero if the sample variance equalled the sample mean. However, if this is not possible, there remains the possibility that the denominator may take on extremely small values, be it with very small probability for large $K$, causing the variance of $\hat{\alpha}_i$ , $i = 1, 2$ to explode. This would then make comparisons of the two estimators based on these quantities dubious.

With this in mind we instead obtain results below for two random variables $\hat{\alpha}_1^*$ and $\hat{\alpha}_2^*$ that differ from $\hat{\alpha}_1$ and $\hat{\alpha}_2$, respectively, by $O_p(K^{-1})$. Results will then be obtained for these two new random variables, which incidently will have the same asymptotic distribution as their counterparts.

To obtain these we need the following results for $Z_i$ defined in (3.4).

$$E\left(Z_i^2\right) = \frac{1}{n_i^2}\left(n_i\mu + \frac{n_i^2\mu^2}{\alpha}\right) , \tag{3.9}$$

$$E\left(Z_i^3\right) = \frac{1}{n_i^3}\left(n_i\mu + \frac{3n_i^2\mu^2}{\alpha} + \frac{2n_i^3\mu^3}{\alpha^2}\right) , \text{ and} \tag{3.10}$$

$$E\left(Z_i^4\right) = \frac{1}{n_i^4}\left(n_i\mu + \frac{(3\alpha+7)n_i^2\mu^2}{\alpha} + \frac{6(\alpha+2)n_i^3\mu^3}{\alpha^2} + \frac{3(\alpha+2)n_i^4\mu^4}{\alpha^3}\right) . \tag{3.11}$$

These results are readily obtained from the moments of the negative binomial distribution.

Also, for the MME-1 results we need the following lemmas.

**Lemma 3.3**

As $K \to \infty$, assuming the $n_i$ are bounded,

$$V\left(S_1\right) = \frac{(2\alpha+6)\sum_{i=1}^{K} n_i^2 \, \mu^4}{\alpha^3} + \frac{(4\alpha+12)N\mu^3}{\alpha^2} + \frac{(2\alpha+7)K\mu^2}{\alpha} + \sum_{i=1}^{K} n_i^{-1}\mu + O(1) . \tag{3.12}$$

**Proof.**

$$E\left(S_1^2\right) = \frac{1}{N^2}E\left(\mathbf{Z}'\mathbf{A}_1\mathbf{Z}\mathbf{Z}'\mathbf{A}_1\mathbf{Z}\right)$$

$$= \frac{1}{N^2}\sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij}^{(1)} \sum_{k=1}^{K}\sum_{l=1}^{K} a_{kl}^{(1)} \; E\left(Z_iZ_jZ_kZ_l\right)$$

$$= \frac{1}{N^2}\left(\sum_{i=1}^{K}(a_{ii}^{(1)})^2 E\,(Z_i^4) + \sum_{i=1}^{K}\sum_{j\neq i}(a_{ii}^{(1)}a_{jj}^{(1)} + 2(a_{ij}^{(1)})^2)E\,(Z_i^2)E\,(Z_j^2)\right)$$

$$= \frac{(N^2\alpha + 6\sum_{i=1}^{K} n_i^2)\mu^4}{\alpha^3} + \frac{2(K\alpha(N - \sum_{i=1}^{K} n_i^2/N) + \alpha N + 6N)\mu^3}{\alpha^2}$$

$$+ \frac{(K^2\alpha + 7K)\mu^2}{\alpha} + \sum_{i=1}^{K} n_i^{-1}\ \mu + O(1)\ ,$$

where the last line follows from (3.9), (3.11) and using

$$\sum_{i=1}^{K} n_i^r(N - n_i)^2 \ = \ N^2\sum_{i=1}^{K} n_i^r + O(K^2) \quad \text{as the } n_i \text{ are bounded.}$$

The lemma is complete once the above result is combined with $E\,(S_1)$ given in (3.6).

•

**Lemma 3.4**

As $K \to \infty$, assuming the $n_i$ are bounded,

$$\text{Cov}\,(\hat{\mu}_1, S_1) \ = \ \frac{1}{N}\left(\frac{2\sum_{i=1}^{K} n_i^2\ \mu^3}{\alpha^2} + \frac{3N\mu^2}{\alpha} + K\mu\right) + O(K^{-1})\ . \qquad (3.13)$$

**Proof.**

$$\begin{aligned}\text{Cov}\,(\hat{\mu}_1, S_1) &= E\,((\hat{\mu}_1 - \mu)S_1) \\ &= E\left(\frac{1}{N}\sum_{i=1}^{K} n_i Z_i\ .\ \frac{1}{N}\mathbf{Z}'\mathbf{A}_1\mathbf{Z}\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{K} n_i a_{ii}^{(1)} E\,(Z_i^3)\ ,\end{aligned}$$

and hence the result follows from (3.10).

•

**Lemma 3.5**

For bounded $n_i$ we have, as $K \to \infty$,

$$\frac{1}{K^s}(\hat{\mu}_1 - \mu)^r(S_1 - E\,(S_1))^s \ = \ O_p(K^{-(r+s)/2})\ , \qquad (3.14)$$

where $r$ and $s$ are non-negative integers. (See Appendix A for definition of $o_p$ and $O_p$ notation.)

**Proof.**

Let $W_i = n_i\,(Z_i^2 - E\,(Z_i^2))$ . Then from equations (3.4) and (3.5) and from Lemma 3.1 we have

$$\begin{aligned} & S_1 - E\,(S_1) \\ &= \sum_{i=1}^{K} n_i \left(\bar{Y}_{i\cdot} - \frac{1}{N}\sum_{i'=1}^{K} n_{i'}\bar{Y}_{i'\cdot}\right)^2 - \frac{1}{N}\sum_{i=1}^{K} n_i(N-n_i)E(Z_i^2) \\ &= \sum_{i=1}^{K} n_i \left(Z_i - \frac{1}{N}\sum_{i'=1}^{K} n_{i'}Z_{i'}\right)^2 - \frac{1}{N}\sum_{i=1}^{K} n_i(N-n_i)E(Z_i^2) \\ &= \frac{1}{N}\left(N\sum_{i=1}^{K} W_i + C - \left(\sum_{i=1}^{K} n_i Z_i\right)^2\right), \end{aligned} \tag{3.15}$$

where $C = \sum_{i=1}^{K} n_i^2\ E\,(Z_i^2) = \sum_{i=1}^{K}(n_i\mu + n_i^2\mu^2/\alpha) = O(K)$.

Letting $U_i = n_i Z_i$ then gives

$$\begin{aligned} & \frac{1}{K^s} E\ (\hat{\mu}_1 - \mu)^r (S_1 - E\,(S_1))^s \\ &= \frac{1}{K^s N^{r+s}} E\left(\sum_{i=1}^{K} U_i\right)^r \left(N\sum_{i=1}^{K} W_i + C - \left(\sum_{i=1}^{K} U_i\right)^2\right)^s \\ &= \sum_{j=0}^{s} \frac{N^s}{K^s N^{r+s}} \binom{s}{j} \sum_{l=0}^{j} (-1)^l \binom{j}{l} N^{-j} C^l\ E\left(\sum_{i=1}^{K} U_i\right)^{r+2(j-l)} \left(\sum_{i=1}^{K} W_i\right)^{s-j}. \end{aligned}$$

As the $U_i$ $(i = 1, \cdots, K)$ are independent random variables with zero means and absolute moments of all order bounded uniformly in $i$ and $K$, the $W_i$ also have zero means and have absolute moments of all order, again bounded uniformly in $i$ and $K$. Therefore, from Theorem B1 in Appendix B and using $N = O(K)$, we have

$$\frac{1}{K^s} E\ (\hat{\mu}_1 - \mu)^r (S_1 - E\,(S_1))^s \ = \ O(K^{[(r+s)/2]-(r+s)}),$$

where $[x]$ denotes the integer part of $x$.

Thus the mean and variance of $\frac{1}{K^s}\,(\hat{\mu}_1 - \mu)^r (S_1 - E\,(S_1))^s$ are $O(K^{-(r+s)+[(r+s)/2]})$ and $O(K^{-(r+s)})$ respectively and hence the lemma follows from Chebyshev's Inequality.

•

We now introduce some further notation. Let

$$\begin{aligned} m &= N - \sum_{i=1}^{K} n_i^2/N \ , \\ \sigma &= (K-1)\mu + \frac{m\mu^2}{\alpha} \ , \quad \text{and} \\ f(x,y) &= \frac{mx^2}{y-(K-1)x} \ . \end{aligned} \tag{3.16}$$

Note: $\sigma = E(S_1)$.

Define

$$\hat{\alpha}_1^* = \alpha + f_x'(\mu,\sigma).(\hat{\mu}_1 - \mu) + f_y'(\mu,\sigma).(S_1 - \sigma) \ , \tag{3.17}$$

where $f_x'(\mu,\sigma) = \partial f(x,y)/\partial x$ evaluated at $x = \mu, \ \ y = \sigma$ and similarly for $f_y'(\mu,\sigma)$.

**Lemma 3.6**

If the $n_i$ are bounded and $\alpha$ is not zero, we have

$$\hat{\alpha}_1^* = \hat{\alpha}_1 + O_p(K^{-1}) \ .$$

**Proof.**

Firstly, for $K > 1$,

$$\begin{aligned} \frac{m}{K-1} &= \frac{1}{N(K-1)} \sum_{i \neq j} n_i n_j \\ \Rightarrow \ \left(\max_i\{n_i\}\right)^{-1} &\leq \frac{m}{K-1} \leq \left(\max_i\{n_i\}\right)^2 \ . \end{aligned} \tag{3.18}$$

Now, since $\hat{\alpha}_1 = f(\hat{\mu}_1, S_1)$, from Taylor's theorem we have,

$$\hat{\alpha}_1^* - \hat{\alpha}_1 = \frac{1}{2} \sum_{j=0}^{2} D^{(2-j,j)} f(\mu^*,\sigma^*).(\hat{\mu}_1 - \mu)^{2-j}(S_1 - \sigma)^j \ , \tag{3.19}$$

where

$$D^{(2-j,j)} f(\mu^*,\sigma^*) = \left[\frac{\partial^2 f(x,y)}{\partial x^{2-j} \partial y^j}\right]_{x=\mu^*, y=\sigma^*} \quad ; \ j = 0,1,2.$$

for some $\mu^*, \sigma^*$ satisfying $|\mu^* - \mu| \leq |\hat{\mu}_1 - \mu|$ and $|\sigma^* - \sigma| \leq |S_1 - \sigma|$ . Then letting

$$\mathcal{S} = \{(x,y) : |\mu - x| \leq \tfrac{1}{4} \tfrac{m}{K-1} \tfrac{\mu^2}{\alpha} \ , \ |\tfrac{\sigma}{K-1} - \tfrac{y}{K-1}| \leq \tfrac{1}{4} \tfrac{m}{K-1} \tfrac{\mu^2}{\alpha}\} \ ,$$

we have, for all $(x, y) \in \mathcal{S}$,

$$\begin{aligned} D^{(2-j,j)} f(x,y) &= \frac{1}{(K-1)^j} \frac{\partial^2 \left( \frac{m}{K-1} x^2 / \left( \frac{y}{K-1} - x \right) \right)}{\partial x^{2-j} \partial (\frac{y}{K-1})^j} \\ &= \frac{1}{K^j} \frac{O(1)}{(\frac{y}{K-1} - x)^4} \quad ; \; j = 0, 1, 2. \end{aligned}$$

Since for all $(x, y) \in \mathcal{S}$,

$$\begin{aligned} \frac{y}{K-1} &\geq \frac{\sigma}{K-1} - \frac{1}{4} \frac{m}{(K-1)} \frac{\mu^2}{\alpha} \\ &= \mu + \frac{3}{4} \frac{m}{(K-1)} \frac{\mu^2}{\alpha} \end{aligned}$$

and

$$x \leq \mu + \frac{1}{4} \frac{m}{(K-1)} \frac{\mu^2}{\alpha}$$

we have

$$\frac{y}{K-1} - x \geq \frac{1}{2} \frac{m}{(K-1)} \frac{\mu^2}{\alpha}$$

for all $(x, y) \in \mathcal{S}$ and hence, since $\frac{m}{K-1}$ is bounded, $D^{(2-j,j)} f(x, y) = O(K^{-j})$ for all $(x, y) \in \mathcal{S}$.

So, since $E\,(\hat{\mu}_1 - \mu)^2$ and $E\,\left( \frac{1}{K-1} (S_1 - \sigma) \right)^2$ are both $O_p(K^{-1})$ (see Lemma 3.5) we have

$$\begin{aligned} & P\,((\hat{\mu}_1, S_1) \in \mathcal{S}) \\ & \geq 1 - P(|\hat{\mu}_1 - \mu| > \tfrac{1}{4} \tfrac{m}{K-1} \tfrac{\mu^2}{\alpha}) - P(\tfrac{1}{K-1} |S_1 - \sigma| > \tfrac{1}{4} \tfrac{m}{K-1} \tfrac{\mu^2}{\alpha}) \\ & = 1 - O(K^{-1}) \,, \end{aligned}$$

and hence obtain

$$D^{(2-j,j)} f(\mu^\star, \sigma^\star) = O_p(K^{-j}). \tag{3.20}$$

The lemma now follows from (3.19), (3.20) and (3.14).

•

**Theorem 3.2**

**If the $n_i$ are bounded and $\alpha$ is not zero,**

$$\begin{aligned}
\text{Cov}\,(\sqrt{K}\hat{\mu}_1, \sqrt{K}\hat{\alpha}_1^*) &= \frac{K^2\alpha}{N^3}\sum_{i=1}^{K} n_i^2 - \frac{K\alpha}{N} + O(K^{-1}) \,, \quad \text{and} \\
V\,(\sqrt{K}\hat{\alpha}_1^*) &= \frac{2\alpha(\alpha+1)K}{N^2}\sum_{i=1}^{K} n_i^2 + \frac{4(\alpha+1)\alpha^2 K}{N\mu} + \frac{K^3\alpha^3}{N^4\mu^2}\sum_{i=1}^{K} n_i^2 \\
&\quad + \frac{(2\alpha+1)K^2\alpha^3}{N^2\mu^2} + \frac{\alpha^4 K}{N^2\mu^3}\sum_{i=1}^{K} n_i^{-1} - \frac{K^3\alpha^4}{N^3\mu^3} + O(K^{-1}) \,.
\end{aligned}$$

**Proof.**

The result follows, after substantial algebra, from

$$\begin{aligned}
V\,(\hat{\alpha}_1^*) &= \left(f_x'(\mu,\sigma)\right)^2 V(\hat{\mu}_1) \\
&\quad + 2f_x'(\mu,\sigma)f_y'(\mu,\sigma)\,\text{Cov}\,(\hat{\mu}_1, S_1) + \left(f_y'(\mu,\sigma)\right)^2 V\,(S_1) \,, \\
\text{Cov}\,(\hat{\mu}_1, \hat{\alpha}_1^*) &= f_x'(\mu,\sigma)\,V\,(\hat{\mu}_1) + f_y'(\mu,\sigma)\,\text{Cov}\,(\hat{\mu}_1, S_1) \,,
\end{aligned}$$

where

$$f_x'(\mu,\sigma) = \frac{\alpha}{m\mu^2}\left(2m\mu + (K-1)\alpha\right) \,, \tag{3.21}$$

$$f_y'(\mu,\sigma) = -\frac{\alpha^2}{m\mu^2} \,, \tag{3.22}$$

$$m^{-1} = N^{-1} + O(K^{-2}) \,, \tag{3.23}$$

together with the results in (3.7), (3.12) and (3.13).

•

Similarly we can derive the following results for MME-2.

Let

$$\begin{aligned}
g(x,y) &= \frac{(K-1)x^2}{y - \frac{K-1}{K}(\sum_{i=1}^{K} n_i^{-1})x} \,, \\
\tau &= (K-1)\left(\mu\sum_{i=1}^{K} n_i^{-1}/K + \mu^2/\alpha\right) \,,
\end{aligned}$$

(that is, $\tau = E\,S_2$ ) and define

$$\hat{\alpha}_2^* = \alpha + g_x'(\mu,\tau).(\hat{\mu}_1 - \mu) + g_y'(\mu,\tau).(S_2 - \tau) \,,$$

where $g_x'(\mu,\tau) = \partial g(x,y)/\partial x$ evaluated at $x=\mu,\ \ y=\tau$ and similarly for $g_y'(\mu,\sigma)$.

**Lemma 3.7**

Provided the $n_i$ are bounded,

$$V\ (S_2) \ = \ \frac{2K(\alpha+3)\mu^4}{\alpha^3} + \frac{4(\alpha+3)\mu^3}{\alpha^2}\sum_{i=1}^{K} n_i^{-1} + \\ + \frac{(2\alpha+7)\mu^2}{\alpha}\sum_{i=1}^{K} n_i^{-2} + \mu\sum_{i=1}^{K} n_i^{-3} + O(1)\ , \qquad (3.24)$$

$$\text{Cov}\ (\hat{\mu}_2, S_2) \ = \ \frac{2\mu^3}{\alpha^2} + \frac{3\mu^2}{K\alpha}\sum_{i=1}^{K} n_i^{-1} + \frac{\mu}{a}\sum_{i=1}^{K} n_i^{-2} + O(K^{-1})\ . \qquad (3.25)$$

$$\frac{1}{K^s}(\hat{\mu}_2-\mu)^r(S_2 - E\ (S_2))^s \ = \ O_p(K^{-(r+s)/2})\ , \qquad (3.26)$$

$$\hat{\alpha}_2^* \ = \ \hat{\alpha}_2 + O_p(K^{-1})\ . \qquad (3.27)$$

As the proof proceeds in exactly the same manner as for the proofs of Lemmas 3.3 to 3.6 it is not given.

•

**Theorem 3.3**

**If the $n_i$ are bounded and $\alpha$ is not zero**

$$\text{Cov}\ (\sqrt{K}\hat{\mu}_2, \sqrt{K}\hat{\alpha}_2^*) \ = \ \frac{\alpha^2}{\mu}\left(N_{K,-1}^2 - N_{K,-2}\right) + O(K^{-1})\ , \ \text{and}$$

$$V\ (\sqrt{K}\hat{\alpha}_2^*) \ = \ 2(\alpha+1)\alpha + \frac{4(\alpha+1)\alpha^2 N_{K,-1}}{\mu} \\ + \frac{\alpha^3\left((3+2\alpha)N_{K,-2} - N_{K,-1}^2\right)}{\mu^2} \\ + \frac{\alpha^4\left(N_{K,-1}^3 - 2N_{K,-1}\ N_{K,-2} + N_{K,-3}\right)}{\mu^3} + O(K^{-1})\ ,$$

where

$$N_{K,r} \ = \ \frac{1}{K}\sum_{i=1}^{K} n_i^r. \qquad (3.28)$$

Again, as the proof corresponds to the proof of Theorem 3.2 but using the results in Lemma 3.7 instead of the results in Lemmas 3.3 to 3.6, the proof is not given.

•

As a consequence of the above results we have the following corollary.

**Corollary 3.1**

For the case $n_i = n$ for all $i$ the two sets of moment estimators are equivalent with

$$\begin{aligned} V\left(\sqrt{K}\hat{\mu}_j\right) &= \frac{\mu}{n} + \frac{\mu^2}{\alpha}\,, \\ \text{Cov}\left(\sqrt{K}\hat{\mu}_j, \sqrt{K}\hat{\alpha}_j^*\right) &= 0 + O(K^{-1})\,, \\ V\left(\sqrt{K}\hat{\alpha}_j^*\right) &= \frac{2\alpha(\alpha+1)(n\mu+\alpha)^2}{n^2\mu^2} + O(K^{-1})\,, \end{aligned} \tag{3.29}$$

for $j = 1$ and 2.

•

To conclude this section we obtain the limiting distribution of the two moment estimators of $\alpha$ when both $\mu$ and $\alpha$ are finite and not zero.

**Theorem 3.4**

If both $\mu$ and $\alpha$ are finite and non zero and the $n_i$ $(i = 1, \cdots, K)$ are uniformly bounded, then as $K \to \infty$ ,

$$\sqrt{K}(\hat{\alpha}_i - \alpha) \text{ is asymptotically normal for } i = 1, 2.$$

**Proof.**

Let $a_1 = \frac{K}{N} f'_x(\mu, \sigma)$ and $a_2 = K\, f'_y(\mu, \sigma)$, where $f(x, y)$ is defined in (3.16). From (3.21), (3.22) and (3.23) we therefore have that $a_1$ and $a_2$ are constants uniformly bounded for all $K$.

Then using (3.17), $\hat{\mu}_1 - \mu = \frac{1}{N}\sum_{i=1}^{K} n_i Z_i$ and (3.15) we obtain

$$\begin{aligned}
&\sqrt{K}\,(\hat{\alpha}_1^* - \alpha) \\
&= \frac{1}{\sqrt{K}} \sum_{i=1}^{K} \left( a_1\, n_i\, Z_i + a_2\, n_i\, Z_i^2 - a_2\, n_i\, E\, Z_i^2 \right) \\
&\qquad + \frac{a_2\, C}{\sqrt{K}\, N} - \frac{a_2}{\sqrt{K}\, N} \left( \sum_{i=1}^{K} n_i Z_i \right)^2 \qquad (3.30) \\
&= \frac{1}{\sqrt{K}} \sum_{i=1}^{K} \left( a_1\, n_i\, Z_i + a_2\, n_i\, Z_i^2 - a_2\, n_i\, E\, Z_i^2 \right) + o_p(1)\,,
\end{aligned}$$

since $C = O(K)$ and $\sum_{i=1}^{K} n_i\, Z_i = N(\hat{\mu}_1 - \mu) = O_p(\sqrt{K})$.

Now since $Z_i = \frac{1}{n_i}(Y_{i\cdot} - n_i \mu)$ where the $Y_{i\cdot}$ are independent $NBD(n_i\, \mu, \alpha)$ random variables for which all moments exist, we have letting

$$T_i \;=\; \left( a_1\, n_i\, Z_i + a_2\, n_i\, Z_i^2 - a_2\, n_i\, E\, Z_i^2 \right),$$

that $E\, T_i = 0\ \forall\ i$ whilst $E\, T_i^2$ and $E\, T_i^4$ are constants uniformly bounded for all $i$. This latter result holds since $|a_1|, |a_2|$ and $n_i$ are uniformly bounded (uniformly in $i$) with $n_i \geq 1$ and the moments of $Y_{i\cdot}$ , which are functions of $n_i, \mu$ and $\alpha$, are also uniformly bounded .

Therefore

$$\frac{\sum_{i=1}^{K} E \left\{ \frac{1}{\sqrt{K}} \left( a_1\, n_i\, Z_i + a_2\, n_i\, Z_i^2 - a_2\, n_i\, E\, Z_i^2 \right) \right\}^4}{\left\{ \sum_{i=1}^{K} E \left\{ \frac{1}{\sqrt{K}} \left( a_1\, n_i\, Z_i + a_2\, n_i\, Z_i^2 - a_2\, n_i\, E\, Z_i^2 \right) \right\}^2 \right\}^2} \;=\; o(1)\,,$$

and hence, from Liapounov's theorem, $\sqrt{K}(\hat{\alpha}_1^* - \alpha)$ is asymptotically normal. That $\sqrt{K}(\hat{\alpha}_1 - \alpha)$ is asymptotically normally distributed then follows from Lemma 3.6 and Slutsky's theorem.

The corresponding result for MME-2 can be obtained similarily.

•

## 3.4 Asymptotic comparison of the estimates.

Although the formulae above for the variances and covariances of the different estimators provide a succinct summary of their asymptotic ($K \to \infty$) performance they

unfortunately give little insight, as they stand, into their relative performance. Following Anscombe (1950) we provide contour plots comparing the asymptotic efficiencies of the different estimators. As for Anscombe (1950), the large sample efficiency is taken as the square root of the ratio of the generalised variance for the maximum likelihood estimators and for the alternative estimators (MME-1 or MME-2), where the generalised variance is the determinant of the variance-covariance matrix for the limiting distribution.

Here however, unlike the situation considered by Anscombe (1950), there is the added complication of an unlimited choice of possible values for $n_i$, $i = 1, \cdots, K$. To overcome this we consider only the case where the $n_i$ are *iid* random variables, independent of the $Y_{ij}$, and restrict ourselves to two forms of distributions for $n_1$. One form will have $n_1 - 1$ distributed uniformly on the integer values between 0 and $a$ (inclusive), the other with $n_1 - 1$ distributed as a truncated Poisson with untruncated mean $\nu$, truncating values of 20 and above. Thus in both cases the $n_i$ are bounded random variables.

Before presenting the graphical comparisons though we need the following results.

For $r = -1, 1$ and 2 let

$$\begin{aligned} \theta_r &= E\left(n_1^r\right), \\ \underline{\theta} &= (\theta_{-1}, \theta_1, \theta_2)', \quad \text{and} \\ \underline{\hat{\theta}}_K &= (N_{K,-1}, N_{K,1}, N_{K,2})', \end{aligned}$$

where $N_{K,r}$ is defined in (3.28). Now if

$$V\left(\sqrt{K}\hat{\mu}_1\right) V\left(\sqrt{K}\hat{\alpha}_1^*\right) - \left(Cov\left(\sqrt{K}\hat{\mu}_1, \sqrt{K}\hat{\alpha}_1^*\right)\right)^2 = h_1(\underline{\hat{\theta}}_K) + O(K^{-1})$$

then

$$\begin{aligned} h_1(\underline{\hat{\theta}}_K) = & \left(\frac{\mu}{N_{K,1}} + \frac{\mu^2 N_{K,2}}{\alpha N_{K,1}^2}\right)\left(\frac{2\alpha(\alpha+1) N_{K,2}}{N_{K,1}^2} + \frac{4(\alpha+1)\alpha^2}{\mu N_{K,1}} + \frac{\alpha^3 N_{K,2}}{\mu^2 N_{K,1}^4}\right. \\ & \left. + \frac{(2\alpha+1)\alpha^3}{\mu^2 N_{K,1}^2} + \frac{\alpha^4 N_{K,-1}}{\mu^3 N_{K,1}^2} - \frac{\alpha^4}{\mu^3 N_{K,1}^3}\right) - \left(\frac{\alpha N_{K,2}}{N_{K,1}^3} - \frac{\alpha}{N_{K,1}}\right)^2. \end{aligned}$$

(**Note:** Here, and below, $\hat{\mu}_1, \hat{\alpha}_1$ and $\hat{\alpha}_1^*$ are functions of $K$ but we supress the extra subscript for clarity.)

**Theorem 3.5**

If the $n_i$ , $i = 1, \cdots, K$ , are independent, identically distributed, uniformly bounded random variables we have

$$\mathcal{GV}\left(\lim_{K\to\infty}\sqrt{K}\hat{\mu}_1, \lim_{K\to\infty}\sqrt{K}\hat{\alpha}_1\right) = h_1(\underline{\theta}) ,$$

where $\mathcal{GV}$ denotes the generalized variance.

**Proof.**

Let $X_K$ equal $V\ (\sqrt{K}\hat{\alpha}_1^*)$ omitting terms of order $K^{-1}$, that is

$$X_K = \frac{2\alpha(\alpha+1)N_{K,2}}{N_{K,1}^2} + \frac{4(\alpha+1)\alpha^2}{\mu\ N_{K,1}} + \frac{\alpha^3 N_{K,2}}{\mu^2\ N_{K,1}^4} + \frac{(2\alpha+1)\alpha^3}{\mu^2\ N_{K,1}^2} + \frac{\alpha^4 N_{K,-1}}{\mu^3\ N_{K,1}^2} - \frac{\alpha^4}{\mu^3\ N_{K,1}^3}$$

and

$$X = \frac{2\alpha(\alpha+1)\theta_2}{\theta_1^2} + \frac{4(\alpha+1)\alpha^2}{\mu\ \theta_1} + \frac{\alpha^3\theta_2}{\mu^2\ \theta_1^4} + \frac{(2\alpha+1)\alpha^3}{\mu^2\ \theta_1^2} + \frac{\alpha^4\theta_{-1}}{\mu^3\ \theta_1^2} - \frac{\alpha^4}{\mu^3\ \theta_1^3} .$$

Then

$$\begin{aligned}
& V\ (\lim_{K\to\infty}\sqrt{K}\hat{\alpha}_1) \\
&= V\ (\lim_{K\to\infty}\sqrt{K}\hat{\alpha}_1^*) && \text{(by Lemma 3.6)} \\
&= \lim_{K\to\infty} V\ (\sqrt{K}\hat{\alpha}_1^*) && \text{(See Comment 1 below)} \\
&= \lim_{K\to\infty}\left[E\ V\ (\sqrt{K}\hat{\alpha}_1^* \mid \underline{\hat{\theta}}) + V\ E(\sqrt{K}\hat{\alpha}_1^* \mid \underline{\hat{\theta}})\right] \\
&= \lim_{K\to\infty}\left[E\ \left(X_K + O(K^{-1})\right) + 0\right] \\
&= \lim_{K\to\infty} E\ (X_K) \\
&= E\ (X) && \text{(See Comment 2 below)} \\
&= X\ .
\end{aligned}$$

**Comment 1:** Using an expansion for $\sqrt{K}\hat{\alpha}_1^*$ based on (3.30) we can show, by first conditioning on the $n_i$ and then using the fact that the $n_i$ are uniformly bounded,

that all moments of $\sqrt{K}\hat{\alpha}_1^*$, in particular the third moment, are bounded uniformly in K. Hence $V\left(\lim_{K\to\infty}\sqrt{K}\hat{\alpha}_1^*\right) = \lim_{K\to\infty} V\left(\sqrt{K}\hat{\alpha}_1^*\right)$ using Theorem 4.5.2 of Chung (1974).

**Comment 2:** Since the $N_{K,r} \xrightarrow{p} \theta_r \ \forall\ r$ we have, by the corollary of Theorem 2.3.3 of Lukacs (1975) that $X_K \xrightarrow{p} X$. Also, as the $X_K$ are absolutely bounded uniformly in $K$ (which follows since the $n_i$ , $i = 1, \cdots, K$ are uniformly bounded and hence $1 \leq N_{K,r} \leq (\max_i n_i)^r$) we have $\sup_K E\ |X_K|^p < \infty\ \forall\ p$. Thus from Theorem 4.5.2 of Chung (1974), $\lim_{K\to\infty} E\ X_K^p = E\ X^p < \infty$ for all positive integer values of $p$.

In precisely the same way we can also show that

$$\begin{aligned} V\left(\lim_{K\to\infty}\sqrt{K}\hat{\mu}_1\right) &= \frac{\mu}{\theta_1} + \frac{\mu^2\ \theta_2}{\alpha\ \theta_1^2}\ , \\ Cov\left(\lim_{K\to\infty}\sqrt{K}\hat{\mu}_1, \lim_{K\to\infty}\sqrt{K}\hat{\alpha}_1\right) &= \frac{\alpha\ \theta_2}{\theta_1^3} - \frac{\alpha}{\theta_1}\ , \end{aligned}$$

from which the result given in the theorem immediately follows.

•

**Theorem 3.6**

$$\begin{aligned} &\mathcal{GV}\left(\lim_{K\to\infty}\sqrt{K}\hat{\mu}_2, \lim_{K\to\infty}\sqrt{K}\hat{\alpha}_2\right) = \\ &\left(\mu\ \theta_{-1} + \frac{\mu^2}{\alpha}\right)\left(2(\alpha+1)\alpha + \frac{4(\alpha+1)\alpha^2\ \theta_{-1}}{\mu} + \frac{\alpha^3\left((3+2\alpha)\ \theta_{-2} - \theta_{-1}^2\right)}{\mu^2} + \right. \\ &\left. + \frac{\alpha^4\left(\theta_{-1}^3 - 2\theta_{-1}\ \theta_{-2} + \theta_{-3}\right)}{\mu^3}\right) - \frac{\alpha^4}{\mu^2}\left(\theta_{-1}^2 - \theta_{-2}\right)^2\ , \end{aligned}$$

with the result following in precisely the same manner as the previous theorem.

•

For the MLEs we have from (3.3) for given $n_i$, $i = 1, 2, \cdots$ that $[K^{1/2}(\hat{\mu} - \mu), K^{1/2}(\hat{\alpha} - \alpha)]$ has limiting distribution that is bivariate normal with zero means and variance-covariance matrix

$$\lim_{K\to\infty}\begin{pmatrix} \left(\frac{1}{K}\sum_{i=1}^K \frac{\alpha}{\mu^2}\left(\frac{n_i\mu}{n_i\mu+\alpha}\right)\right)^{-1} & 0 \\ 0 & \left(\frac{1}{K}\sum_{i=1}^K \sum_{r=2}^{\infty} \frac{r!\,\Gamma(\alpha)}{r^2\,\Gamma(\alpha+r)}\left(\frac{n_i\mu}{n_i\mu+\alpha}\right)^r\right)^{-1} \end{pmatrix}.$$

Since the limiting variance-covariance matrix is positive definite for all sequences $\{n_i\}$ we therefore have, by the strong law of large numbers, that as $K \to \infty$, $[K^{1/2}(\hat{\mu} - \mu), K^{1/2}(\hat{\alpha} - \alpha)]$ has unconditional limiting distribution that is bivariate normal with zero means and variance-covariance matrix given by

$$\begin{pmatrix} \left\{\frac{\alpha}{\mu^2} E\left(\frac{n_1^2 \mu}{n_1 \mu + \alpha}\right)\right\}^{-1} & 0 \\ 0 & \left\{\sum_{r=2}^{\infty} \frac{r!\,\Gamma(\alpha)}{r^2\,\Gamma(\alpha+r)} E\left(\frac{n_1 \mu}{n_1 \mu + \alpha}\right)^r\right\}^{-1} \end{pmatrix}.$$

From here the $\mathcal{GV}$ for the MLEs is readily calculated.

For the contour plots of the efficiencies we now take for the $a$ and $\nu$ parameters of the uniform and truncated Poisson distributions for $n_1 - 1$ the following values.

| | | | | $V(n_1)$ | |
|---|---|---|---|---|---|
| Figure | $a$ | $\nu$ | $E(n_1)$ | Uniform | Trunc. Poisson |
| 1 | 1 | 0.5 | 1.5 | 0.25 | 0.50 |
| 2 | 2 | 1.0 | 2.0 | 0.67 | 1.00 |
| 3 | 10 | 5.0 | 6.0 | 10.00 | 5.00 |

**Notes:**

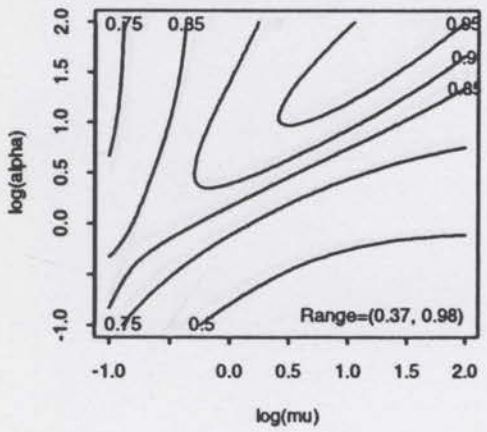
(i) Here $E\, n_1$ is not exact for the truncated Poisson but is correct to the decimal places reported; and

(ii) for the case $a = 0$, which corresponds to $n_i = 1\ \forall\, i$, the two moment estimators are identical. This situation corresponds to the standard negative binomial distribution, studied by Anscombe (1950), and hence is not considered here.

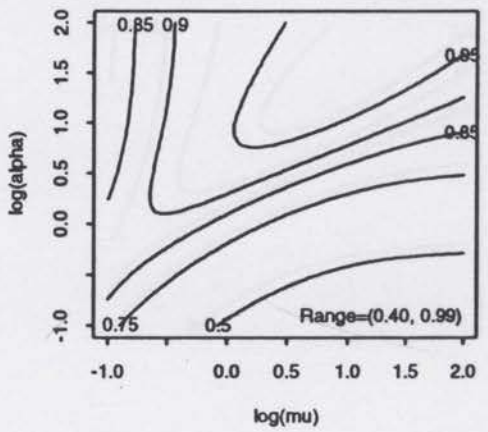
Each of the three figures (Figure 3.1, Figure 3.2 and Figure 3.3) comprises six plots in two columns of three, with the column on the left for the plots associated with the truncated Poisson distribution for $n_1 - 1$. The column on the right corresponds to the uniform distribution for $n_1 - 1$. Within each column the top two plots give relative efficiency contours for the MME-1 and MME-2 estimators respectively whilst the bottom plot gives contours for the ratio of the relative efficiency of MME-2 to MME-1. In addition, in the bottom right hand corner of each plot is given the (minimum, maximum) that the plotted value took on the region $-1 \leq \log_{10}(\mu), \log_{10}(\alpha) \leq 2$.

## Figure 3.1: Relative efficiencies when average n = 1.5

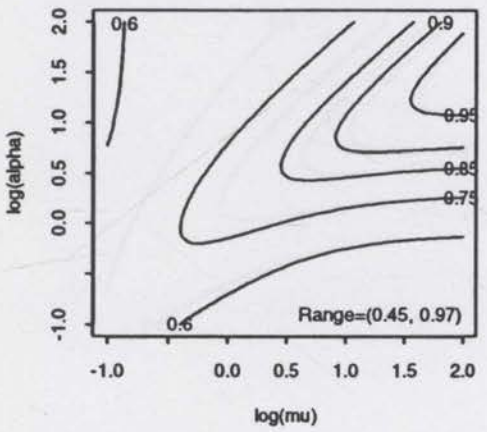
MME-1 for truncated Poisson.



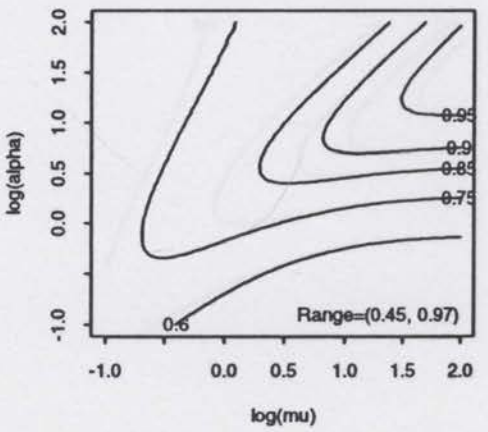

MME-1 for uniform.



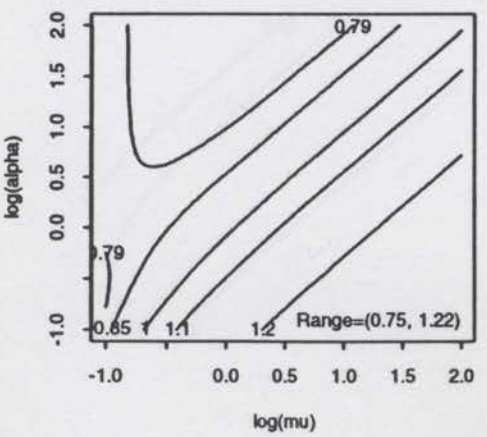

MME-2 for truncated Poisson.



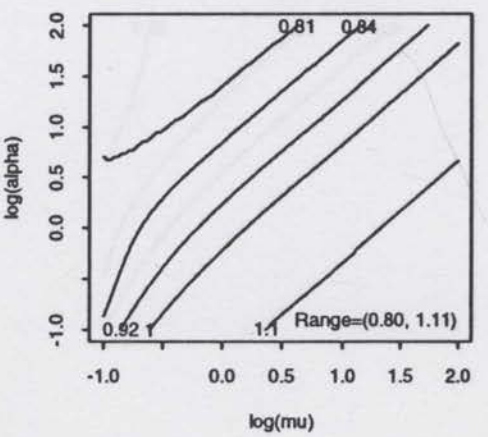

MME-2 for uniform.




MME-2 / MME-1 for truncated Poisson.

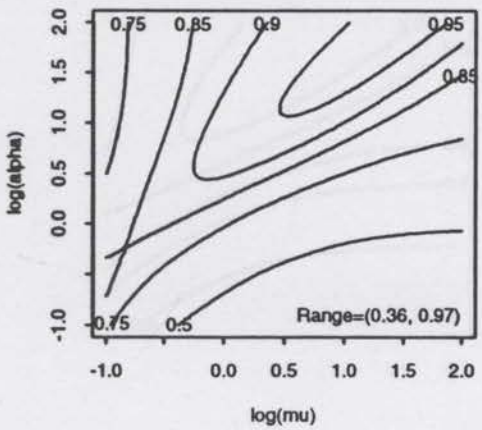



MME-2 / MME-1 for uniform.

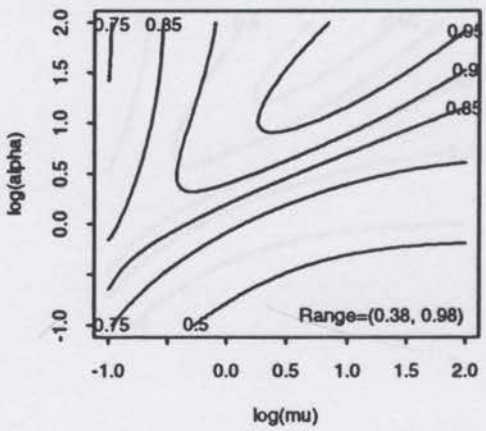


## Figure 3.2: Relative efficiencies when average n = 2

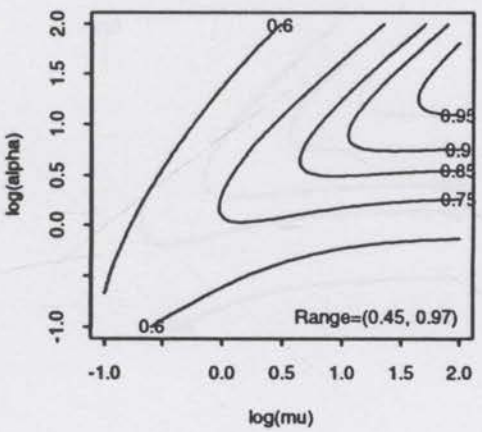
MME-1 for truncated Poisson.



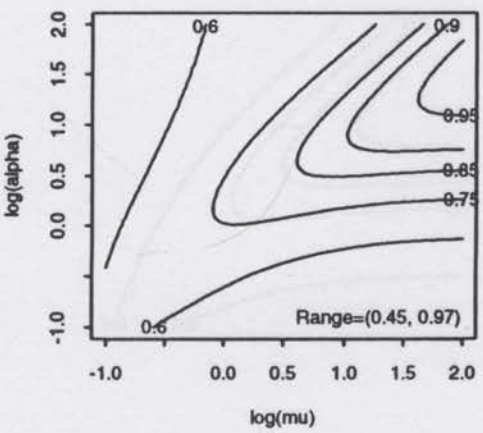

MME-1 for uniform.



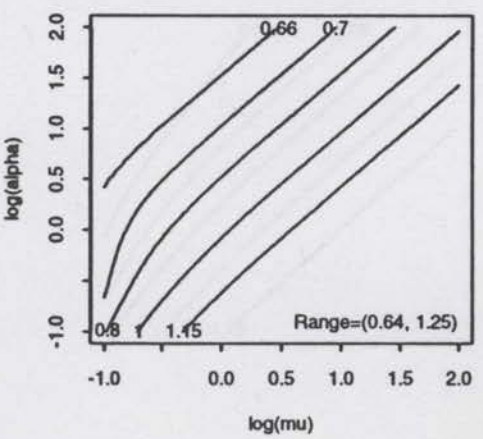

MME-2 for truncated Poisson.




MME-2 for uniform.




MME-2 / MME-1 for truncated Poisson.




MME-2 / MME-1 for uniform.

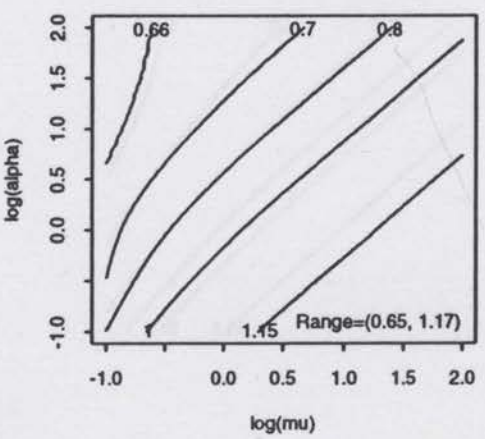
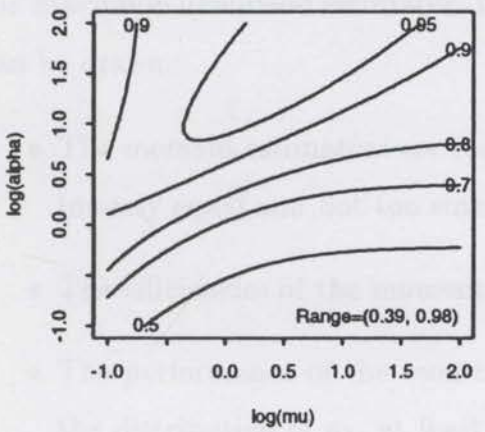
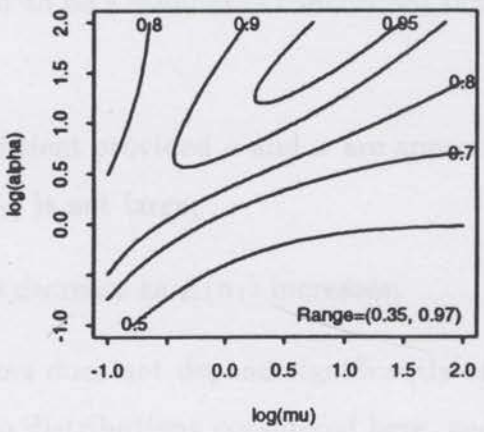


Figure 3.3: Relative efficiencies when average n = 6

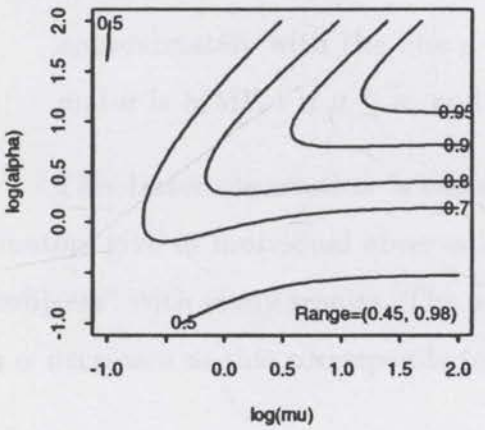
MME-1 for truncated Poisson.



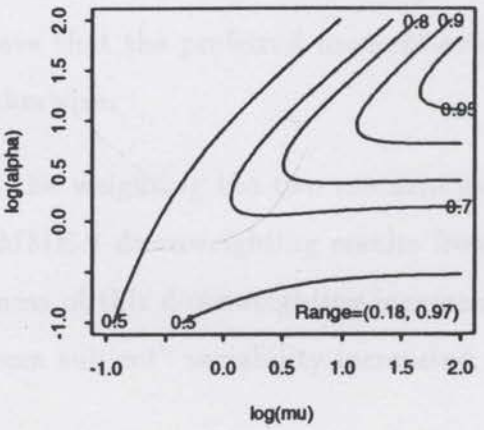

MME-1 for uniform.



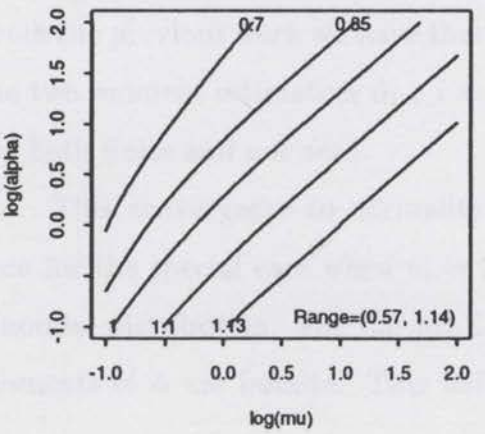

MME-2 for truncated Poisson.



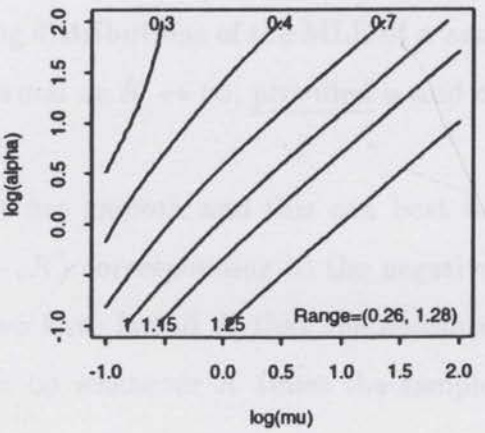

MME-2 for uniform.




MME-2 / MME-1 for truncated Poisson.




MME-2 / MME-1 for uniform.



From this limited graphical comparison of the two moment estimators with the maximum likelihood estimators there appear to be a number of conclusions that can be drawn.

- The moment estimators are reasonably efficient provided $\mu$ and $\alpha$ are approximately equal and not too small and $E(n_1)$ is not large;
- The efficiencies of the moment estimators decrease as $E(n_1)$ increases;
- The performance of the moment estimators does not depend significantly on the distribution of $n_1$, at least for the two distributions considered here, and particularily so for small $E(n_1)$ (in which case the two estimators are probably almost equivalent);
- From the bottom plots in each figure, for which the unit contour corresponds approximately with the line $\mu = \alpha$, we have that the preferred moment estimator is MME-1 if $\mu \leq \alpha$, and MME-2 otherwise.

This latter observation is consistent with the weighting the two moment estimators give to individual observations, with MME-2 downweighting results from "subjects" with many results. The appropriateness of this downweighting increases as $\alpha$ decreases as this corresponds to the "between subject" variability increasing.

## 3.5 Erratic behaviour of the moment estimates for $\alpha$.

From the previous work we have that the limiting distributions of the MLE of $\alpha$ and the two moment estimators $\hat{\alpha}_i$ , $i = 1, 2$ are normal as $K \to \infty$, provided $\mu$ and $\alpha$ are both finite and not zero.

This convergence to normality is however not smooth and this can best be seen for the special case when $n_i = 1$ $(i = 1, \cdots, K)$ corresponding to the negative binomial distribution. For the MLE of $\alpha$ $(\hat{\alpha})$ we have for all $K$ that the absolute moments of $\hat{\alpha}$ are infinite. This follows as $\hat{\alpha} = \infty$ whenever $K$ times the sample

mean exceeds $(K-1)$ times the sample variance, an event that occurs with positive probability. For the moment estimator on the other hand we obtain very large, and possibly infinite, estimates for $\alpha$ when the absolute difference between the sample variance and sample mean is small. This tendency towards large (absolute) estimates for $\alpha$ increases as $\alpha$ increases, as will be shown in Chapter 4.

To gain some insight into the "erratic" behavior for our two moment estimators for $\alpha$ in the general case we adopt the approach of Hall (1994). (Recall that these estimators reduce to the usual moment estimator of $\alpha$ for the negative binomial distribution.) Here we consider a sequence of estimators $\hat{\alpha}_{K,i}$ $(i = 1, 2)$ depending on $K$, the number of "subjects", and allow the two parameters $(\mu_K, \alpha_K)$ to vary with $K$ (hence the additional subscripts). We then let $\alpha_K$ increase with $K$, $\mu_K/\alpha_K$ decrease with $K$, and show that the limiting distribution of $\hat{\alpha}_i$ , $i = 1, 2$ depends on the limit of $K\mu_K^2/\alpha_K^2$.

Letting $\lambda_K = K\mu_K^2/\alpha_K^2$ and $\lambda = \lim_{K\to\infty} \lambda_K$ what our main result in this section will show is that if $\lambda = \infty$ then the moment estimators have asymptotic normal distributions whilst if $\lambda$ is finite the limiting distribution is very heavy-tailed. Actually, in this latter case all moments of the limiting distribution are infinite and the smaller the value of $\lambda$ the heavier the tails of this distribution.

We now formally state the conditions under which the results in this section hold.

For $K = 1, 2, \cdots$ we have

$$\begin{aligned} \alpha_K &\uparrow \infty , \\ \mu_K &\to \mu \in (0, \infty] \text{ monotonically,} \\ \mu_K/\alpha_K &\downarrow 0 , \text{ and} \\ \lambda_K &\to \lambda \in [0, \infty] \text{ where } \lambda_K = K\mu_K^2/\alpha_K^2. \end{aligned}$$

Also, throughout this section we use the following notation:

$$\begin{aligned} m_K &= N - KN_{K,2}/N, \\ \psi_{K,1} &= 2\frac{K^2}{N^2} + \mu_K^{-1}\frac{K^2}{N^2}\left(N_{K,-1} - \frac{K}{N}\right), \end{aligned}$$

$$\psi_{K,2} = 2\,N_{K,-2} + \mu_K^{-1}\left(N_{K,-3} - 2\,N_{K,-1}\,N_{K,-2} + N_{K,-1}^3\right), \text{ and}$$

$$\psi_i = \lim_{K\to\infty} \psi_{K,i} \qquad i = 1,2\,,$$

where $N_{K,r}$ is defined in (3.28) and $n_i$ , $i = 1,2,\cdots$ is a sequence, uniformly bounded for which $\psi_i$ $(i = 1,2)$ exist.

The remainder of the notation is as for the previous sections but with an added subscript $K$, except on $N$. Here a subscript is omitted so as to avoid confusing it with $N_{K,1}$. Note that $N = KN_{K,1}$.

**Lemma 3.8**

Under the conditions specified above,

$$\frac{K}{N}\frac{1}{\sqrt{K}\,\mu_K}\sum_{i=1}^{K} n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right) \xrightarrow{\mathcal{L}} N(0,\psi_1)\,, \qquad (3.31)$$

$$\text{and} \quad \frac{1}{\sqrt{K}\,\mu_K}\sum_{i=1}^{K}\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - N_{K,-1}Z_{K,i}\right) \xrightarrow{\mathcal{L}} N(0,\psi_2)\,, \qquad (3.32)$$

as $K \to \infty$.

**Proof.**

We only show the result for the former as the proof of the latter follows in precisely the same way.

Since $E\,Z_{K,i} = 0\ \forall\ K$ we immediately have

$$E\,\frac{1}{\sqrt{K}\,\mu_K}\sum_{i=1}^{K} n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right) = 0\ \forall\ K\,,$$

whilst using (3.9), (3.10) and (3.11) we obtain

$$V\left\{\frac{1}{\sqrt{K}\,\mu_K}\sum_{i=1}^{K} n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)\right\}$$
$$= \frac{1}{K\,\mu_K^2}\sum_{i=1}^{K} n_i^2\left(E\,Z_{K,i}^4 - 2\,\frac{K}{N}E\,Z_{K,i}^3 - E^2\,Z_{K,i}^2 + \frac{K^2}{N^2}E\,Z_{K,i}^2\right)$$

$$= \left(\frac{N_{K,-1}}{\mu_K} + 3 + \frac{7}{\alpha_K} + \frac{6\,N\,\mu_K}{K\,\alpha_K} + \frac{12\,N\,\mu_K}{K\,\alpha_K^2} + \frac{3\,N_{K,2}\,\mu_K^2}{\alpha_K^2} + \frac{6\,N_{K,2}\,\mu_K^2}{\alpha_K^3}\right)$$
$$-\frac{K}{N}\left(\frac{2}{\mu_K} + \frac{6\,N}{K\,\alpha_K} + \frac{4\,N_{K,2}\,\mu_K}{\alpha_K^2}\right) - \left(1 + \frac{2\,N\,\mu_K}{K\,\alpha_K} + \frac{N_{K,2}\,\mu_K^2}{\alpha_K^2}\right)$$
$$+\frac{K^2}{N^2}\left(\frac{N}{K\,\mu_K} + \frac{N_{K,2}}{\alpha_K}\right)$$
$$= 2 + \mu_K^{-1}\left(N_{K,-1} - \frac{K}{N}\right) + o(1)\,,$$

where the last line follows as $\mu_K/\alpha_K = o(1)$ , $1/\alpha_K = o(1)$ and $N_{K,r}/N_{K,s} = O(1)\ \forall\ r, s$.

Therefore, since $\mu_K$ is bounded away from zero, we have

$$\frac{1}{K\,\mu_K^2}\sum_{i=1}^{K} E\left\{n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)\right\}^2 = O(1)\,. \tag{3.33}$$

On the other hand we have

$$\frac{1}{K^2\,\mu_K^4}\sum_{i=1}^{K} E\left\{n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)\right\}^4 = O(K^{-1})\,. \tag{3.34}$$

To see that (3.34) holds observe that $E\left\{n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)\right\}^4$ equals the sum of a <u>finite</u> number of terms of the form $c_{rs}E\,(Y_{K,i\cdot} - n_i\,\mu_K)^r.E^s(Y_{K,i\cdot} - n_i\,\mu_K)^2$ where the $c_{rs}$, which are functions of $n_i$ and $K/N$, are absolutely bounded, uniformly in $K$, for the different values of $r$ and $s$. Here the possible values of $r$ and $s$ satisfy $0 \leq r \leq 8, 0 \leq s \leq 4$ and $[r/2]+s \leq 4$. Now since $Y_{K,i\cdot} \sim NBD(n_i\mu_K, \alpha_K)$ we have from Theorem B2 in the Appendix and the relationship between the central moments and the cumulants (see for example Kendall and Stuart (1977) page 72) that $|E\,(Y_{K,i\cdot} - n_i\,\mu_K)^r| \leq \mu_k^{[r/2]}\,L_r$ for $0 \leq r \leq 8$, where $L_r$ is a constant independent of $K$. Therefore, from the conditions on $r$ and $s$, we have $|c_{rs}E\,(Y_{K,i\cdot} - n_i\,\mu_K)^r.E^s(Y_{K,i\cdot} - n_i\,\mu_K)^2| \leq \max\{1, \mu_K^4\}\,L^\star$, where $L^\star$ is independent of $K$, for all terms in the expansion of $E\left\{n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)\right\}^4$ and hence we obtain (3.34).

From (3.34) and (3.33) we therefore have

$$\frac{\sum_{i=1}^{K} E\left\{\frac{n_i}{\sqrt{K}\,\mu_K}\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)\right\}^4}{\left\{\sum_{i=1}^{K} E\left\{\frac{n_i}{\sqrt{K}\,\mu_K}\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)\right\}^2\right\}^2} \;=\; o(1)\;,$$

and hence the lemma follows from Liapounov's theorem.

•

We are now in a position to give the main result for this section.

**Theorem 3.7**

If $\hat{\alpha}_{K,i}$ , $K = 2, 3, \cdots$ is a sequence of MME-$I$ $(I = 1, 2)$ estimators for $\alpha_K$ based on $K$ "subjects" satisfying the above conditions and $\mathcal{Z}$ is a standard normal random variable then, under the assumption that the $n_i$ are bounded, we have

$$\sqrt{\frac{\psi_{K,i}}{\lambda_K}}\,\frac{\hat{\alpha}_{K,i}}{\alpha_K} \;\xrightarrow{\mathcal{L}}\; \frac{1}{\mathcal{Z}} \qquad \text{if } \lambda = 0\,, \tag{3.35}$$

$$\frac{\hat{\alpha}_{K,i}}{\alpha_K} \;\xrightarrow{\mathcal{L}}\; \frac{1}{1+\sqrt{\frac{\psi_i}{\lambda}}\,\mathcal{Z}} \qquad \text{if } \lambda \in (0,\infty)\,, \tag{3.36}$$

$$\sqrt{\frac{\lambda_K}{\psi_{K,i}}}\left(\frac{\hat{\alpha}_{K,i}-\alpha_K}{\alpha_K}\right) \;\xrightarrow{\mathcal{L}}\; \mathcal{Z} \qquad \text{if } \lambda = \infty\,, \tag{3.37}$$

for $i = 1, 2$.

**Proof.**

<u>Consider MME-1</u>

Here we have

$$\hat{\mu}_{K,1} \;=\; \mu_K\left(1 + \frac{1}{N\,\mu_K}\sum_{i=1}^{K} n_i Z_{K,i}\right)\,, \tag{3.38}$$

whilst

$$\begin{aligned} & S_{K,1} - (K-1)\hat{\mu}_{K,1} \\ &= (S_{K,1} - E(S_{K,1})) - (K-1)\,(\hat{\mu}_{K,1} - \mu_K) + (E(S_{K,1}) - (K-1)\mu_K) \\ &= \left(\sum_{i=1}^{K} n_i(Z_{K,i}^2 - E\,Z_{K,i}^2) + \frac{C_K}{N} - \frac{1}{N}\left(\sum_{i=1}^{K} n_i\,Z_{K,i}\right)^2\right) \\ &\qquad - \frac{K-1}{N}\sum_{i=1}^{K} n_i\,Z_{K,i} + \frac{m_K\,\mu_K^2}{\alpha_K} \end{aligned}$$

from (3.15) and where $C_K = \sum_{i=1}^{K} E\,(n_i^2\, Z_{K,i}^2)$

$$= \frac{m_K\,\mu_K^2}{\alpha_K}\left\{1+\frac{\alpha_K}{m_K\,\mu_K^2}\sum_{i=1}^{K} n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)+\right.$$

$$\left.+\frac{\alpha_K}{m_K\,\mu_K^2}\,\frac{1}{N}\sum_{i=1}^{K} n_i\,Z_{K,i} + \frac{\alpha_K}{m_K\,N\,\mu_K^2}\left(C_K - \left(\sum_{i=1}^{K} n_i Z_{K,i}\right)^2\right)\right\}. \tag{3.39}$$

Now

$$E\sum_{i=1}^{K}(n_i\,Z_{K,i}) \;=\; 0 \quad \text{and}$$

$$V\sum_{i=1}^{K}(n_i\,Z_{K,i}) \;=\; N\,\mu_K + \mu_K^2\,K\,N_{K,2}/\alpha_K = N\,\mu_K(1+o(1))\;,$$

hence $\sum_{i=1}^{K} n_i Z_{K,i} = O_p(\sqrt{K\,\mu_K})$ and $C_K = O(K\,\mu_K)$.

From this and (3.18) we have

$$\frac{1}{N\,\mu_K}\sum_{i=1}^{K} n_i Z_{K,i} \;=\; o_p(1)\;,$$

$$\frac{\sqrt{\lambda_K}\,\alpha_K}{m_K\,\mu_K^2}\,\frac{1}{N}\sum_{i=1}^{K} n_i Z_{K,i} \;=\; o_p(1)\;,$$

$$\frac{\sqrt{\lambda_K}\,\alpha_K}{m_K\,N\,\mu_K^2}\,C_K \;=\; o(1)\;,$$

$$\frac{\sqrt{\lambda_K}\,\alpha_K}{m_K\,N\,\mu_K^2}\left(\sum_{i=1}^{K} n_i Z_{K,i}\right)^2 \;=\; o_p(1)\;.$$

Therefore, from the above and the definition of $\hat{\alpha}_{K,1}$ we obtain

$$\frac{\hat{\alpha}_{K,1}}{\alpha_K} = \frac{m_K\hat{\mu}_{K,1}^2}{\alpha_k[S_{K,1}-(K-1)\hat{\mu}_{K,1}]}$$

$$\underline{\mathcal{L}} \;\frac{(1+o_p(1))^2}{1+\frac{1}{\sqrt{\lambda_K}}\,\frac{K}{m_K}\,\frac{1}{\sqrt{K}\,\mu_K}\sum_{i=1}^{K} n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)+\frac{1}{\sqrt{\lambda_K}}o_p(1)}\;, \tag{3.40}$$

with the last line following from (3.38), (3.39) and the results immediately above. Letting

$$V_K \;=\; \frac{K}{m_K}\,\frac{1}{\sqrt{K}\,\mu_K}\sum_{i=1}^{K} n_i\left(Z_{K,i}^2 - E\,Z_{K,i}^2 - \frac{K}{N}Z_{K,i}\right)$$

we have

$$\frac{\hat{\alpha}_{K,1}}{\alpha_K} \stackrel{\mathcal{L}}{=} \frac{1+o_p(1)}{1+\frac{1}{\sqrt{\lambda_K}}V_K+\frac{1}{\sqrt{\lambda_K}}o_p(1)} \tag{3.41}$$

where, from $K/m_K = K/N + O(K^{-1})$ and Lemma 3.8, $V_K$ has variance $\psi_{K,1}+o(1)$ and is asymptotically $N(0,\psi_1)$.

To complete the proof for MME-1 we consider the three cases separately.

$\underline{\lambda = 0}$.

$$\sqrt{\frac{\psi_{K,1}}{\lambda_K}}\frac{\hat{\alpha}_{K,1}}{\alpha_K} \stackrel{\mathcal{L}}{=} \frac{1+o_p(1)}{\sqrt{\lambda_K/\psi_{K,1}}+V_K/\sqrt{\psi_{K,i}}+o_p(1)/\sqrt{\psi_{K,1}}} \stackrel{\mathcal{L}}{\to} \frac{1}{\mathcal{Z}}$$

$\underline{\lambda \in (0,\infty)}$

Here the result follows directly from (3.41).

$\underline{\lambda = \infty}$.

$$\frac{\hat{\alpha}_{K,1}}{\alpha_K}-1 \stackrel{\mathcal{L}}{=} \frac{1}{1+V_K/\sqrt{\lambda_K}}-1 \stackrel{\mathcal{L}}{=} \frac{V_K/\sqrt{\lambda_K}}{1+V_K/\sqrt{\lambda_K}}$$

giving

$$\sqrt{\frac{\lambda_K}{\psi_{K,1}}}\left(\frac{\hat{\alpha}_{K,1}-\alpha_K}{\alpha_K}\right) \stackrel{\mathcal{L}}{=} \frac{V_K/\sqrt{\psi_{K,1}}}{1+V_K/\sqrt{\lambda_K}} \stackrel{\mathcal{L}}{\to} \mathcal{Z}$$

The proof for MME-2 proceeds in precisely the same manner and hence is not included.

•

To conclude this section we note that although the results obtained here are limiting results as $K \to \infty$, they reflect the behaviour for finite $K$, as will be seen in the simulation study in Chapter 4. There we see that for the negative binomial distribution with finite $K$, the erratic behaviour of the moment estimator of $\alpha$ increases, as measured by the number of "failed" estimators and the $\sqrt{MSE}$ of "successful" estimators, as $\alpha$ increases for a fixed $\mu$ or as $\mu$ decreases for a fixed $\alpha$. This then corresponds to the result that the erratic behaviour increases as $\lambda_K$ decreases.

## 3.6 Concluding remarks.

In this chapter we have obtained two moment estimators for the parameters of a simple mixed gamma-Poisson model with repeated observations. The two moment estimators are equivalent when the number of repeat observations per "subject" are the same, in which case they correspond to the usual moment estimators for the negative binomial distribution.

To compare the two moment estimators, both between themselves and with the corresponding MLE, we have based comparisons on the asymptotic efficiencies. In determining these quantities we made the assumption that the number of observations per "subject" are bounded, a condition that assures the asymptotic normality of the two moment estimators and of the MLEs. This assumption though is most likely much stronger than is necessary to obtain the results in Sections 3.4 and 3.5. In practice, what is required for the result we obtain is that no "subject" dominates the information. We therefore conjecture that the results in Sections 3.4 and 3.5 continue to hold provided the number of repeat observations per "subject" is bounded in probability, that is $n_i = O_p(1)$, particularly for MME-1. For MME-2 we conjecture that the results hold for any $n_i$ since it gives equal weight to means instead of observations.

To give some credence to the above conjectures consider the results in the example considered in Chapter 2. Now though restrict attention only to those results reported by counters with counter status 2 or 3 (29 counters in total) and assume that the six slides have identical fibre loadings. This assumption can be justified for two reasons, one being that the six slides were from a batch of 14 nominally identical slides (see Section 1.1) and second, the equality of the means is supported by the results in Section 2.5. The model we then have for these results is that the $j^{th}$ count for the $i^{th}$ counter is Poisson with mean $\gamma_i \mu$ with the $\gamma_i$ independent gamma random variables with mean 1 and variance $1/\alpha$. Estimating the $\mu$ and $\alpha$ parameters using the two moment methods and the MLE above gives:

| Estimator | $\hat{\mu}$ | $\hat{\alpha}$ | $1/\hat{\alpha}$ |
|---|---|---|---|
| MME-1 | 26.08 (3.04) | 21.85 (17.75) | 0.046 |
| MME-2 | 23.78 (1.44) | 14.87 ( 6.41) | 0.067 |
| MLE | 24.05 (1.42) | 15.69 ( 6.47) | 0.064 |

where the term in brackets is the estimate of the standard error of the corresponding estimator based on the results in Theorems 3.1, 3.2 and 3.3 and from (3.3) evaluated at the current estimates.

We thus see that the standard errors of the MME-1 estimators are relatively large compared with MME-2 and MLE. This is brought about by the large number of results from one "subject", counter 4, who contributed more that 50 percent of the results. For this example, and in all examples where a few "subjects" dominate, we would suggest MME-2 over MME-1.

# Chapter 4

# Estimating the shape parameter for the negative binomial distribution.

In the previous chapter a special case of a mixed gamma-Poisson model of Chapter 2 was studied which, for equal number of observations per "subject" corresponds to a negative multinomial distribution (see Sibuya, Yoshimura and Shimizu (1964)) or, for a single observation per "subject", a negative binomial distribution (see Johnson, Kotz and Kemp (1992), Chapter 5). There two moment estimators and the MLEs were obtained for the case of unequal observations per "subject" and it was shown that asymptotically and under certain conditions the moment estimators performed favourably compared to the MLE.

In view of this favourable asymptotic comparison of the moment estimators with the MLEs, in this chapter we consider a comparison of the moment estimator and the MLEs for finite samples. Here the comparison will be based on a simulation study and restricted to a very special case of the model in Chapter 3, that being the negative binomial distribution with parameters mean $\mu$ and shape $\alpha$, denoted by $NBD(\mu, \alpha)$ and with probability distribution given in Appendix A. Since the two moment estimators MME-1 and MME-2 of Chapter 3 are identical in the negative binomial case we will refer throughout this chapter to the moment estimator as

MME. Also, as the MME and MLE of $\mu$ are both given by the sample mean, our simulation study only considers the estimation of $\alpha$. Included in the simulation study are two alternative estimates for the shape parameter.

We begin in Section 4.1 with a review of previous studies comparing various estimators of $\alpha$ and point out the limitations of these studies. Section 4.2 then introduces a new measure, the percentile measure, that is included in our simulation study to compare the estimators, together with a justification of this new measure. This is followed by Section 4.3 giving details of the computer simulation. The results of the simulation study are then reported in Section 4.4 and a discussion of the results is given in Section 4.5.

## 4.1 Review of estimation of $\alpha$ for the negative binomial distribution.

Because the negative binomial distribution has found wide applicability across a field of disciplines (see Tripathi (1985) for examples) owing to its ability to model overdispersed count data relative to the Poisson distribution, the estimation of its parameters has been extensively studied. Further, the family of NBDs has the advantage that it can be extended to include the family of Poisson distributions simply by extending the domain of $\alpha$ to $(0, \infty]$.

To estimate the parameters of the negative binomial distribution there is little contention regarding $\mu$, with the sample mean generally accepted as the best choice, but the same is not the case for $\alpha$. Here a number of choices exist including the method of moments estimator (MME), maximum likelihood estimator (MLE), zero-class estimator, digamma estimator (see Anscombe (1950) and Pieters, Gates, Matis and Sterling (1977) for details of these), maximum quasi-likelihood estimator (MQLE) (Clark and Perry (1989)) and the conditional likelihood estimator (CLE) (Anraku and Yanagimoto (1990)). Of these the asymptotic behaviour of the first four have been studied by Anscombe (1950). Simulation studies have been undertaken to study the small sample behaviour of all six (see Pieters et al. (1977);

Willson, Folks and Young (1984); Clark and Perry (1989); Anraku and Yanagimoto (1990); Piegorsch (1990)), with each simulation study examining only a subset of the above six estimators.

Among the other estimators for the parameters of the NBD are minimum $\chi^2$ estimators proposed by Katti and Gurland (1962). As these have not been specifically formulated for the estimation of $\alpha$ they are not discussed further here.

Based on asymptotic results given in Anscombe (1950) and the results of the simulation study by Pieters et al. (1977) it appears to be generally accepted that the zero-class and digamma estimators are no longer contenders for the estimation of $\alpha$ leaving only the MME, MLE, MQLE and CLE from those above. Of these remaining four it is not clear from the simulation studies available which is superior/inferior under particular conditions for two main reasons.

First of these is a consequence of the fact that the MME gives negative estimates with positive probability whilst each of the remaining three estimators give infinite estimates, again with positive probability. To overcome the problems associated with this a number of different strategies have been employed. Of these the most common is to discard samples giving infinite or negative estimates and to generate replacement samples. This is the procedure employed by Pieters et al. (1977), Willson et al. (1984) and Anraku and Yanagimoto (1990). It should be noted here that when using computer simulations infinite estimates are taken to correspond to estimates exceeding a pre-specified upper bound. Only Willson et al. (1984) document the value they chose.

An alternative strategy is to estimate a function of $\alpha$ that overcomes the problems associated with infinite estimates. This second strategy was employed by Clark and Perry (1989) and Piegorsch (1990) who considered estimation of $1/\alpha$ instead of $\alpha$ and also by Anraku and Yanagimoto (1990) who considered estimation of $1/\alpha$ and $1/(1+\alpha)$ in addition to $\alpha$. Using this latter strategy, whilst overcoming the problems associated with infinite estimates, does not remove the problem of negative MMEs and now gives rise to negative MQLEs and MLEs (negative estimates in this case correspond to under-dispersion). Clark and Perry (1989) and Piegorsch (1990)

have retained these negative estimates in their simulation. Anraku and Yanagimoto (1990) on the other hand, when estimating $1/\alpha$ or $1/(1+\alpha)$, set the estimate to zero for samples giving infinite or negative values for the corresponding estimate of $\alpha$.

A third strategy, used by Anraku and Yanagimoto (1990) is to compare the behaviour of the estimated probability distribution functions via the Kullback-Leibler risk. The basis of this idea will be utilised in this chapter.

The second reason for difficulty associated with comparing the four estimators based on the studies already undertaken is that none of the studies included all four estimators. By itself this would not be a problem were it not that most of the five simulation studies chose parameter values and sample sizes for their study which offered only limited opportunity for comparison across studies. Piegorsch (1990) is the only exception, choosing values to enable a comparison with the results of Clark and Perry (1989).

This chapter aims to rectify this situation by reporting the results of a simulation study involving the four estimators MME, MLE, MQLE and CLE. Here the results are derived using the most commonly used procedure for comparing alternative estimators of $\alpha$, that being to discard samples giving one or more estimates of $\alpha$ outside a given interval. The ranges of parameter values $\alpha$ and $\mu$ and sample sizes $n$ included were chosen to cover most of the values examined by others. In addition, we introduce an alternative and arguably more appropriate measure whereby the performances of the four estimators can be compared. This measure will be referred to as the percentile measure ($P_M$).

## 4.2 Percentile measure ($P_M$).

Unlike the mean parameter of the NBD, the shape parameter in most applications does not have an immediately obvious interpretation. Waters (1959) gives an example where $\alpha$ has a physical interpretation and hence a justification of its estimation in its own right, but generally the reason for the estimation of $\alpha$ is in order to quan-

tify, via the distribution function, the random behaviour underlying the process. In such cases the estimation of a function of a parameter rather than the parameter itself serves no practical purpose.

Therefore, to assess the performance of the estimator it may be more appropriate to compare the "distribution" of the estimated distribution functions, generally through some metric such as Kolmogorov's metric or the total variation metric or, as considered by Anraku and Yanagimoto (1990), the Kullback-Leibler risk, rather than the distribution of the parameter estimates themselves. This alternative procedure has the advantage that it provides a mechanism to overcome the problems associated with negative MMEs or infinite MLEs, MQLEs and CLEs of $\alpha$ in the case where the true underlying distribution is assumed to be a member of the extended NBD family (extended in the sense of including the Poisson family, which is the limit as $\alpha \to \infty$). In such cases an appropriate estimate of the underlying distribution is a Poisson distribution with mean equal the sample mean. The case for assuming an underlying Poisson for negative MMEs of $\alpha$ is based on a number of grounds, not least the under-dispersion of the data relative to expectation.

Unfortunately many of the metrics available give little insight into the relative "closeness" of distributions particularly when, as will often be the case, tolerence intervals for subsequent observations are required. For example, it is a simple exercise to construct two distributions for which the Kolmogorov metric is less than $\epsilon$ ($> 0$) for which specified percentiles differ by more than K ($> 0$).

To overcome this the measure proposed in this chapter is:

$$P_M(F,G) \;=\; \sup_{t\in A}\left|F^{-1}(t) - G^{-1}(t)\right| \tag{4.1}$$

where $F$ and $G$ denote distribution functions, $A = [0.01, 0.99]$ and $F^{-1}(t) = \inf\{x : F(x) \geq t\}$. Here $F$ will be taken as the true underlying NBD and $G$ will be taken to be the estimated distribution function, obtained by replacing the parameters of the distribution by their estimates.

The reason for the restriction of the above measure to a set A of the form $[\beta, 1-\gamma]$ for $0 < \beta < 1-\gamma < 1$ is twofold. Firstly, the choice A = [0, 1] has the problem that the difference between the inverse of two distribution functions is not

defined at 0, nor is it at 1 for certain distributions. The second and more practical reason hinges on the fact that estimates of percentiles corresponding to the extreme tails of the distribution are rarely sought owing to the uncertainty associated with such estimates.

The choice of A therefore needs to reflect the range of values for which percentile estimates may be required whilst at the same time providing a measure of "closeness" of the distributions. The choice taken for this study is A = [.01, .99]. Scope does however exist for varying A without significantly altering the conclusions of the study.

A point worth noting here is that although the percentile measure is not a metric on the space of distribution functions it is a metric on the restricted space of doubly censored distribution functions with $100\beta\%$ and $100\gamma\%$ censored on the left and right respectively. Hence a percentile measure of zero corresponds to effectively equivalent distributions in the case where the tails of the distribution beyond the $\beta$ and $(1-\gamma)$ percentiles are of no interest.

## 4.3 Computer simulation.

For the Monte Carlo study, conducted on Apollo DN2500 and DN3500 workstations, 10000 samples of size $n$ ($n$ = 20, 30, 50 and 100) were randomly generated for each of the 24 negative binomial distributions with parameters $\mu$ = 1, 3, 5, 10, 20 and 50 crossed with $\alpha$ = .5, 1, 3 and 5. The four estimates of $\alpha$, and the corresponding estimates of the distribution functions, for each combination of $n$, $\alpha$ and $\mu$, were then based on the same 10000 samples.

To generate the NBD random variables a two stage process was used. Firstly, a gamma random variable W with parameter $\alpha$ was generated. Ahren's gamma algorithm GS, given in Atkinson and Pearce (1976), was used for $\alpha$ less than 1 whilst an algorithm proposed by Cheng (1977) was used for $\alpha$ greater than or equal one. A Poisson random variable (X) with mean $W\mu/\alpha$ was then generated using the idea of mimicking the Poisson process outlined in Mihram (1972). The resultant

random variable X has a NBD with mean $\mu$ and shape parameter $\alpha$.

For each sample of size n the mean $\mu$ was estimated by the sample mean. Denoting the sample mean and sample variance by $\bar{x}$ and $s^2$ respectively the four estimates for $\alpha$ are given by:

$$\text{MME:} \quad \hat{\alpha}_{MME} = \frac{\bar{x}^2}{s^2 - \bar{x}}$$

$$\text{MLE:} \quad \hat{\alpha}_{MLE} = \text{the value } \alpha \text{ that maximises } \prod_{i=1}^{n} \left[ \frac{\Gamma(\alpha + x_i)}{\Gamma(\alpha)\, x_i!} \left( \frac{\alpha}{\alpha + \bar{x}} \right)^{\alpha} \left( \frac{\bar{x}}{\alpha + \bar{x}} \right)^{x_i} \right]$$

$$\text{MQLE:} \quad \hat{\alpha}_{MQLE} = \text{the value } \alpha \text{ that maximises } \sum_{i=1}^{n} \left[ x_i \ln \left( \frac{\bar{x}}{x_i} \right) - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \left( x_i + \frac{1}{6} \right) \right.$$
$$\left. -(\alpha + x_i) \ln \left( \frac{\alpha + \bar{x}}{\alpha + x_i} \right) - \ln \left( \frac{\alpha + x_i}{\alpha} \right) + \frac{1}{2} \ln \left( \frac{6\alpha + 6x_i + 1}{6\alpha + 1} \right) \right]$$

$$\text{CLE:} \quad \hat{\alpha}_{CLE} = \text{the value } \alpha \text{ that maximises } \frac{\Gamma(n\alpha)\ (n\bar{x})!}{\Gamma(n(\alpha + \bar{x}))} \prod_{i=1}^{n} \left[ \frac{\Gamma(\alpha + x_i)}{\Gamma(\alpha)\, x_i!} \right]$$

Of the four estimators, it can immediately be seen that the MME is negative for $s^2 < \bar{x}$ and infinite for $s^2 = \bar{x}$, whilst Anraku and Yanagimoto (1990) note that the CLE is infinite for $s^2 \leq \bar{x}$ and is positive and unique otherwise. On the other hand, the MLE is infinite if $(n-1)s^2 \leq n\bar{x}$ and is positive and unique otherwise. This latter result, and the corresponding result for the CLE, follow from a theorem by Levin and Reeds (1977). It follows then that the MLE is infinite whenever the CLE is infinite which in turn corresponds to a non-positive MME. This will be borne out in the simulations. For the MQLE, obtained from the extended quasi-likelihood function for the NBD (see Nelder and Pregibon (1987)), no results are available as to when the estimate is positive or infinite, nor if the estimate is unique.

To obtain the MLE, MQLE and CLE a NAG minimisation routine (E04BBF) was used with boundary conditions on $\alpha$ set as 0.0001 and 10000. (Solutions for the estimate of $\alpha$ beyond the boundary limits were assigned the corresponding boundary

values.) These limits were chosen for a number of reasons, one being that they were the values used by Willson et al. (1984). Another reason was that, for the values of $\mu$ considered in this paper, a negative binomial distribution with $\alpha$ greater than 10000 is effectively identical to a Poisson distribution with mean $\mu$. As for the lower limit, none of the estimates of $\alpha$ in the study fell in the range (0, 0.0001).

In view of the above comments MMEs for $\alpha$ outside the interval $(0, 10000]$ were reassigned the value 10000.

## 4.4 Results of computer simulation.

For the first stage the bias and mean square error (MSE) of the four estimators for $\alpha$ were calculated from samples giving all four estimators in the interval (0, 10000). (Unlike some previous studies, discarded samples were not replaced.) The results of this study are given in Tables 4.1, 4.2, 4.3 and 4.4, corresponding to $n$ = 20, 30, 50 and 100 respectively. Each table contains, for each estimator and for each $\mu$ and $\alpha$ combination considered, the number of times the particular estimator failed (i.e. outside the interval (0, 10000)), the bias and the square root of the mean square error of the estimator. Also, in the third column of each table are the number of samples, for the particular combination of $n$, $\alpha$ and $\mu$, that had to be discarded for this comparison because one or more of the estimators gave an estimate outside the range (0, 10000).

For the second comparison none of the 10000 samples for each combination of $n$, $\alpha$ and $\mu$ were discarded. Here the estimate of $\alpha$ was taken as 10000 whenever the estimate fell outside the range (0, 10000). The percentile measure was calculated for each estimator for each sample. Then, for each combination of $n$, $\alpha$ and $\mu$, the mean and standard deviation of the $P_M$ values were calculated. These results are given in Tables 4.5, 4.6, 4.7 and 4.8, corresponding to $n$ = 20, 30, 50 and 100 respectively.

**Table 4.1** Moment comparison of estimators for $\alpha$ for sample size **n=20**

| | | Total | MME | | | MLE | | | MQLE | | | CLE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | fail | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ |
| 0.5 | 1 | 374 | 275 | 0.455 | 1.112 | 374 | 0.495 | 2.407 | 374 | 0.469 | 2.387 | 275 | 0.207 | 0.815 |
| 1.0 | 1 | 1080 | 759 | 0.805 | 2.033 | 1076 | 1.248 | 5.461 | 1080 | 1.207 | 5.438 | 759 | 0.429 | 1.591 |
| 3.0 | 1 | 3524 | 2766 | 0.211 | 2.900 | 3515 | 1.956 | 11.682 | 3524 | 1.902 | 11.662 | 2766 | -0.301 | 2.618 |
| 5.0 | 1 | 4488 | 3590 | -1.201 | 3.501 | 4476 | 1.471 | 15.060 | 4488 | 1.412 | 15.042 | 3590 | -1.797 | 3.459 |
| 0.5 | 3 | 3 | 1 | 0.236 | 0.542 | 3 | 0.147 | 0.505 | 3 | 0.134 | 0.497 | 1 | 0.087 | 0.389 |
| 1.0 | 3 | 19 | 15 | 0.510 | 1.513 | 18 | 0.493 | 2.856 | 19 | 0.474 | 2.851 | 15 | 0.289 | 1.254 |
| 3.0 | 3 | 645 | 477 | 2.158 | 6.249 | 626 | 4.860 | 29.766 | 645 | 4.837 | 29.768 | 477 | 1.681 | 5.639 |
| 5.0 | 3 | 1586 | 1194 | 2.577 | 8.187 | 1555 | 8.513 | 46.724 | 1586 | 8.489 | 46.743 | 1194 | 2.012 | 7.572 |
| 0.5 | 5 | 0 | 0 | 0.200 | 0.409 | 0 | 0.105 | 0.308 | 0 | 0.094 | 0.302 | 0 | 0.062 | 0.272 |
| 1.0 | 5 | 1 | 0 | 0.369 | 1.180 | 1 | 0.414 | 12.306 | 1 | 0.399 | 12.305 | 0 | 0.192 | 0.971 |
| 3.0 | 5 | 114 | 76 | 1.825 | 5.948 | 111 | 3.325 | 31.114 | 114 | 3.308 | 31.106 | 76 | 1.464 | 5.461 |
| 5.0 | 5 | 553 | 401 | 3.241 | 9.902 | 546 | 7.944 | 57.296 | 553 | 7.890 | 55.773 | 401 | 2.754 | 9.251 |
| 0.5 | 10 | 0 | 0 | 0.170 | 0.347 | 0 | 0.080 | 0.243 | 0 | 0.071 | 0.238 | 0 | 0.047 | 0.221 |
| 1.0 | 10 | 0 | 0 | 0.262 | 0.631 | 0 | 0.184 | 0.527 | 0 | 0.172 | 0.523 | 0 | 0.116 | 0.474 |
| 3.0 | 10 | 3 | 2 | 1.031 | 3.358 | 3 | 1.170 | 4.245 | 3 | 1.159 | 4.243 | 2 | 0.802 | 3.115 |
| 5.0 | 10 | 51 | 35 | 2.305 | 8.103 | 50 | 3.419 | 21.864 | 51 | 3.408 | 21.838 | 35 | 1.992 | 7.632 |
| 0.5 | 15 | 0 | 0 | 0.174 | 0.326 | 0 | 0.076 | 0.221 | 0 | 0.067 | 0.216 | 0 | 0.046 | 0.202 |
| 1.0 | 15 | 0 | 0 | 0.254 | 0.601 | 0 | 0.175 | 0.491 | 0 | 0.163 | 0.487 | 0 | 0.113 | 0.446 |
| 3.0 | 15 | 1 | 0 | 0.784 | 2.485 | 1 | 0.850 | 2.806 | 1 | 0.840 | 2.804 | 0 | 0.587 | 2.271 |
| 5.0 | 15 | 9 | 7 | 1.608 | 6.234 | 9 | 2.115 | 12.490 | 9 | 2.108 | 12.495 | 7 | 1.366 | 5.932 |
| 0.5 | 20 | 0 | 0 | 0.200 | 0.330 | 0 | 0.082 | 0.216 | 0 | 0.073 | 0.211 | 0 | 0.053 | 0.197 |
| 1.0 | 20 | 0 | 0 | 0.249 | 0.573 | 0 | 0.166 | 0.466 | 0 | 0.155 | 0.463 | 0 | 0.107 | 0.425 |
| 3.0 | 20 | 0 | 0 | 0.694 | 2.198 | 0 | 0.745 | 2.381 | 0 | 0.737 | 2.379 | 0 | 0.515 | 2.014 |
| 5.0 | 20 | 0 | 0 | 1.320 | 4.050 | 0 | 1.585 | 4.500 | 0 | 1.578 | 4.499 | 0 | 1.111 | 3.811 |

**Table 4.2** Moment comparison of estimators for $\alpha$ for sample size **n=30**

| | | Total | MME | | | MLE | | | MQLE | | | CLE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | fail | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ |
| 0.5 | 1 | 90 | 76 | 0.339 | 0.938 | 90 | 0.298 | 1.336 | 90 | 0.274 | 1.318 | 76 | 0.168 | 0.709 |
| 1.0 | 1 | 431 | 341 | 0.752 | 2.065 | 427 | 0.952 | 3.839 | 431 | 0.913 | 3.818 | 341 | 0.479 | 1.750 |
| 3.0 | 1 | 2504 | 2021 | 0.890 | 3.804 | 2473 | 2.311 | 9.754 | 2504 | 2.255 | 9.733 | 2021 | 0.424 | 3.394 |
| 5.0 | 1 | 3681 | 3079 | -0.430 | 4.043 | 3631 | 1.459 | 10.880 | 3681 | 1.399 | 10.865 | 3079 | -0.954 | 3.831 |
| 0.5 | 3 | 0 | 0 | 0.152 | 0.359 | 0 | 0.081 | 0.281 | 0 | 0.068 | 0.274 | 0 | 0.048 | 0.246 |
| 1.0 | 3 | 2 | 1 | 0.310 | 0.869 | 2 | 0.243 | 0.941 | 2 | 0.224 | 0.934 | 1 | 0.161 | 0.733 |
| 3.0 | 3 | 222 | 161 | 1.810 | 5.934 | 220 | 2.996 | 19.308 | 222 | 2.973 | 19.306 | 161 | 1.476 | 5.439 |
| 5.0 | 3 | 850 | 690 | 3.282 | 10.153 | 839 | 7.922 | 41.992 | 850 | 7.896 | 41.974 | 690 | 2.789 | 9.566 |
| 0.5 | 5 | 0 | 0 | 0.132 | 0.289 | 0 | 0.064 | 0.210 | 0 | 0.053 | 0.205 | 0 | 0.038 | 0.194 |
| 1.0 | 5 | 0 | 0 | 0.220 | 0.565 | 0 | 0.152 | 0.478 | 0 | 0.137 | 0.472 | 0 | 0.097 | 0.436 |
| 3.0 | 5 | 22 | 15 | 1.129 | 4.379 | 22 | 1.574 | 16.900 | 22 | 1.557 | 16.890 | 15 | 0.888 | 4.036 |
| 5.0 | 5 | 171 | 131 | 2.845 | 9.817 | 167 | 5.423 | 44.692 | 171 | 5.406 | 44.671 | 131 | 2.475 | 9.257 |
| 0.5 | 10 | 0 | 0 | 0.120 | 0.257 | 0 | 0.052 | 0.173 | 0 | 0.042 | 0.169 | 0 | 0.031 | 0.162 |
| 1.0 | 10 | 0 | 0 | 0.180 | 0.483 | 0 | 0.117 | 0.382 | 0 | 0.105 | 0.378 | 0 | 0.075 | 0.356 |
| 3.0 | 10 | 0 | 0 | 0.580 | 3.088 | 0 | 0.734 | 12.494 | 0 | 0.723 | 12.474 | 0 | 0.426 | 2.863 |
| 5.0 | 10 | 5 | 3 | 1.381 | 5.683 | 5 | 2.375 | 75.282 | 5 | 2.277 | 66.503 | 3 | 1.168 | 5.196 |
| 0.5 | 15 | 0 | 0 | 0.129 | 0.250 | 0 | 0.050 | 0.163 | 0 | 0.041 | 0.159 | 0 | 0.031 | 0.153 |
| 1.0 | 15 | 0 | 0 | 0.168 | 0.450 | 0 | 0.108 | 0.355 | 0 | 0.097 | 0.351 | 0 | 0.070 | 0.332 |
| 3.0 | 15 | 0 | 0 | 0.479 | 1.488 | 0 | 0.493 | 1.442 | 0 | 0.484 | 1.440 | 0 | 0.346 | 1.322 |
| 5.0 | 15 | 1 | 1 | 0.941 | 3.965 | 1 | 1.131 | 5.497 | 1 | 1.123 | 5.495 | 1 | 0.796 | 3.734 |
| 0.5 | 20 | 0 | 0 | 0.164 | 0.260 | 0 | 0.057 | 0.160 | 0 | 0.048 | 0.156 | 0 | 0.039 | 0.150 |
| 1.0 | 20 | 0 | 0 | 0.180 | 0.431 | 0 | 0.108 | 0.335 | 0 | 0.097 | 0.332 | 0 | 0.072 | 0.314 |
| 3.0 | 20 | 0 | 0 | 0.398 | 1.297 | 0 | 0.412 | 1.250 | 0 | 0.404 | 1.248 | 0 | 0.280 | 1.157 |
| 5.0 | 20 | 0 | 0 | 0.727 | 2.629 | 0 | 0.852 | 2.759 | 0 | 0.845 | 2.757 | 0 | 0.590 | 2.483 |

**Table 4.3** Moment comparison of estimators for $\alpha$ for sample size **n=50**

| | | Total | MME | | | MLE | | | MQLE | | | CLE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | fail | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ |
| 0.5 | 1 | 5 | 5 | 0.183 | 0.516 | 5 | 0.131 | 0.620 | 5 | 0.110 | 0.606 | 5 | 0.085 | 0.403 |
| 1.0 | 1 | 97 | 78 | 0.543 | 1.796 | 94 | 0.633 | 4.793 | 97 | 0.597 | 4.794 | 78 | 0.374 | 1.588 |
| 3.0 | 1 | 1598 | 1263 | 1.594 | 5.169 | 1540 | 3.090 | 14.413 | 1598 | 3.034 | 14.407 | 1263 | 1.209 | 4.743 |
| 5.0 | 1 | 2693 | 2201 | 1.039 | 6.270 | 2625 | 4.470 | 32.634 | 2693 | 4.407 | 32.571 | 2201 | 0.558 | 5.863 |
| 0.5 | 3 | 0 | 0 | 0.095 | 0.231 | 0 | 0.047 | 0.172 | 0 | 0.035 | 0.166 | 0 | 0.029 | 0.162 |
| 1.0 | 3 | 0 | 0 | 0.161 | 0.469 | 0 | 0.113 | 0.399 | 0 | 0.095 | 0.392 | 0 | 0.074 | 0.372 |
| 3.0 | 3 | 39 | 28 | 1.069 | 4.524 | 38 | 1.994 | 52.277 | 39 | 1.978 | 52.609 | 28 | 0.883 | 4.256 |
| 5.0 | 3 | 296 | 237 | 2.921 | 10.009 | 289 | 5.250 | 49.650 | 296 | 5.231 | 50.129 | 237 | 2.591 | 9.592 |
| 0.5 | 5 | 0 | 0 | 0.082 | 0.206 | 0 | 0.035 | 0.141 | 0 | 0.024 | 0.137 | 0 | 0.020 | 0.135 |
| 1.0 | 5 | 0 | 0 | 0.131 | 0.388 | 0 | 0.083 | 0.314 | 0 | 0.068 | 0.309 | 0 | 0.053 | 0.297 |
| 3.0 | 5 | 0 | 0 | 0.584 | 2.782 | 0 | 0.625 | 4.616 | 0 | 0.609 | 4.615 | 0 | 0.441 | 2.433 |
| 5.0 | 5 | 27 | 22 | 1.874 | 8.027 | 27 | 2.999 | 40.697 | 27 | 2.984 | 40.873 | 22 | 1.647 | 7.569 |
| 0.5 | 10 | 0 | 0 | 0.078 | 0.188 | 0 | 0.029 | 0.121 | 0 | 0.020 | 0.118 | 0 | 0.017 | 0.117 |
| 1.0 | 10 | 0 | 0 | 0.108 | 0.341 | 0 | 0.063 | 0.258 | 0 | 0.050 | 0.254 | 0 | 0.039 | 0.247 |
| 3.0 | 10 | 0 | 0 | 0.319 | 1.122 | 0 | 0.313 | 1.055 | 0 | 0.302 | 1.052 | 0 | 0.220 | 0.994 |
| 5.0 | 10 | 0 | 0 | 0.680 | 2.312 | 0 | 0.763 | 2.354 | 0 | 0.753 | 2.352 | 0 | 0.566 | 2.179 |
| 0.5 | 15 | 0 | 0 | 0.097 | 0.182 | 0 | 0.030 | 0.114 | 0 | 0.021 | 0.111 | 0 | 0.019 | 0.110 |
| 1.0 | 15 | 0 | 0 | 0.102 | 0.315 | 0 | 0.060 | 0.240 | 0 | 0.048 | 0.237 | 0 | 0.038 | 0.230 |
| 3.0 | 15 | 0 | 0 | 0.267 | 0.984 | 0 | 0.267 | 0.920 | 0 | 0.257 | 0.918 | 0 | 0.187 | 0.873 |
| 5.0 | 15 | 0 | 0 | 0.485 | 1.865 | 0 | 0.547 | 1.846 | 0 | 0.538 | 1.844 | 0 | 0.389 | 1.737 |
| 0.5 | 20 | 0 | 0 | 0.139 | 0.199 | 0 | 0.040 | 0.115 | 0 | 0.032 | 0.111 | 0 | 0.030 | 0.110 |
| 1.0 | 20 | 0 | 0 | 0.123 | 0.306 | 0 | 0.064 | 0.235 | 0 | 0.053 | 0.232 | 0 | 0.043 | 0.226 |
| 3.0 | 20 | 0 | 0 | 0.229 | 0.918 | 0 | 0.225 | 0.842 | 0 | 0.216 | 0.840 | 0 | 0.151 | 0.803 |
| 5.0 | 20 | 0 | 0 | 0.458 | 1.663 | 0 | 0.511 | 1.630 | 0 | 0.503 | 1.628 | 0 | 0.368 | 1.540 |

**Table 4.4** Moment comparison of estimators for $\alpha$ for sample size **n=100**

| | | Total | MME | | | MLE | | | MQLE | | | CLE | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | fail | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ | fail | bias | $\sqrt{MSE}$ |
| 0.5 | 1 | 0 | 0 | 0.083 | 0.243 | 0 | 0.050 | 0.211 | 0 | 0.032 | 0.202 | 0 | 0.034 | 0.197 |
| 1.0 | 1 | 2 | 1 | 0.221 | 0.782 | 1 | 0.184 | 0.874 | 2 | 0.151 | 0.861 | 1 | 0.139 | 0.672 |
| 3.0 | 1 | 503 | 406 | 1.958 | 6.702 | 481 | 3.528 | 48.300 | 503 | 3.438 | 46.244 | 406 | 1.703 | 6.282 |
| 5.0 | 1 | 1576 | 1347 | 2.705 | 9.764 | 1530 | 6.689 | 74.635 | 1576 | 6.465 | 64.416 | 1347 | 2.346 | 9.311 |
| 0.5 | 3 | 0 | 0 | 0.043 | 0.143 | 0 | 0.018 | 0.104 | 0 | 0.006 | 0.100 | 0 | 0.009 | 0.101 |
| 1.0 | 3 | 0 | 0 | 0.082 | 0.296 | 0 | 0.053 | 0.242 | 0 | 0.035 | 0.237 | 0 | 0.035 | 0.233 |
| 3.0 | 3 | 0 | 0 | 0.406 | 1.650 | 0 | 0.416 | 1.765 | 0 | 0.393 | 1.759 | 0 | 0.324 | 1.557 |
| 5.0 | 3 | 17 | 12 | 1.368 | 5.869 | 17 | 1.618 | 9.560 | 17 | 1.594 | 9.557 | 12 | 1.227 | 5.551 |
| 0.5 | 5 | 0 | 0 | 0.043 | 0.138 | 0 | 0.018 | 0.092 | 0 | 0.008 | 0.090 | 0 | 0.011 | 0.090 |
| 1.0 | 5 | 0 | 0 | 0.067 | 0.257 | 0 | 0.040 | 0.201 | 0 | 0.025 | 0.198 | 0 | 0.026 | 0.196 |
| 3.0 | 5 | 0 | 0 | 0.237 | 0.967 | 0 | 0.234 | 0.923 | 0 | 0.218 | 0.919 | 0 | 0.175 | 0.885 |
| 5.0 | 5 | 0 | 0 | 0.620 | 2.444 | 0 | 0.663 | 2.487 | 0 | 0.647 | 2.483 | 0 | 0.522 | 2.310 |
| 0.5 | 10 | 0 | 0 | 0.043 | 0.124 | 0 | 0.015 | 0.080 | 0 | 0.006 | 0.078 | 0 | 0.009 | 0.079 |
| 1.0 | 10 | 0 | 0 | 0.057 | 0.227 | 0 | 0.031 | 0.169 | 0 | 0.018 | 0.167 | 0 | 0.019 | 0.165 |
| 3.0 | 10 | 0 | 0 | 0.161 | 0.705 | 0 | 0.148 | 0.640 | 0 | 0.137 | 0.638 | 0 | 0.105 | 0.620 |
| 5.0 | 10 | 0 | 0 | 0.321 | 1.364 | 0 | 0.338 | 1.327 | 0 | 0.328 | 1.325 | 0 | 0.252 | 1.279 |
| 0.5 | 15 | 0 | 0 | 0.070 | 0.124 | 0 | 0.018 | 0.078 | 0 | 0.010 | 0.075 | 0 | 0.013 | 0.076 |
| 1.0 | 15 | 0 | 0 | 0.060 | 0.215 | 0 | 0.030 | 0.159 | 0 | 0.018 | 0.157 | 0 | 0.019 | 0.156 |
| 3.0 | 15 | 0 | 0 | 0.126 | 0.626 | 0 | 0.121 | 0.559 | 0 | 0.111 | 0.558 | 0 | 0.083 | 0.544 |
| 5.0 | 15 | 0 | 0 | 0.236 | 1.157 | 0 | 0.264 | 1.107 | 0 | 0.256 | 1.105 | 0 | 0.191 | 1.072 |
| 0.5 | 20 | 0 | 0 | 0.123 | 0.157 | 0 | 0.029 | 0.079 | 0 | 0.020 | 0.076 | 0 | 0.024 | 0.077 |
| 1.0 | 20 | 0 | 0 | 0.087 | 0.203 | 0 | 0.037 | 0.151 | 0 | 0.026 | 0.149 | 0 | 0.027 | 0.148 |
| 3.0 | 20 | 0 | 0 | 0.116 | 0.610 | 0 | 0.112 | 0.540 | 0 | 0.104 | 0.539 | 0 | 0.077 | 0.527 |
| 5.0 | 20 | 0 | 0 | 0.219 | 1.082 | 0 | 0.238 | 1.015 | 0 | 0.231 | 1.014 | 0 | 0.172 | 0.986 |

**Table 4.5** Moment comparison of $P_M$ for sample size n=20

| | | MME | | MLE | | MQLE | | CLE | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. |
| 0.5 | 1 | 2.887 | 1.757 | 2.981 | 1.909 | 2.996 | 1.937 | 3.090 | 2.103 |
| 1.0 | 1 | 1.956 | 1.102 | 2.022 | 1.151 | 2.042 | 1.180 | 2.087 | 1.286 |
| 3.0 | 1 | 1.431 | 0.680 | 1.444 | 0.696 | 1.457 | 0.712 | 1.515 | 0.777 |
| 5.0 | 1 | 1.256 | 0.580 | 1.263 | 0.604 | 1.273 | 0.620 | 1.333 | 0.698 |
| 0.5 | 3 | 6.845 | 4.764 | 6.749 | 4.870 | 6.787 | 4.946 | 6.899 | 5.214 |
| 1.0 | 3 | 4.382 | 2.927 | 4.303 | 2.832 | 4.340 | 2.879 | 4.376 | 3.008 |
| 3.0 | 3 | 2.420 | 1.412 | 2.419 | 1.376 | 2.431 | 1.389 | 2.452 | 1.449 |
| 5.0 | 3 | 2.032 | 1.113 | 2.030 | 1.079 | 2.037 | 1.086 | 2.072 | 1.149 |
| 0.5 | 5 | 11.080 | 7.845 | 10.807 | 7.804 | 10.847 | 7.915 | 10.992 | 8.253 |
| 1.0 | 5 | 6.466 | 4.568 | 6.250 | 4.318 | 6.285 | 4.376 | 6.338 | 4.533 |
| 3.0 | 5 | 3.222 | 2.033 | 3.180 | 1.931 | 3.192 | 1.943 | 3.221 | 2.030 |
| 5.0 | 5 | 2.580 | 1.535 | 2.563 | 1.466 | 2.569 | 1.475 | 2.607 | 1.564 |
| 0.5 | 10 | 20.410 | 13.706 | 19.653 | 14.272 | 19.729 | 14.472 | 19.929 | 14.979 |
| 1.0 | 10 | 11.651 | 8.322 | 11.071 | 7.810 | 11.126 | 7.904 | 11.179 | 8.150 |
| 3.0 | 10 | 5.523 | 3.615 | 5.387 | 3.433 | 5.399 | 3.448 | 5.418 | 3.558 |
| 5.0 | 10 | 4.051 | 2.537 | 4.005 | 2.431 | 4.010 | 2.438 | 4.033 | 2.520 |
| 0.5 | 15 | 27.072 | 16.831 | 27.264 | 19.117 | 27.346 | 19.324 | 27.511 | 19.809 |
| 1.0 | 15 | 16.853 | 11.531 | 15.957 | 11.226 | 16.029 | 11.364 | 16.123 | 11.684 |
| 3.0 | 15 | 7.679 | 5.290 | 7.416 | 4.930 | 7.436 | 4.956 | 7.454 | 5.113 |
| 5.0 | 15 | 5.534 | 3.599 | 5.417 | 3.385 | 5.425 | 3.391 | 5.435 | 3.498 |
| 0.5 | 20 | 31.565 | 21.974 | 33.559 | 22.990 | 33.572 | 23.142 | 33.605 | 23.477 |
| 1.0 | 20 | 20.947 | 13.572 | 20.163 | 14.274 | 20.247 | 14.402 | 20.306 | 14.698 |
| 3.0 | 20 | 9.864 | 6.750 | 9.454 | 6.286 | 9.473 | 6.307 | 9.520 | 6.485 |
| 5.0 | 20 | 7.071 | 4.626 | 6.902 | 4.423 | 6.910 | 4.432 | 6.936 | 4.543 |

**Table 4.6** Moment comparison of $P_M$ for sample size n=30

| | | MME | | MLE | | MQLE | | CLE | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. |
| 0.5 | 1 | 2.534 | 1.504 | 2.549 | 1.541 | 2.559 | 1.565 | 2.591 | 1.633 |
| 1.0 | 1 | 1.759 | 0.925 | 1.783 | 0.950 | 1.800 | 0.975 | 1.817 | 1.025 |
| 3.0 | 1 | 1.310 | 0.562 | 1.316 | 0.555 | 1.325 | 0.568 | 1.350 | 0.601 |
| 5.0 | 1 | 1.174 | 0.463 | 1.177 | 0.466 | 1.183 | 0.474 | 1.215 | 0.520 |
| 0.5 | 3 | 5.856 | 4.099 | 5.614 | 4.012 | 5.648 | 4.085 | 5.700 | 4.194 |
| 1.0 | 3 | 3.738 | 2.450 | 3.594 | 2.284 | 3.619 | 2.317 | 3.637 | 2.378 |
| 3.0 | 3 | 2.080 | 1.159 | 2.068 | 1.135 | 2.076 | 1.147 | 2.079 | 1.173 |
| 5.0 | 3 | 1.799 | 0.942 | 1.801 | 0.918 | 1.806 | 0.923 | 1.817 | 0.956 |
| 0.5 | 5 | 9.404 | 6.655 | 8.931 | 6.336 | 8.976 | 6.424 | 9.033 | 6.557 |
| 1.0 | 5 | 5.429 | 3.634 | 5.160 | 3.431 | 5.187 | 3.478 | 5.205 | 3.548 |
| 3.0 | 5 | 2.756 | 1.705 | 2.700 | 1.611 | 2.710 | 1.625 | 2.734 | 1.679 |
| 5.0 | 5 | 2.195 | 1.249 | 2.190 | 1.208 | 2.194 | 1.214 | 2.210 | 1.255 |
| 0.5 | 10 | 17.206 | 11.424 | 16.305 | 11.729 | 16.376 | 11.891 | 16.440 | 12.080 |
| 1.0 | 10 | 9.878 | 7.031 | 9.121 | 6.359 | 9.170 | 6.446 | 9.177 | 6.542 |
| 3.0 | 10 | 4.633 | 3.028 | 4.460 | 2.816 | 4.471 | 2.826 | 4.479 | 2.881 |
| 5.0 | 10 | 3.439 | 2.132 | 3.372 | 2.046 | 3.378 | 2.053 | 3.389 | 2.092 |
| 0.5 | 15 | 21.912 | 14.499 | 22.229 | 15.768 | 22.303 | 15.914 | 22.358 | 16.082 |
| 1.0 | 15 | 14.143 | 9.568 | 13.245 | 9.311 | 13.311 | 9.416 | 13.322 | 9.548 |
| 3.0 | 15 | 6.364 | 4.222 | 6.058 | 3.930 | 6.074 | 3.945 | 6.079 | 4.001 |
| 5.0 | 15 | 4.657 | 2.965 | 4.531 | 2.789 | 4.538 | 2.798 | 4.550 | 2.843 |
| 0.5 | 20 | 26.913 | 18.775 | 27.756 | 19.271 | 27.718 | 19.362 | 27.654 | 19.462 |
| 1.0 | 20 | 16.903 | 11.205 | 16.516 | 11.477 | 16.576 | 11.584 | 16.568 | 11.700 |
| 3.0 | 20 | 8.163 | 5.562 | 7.794 | 5.129 | 7.807 | 5.145 | 7.828 | 5.243 |
| 5.0 | 20 | 5.904 | 3.917 | 5.744 | 3.686 | 5.750 | 3.693 | 5.764 | 3.769 |

**Table 4.7** Moment comparison of $P_M$ for sample size n=50

| | | MME | | MLE | | MQLE | | CLE | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. |
| 0.5 | 1 | 2.146 | 1.216 | 2.101 | 1.205 | 2.113 | 1.226 | 2.128 | 1.248 |
| 1.0 | 1 | 1.527 | 0.744 | 1.516 | 0.715 | 1.528 | 0.734 | 1.526 | 0.743 |
| 3.0 | 1 | 1.193 | 0.431 | 1.196 | 0.431 | 1.203 | 0.440 | 1.210 | 0.449 |
| 5.0 | 1 | 1.089 | 0.315 | 1.088 | 0.309 | 1.094 | 0.321 | 1.107 | 0.340 |
| 0.5 | 3 | 4.807 | 3.255 | 4.523 | 3.044 | 4.552 | 3.093 | 4.560 | 3.126 |
| 1.0 | 3 | 3.047 | 1.927 | 2.898 | 1.769 | 2.916 | 1.798 | 2.909 | 1.810 |
| 3.0 | 3 | 1.772 | 0.950 | 1.744 | 0.900 | 1.752 | 0.906 | 1.752 | 0.914 |
| 5.0 | 3 | 1.546 | 0.724 | 1.543 | 0.709 | 1.548 | 0.713 | 1.545 | 0.723 |
| 0.5 | 5 | 7.621 | 5.235 | 7.048 | 4.908 | 7.082 | 4.992 | 7.089 | 5.016 |
| 1.0 | 5 | 4.451 | 2.926 | 4.130 | 2.652 | 4.153 | 2.686 | 4.150 | 2.701 |
| 3.0 | 5 | 2.288 | 1.323 | 2.231 | 1.253 | 2.240 | 1.266 | 2.241 | 1.284 |
| 5.0 | 5 | 1.854 | 0.980 | 1.833 | 0.953 | 1.838 | 0.958 | 1.843 | 0.973 |
| 0.5 | 10 | 13.646 | 9.047 | 12.644 | 9.004 | 12.725 | 9.124 | 12.729 | 9.146 |
| 1.0 | 10 | 7.980 | 5.628 | 7.203 | 4.963 | 7.241 | 5.023 | 7.233 | 5.054 |
| 3.0 | 10 | 3.750 | 2.365 | 3.592 | 2.203 | 3.603 | 2.218 | 3.606 | 2.240 |
| 5.0 | 10 | 2.756 | 1.624 | 2.704 | 1.558 | 2.707 | 1.564 | 2.704 | 1.581 |
| 0.5 | 15 | 17.233 | 11.687 | 17.430 | 12.230 | 17.470 | 12.355 | 17.459 | 12.368 |
| 1.0 | 15 | 11.026 | 7.409 | 10.162 | 7.049 | 10.204 | 7.133 | 10.193 | 7.149 |
| 3.0 | 15 | 5.167 | 3.422 | 4.903 | 3.152 | 4.914 | 3.161 | 4.912 | 3.191 |
| 5.0 | 15 | 3.770 | 2.333 | 3.659 | 2.199 | 3.662 | 2.201 | 3.658 | 2.228 |
| 0.5 | 20 | 23.171 | 15.334 | 21.912 | 15.256 | 21.781 | 15.260 | 21.740 | 15.245 |
| 1.0 | 20 | 13.177 | 8.756 | 12.890 | 8.945 | 12.934 | 9.034 | 12.894 | 9.047 |
| 3.0 | 20 | 6.577 | 4.398 | 6.159 | 4.043 | 6.175 | 4.060 | 6.176 | 4.097 |
| 5.0 | 20 | 4.693 | 2.994 | 4.550 | 2.863 | 4.553 | 2.868 | 4.550 | 2.890 |

**Table 4.8** Moment comparison of $P_M$ for sample size n=100

| | | MME | | MLE | | MQLE | | CLE | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. | Mean | St. Dev. |
| 0.5 | 1 | 1.717 | 0.893 | 1.654 | 0.836 | 1.665 | 0.852 | 1.664 | 0.850 |
| 1.0 | 1 | 1.267 | 0.497 | 1.250 | 0.480 | 1.258 | 0.489 | 1.253 | 0.484 |
| 3.0 | 1 | 1.073 | 0.269 | 1.073 | 0.264 | 1.078 | 0.273 | 1.077 | 0.273 |
| 5.0 | 1 | 1.026 | 0.161 | 1.024 | 0.154 | 1.027 | 0.162 | 1.029 | 0.168 |
| 0.5 | 3 | 3.542 | 2.341 | 3.259 | 2.095 | 3.282 | 2.136 | 3.271 | 2.124 |
| 1.0 | 3 | 2.336 | 1.374 | 2.197 | 1.259 | 2.211 | 1.279 | 2.200 | 1.272 |
| 3.0 | 3 | 1.421 | 0.637 | 1.398 | 0.611 | 1.403 | 0.617 | 1.398 | 0.615 |
| 5.0 | 3 | 1.284 | 0.511 | 1.272 | 0.493 | 1.274 | 0.498 | 1.274 | 0.498 |
| 0.5 | 5 | 5.723 | 3.904 | 5.100 | 3.414 | 5.136 | 3.472 | 5.117 | 3.445 |
| 1.0 | 5 | 3.363 | 2.133 | 3.084 | 1.927 | 3.106 | 1.956 | 3.091 | 1.947 |
| 3.0 | 5 | 1.760 | 0.928 | 1.713 | 0.875 | 1.719 | 0.881 | 1.722 | 0.886 |
| 5.0 | 5 | 1.469 | 0.689 | 1.446 | 0.659 | 1.448 | 0.661 | 1.452 | 0.668 |
| 0.5 | 10 | 9.878 | 6.629 | 9.085 | 6.348 | 9.143 | 6.430 | 9.115 | 6.387 |
| 1.0 | 10 | 5.818 | 3.942 | 5.224 | 3.493 | 5.258 | 3.549 | 5.236 | 3.530 |
| 3.0 | 10 | 2.793 | 1.671 | 2.670 | 1.563 | 2.677 | 1.570 | 2.676 | 1.578 |
| 5.0 | 10 | 2.123 | 1.166 | 2.073 | 1.125 | 2.075 | 1.129 | 2.078 | 1.134 |
| 0.5 | 15 | 12.851 | 8.785 | 12.560 | 8.767 | 12.548 | 8.804 | 12.542 | 8.780 |
| 1.0 | 15 | 8.017 | 5.356 | 7.347 | 5.045 | 7.372 | 5.103 | 7.348 | 5.080 |
| 3.0 | 15 | 3.758 | 2.346 | 3.519 | 2.163 | 3.526 | 2.164 | 3.519 | 2.173 |
| 5.0 | 15 | 2.782 | 1.633 | 2.692 | 1.557 | 2.698 | 1.560 | 2.700 | 1.571 |
| 0.5 | 20 | 20.190 | 11.728 | 16.131 | 11.108 | 15.872 | 11.002 | 15.962 | 11.039 |
| 1.0 | 20 | 9.549 | 6.522 | 9.314 | 6.490 | 9.344 | 6.542 | 9.312 | 6.517 |
| 3.0 | 20 | 4.814 | 3.119 | 4.488 | 2.853 | 4.501 | 2.865 | 4.507 | 2.874 |
| 5.0 | 20 | 3.517 | 2.126 | 3.354 | 2.022 | 3.357 | 2.025 | 3.352 | 2.035 |

## 4.5 Discussion of results.

An examination of Tables 4.1, 4.2, 4.3 and 4.4 shows the results obtained from the simulation study are generally in broad agreement with the results of the other studies when the situations are comparable. On the occasion that there is some discrepancy (e.g. $n = 100$, $\alpha = 5$ and $\mu = 1$ where the results here differ somewhat from Willson et al. (1984) but agree with Anraku and Yanagimoto (1990)) the reason for the difference is not clear but a possible explanation is the number of estimators considered. As was noted earlier, a sample is discarded if one or more of the estimators of $\alpha$ fails for that sample. Hence, the more estimators one includes in a study the more samples that will be discarded, inevitably altering the estimates of the bias and MSE of the estimators. Now, of the estimators considered in this paper, Willson et al. (1984) include only the MME and MLE in their study whilst Anraku and Yanagimoto (1990) only exclude the MQLE. A better agreement would thus be expected with the latter's results.

With regard to the failure rate of the four estimators, the results obtained here agree with the results of others, bearing in mind that here the failure count corresponds to the number of failures in 10000 samples, not the number of samples that failed before 10000 successful samples were observed. For the MQLE it can be observed from Tables 4.1 to 4.4 that it failed for all samples for which the MLE and hence for which the MME and CLE failed. This is somewhat contrary to expectation, given that the MQLE always has smaller bias than the MLE. The reason for this appears, from a limited examination of the individual estimates, to be because the MQLE differs from $\alpha$ more than the MLE only when the latter is significantly larger than $\alpha$.

Before comparing the four estimators on the other criteria one further comment on the failure rate is warranted in light of the results of Chapter 3. In Section 3.5 we showed that the limiting distribution of the moment estimator for $\alpha$ when $\alpha$ and $\mu$ change with $n$ depends on the limit of $\lambda_n = n\mu_n^2/\alpha_n^2$, with the limiting distribution of the estimator having no finite moments if $\lim_{n\to\infty} \lambda_n < \infty$. This then is in agreement

with the results observed in Tables 4.1 to 4.4, that being that the erractic behaviour of the MME for $\alpha$ increases as $\mu$ decreases and $\alpha$ increases. Also, for fixed $\mu$ and $\alpha$ the erractic behaviour decreases as $n$ increases.

If an overall comparison of the four estimators is based on the absolute bias of the estimators alone, over the range of parameter values considered, it would appear that the CLE is generally superior but with the MQLE better for $n = 100$ and $\alpha < 1$. The MME on the otherhand, which is superior in terms of ease of computation, tends to perform worst for smaller values of $\alpha$ but outperforms the MLE and the MQLE as $\alpha$ increases. These results can be readily observed from Table 4.9 where the absolute bias of the estimators has been ranked from smallest to largest within each parameter and sample size configuration.

Similar comparisons of the four estimators can be based on the variance, mean square error and coefficient of variation. Again the CLE appears to generally outperform the other three estimators in all three categories over the range of parameter values considered. For those parameter configurations for which it is outperformed on a particular category it is then ranked second.

As a further comment on the specific results given in Tables 4.1 to 4.4 it can be seen that for certain parameter values (e.g. $n = 50$, $\mu = \alpha = 5$) the bias and MSE of the MLE and MQLE are significantly worse than for the other two estimators. The reason for this is that the MLE and MQLE have a higher tendency than the other two estimators considered to return the occasional very extreme finite estimate for $\alpha$. The results in Tables 4.1 - 4.4 are therefore dependent on the upper limit chosen to correspond to effectively infinite estimates of $\alpha$. In practice this upper limit would often be chosen significantly less than 10000 in which case the differences referred to in the tables would be reduced. Hence the need for an alternative assessment criterion.

The conclusions change when assessment is based on the second criterion, the percentile measure, the summary statistics for which are given in Tables 4.5 to 4.8.

**Table 4.9** Ranked absolute biases for different sample sizes.

| | | n=20 | | | | n=30 | | | | n=50 | | | | n=100 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | A | B | C | D | A | B | C | D | A | B | C | D | A | B | C | D |
| 0.5 | 1 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 1.0 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 1 |
| 3.0 | 1 | 1 | 4 | 3 | 2 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 |
| 5.0 | 1 | 1 | 3 | 2 | 4 | 1 | 4 | 3 | 2 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 |
| 0.5 | 3 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 1.0 | 3 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 3.0 | 3 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 3 | 4 | 2 | 1 |
| 5.0 | 3 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 |
| 0.5 | 5 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 1.0 | 5 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 3.0 | 5 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 1 |
| 5.0 | 5 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 |
| 0.5 | 10 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 1.0 | 10 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 3.0 | 10 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 |
| 5.0 | 10 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 |
| 0.5 | 15 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 1.0 | 15 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 3.0 | 15 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 3 | 3 | 2 | 1 | 4 | 3 | 2 | 1 |
| 5.0 | 15 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 |
| 0.5 | 20 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 1.0 | 20 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 3.0 | 20 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 4 | 3 | 2 | 1 | 4 | 3 | 2 | 1 |
| 5.0 | 20 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 3 | 1 |

LEGEND: A-$\hat{\alpha}_{MME}$; B-$\hat{\alpha}_{MLE}$; C-$\hat{\alpha}_{MQLE}$; D-$\hat{\alpha}_{CLE}$

**Table 4.10** Ranked average $P_M$ for different sample sizes.

| | | n=20 | | | | n=30 | | | | n=50 | | | | n=100 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | A | B | C | D | A | B | C | D | A | B | C | D | A | B | C | D |
| 0.5 | 1 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 |
| 1.0 | 1 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 3 | 1 | 4 | 2 | 4 | 1 | 3 | 2 |
| 3.0 | 1 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 1 | 4 | 3 |
| 5.0 | 1 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 2 | 1 | 3 | 4 | 2 | 1 | 3 | 4 |
| 0.5 | 3 | 3 | 1 | 2 | 4 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 |
| 1.0 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 | 4 | 1 | 3 | 2 |
| 3.0 | 3 | 2 | 1 | 3 | 4 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 2 | 4 | 1 | 3 | 1 |
| 5.0 | 3 | 2 | 1 | 3 | 4 | 1 | 2 | 3 | 4 | 3 | 1 | 4 | 2 | 4 | 1 | 2 | 2 |
| 0.5 | 5 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 |
| 1.0 | 5 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 | 4 | 1 | 3 | 2 |
| 3.0 | 5 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 |
| 5.0 | 5 | 3 | 1 | 2 | 4 | 3 | 1 | 2 | 4 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 |
| 0.5 | 10 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 |
| 1.0 | 10 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 | 4 | 1 | 3 | 2 |
| 3.0 | 10 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 |
| 5.0 | 10 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 1 | 4 | 1 | 2 | 3 |
| 0.5 | 15 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 4 | 3 | 4 | 3 | 2 | 1 |
| 1.0 | 15 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 | 4 | 1 | 3 | 2 |
| 3.0 | 15 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 | 4 | 1 | 3 | 1 |
| 5.0 | 15 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 2 | 3 | 1 | 4 | 1 | 2 | 3 |
| 0.5 | 20 | 1 | 2 | 3 | 4 | 1 | 4 | 3 | 2 | 4 | 3 | 2 | 1 | 4 | 3 | 1 | 2 |
| 1.0 | 20 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 | 4 | 1 | 3 | 2 | 4 | 2 | 3 | 1 |
| 3.0 | 20 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 |
| 5.0 | 20 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 1 | 4 | 2 | 3 | 1 |

LEGEND: A-$\hat{\alpha}_{MME}$; B-$\hat{\alpha}_{MLE}$; C-$\hat{\alpha}_{MQLE}$; D-$\hat{\alpha}_{CLE}$

Consider the average $P_M$ values, the ranked values of which are given in Table 4.10 for the different parameter and sample size configurations. Here the MME performs the worst of the four estimators except for the smaller values of $n$ and $\mu$, in which case it performs best. The MQLE and the CLE generally perform worse than the MLE based on this criterion. Of possibly greater interest however is how small the actual differences are between the averages for the four estimators for most values of $n$, $\alpha$ and $\mu$ using this criterion. Therefore, if ease of computation is a consideration the MME which can be obtained explicitly, must be considered.

In terms of the standard deviations of the $P_M$ the MLE again generally tends to perform the best of the four estimators but here, as before, the MME estimator can not be ruled out owing to the small differences.

One may now ask if the findings based on the percentile measure would have been the same had a different choice of A been used. To help answer this question a limited, independent, simulation study was conducted with A = [0.1, 0.9] comprising only 1000 simulations for the case $n = 20$ and for a restricted set of the $\alpha$ and $\mu$ values considered above. The results of this limited study gave almost identical conclusions to those above.

As a final comment, it is difficult on the basis of the study to find any grounds upon which to recommend the MQLE. As an estimator of $\alpha$ it cannot be solved for explicitly, fails more often than the three other estimators and is generally outperformed by at least one of the other three estimators on either criterion.

# Chapter 5

# Miscellany on the negative binomial distribution.

In the previous chapters we considered estimation for mixed gamma-Poisson models for which the negative binomial distribution (NBD) is a special case. In all that work it was assumed implicitly that the sample contained no contamination, that is, data from alternative distributions. For the remaining chapters we focus attention on robust estimation and related matters.

In this chapter we retain our focus on the NBD and give some results associated with robustness. Firstly, given the close relationship between the median and robustness we obtain in Section 5.1 bounds for the median of a NBD. From these bounds it follows that the median is not necessarily an appropriate robust estimate for the mean of a NBD and instead we will show that a quantile other than the $50^{th}$ percentile should be used. In Section 5.2 we obtain bounds for the mean in terms of this alternative quantile which depends on $\alpha$. This bound is then used in Section 5.3 to obtain a robust estimator for the NBD mean when $\alpha$ is known. In the concluding remarks, Section 5.4, we discuss briefly the robust estimation problem for the negative binomial distribution.

## 5.1 Bounds for the Median of the Negative Binomial Distribution.

For the beta distribution Groeneveld and Meeden (1977) outlined a procedure by which it can be shown simply that the median lies between the mean and mode, provided that the two parameters of the distribution are strictly greater than 1. Using this result Payton, Young and Young (1989) obtained bounds for the "median" of the negative binomial distribution (NBD) when the scale parameter is strictly greater than one.

Unfortunately, the value(s) treated by Payton et al. (1989) as median values for the NBD do not, except in the case where there exists a value of $x$ such $P(X \leq x) = \frac{1}{2}$, satisfy the generally accepted definition of the median of a random variable X, that being any value ($m$ say) satisfying $P(X < m) \leq \frac{1}{2} \leq P(X \leq m)$. The bounds thus obtained are not necessarily correct for this definition of the median (e.g. see Payton et al. (1989), *Remark 2*).

This section hence has two objectives. Firstly, we modify the results of Payton et al. (1989) so that they hold for a rigorously defined median. The definition we employ is

$$\text{Med}(X) \;=\; \inf\{x : P(X \leq x) \geq \tfrac{1}{2}\} \,. \tag{5.1}$$

All reference hereafter to the median will refer to this value. Secondly, we extend the results obtained by Payton et al. (1989) to cover the complete family of NBDs. In achieving this we extend the results available for the median of the beta distribution.

Let $Y = G_{p,q}$ be a beta$(p, q)$ random variable with distribution function $G_{p,q}(y)$ and density function $g_{p,q}(y)$ given by

$$g_{p,q}(y) \;=\; \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1} \quad ; \quad 0 \leq y \leq 1;\; p > 0,\; q > 0 \,.$$

**Theorem 5.1**

If Y has a beta(p,q) distribution then

$$0 < \text{Med}(Y) < E(Y) < \tfrac{1}{2} \;\Leftrightarrow\; p < q \,,$$

$$\text{Med}(Y) = E(Y) = \tfrac{1}{2} \;\Leftrightarrow\; p = q\,,$$
$$\tfrac{1}{2} < E(Y) < \text{Med}(Y) < 1 \;\Leftrightarrow\; p > q\,.$$

**Proof**

To obtain this result we separate the parameter space into the disjoint regions $p = q$, $p < q$ and $p > q$. By symmetry, if $p = q$ then $\text{Med}(Y) = E(Y) = \frac{1}{2}$. For the region $p < q$ we now show, by considering each of the following sub-regions $1 < p < q$, $1 = p < q$, $p < 1 \leq q$ and $p < q < 1$ separately, that

$$0 < \text{Med}(Y) < E(Y) < \tfrac{1}{2} \tag{5.2}$$

for $p < q$.

For the sub-region $1 < p < q$, (5.2) immediately follows from the result of Groeneveld and Meeden(1977) that

$$0 < \frac{p-1}{p+q-2} = \text{Mode}(Y) < \text{Med}(Y) < E(Y) < \tfrac{1}{2} \;\;;\;\; 1 < p < q\,. \tag{5.3}$$

For $1 = p < q$, $E(Y) = 1/(1+q) < 1/2$ and $\text{Med}(Y) = 1 - (1/2)^q$ and then (5.2) follows from $(q/(1+q))^q < 1/2$.

For the remaining two sub-regions the approach outlined in Groeneveld and Meeden (1977) is used, beginning with $p < 1 \leq q$. Let $W$ be a continous random variable with density and cumulative distribution $h(w)$ and $H(w)$ respectively, with

$$h(w) = I\,(w \in (0,m))\, g_{p,q}(w) + I\,(w \in (m,2m))\, g_{p,q}(2m-w)$$

where $m = \text{Med}(Y) < \frac{1}{2}$. The latter inequality follows since $g_{p,q}(y)$ is decreasing on [0,1]. Obviously $h(w)$ depends on $p$ and $q$ but this dependence has been supressed. Now, since $G_{p,q}(y) \leq H(y)\ \forall\ y$ and $G_{p,q}(y) < H(y)$ for some $y$, we have

$$\begin{aligned} E(Y) &= \int_0^\infty (1 - G_{p,q}(y))\, dy \\ &> \int_0^\infty (1 - H(w))\, dw \\ &= E(W) \end{aligned}$$

That (5.2) is thus satisfied for $p < 1 \leq q$ follows as $E(W) = \text{Med}(Y)$ by symmetry.

For $p < q < 1$ we have $E(Y) = \frac{p}{p+q} < \frac{1}{2}$. Further, for all $y \in (0, 0.5)$

$$\begin{aligned} \frac{g_{p,q}(0.5-y)}{g_{p,q}(0.5+y)} &= \left(\frac{0.5-y}{0.5+y}\right)^{p-1}\left(\frac{0.5+y}{0.5-y}\right)^{q-1} \\ &> \left(\frac{0.5-y}{0.5+y}\right)^{q-1}\left(\frac{0.5+y}{0.5-y}\right)^{q-1} \\ &= 1 \end{aligned}$$

implying that $\text{Med}(Y) < \frac{1}{2}$. Equation (5.2) can then be shown to hold for this sub-region by defining a random variable W as for the sub-region $p < 1 \leq q$ above and repeating the same argument.

Hence we have shown (5.2) holds for all $p < q$. That the result

$$\tfrac{1}{2} < E(Y) < \text{Med}(Y) < 1 \quad \Leftrightarrow \quad p > q$$

holds then follows by symmetry and hence we have proved our result.

•

Now let $X$ have a negative binomial distribution with mean $\mu$ $(> 0)$, shape parameter $\alpha$ $(> 0)$ and cumulative distribution function $F(x)$. Using the result of Patil (1960)

$$F(x) = G_{\alpha,\,[x]+1}\left(\tfrac{\alpha}{\alpha+\mu}\right) \quad ; \quad x \geq 0\,, \tag{5.4}$$

where $[x]$ denotes the largest integer not greater than $x$, together with a tightening of the procedure employed by Payton et al. (1989) we obtain the following result for the median of a NBD$(\mu, \alpha)$.

**Theorem 5.2**

If $\eta = \text{Med}(X)$ then for all $\alpha > 0$ and $\mu > 0$

$$\begin{aligned} [(\tfrac{\alpha-1}{\alpha})\mu] + 1 \;\leq\; \eta \;<\; \mu \qquad & \text{if } \mu > \alpha \\ ]\mu[ - 1 \;\leq\; \eta \;<\; (\tfrac{\alpha-1}{\alpha})\mu + 1 \qquad & \text{if } \mu \leq \alpha \end{aligned}$$

where $]x[$ $(= -[-x])$ denotes the smallest integer not less than $x$.

**Proof**

In the proof below use is made of the fact that, by definition, $\eta$ is an integer.

**Consider the case $\mu > \alpha$.**

From (5.1) and (5.4) we have $\frac{1}{2} \leq \mathrm{F}(\eta) = G_{\alpha,\eta+1}(\frac{\alpha}{\alpha+\mu})$ giving

$$\mathrm{Med}(G_{\alpha,\eta+1}) \leq (\frac{\alpha}{\alpha+\mu}) < \frac{1}{2}\,. \tag{5.5}$$

So by Theorem 5.1 we obtain $\mathrm{E}(G_{\alpha,\eta+1}) < \frac{1}{2}$ and hence

$$\eta > \alpha - 1 \qquad \text{for all } \mu > \alpha\,. \tag{5.6}$$

For $\eta \geq \alpha$ we have, again by Theorem 5.1,

$$\mathrm{Med}(G_{\alpha,\eta}) \leq E(G_{\alpha,\eta}) = (\tfrac{\alpha}{\alpha+\eta})\,,$$

where the last term is strictly greater than $(\frac{\alpha}{\alpha+\mu})$ because

$$\tfrac{1}{2} > F(\eta-1) = G_{\alpha,\eta}(\tfrac{\alpha}{\alpha+\mu}) \;\Rightarrow\; \mathrm{Med}(G_{\alpha,\eta}) > (\tfrac{\alpha}{\alpha+\mu})\,.$$

From here we obtain the result $\eta < \mu$ whenever $\eta \geq \alpha$ and hence

$$\eta < \mu \qquad \text{for all } \mu > \alpha\,. \tag{5.7}$$

For the case $\alpha > 1$, we have, from (5.6), (5.3) and (5.5)

$$(\tfrac{\alpha-1}{\alpha+\eta-1}) = \mathrm{Mode}(G_{\alpha,\eta+1}) < \mathrm{Med}(G_{\alpha,\eta+1}) \leq (\tfrac{\alpha}{\alpha+\mu})$$

from which it follows that

$$\eta > (\tfrac{\alpha-1}{\alpha})\mu \qquad \text{for all } 1 < \alpha < \mu\,. \tag{5.8}$$

Equations (5.6), (5.7) and (5.8) can be combined to give

$$[(\tfrac{\alpha-1}{\alpha})\mu] + 1 \leq \eta < \mu \qquad \text{if } \mu > \alpha\,.$$

**For the case** $\mu \leq \alpha$ the result is obtained similarly. First, since $\frac{1}{2} \leq F(\eta) = G_{\alpha,\eta+1}(\frac{\alpha}{\alpha+\mu})$ we have

$$\text{Med}(G_{\alpha,\eta+1}) \leq \frac{\alpha}{\alpha+\mu} \tag{5.9}$$

whilst $\frac{1}{2} > F(\eta-1) = G_{\alpha,\eta}(\frac{\alpha}{\alpha+\mu})$ implies that

$$\text{Med}(G_{\alpha,\eta}) > \frac{\alpha}{\alpha+\mu} \geq \frac{1}{2} \tag{5.10}$$

Hence, by Theorem 5.1 we have $\text{E}(G_{\alpha,\eta}) = \frac{\alpha}{\alpha+\eta} > \frac{1}{2}$, since $\text{Med}(G_{\alpha,\eta}) > \frac{1}{2}$, and thus

$$\eta \ < \ \alpha \quad \text{for all} \quad \mu \leq \alpha \, . \tag{5.11}$$

For $\eta \leq \alpha - 1$ we have $\text{E}(G_{\alpha,\eta+1}) = \frac{\alpha}{\alpha+\eta+1} \geq \frac{1}{2}$ so that by Theorem 5.1 and (5.9)

$$\frac{1}{2} \leq \frac{\alpha}{\alpha+\eta+1} = \text{E}(G_{\alpha,\eta+1}) \leq \text{Med}(G_{\alpha,\eta+1}) \leq \frac{\alpha}{\alpha+\mu} \tag{5.12}$$

giving $\eta \geq \mu - 1$. On the otherhand, $\eta > \alpha - 1$ implies $\eta > \mu - 1$, thus giving

$$\mu - 1 \leq \eta < \alpha \quad \text{for all} \quad \mu \leq \alpha \, . \tag{5.13}$$

To tighten the upper bound note that for $\alpha > 1$, using the above results and the complementary result to (5.3) given by Groeneveld and Meeden (1977), that being,

$$\frac{1}{2} < E(Y) < \text{Med}(Y) < \text{Mode}(Y) = \frac{p-1}{p+q-2} \quad ; \quad 1 < q < p$$

that

$$\frac{1}{2} < \text{E}(G_{\alpha,\eta}) < \text{Med}(G_{\alpha,\eta}) \leq \frac{\alpha-1}{\alpha+\eta-2}$$

since $\alpha > \eta$. This combined with (5.10) gives $\eta < (\frac{\alpha-1}{\alpha})\mu + 1$. For $\alpha \leq 1$ we have by (5.11) that $\eta = 0$. Combining the above and using the information that $\eta$ is an integer gives the second part of the result in Theorem 5.2 .

•

**Corollary 5.1**

If $\mu < \alpha$ and both $\mu$ and $\alpha$ are integers then $\eta = \mu$.

**Proof**

In this case, from (5.11) and since $\alpha$ is an integer we have $\eta \leq \alpha - 1$.

If $\eta = \alpha - 1$ we have $\eta \geq \mu$ since $\mu \leq \alpha - 1$.

For $\eta < \alpha - 1$, using a similar argument to (5.12), we obtain

$$\frac{1}{2} < \frac{\alpha}{\alpha + \eta + 1} = \mathrm{E}(G_{\alpha,\eta+1}) < \mathrm{Med}(G_{\alpha,\eta+1}) \leq \frac{\alpha}{\alpha + \mu}$$

from which it follows that $\eta > \mu - 1$.

Hence $\eta \geq \mu$ and thus the corollary follows from Theorem 5.2 .

•

Bounds for the median of the negative binomial distribution in the restricted case of $\alpha$ an integer are also given in Göb (1994) using a different parametrization. The method used to obtain these bounds is different from the method above but the bounds are equivalent to those given in Theorem 5.2 and Corollary 5.1.

### 5.1.1 Remarks on Theorem 5.2 .

Let $X \sim NBD(\mu, \alpha)$ and $\eta = \mathrm{Med(X)}$.

**1.** If $\mu = \alpha$ then $\eta = ]\mu[ - 1$.

**2.** Let $M = (\frac{\alpha - 1}{\alpha})\mu$. The mode of $X$ is $([M] \vee 0)$ if $M$ is not a positive integer whilst the modal values are $M$ and $M - 1$ if $M$ is a positive integer. From the lower bounds in Theorem 5.2 we have $\eta$ always bounded below by the mode. Moreover, if $M$ is a positive integer and $\alpha \geq \mu$ then $\eta = M$.

**3.** It is not the case that the median of X is always bounded above by the mean of $X$. For example, if $X \sim NBD(1.9, 3)$ then $\eta = 2$.

**4.** If $\alpha$ and $\eta$ both take values in (0,1] then $\eta = 0$.

**5.** For the case $\mu > \alpha$ it can be seen from Theorem 5.2 that the bounds for the median of $X \sim NBD(\mu, \alpha)$ are not very tight. For example, for a $NBD(20, 1.4)$ distribution the bounds for the median are $6 \leq \eta \leq 19$, whilst the actual median is 15. However, as

$$\frac{X}{\mu} \xrightarrow{d} \Gamma(\alpha, \tfrac{1}{\alpha}) \quad \text{as } \mu \to \infty \ (\alpha \text{ fixed}),$$

there would appear little scope for relative tightening if the bounds obtained by Chen and Rubin (1986) for a gamma random variable

$$(\alpha - 1/3) < \text{Med}(\Gamma(\alpha, 1)) < \alpha$$

are tight as these give the following bounds

$$\left(\frac{\alpha - \frac{1}{3}}{\alpha}\right) < \lim_{\mu \to \infty} \text{median} \left(\frac{X}{\mu}\right) < 1 \, .$$

## 5.2 Percentile related bounds for the mean of the negative binomial distribution.

For the Poisson and binomial distributions bounds are available for the median in terms of the mean, where the median is defined as $\inf\{x : P(X \leq x) \geq 0.5\}$. In particular, if X has a Poisson distribution with mean $\lambda$ and median $m$, then from Teicher (1955) we obtain the result

$$[\lambda] \ \leq m < \ \lambda + 1 \, .$$

The corresponding result for the binomial random variable $B(n, p)$, obtained by Choi and He (1991), is as above but with $\lambda$ replaced by $np$.

Reversing the above result we obtain the following bounds for the mean of the Poisson random variable:

$$m - 1 \ < \lambda < \ m + 1 \, , \tag{5.14}$$

with an equivalent result holding for the binomial random variable. It thus follows that in both cases the mean and median are equal when the former is an integer.

(Note: For the Poisson case Chen and Rubin (1986) obtain the tighter bounds $m - \frac{1}{3} < \lambda < m + 1$ .)

Using (5.14), the sample median can be justified as a robust estimator of the population mean for either a Poisson or binomial random variable, with the estimator having bounded influence and finite sample breakdown point of $\frac{1}{2}$. The estimator is however not very efficient and, as can be seen from (5.14), may exhibit some bias. The sample median will though be useful for those cases where there may be outliers or where an initial choice of robust estimator is required for the development of a more efficient robust estimator. This second situation includes the case when the latter estimator has influence function dependent on the initial choice of estimator.

If consideration is now given to the negative binomial distribution having mean $\mu$ $(> 0)$ shape parameter $\alpha$ $(> 0)$, written NBD$(\mu, \alpha)$, the sample median is no longer a satisfactory estimator for the population mean. This follows from the result that the population median and mean may differ significantly, as can be seen from Theorem 5.2.

To overcome this problem, we obtain equivalent bounds to (5.14) for the mean $\mu$ of an NBD$(\mu, \alpha)$. Here though it is necessary to replace the median in (5.14) with a percentile dependent on $\alpha$.

Let $Y$ have a gamma distribution with parameter $\alpha$ such that Y has density $g(y) = y^{\alpha-1}e^{-y}/\Gamma(\alpha)$ for $y > 0$ , and let

$$p_\alpha \;=\; P(Y \leq \alpha)\,, \tag{5.15}$$

where $p_\alpha > \frac{1}{2}$ since the median of Y is less than its mean.

**Theorem 5.3**

If $X \sim NBD(\mu, \alpha)$ and $m_\alpha = \inf\{x : P(X \leq x) \geq p_\alpha\}$ , then

$$m_\alpha - 1 \;<\mu<\; m_\alpha + 1\,, \tag{5.16}$$

with $\mu = m_\alpha$ if $\mu$ is an integer.

•

Before proving Theorem 5.3 it is necessary to obtain a few minor results pertaining to a function $A(x;\mu,\alpha)$ defined as

$$A(x;\mu,\alpha) = P(X \leq x) \quad \text{where} \quad X \sim NBD(\mu,\alpha)\,, \tag{5.17}$$

and which, using (5.4), satisfies

$$A(x;\mu,\alpha) \;=\; \int_0^{\frac{\alpha}{\alpha+\mu}} \frac{\Gamma(\alpha+[x]+1)}{\Gamma(\alpha)\,[x]!}\; t^{\alpha-1}\,(1-t)^{[x]}\,dt\,, \tag{5.18}$$

provided $x$ is non-negative. These results, and the method of proof of the above theorem, are in effect an extension of the result obtained by Teicher (1955).

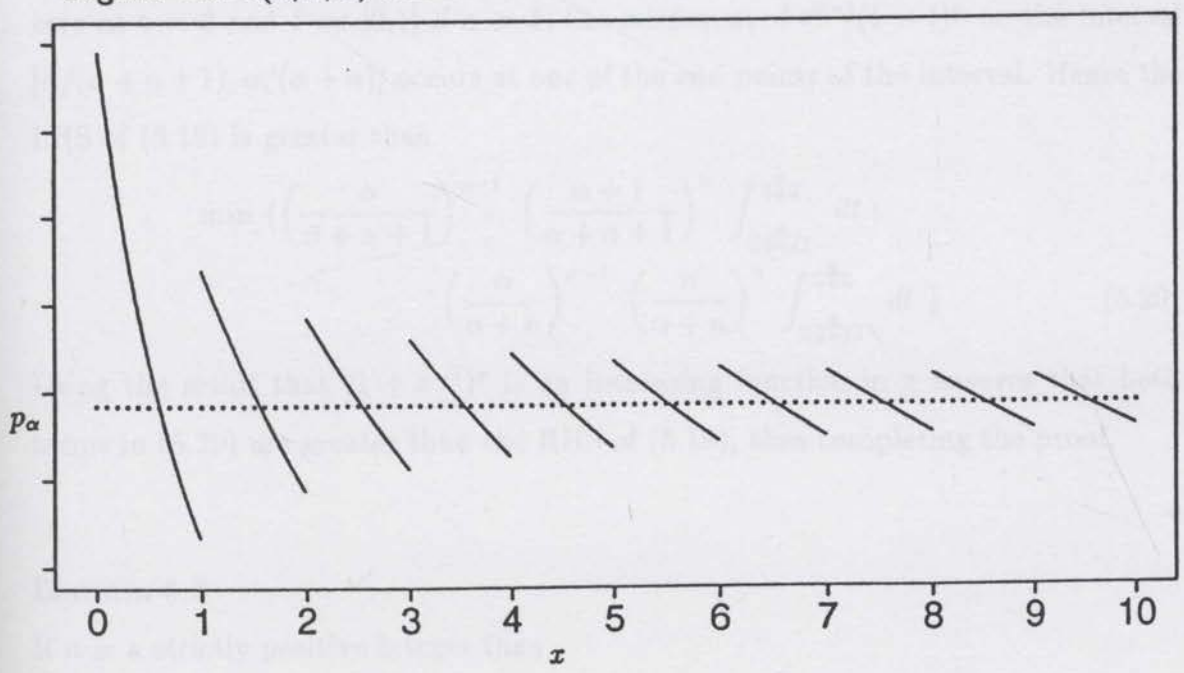
**Lemma 5.1**

For fixed $x$ $(\geq 0)$ and $\alpha$ $(> 0)$ the function $A(x;\mu,\alpha)$ is monotonically decreasing in $\mu$.

•

The remaining lemmas refer to the function $A(x;x,\alpha)$ as a function in $x$ for a given $\alpha$. This is a piecewise continuous function and is illustrated in Figure 5.1.

**Figure 5.1** $A(x;x,\alpha)$ versus $x$



**Lemma 5.2**

Define $A(0;0,\alpha)$ to be 1 and let $n$ be a non-negative integer. Then

$$A(n;n,\alpha) \;>\; A(n+1;n+1,\alpha)\,.$$

**Proof**

As the result obviously holds for $n = 0$ we need only consider the case $n \geq 1$.

Let

$$\begin{aligned} c_n &= A(n;n,\alpha) - A(n;n+1,\alpha) \\ &= \int_{\frac{\alpha}{\alpha+n+1}}^{\frac{\alpha}{\alpha+n}} \frac{\Gamma(\alpha+n+1)}{\Gamma(\alpha)\,n!}\, t^{\alpha-1}\,(1-t)^n\,dt \end{aligned}$$

and

$$\begin{aligned} d_n &= A(n+1;n+1,\alpha) - A(n;n+1,\alpha) \\ &= \frac{\Gamma(\alpha+n+1)}{\Gamma(\alpha)\,(n+1)!}\left(\frac{\alpha}{\alpha+n+1}\right)^{\alpha}\left(\frac{n+1}{\alpha+n+1}\right)^{n+1}. \end{aligned}$$

To prove Lemma 5.2 it suffices to prove $c_n > d_n$ , or equivalently,

$$\int_{\frac{\alpha}{\alpha+n+1}}^{\frac{\alpha}{\alpha+n}} t^{\alpha-1}\,(1-t)^n\,dt \;>\; \frac{1}{n+1}\left(\frac{\alpha}{\alpha+n+1}\right)^{\alpha}\left(\frac{n+1}{\alpha+n+1}\right)^{n+1}. \qquad (5.19)$$

Now since $t^{\alpha-1}(1-t)^n$ is decreasing on [0,1] if $\alpha \leq 1$, and is unimodal with the value zero at t = 0 and 1 on [0,1] if $\alpha > 1$, the minimum of $t^{\alpha-1}(1-t)^n$ on the interval $[\alpha/(\alpha+n+1), \alpha/(\alpha+n)]$ occurs at one of the end points of the interval. Hence the LHS of (5.19) is greater than

$$\begin{aligned} \min\{&\left(\frac{\alpha}{\alpha+n+1}\right)^{\alpha-1}\left(\frac{n+1}{\alpha+n+1}\right)^{n}\int_{\frac{\alpha}{\alpha+n+1}}^{\frac{\alpha}{\alpha+n}} dt\,; \\ &\left(\frac{\alpha}{\alpha+n}\right)^{\alpha-1}\left(\frac{n}{\alpha+n}\right)^{n}\int_{\frac{\alpha}{\alpha+n+1}}^{\frac{\alpha}{\alpha+n}} dt\,\}\,. \end{aligned} \qquad (5.20)$$

Using the result that $(1+x^{-1})^x$ is an increasing function in x ensures that both terms in (5.20) are greater than the RHS of (5.19), thus completing the proof.

•

**Lemma 5.3**

If $n$ is a strictly positive integer then

$$A(n-1;n,\alpha) \;<\; A(n;n+1,\alpha)\,.$$

**Proof**

With $c_n$ and $d_n$ defined as in Lemma 5.2, the result follows if $c_n < d_{n-1}$, or equivalently

$$(\alpha+n)\int_{\frac{\alpha}{\alpha+n+1}}^{\frac{\alpha}{\alpha+n}} t^{\alpha-1}(1-t)^n\,dt \;<\; \left(\frac{\alpha}{\alpha+n}\right)^{\alpha}\left(\frac{n}{\alpha+n}\right)^{n}. \qquad (5.21)$$

Now, for $\alpha \leq n+1$ equation (5.21) follows readily by noting that $t^{\alpha-1}(1-t)^n$ has its maximum at $t = \alpha/(\alpha+n+1)$ on the interval $[\alpha/(\alpha+n+1), \alpha/(\alpha+n)]$. For $\alpha > n+1$ the integrand in (5.21) attains its maximum on the interval of integration at the point $t = (\alpha-1)/(\alpha+n-1)$. Hence, for $\alpha > n+1$ the LHS of (5.21) is less than

$$\begin{aligned}
&(\alpha+n)\left(\frac{\alpha-1}{\alpha+n-1}\right)^{\alpha-1}\left(\frac{n}{\alpha+n-1}\right)^{n}\int_{\frac{\alpha}{\alpha+n+1}}^{\frac{\alpha}{\alpha+n}} dt \\
= &\left(\frac{\alpha+n}{\alpha+n+1}\right)\left(\frac{\alpha+n}{\alpha+n-1}\right)^{\alpha+n-1}\left(\frac{\alpha-1}{\alpha}\right)^{\alpha-1}\left(\frac{\alpha}{\alpha+n}\right)^{\alpha}\left(\frac{n}{\alpha+n}\right)^{n} \\
< &\left(\frac{2\alpha-1}{2\alpha}\right)\left(\frac{2\alpha-1}{2\alpha-2}\right)^{2\alpha-2}\left(\frac{\alpha-1}{\alpha}\right)^{\alpha-1}\left(\frac{\alpha}{\alpha+n}\right)^{\alpha}\left(\frac{n}{\alpha+n}\right)^{n}.
\end{aligned}$$

This last inequality holds since $n < \alpha - 1$. The lemma then follows on noting that the last expression is less than the RHS of (5.21).

•

**Lemma 5.4**

$$\lim_{\mu\to\infty} A(\mu;\mu,\alpha) \;=\; p_\alpha\,,$$

where $p_\alpha$ is defined in (5.15).

**Proof**

The result follows directly from the result that if $X \sim NBD(\mu,\alpha)$ with $\alpha$ fixed, then

$$\alpha\,\frac{X}{\mu} \;\xrightarrow{d}\; Y \quad \text{as } \mu\to\infty\,,$$

where $Y$ is a gamma random variable with parameter $\alpha$. See Pessin (1961).

•

We are now in a position to prove main result.

**Proof of Theorem 5.3**

If $\mu$ is an integer, then by Lemmas 5.2 and 5.4, $A(\mu;\mu,\alpha) > p_\alpha$. For $\mu$ not an integer $A([\mu]+1;\mu,\alpha) > A([\mu]+1;[\mu]+1,\alpha)$, by Lemma 5.1, which in turn is greater than $p_\alpha$ since $[\mu]+1$ is an integer. Hence $A([\mu]+1,\mu,\alpha) > p_\alpha$ . Using a similar argument, but employing Lemma 5.3 instead of Lemma 5.2 we also have $A([\mu]-1;\mu,\alpha) < p_\alpha$. Hence $[\mu] \leq m_\alpha < \mu+1$ and from this Theorem 5.3 follows.

•

## 5.3 A percentile based estimator for the mean when $\alpha$ known.

Based on (5.16), the logical extension of using the median to estimate the mean of a binomial or Poisson random variable is to use the $p_\alpha^{th}$ sample percentile to estimate the mean of a NBD$(\mu,\alpha)$ random variable when $\alpha$ is known. This would then give a robust estimator with bounded influence function and finite sample breakdown point equal to $1-p_\alpha$.

However, instead of using $\hat{m}_\alpha$ as the estimator of $\mu$ we propose the following modification which sacrifices none of the robustness properties mentioned earlier:

$$\begin{aligned} \hat{\mu} = \;& \alpha(\hat{F}^{-\frac{1}{\alpha}}(0)-1) \quad \text{if } \hat{m}_\alpha = 0 \\ = \;& \hat{\xi}+.5 \qquad \text{if } \hat{m}_\alpha > 0, \end{aligned} \tag{5.22}$$

where

$$\hat{\xi} = \max\{0\ ,\ (\hat{m}_\alpha-1)+\frac{p_\alpha-\hat{F}(\hat{m}_\alpha-1)}{\hat{F}(\hat{m}_\alpha)-\hat{F}(\hat{m}_\alpha-1)}\}; \tag{5.23}$$

$\hat{F}(x)$ denotes the usual empirical distribution function; and

$$\hat{m}_\alpha = \inf\{x : \hat{F}(x) \geq p_\alpha\}.$$

(Note: The value $\xi$, obtained by replacing estimated values by population values in (5.23), is the $p_\alpha^{th}$ percentile of $F_S(x)$, a distribution function related to $F(x)$ by $F_S(x) = F([x]) + (x-[x])\,(F([x]+1)-F([x]))$ for $x \geq 0$ and zero otherwise.)

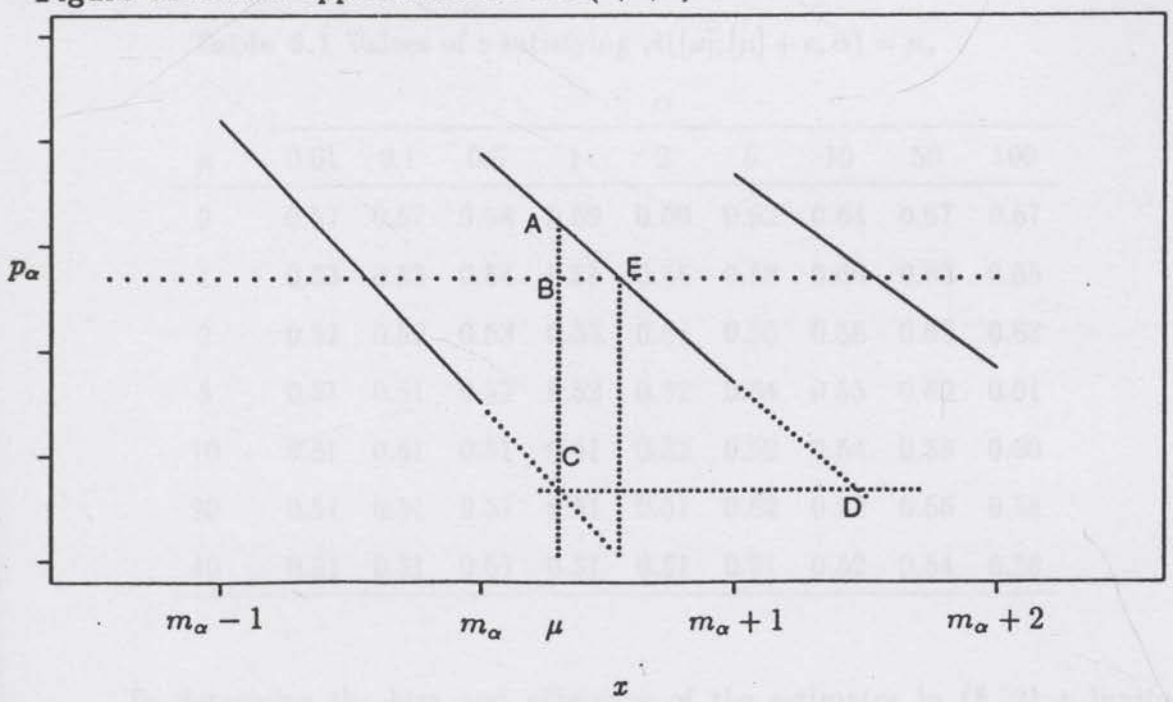
This estimator has two advantages over $\hat{m}_\alpha$. Firstly, it avoids giving a zero estimate, except in the case when all data values are zero. This is achieved by using a zero-class estimator whenever $\hat{m}_\alpha = 0$. Secondly, it contains an adjustment to $\hat{m}_\alpha$ which attempts to correct for its bias, the justification for which is based on the following argument.

Let

$$A([\mu]; [\mu] + c, \alpha) \;=\; p_\alpha \tag{5.24}$$

where $c$, using the results in Section 5.2, can be shown to lie in the interval $(0, 1)$. Also, approximate the continuous segments of $A(x; x, \alpha)$, considered as a function of $x$, by straight line segments. We now consider the case $\mu \geq m_\alpha$ (and hence $\mu \leq m_\alpha + c$ ) and use a geometrical argument which is illustrated with the aid of Figure 5.2, wherein the solid lines represent the linear approximations to $A(x; x, \alpha)$

**Figure 5.2** Linear approximations to $A(x; x, \alpha)$ versus $x$



on the interval $[m_\alpha - 1, m_\alpha + 2)$. Since the lengths AB, BE, AC and CD are approximately $F(m_\alpha) - p_\alpha$, $m_\alpha + c - \mu$, $F(m_\alpha) - F(m_\alpha - 1)$ and 1 respectively, where $F(x) = A(x; \mu, \alpha)$, we obtain the result that

$$\mu \approx \xi + c, \tag{5.25}$$

where $\xi$ is defined as above. For the case $\mu < m_\alpha$ ( and hence $\mu > m_\alpha - 1 + c$) equation (5.25) also holds and can be shown using an equivalent argument.

It remains now to determine the value of $c$ which satisfies (5.24). This value is a function of both $\mu$ and $\alpha$ but appears from examination over a wide range of values for both parameters, some of which are given in Table 5.1, to decrease as $\mu$ increases and as $\alpha$ decreases, quickly approaching a limit of $\frac{1}{2}$. Also the value of $c$ appears to be restricted to the range $\frac{1}{2}$ and ln(2). Here the ln(2) upper bound is based on Conjecture 1 of Chen and Rubin (1986) that for a Poisson random variable $X_\mu$ with mean $\mu$, a limiting case of the negative binomial, that $\mu - \text{Med}(X_\mu) < \ln(2)$. Hence $c$ is set to 0.5 in (5.22).

**Table 5.1** Values of $c$ satisfying $A([\mu]; [\mu] + c, \alpha) = p_\alpha$

| | $\alpha$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\mu$ | 0.01 | 0.1 | 0.5 | 1 | 2 | 5 | 10 | 50 | 100 |
| 0 | 0.57 | 0.57 | 0.58 | 0.59 | 0.60 | 0.62 | 0.64 | 0.67 | 0.67 |
| 1 | 0.53 | 0.53 | 0.54 | 0.55 | 0.56 | 0.58 | 0.60 | 0.63 | 0.65 |
| 2 | 0.52 | 0.52 | 0.53 | 0.53 | 0.54 | 0.56 | 0.58 | 0.62 | 0.63 |
| 5 | 0.51 | 0.51 | 0.52 | 0.52 | 0.52 | 0.54 | 0.55 | 0.60 | 0.61 |
| 10 | 0.51 | 0.51 | 0.51 | 0.51 | 0.52 | 0.52 | 0.54 | 0.58 | 0.60 |
| 20 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.52 | 0.52 | 0.56 | 0.58 |
| 40 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.52 | 0.54 | 0.56 |

To determine the bias and efficiency of the estimator in (5.22) a limited simulation study was undertaken. For each combination of $\alpha = 0.5, 1, 3$ & $5$, $\mu = .5, 1, 3, 5, 10, 15$ & $20$ and sample size $n = 20, 30, 50$ & $100$ a total of 10000 simulations were performed. From these 10000 simulations estimates of the bias and

efficiency were obtained, where the latter is defined as $(\mu + \frac{\mu^2}{\alpha})/(nVar(\hat{\mu}))$ . The results of this simulation study are given in Table 5.2.

From Table 5.2 it can be seen that as the mean increases the efficiency of the percentile based estimator is similar to the corresponding percentile estimator of the mean of a gamma random variable. (For the gamma distribution the asymptotic relative efficiencies of the $p_\alpha^{th}$ percentile estimator to the sample mean when $\alpha =$ .5, 1, 3 and 5 are .541, .582, .617 and .625 respectively.) This is to be expected as an $NBD(\mu, \alpha)$ approaches a gamma distribution when $\alpha$ is fixed and $\mu$ increases.

Also apparent from Table 5.2 is the fact that the estimator appears to retain some bias under certain parameter combinations. This does appear however to decrease in magnitude as the sample size increases.

## 5.4 Concluding remarks.

In this chapter a robust estimator for the mean for the NBD has been proposed and studied when the shape parameter is known. The more general case where the mean and shape are both unknown still needs consideration. A number of possibilities for the joint robust estimation of the two parameters are available. One would be the maximum likelihood estimates based on trimmed or censored samples, trimming or censoring proportions $p_1$ and $p_2$ from the extremes of the sample. These estimates need to be obtained using iterative methods, solving complicated expressions.

As an alternative to this one may exploit the close relationship between the NBD and the two parameter gamma distribution, as has been done by Guenther (1972) and Best and Gipps (1974) when approximating cumulative probabilities. In the first of the two papers the NBD is approximated with a gamma having the same first two moments whilst the second paper equates second and third moments. Using this idea it may be possible to consider a sample supposedly from a negative binomial distribution as that from a gamma sample with appropriate parameters, use one of the available robust methods for estimating the parameters of that gamma (e.g. Kimber (1983), Windham (1995)) and then obtain estimates of the NBD

**Table 5.2** Estimates of Bias and Efficiency of $\hat{\mu}$

| | | Bias | | | | Efficiency | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\mu$ | n = 20 | 30 | 50 | 100 | n = 20 | 30 | 50 | 100 |
| 0.5 | 0.5 | 0.04 | 0.02 | 0.01 | 0.00 | 0.58 | 0.60 | 0.60 | 0.68 |
| 0.5 | 1.0 | 0.08 | 0.06 | 0.04 | 0.04 | 0.54 | 0.57 | 0.58 | 0.57 |
| 0.5 | 3.0 | 0.16 | 0.12 | 0.09 | 0.05 | 0.55 | 0.56 | 0.55 | 0.57 |
| 0.5 | 5.0 | 0.21 | 0.18 | 0.15 | 0.05 | 0.55 | 0.54 | 0.53 | 0.57 |
| 0.5 | 10.0 | 0.32 | 0.42 | 0.32 | 0.12 | 0.54 | 0.53 | 0.50 | 0.54 |
| 0.5 | 15.0 | 0.51 | 0.68 | 0.51 | 0.24 | 0.53 | 0.51 | 0.51 | 0.54 |
| 0.5 | 20.0 | 0.62 | 0.80 | 0.77 | 0.35 | 0.53 | 0.53 | 0.50 | 0.55 |
| 1.0 | 0.5 | 0.02 | 0.01 | 0.00 | 0.00 | 0.71 | 0.70 | 0.78 | 0.86 |
| 1.0 | 1.0 | 0.04 | 0.03 | 0.02 | 0.02 | 0.65 | 0.65 | 0.67 | 0.65 |
| 1.0 | 3.0 | 0.09 | 0.05 | 0.04 | 0.02 | 0.61 | 0.60 | 0.63 | 0.61 |
| 1.0 | 5.0 | 0.11 | 0.06 | 0.05 | 0.02 | 0.60 | 0.60 | 0.60 | 0.60 |
| 1.0 | 10.0 | 0.13 | -0.02 | 0.11 | 0.06 | 0.61 | 0.63 | 0.60 | 0.60 |
| 1.0 | 15.0 | 0.18 | -0.08 | 0.12 | 0.11 | 0.59 | 0.61 | 0.59 | 0.59 |
| 1.0 | 20.0 | 0.28 | -0.11 | 0.19 | 0.16 | 0.60 | 0.61 | 0.58 | 0.59 |
| 3.0 | 0.5 | 0.00 | 0.00 | 0.00 | 0.00 | 0.89 | 0.93 | 0.94 | 1.00 |
| 3.0 | 1.0 | 0.00 | -0.01 | -0.01 | -0.01 | 0.72 | 0.71 | 0.73 | 0.74 |
| 3.0 | 3.0 | 0.02 | 0.00 | 0.00 | -0.01 | 0.69 | 0.68 | 0.65 | 0.63 |
| 3.0 | 5.0 | 0.03 | 0.03 | 0.02 | 0.00 | 0.68 | 0.65 | 0.65 | 0.63 |
| 3.0 | 10.0 | 0.09 | 0.07 | 0.03 | 0.00 | 0.62 | 0.63 | 0.64 | 0.63 |
| 3.0 | 15.0 | 0.08 | 0.13 | 0.05 | 0.04 | 0.63 | 0.63 | 0.64 | 0.64 |
| 3.0 | 20.0 | 0.15 | 0.20 | 0.00 | 0.00 | 0.63 | 0.62 | 0.64 | 0.65 |
| 5.0 | 0.5 | -0.01 | 0.00 | 0.00 | 0.00 | 1.02 | 0.96 | 0.99 | 1.01 |
| 5.0 | 1.0 | -0.03 | -0.03 | -0.03 | -0.03 | 0.73 | 0.76 | 0.75 | 0.76 |
| 5.0 | 3.0 | 0.00 | -0.02 | -0.02 | -0.02 | 0.68 | 0.67 | 0.65 | 0.67 |
| 5.0 | 5.0 | 0.01 | 0.00 | 0.00 | -0.02 | 0.67 | 0.65 | 0.67 | 0.66 |
| 5.0 | 10.0 | 0.09 | 0.00 | 0.00 | 0.00 | 0.65 | 0.66 | 0.64 | 0.64 |
| 5.0 | 15.0 | 0.19 | 0.00 | 0.02 | 0.03 | 0.65 | 0.66 | 0.66 | 0.64 |
| 5.0 | 20.0 | 0.22 | -0.01 | 0.00 | 0.03 | 0.64 | 0.65 | 0.64 | 0.63 |

parameters from the robust estimates of the gamma parameters. This idea could also be used when one of the parameters is known.

This estimation problem is not pursued further here. Instead, in the next two chapters robust estimation for more general models is developed, firstly for normal linear mixed models and then for generalized linear mixed models (GLMMs). Using the results for GLMMs in Chapter 7 approximate robust parameter estimates for a NBD can be obtained as a special case. There though, instead of approximating a NBD by a two parameter gamma as suggested above, the NBD will be approximated by a mixed lognormal-Poisson model. This approximation should be reasonable since the gamma and lognormal distributions have similar shaped distributions (provided the gamma parameter is greater than 1). Also, the variance of both the NBD and the mixed lognormal-Poisson distribution are quadratic in the mean.

# Chapter 6

# Robust estimation in linear mixed models.

To limit the influence of outlying results in the mixed-model analysis Fellner (1986) developed a robust procedure based on the mixed-model equations of Henderson (also referred to as the BLUP (Best Linear Unbiased Prediction) equations) combined with the restricted-maximum-likelihood (REML) procedure, both of which are detailed in Searle, Casella and McCulloch (1992). An attractive feature of the procedure, as mentioned by Fellner (1986), is that it gives robust analogs of the empirical Bayes estimates of the random effects. These can then be used to identify outlying results in the random effects and to prepare diagnostic plots.

As an alternative to this procedure of Fellner, Rocke (1983) and (1991) developed moment based procedures, in particular procedures based on Henderson's methods 1 and 3 for estimating variance components respectively. Again, details of Henderson's methods 1 and 3 can be found in Searle et. al. (1992). Unlike the method of Fellner (1986), the methods of Rocke do not use the BLUP equations to estimate (predict) the parameters in the regression model, rather they use the fixed effects model approach. That is, when estimating the regression parameters they are all considered fixed. This is unfortunate as the BLUP equations are generally considered preferable for estimating (predicting) the regression parameters to estimates based on the fixed effects model (see Robinson (1991)). Here the term

predict is used when referring to estimates of random effects as this seems to be a commonly used convention.

In this chapter we have two goals. First, we propose a modification to the procedure of Fellner (1986). The aim is to reduce the positive biasedness, significant in certain cases, of the variance component estimates using the Fellner method as it is currently formulated. Here the biasedness is brought about in the procedure by the use of inappropriate scaling factors for scaling predicted values of random parameters. Our modification suggests alternative scaling factors based on the variance-covariance matrices of the BLUP estimates of the regression parameters and residuals. In the process our standardization of random components for diagnostic purposes differs from that recommended by Fellner (1986).

The second purpose of this chapter is to give an alternative robust procedure to that of Rocke (1991) which estimates the variance components using method of moments. This alternative, as with the Fellner method but unlike the Rocke procedure, estimates (predicts) the regression parameters in the model using the BLUP procedure. But unlike the Fellner method which uses REML methods to estimate the variance components we use method of moments estimates, in particular Henderson's method 3.

In summary, in Section 6.1 we give the model in the general form and summarize the robust procedure as developed by Fellner (1986). In doing so we do not restrict, as Fellner did, the $\psi$ function associated with M estimation to the Huber-psi function but rather leave it general. In Section 6.2 the issue of the inappropriate scaling factor mentioned above is addressed and a modification to the Fellner method is proposed. Some comments are then offered as to when this modification will lead to significantly improved estimators. Section 6.3 briefly outlines the general procedures for estimating the variance components using Henderson's method 3 and then gives a robust version of this method as an alternative to the method given in Rocke (1991). The results of a limited simulation study for a balanced one-way random effects model are given in Section 6.4. In Section 6.5 we reconsider the example given in Fellner (1986). Some concluding remarks are offered in Section 6.6.

Throughout this and the following chapter the matrix notation of Searle et. al. (1992), Appendix M3 will be used. A summary of this notation is given in Appendix A of this thesis.

## 6.1 Linear mixed-model and Fellner's robust procedure.

Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}_1\,\mathbf{b}_1 + \cdots + \mathbf{U}_c\,\mathbf{b}_c + \mathbf{e}$$

where $\mathbf{y}$ is an $n \times 1$ response vector, $\mathbf{X}$ is an $n \times p$ known matrix of rank $p$, $\mathbf{U}_i$ $(i = 1, \cdots, c)$ is an $n \times q_i$ known matrix, $\boldsymbol{\alpha}$ is a $p$ component vector of unknown fixed effects and $\mathbf{b}_i$ $(i = 1, \cdots, c)$ is a $q_i$ component vector of unknown random effects.

We now assume that $E(\mathbf{b}_i) = \mathbf{0},\ i = 1, \cdots, c\ ,\quad E(\mathbf{e}) = \mathbf{0}\ ,\quad Cov(\mathbf{b}_i, \mathbf{e}) = \mathbf{0},\ i = 1, \cdots, c\ ,\quad Cov(\mathbf{b}_i, \mathbf{b}_j) = \mathbf{0},\ i \neq j\ ,$

$$V(\mathbf{b}_i) = \mathbf{D}_i = \sigma_i^2\,\mathbf{I}_{q_i}\ ,\quad (i = 1, \cdots, c)\ ,\ \text{and}$$
$$V(\mathbf{e}) = \mathbf{R} = \sigma^2\,\mathbf{I}_n\ .$$

Now let $\mathbf{b} = \{_c\ \mathbf{b}_i\}_{i=1}^c$, $\mathbf{U} = \{_r\ \mathbf{U}_i\}_{i=1}^c$ and $\mathbf{D} = \{_d\ \mathbf{D}_i\}_{i=1}^c$. Thus $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}\,\mathbf{b} + \mathbf{e}$, $V(\mathbf{b}) = \mathbf{D}$ and $V(\mathbf{y}) = \mathbf{U}\,\mathbf{D}\,\mathbf{U}' + \mathbf{R}$.

Also, let

$$\mathbf{C} = \begin{pmatrix} \mathbf{R}^{-1/2}\mathbf{X} & \mathbf{R}^{-1/2}\mathbf{U} \\ \mathbf{0} & \mathbf{D}^{-1/2} \end{pmatrix}\ , \tag{6.1}$$

$$\mathbf{w} = \begin{pmatrix} \mathbf{R}^{-1/2}\mathbf{y} \\ \mathbf{D}^{-1/2}\mathbf{0} \end{pmatrix}\ , \tag{6.2}$$

and $\boldsymbol{\theta}' = (\boldsymbol{\alpha}', \boldsymbol{\beta}')$ where $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_c')$ and $\boldsymbol{\beta}_i$ is the realized value of $\mathbf{b}_i$, $i = 1, \cdots, c$. Here the vector $\mathbf{0} = (\mathbf{0}_1', \cdots, \mathbf{0}_c')'$ where $\mathbf{0}_i$ is a vector of $q_i$ zeros is retained in (6.2) to facilitate subsequent development.

Then, for a suitably chosen $\psi_k$ function associated with M estimation (Huber (1981)) where $k$ a tuning constant, the Fellner (1986) robust parameter estimates

of $\alpha$, $\beta$, $\sigma_i^2$ $(i = 1, \cdots, c)$ and $\sigma^2$ are given as the solutions to the following three equations:

$$\mathbf{C}' \, \psi_k(\mathbf{w} - \mathbf{C}\,\hat{\boldsymbol{\theta}}) \;=\; \mathbf{0}\,; \tag{6.3}$$

$$\hat{\sigma_i^2} \;=\; \frac{\| \, \mathbf{D}_i^{1/2} \, \psi_k(\mathbf{D}_i^{-1/2} \, \hat{\boldsymbol{\beta}}_i) \, \|^2}{m_k \, (q_i - v_i)} \;, \quad i = 1, \cdots, c \,; \text{ and} \tag{6.4}$$

$$\hat{\sigma^2} \;=\; \frac{\|\mathbf{R}^{1/2} \psi_k(\mathbf{R}^{-1/2} \hat{\boldsymbol{\epsilon}})\|^2}{m_k \, (n - p - \sum_{i=1}^{c} (q_i - v_i))} \;, \tag{6.5}$$

where $\mathbf{C}$, $\mathbf{D}$, $\mathbf{R}$ (and $\mathbf{T}$ below) are evaluated at the most recent variance component estimates,

$$\hat{\boldsymbol{\epsilon}} \;=\; \mathbf{y} - \mathbf{X}\,\hat{\boldsymbol{\alpha}} - \mathbf{U}\,\hat{\boldsymbol{\beta}} \;, \tag{6.6}$$

$$(\psi_k(\mathbf{x}))' \;=\; (\psi_k(x_1), \psi_k(x_2), \cdots) \;,$$

$$m_k \;=\; \int \psi_k^2(x) \, d\Phi(x) \;, \text{ and} \tag{6.7}$$

$$v_i \;=\; Tr(\mathbf{T}_{ii}) \;,$$

where $\mathbf{T} = [\mathbf{I} + \mathbf{U}' \, \mathbf{R}^{-1} \, \mathbf{U} \, \mathbf{D} - \mathbf{U}' \, \mathbf{R}^{-1} \, \mathbf{X} \, (\mathbf{X}' \, \mathbf{R}^{-1} \, \mathbf{X})^{-1} \, \mathbf{X}' \, \mathbf{R}^{-1} \, \mathbf{U} \, \mathbf{D}]^{-1}$ and which is partitioned conformally with $\mathbf{D}$ as $\{_m \, \mathbf{T}_{ij}\}_{i,j=1}^{c}$. The value $m_k$ is chosen to adjust for bias under normality.

Alternative ways exist to write (6.3) which enable the equations to be solved. One is to introduce data-determined weights as in Rocke (1991). Another, used by Fellner, is to write (6.3) in terms of pseudo-observations giving

$$\begin{pmatrix} \mathbf{X}' \, \mathbf{R}^{-1} \, \mathbf{X} & \mathbf{X}' \, \mathbf{R}^{-1} \, \mathbf{U} \\ \mathbf{U}' \, \mathbf{R}^{-1} \, \mathbf{X} & \mathbf{D}^{-1} + \mathbf{U}' \, \mathbf{R}^{-1} \, \mathbf{U} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} \;=\; \begin{pmatrix} \mathbf{X}' \, \mathbf{R}^{-1} \, \tilde{\mathbf{y}} \\ \mathbf{U}' \, \mathbf{R}^{-1} \, \tilde{\mathbf{y}} + \mathbf{D}^{-1} \, \tilde{\mathbf{0}} \end{pmatrix} \tag{6.8}$$

with

$$\tilde{\mathbf{y}} \;=\; \mathbf{X}\,\hat{\boldsymbol{\alpha}} + \mathbf{U}\,\hat{\boldsymbol{\beta}} + \mathbf{R}^{1/2} \, \psi_k(\mathbf{R}^{-1/2} \, \hat{\boldsymbol{\epsilon}}) \quad \text{and} \tag{6.9}$$

$$\tilde{\mathbf{0}} \;=\; \hat{\boldsymbol{\beta}} - \mathbf{D}^{1/2} \, \psi_k(\mathbf{D}^{-1/2} \, \hat{\boldsymbol{\beta}}) \;. \tag{6.10}$$

To obtain estimates of the parameters we need to iterate through equations (6.8), (6.4) and (6.5), replacing $\mathbf{R}$ and $\mathbf{D}$ at each iteration by their most recent estimates, until we have the required convergence.

For $\psi_k(x)$ Fellner chose the Huber-psi function

$$\psi_k(x) \;=\; \max(-k,\ \min(x,k)) \text{ for some suitably chosen } k\ ;$$

for which Fellner suggests $k = 2$ as a compromise between efficiency and robustness. The method above could however be used for different choices of $\psi_k(x)$. One alternative is the Tukey biweight given by

$$\psi_k(x) = x\left(\frac{k^2 - x^2}{k^2}\right)^2 I(\mid x \mid \leq k)\ . \tag{6.11}$$

This choice of $\psi_k(x)$ with $k = 9$ was studied by Rocke (1991), along with the Huber-psi function with $k = 2$, for robustly estimating the variance components in the mixed model using a different estimating procedure to that of Fellner. Based on the results of Rocke (1991) the choice of Tukey biweight for $\psi$ would appear to have merit, particularly when the data contains very extreme results.

## 6.2 Modification to the Fellner procedure.

We now propose a modification to the procedure of Fellner (1986) in order to improve its robustness by further reducing the influence of outlying results on the variance component estimates.

Consider the REML estimates of $\sigma_i^2$ satisfying

$$\hat{\sigma_i^2} \;=\; \frac{||\hat{\boldsymbol{\beta}}_i||^2}{q_i - v_i} \;=\; \frac{||\hat{\boldsymbol{\beta}}_i - \mathbf{0}_i||^2}{q_i - v_i}\ , \quad i = 1, \cdots, c$$

where $q_i$ and $v_i$ are as given in Section 6.1. That is, equations (6.3) and (6.4) with $\psi_k(x) = x$. To robustify this Fellner proposed moving the mean of $\boldsymbol{\beta}$, that is $\mathbf{0}$, closer to its predicted value $\hat{\boldsymbol{\beta}}$ when the two are "far" apart. Here "far" is based on a standardized value of $\hat{\boldsymbol{\beta}}$, this being $\mathbf{D}^{-1/2}\,\hat{\boldsymbol{\beta}}$. What this fails to take into account is that $\hat{\boldsymbol{\beta}}$, the BLUP estimate of $\boldsymbol{\beta}$, is a shrinkage estimator (see Robinson (1991)). Elements of $\hat{\boldsymbol{\beta}}$ will therefore have smaller variance than the corresponding elements of $\mathbf{b}$. Likewise, elements of $\hat{\boldsymbol{\epsilon}}$ will have smaller variance than the corresponding elements of $\mathbf{e}$. Fellner does make passing mention to a related matter (see Fellner (1986), page 55) though only in the context of modifying the unbiasing constant.

We therefore propose replacing (6.4) and (6.5) by

$$\hat{\sigma_i^2} = \frac{\| \mathbf{K}_i^{1/2} \psi_k(\mathbf{K}_i^{-1/2} \hat{\boldsymbol{\beta}}_i) \|^2}{m_k (q_i - v_i)}, \quad i = 1, \cdots, c \text{ ; and} \tag{6.12}$$

$$\hat{\sigma^2} = \frac{\|\mathbf{L}^{1/2}\psi_k(\mathbf{L}^{-1/2}\hat{\boldsymbol{\epsilon}})\|^2}{m_k \ (n - p - \sum_{i=1}^{c}(q_i - v_i))}, \tag{6.13}$$

with $q_i$ and $v_i$ defined earlier and for $\mathbf{K}_i$ $(i = 1, \cdots, c)$ and $\mathbf{L}$ yet to be defined.

At first glance an appropriate choice for $\mathbf{K}_i$ might be $Var(\hat{\boldsymbol{\beta}}_i)$ $(i = 1, \cdots, c)$ evaluated at the current variance component estimates and $\mathbf{L} = Var(\hat{\boldsymbol{\epsilon}})$ evaluated at the current estimates. This would take account of the correlation between the estimates. The problem here is that these matrices in general are singular. Instead we take $\mathbf{K}_i$ be a diagonal matrix with diagonal elements given by the diagonal elements of $Var(\hat{\boldsymbol{\beta}}_i)$ $i = 1, \cdots, c$ evaluated at the estimates of the variance components, thus treating the correlations as zero. Similarly, we take $\mathbf{L}$ equal the diagonal matrix corresponding to $Var(\hat{\boldsymbol{\epsilon}})$. That is, letting $\hat{\beta}_{ij}$ $(j = 1, \cdots, q_i)$ equal the $j^{th}$ element of $\hat{\boldsymbol{\beta}}_i$ $(i = 1, \cdots, c)$ and $\hat{\epsilon}_j$ $(j = 1, \cdots, n)$ the $j^{th}$ element of $\hat{\boldsymbol{\epsilon}}$

$$\begin{aligned} \mathbf{K}_i &= \{_d\, Var(\hat{\beta}_{ij})\}_{j=1}^{q_i} \quad (i = 1, \cdots, c), \quad \text{and} \\ \mathbf{L} &= \{_d\, Var(\hat{\epsilon}_j)\}_{j=1}^{n}. \end{aligned}$$

It remains to determine $Var(\hat{\boldsymbol{\beta}}_i)$ $(i = 1, \cdots, c)$ and $Var(\hat{\boldsymbol{\epsilon}})$. We approximate these by using the expression for the standard BLUP estimates given by (6.8) with $k = \infty$ when $\psi_k$ is either the Huber-psi or the Tukey biweight. This approximation should be reasonable provided $k$ is not chosen too small (e.g. $k > 1.5$) as the asymptotic relative efficiency in terms of variance for M estimators is large (see for example Staudte and Sheather (1990), Table 4.5).

The standard BLUP estimates are then

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{U} \\ \mathbf{U}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{D}^{-1} + \mathbf{U}'\mathbf{R}^{-1}\mathbf{U} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1} \\ \mathbf{U}'\mathbf{R}^{-1} \end{pmatrix} \mathbf{y}.$$

With $\boldsymbol{\Sigma} = \mathbf{R} + \mathbf{U}\,\mathbf{D}\,\mathbf{U}'$ (the unconditional variance of $\mathbf{y}$), the standard BLUP estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and the related estimate of $\boldsymbol{\epsilon}$ satisfy

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\,\boldsymbol{\Sigma}^{-1}\,\mathbf{X})^{-1}\mathbf{X}'\,\boldsymbol{\Sigma}^{-1}\,\mathbf{y},$$

$$\hat{\boldsymbol{\beta}} = \mathbf{D}\,\mathbf{U}'\,\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\,\hat{\boldsymbol{\alpha}}) \text{ and}$$

$$\hat{\boldsymbol{\varepsilon}} = Y - \mathbf{X}\,\hat{\boldsymbol{\alpha}} - \mathbf{U}\,\hat{\boldsymbol{\beta}}.$$

From these we obtain,

$$Var(\hat{\boldsymbol{\alpha}}) = (\mathbf{X}'\,\boldsymbol{\Sigma}^{-1}\,\mathbf{X})^{-1}\,, \tag{6.14}$$

$$Var(\hat{\boldsymbol{\beta}}) = \mathbf{D}\,\mathbf{U}'\,\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \mathbf{X}\,(\mathbf{X}'\,\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\right)\boldsymbol{\Sigma}^{-1}\mathbf{U}\,\mathbf{D} \text{ and} \tag{6.15}$$

$$Var(\hat{\boldsymbol{\varepsilon}}) = \left(\mathbf{I} - \mathbf{U}\,\mathbf{D}\,\mathbf{U}'\,\boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{\Sigma} - \mathbf{X}\,(\mathbf{X}'\,\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\right)\left(\mathbf{I} - \boldsymbol{\Sigma}^{-1}\,\mathbf{U}\,\mathbf{D}\,\mathbf{U}'\right). \tag{6.16}$$

To invert $\boldsymbol{\Sigma}$, an $n \times n$ matrix, it is sometimes easier to use the form

$$\boldsymbol{\Sigma}^{-1} = \left(\mathbf{I} - \mathbf{R}^{-1}\mathbf{U}\,\mathbf{D}(\mathbf{I} + \mathbf{U}'\mathbf{R}^{-1}\mathbf{U}\,\mathbf{D})^{-1}\mathbf{U}'\right)\mathbf{R}^{-1}\,.$$

To obtain the modified Fellner estimates of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\sigma_i^2$ $(i = 1, \cdots, c)$ and $\sigma^2$ we iterate through the three equations (6.8), (6.12) and (6.13), using (6.15) and (6.16) to obtain $\mathbf{K}_i$ $(i = 1, \cdots, c)$ and $\mathbf{L}$ and substituting for $\mathbf{R}$ and $\mathbf{D}$ at each stage of the iterative process their most recent estimates. Iteration is continued until the required convergence criteria are met.

Using the above modification to the Fellner procedure we can expect the variance component estimates to be at least as good as those of Fellner. Where the modified method will perform better is in estimating variance components for which there is only little information in the sample to predict the random parameters associated with that variance component, in particular the random error variance. This follows as the extent of the shrinkage of the BLUP estimates depends on the amount of information in the sample pertaining to the particular parameter, with the shrinkage increasing as the amount of information decreases. These comments will be supported by the simulations in Section 6.4 .

## 6.3 Robust variance components estimation using method of moments.

In this section we propose a robust estimation procedure for the linear mixed model analysis based on the BLUP method for estimating (predicting) the regression parameters combined with method of moments based estimates for the variance components. Method of moments estimates for the variance components are worthy of consideration as they have been shown to perform reasonably well in simulation studies (e.g. Swallow and Monahan (1984)) when compared with more contemporary methods such as REML and MINQUE, with the latter described in Rao and Kleffe (1988). Our robust procedure is based on Henderson's method 3.

### 6.3.1 Standard Henderson method 3 estimates.

The basis of the Henderson method 3 procedure for estimating variance components associated with a linear mixed model is to equate changes in residual sums of squares after omitting terms from the model with the expectations of these changes. Thus, by a suitable choice of deletions a set of linear equations, not necessarily unique, can be obtained in the variance components. We now provide a brief outline of one possible solution. A more extensive coverage of the method can be found in Searle et. al. (1992), Chapter 5.

For convenience of notation replace $\mathbf{X}$ and $\boldsymbol{\alpha}$ by $\mathbf{U}_0$ and $\mathbf{b}_0$ respectively to give

$$\mathbf{y} = \sum_{i=0}^{c} \mathbf{U}_i \, \mathbf{b}_i + \mathbf{e} \, .$$

Also introduce the following notation

$$\begin{aligned}
\mathbf{U}_{\leq i} &= \{_r \, \mathbf{U}_j\}_{j=0}^{i} \, , \; i = 0, 1, \cdots, c \, , \\
\mathbf{b}_{\leq i} &= \{_c \, \mathbf{b}_j\}_{j=0}^{i} \, , \; i = 0, 1, \cdots, c \, , \\
\mathbf{Q}_i &= \mathbf{U}_{\leq i} \left( \mathbf{U}'_{\leq i} \, \mathbf{U}_{\leq i} \right)^{-} \mathbf{U}'_{\leq i}, \; i = 0, \cdots, c \quad \text{and} \\
\mathbf{H}_i &= \mathbf{I} - \mathbf{Q}_i \, .
\end{aligned}$$

Here $\mathbf{A}^{-}$ denotes any g-inverse of $\mathbf{A}$ satisfying $\mathbf{A} \, \mathbf{A}^{-} \, \mathbf{A} = \mathbf{A}$.

Without loss of generality we now assume that $\rho(\mathbf{U}_{\leq i}) < \rho(\mathbf{U}_{\leq i'})$ for all $i < i'$ where $\rho(\mathbf{A})$ is the rank of $\mathbf{A}$. Sequentially fitting the models $\mathbf{y} = \mathbf{U}_{\leq i}\,\mathbf{b}_{\leq i} + \mathbf{e}$ $(i = 0, \cdots, c)$ we obtain the following Henderson method 3 estimates for the variance components:

$$\begin{aligned} \hat{\sigma^2} &= \frac{SSE}{n - \rho(\mathbf{U}_{\leq c})}\,, \\ \hat{\boldsymbol{\sigma}}_v^2 &= \mathbf{P}^{-1}\,(\mathbf{d} - \hat{\sigma^2}\,\mathbf{f}), \end{aligned}$$

where $SSE = \mathbf{y}'\,\mathbf{H}_c\,\mathbf{y}$, $\hat{\boldsymbol{\sigma}}_v^2 = \{_c\ \hat{\sigma}_i^2\}_{i=1}^c$,

$$\begin{aligned} \mathbf{P} = \{_m\ p_{ij}\}_{i,j=1}^c \text{ with } p_{ij} &= tr(\mathbf{U}_j'\,\mathbf{H}_{i-1}\,\mathbf{U}_j)\ ,\ 1 \leq i \leq j \leq c\ ; \\ &= 0 \text{ otherwise}\ ; \end{aligned}$$

$$\mathbf{f} = \{_c\ f_i\}_{i=1}^c \text{ with } f_i = \rho(\mathbf{U}_{\leq c}) - \rho(\mathbf{U}_{\leq i-1})\ ,\ i = 1, \cdots, c\ ;$$

$$\mathbf{d} = \{_c\ d_i\}_{i=1}^c \text{ with } d_i = \mathbf{y}'\,(\mathbf{H}_{i-1} - \mathbf{H}_c)\mathbf{y}\ ,i = 1, \cdots, c\ .$$

To assist with the computations a useful recurrence relationship is

$$\begin{aligned} \mathbf{H}_0 &= \mathbf{I} - \mathbf{X}\,(\mathbf{X}'\,\mathbf{X})^{-1}\,\mathbf{X}' \\ \mathbf{H}_i &= \mathbf{H}_{i-1} - \mathbf{M}_i\,(\mathbf{M}_i'\,\mathbf{M}_i)^-\,\mathbf{M}_i' \quad \text{for} \quad i = 1, \cdots, c \\ &\quad \text{where} \quad \mathbf{M}_i = \mathbf{H}_{i-1}\,\mathbf{U}_i\ . \end{aligned}$$

This latter equation follows from equation (23) on page 450 of Searle et. al. (1992).

### 6.3.2 Robust Henderson method 3 estimates.

We now propose a more robust procedure to that outlined in Section 6.3.1. Consider the general model in Section 6.1 and assume all random components are Normal random variables. Let

$$\begin{aligned} \tilde{\mathbf{b}} &= \mathbf{D}^{1/2}\,\psi_k(\mathbf{D}^{-1/2}\,\mathbf{b})\,, \\ \tilde{\mathbf{0}}^{(b)} &= \mathbf{b} - \tilde{\mathbf{b}}\,, \text{ and} \\ \tilde{\mathbf{e}} &= \mathbf{R}^{1/2}\,\psi_k(\mathbf{R}^{-1/2}\,\mathbf{e})\,. \end{aligned}$$

Considering the (unobservable) random variable

$$\mathbf{y}^* \;=\; \mathbf{X}\,\boldsymbol{\alpha} + \mathbf{U}\,\tilde{\mathbf{b}} + \tilde{\mathbf{e}}$$

we have $\mathcal{E}(\tilde{\mathbf{b}}) = \mathbf{0},\ Var(\tilde{\mathbf{b}}) = m_k\,\mathbf{D},\ \mathcal{E}(\tilde{\mathbf{e}}) = \mathbf{0},\ Var(\tilde{\mathbf{e}}) = m_k\,\mathbf{R}$ and $Cov(\tilde{\mathbf{b}}, \tilde{\mathbf{e}}) = \mathbf{0}$, where $m_k$ is given in (6.7).

If $\mathbf{y}^*$ was observable robust estimates of $\sigma_i^2$ , $i = 1, \cdots, c$ and $\sigma^2$ could then be obtained using Henderson's method 3 above. Here the estimates would be robust as the elements of $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{e}}$ are bounded.

In practice, since $\mathbf{y}^* = \mathbf{X}\,\boldsymbol{\alpha} + \mathbf{U}\,\mathbf{b} + \tilde{\mathbf{e}} - \mathbf{U}\,\tilde{\mathbf{0}}^{(b)}$ is unobservable, we replace it by $\tilde{\mathbf{y}}^*$ given by

$$\begin{aligned} \tilde{\mathbf{y}}^* &= \mathbf{X}\,\hat{\boldsymbol{\alpha}} + \mathbf{U}\,\hat{\boldsymbol{\beta}} + \mathbf{R}^{1/2}\,\psi_k(\mathbf{R}^{-1/2}\,\hat{\boldsymbol{\epsilon}}) - \mathbf{U}\,\tilde{\mathbf{0}} \\ &= \tilde{\mathbf{y}} - \mathbf{U}\,\tilde{\mathbf{0}}\,, \end{aligned}$$

where $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{0}}$ are as defined in (6.9) and (6.10) respectively.

Robust Henderson's method 3 estimates of the variance components are then obtained from

$$\hat{\sigma^2} \;=\; \frac{SSE^*}{m_k(n - \rho(\mathbf{U}_{\leq c}))}\,, \tag{6.17}$$

$$\hat{\boldsymbol{\sigma}_v^2} \;=\; m_k^{-1}\mathbf{P}^{-1}\,(\mathbf{d}^* - m_k\,\hat{\sigma^2}\,\mathbf{f})\;\;, \tag{6.18}$$

where $SSE^*$ and $\mathbf{d}^*$ are obtained as for $SSE$ and $\mathbf{d}$ but with $\mathbf{y}$ replaced by $\tilde{\mathbf{y}}^*$.

Unlike the method proposed by Fellner (1986) which always gives non-negative estimates of the variance components, the method above based on Henderson's method 3 can give negative estimates. A common procedure used in practice for dealing with negative estimates, which we recommend here, is to set these values to zero and continue with the iterative procedure. However, if any variance component is zero then $\mathbf{D}^{-1}$ does not exist. In such cases replace (6.8) by a robust version based on the Harville (1977) form of the Henderson mixed model equations, that being

$$\begin{pmatrix} \mathbf{X}'\,\mathbf{R}^{-1}\,\mathbf{X} & \mathbf{X}'\,\mathbf{R}^{-1}\,\mathbf{U}\,\mathbf{D} \\ \mathbf{U}'\,\mathbf{R}^{-1}\,\mathbf{X} & \mathbf{I} + \mathbf{U}'\,\mathbf{R}^{-1}\,\mathbf{U}\,\mathbf{D} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\nu}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\,\mathbf{R}^{-1}\,\tilde{\mathbf{y}} \\ \mathbf{U}'\,\mathbf{R}^{-1}\,\tilde{\mathbf{y}} + \mathbf{D}^{-}\,\tilde{\mathbf{0}} \end{pmatrix} \tag{6.19}$$

where $\boldsymbol{\beta} = \mathbf{D}\,\boldsymbol{\nu}$ and $\mathbf{D}^-$ is a diagonal matrix with elements the inverse of the corresponding element of $\mathbf{D}$ if that element is positive and zero otherwise. We then let $\hat{\boldsymbol{\beta}} = \mathbf{D}\,\hat{\boldsymbol{\nu}}$ and replace (6.10) by

$$\tilde{\mathbf{0}} \;=\; \hat{\boldsymbol{\beta}} - \mathbf{D}^{1/2}\,\psi_k((\mathbf{D}^-)^{1/2}\,\hat{\boldsymbol{\beta}})\;. \tag{6.20}$$

Estimates of the regression parameters and the variance components are then obtained by iterating through equations (6.19), (6.17) and (6.18) until convergence.

In the above formulation of this robust moment method it was assumed that all random components were Normally distributed, thus ensuring that the constant $m_k$ used throughout was correct. However, $m_k$ should be a reasonable approximation for the bias adjustment provided deviations from Normality are not too significant.

## 6.4 Simulations.

For our limited simulation studies we consider the balanced one-way random effects model given by

$$Y_{ij} \;=\; \alpha + b_i + e_{ij} \;\; (i = 1, \cdots, l \;\; ; \;\; j = 1, \cdots, r) \tag{6.21}$$

where the $b_i$ are *i.i.d.* with means zero and variances $\sigma_L^2$, independent of the $e_{ij}$ which are *i.i.d.* with means zero and variances $\sigma^2$.

For this model we have from (6.15) and (6.16), assuming $\sigma_L^2$ and $\sigma^2$ are known, the following approximations:

$$var(\hat{\beta}_i) \;=\; \left(\frac{(l-1)}{l}\,\frac{r\sigma_L^2}{(\sigma^2 + r\sigma_L^2)}\right)\sigma_L^2\;, \text{ and} \tag{6.22}$$

$$var(\hat{\epsilon}_{ij}) \;=\; \left(\frac{(r-1)}{r} + \frac{1}{r}\,\frac{(l-1)}{l}\,\frac{\sigma^2}{(\sigma^2 + r\sigma_L^2)}\right)\sigma^2\;. \tag{6.23}$$

The square root of these values are the scaling factors for the modified Fellner method.

To link in with the simulations of Rocke (1991), we refer to the $b_i$ as laboratory effects and consider three distributions for the laboratory and error distributions, these being standard normal (N), a mixture of 90% $N(0,1)$ with 10% $N(0,3^2)$ (LT)

and a mixture of 95% $N(0,1)$ with 5% $N(0,10^2)$ (VLT). For choices of $l$ and $r$ we let both equal five, both equal ten and finally $l = 10$ with $r = 2$. In all cases we set $\alpha = 0$ but estimate it from the data.

Three sets of robust estimators for $\alpha$, $\sigma^2$ and $\sigma_L^2$ are considered. These are

| | | |
|---|---|---|
| Robust method 1 (RM1) | - | Fellner method as outlined in Section 6.1 |
| Robust method 2 (RM2) | - | Modified Fellner method as outlined in Section 6.2 |
| Robust method 3 (RM3) | - | Moment method as outlined in Section 6.3 |

In all cases $\psi_k$ is taken as the Huber-psi function with tuning constant $k = 2$.

For each combination of $l$, $r$, laboratory distribution and error distribution considered a total of 500 data sets were simulated and the parameters of the model estimated using each of the three procedures. Initial values for iteration were taken as the non-robust estimates obtained using BLUP equations and Henderson's method 3 but with negative estimates of $\sigma_L^2$ set to 0.001 for RM1 and RM2 and 0 for RM3. Estimators were considered to have converged if the maximum absolute difference between subsequent estimates of $\alpha$, $\sigma_L^2$, $\sigma^2$ and $\beta_i$, $i = 1, \cdots, l$ was less than 0.001. Note here that this convergence criterion is more restrictive than that of Fellner (1986) wherein tests for convergence are based on subsequent estimates of the variance components alone.

In the first set of simulations the combinations of laboratory and error distributions considered, together with the choice $l$ and $r$ (both five or both ten), correspond to those of Rocke (1991), thus allowing a direct comparison. The results of the simulation are in Table 6.1 wherein are given the mean and standard deviation (the latter in brackets) of $n_{it}$ (the number of iterations to convergence), $\hat{\sigma^2}$ and $\hat{\sigma_L^2}$ based on the 500 data sets generated.

Comparing the results in Table 6.1 for $\hat{\sigma_L^2}$ with the results for the corresponding estimators in Rocke (1991) (that is $\hat{V}^H$ with Rocke's Robust methods 1 and 2) tends to indicate that when outliers are present in the laboratories the three estimators considered here perform as well as, if not better than the estimators in Rocke (1991) with respect to giving estimates for $\sigma_L^2$ closer to 1 on average and with respect to

having smaller variance. When there are no outliers in the laboratories none of the three estimators considered in Table 6.1 and the two Rocke estimators mentioned above appears uniformily better than the others for estimating $\sigma_L^2$.

If we now consider the joint estimation of $\sigma^2$ and $\sigma_L^2$ for the simulations reported in Table 6.1 it would appear that based on the mean and variance of these estimates, RM2 is a better estimator than either RM1 or RM3, whereas the latter two are effectively equivalent. This claim on behalf of RM2 follows because on those occasions when RM2 is not best with respect to a specific variance component it easily compensates with respect the other variance component. The main disadvantage of the RM2 over the other two methods based on the results in Table 6.1 is that it requires more iterations to converge on average than either RM1 or RM3, with RM3 the preferred estimator based on this criteria.

As was mentioned in Section 6.2, where we would expect a significant improvement in the modified Fellner method (RM2) over the the Fellner method (RM1) is when there is little sample information to predict the corresponding random components associated with the particular variance component. We can see this to some extent in the results for $\sigma^2$ in Table 6.1, particularly for $r = 5$, but consider again the one-way model given in (6.21). For this model we have from (6.22) that the shrinkage of the BLUP estimates $\hat{\beta}_i$ $(i = 1, \cdots, l)$ increases as $r$ decreases. We would therefore expect RM2 to be a better estimator of $\sigma_L^2$ than RM1 when $r$ is small. Therefore, for the second set of simulations we have considered all nine combinations of laboratory and error distribution together with $l = 10$ and $r = 2$. The results are given in Table 6.2 . From this table we can see that, as expected, RM2 is the preferred estimator to either RM1 and RM3 based on mean and variance of the variance components estimates whereas once again the latter two estimators are essentially equivalent. In fact RM2 is relatively much better in certain cases. The one anomaly is for VLT Error distribution and N Lab distribution in which case RM2 gives a larger (marginally) average estimate for $\sigma_L^2$ than RM1 and RM3. This is concluded to have been brought about by a large number of samples in the actual 500 data sets generated having extreme results in the Error distribution. RM2 has

**Table 6.1:** Robust variance component estimates - Balanced one-way model ($l = r$).

| | | | RM1 | | | RM2 | | | RM3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $l$ | Error | Labs | $n_{it}$ | $\hat{\sigma^2}$ | $\hat{\sigma_L^2}$ | $n_{it}$ | $\hat{\sigma^2}$ | $\hat{\sigma_L^2}$ | $n_{it}$ | $\hat{\sigma^2}$ | $\hat{\sigma_L^2}$ |
| 5 | N | N | 3.80 (3.39) | 1.03 (0.34) | 1.18 (0.93) | 4.06 (2.95) | 1.01 (0.34) | 1.18 (0.93) | 3.62 (1.18) | 1.03 (0.34) | 1.17 (0.93) |
| 10 | N | N | 4.24 (1.39) | 1.03 (0.16) | 1.04 (0.54) | 4.90 (2.37) | 1.02 (0.16) | 1.00 (0.53) | 4.04 (1.21) | 1.03 (0.16) | 1.03 (0.54) |
| 5 | LT | N | 7.11 (5.79) | 1.47 (0.63) | 1.02 (0.95) | 7.35 (5.89) | 1.41 (0.58) | 1.03 (0.95) | 5.40 (2.59) | 1.48 (0.64) | 1.00 (0.95) |
| 10 | LT | N | 6.49 (1.90) | 1.35 (0.26) | 1.08 (0.59) | 7.03 (2.53) | 1.33 (0.26) | 1.05 (0.58) | 6.03 (1.58) | 1.35 (0.26) | 1.07 (0.59) |
| 5 | VLT | N | 10.4 (8.59) | 1.56 (1.04) | 1.12 (0.98) | 10.4 (7.81) | 1.46 (0.86) | 1.13 (0.97) | 7.89 (4.82) | 1.58 (1.08) | 1.09 (0.98) |
| 10 | VLT | N | 8.67 (2.53) | 1.31 (0.24) | 1.05 (0.59) | 9.61 (3.91) | 1.30 (0.24) | 1.00 (0.59) | 7.89 (2.09) | 1.31 (0.24) | 1.04 (0.59) |
| 5 | N | LT | 4.07 (3.40) | 1.06 (0.34) | 1.98 (2.56) | 4.34 (3.22) | 1.04 (0.34) | 1.98 (2.56) | 3.77 (1.46) | 1.06 (0.34) | 1.96 (2.56) |
| 10 | N | LT | 5.31 (3.10) | 1.02 (0.15) | 1.56 (1.17) | 6.19 (4.41) | 1.02 (0.15) | 1.47 (1.11) | 5.09 (2.87) | 1.02 (0.15) | 1.55 (1.17) |
| 5 | LT | LT | 6.99 (5.40) | 1.46 (0.62) | 1.96 (2.70) | 6.90 (4.77) | 1.41 (0.58) | 1.97 (2.70) | 5.46 (2.56) | 1.47 (0.63) | 1.94 (2.70) |
| 10 | LT | LT | 7.06 (2.72) | 1.34 (0.24) | 1.73 (1.51) | 7.71 (3.98) | 1.33 (0.24) | 1.62 (1.41) | 6.71 (2.51) | 1.34 (0.24) | 1.71 (1.50) |
| 5 | N | VLT | 3.79 (2.84) | 1.09 (0.37) | 8.01 (26.7) | 3.97 (2.27) | 1.07 (0.37) | 8.01 (26.7) | 3.62 (1.31) | 1.09 (0.37) | 7.99 (26.7) |
| 10 | N | VLT | 7.14 (4.89) | 1.02 (0.16) | 2.04 (3.63) | 8.04 (6.89) | 1.02 (0.16) | 1.82 (3.21) | 6.81 (4.69) | 1.02 (0.16) | 2.03 (3.63) |
| 5 | VLT | VLT | 9.25 (7.78) | 1.49 (0.97) | 5.19 (14.2) | 9.29 (7.13) | 1.41 (0.86) | 5.20 (14.2) | 7.45 (4.72) | 1.50 (0.99) | 5.16 (14.2) |
| 10 | VLT | VLT | 11.4 (10.2) | 1.31 (0.26) | 2.08 (3.88) | 12.0 (7.26) | 1.30 (0.26) | 1.72 (2.85) | 10.7 (10.2) | 1.31 (0.26) | 2.06 (3.88) |

*Note:* Results for robust estimation of the parameters for the balanced one-way random effects model with $l$ groups and $r = l$ results from each group, based on 500 Monte Carlo replications. Here N denotes standard normal, LT denotes 90% N with 10% $N(0, 3^2)$ and VLT denotes 95% N with 5% $N(0, 10^2)$. Each pair of entries under $n_{it}$, $\hat{\sigma^2}$, and $\hat{\sigma_L^2}$ is the mean (standard deviation) over the 500 replications.

**Table 6.2:** Robust variance component estimates - Balanced one-way model ($l = 10, \quad r = 2$).

| | | RM1 | | | RM2 | | | RM3 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Error | Labs | $n_{it}$ | $\hat{\sigma^2}$ | $\hat{\sigma_L^2}$ | $n_{it}$ | $\hat{\sigma^2}$ | $\hat{\sigma_L^2}$ | $n_{it}$ | $\hat{\sigma^2}$ | $\hat{\sigma_L^2}$ |
| N | N | 4.95 (7.85) | 1.05 (0.45) | 1.04 (0.77) | 9.04 (9.96) | 0.99 (0.41) | 0.99 (0.75) | 3.61 (1.29) | 1.06 (0.47) | 1.00 (0.77) |
| N | LT | 5.23 (6.43) | 1.04 (0.44) | 1.81 (1.51) | 9.54 (9.51) | 1.01 (0.43) | 1.60 (1.34) | 4.45 (2.56) | 1.05 (0.46) | 1.76 (1.50) |
| N | VLT | 7.35 (9.20) | 1.08 (0.47) | 2.24 (3.22) | 14.2 (32.3) | 1.06 (0.48) | 1.81 (2.37) | 6.43 (6.84) | 1.09 (0.49) | 2.18 (3.20) |
| LT | N | 8.44 (10.1) | 1.57 (0.84) | 1.15 (0.95) | 12.2 (11.1) | 1.38 (0.67) | 1.15 (0.90) | 4.86 (2.60) | 1.62 (0.99) | 1.10 (0.94) |
| LT | LT | 8.35 (9.84) | 1.68 (0.95) | 1.60 (1.45) | 13.8 (16.9) | 1.51 (0.81) | 1.45 (1.24) | 5.13 (2.95) | 1.70 (0.98) | 1.53 (1.43) |
| LT | VLT | 10.6 (10.6) | 1.62 (0.94) | 2.66 (5.68) | 16.1 (16.8) | 1.48 (0.83) | 2.16 (4.71) | 7.21 (5.23) | 1.65 (0.98) | 2.58 (5.66) |
| VLT | N | 13.5 (14.2) | 2.01 (2.88) | 1.15 (0.97) | 15.0 (11.8) | 1.54 (1.53) | 1.19 (1.04) | 7.31 (5.28) | 2.15 (3.29) | 1.07 (0.95) |
| VLT | LT | 13.5 (13.5) | 1.98 (2.25) | 1.86 (2.25) | 15.4 (12.1) | 1.63 (1.44) | 1.73 (2.07) | 7.59 (4.95) | 2.03 (2.28) | 1.78 (2.19) |
| VLT | VLT | 14.8 (13.8) | 2.03 (1.95) | 3.46 (6.88) | 18.9 (20.7) | 1.71 (1.39) | 2.51 (4.13) | 10.6 (10.8) | 2.10 (2.26) | 3.34 (6.80) |

*Note:* As for *Note* Table 6.1 except that here $l = 10, \quad r = 2$ for all simulations.

been better at reducing the influence of these outliers, as reflected by the estimate of $\sigma^2$ but as a consequence has marginally increased the average estimate of $\sigma_L^2$. Also apparent from Table 6.2 as for the first set of simulations, is that RM2 requires more iterations to converge on average than RM1 which in turn requires more than RM3.

With respect to the estimation of $\alpha$ all three estimators are essentially equivalent. All exhibit no noticeable bias and all estimates of $\alpha$ have similar variances.

## 6.5 Example.

Here we reconsider the example included in Fellner (1986). Briefly, the results pertain to the determination of metal content in two types of material, Types 1 and 2, from which 18 and 13 lots were respectively randomly selected. From each lot two samples were then taken and each divided into two sub-samples. Each sub-sample was analysed by a different chemist in duplicate. Hence a nested model is appropriate.

Let $y_{ijklm}$ denote the $m^{th}$ $(m = 1,2)$ result for the $l^{th}$ $(l = 1,2)$ sub-sample from the $k^{th}$ $(k = 1,2)$ sample drawn from the $j^{th}$ $(j = 1,\cdots,n_i)$ lot of material type $i$ $(i = 1,2)$, where $n_1 = 18$ and $n_2 = 13$. The model for the results is then

$$y_{ijklm} = \mu_i + a_{ij} + b_{ijk} + c_{ijkl} + e_{ijklm}$$

where $a_{ij}$ has zero mean and variance $\sigma_L^2$, that is $a_{ij} \sim (0, \sigma_L^2)$, $b_{ijk} \sim (0, \sigma_S^2)$, $c_{ijkl} \sim (0, \sigma_C^2)$ and $e_{ijklm} \sim (0, \sigma^2)$ with all random components uncorrelated. Here $\sigma_L^2$ denotes the variances of lots, $\sigma_S^2$ the variance of sample within lots (sample variance) whilst $\sigma_C^2$ is a combination of variation between chemists and the variation of sub-samples within samples and referred to as chemist variance. Finally, $\sigma^2$ denotes the remaining source of variability, that is analysis variance.

Fitting the model using Robust Methods 1, 2 and 3 above, taking $\psi_k(x)$ in each case as the Huber-psi function with $k = 2$, gave the fixed parameters and the variance components estimates given in Table 6.3 . Also given in Table 6.3, in brackets, are estimates of the standard errors of the fixed parameters in the model

based on the approximation to the variance of $\hat{\alpha}$ given in (6.14). This approximation should be satisfactory as $k$ is not too small.

From Table 6.3 we immediately notice that, to the number of figures reported, Robust Methods 1 and 3 give the same parameter estimates. For both of these methods twelve iterations were required to achieve convergence using the same criterion as used in the simulations. On the otherhand Robust Method 2, which required twenty iterations to converge, gives lower variance component estimates for each component and particularly so for the "Analysis variance" component. Based on the observations in the previous sections it is recommended here that the estimates obtained using Robust method 2 are to be preferred.

**Table 6.3:** Metallic oxide summary results.

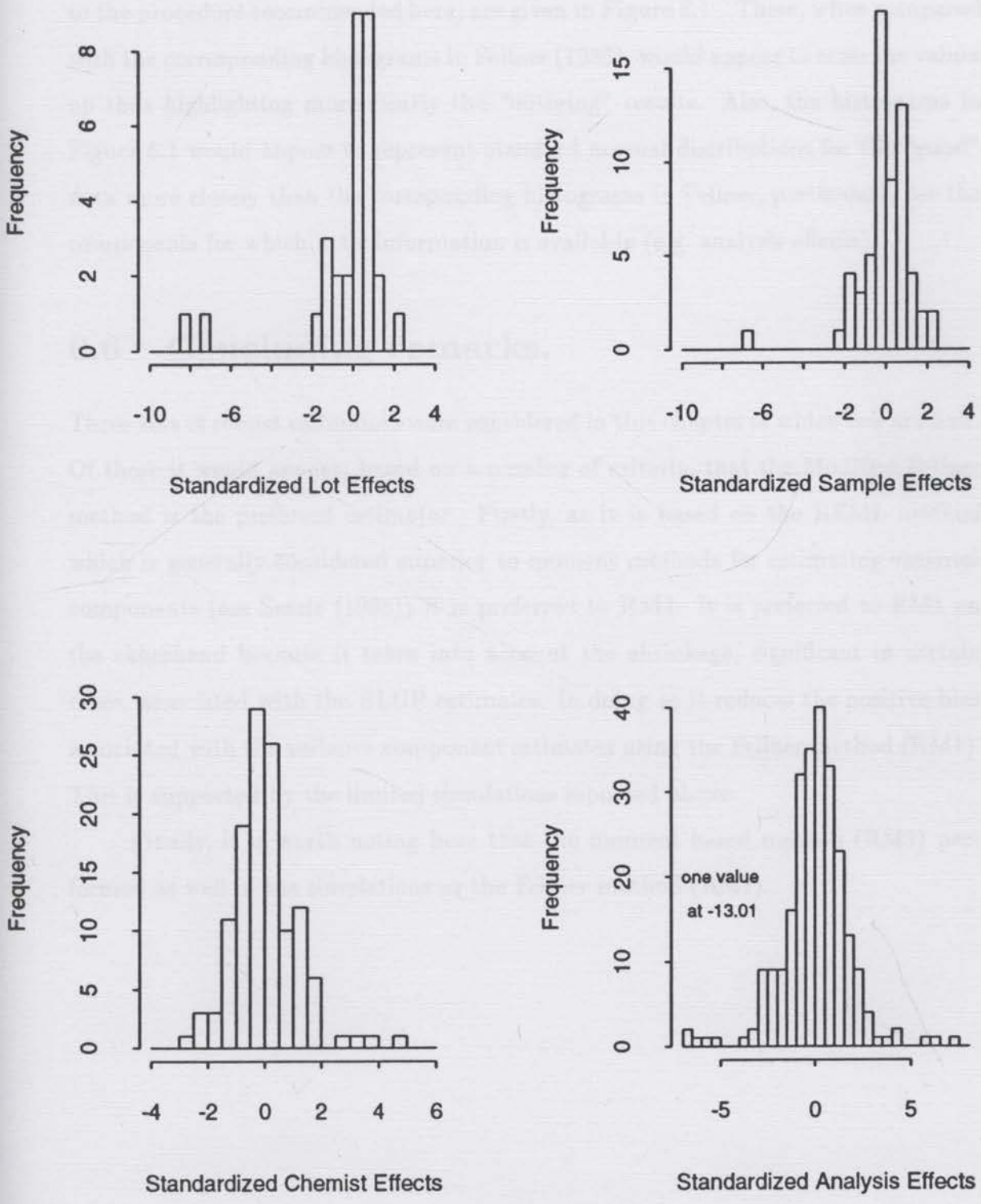
| | Robust Method Estimates | | |
|---|---|---|---|
| Parameter | RM1 | RM2 | RM3 |
| Type 1 mean | 3.86 (0.11) | 3.87 (0.10) | 3.86 (0.11) |
| Type 2 mean | 3.34 (0.13) | 3.35 (0.12) | 3.34 (0.13) |
| Lot variance | 0.176 | 0.153 | 0.176 |
| Sample variance | 0.037 | 0.026 | 0.037 |
| Chemist variance | 0.034 | 0.030 | 0.034 |
| Analysis variance | 0.037 | 0.012 | 0.037 |

Also as part of the analysis we obtained the "standardized values" for each of the random components in the model. In Fellner (1986) the recommended procedure for obtaining the "standardized values" is to divide each element of $\hat{\beta}$ and $\hat{\epsilon}$ by

$$\left[\hat{\sigma^2}(1-s^2)\right]^{1/2}$$

where, for the matrix $\mathbf{C}$ defined in (6.1), $s^2$ is the corresponding diagonal element of $\mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'$ and $\hat{\sigma^2}$ is the appropriate estimated variance. Here though, instead

# Figure 6.1: Standardized Effects for Metallic Oxide Data



of standardizing according to the procedure given in Fellner (1986), we recommend standardizing the elements of $\hat{\beta}$ and $\hat{\epsilon}$ based on the corresponding diagonal elements of the variances as given in (6.15) and (6.16).

Histograms of the standardized random components, standardized according to the procedure recommended here, are given in Figure 6.1 . These, when compared with the corresponding histograms in Fellner (1986), would appear to scale the values up thus highlighting more clearly the "outlying" results. Also, the histograms in Figure 6.1 would appear to represent standard normal distributions for the "good" data more closely than the coresponding histograms in Fellner, particularly for the components for which little information is available (e.g. analysis effects).

## 6.6 Concluding remarks.

Three sets of robust estimators were considered in this chapter of which two are new. Of these it would appear, based on a number of criteria, that the Modified Fellner method is the preferred estimator. Firstly, as it is based on the REML method which is generally considered superior to moment methods for estimating variance components (see Searle (1995)) it is preferred to RM3. It is preferred to RM1 on the otherhand because it takes into account the shrinkage, significant in certain cases, associated with the BLUP estimates. In doing so it reduces the positive bias associated with the variance component estimates using the Fellner method (RM1). This is supported by the limited simulations reported above.

Finally, it is worth noting here that the moment based method (RM3) performed as well in the simulations as the Fellner method (RM1).

# Chapter 7

# Robust estimation in generalized linear mixed models.

A natural extension to the linear mixed model considered in the previous chapter is the generalized linear mixed model (GLMM). Here the model, conditional on the random components which are assumed to be realizations from Normal distributions, corresponds to the generalized linear model covered in the text by McCullagh and Nelder (1989). One advantage of the inclusion of random components in the linear regression model, for example time effects associated with repeated measures, is that it allows for the modelling of correlations in the data.

Particular cases of GLMMs are represented in the literature: see for example Gilmour, Anderson and Rae (1985) and Karim and Zeger (1992). It has only been over the last few years however that more general procedures have been developed for GLMMs, both for "subject specific" (SS) and "population averaged" (PA) models as defined in Zeger, Liang and Albert (1988). Schall (1991) proposed an approximate procedure based on a first order Taylor's series approximation, with the approach appropriate to SS models. Breslow and Clayton (1993) obtain the same solution as Schall (1991) from a different approach which utilises the penalized quasi-likelihood. Also given in Breslow and Clayton (1993) is an estimating procedure appropriate to the PA model. Here the procedure is obtained using a marginal quasi-likelihood but the procedure could equally well have been obtained using a linearization argu-

ment as in Schall (1991) applied to the marginal model for the data. An alternative approach is given in McGilchrist (1994) wherein the solution is obtained by approximating the likelihood in the region of the maximum by a quadratic. This approach corresponds to the method of Schall (1991) for GLMMs but can be extended to a wider class of models. Also to have considered this problem, both for SS and PA models, has been Wolfinger and O'Connell (1993). Their approach is based on a (restricted) pseudo-likelihood.

For the above estimating procedures it has been implicitly assumed that all the data correspond to the specified GLMM. In general this will not be the case and instead it can be expected that some of the data will have arisen from an alternative model. In such cases these outlying results may have a significant influence on the parameter estimates for the model appropriate to the bulk of the data. Hence the need for robust methods for GLMMs analogous to methods for normal linear mixed models.

In this chapter we develop an approximate robust procedure for GLMMs, based on the estimating procedure of Schall (1991) and the results from Chapter 6. We begin in Section 7.1 by giving a representation for the generalized linear mixed model. In Section 7.2 robust estimates of the regression parameters, both fixed and random, are obtained for the case where the variance components and the dispersion parameter are known. Section 7.3 considers the robust estimation of the variance components and the dispersion parameter. A small simulation study is reported in Section 7.4 and in the following section an example of the robust estimation procedure is considered. This example corresponds to a data set from the National Asbestos Program discussed in Chapter 1 with the data set containing some obvious and some not so obvious outlying results. Again we finish off with some concluding remarks contained in Section 7.6.

## 7.1 General model.

The generalized linear mixed model (GLMM) has the form

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{e} \tag{7.1}$$

where $\mathbf{y}$ is an $(n \times 1)$ response vector and $\boldsymbol{\mu}$ and $\mathbf{e}$ are the systematic and random components of the model respectively. The systematic component is linked via a link function $g(\cdot)$ to a linear combination of fixed and random effects by

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}_1\mathbf{b}_1 + \cdots + \mathbf{U}_c\mathbf{b}_c \tag{7.2}$$

where $g(\cdot)$ is a differentiable, strictly monotone function with inverse denoted by $h(\cdot)$. We assume $\mathbf{X}$ is a full column rank matrix and let $\boldsymbol{\alpha}$ be a $(p \times 1)$ vector of fixed effects whilst $\mathbf{b}_i$ , $i = 1, \cdots, c$ is a vector of random effects of length $q_i$. We also assume that the $\mathbf{b}_i$ are independent with $\mathbf{b}_i \sim N(\mathbf{0}, \sigma_i^2\mathbf{I})$ , $i = 1, \cdots, c$ and that they are uncorrelated with $\mathbf{e}$.

To simplify later expressions we let $q = \sum_{i=1}^{c} q_i$ and as in Chapter 6 let $\mathbf{U} = \{_r\ \mathbf{U}_i\}_{i=1}^{c}$, $\mathbf{b} = \{_c\ \mathbf{b}_i\}_{i=1}^{c}$, $\mathbf{D}_i = \sigma_i^2\,\mathbf{I}_{q_i}$ , $i = 1, \cdots, c$ and $\mathbf{D} = \{_d\ \mathbf{D}_i\}_{i=1}^{c}$. Thus $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}\mathbf{b}$ and $\mathrm{Var}(\mathbf{b}) = \mathbf{D}$.

To complete the model we assume that, conditional on a fixed $\mathbf{b}$, the elements of $\mathbf{y}$ are independent and have distributions in the exponential dispersion family of the form

$$f_{y_i}(y; \theta_i, \phi) = \exp\left\{\frac{y\theta_i - K(\theta_i)}{\phi\, a_i} + c(y, \phi)\right\} \tag{7.3}$$

where $K(\cdot)$ and $c(\cdot)$ are specific functions, $a_i$ are known constants and $\phi$ is the dispersion parameter which may or may not be known. Conditional on $\mathbf{b}$, we denote the mean and variance-covariance matrix of $\mathbf{y}$ by $\boldsymbol{\mu}$ and $\phi\mathbf{V}(\boldsymbol{\mu})$ respectively, with $\mathbf{V}(\boldsymbol{\mu})$ a diagonal matrix with diagonal elements $a_i v(\mu_i) = a_i \ddot{K}(\theta_i)$, where $\ddot{K}$ denotes the second derivative of $K(\cdot)$. Therefore

$$\mathbf{V}(\boldsymbol{\mu}) = \left\{_d\ a_i v(\mu_i)\right\}_{i=1}^{n} .$$

To obtain approximate estimates of the parameters for the above model we can use the procedure of Schall (1991). Here the data vector $\mathbf{y}$ is transformed, using

the link function, to a new dependent variable $\mathbf{z} = g(\mathbf{y})$ which, to a first order approximation, is given by $\mathbf{z} = g(\boldsymbol{\mu}) + \{_d\, \dot{g}(\mu_i)\}_{i=1}^n\ (\mathbf{y} - \boldsymbol{\mu})$, where $\dot{g}(\mu_i)$ denotes the first derivatives of $g(\cdot)$ evaluated at $\mu_i$. Hence $\mathbf{z}$, termed the adjusted dependent variable, approximately satisfies

$$\begin{aligned} \mathbf{z} &= \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}\mathbf{b} + \{_d\, \dot{g}(\mu_i)\}_{i=1}^n\ \mathbf{e}\,, \\ &= \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}\mathbf{b} + \boldsymbol{\eta}. \end{aligned} \tag{7.4}$$

Letting

$$\begin{aligned} w_i(\mu_i) &= \left( a_i\, [\dot{g}(\mu_i)]^2 v(\mu_i) \right)^{-1}\ ,\ i = 1, \cdots, n\ ,\ \text{and} \\ \mathbf{W}(\boldsymbol{\mu}) &= \left\{ _d\, w_i(\mu_i) \right\}_{i=1}^n\ , \end{aligned} \tag{7.5}$$

we have based on the first order approximation above, that the mean and variance of $\mathbf{z}$, conditional on $\mathbf{b}$, are

$$\begin{aligned} E(\mathbf{z}|\mathbf{b}) &= \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}\mathbf{b}\,, \\ V(\mathbf{z}|\mathbf{b}) &= \phi\, \mathbf{W}^{-1}(\boldsymbol{\mu}) \end{aligned}$$

which leads to

$$\begin{aligned} E(\mathbf{z}) &= \mathbf{X}\boldsymbol{\alpha}\,, \\ V(\mathbf{z}) &= \phi\, E\left[\mathbf{W}^{-1}(\boldsymbol{\mu})\right] + \mathbf{U}\mathbf{D}\mathbf{U}' \end{aligned}$$

as an approximation. Further, based on a first order approximation we have

$$E\left[\mathbf{W}^{-1}(\boldsymbol{\mu})\right] = \mathbf{W}^{-1}\left(h(\mathbf{X}\,\boldsymbol{\alpha} + \mathbf{U}\,\mathbf{0})\right) \tag{7.6}$$

where $\mathbf{0}$, the expectation of $\mathbf{b}$, is retained for later development of robust estimators. For the case where the link function $g(\cdot)$ corresponds to the canonical link (i.e. $g(\mu_i) = \theta_i$) we have

$$E\left[\mathbf{W}^{-1}(\boldsymbol{\mu})\right] = \left\{ _d\, a_i\, E\left[v^{-1}(\mu_i)\right] \right\}_{i=1}^n \tag{7.7}$$

since in this case $\dot{g}(\mu_i) = \partial\, \theta_i / \partial\, \mu_i = [v(\mu_i)]^{-1}$ .

From here on we denote $\mathbf{W}(\boldsymbol{\mu})$ evaluated at a realized value of $\boldsymbol{\mu}$ by $\mathbf{W}_\mu$ and let

$$\mathbf{W}_e = \left(E\left[\mathbf{W}^{-1}(\boldsymbol{\mu})\right]\right)^{-1} . \tag{7.8}$$

We now assume that $w_i^{-1}(\mu_i)$ $(i = 1, \cdots, n)$ changes slowly as a function of $\mu_i$ about $E\left[w_i^{-1}(\mu_i)\right]$ thus allowing us to treat the conditional variance of $\mathbf{z}$ given $\mathbf{b} = \boldsymbol{\beta}$, that is $\phi\, \mathbf{W}_\mu^{-1}$, as a constant matrix which depends on the unknowns $\boldsymbol{\alpha}$, the variance components $(\sigma_i^2,\ i = 1, \cdots, c)$ and $E(\mathbf{b})$. Here the latter may contain elements other than zero if outliers are present in the data. Under this assumption $\mathbf{W}_\mu^{-1}$ and $\mathbf{W}_e^{-1}$ are approximately equivalent. A variation of this slowly varying assumption is also assumed by Breslow and Clayton (1993) in their derivation of estimators via the penalized quasilikelihood approach. Justification of the work of Schall (1991) is also contingent on this assumption though this is not stated.

## 7.2 Robust estimation of regression parameters - variance components and $\phi$ known.

### 7.2.1 Outline of a standard non-robust procedure.

Before developing robust methods for estimating the regression parameters when the variance components and $\phi$ are known consider the Schall (1991) procedure in the corresponding situation. That procedure is to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, where the latter is the realized value of $\mathbf{b}$, via the best linear unbiased prediction (BLUP) method of Henderson (1963). These estimates of Schall, denoted by $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are, with the exception for one modification to be mentioned below, the least squares solutions to the following overdetermined set of equations:

$$\begin{pmatrix} \sqrt{\phi^{-1}}\, \mathbf{W}_\mu^{1/2}\mathbf{X} & \sqrt{\phi^{-1}}\, \mathbf{W}_\mu^{1/2}\mathbf{U} \\ \mathbf{0} & \mathbf{D}^{-1/2} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \sqrt{\phi^{-1}}\, \mathbf{W}_\mu^{1/2}\mathbf{z} \\ \mathbf{D}^{-1/2}\mathbf{0} \end{pmatrix} \tag{7.9}$$

where $\mathbf{z} = g(\mathbf{y})$ is the realized vector of the adjusted dependent variable and, as in Chapter 6, $\mathbf{0} = (\mathbf{0}_1', \cdots, \mathbf{0}_c')'$. Here $\mathbf{0} = E(\mathbf{b})$. Alternatively, the estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$

satisfy the following equations

$$\begin{pmatrix} \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{X} & \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{U} \\ \phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{X} & \mathbf{D}^{-1}+\phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{U} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{z} \\ \phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{z}+\mathbf{D}^{-1}\mathbf{0} \end{pmatrix} \tag{7.10}$$

where, as in Chapter 6, we retain $\mathbf{D}^{-1}\mathbf{0}$ to facilitate the later development of robust estimates of the parameters.

The modification to the Schall procedure mentioned above is that in the equations for $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ the matrix $\mathbf{W}$ has a subscript to indicate that it is to be evaluated at a realized value of $\boldsymbol{\mu}$. This modification needs some comments. Firstly, if the slowly varying weight assumption mentioned at the end of the previous section holds and there are no outliers in the data, then the subscript on $\mathbf{W}$ above is superfluous and the equations correspond to those of Schall (1992).

Secondly, the matrix $\mathbf{W}_\mu$ is used in the above equations and not $\mathbf{W}_e$ as a consequence of the formulation of the BLUP procedure. Here, the first equation in (7.9) is based on the mean and variance of $\mathbf{z}$ given $\mathbf{b} = \boldsymbol{\beta}$, that is $\mathbf{X}\boldsymbol{\alpha} + \mathbf{U}\boldsymbol{\beta}$ and $\phi\,\mathbf{W}_\mu^{-1}$ respectively whilst the second equation in (7.9) is based on the mean and variance of the distribution of $\mathbf{b}$. Now, according to our earlier assumption that $w_i^{-1}(\mu_i)$ ; $(i = 1, \cdots, n)$ changes slowly as a function of $\mu_i$ about $E\left[w_i^{-1}(\mu_i)\right]$ we should be able to interchange $\mathbf{W}_\mu$ with $\mathbf{W}_e$. However we retain $\mathbf{W}_\mu$ because when we later replace this weight matrix by plug in estimates, that is plugging in estimates of $\boldsymbol{\alpha}, \boldsymbol{\beta}$, etc. we expect estimates based on $\mathbf{W}_\mu$ to be better than estimates based on $\mathbf{W}_e$. This is because $\mathbf{W}_e$ is a function of $E(\mathbf{b})$ where the elements of the latter will be assumed to be zero and only modified, reluctantly at that, if the corresponding data values are outliers. Hence, if the elements of $\mathbf{W}_e$ are not slowly varying functions of the elements of $E(\mathbf{b})$ then $\mathbf{W}_\mu$ will be more sensitive to outliers in the element $\mathbf{b}$ than $\mathbf{W}_e$.

In the situation where $\mathbf{D}$ is not positive definite (e.g. if some variance components are zero), an alternative representation of (7.10) based on the form given in

Harville (1977), is

$$\begin{pmatrix} \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{X} & \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{U}\,\mathbf{D} \\ \phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{X} & \mathbf{I}+\phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{U}\,\mathbf{D} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\nu}} \end{pmatrix} = \begin{pmatrix} \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{z} \\ \phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{z}+\mathbf{D}^{-}\mathbf{0} \end{pmatrix} \tag{7.11}$$

where $\boldsymbol{\beta} = \mathbf{D}\,\boldsymbol{\nu}$ and $\mathbf{D}^{-}$ is a diagonal matrix with elements the inverse of the corresponding element of $\mathbf{D}$ if that element is positive and zero otherwise. We then let $\hat{\boldsymbol{\beta}} = \mathbf{D}\,\hat{\boldsymbol{\nu}}$. This form will be useful when we replace $\mathbf{D}$ with moment based estimates in which case it is possible that some variance components are estimated as zero.

### 7.2.2 Robustification of the standard procedure.

The procedure we now give is based on the method of Fellner (1986) as outlined in Chapter 6 but with appropriate modification to take into account that the distribution of the random error may differ significantly from a normal distribution. The procedure will be to move those "observed" values differing significantly from their expectation closer to the expected value. Here we include $\mathbf{0}$ in (7.10) as "observed" as the BLUP procedure treats it as the observed value of $\mathbf{b}$.

To begin with we consider $\mathbf{b}$ and hence $\boldsymbol{\mu}$ fixed and replace $\mathbf{z}$ by $\tilde{\mathbf{z}}$, where outlying results in the former are downweighted in the latter. One way of doing this is simply to apply the procedure of Fellner (1986) to $\mathbf{z}$. However, as $g(\cdot)$ is a linearizing rather than a normalizing transformation we instead recommend replacing $\mathbf{y}$ by $\tilde{\mathbf{y}}$, where outlying results in $\mathbf{y}$ are replaced by values more in line with their expectation whilst taking account of any asymmetry. Then we set $\tilde{\mathbf{z}} = g(\tilde{\mathbf{y}})$. To achieve this we define generalized residuals $r_i = r_i(y_i;\mu_i)$ corresponding to the model in (7.1), form pseudo generalized residuals by shrinking "large" generalized residuals and then form pseudo observations to correspond to the pseudo generalized residuals.

Here the most appropriate generalized residual would appear to be the Anscombe residual defined below (so named following the convention of McCullagh and Nelder (1989)), given that the generally accepted alternative, the deviance residual,

is not appropriate in our context as the BLUP method does not correspond to maximizing a likelihood.

Given $\mu$, the standard form of the Anscombe residual for an observation $y_i$ from a distribution of the form (7.3) is given by

$$r_i(y_i;\mu_i) \;=\; \frac{t(y_i)-t(\mu_i)}{\dot{t}(\mu_i)\sqrt{\phi\, a_i\, v(\mu_i)}} \tag{7.12}$$

where $t(y)$ is the "normalizing" transformation given up to a proportionality constant by the indefinite integral $\int v^{-1/3}(y)\,dy$ (see Barndorff-Nielsen (1978) pp 177-9) and $\dot{t}(y)$ is the derivative of $t(y)$. These residuals are approximately standard normal random variables in that they have zero skewness with approximately zero mean and unit variance. An adjusted version of the Anscombe residual, studied by Pierce and Schafer (1986) and shown by them to approximate standard normal random variables very well, includes a second order correction for the expected value of $t(Y)$. This adjusted version of the Anscombe residual is given by

$$r_i(y_i,\mu_i) \;=\; \frac{t(y_i)-t(\mu_i)-\frac{\phi\, a_i}{2}\ddot{t}(\mu_i)\, v(\mu_i)}{\dot{t}(\mu_i)\sqrt{\phi\, a_i\, v(\mu_i)}} \tag{7.13}$$

where $\ddot{t}(\mu)$ denotes the second order derivative of $t(\cdot)$ evaluated at $\mu$. The two forms of the Anscombe residual, standard and adjusted, for commonly encountered distributions in the exponential family are given in Table 7.1.

One modification recommended in practice for the Anscombe residual, in either standard or adjusted form, is to include a correction for continuity when $y$ is integer valued. In this case Pierce and Schafer (1986) recommend replacing $y$ by $y \pm \frac{1}{2}$, moving towards the center of the distribution.

We now define the pseudo-observations more specifically. Let $\psi_k(x)$ be an odd, bounded, not identically zero function associated with M estimation and satisfying

$$\psi_k(cx) \;=\; c\,\psi_{k/c}(x) \qquad \text{for all } \; c>0\,. \tag{7.14}$$

Also, let $\hat{\mu}_i = h(\mathbf{x}_i'\,\hat{\boldsymbol{\alpha}} + \mathbf{u}_i'\,\hat{\boldsymbol{\beta}})$, the current estimate of $\mu_i$, where $\mathbf{x}_i'$ and $\mathbf{u}_i'$ are the $i^{th}$ rows of $\mathbf{X}$ and $\mathbf{U}$ respectively. Then with $r_i^{-1}(y;\mu)$ denoting the inverse

**Table 7.1: Generalized (Anscombe) residuals**

| Distribution | Standard form | Adjusted form |
|---|---|---|
| Binomial$(m,p)$ $(\mu = mp)$ | $\frac{t(y/m)-t(p)}{(pq)^{1/6}/\sqrt{m}}$ | $\frac{t(y/m)-[t(p)+(pq)^{-1/3}(2p-1)/(6m)]}{(pq)^{1/6}/\sqrt{m}}$ |
| Poisson$(\mu)$ | $\frac{y^{2/3}-\mu^{2/3}}{2\mu^{1/6}/3}$ | $\frac{y^{2/3}-[\mu^{2/3}-\mu^{-1/3}/9]}{2\mu^{1/6}/3}$ |
| Gamma$(\alpha,\beta)$ $(\mu = \alpha\beta)$ | $\frac{y^{1/3}-\mu^{1/3}}{\mu^{1/3}/(3\sqrt{\alpha})}$ | $\frac{y^{1/3}-\mu^{1/3}[1-1/(9\alpha)]}{\mu^{1/3}/(3\sqrt{\alpha})}$ |
| Inverse Gaussian$(\mu,\lambda)$ | $\frac{\ln(y)-\ln(\mu)}{\sqrt{\mu/\lambda}}$ | $\frac{\ln(y)-[\ln(\mu)-\mu/(2\lambda)]}{\sqrt{\mu/\lambda}}$ |
| Normal$(\mu,\sigma^2)$ | $\frac{y-\mu}{\sigma}$ | $\frac{y-\mu}{\sigma}$ |

For binomial, $t(u) = \int_0^u s^{-1/3}(1-s)^{-1/3}ds$.

Gamma$(\alpha,\beta)$ density: $\beta^{-\alpha}y^{\alpha-1}\exp(y/\beta)/\Gamma(\alpha)$ ; $y > 0$.

Inverse Gaussian$(\mu,\lambda)$ density: $\sqrt{\lambda/(2\pi y^3)}\exp[-\lambda(y-\mu)^2/(2\mu^2 y)]$ ; $y > 0$.

function associated with $r_i(y;\mu)$, which is well defined since by definition $t(y)$ is non-decreasing, we define the pseudo observation $\tilde{y}_i$ as

$$\tilde{y}_i \;=\; r_i^{-1}\left(\psi_k[r_i(y_i;\hat{\mu}_i)]\right) \tag{7.15}$$

or, if $\tilde{y}_i$ so defined is not realizable, then its nearest realizable value closest to $y_i$.

For example, if $\psi_k(x)$ is chosen as the Huber-psi function, then $\tilde{y}_i$ is given as

$$\tilde{y}_i \;=\; \max\{y_L(\hat{\mu}_i), \min\{y_i, y_U(\hat{\mu}_i)\}\}\;,$$

where

$$\begin{aligned} y_L(\mu) \;&=\; \max\{y \;:\; r_i(y;\mu) \le -k\}\;, \text{ and} \\ y_U(\mu) \;&=\; \min\{y \;:\; r_i(y;\mu) \ge k\}\;. \end{aligned}$$

An alternative choice of $\psi_k(x)$ is the Tukey biweight defined in (6.11).

The pseudo-adjusted dependent variable is then given by

$$\tilde{\mathbf{z}} \;=\; g(\tilde{\mathbf{y}}) \tag{7.16}$$

It remains now to modify the $\mathbf{0}$, the "observed" value of $\mathbf{b}$, in (7.10) or the alternative version (7.11). Here, as in Chapter 6, we replace $\mathbf{0}$ by $\tilde{\mathbf{0}}$, with the latter defined in (6.10) or, if $\mathbf{D}$ is singular, (6.20) and where $\psi_k(x)$ satisfies the conditions above. Here it should be noted that it is not necessary to use the same $\psi$ function nor the same tuning constant $k$ when downweighting the elements of $\mathbf{y}$ and $\boldsymbol{\beta}_i$ $(i = 1, \cdots, c)$. We will assume from here on though that the same function and tuning constant is used throughout.

The outlier resistant normal equations then become,

$$\begin{pmatrix} \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{X} & \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{U} \\ \phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{X} & \mathbf{D}^{-1}+\phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{U} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} \;=\; \begin{pmatrix} \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\tilde{\mathbf{z}} \\ \phi^{-1}\mathbf{U}'\mathbf{W}_\mu\tilde{\mathbf{z}}+\mathbf{D}^{-1}\tilde{\mathbf{0}} \end{pmatrix} \tag{7.17}$$

or using the Harville form,

$$\begin{pmatrix} \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{X} & \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\mathbf{U}\,\mathbf{D} \\ \phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{X} & \mathbf{I}+\phi^{-1}\mathbf{U}'\mathbf{W}_\mu\mathbf{U}\,\mathbf{D} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\nu}} \end{pmatrix} \;=\; \begin{pmatrix} \phi^{-1}\mathbf{X}'\mathbf{W}_\mu\tilde{\mathbf{z}} \\ \phi^{-1}\mathbf{U}'\mathbf{W}_\mu\tilde{\mathbf{z}}+\mathbf{D}^{-}\tilde{\mathbf{0}} \end{pmatrix} \tag{7.18}$$

where $\hat{\boldsymbol{\beta}} = \mathbf{D}\,\hat{\boldsymbol{\nu}}$. Here the solution will need to be obtained iteratively with the estimate of $\mathbf{W}_{\mu}$ given by $\mathbf{W}_{\hat{\mu}} = \mathbf{W}\left(h(\mathbf{X}\,\hat{\boldsymbol{\alpha}} + \mathbf{U}\,\hat{\boldsymbol{\beta}})\right)$.

Two points need to be made now. Firstly, the procedure above corresponds to the procedure of Fellner (1986) when the conditional distribution of $\mathbf{y}$ is Gaussian and the link function $g(\cdot)$ is the identity. This follows as the "normalizing" transformation $t(\cdot)$ for the Gaussian distribution is simply the identity function. The second and more important point concerns the definition of the pseudo-adjusted dependent variable, in particular when $\psi_k(x)$ corresponds to the Huber-psi fuction, but the comments should carry over to other $\psi_k(x)$ fuctions satisfying the required conditions. For asymmetric members of the exponential dispersion family the procedure will introduce some bias into the estimation. However, by using asymmetric Winsorization via the Anscombe residual, rather than symmetric Winsorization, this will reduce this bias. Also, as the more commonly encountered members of this family have short tails, this bias will be relatively small, provided the Winsorization points are not too close to the mean. Examples to support this claim are given in the following sub-section.

### 7.2.3 Estimation of the mean using Winsorization.

To robustly estimate the parameters of the generalized linear mixed model our procedure is to use Winsorization to reduce the influence of "outlying" results. In the case of the data conditionally having a normal distribution, Winsorization is reasonably straight forward with the Winsorizing points generally chosen to be symmetric about the mean. In this case, that is symmetric Winsorization for a normal population, the mean of the Winsorized population corresponds to the mean of the underlying population.

This is no longer the case for asymmetric (skewed) distributions as is the case for many members of the exponential dispersion family of distributions. In such cases, particularly with symmetric Winsorization, the mean of the Winsorized and un-Winsorized populations will differ. Hence, use of Winsorization to limit the influence of outlying results may introduce bias into the modelling of the mean.

To examine the extent of this bias, let $F(x)$ denote the underlying distribution with mean $\mu$ and assume that the distribution is Winsorized at the points $a$ and $b$ with $(a < b)$. The mean of the Winsorized distribution $\mu_W$ is then

$$\mu_W = aP(X \leq a) + \int_{a^+}^{b^-} x\, dF(x) + bP(X \geq b). \tag{7.19}$$

The bias introduced by Winsorizing is $\mu_W - \mu$.

Now, for a given $F(x)$ it will be possible to chose $a$ and $b$ so that $\mu_W = \mu$. The solution though for $a$ and $b$ will generally be intractable and hence some approximations will be needed. One approximation that takes into account the skewness of the distribution is to use symmetric-quantile Winsorization, that is choose $a$ and $b$ to correspond to $F^{-1}(p)$ and $F^{-1}(1-p)$ for some $p \in [0, 0.5]$. This method will not eliminate the bias but, by taking account of the skewness of the distribution, should reduce it compared to standard symmetric Winsorization. Also, provided $p$ is not chosen too large (i.e. $p$ not choosen near 0.5) the bias should be small relative to $\mu$ owing to the "shortness" of the tails for most members of the exponential dispersion family. In practice, for members of this family of distributions, an approximately equivalent and simpler procedure to symmetric-quantile Winsorization is to Winsorize according to the Anscombe residuals defined earlier.

To examine the bias introduced by Winsorization we now consider two members of the exponential dispersion family, these being the Poisson and Gamma families. For both families we will consider symmetric Winsorization at the points $\mu \pm 2\sqrt{\text{Var}(X)}$, symmetric-quantile Winsorization with $p = 0.0228$ and Winsorization based on the adjusted Anscombe residuals. In the latter case values of $x$ giving an adjusted Anscombe residual larger than 2 in absolute value are drawn in as far as possible so that the Anscombe residual is not less than 2 in magnitude. For the symmetric-quantile Winsorization, $p$ is taken as 0.0228 so that all three estimators are equivalent in the normal case.

#### 7.2.3.1 Poisson Family: $P(\lambda)$ , $\lambda > 0$.

Letting $F_\lambda(x) = P(X \leq x)$ with $X \sim P(\lambda)$ we have from (7.19) for integer $a$

and $b$ satisfying $a < \lambda < b$, $b \geq 1$

$$\mu_W = aF_\lambda(a) + I(b \geq a+2)\,\lambda\{F_\lambda(b-2) - F_\lambda(a-1)\} + b\{1 - F_\lambda(b-1)\} .$$

This result follows the result that for a member of the Poisson family and integers $n \geq 1$,

$$\int_{-\infty}^{n^+} x\, dF_\lambda(x) = \lambda\, F(n-1) .$$

For symmetric Winsorizing $a(\mu)$ and $b(\mu)$ are $\left[(\lambda - 2\sqrt{\lambda})\right] \vee 0$ and $\left]\lambda + 2\sqrt{\lambda}\right[$ respectively, where $[x]$ and $]x[$ are defined in Appendix A. For Winsorization based on the adjusted Anscombe residuals we have:

$$\begin{aligned} a(\mu) &= \left[\left\{\left(\lambda^{2/3} - \lambda^{-1/3}/9 - 4\lambda^{1/6}/3\right) \vee 0\right\}^{3/2} - 0.5\right] \vee 0 \\ b(\mu) &= \left]\left\{\left(\lambda^{2/3} - \lambda^{-1/3}/9 + 4\lambda^{1/6}/3\right)\right\}^{3/2} + 0.5\right[ . \end{aligned}$$

Graphed in Figures 7.1(a) and 7.1(b) are the biases $(\mu_W - \mu)$ for the three Winsorization schemes for $\lambda = .001(.01)1$ and $\lambda = 1(.01)10$ respectively, with plots within each figure having the same vertical scale. From these plots, for the range of $\lambda$ values considered, we firstly observe that the bias is not large for any Winsorization scheme, except possibly relatively so for symmetric Winsorization with small $\lambda$. Secondly, the bias for symmetric Winsorization is larger in magnitude than for the other two Winsorization schemes for almost all $\lambda$ considered. Finally, the bias for adjusted Anscombe Winsorization is generally of the same magnitude as for symmetric-quantile Winsorization. We therefore conclude that of the three procedures Winsorization based on the adjusted Anscombe residuals would be preferred given that symmetric-quantile Winsorization is computationally more intensive.

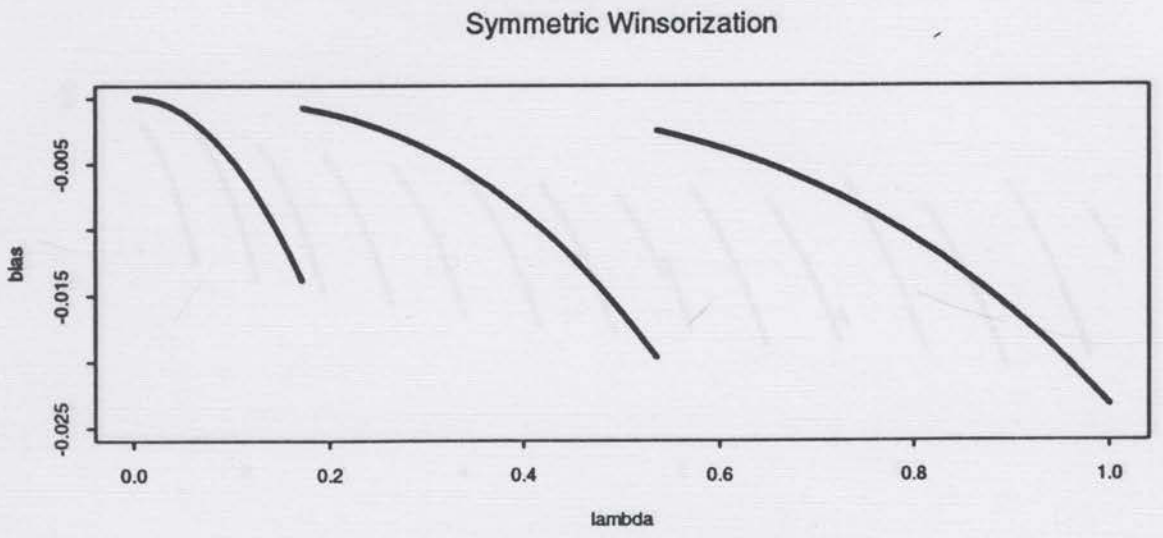
#### 7.2.3.2 Gamma Family: $\Gamma(\alpha, \beta)$ , $\alpha > 0$ , $\beta > 0$ ($\beta$ known).

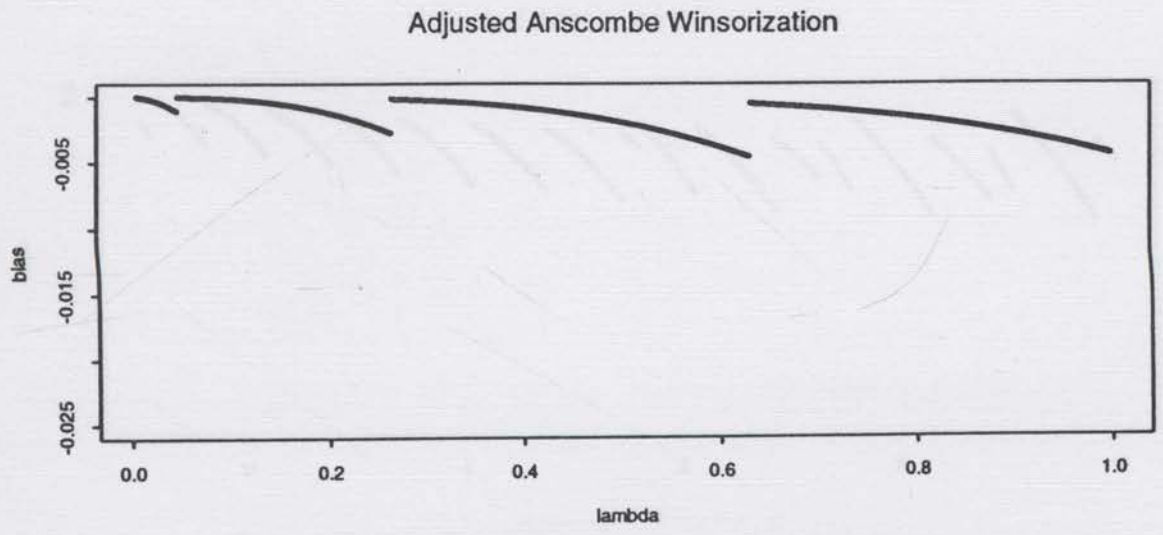
Let $F_{\alpha,\beta}(x) = P(X \leq x)$ for $X \sim \Gamma(\alpha, \beta)$.

Then from (7.19)

$$\mu_W = aF_{\alpha,1}(a/\beta) + \alpha\beta\{F_{\alpha+1,1}(b/\beta) - F_{\alpha+1,1}(a/\beta)\} + b(1 - F_{\alpha,1}(b/\beta) .$$

FIGURE 7.1(a) Poisson(lambda) distribution; lambda in [.001, 1]



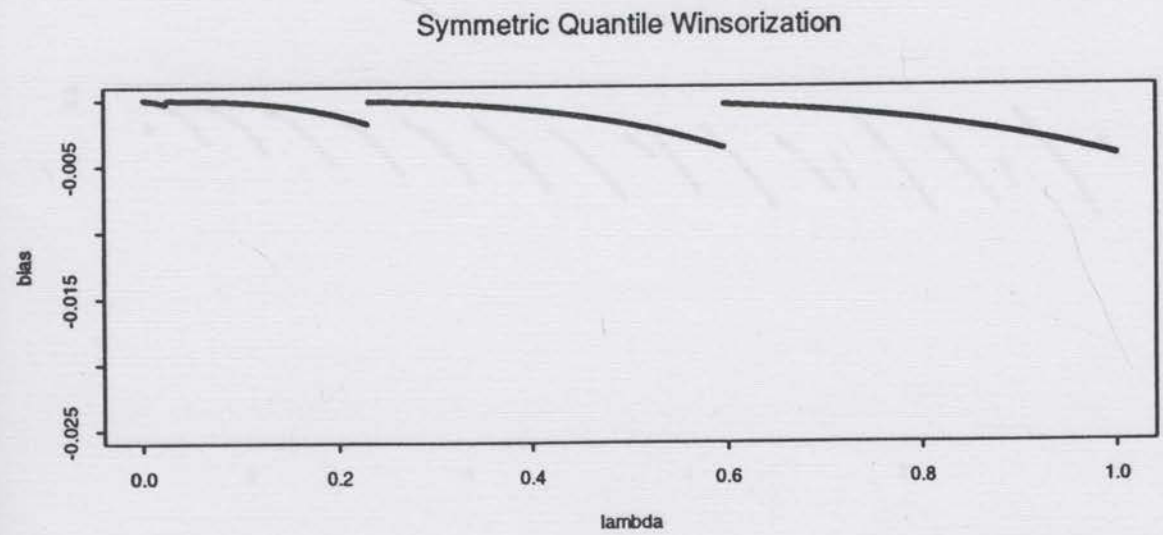
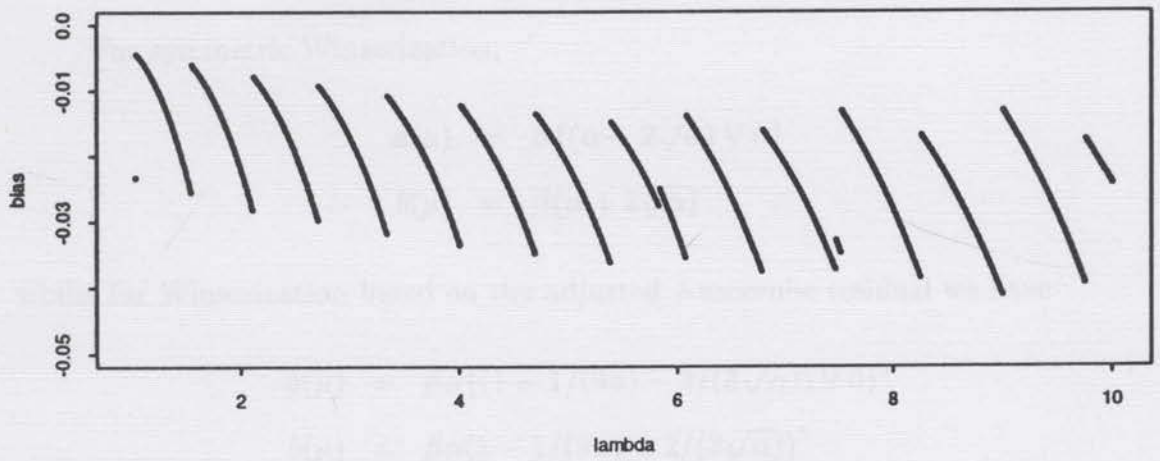
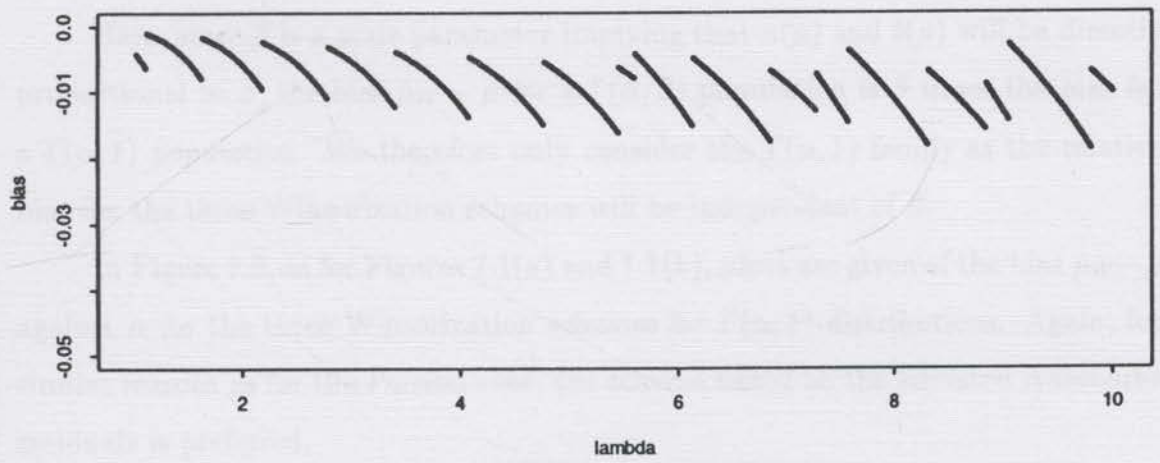



FIGURE 7.1(b) Poisson(lambda) distribution; lambda in [1, 10]

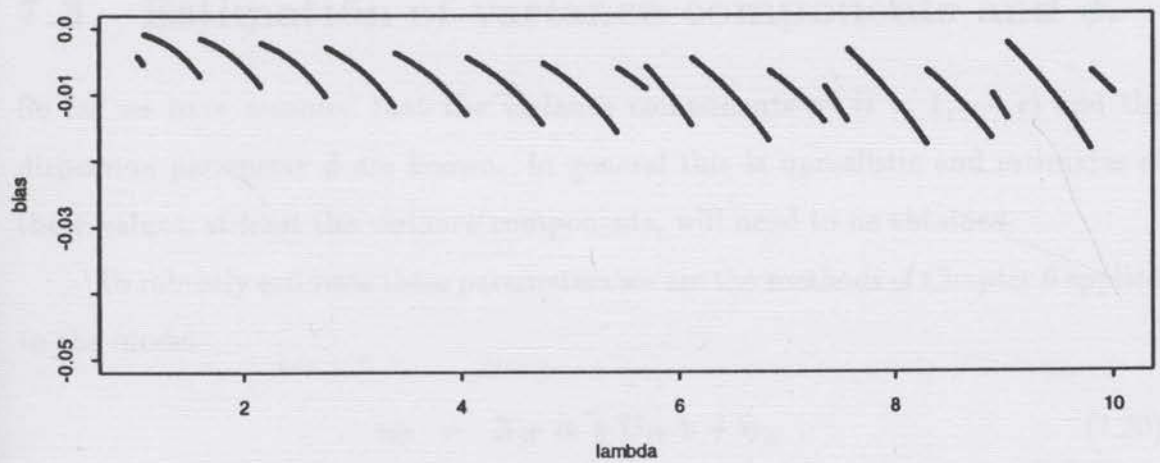
Symmetric Winsorization




Adjusted Anscombe Winsorization




Symmetric Quantile Winsorization



Here the result follows from the equality

$$\int_{-\infty}^{y} x\, dF_{\alpha,\beta}(x) \;=\; \alpha\beta\, F_{\alpha+1,1}(y/\beta)\,.$$

For symmetric Winsorization,

$$\begin{aligned} a(\mu) &= \beta\{(\alpha - 2\sqrt{\alpha}) \vee 0\} \\ b(\mu) &= \beta(\alpha + 2\sqrt{\alpha}) \end{aligned}$$

whilst for Winsorization based on the adjusted Anscombe residual we have

$$\begin{aligned} a(\mu) &= \beta\alpha\{(1 - 1/(9\alpha) - 2/(3\sqrt{\alpha})) \vee 0\}^3 \\ b(\mu) &= \beta\alpha(1 - 1/(9\alpha) + 2/(3\sqrt{\alpha}))^3 \end{aligned}$$

provided $(1 - 1/(9\alpha) + 2/(3\sqrt{\alpha})) > 0$, that is $\alpha > (\sqrt{2} - 1)^2/9 = 0.0191$.

Here, since $\beta$ is a scale parameter implying that $a(\mu)$ and $b(\mu)$ will be directly proportional to $\beta$, the bias $\mu_W - \mu$ for a $\Gamma(\alpha, \beta)$ population is $\beta$ times the bias for a $\Gamma(\alpha, 1)$ population. We therefore only consider the $\Gamma(\alpha, 1)$ family as the relative bias for the three Winsorization schemes will be independent of $\beta$.

In Figure 7.2, as for Figures 7.1(a) and 7.1(b), plots are given of the bias $\mu_W - \mu$ against $\alpha$ for the three Winsorization schemes for $\Gamma(\alpha, 1)$ distributions. Again, for similar reasons as for the Poisson case, the scheme based on the adjusted Anscombe residuals is preferred.

## 7.3 Estimation of variance components and $\phi$.

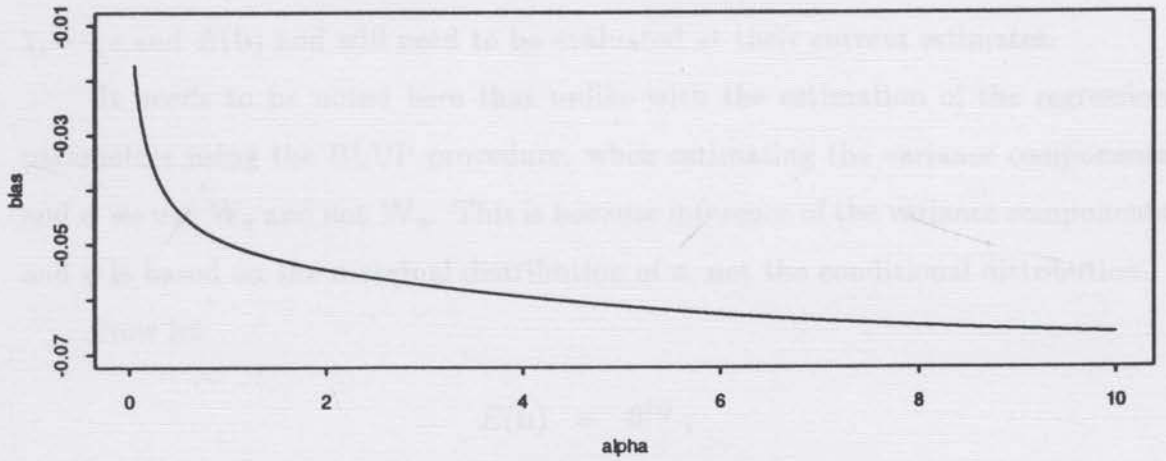
So far we have assumed that the variance components $\sigma_i^2$ $(i = 1, \cdots, c)$ and the dispersion parameter $\phi$ are known. In general this is unrealistic and estimates of these values, at least the variance components, will need to be obtained.

To robustly estimate these parameters we use the methods of Chapter 6 applied to the model

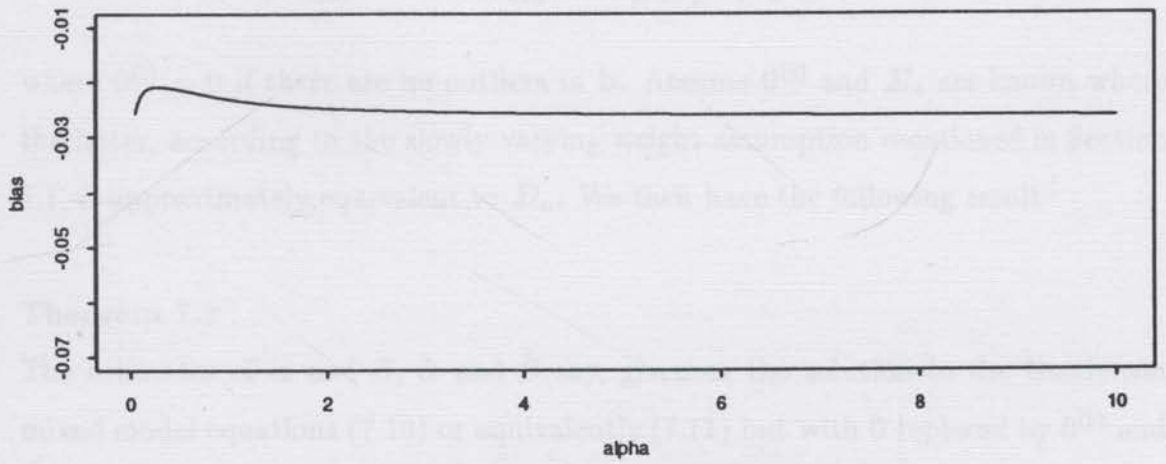
$$\mathbf{z}_W \;=\; \mathbf{X}_W\, \boldsymbol{\alpha} + \mathbf{U}_W\, \mathbf{b} + \boldsymbol{\eta}_W\,, \qquad (7.20)$$

FIGURE 7.2 Gamma(alpha,1) distribution; alpha in [.05,10]

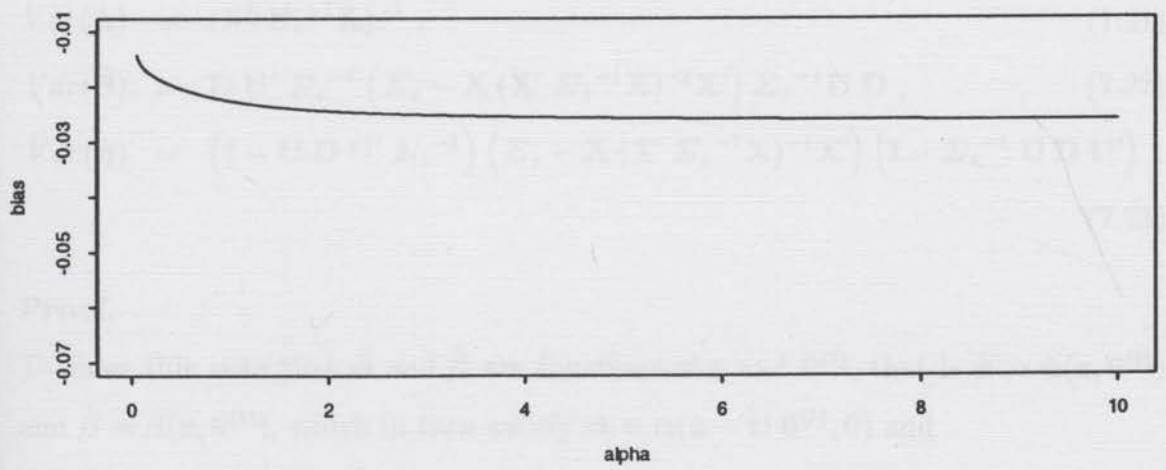
Symmetric Winsorization




Adjusted Anscombe Winsorization




Symmetric Quantile Winsorization



where $\mathbf{z}_W = \mathbf{W}_e^{1/2}\mathbf{z}$, $\mathbf{X}_W = \mathbf{W}_e^{1/2}\mathbf{X}$, $\mathbf{U}_W = \mathbf{W}_e^{1/2}\mathbf{U}$ and $\boldsymbol{\eta}_W = \mathbf{W}_e^{1/2}\boldsymbol{\eta}$. This model follows from (7.4). Here we have $E(\boldsymbol{\eta}_W) = \mathbf{0}$, $Cov(\boldsymbol{\eta}_W) = \phi\,\mathbf{I}$ and $Cov(\mathbf{b}, \boldsymbol{\eta}_W) = \mathbf{0}$. In practice $\mathbf{W}_e$ is unknown as it depends on $\boldsymbol{\alpha}$, the variance components $\sigma_i^2$ $i = 1, \cdots, c$ and $E(\mathbf{b})$ and will need to be evaluated at their current estimates.

It needs to be noted here that unlike with the estimation of the regression parameters using the BLUP procedure, when estimating the variance components and $\phi$ we use $\mathbf{W}_e$ and not $\mathbf{W}_\mu$. This is because inference of the variance components and $\phi$ is based on the marginal distribution of $\mathbf{z}$, not the conditional distribution.

Now let

$$\begin{aligned} E(\mathbf{b}) &= \mathbf{0}^{(b)}\,, \\ \boldsymbol{\Sigma}_\mu &= \phi\,\mathbf{W}_\mu^{-1} + \mathbf{U}\,\mathbf{D}\,\mathbf{U}'\,, \\ \text{and} \qquad \boldsymbol{\Sigma}_e &= \phi\,\mathbf{W}_e^{-1} + \mathbf{U}\,\mathbf{D}\,\mathbf{U}' \end{aligned}$$

where $\mathbf{0}^{(b)} = \mathbf{0}$ if there are no outliers in $\mathbf{b}$. Assume $\mathbf{0}^{(b)}$ and $\boldsymbol{\Sigma}_e$ are known where the latter, according to the slowly varying weight assumption mentioned in Section 7.1, is approximately equivalent to $\boldsymbol{\Sigma}_\mu$. We then have the following result.

**Theorem 7.1**

The estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ say, given as the solution to the Henderson mixed model equations (7.10) or equivalently (7.11) but with $\mathbf{0}$ replaced by $\mathbf{0}^{(b)}$ and $\mathbf{W}_\mu$ replaced by the constant matrix $\mathbf{W}_e$, and with $\hat{\boldsymbol{\eta}} = \mathbf{z} - \mathbf{X}\,\hat{\boldsymbol{\alpha}} - \mathbf{U}\,\hat{\boldsymbol{\beta}}$, satisfy

$$Var(\hat{\boldsymbol{\alpha}}) = (\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\mathbf{X})^{-1}\,, \tag{7.21}$$

$$Var(\hat{\boldsymbol{\beta}}) = \mathbf{D}\,\mathbf{U}'\,\boldsymbol{\Sigma}_e{}^{-1}\left(\boldsymbol{\Sigma}_e - \mathbf{X}\,(\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\mathbf{X})^{-1}\mathbf{X}'\right)\boldsymbol{\Sigma}_e{}^{-1}\mathbf{U}\,\mathbf{D}\,, \tag{7.22}$$

$$Var(\hat{\boldsymbol{\eta}}) = \left(\mathbf{I} - \mathbf{U}\,\mathbf{D}\,\mathbf{U}'\,\boldsymbol{\Sigma}_e{}^{-1}\right)\left(\boldsymbol{\Sigma}_e - \mathbf{X}\,(\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\mathbf{X})^{-1}\mathbf{X}'\right)\left(\mathbf{I} - \boldsymbol{\Sigma}_e{}^{-1}\,\mathbf{U}\,\mathbf{D}\,\mathbf{U}'\right)\,. \tag{7.23}$$

**Proof.**

To show this note that $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are functions of $\mathbf{z}$ and $\mathbf{0}^{(b)}$, that is $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(\mathbf{z}, \mathbf{0}^{(b)})$ and $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\mathbf{z}, \mathbf{0}^{(b)})$, which in turn satisfy $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(\mathbf{z} - \mathbf{U}\,\mathbf{0}^{(b)}, \mathbf{0})$ and

$\hat{\boldsymbol{\beta}} = \mathbf{0}^{(b)} + \boldsymbol{\beta}(\mathbf{z} - \mathbf{U}\,\mathbf{0}^{(b)}, \mathbf{0})$. Hence we have, replacing $\boldsymbol{\Sigma}_\mu$ with $\boldsymbol{\Sigma}_e$

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &= (\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\,\mathbf{X})^{-1}\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\left(\mathbf{z} - \mathbf{U}\,\mathbf{0}^{(b)}\right), \\
\hat{\boldsymbol{\beta}} &= \mathbf{0}^{(b)} + \mathbf{D}\,\mathbf{U}'\,\boldsymbol{\Sigma}_e{}^{-1}\left(\mathbf{z} - \mathbf{U}\,\mathbf{0}^{(b)} - \mathbf{X}\,\hat{\boldsymbol{\alpha}}\right) \\
&= \mathbf{0}^{(b)} + \mathbf{D}\,\mathbf{U}'\,\boldsymbol{\Sigma}_e{}^{-1}\left(\mathbf{I} - \mathbf{X}\,(\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\mathbf{X})^{-1}\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\right)\left(\mathbf{z} - \mathbf{U}\,\mathbf{0}^{(b)}\right), \text{ and} \\
\hat{\boldsymbol{\eta}} &= \mathbf{z} - \mathbf{X}\,\hat{\boldsymbol{\alpha}} - \mathbf{U}\,\hat{\boldsymbol{\beta}} \\
&= \left(\mathbf{I} - \mathbf{U}\,\mathbf{D}\,\mathbf{U}'\,\boldsymbol{\Sigma}_e{}^{-1}\right)\left(\mathbf{I} - \mathbf{X}\,(\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\mathbf{X})^{-1}\mathbf{X}'\,\boldsymbol{\Sigma}_e{}^{-1}\right)\left(\mathbf{z} - \mathbf{U}\,\mathbf{0}^{(b)}\right).
\end{aligned}$$

The result then follows on noting that the unconditional variance of $(\mathbf{z} - \mathbf{U}\,\mathbf{0}^{(b)})$, based on a first order approximation, is $\boldsymbol{\Sigma}_e$.

•

In practice $\mathbf{0}^{(b)}$, the variance components and possibly $\phi$ are unknown and these will need to be replaced in the above by their estimates. For $\mathbf{0}^{(b)}$ we propose to use $\tilde{\mathbf{0}}$ defined in (6.20) whilst for estimates of the variance components and $\phi$ we propose estimates based on the variance components estimates given in Chapter 6, the derivations of which are now given.

### 7.3.1 (Modified) Fellner estimates.

In this section we obtain robust analogues of REML based estimates of the variance components and $\phi$ corresponding to the procedure given in Section 6.2 .

Consider $\mathbf{0}^{(b)}$ and $\mathbf{W}_e$, and hence $\mathbf{X}_W$ and $\mathbf{U}_W$, as known. In practice these will be replaced by the currently available estimates.

Let

(i)

$$\mathbf{T} = \left(\mathbf{I} + \phi^{-1}\,\mathbf{U}'\,\mathbf{W}_e\,\mathbf{U}\,\mathbf{D} - \phi^{-1}\,\mathbf{U}'\,\mathbf{W}_e\,\mathbf{X}\,(\mathbf{X}'\,\mathbf{W}_e\,\mathbf{X})^{-1}\,\mathbf{X}'\,\mathbf{W}_e\,\mathbf{U}\,\mathbf{D}\right)^{-1} \tag{7.24}$$

with $\mathbf{T}$ partitioned conformally with $\mathbf{D}$ as $\{{}_m\,\mathbf{T}_{ij}\}_{i,j=1}^{c}$,

(ii) $\mathbf{K}_i$ is the diagonal matrix with diagonal elements corresponding to $Var(\hat{\boldsymbol{\beta}}_i)$ $i = 1, \cdots, c$ as given by (7.22), where $\hat{\boldsymbol{\beta}}_i$ corresponds to the estimate of the realized value of $\mathbf{b}_i$ $(i = 1, \cdots, c)$. That is $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1', \cdots, \hat{\boldsymbol{\beta}}_c')'$

(iii) $\mathbf{L}_W$ equals the diagonal matrix with diagonal elements corresponding to $Var(\hat{\boldsymbol{\eta}}_W)$ where $\hat{\boldsymbol{\eta}}_W = \mathbf{W}_e^{1/2}\,\hat{\boldsymbol{\eta}}$. Hence $\mathbf{L}_W = \mathbf{W}_e^{1/2}\,\mathbf{L}\,\mathbf{W}_e^{1/2}$ where $\mathbf{L}$ equals the diagonal matrix with diagonal elements corresponding to $Var(\hat{\boldsymbol{\eta}})$ which is given by (7.23).

The modified Fellner robust estimates of the variance components and $\phi$ corresponding to (6.12) and (6.13) but for the model (7.20) satisfy

$$\hat{\sigma}_i^2 = \frac{\| \mathbf{K}_i^{1/2}\,\psi_k(\mathbf{K}_i^{-1/2}\,\hat{\boldsymbol{\beta}}_i) \|^2}{m_k\,(q_i - v_i)}\,, \quad i = 1, \cdots, c\,; \text{ and} \tag{7.25}$$

$$\begin{aligned} \hat{\phi} &= \frac{\|\mathbf{L}_W^{1/2}\psi_k\left(\mathbf{L}_W^{-1/2}\hat{\boldsymbol{\eta}}_W\right)\|^2}{m_k\,(n - p - \sum_{i=1}^{c}(q_i - v_i))}\,, \\ &= \frac{\|\mathbf{W}_e^{1/2}\mathbf{L}^{1/2}\psi_k\left(\mathbf{L}^{-1/2}\hat{\boldsymbol{\eta}}\right)\|^2}{m_k\,(n - p - \sum_{i=1}^{c}(q_i - v_i))}\,, \end{aligned} \tag{7.26}$$

where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are the solutions to (7.10) and

$$\mathbf{z}_W = \mathbf{W}_e^{1/2}\,\mathbf{z}\,, \tag{7.27}$$

$$\hat{\boldsymbol{\eta}}_W = \mathbf{z}_W - \mathbf{X}_W\,\hat{\boldsymbol{\alpha}} - \mathbf{U}_W\,\hat{\boldsymbol{\beta}}\,, \tag{7.28}$$

$$m_k = \int \psi_k^2(x)\,d\Phi(x), \text{ and} \tag{7.29}$$

$$v_i = Tr(\hat{\mathbf{T}}_{ii}), \tag{7.30}$$

where $\hat{\mathbf{T}}$ corresponds to $\mathbf{T}$ evaluated at the estimates of the variance components and $\psi_k$ is as defined Section 7.2.2.

To obtain solutions $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}_i^2$ $(i = 1, \cdots, c)$ and $\hat{\phi}$ we iterate between equations (7.18), (7.25) and (7.26), starting at initial estimates of the parameters, and updating our estimates of $\mathbf{0}^{(b)}$ and $\mathbf{W}_e$ as we proceed. Here we estimate $\mathbf{0}^{(b)}$ by $\tilde{\mathbf{0}}$ given in (6.20) whilst for $\mathbf{W}_e$ we have two choices. One choice is simply to estimate $\mathbf{W}_e$ using (7.6) evaluated at $\hat{\boldsymbol{\alpha}}$ and $\tilde{\mathbf{0}}$, that is $\mathbf{W}\left(h(\mathbf{X}\,\hat{\boldsymbol{\alpha}} + \mathbf{U}\,\tilde{\mathbf{0}})\right)$. Alternatively, for the case where $g(\cdot)$ is the canonical link, an expression for $\mathbf{W}_e$ based on (7.7) can be

used. This expression, in terms of $\alpha$, $\mathbf{0}^{(b)}$, the variance components and $\phi$, can be evaluated at the most recent estimates. Iteration is continued until the required convergence criteria are met. Where $\phi$ is known we exclude equation (7.26) from the iteration and use the actual value of $\phi$ in the others.

An important point to note here is that we evaluate $\mathbf{W}_e$ at the estimate of $\mathbf{0}^{(b)}$ and do not simply plug $\mathbf{0}$ into the estimate. This is because the variances of the random components in the model may vary significantly as a function of $\mathbf{0}^{(b)}$ in which case outliers in the elements of $\beta$ could significantly influence the estimates.

We now note two variations on the above procedure.

(i) If maximum likelihood estimates for the variance components are preferred to REML estimates we use exactly the same procedure as above but for the following two changes. First we replace $\mathbf{T}$ above by

$$\left(\mathbf{I} + \phi^{-1}\,\mathbf{U}'\,\mathbf{W}_e\,\mathbf{U}\,\mathbf{D}\right)^{-1}$$

and second, we replace $n - p$ by $n$ in the divisor on the right hand side of (7.26).

(ii) If we replace $\mathbf{K}_i$ in (7.25) by $\sigma_i^2\,\mathbf{I}_{q_i}$ , $i = 1, \cdots, c$ and $\mathbf{L}$ in (7.26) by $\phi\mathbf{W}_e^{-1}$ we obtain estimates of the variance components based on the extension of the Fellner method of Chapter 6 to the GLMM.

### 7.3.2 Method of moments estimates.

The method we now give corresponds to the method given in Section 6.3. Before outlining the method for the GLMM though we begin with the following result.

Let $\tilde{z}_i = g(\tilde{y}_i)$ where $\tilde{y}_i$ is defined in (7.15).

**Theorem 7.2**

Based on first order approximations we have, conditional on $\mu_i$,

$$\tilde{z}_i \;=\; g(\mu_i) + \tilde{\eta}_i \qquad (7.31)$$

where $E(\tilde{\eta}_i) = 0$ and $Var(\tilde{\eta}_i) = m_k\,\phi\,w_i(\mu_i)$ where $m_k$ is the unbiasing constant given in (7.29).

**Proof.**

Let $T_i = t(y_i)$ where $t(\cdot)$ is the normalizing transform associated with the random variable $y_i$. Then, conditional on $\mu_i$ the statistic $T_i$, based on a first order approximation, is approximately $N\left(t(\mu_i), \sigma_T^2\right)$ where $\sigma_T^2 = \phi\, a_i\, [\dot{t}(\mu_i)]^2\, v(\mu_i)$. Hence,

$$\tilde{T}_i \;=\; t(\mu_i) + \sigma_T\,\psi_k\left(\frac{T_i - t(\mu_i)}{\sigma_T}\right)$$

has, from the theory for a Winsorized normal random variable, mean $t(\mu_i)$ and variance $m_k\,\sigma_T^2$. Now, for $s(\cdot)$ corresponding to the inverse function of $t(\cdot)$, we have $\tilde{y}_i \approx s(\tilde{T}_i)$ giving

$$\tilde{Y}_i \;=\; \mu_i + \dot{s}\left(t(\mu_i)\right)\,\left(\tilde{T}_i - t(\mu_i)\right)$$

to first order. Therefore, to first order $\tilde{y}_i$ has mean $\mu_i$ and variance $m_k\,\phi\,a_i\,v(\mu_i)$, using the result that $\dot{s}\left(t(x)\right) = 1/\dot{t}(x)$. Equation (7.31) then follows from $\tilde{z}_i = g(\mu_i) + \dot{g}(\mu_i)\,(\tilde{y}_i - \mu_i)$, to first order, on letting $\tilde{\eta}_i = \dot{g}(\mu_i)\,(\tilde{Y}_i - \mu_i)$.

•

To estimate the variance components and $\phi$ using the method of moments we let

$$\begin{aligned} \tilde{\mathbf{z}}^* \;&=\; \tilde{\mathbf{z}} - \mathbf{U}\,\tilde{\mathbf{0}}\,, \quad \text{and} \\ \tilde{\mathbf{z}}^*_W \;&=\; \mathbf{W}_e^{1/2}\,\tilde{\mathbf{z}}^*\,. \end{aligned} \tag{7.32}$$

where $\tilde{\mathbf{0}}$ is as defined in (6.20) and base our inference on the model

$$\tilde{\mathbf{z}}^*_W \;=\; \mathbf{X}_W\,\boldsymbol{\alpha} + \mathbf{U}_W\,\mathbf{b}^* + \tilde{\boldsymbol{\eta}}_W \tag{7.33}$$

where $\mathbf{b}^* = \mathbf{b} - \tilde{\mathbf{0}}$. We then have approximately $\mathcal{E}(\mathbf{b}^*) = \mathbf{0}$, $Var(\mathbf{b}^*) = m_k\,\mathbf{D}$, $Var(\tilde{\boldsymbol{\eta}}_W) = m_k\,\phi\,\mathbf{I}$ and $Cov(\mathbf{b}^*, \tilde{\boldsymbol{\eta}}_W) = \mathbf{0}$.

Robust estimates of the variance components can now be obtained based on the mixed model in (7.33) using one of the available methods for estimating variance

components. In particular, robust moment methods based on Henderson's method 3 estimates of the variance components as given in Chapter 6, Section 6.3.2, can be obtained from (6.17) and (6.18) but with $\mathbf{y}^*$ replaced by $\tilde{\mathbf{z}}_W^*$, with $\mathbf{X}$ and $\mathbf{U}$ replaced by $\mathbf{X}_W$ and $\mathbf{U}_W$ respectively and finally $\sigma^2$ replaced by $\phi$. Also, if $\phi$ is known we omit equation (6.17) and evaluate (6.18) at the known $\phi$.

One iterative procedure to obtain the robust method of moments estimates is as follows. First, obtain a non-weighted least squares estimate of $\boldsymbol{\alpha}$ and set the initial estimates of $\boldsymbol{\beta}$ and $E(\mathbf{b})$ both to the zero vector. From these an initial estimate of $\mathbf{W}_e$ can be obtained and then, using the moment procedure outlined in this subsection, initial estimates of $\sigma_i^2$ $(i = 1, \cdots, c)$ and $\phi$ can be obtained. The procedure is now simply to iterate between equations (7.18) to obtain new estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ and the equations giving estimates of the variance components based on the model (7.33). At each stage we calculate $\tilde{\mathbf{0}}$ and $\tilde{\mathbf{z}}$ and upgrade $\mathbf{W}_\mu$ and $\mathbf{W}_e$. Iteration is continued until the convergence criteria are satisfied.

## 7.4 Simulations.

To obtain some appreciation of the relative performance of the robust estimators for the GLMM considered in this chapter a simple simulation, similar to that conducted in Chapter 6, was undertaken.

The basic uncontaminated model for the study corresponds to a "balanced" one-way log linear Poisson model containing one random effect in the linear model. Formally, letting $y_{ij}$ $(i = 1, \cdots, l\,;\ j = 1, \cdots, r)$ denote the data, we have that the $y_{ij}$, conditional on $\mu_{ij}$, are independent Poisson random variables with mean $\mu_{ij}$. That is,

$$y_{ij}|\mu_{ij} \ \sim \ P(\mu_{ij}) \ ;\ i = 1, \cdots, l\,;\ j = 1, \cdots, r\,. \tag{7.34}$$

We then let

$$\log(\mu_{ij}) \ = \ \alpha + b_i \tag{7.35}$$

where $\alpha = 2$ and the $b_i$ $(i = 1, \cdots, l)$ are independent Normal random variables with zero mean and variance $\sigma_1^2 = 0.25$. That is $b_i \sim N(0, 0.25)$ , $(i = 1, \cdots, l)$.

To have notation consistent with the notation of the previous sections we let

$$\begin{aligned} \mathbf{y} &= (y_{11}, \cdots, y_{1r}, y_{21}, \cdots, y_{2r}, \cdots \quad \cdots, y_{l1}, \cdots, y_{lr})' , \\ &= \left\{ {}_c \{ {}_c\, y_{ij} \}_{j=1}^{r} \right\}_{i=1}^{l} \end{aligned}$$

where the latter form is simply a nested application of the notation of Searle et. al. (1992) and follows from the brief description given in Appendix A. We will avoid referring specifically to the $i^{th}$ element of $\mathbf{y}$ so as not to confuse the situation by having $y$ values with both single and double subscripts. Similarly we will not refer specifically to the $i^{th}$ elements of $\mathbf{z} = \log(\mathbf{y})$ but instead let $z_{ij} = \log(y_{ij})$ or, if $y_{ij} = 0$, $z_{ij} = \log(0.5)$.

For the model in the form (7.1) and (7.2) we have $c = 1$ and

$$\mathbf{X} = \{ {}_c\, 1 \}_{i=1}^{lr} \quad ; \quad \mathbf{U} = \left\{ {}_d \{ {}_c\, 1 \}_{j=1}^{r} \right\}_{i=1}^{l} \quad \text{and} \quad \mathbf{b} = \{ {}_c\, b_i \}_{i=1}^{l} \, .$$

Further, from Section 7.1

$$\begin{aligned} \mathbf{V}(\boldsymbol{\mu}) &= \left\{ {}_d \{ {}_d\, \phi\, \mu_{ij} \}_{j=1}^{r} \right\}_{i=1}^{l} , \\ \mathbf{W}_{\mu} &= \left\{ {}_d \{ {}_d\, \mu_{ij} \}_{j=1}^{r} \right\}_{i=1}^{l} , \end{aligned}$$

where $\phi = 1$ but is retained in $\mathbf{V}(\boldsymbol{\mu})$ as later we will assume it is unknown. Also, from (7.7) and (7.8) we have, using the result $E(\mu_{ij}^{-1}) = e^{-\alpha} E(e^{-b_i})$, that

$$\mathbf{W}_e = \left\{ {}_d \{ {}_d\, e^{\alpha + E(b_i) - \sigma_1^2/2} \}_{j=1}^{r} \right\}_{i=1}^{l}$$

where $E(b_i)$ is retained so as later to accommodate outlying results in the $b_i$ $(i = 1, \cdots, l)$.

For all the simulations below $\psi_k(\cdot)$ is taken as the Huber-psi function with tuning constant $k = 2$, in which case $m_k = 0.921$.

To obtain expressions for estimates of the parameters we first need to obtain an expression for the pseudo variable defined in (7.15) based on the current estimate of $\hat{\mu}_{ij}$. Using the results from Section 7.2.3.1 we have

$$\tilde{y}_{ij} = \max\{a(\hat{\mu}_{ij}), \min\{y_{ij}, b(\hat{\mu}_{ij})\}\} \tag{7.36}$$

where

$$a(\hat{\mu}_{ij}) = \left[\left\{\left(\hat{\mu}_{ij}^{2/3} - \hat{\mu}_{ij}^{-1/3}/9 - 4\hat{\mu}_{ij}^{1/6}/3\right) \vee 0\right\}^{3/2} - 0.5\right] \vee 0$$

$$b(\hat{\mu}_{ij}) = \left]\left\{\left(\hat{\mu}_{ij}^{2/3} - \hat{\mu}_{ij}^{-1/3}/9 + 4\hat{\mu}_{ij}^{1/6}/3\right)\right\}^{3/2} + 0.5\right[ .$$

The pseudo adjusted dependent variable is $\tilde{z}_{ij} = \log(\max\{\tilde{y}_{ij}, 0.5\})$.

Letting $w_{ij}$ equal that diagonal element of $\mathbf{W}_\mu$ corresponding to the element $y_{ij}$ of $\mathbf{y}$, that is $w_{ij} = \mu_{ij} = e^{\alpha+b_i}$ which is independent of $j$, the solutions to (7.17) can be expressed as

$$\hat{\alpha} = \frac{\sum_{i=1}^{l} c_i \left(\frac{1}{r}\sum_{j=1}^{r} \tilde{z}_{ij} - \tilde{0}_i\right)}{\sum_{i=1}^{l} c_i} , \tag{7.37}$$

$$\hat{\beta}_i = \tilde{0}_i + c_i \left(\tfrac{1}{r}\textstyle\sum_{j=1}^{r} \tilde{z}_{ij} - \tilde{0}_i - \hat{\alpha}\right) \; ; \quad (i = 1, \cdots, l) , \tag{7.38}$$

where

$$c_i = \frac{r\, w_{i1}\, \sigma_1^2}{\phi + r\, w_{i1}\, \sigma_1^2} , \tag{7.39}$$

$$\tilde{0}_i = \hat{\beta}_i - \sigma_1 \max\left\{-2, \min\{\hat{\beta}_i/\sigma_1, 2\}\right\} \; ; \quad (i = 1, \cdots, l) . \tag{7.40}$$

If $\phi$ and $\sigma_1^2$ were known we would simply iterate through (7.36), (7.37) and (7.38), changing $w_{ij}$ at each stage so as to base it on the most recent estimate of $\mu_{ij}$, and iterate until we have the required convergence. In our simulations however we consider the more realistic situation, that being where at least $\sigma_1^2$ is not known.

To obtain the additional equations required for the modified Fellner estimates of $\sigma_1^2$ and $\phi$ we need approximate expressions firstly for the $Var(\hat{\beta}_i)$ and $Var(\hat{\eta}_{ij})$. Here the result can be obtained from Theorem 7.1. Alternatively, and more simply in this case, they can be obtained from the standard BLUP estimates given by (7.37) and (7.38) but with $\tilde{z}_{ij}$ and $\tilde{0}_i$ replaced by $z_{ij}$ and $E(b_i) = 0_i^{(b)}$ respectively, where the latter is assumed known. Note here that for this calculation we treat the $w_{ij}$ as constants which, in Theorem 7.1 have been replaced by the $\left(E(\mu_{ij}^{-1})\right)^{-1} =$

$e^{\alpha+0_i^{(b)}-\sigma_1^2/2} = w'_{ij}$, say. This then gives,

$$Var(\hat{\beta}_i) = c'_i\left(1-\frac{c'_i}{c'_.}\right)\sigma_1^2\ ;\ (i=1,\cdots,l)\ , \tag{7.41}$$

$$Var(\hat{\eta}_{ij}) = \left(1-\frac{c'_i}{r}\left(1+\frac{1-c'_i}{c'_.}\right)\right)\frac{\phi}{w'_{ij}}\ ;\ (i=1,\cdots,l\ ;\ j=1,\cdots,r) \tag{7.42}$$

where

$$c'_i = \frac{r\,w'_{i1}\,\sigma_1^2}{\phi + r\,w'_{i1}\,\sigma_1^2} \quad \text{and}$$

$$c'_. = \sum_{i=1}^{l} c'_i\ .$$

Two comments are worth making before continuing. First, the assumption that $w_{ij}^{-1} = \mu_{ij}^{-1}$ is slowing changing is not unreasonable, provided that the $\mu_{ij}$ are not too small and $\sigma_1^2$ is not too large. Second, we can see from the variance expressions above that the standard BLUP estimates are shrinkage estimates.

In addition to the above we also need $v_1$ as defined in (7.30), which is given by

$$v_1 = \sum_{i=1}^{l}(1-c'_i)\left\{1+\frac{c_{i'}\,w'_{i1}}{w'_{.1}}\left[1-\sum_{i'=1}^{l}\frac{c_{i'}\,w'_{i'1}}{w'_{.1}}\right]^{-1}\right\}$$

where $w'_{.1} = \sum_{i=1}^{l} w'_{i1}$.

To obtain modified Fellner estimates of $\sigma_1^2$ and $\phi$ in conjunction with the estimates of $\alpha$ and $\beta_i\,(i=1,\cdots,l)$ we include the following equations, corresponding to (7.26) and (7.25) and simplified using (7.14), into the iterative procedure:

$$\hat{\phi} = \frac{\sum_{i=1}^{l}\sum_{j=1}^{r}\left[\sqrt{\phi}\,\psi_{a_1k}\left(\hat{\eta}_{ij}\,\sqrt{w'_{ij}/\phi}\,\right)\right]^2}{m_k(lr-1-l+v_1)}\ , \quad \text{and} \tag{7.43}$$

$$\hat{\sigma_1^2} = \frac{\sum_{i=1}^{l}\left[\sigma_1\,\psi_{a_2k}(\hat{\beta}_i/\sigma_1)\right]^2}{m_k(l-v_1)}\ , \tag{7.44}$$

where

$$k = 2\ ,$$

$$a_1 = \sqrt{\left(1-\frac{c'_i}{r}\left(1+\frac{1-c'_i}{c'_.}\right)\right)}\ , \quad \text{and}$$

$$a_2 = \sqrt{c'_i\left(1-\frac{c'_i}{c'_.}\right)}\ .$$

Here the right hand sides of (7.43) and (7.44) are evaluated at the most recent estimates of $\sigma_1^2$ and $\phi$. The robust estimation procedure then requires that one iterates through (7.36), (7.37), (7.38), (7.43) and (7.44) until convergence. For our simulations the procedure is concluded to have converged if the maximum absolute difference beween subsequent estimates of $\alpha$, $\beta_i$, $(i = 1, \cdots, l)$, $\sigma_1^2$ and $\phi$ is less than 0.001 .

For Fellner based estimates, that is ignoring the shrinkage of the BLUP estimates, the procedure is as above except that we set $a_1 = 1$ and $a_2 = 1$ in (7.43) and (7.44) respectively.

Using moment based method as developed in Section 7.3.2 in place of the (modified) Fellner method, we replace (7.43) and (7.44) by

$$\hat{\phi} = \frac{\sum_{i=1}^{l}\sum_{j=1}^{r} w'_{ij}\left(\tilde{z}_{ij} - \frac{1}{r}\sum_{j=1}^{r}\tilde{z}_{ij}\right)^2}{m_k\, l\,(r-1)}\,, \tag{7.45}$$

$$\hat{\sigma_1^2} = \frac{\sum_{i=1}^{l} r w'_{i1}\left(\frac{1}{r}\sum_{j=1}^{r}(\tilde{z}_{ij} - \tilde{0}_i) - \sum_{i'=1}^{l}\frac{w'_{i'1}}{w'_{\cdot 1}}\frac{1}{r}\sum_{j=1}^{r}(\tilde{z}_{i'j} - \tilde{0}_i)\right)^2 - m_k\,(l-1)\hat{\phi}}{m_k\, r\,(w'_{\cdot 1} - \sum_{i=1}^{l} w'^2_{i1}/w'_{\cdot 1})}\,. \tag{7.46}$$

For the situation where $\phi$ is known, that is $\phi = 1$, equation (7.43), or equation (7.45), is left out of the iterative process.

The forms of contamination we consider in the simulations reported below are as follows:

Distribution of the $b_i$ (i.e. Systematic Error)

| | |
|---|---|
| $\mathcal{S}_1$ | $b_i \sim N(0, 0.25)$ i.e. no contamination |
| $\mathcal{S}_2$ | $b_i \sim N(0, 0.25)$ *w.p.* 0.9, $b_i \sim N(0, 1.0)$ *w.p.* 0.1 |
| $\mathcal{S}_3$ | $b_i \sim N(0, 0.25)$ *w.p.* 0.95, $b_i \sim N(0, 2.25)$ *w.p.* 0.05 |
| $\mathcal{S}_4$ | $b_i \sim N(0, 0.25)$ *w.p.* 0.95, $b_i = (-3, 3)$ *w.p.* (0.025, 0.025) resp. |
| $\mathcal{S}_5$ | $b_i \sim N(0, 0.25)$ *w.p.* 0.90, $b_i = (-4, -3, 3, 4)$ *w.p.* (.025, .025, .025, .025) resp. |

*w.p. denotes "with probability".*

Here $\mathcal{S}_4$ and $\mathcal{S}_5$ represent extreme forms of contamination whilst $\mathcal{S}_2$ and $\mathcal{S}_3$ represent more mild forms.

To introduce contamination at the level of the individual result (i.e. Random Error) the parameter $\alpha$, which equals 2 in the uncontaminated model, was allowed to change at random for each result according to the following regime:

Random Error contamination

| | |
|---|---|
| $\mathcal{R}_1$ | $\alpha = 2$ *w.p.* 1. i.e. no contamination |
| $\mathcal{R}_2$ | $\alpha = (2,3)$ *w.p.* $(0.9, 0.1)$ resp. |
| $\mathcal{R}_3$ | $\alpha = (2,3,4)$ *w.p.* $(0.9, 0.05, 0.05)$ resp. |

The estimators considered in the simulations are:

RM0 - Non robust procedure (i.e. procedure of Schall (1991))

RM1 - Fellner procedure

RM2 - Modified Fellner procedure

RM3 - Moment based method

For the simulations various combinations of the above contamination models were considered. For each combination of random and systematic error contamination 500 data sets were generated according to the model given in (7.34) and (7.35) and the parameters were estimated using each of the four methods above. The starting values in each case for $\phi$ and $\sigma_1^2$ were obtained using the non robust version of the moment method above, that is Henderson's method 3. As for the simulations in Chapter 6, the iterative procedure for each estimator was considered to have converged when the maximum absolute difference between subsequent estimates of $\alpha$, $\phi$, $\sigma_1^2$ and $\beta_i$ $(i = 1, \cdots, l)$ was less than 0.001 . In Tables 7.2 and 7.4 are given the means and standard deviations (the latter in brackets) of $n_{it}$, the number of iterations to achieve convergence, $\hat{\alpha}$, $\hat{\phi}$ and $\hat{\sigma_1^2}$ for each of the four estimators over the 500 data sets. For the results in Table 7.2, $l = 10$ and $r = 2$ whilst in Table 7.4, $l = 10$ and $r = 10$. The results in Tables 7.3 and 7.5 correspond to those in Tables 7.2 and 7.4 respectively except that in this case $\phi$ was treated as known and equal to 1.

**Table 7.2: Robust parameter estimates; $l = 10, r = 2$; $\phi$ estimated**

| Random | Systematic | Estimate | $n_{it}$ | $\alpha$ | $\phi$ | $\sigma_1^2$ |
|---|---|---|---|---|---|---|
| $\mathcal{R}_1$ | $\mathcal{S}_1$ | RM0 | 6.42 (3.39) | 1.98 (0.20) | 1.14 (0.55) | 0.27 (0.17) |
| | | RM1 | 6.90 (2.77) | 1.98 (0.20) | 1.16 (0.56) | 0.31 (0.19) |
| | | RM2 | 10.5 (11.6) | 1.97 (0.20) | 0.92 (0.26) | 0.31 (0.19) |
| | | RM3 | 5.10 (1.72) | 1.98 (0.20) | 1.26 (0.72) | 0.31 (0.20) |
| $\mathcal{R}_1$ | $\mathcal{S}_2$ | RM0 | 6.38 (3.90) | 1.96 (0.20) | 1.18 (0.63) | 0.33 (0.23) |
| | | RM1 | 7.13 (3.01) | 1.96 (0.20) | 1.18 (0.60) | 0.37 (0.24) |
| | | RM2 | 9.81 (6.86) | 1.96 (0.20) | 0.92 (0.27) | 0.37 (0.23) |
| | | RM3 | 5.23 (1.67) | 1.96 (0.21) | 1.26 (0.77) | 0.38 (0.27) |
| $\mathcal{R}_1$ | $\mathcal{S}_3$ | RM0 | 6.43 (4.55) | 1.96 (0.20) | 1.13 (0.56) | 0.37 (0.30) |
| | | RM1 | 7.14 (3.40) | 1.95 (0.20) | 1.15 (0.56) | 0.39 (0.27) |
| | | RM2 | 10.9 (12.5) | 1.95 (0.20) | 0.91 (0.26) | 0.39 (0.26) |
| | | RM3 | 5.55 (1.86) | 1.95 (0.20) | 1.21 (0.72) | 0.43 (0.35) |
| $\mathcal{R}_1$ | $\mathcal{S}_4$ | RM0 | 6.09 (3.14) | 2.00 (0.26) | 1.12 (0.51) | 0.65 (0.63) |
| | | RM1 | 8.48 (4.32) | 1.98 (0.25) | 1.15 (0.63) | 0.61 (0.63) |
| | | RM2 | 11.3 (7.88) | 1.97 (0.25) | 0.90 (0.28) | 0.57 (0.59) |
| | | RM3 | 5.65 (2.01) | 1.98 (0.26) | 1.07 (0.68) | 0.75 (0.70) |
| $\mathcal{R}_1$ | $\mathcal{S}_5$ | RM0 | 5.76 (2.46) | 2.02 (0.34) | 1.09 (0.55) | 1.11 (1.00) |
| | | RM1 | 9.41 (5.37) | 1.99 (0.33) | 1.09 (0.72) | 0.99 (1.02) |
| | | RM2 | 11.2 (8.26) | 1.98 (0.33) | 0.84 (0.32) | 0.93 (0.98) |
| | | RM3 | 6.04 (2.21) | 1.99 (0.34) | 0.85 (0.67) | 1.30 (1.12) |
| $\mathcal{R}_2$ | $\mathcal{S}_1$ | RM0 | 7.48 (4.93) | 2.03 (0.19) | 1.84 (1.01) | 0.26 (0.16) |
| | | RM1 | 8.44 (5.43) | 2.02 (0.19) | 1.66 (0.85) | 0.30 (0.18) |
| | | RM2 | 12.1 (8.31) | 2.01 (0.18) | 1.08 (0.26) | 0.32 (0.18) |
| | | RM3 | 5.38 (1.67) | 2.02 (0.19) | 1.75 (1.02) | 0.30 (0.20) |
| $\mathcal{R}_2$ | $\mathcal{S}_2$ | RM0 | 6.84 (6.57) | 2.03 (0.21) | 1.77 (1.05) | 0.34 (0.23) |
| | | RM1 | 7.91 (4.68) | 2.02 (0.21) | 1.60 (0.89) | 0.38 (0.25) |
| | | RM2 | 12.6 (18.4) | 2.01 (0.21) | 1.05 (0.26) | 0.39 (0.25) |
| | | RM3 | 5.78 (1.86) | 2.02 (0.21) | 1.56 (0.98) | 0.40 (0.27) |
| $\mathcal{R}_2$ | $\mathcal{S}_3$ | RM0 | 7.38 (7.48) | 2.04 (0.23) | 1.90 (1.24) | 0.36 (0.31) |
| | | RM1 | 8.32 (4.40) | 2.02 (0.22) | 1.64 (0.89) | 0.38 (0.26) |
| | | RM2 | 12.1 (9.37) | 2.01 (0.22) | 1.06 (0.26) | 0.39 (0.24) |
| | | RM3 | 6.02 (6.25) | 2.02 (0.22) | 1.62 (1.05) | 0.41 (0.34) |

*Table 7.2 continued on next page*

**Table 7.2 (Continued)**

| Random | Systematic | Estimate | $n_{it}$ | $\alpha$ | $\phi$ | $\sigma_1^2$ |
|---|---|---|---|---|---|---|
| $\mathcal{R}_2$ | $\mathcal{S}_4$ | RM0 | 6.91 (3.47) | 2.05 (0.25) | 2.10 (2.07) | 0.59 (0.55) |
| | | RM1 | 9.10 (4.48) | 2.02 (0.23) | 1.66 (0.94) | 0.57 (0.53) |
| | | RM2 | 12.6 (8.96) | 2.00 (0.23) | 1.06 (0.27) | 0.55 (0.49) |
| | | RM3 | 7.06 (5.26) | 2.01 (0.24) | 1.40 (0.88) | 0.69 (0.59) |
| $\mathcal{R}_2$ | $\mathcal{S}_5$ | RM0 | 6.96 (3.76) | 2.10 (0.36) | 2.66 (4.77) | 1.06 (1.01) |
| | | RM1 | 11.7 (12.0) | 2.04 (0.32) | 1.49 (0.96) | 0.96 (1.03) |
| | | RM2 | 14.2 (13.7) | 2.02 (0.32) | 0.99 (0.34) | 0.90 (0.99) |
| | | RM3 | 9.45 (20.8) | 2.04 (0.34) | 1.10 (0.94) | 1.26 (1.12) |
| $\mathcal{R}_3$ | $\mathcal{S}_1$ | RM0 | 8.88 (6.05) | 2.18 (0.24) | 4.80 (3.34) | 0.28 (0.24) |
| | | RM1 | 13.9 (12.6) | 2.11 (0.22) | 3.35 (2.33) | 0.29 (0.26) |
| | | RM2 | 17.4 (13.1) | 2.09 (0.21) | 1.25 (0.26) | 0.35 (0.23) |
| | | RM3 | 6.53 (2.20) | 2.11 (0.22) | 2.46 (1.58) | 0.32 (0.27) |
| $\mathcal{R}_3$ | $\mathcal{S}_2$ | RM0 | 8.81 (6.28) | 2.16 (0.25) | 4.78 (3.34) | 0.34 (0.26) |
| | | RM1 | 12.4 (11.7) | 2.09 (0.23) | 3.24 (2.35) | 0.36 (0.29) |
| | | RM2 | 19.0 (33.4) | 2.06 (0.22) | 1.23 (0.27) | 0.41 (0.25) |
| | | RM3 | 6.77 (2.21) | 2.08 (0.23) | 2.43 (1.63) | 0.39 (0.30) |
| $\mathcal{R}_3$ | $\mathcal{S}_3$ | RM0 | 8.19 (5.34) | 2.16 (0.25) | 4.82 (3.99) | 0.35 (0.33) |
| | | RM1 | 12.5 (10.5) | 2.10 (0.23) | 3.20 (2.21) | 0.36 (0.32) |
| | | RM2 | 17.2 (16.3) | 2.07 (0.22) | 1.24 (0.26) | 0.41 (0.28) |
| | | RM3 | 6.78 (2.24) | 2.09 (0.23) | 2.36 (1.78) | 0.41 (0.36) |
| $\mathcal{R}_3$ | $\mathcal{S}_4$ | RM0 | 8.43 (4.99) | 2.22 (0.31) | 5.84 (8.62) | 0.60 (0.61) |
| | | RM1 | 12.8 (10.8) | 2.14 (0.28) | 3.04 (2.31) | 0.57 (0.59) |
| | | RM2 | 18.8 (31.2) | 2.10 (0.26) | 1.20 (0.28) | 0.56 (0.52) |
| | | RM3 | 8.72 (14.5) | 2.12 (0.28) | 2.08 (1.65) | 0.68 (0.63) |
| $\mathcal{R}_3$ | $\mathcal{S}_5$ | RM0 | 9.12 (5.23) | 2.25 (0.40) | 9.32 (22.7) | 1.13 (1.11) |
| | | RM1 | 14.6 (16.9) | 2.14 (0.34) | 2.85 (2.26) | 1.06 (1.10) |
| | | RM2 | 20.3 (22.1) | 2.09 (0.34) | 1.16 (0.33) | 1.01 (1.04) |
| | | RM3 | 12.4 (26.0) | 2.11 (0.35) | 1.61 (1.48) | 1.35 (1.17) |

**Table 7.3: Robust parameter estimates; $l = 10, r = 2$; $\phi$ set equal 1**

| Random | Systematic | Estimate | $n_{it}$ | $\alpha$ | $\sigma_1^2$ |
|---|---|---|---|---|---|
| $\mathcal{R}_1$ | $\mathcal{S}_1$ | RM0 | 4.45 (1.01) | 1.97 (0.19) | 0.28 (0.16) |
| | | RM1 | 4.79 (1.25) | 1.97 (0.19) | 0.31 (0.18) |
| | | RM2 | 5.67 (3.22) | 1.97 (0.19) | 0.30 (0.18) |
| | | RM3 | 5.79 (1.45) | 1.96 (0.19) | 0.33 (0.20) |
| $\mathcal{R}_1$ | $\mathcal{S}_2$ | RM0 | 4.52 (1.10) | 1.95 (0.20) | 0.36 (0.23) |
| | | RM1 | 5.00 (1.81) | 1.95 (0.20) | 0.40 (0.25) |
| | | RM2 | 5.79 (2.73) | 1.95 (0.21) | 0.38 (0.25) |
| | | RM3 | 6.09 (1.79) | 1.95 (0.20) | 0.43 (0.27) |
| $\mathcal{R}_1$ | $\mathcal{S}_3$ | RM0 | 4.56 (1.07) | 1.95 (0.20) | 0.37 (0.26) |
| | | RM1 | 5.19 (1.99) | 1.94 (0.20) | 0.40 (0.24) |
| | | RM2 | 5.89 (2.78) | 1.95 (0.20) | 0.37 (0.23) |
| | | RM3 | 6.06 (1.88) | 1.94 (0.20) | 0.44 (0.31) |
| $\mathcal{R}_1$ | $\mathcal{S}_4$ | RM0 | 4.77 (1.34) | 1.98 (0.26) | 0.60 (0.54) |
| | | RM1 | 6.56 (3.07) | 1.96 (0.25) | 0.56 (0.50) |
| | | RM2 | 7.28 (3.49) | 1.96 (0.25) | 0.51 (0.46) |
| | | RM3 | 6.62 (2.71) | 1.97 (0.26) | 0.70 (0.57) |
| $\mathcal{R}_1$ | $\mathcal{S}_5$ | RM0 | 5.10 (1.48) | 2.04 (0.33) | 1.14 (0.95) |
| | | RM1 | 7.79 (3.69) | 1.99 (0.31) | 0.99 (0.97) |
| | | RM2 | 8.71 (4.85) | 1.99 (0.31) | 0.89 (0.92) |
| | | RM3 | 7.27 (3.00) | 2.01 (0.32) | 1.31 (1.01) |
| $\mathcal{R}_2$ | $\mathcal{S}_1$ | RM0 | 4.48 (1.17) | 2.02 (0.21) | 0.31 (0.18) |
| | | RM1 | 4.94 (1.48) | 2.01 (0.21) | 0.34 (0.19) |
| | | RM2 | 5.74 (2.63) | 2.01 (0.21) | 0.32 (0.19) |
| | | RM3 | 5.69 (1.40) | 2.01 (0.21) | 0.35 (0.21) |
| $\mathcal{R}_2$ | $\mathcal{S}_2$ | RM0 | 4.56 (1.10) | 2.01 (0.22) | 0.37 (0.22) |
| | | RM1 | 5.15 (1.83) | 1.99 (0.22) | 0.40 (0.23) |
| | | RM2 | 5.86 (2.90) | 1.99 (0.22) | 0.38 (0.23) |
| | | RM3 | 5.92 (1.52) | 1.99 (0.22) | 0.42 (0.25) |
| $\mathcal{R}_2$ | $\mathcal{S}_3$ | RM0 | 4.47 (0.98) | 2.03 (0.23) | 0.38 (0.29) |
| | | RM1 | 5.30 (1.86) | 2.01 (0.22) | 0.39 (0.27) |
| | | RM2 | 5.86 (2.46) | 2.02 (0.22) | 0.38 (0.26) |
| | | RM3 | 5.95 (1.80) | 2.01 (0.22) | 0.44 (0.34) |

*Table 7.3 continued on next page*

**Table 7.3 (Continued)**

| Random | Systematic | Estimate | $n_{it}$ | $\alpha$ | $\sigma_1^2$ |
|---|---|---|---|---|---|
| $\mathcal{R}_2$ | $\mathcal{S}_4$ | RM0 | 4.77 (1.37) | 2.04 (0.27) | 0.66 (0.54) |
| | | RM1 | 6.95 (4.12) | 2.01 (0.25) | 0.61 (0.51) |
| | | RM2 | 7.80 (4.84) | 2.01 (0.25) | 0.56 (0.49) |
| | | RM3 | 7.18 (3.98) | 2.02 (0.26) | 0.76 (0.57) |
| $\mathcal{R}_2$ | $\mathcal{S}_5$ | RM0 | 4.96 (1.42) | 2.08 (0.34) | 1.14 (1.01) |
| | | RM1 | 9.66 (12.5) | 2.04 (0.31) | 0.98 (1.03) |
| | | RM2 | 10.4 (12.5) | 2.03 (0.31) | 0.90 (1.00) |
| | | RM3 | 9.31 (12.5) | 2.05 (0.33) | 1.30 (1.05) |
| $\mathcal{R}_3$ | $\mathcal{S}_1$ | RM0 | 4.27 (0.92) | 2.12 (0.21) | 0.41 (0.24) |
| | | RM1 | 6.70 (2.72) | 2.07 (0.19) | 0.36 (0.21) |
| | | RM2 | 7.43 (3.09) | 2.07 (0.19) | 0.35 (0.21) |
| | | RM3 | 6.77 (2.14) | 2.07 (0.19) | 0.38 (0.24) |
| $\mathcal{R}_3$ | $\mathcal{S}_2$ | RM0 | 4.33 (0.91) | 2.12 (0.23) | 0.46 (0.28) |
| | | RM1 | 6.85 (3.31) | 2.06 (0.22) | 0.42 (0.26) |
| | | RM2 | 7.68 (4.31) | 2.06 (0.22) | 0.40 (0.25) |
| | | RM3 | 6.99 (2.40) | 2.06 (0.22) | 0.44 (0.29) |
| $\mathcal{R}_3$ | $\mathcal{S}_3$ | RM0 | 4.40 (0.91) | 2.10 (0.23) | 0.47 (0.33) |
| | | RM1 | 6.73 (2.64) | 2.04 (0.21) | 0.43 (0.29) |
| | | RM2 | 7.62 (3.63) | 2.04 (0.21) | 0.41 (0.28) |
| | | RM3 | 7.04 (2.16) | 2.04 (0.21) | 0.48 (0.37) |
| $\mathcal{R}_3$ | $\mathcal{S}_4$ | RM0 | 4.58 (1.23) | 2.14 (0.27) | 0.73 (0.56) |
| | | RM1 | 10.3 (12.1) | 2.06 (0.25) | 0.60 (0.50) |
| | | RM2 | 11.0 (11.9) | 2.06 (0.25) | 0.55 (0.47) |
| | | RM3 | 10.4 (14.0) | 2.07 (0.25) | 0.72 (0.56) |
| $\mathcal{R}_3$ | $\mathcal{S}_5$ | RM0 | 5.00 (1.55) | 2.15 (0.35) | 1.22 (0.98) |
| | | RM1 | 15.9 (33.6) | 2.07 (0.32) | 1.01 (0.98) |
| | | RM2 | 15.8 (30.1) | 2.07 (0.31) | 0.91 (0.93) |
| | | RM3 | 15.9 (37.2) | 2.08 (0.33) | 1.30 (1.00) |

**Table 7.4: Robust parameter estimates; $l = 10, r = 10$; $\phi$ estimated**

| Random | Systematic | Estimate | $n_{it}$ | $\alpha$ | $\phi$ | $\sigma_1^2$ |
|---|---|---|---|---|---|---|
| $\mathcal{R}_1$ | $\mathcal{S}_1$ | RM0 | 3.34 (0.48) | 1.93 (0.18) | 1.16 (0.21) | 0.29 (0.14) |
| | | RM1 | 5.86 (1.12) | 1.93 (0.18) | 0.99 (0.17) | 0.31 (0.15) |
| | | RM2 | 6.61 (1.95) | 1.93 (0.18) | 0.97 (0.10) | 0.30 (0.15) |
| | | RM3 | 3.62 (1.22) | 1.93 (0.18) | 1.25 (0.24) | 0.31 (0.15) |
| $\mathcal{R}_1$ | $\mathcal{S}_4$ | RM0 | 3.46 (0.53) | 1.96 (0.26) | 1.12 (0.22) | 0.66 (0.53) |
| | | RM1 | 7.37 (2.36) | 1.94 (0.25) | 0.87 (0.25) | 0.58 (0.50) |
| | | RM2 | 8.14 (4.26) | 1.94 (0.24) | 0.88 (0.17) | 0.53 (0.48) |
| | | RM3 | 4.64 (2.27) | 1.95 (0.25) | 1.04 (0.36) | 0.70 (0.60) |
| $\mathcal{R}_1$ | $\mathcal{S}_5$ | RM0 | 3.52 (0.55) | 1.98 (0.34) | 1.07 (0.21) | 1.19 (1.00) |
| | | RM1 | 8.21 (2.76) | 1.94 (0.32) | 0.72 (0.29) | 0.98 (0.98) |
| | | RM2 | 8.87 (5.68) | 1.94 (0.31) | 0.78 (0.23) | 0.90 (0.95) |
| | | RM3 | 5.31 (2.76) | 1.97 (0.33) | 0.80 (0.41) | 1.31 (1.19) |
| $\mathcal{R}_2$ | $\mathcal{S}_1$ | RM0 | 3.40 (0.57) | 1.98 (0.16) | 1.64 (0.33) | 0.27 (0.14) |
| | | RM1 | 6.60 (1.21) | 1.96 (0.16) | 1.31 (0.27) | 0.29 (0.15) |
| | | RM2 | 9.16 (2.75) | 1.96 (0.16) | 1.10 (0.10) | 0.29 (0.15) |
| | | RM3 | 4.19 (1.02) | 1.96 (0.16) | 1.47 (0.32) | 0.30 (0.15) |
| $\mathcal{R}_2$ | $\mathcal{S}_4$ | RM0 | 3.59 (0.69) | 2.01 (0.24) | 1.83 (0.79) | 0.63 (0.53) |
| | | RM1 | 8.07 (2.44) | 1.97 (0.22) | 1.19 (0.35) | 0.52 (0.48) |
| | | RM2 | 9.70 (3.93) | 1.97 (0.22) | 1.04 (0.16) | 0.48 (0.44) |
| | | RM3 | 5.17 (2.75) | 1.99 (0.23) | 1.25 (0.43) | 0.66 (0.59) |
| $\mathcal{R}_2$ | $\mathcal{S}_5$ | RM0 | 3.85 (0.96) | 2.05 (0.36) | 2.28 (1.91) | 1.19 (1.00) |
| | | RM1 | 8.83 (2.71) | 1.99 (0.34) | 0.99 (0.45) | 1.00 (1.04) |
| | | RM2 | 10.0 (5.33) | 1.99 (0.33) | 0.93 (0.26) | 0.93 (1.02) |
| | | RM3 | 5.84 (2.98) | 2.02 (0.34) | 0.95 (0.51) | 1.31 (1.16) |
| $\mathcal{R}_3$ | $\mathcal{S}_1$ | RM0 | 4.01 (0.57) | 2.11 (0.17) | 3.83 (1.19) | 0.28 (0.15) |
| | | RM1 | 9.98 (2.22) | 2.03 (0.17) | 2.01 (0.63) | 0.29 (0.16) |
| | | RM2 | 13.2 (3.08) | 2.01 (0.17) | 1.21 (0.10) | 0.30 (0.16) |
| | | RM3 | 5.74 (1.21) | 2.03 (0.17) | 1.83 (0.45) | 0.30 (0.16) |
| $\mathcal{R}_3$ | $\mathcal{S}_4$ | RM0 | 4.41 (1.10) | 2.15 (0.27) | 5.04 (3.76) | 0.64 (0.55) |
| | | RM1 | 10.8 (2.71) | 2.04 (0.25) | 1.83 (0.74) | 0.55 (0.53) |
| | | RM2 | 13.0 (4.12) | 2.03 (0.25) | 1.17 (0.14) | 0.52 (0.50) |
| | | RM3 | 6.51 (2.55) | 2.05 (0.26) | 1.53 (0.57) | 0.68 (0.60) |
| $\mathcal{R}_3$ | $\mathcal{S}_5$ | RM0 | 5.18 (2.50) | 2.19 (0.38) | 8.38 (10.9) | 1.16 (0.97) |
| | | RM1 | 11.5 (2.85) | 2.04 (0.34) | 1.55 (0.75) | 0.95 (0.96) |
| | | RM2 | 12.7 (4.89) | 2.03 (0.33) | 1.09 (0.20) | 0.88 (0.93) |
| | | RM3 | 7.21 (4.00) | 2.06 (0.35) | 1.17 (0.66) | 1.27 (1.12) |

**Table 7.5: Robust parameter estimates; $l = 10, r = 10$; $\phi$ set equal 1**

| Random | Systematic | Estimate | $n_{it}$ | $\alpha$ | $\sigma_1^2$ |
|---|---|---|---|---|---|
| $\mathcal{R}_1$ | $\mathcal{S}_1$ | RM0 | 2.90 (0.34) | 1.93 (0.18) | 0.29 (0.14) |
| | | RM1 | 3.39 (1.11) | 1.93 (0.18) | 0.31 (0.16) |
| | | RM2 | 3.73 (1.61) | 1.93 (0.18) | 0.30 (0.16) |
| | | RM3 | 3.57 (1.08) | 1.93 (0.18) | 0.31 (0.16) |
| $\mathcal{R}_1$ | $\mathcal{S}_4$ | RM0 | 3.00 (0.62) | 1.95 (0.25) | 0.62 (0.55) |
| | | RM1 | 5.30 (3.22) | 1.94 (0.23) | 0.54 (0.51) |
| | | RM2 | 5.73 (4.16) | 1.94 (0.23) | 0.50 (0.48) |
| | | RM3 | 4.44 (2.62) | 1.95 (0.24) | 0.66 (0.61) |
| $\mathcal{R}_1$ | $\mathcal{S}_5$ | RM0 | 3.15 (0.72) | 1.96 (0.33) | 1.17 (1.02) |
| | | RM1 | 6.44 (3.73) | 1.93 (0.31) | 0.99 (1.03) |
| | | RM2 | 7.18 (5.46) | 1.93 (0.30) | 0.92 (1.00) |
| | | RM3 | 5.09 (3.45) | 1.96 (0.32) | 1.28 (1.17) |
| $\mathcal{R}_2$ | $\mathcal{S}_1$ | RM0 | 2.90 (0.32) | 1.98 (0.18) | 0.28 (0.13) |
| | | RM1 | 3.53 (1.24) | 1.97 (0.18) | 0.29 (0.14) |
| | | RM2 | 3.84 (1.84) | 1.97 (0.18) | 0.28 (0.14) |
| | | RM3 | 3.68 (1.35) | 1.97 (0.18) | 0.30 (0.14) |
| $\mathcal{R}_2$ | $\mathcal{S}_4$ | RM0 | 2.97 (0.62) | 2.00 (0.25) | 0.61 (0.53) |
| | | RM1 | 5.52 (3.29) | 1.97 (0.23) | 0.51 (0.47) |
| | | RM2 | 6.03 (4.67) | 1.97 (0.23) | 0.47 (0.43) |
| | | RM3 | 4.62 (2.62) | 1.98 (0.24) | 0.63 (0.57) |
| $\mathcal{R}_2$ | $\mathcal{S}_5$ | RM0 | 3.07 (0.75) | 2.05 (0.33) | 1.22 (1.01) |
| | | RM1 | 7.38 (5.07) | 1.99 (0.30) | 0.94 (0.99) |
| | | RM2 | 7.99 (5.80) | 1.99 (0.30) | 0.87 (0.96) |
| | | RM3 | 5.36 (3.11) | 2.03 (0.31) | 1.29 (1.15) |
| $\mathcal{R}_3$ | $\mathcal{S}_1$ | RM0 | 2.81 (0.40) | 2.09 (0.18) | 0.32 (0.17) |
| | | RM1 | 4.44 (1.37) | 2.01 (0.18) | 0.31 (0.16) |
| | | RM2 | 4.79 (1.70) | 2.01 (0.18) | 0.30 (0.16) |
| | | RM3 | 4.40 (1.42) | 2.01 (0.18) | 0.32 (0.16) |
| $\mathcal{R}_3$ | $\mathcal{S}_4$ | RM0 | 2.96 (0.67) | 2.11 (0.27) | 0.69 (0.58) |
| | | RM1 | 6.38 (3.22) | 2.02 (0.24) | 0.57 (0.55) |
| | | RM2 | 6.81 (4.07) | 2.02 (0.24) | 0.53 (0.52) |
| | | RM3 | 5.45 (3.19) | 2.03 (0.25) | 0.68 (0.63) |
| $\mathcal{R}_3$ | $\mathcal{S}_5$ | RM0 | 3.08 (0.77) | 2.13 (0.33) | 1.18 (0.96) |
| | | RM1 | 7.53 (3.79) | 2.01 (0.29) | 0.89 (0.89) |
| | | RM2 | 8.43 (6.10) | 2.01 (0.29) | 0.82 (0.85) |
| | | RM3 | 5.89 (3.39) | 2.05 (0.31) | 1.22 (1.07) |

The results of the simulations are summarized below:

1. When there is no contamination in the model all four estimators are approximately equivalent w.r.t. the estimation of $\alpha$. For estimating $\phi$ in this case RM2 appears to exhibit a negative bias whilst RM0 and RM3 exhibit positive biases, with the magnitude of the bias smallest for RM2. The standard error of the RM2 estimate of $\phi$ is also less than each of the other three estimators. To estimate $\sigma_1^2$ when there is no contamination it would appear that RM0 is the preferred estimator as it has the smallest positive bias. The bias for each of the other three estimators though is not large and is smallest for RM2.

2. The moment method (RM3) breaks down with respect to the estimation of $\sigma_1^2$ whether $\phi$ is estimated or known. This can be seen in that RM3 gives estimates of $\sigma_1^2$ with positive bias generally larger than even the non robust method RM0. Exactly why this is the case is not clear at this time but it may be related to the starting values for $\phi$ and $\sigma_1^2$ (see further comments below) and to the "unbalance" introduced by contamination in the systematic components. In a simulation study by Swallow and Monahan (1984), moment based methods for estimating variance components were concluded to be adequate except in certain cases when the data are severely unbalanced.

   RM3 appears to have merit over RM0 w.r.t. estimation of $\alpha$ and/or $\phi$ under some forms of contamination but in view if its performance in estimating $\sigma_1^2$ it would appear to be unsuitable in the GLMM case, at least with non robust starting values for the variance components.

3. For RM1 and RM2 it would appear that for estimating $\alpha$ there is not a big difference between the two procedures based on bias and standard error but when there is a difference, RM2 is generally the preferred estimator. For estimating $\phi$ on the otherhand there is a significant difference. From the simulations we see that based on the magnitude of the bias, RM2 is generally preferred to RM1, giving major gains particularly when there is contamination in the random component. For those few incidences when RM1 is better than

RM2 for estimating $\phi$, the difference is not large or else the worse performance of RM2 for estimating $\phi$ is compensated for by its superior performance w.r.t. estimation of $\sigma_1^2$.

Considering the estimation of $\sigma_1^2$ using RM1 and RM2 based on the simulations conducted, the conclusions drawn from the summary statistics in Tables 7.2 to 7.5 depend on whether $\phi$ is known or not known and on the form of contamination. Beginning with the case $\phi$ known to be 1 we see that RM2 is always preferred to RM1. This is also the case if $\phi$ is estimated from the data and the systematic contamination is heavy (i.e. the distribution of $b_i$ is either $\mathcal{S}_4$ or $\mathcal{S}_5$). Only when $\phi$ is estimated and the systematic contamination is not heavy is RM1 better than RM2 at estimating $\sigma_1^2$ and then only marginally better. In this case though the superiority of RM1 over RM2 is lost if estimation of $\phi$ is also taken into account.

Thus, based on the estimation of $\alpha$ and the joint estimation of $\phi$ and $\sigma_1^2$, RM2 is concluded to be preferable to RM1.

4. It therefore remains to compare RM2 and RM0. For estimating $\alpha$ the two estimators are roughly equivalent when there is no contamination in the random component whereas RM2 is generally better than RM0 otherwise. For estimating $\phi$ the conclusions are the same as for the conclusions when comparing RM2 and RM1, except that now RM2 gives an even more significant improvement over RM0 than it did over RM1.

   For estimating $\sigma_1^2$, based on the summary statistics in Tables 7.2 to 7.5, RM2 is preferable to RM0 if the distribution of the $b_i$, $(i = 1, \cdots, l)$ is either $\mathcal{S}_4$ or $\mathcal{S}_5$, this corresponding to the more extreme forms of systematic contamination considered. The RM2 estimate of $\sigma_1^2$ is also better than the RM0 estimate when $\phi$ is known and the random contamination distribution is $\mathcal{R}_3$, again the extreme form of contamination considered. Otherwise the average RM2 estimate of $\sigma_1^2$ is approximately the same as for the RM0 estimate except for a slightly larger positive bias. This would tend to indicate that the RM2

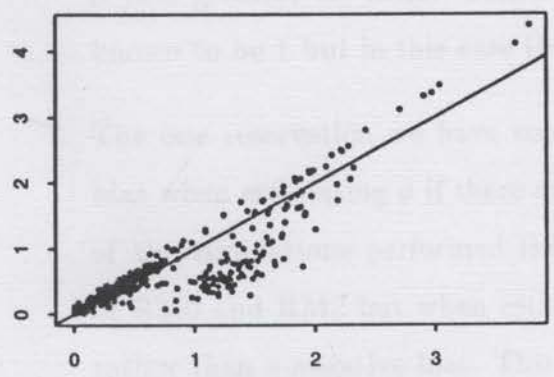
method is not very sensitive to outliers in the $b_i$, $(i = 1, \cdots, l)$ when non robust estimates of $\phi$ and $\sigma_1^2$ are used as starting values but that it can reduce the influence of gross errors in the $b_i$, $(i = 1, \cdots, l)$.

We now illustrate that using the RM2 method as in the simulations, that is with non robust staring values for $\phi$ and $\sigma_1^2$, the estimate for $\sigma_1^2$ can break down. In Figure 7.3 we have six plots each associated with the case $l = 10, r = 2$ and $\phi$ estimated. On each plot are plotted the RM2 versus the RM0 estimates of the $\sigma_1^2$ values for the 500 data sets generated according to a particular contamination model. The six contamination models considered correspond to the combinations of random contamination either $\mathcal{R}_1$, $\mathcal{R}_2$ or $\mathcal{R}_3$ and the systematic contamination either $\mathcal{S}_4$ or $\mathcal{S}_5$. On each of these six graphs the plot can be broken into three regions, particularly so for no random contamination and decreasingly so as the random contamination increases. Towards the left hand region of each plot, corresponding to no outliers in the $b_i$ $(i = 1, \cdots, l)$, RM0 and RM2 estimates are approximately equivalent. In the middle region the RM2 estimates are generally better (smaller) than the RM0 estimates and this region corresponds to 1 or 2 outliers in the $b_i$ $(i = 1, \cdots, l)$. Finally, in the right hand region, generally corresponding to 3 or more outliers in the $b_i$ $(i = 1, \cdots, l)$, the RM2 estimate breaks down and performs comparably with RM0.

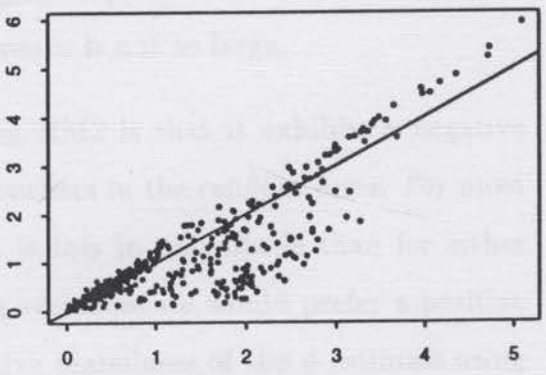
This property of the RM2 estimator when using non robust starting values is a general property of robust procedures (see, for example, Staude and Sheather (1990), *Remark 2*, page 255). It would therefore be of benefit to obtain better starting values in which case RM2 would be expected to perform even better than RM0. RM3 may even perform better in such cases.

In the absence of improved starting values, RM2 is still preferable to RM0 based on the earlier conclusions and should perform reasonably well provided that there are not too many gross errors.

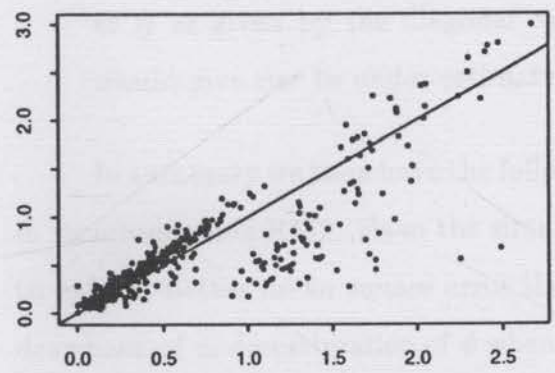
Figure 7.3: Plots of RM2 versus RM0 systematic variance estimates (l=10, r=2)



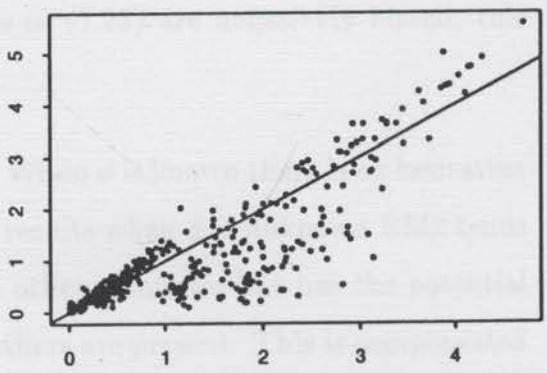

Contamination: Random = 1, Systematic = 4
Figure 7.3(a)



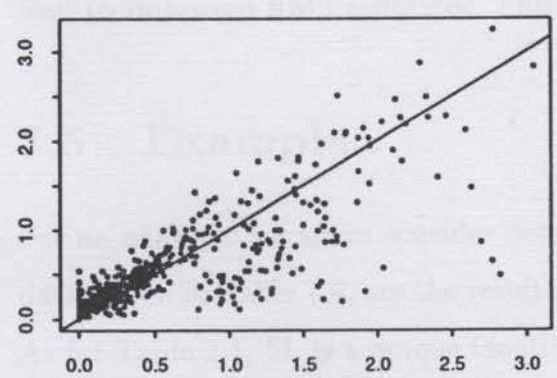

Contamination: Random = 1, Systematic = 5
Figure 7.3(b)



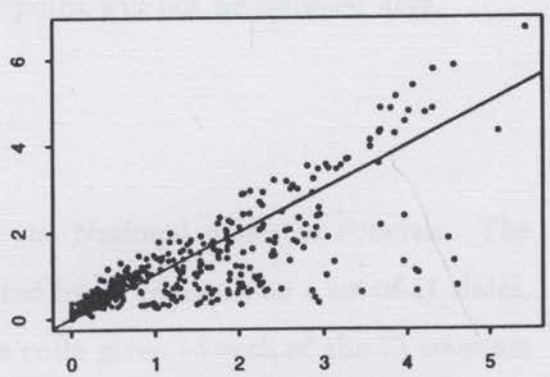

Contamination: Random = 2, Systematic = 4
Figure 7.3(c)




Contamination: Random = 2, Systematic = 5
Figure 7.3(d)




Contamination: Random = 3, Systematic = 4
Figure 7.3(e)




Contamination: Random = 3, Systematic = 5
Figure 7.3(f)

Note: RM0 estimates plotted on horizontal axes

5. As with the simulations in Chapter 6, when the dispersion parameter is unknown RM2 takes more iterations on average to converge than either of the other three estimators. The variability of the number of iterations in this case is also greatest for RM2. This is also generally the case when $\phi$ is taken as known to be 1 but in this case the difference is not so large.

6. The one reservation we have concerning RM2 is that it exhibits a negative bias when estimating $\phi$ if there are no outliers in the random error. For most of the simulations performed this bias is less in magnitude than for either of RM0 and RM2 but when estimating variances we would prefer a positive rather than a negative bias. This negative biasedness of the $\phi$ estimate using the RM2 method may be a consequence of the approximations inherent in the method. For example, if the approximation for the variances of the elements of $\hat{\eta}$ as given by the diagonal elements of (7.23) are negatively biased, this would give rise to under estimates of $\phi$.

In summary we then have the following. When $\phi$ is known there is no hesitation in recommending RM2. From the simulation results when $\phi$ is unknown RM2 tends to exhibit better mean square error than the other estimates but has the potential drawback of underestimation of $\phi$ when no outliers are present. This is compensated to some extent by a positively biased $\sigma_1^2$ estimate. More robust starting values of $\phi$ and $\sigma_1^2$ than the Henderson method 3 estimates would be preferred and such should lead to improved RM2 estimates. This latter point will not be pursued here.

## 7.5 Example.

For an example we again consider data from the National Asbestos Program. The data, given in Table 7.6, are the results reported by 75 counters on a set of 11 slides. As for Table 2.1, Id. is a unique identification code given to each of the 75 counters reporting results on these slides. Each counter's status (CS) is also given in Table 7.6 with M denoting an automatic image analyser, R a registered counter and O

**Table 7.6:** Counts reported by different counters on a set of eleven slides

| Id. | CS | Slide 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | M | | 9 | | 7 | | 5 | | | | 4 | 11 |
| 2 | M | | 13 | | 9 | | 16 | | | | 10 | 11 |
| 3 | O | | | | | | | | | | | 86 |
| 4 | O | | | | | 12 | | | | | | |
| 5 | O | | | | | | 17 | | | | | |
| 6 | O | | | | 10 | | | | | | | |
| 7 | O | 12 | | | | | | | | | | 16 |
| 8 | O | | | | | | | | | | | 15 |
| 9 | O | | | | | | 13 | | 11 | | | |
| " | O | | | | | | | | 11 | | | |
| 10 | O | | | | | 12 | | | | | | |
| 11 | O | 9 | | | | | | | | | 9 | |
| 12 | O | | | | | | 29 | | | | | |
| 13 | O | | 10 | 13 | | | | | | | | |
| 14 | O | | | | | 188 | 10 | | | | | |
| 15 | O | | | | 13 | | | | | | | |
| 16 | O | | | | | | 12 | | 14 | | | 16 |
| 17 | O | | | | 10 | | | | | | | |
| 18 | O | | | | 36 | | | | | | | |
| 19 | O | | | | 17 | | | 14 | | | | |
| 20 | O | | | | | | | 6 | | | | |
| 21 | O | | | | | | | | | | 17 | |
| 22 | O | 28 | | | | | | | | | | |
| 23 | O | | | | | | | 15 | | | | |
| 24 | O | | | | | | | | 7 | | | |
| 25 | O | | | | | | 12 | | | | | |
| 26 | O | | | | | | | | | | 18 | |
| 27 | O | 10 | | 6 | | | | | | | | 16 |
| 28 | O | | | | | | | | | | 14 | |
| 29 | O | | | | | | | | | | 7 | |
| 30 | O | | | | | | | | | | 13 | |
| 31 | O | | | | | | 10 | | | | | |
| 32 | O | 15 | | | | | 18 | | | | | |
| 33 | O | | 20 | | | 14 | | | | | | |
| 34 | O | 10 | | | | | | | | | | |
| 35 | O | | | | | | 16 | | | | | 10 |
| 36 | O | | | | | 28 | 18 | | | | 26 | |
| 37 | O | | | | | | | | | | | 22 |
| 38 | O | | | | | | | | | | 3 | |
| 39 | O | | | 9 | | 6 | | | | | | |
| 40 | O | 14 | | | | | | | | | | |
| 41 | O | | | | | | | | | | 13 | |
| 42 | O | | | | | | | | | | 14 | |
| 43 | O | | | 16 | | | | | | | | |

| Id. | CS | Slide 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 44 | O | | | | 12 | | | | | | | |
| 45 | O | | | 13 | | | | | | | | |
| 46 | O | | 32 | | | | | | | | | 24 |
| 47 | O | | | | 8 | | | | | | | |
| 48 | O | | | | | | | 8 | | | | |
| 49 | O | | | | | | | 21 | | | | |
| 50 | O | | | 12 | | | | | 13 | | | |
| 51 | O | | | | | | | | 13 | | | |
| 52 | O | | | | | | | | | 14 | | |
| 53 | O | | | | | | | | 10 | | | |
| 54 | O | | | | | | 26 | | | | | 25 |
| 55 | O | | | | 62 | | | 52 | | | | |
| 56 | O | | | | | | 6 | | | | | |
| 57 | O | | | | | | | | | | 12 | |
| 58 | O | 12 | | | | | | | | | | |
| 59 | O | | | 10 | | | | | | | | |
| 60 | O | | | | | | 10 | | | | | |
| 61 | O | | | | 6 | | | | | | | |
| 62 | O | | | | | | | | | | | 9 |
| 63 | O | | | | | | | | | 8 | | |
| " | O | | | | | | | | | 16 | | |
| 64 | O | | | | | | | | 10 | | | |
| 65 | O | | | | | | | 9 | 5 | | | |
| 66 | O | | | | | | | | 12 | | | 11 |
| 67 | O | | | | | | | | | 12 | | |
| 68 | O | | | | 10 | | | 17 | | | | |
| 69 | R | | | | | | | | | | 8 | |
| 70 | R | | | | | | | 18 | | 12 | | |
| " | R | | | | | | | 12 | | | | |
| 71 | R | | | 7 | | 10 | | | | | | |
| 72 | R | 12 | 6 | 14 | 17 | 10 | 14 | 19 | 15 | 19 | 10 | 17 |
| " | R | 10 | | 16 | 14 | 22 | 10 | 18 | 8 | 17 | 14 | 14 |
| " | R | 19 | | 21 | 7 | 12 | 10 | 8 | 12 | | 19 | 12 |
| " | R | | | 16 | | | | 8 | | | 14 | |
| " | R | | | | | | | 13 | | | | |
| " | R | | | | | | | 26 | | | | |
| " | R | | | | | | | 22 | | | | |
| " | R | | | | | | | 22 | | | | |
| " | R | | | | | | | 10 | | | | |
| " | R | | | | | | | 12 | | | | |
| 73 | R | | | | 48 | | | | | | | |
| 74 | R | | | | 14 | | | | | | | |
| 75 | R | | | | | | | | | 18 | | |

*Note:* Id. is the counter's identification number, CS denotes the counter's status.

for other counters. Here the correspondence with the categorization given in Table 2.1 is that M and O are equivalent to 1 and 4 respectively whilst R is equivalent to 2 and 3.

The notation we now use to formulate a model for the data in Table 7.6 differs from that given in Chapter 2 and instead corresponds to the notation of this chapter except that here we use multiple subscripts.

Let $y_{ijk}$ denote the $k^{th}$ count by the $i^{th}$ counter on the $j^{th}$ slide ($i = 1, \cdots, 75$; $j = 1, \cdots, 11$; $k = 1, \cdots, n_{ij}$). The model we use for a given counter's results is that the results, conditional on the given means, are independent Poisson random variables with means $\mu_{ijk}$, that is,

$$y_{ijk} \sim P(\mu_{ijk}) \text{ independently for } j = 1, \cdots, 11 \; ; \; k = 1, \cdots, n_{ij}.$$

We now assume that the conditional means can be model as follows:

$$\ln(\mu_{ijk}) = \alpha_j + x_i \, \alpha_{12} + b_i$$

where $\alpha_j$ is the effect for $j^{th}$ slide ($j = 1, \cdots, 11$), $b_i$ is the effect of the $i^{th}$ counter ($i = 1, \cdots, 75$) and $x_i$ equals 1 if the counter's status is M or 0 otherwise. Thus $\alpha_{12}$ is a bias adjustment parameter for automatic image analyser counts (see corresponding parameter in the example in Chapter 2). We then assume that

$$b_i \sim N(0, \sigma_1^2) \text{ independently for } i \in \{1, 2, 69, \cdots, 75\} \text{ independently of}$$

$$b_i \sim N(0, \sigma_2^2) \text{ independently for } i \in \{3, \cdots, 68\} \; .$$

Thus we are allowing the variation of the counters with status M or R to differ from those with status O, expecting $\sigma_1^2$ to be less than $\sigma_2^2$.

For the above model we then have that the conditional variance of $y_{ijk}$ is given by

$$Var(y_{ijk}|\mu_{ijk}) = \phi \, \mu_{ijk} \quad \text{where} \quad \phi = 1 \; . \tag{7.47}$$

Before fitting this model it is worthwhile making some brief comments on the difference between the model employed here and the model employed in the example of Chapter 2. First, in the example in Chapter 2 the model corresponds to a

population average (PA) model since it is the unconditional mean that is linked to the regression model. The model employed in this example on the otherhand corresponds to a subject specific (SS) model. The fixed parameters in the regression models will therefore have slightly different interpretations. Second, the basic difference between the two modelling approaches is in the underlying distribution assumed for the counter effects. If instead of assuming that the $\gamma_i$ in the example in Chapter 2 were gamma random variables with unit mean it was assumed that they were lognormal the model used in that example could be expressed in the form given in the example we are currently considering. However, given that for a gamma distribution with unit mean there is a similarly shaped lognormal distribution, provided the variance of the gamma distribution is less than 1, the conclusions drawn using the two methods should not differ significantly.

Fitting the model above, with $\phi = 1$ in (7.47), using the modified Fellner method (RM2) detailed in Section 7.3.1 for estimating the variance components, gives the following estimates for the fixed regression parameters and the variance components:

| | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | $\alpha_6$ | $\alpha_7$ | $\alpha_8$ | $\alpha_9$ | $\alpha_{10}$ | $\alpha_{11}$ | $\alpha_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimate | 2.58 | 2.63 | 2.63 | 2.62 | 2.69 | 2.52 | 2.66 | 2.42 | 2.73 | 2.53 | 2.70 | −0.41 |
| Std. Err. | 0.12 | 0.15 | 0.11 | 0.10 | 0.11 | 0.10 | 0.09 | 0.11 | 0.14 | 0.10 | 0.10 | 0.28 |

| $\sigma_1^2$ | $\sigma_2^2$ |
|---|---|
| 0.12 | 0.14 |

The standard errors for the estimates of the $\alpha_i$, $i = 1, \cdots, 11$ are approximations based on the expression for the variance for the standard BLUP estimate of $\boldsymbol{\alpha}$ as given in (7.21).

In addition to estimating the parameters the standardized values of the random components in the model were also calculated. Here the estimates of the realized counter effects ($\beta_i$, $i = 1, \cdots, 75$) are standardized based on the corresponding diagonal elements of the Var($\hat{\boldsymbol{\beta}}$) given in (7.22) whilst the (pseudo) residuals

$\hat{\eta} = \ln(y) - \ln(\hat{\mu})$ are similarly standardized based on the $\text{Var}(\hat{\eta})$ given in (7.23). Histograms of these standardized random components, counter and residual, are given in Figures 7.4(a) and 7.4(b) respectively.

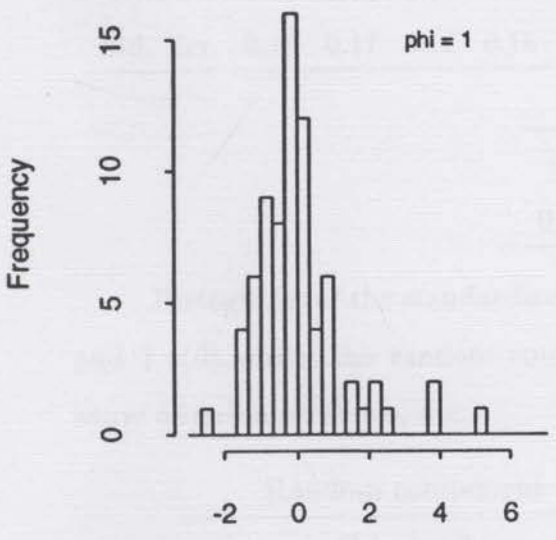
Here the "outliers", basing judgement of "outliers" on the criterion that the absolute standardized random component exceeds 3, are:

| Random component | Standardized random component |
|---|---|
| Counter 3 | 5.06 |
| Counter 55 | 3.91 |
| Counter 73 | 3.95 |
| Counter 14, Slide 5, result 1 | 12.09 |
| Counter 38, Slide 10, result 1 | -4.31 |
| Counter 72, Slide 2, result 1 | -3.50 |

To conclude the above analysis of the data let us make a few comments. First, given that the 11 slides are from a nominally identical batch of 14 artificially produced slides (see Section 1.1) it would be difficult based on an examination of the slide effects ($\alpha_i$, $i = 1, \cdots, 11$) and their approximate standard errors to reject the assumption that all slides have the same mean fibre loadings. Second, the estimate of the extent of proportionately lower counts obtained by the automatic image analysers is consistent with the example in Chapter 2. The final comment concerns the estimates of the variances of the counter effects. Here the estimate of the variance for the effects for counters with status O (i.e. 0.14) is consistent with the results for the corresponding counters (non-accredited counters) in the Chapter 2 example. There is however a difference in the estimate of the corresponding value for the M and R counters in the example here and the equivalent counters (coded 1, 2 and 3) in the example of Chapter 2. This difference is conjectured not to be significant as the estimate of $\sigma_1^2$ here will have a large standard error associated with it given the small number of counters classified as either M or O.

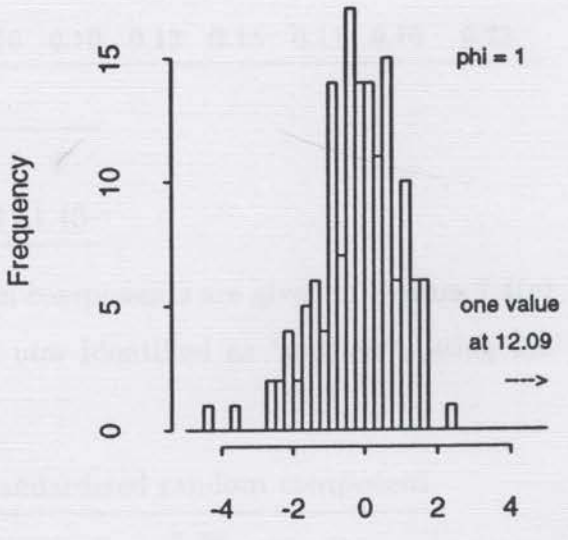
Before leaving this example estimates of the parameters in the model are given when $\phi$ in (7.47) is not assumed to be 1. Using the same procedure as above, that

# Figure 7.4: Standardized Effects for National Asbestos Data



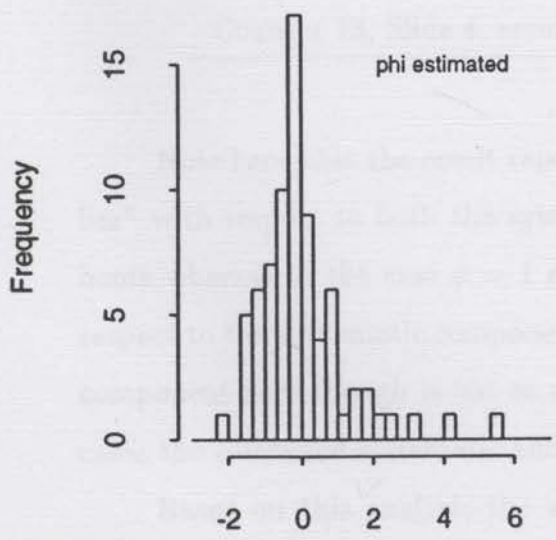

Standardized counter effects

Figure 7.4(a)



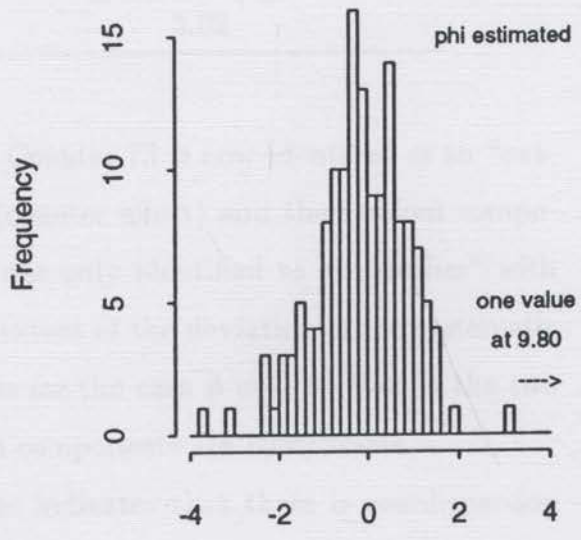

Standardized residuals

Figure 7.4(b)




Standardized counter effects

Figure 7.4(c)




Standardized residuals

Figure 7.4(d)

is the modified Fellner method, we obtain the following estimates:

| | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | $\alpha_6$ | $\alpha_7$ | $\alpha_8$ | $\alpha_9$ | $\alpha_{10}$ | $\alpha_{11}$ | $\alpha_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimate | 2.59 | 2.65 | 2.61 | 2.66 | 2.69 | 2.54 | 2.67 | 2.42 | 2.72 | 2.53 | 2.74 | −0.43 |
| Std. Err. | 0.13 | 0.17 | 0.12 | 0.10 | 0.12 | 0.10 | 0.10 | 0.13 | 0.15 | 0.11 | 0.10 | 0.23 |

| $\sigma_1^2$ | $\sigma_2^2$ | $\phi$ |
|---|---|---|
| 0.07 | 0.13 | 1.40 |

Histograms of the standardized random components are given in Figures 7.4(c) and 7.4(d) whilst the random components now identified as "outliers", using the same criterion as above, are:

| Random component | Standardized random component |
|---|---|
| Counter 3 | 5.36 |
| Counter 55 | 4.12 |
| Counter 73 | 3.06 |
| Counter 14, Slide 5, result 1 | 9.80 |
| Counter 38, Slide 10, result 1 | -3.69 |
| Counter 72, Slide 2, result 1 | -3.06 |
| Counter 73, Slide 4, result 1 | 3.02 |

Note here that the result reported by Counter 73 is now identified as an "outlier" with respect to both the systematic (counter effect) and the random components whereas in the case $\phi = 1$ above it was only identified as an "outlier" with respect to the systematic component. The extent of the deviation of the systematic component here though is not as marked as for the case $\phi = 1$, so that in the two cases the combined systematic and random components are comparable.

Based on this analysis the $\phi$ estimate indicates that there is overdispersion in the data relative to the Poisson but without some indication of the variability of the $\phi$ estimate it is difficult to determine if this is significant. Including this

overdispersion parameter in the model though we obtain estimates of the counter variances more in line with those obtained using the Schall method in the example of Chapter 2, that is $\sigma_1^2 = 0.060$ and $\sigma_2^2 = 0.135$. This agreement of the counter variabilities is what we would expect as the same population of counters is sampled for both examples.

## 7.6 Concluding remarks.

The three robust estimators considered in this chapter correspond to the three estimators in Chapter 6. Unlike the situation in Chapter 6 where the differences between the three estimators was not too large, here they perform very differently. This is considered in part to be caused by the differences in sensitivity of each method to changes in the weight matrix due to "outliers" in the data.

Of the three methods the robust method based on the modification of the Fellner procedure, identified as RM2, is considered the best. More work though needs to be undertaken to improve its performance. One possible way of achieving this goal would be the development of more robust starting values for the variance components than the moment based methods used in the simulations. For example, development of variance components estimates based on quantiles, whilst not efficient, should give good robust starting values for the RM2 method.

# Appendix A

# Notation

**Notation concerning functions**

$I(x \in S)$ denotes the indicator function taking the value 1 if $x \in S$ and 0 otherwise.

$[x]$ denotes the largest integer not greater than x.

$]x[$ denotes the smallest integer not less than x.

**Notation concerning probability distribution and density functions (pdf)**

$NBD(\mu, \alpha)$ denotes a negative binomial random variable with pdf

$$Pr(X = x) = \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)\, x!} \left(\frac{\alpha}{\alpha+\mu}\right)^{\alpha} \left(\frac{\mu}{\alpha+\mu}\right)^{x}$$

for $x = 0, 1, 2, \cdots$ ; $\alpha > 0$ ; $\mu > 0$ .

$\Gamma(\alpha, \beta)$ denotes a gamma random variable with pdf

$$P(X < x) = I(x > 0) \int_0^x \frac{t^{\alpha-1}\, e^{t/\beta}}{\alpha^{\beta}\, \Gamma(\alpha)}\, dt$$

for $\alpha > 0$ and $\beta > 0$.

$\phi(x)$ denotes the standard normal density, $(\sqrt{2\pi})^{-1} \exp(-x^2/2)$ , $-\infty < x < \infty$.

$\Phi(x)$ denotes the cumulative normal distribution, $\int_{-\infty}^{x} \phi(t)\, dt$ , $-\infty < x < \infty$.

**Notation concerning little $o$ and big $O$**

For two non-stochastic sequences $\{a_n\}$ and $\{b_n\}$ we have:

$$a_n = o(b_n) \iff \lim_{n\to\infty} \mid a_n/b_n \mid = 0$$
$$a_n = O(b_n) \iff \limsup_{n\to\infty} \mid a_n/b_n \mid < \infty$$

For a stochastic sequence $\{X_n\}$ and a non-stochastic sequence $\{b_n\}$ we have:

$$X_n = o_p(b_n) \iff \lim_{n\to\infty} P(\mid X_n/b_n \mid > \varepsilon) = 0 \text{ for all } \varepsilon > 0$$
$$X_n = O_p(b_n) \iff \text{For all } \varepsilon > 0 \text{ there exist } M \text{ and } N$$
$$\text{such that } P(\mid X_n/b_n \mid > M) < \varepsilon \text{ for all } n > N.$$

**Notation concerning matrices**

The following notation is taken from Searle et. al. (1992).

If $\mathbf{A}$ is an $m$ by $n$ matrix with elements $a_{i,j}$ $i = 1, \cdots, m$ ; $j = 1, \cdots, n$ then we have

$$\mathbf{A} = \{_m\, a_{ij}\}_{i=1}^{m}{}_{j=1}^{n},$$

which can also be shortened to $\{_m\, a_{ij}\}_{i,j}$ or even $\{_m\, a_{ij}\}$ when there is no chance of misunderstanding. Here the subscript $m$ on the first curly brackets denotes that the array is a matrix.

This notation is easily extended to column vectors (e.g. $\{_c\, b_i\}_{i=1}^{n}$) and row vectors (e.g. $\{_r\, c_j\}_{j=1}^{m}$). For a diagonal array $\mathbf{U}$ with diagonal elements $u_i$ $(i = 1, \cdots, n)$ the notational form is $\{_d\, u_i\}_{i=1}^{n}$.

The notation is also useful for matrices with block structure. In such cases the rules above apply except that elements are replaced by matrices.

# Appendix B

# Miscellaneous results

**Theorem B1.**

Let $X_i$ $(i = 1, \cdots, n)$ be independent random variables and let $g_j$ $(j = 1, \cdots, m)$ be functions such that, for non-negative integers $r_j$ $(j = 1, \cdots, m)$

$$\begin{aligned} E\ g_j(X_i) &= 0 \\ E\ |g_j(X_i)|^{2r_j} &\leq B_j \end{aligned}$$

for all $i = 1, \cdots, n$ and $j = 1, \cdots, m$ where the bounds $B_j$ are independent of $i$ and $n$. Then for

$$F(X_1, \cdots, X_n) = \prod_{j=1}^{m} \left( \sum_{i=1}^{n} g_j(X_i) \right)^{r_j}$$

we have

$$E\ F(X_1, \cdots, X_n) = O(n^{[r./2]})$$

where $r. = \sum_{i=1}^{m} r_j$ and $[x]$ denotes the integer part of $x$.

**Proof.**

Let

$$\mathcal{B}_{j,k} = \{\underline{b} = (b_1, \cdots, b_k) : b_l \in \mathbf{N}\ ;\ b_1 \geq \cdots \geq b_k > 0\ ;\ \sum_{l=1}^{k} b_l = r_j\} \qquad j = 1, \cdots, m$$

$$\mathcal{I}_{n,k} = \{\underline{i} = (i_1, \cdots, i_k) : i_l \in \mathbf{N}\ ;\ i_l \leq n\ ;\ i_l \neq i_{l'}\ \forall\ l \neq l'\}$$

where $\mathbf{N}$ denotes the set of positive integers.

Then, provided $n \geq \max\{r_1, \cdots, r_m\}$,

$$\prod_{j=1}^{m} \left( \sum_{i=1}^{n} g_j(X_i) \right)^{r_j} = \prod_{j=1}^{m} \left( \sum_{k=1}^{r_j} \sum_{\underline{b} \in \mathcal{B}_{j,k}} \sum_{\underline{i} \in \mathcal{I}_{n,k}} \prod_{l=1}^{k} (g_j(X_{i_l}))^{b_l} \right)$$

This latter expression, upon expansion, contains a finite number of terms, independent of $n$, of the form

$$\prod_{j=1}^{m} \left( \sum_{\underline{i}_j \in \mathcal{I}_{n,k_j}} \prod_{l=1}^{k_j} \left( g_j(X_{i_{j,l}}) \right)^{b_l} \right)$$

$$= \sum_{\underline{i}_1 \in \mathcal{I}_{n,k_1}} \cdots \sum_{\underline{i}_m \in \mathcal{I}_{n,k_m}} \left( \prod_{j=1}^{m} \prod_{l=1}^{k_j} \left( g_j(X_{i_{j,l}}) \right)^{b_{j,l}} \right) \qquad \text{(B.1)}$$

corresponding to the different sets of possible values for $\{k_1, \cdots, k_m\}$ and for all $\underline{b}_j \in \mathcal{B}_{j,k_j}$ $(j = 1, \cdots, m)$.

Each of these terms in turn contains a number of terms, each of which has bounded absolute expectation (less than $\prod_{j=1}^{m} B_j$). Hence the order of the expected value of (B.1) is equivalent to the number of terms in this expression that have non-zero expectation. This quantity in turn is less than or equal to the number of terms, $N$ say, that have $b_{j,l} > 1$ for all $j, l$ or have $i_{j,l} = i_{j',l'}$ for some $j \neq j'$ whenever $b_{j,l} = 1$. Since $N$ is independent of $g_j$, $j = 1, \cdots, m$ we can take these functions to be the identity function. Hence $N$ corresponds to the number of terms in the expansion of

$$\prod_{j=1}^{m} \left( \sum_{i=1}^{n} X_i \right)^{r_j} = \left( \sum_{i=1}^{n} X_i \right)^{r_\cdot}$$

that have all $X_i$ raised to at least the power 2. This is easily shown to be $n^{[r_\cdot/2]}$ and hence the proof is complete.

•

**Note:** Above it was assumed that the functions $g_j$ do not depend on $i$. This assumption is not necessary and the results continue to hold provided the $B_j$ do not depend on the $i$.

**Theorem B2.**

If $X_K \sim NBD(\mu_K, \alpha_K)$ with cumulants $\mathcal{K}_{K,r}$ $(r = 1, 2, \cdots)$ and $\mu_K/\alpha_K \downarrow 0$ then

$$|\mathcal{K}_{K,r}| \leq \mu_K M_r \quad \text{for all sufficiently large K}$$

where $M_r$ is a constant depending only on $r$.

**Proof.**

To find the cumulants of $X_K$ we introduce an alternative parameterization for the negative binomial distribution where we keep $\alpha_K$ but let $\beta_K = (\alpha_K + \mu_K)/\alpha_K$. Then, from Johnson and Kotz (1969) we have

$$\begin{aligned} \mathcal{K}_{K,1} &= \alpha_K (\beta_K - 1) \\ \mathcal{K}_{K,r+1} &= \beta_K (\beta_K - 1) \frac{\partial \mathcal{K}_{K,r}}{\partial \beta_K} \qquad r \geq 1. \end{aligned}$$

Hence $\mathcal{K}_{K,r} = \alpha_K \beta_K (\beta_K - 1) h(\beta_K)$ where $h(\beta_K)$ is a polynomial in $\beta_K$ with constant coefficients.

Now $\beta_K = 1 + \mu_K/\alpha_K$ hence there exists a $K^*$ such that $1 \leq \beta_K \leq 2 \; \forall \; K \geq K^*$. From this we have that $|h(\beta_K)|$ is bounded with the bound only depending on $r$. Also, since $\alpha_K \beta_K (\beta_K - 1) = \mu_K (1 + \mu_K/\alpha_K)$, this implies that $\alpha_K \beta_K (\beta_K - 1) \leq 2 \mu_K \; \forall \; K \geq K^*$.

The result then follows immediately.

•

# References


Abramowitz, M. and Stegun, I. A. (1968) *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables.* ($7^{th}$ printing.) U.S. Government Printing Office, Washington.

Anon (1978) *Asbestos: An Information Resource.* U.S. Department of Health, Education, and Welfare. (1978). DHEW Publication Number (NIH) 79-1681.

Anon (1984) *Report on the Health Hazards of Asbestos.* (2nd ed.) National Health and Medical Research Council. Canberra: Australian Government Publishing Service.

Anon (1988) Guidance note on the membrane filter method for estimating airborne asbestos dust [NOHSC:3003]. National Occupational Health and Safety Commission. Canberra: Australian Government Publishing Service.

Anraku, K. and Yanagimoto, T. (1990). Estimation for the negative binomial distribution based on the conditional likelihood. *Communication in Statistics, Simulation and Computation* 19, 771-786.

Anscombe, F.J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika* 37, 358-382.

Atkinson A. C. and Pearce M. C. (1976). The computer generation of beta, gamma and normal random variables. *J. R. Statist. Soc. A* 139, 431-461.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* New York: John Wiley.

Barnwal, R.K. and Paul, S.R. (1988). Analysis of one-way layout of count data with negative binomial variation. *Biometrika* 75, 215-222.

Best, D.J. and Gipps, P.G. (1974). An improved gamma approximation to the negative binomial. *Technometrics* 16, 621-624.

Breslow, N.E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* 33, 38-44.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88, 9-25.

Cheng R. C. H. (1977). The generation of gamma variables with non-integral shape parameters. *Applied Statistics* 26, 71-75.

Chen, J. and Rubin, H.(1986). Bounds for the difference between mean and median of gamma and Poisson distributions. *Statistics and Probability Letters* 4, 281-283.

Choi, K.P. and He, X. A note on the medians of binomial and Poisson distributions. Unpublished manuscript (personal communication, 1991).

Chung, K.L. (1974). *A Course in Probability Theory.* London: Academic Press.

Clark, S. J. and Perry, J. N. (1989). Estimation of the negative binomial parameter k by maximum quasi-likelihood. *Biometrics* 45, 309-316.

Collings, B.J. and Margolin, B.H. (1985). Testing goodness of fit for the Poisson assumption when observations are not identically distributed. *Journal of the American Statistical Association* 80, 411-418.

Cooil, B. (1991). Using medical malpractice data to predict the frequency of claims: A study of Poisson process models with random effects. *Journal of the American Statistical Association* 86, 285-295.

Crawford, N.P. and Cowie, A.J. (1984). Quality control of airborne asbestos fibre counts in the United Kingdom - the present position. *Annals of Occupational Hygiene* 28, 391-398.

Davidian, M. and Carroll, R.J. (1988). A note on extended quasi-likelihood. *J. R. Statist. Soc. B* 50, 74-82.

Dean, C., Lawless, J.F. and Willmot, G.E. (1989). A mixed Poisson-inverse-Gaussian regression model. *The Canadian Journal of Statistics* 17, 171-181.

Fahrmeir, L. (1987). Asymptotic likelihood inference for nonhomogeneous observations. *Statistische Hefte* 28, 81-116.

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

Fisher, R.A. (1941). The negative binomial distribution. *Annals of Eugenics* 11, 182-187.

Gilmour, A.R., Anderson, R.D. and Rae, A.L. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72, 593-599.

Göb, R. (1994). Bounds for median and 50 percentage point of binomial and negative binomial distribution. *Metrika* 41: 43-54.

Groeneveld, R.A. and Meeden, G. (1977). The mode, median, and mean inequality. *The American Statistician* 31, 120-121.

Guenther, W.C. (1972). A simple approximation to the negative binomial (and regular binomial). *Technometrics* 14, 385-389.

Hall, P. (1985). Correcting segment counts for edge effects when estimating intensity. *Biometrika* 72, 459-463.

Hall, P. (1994). On the erratic behavior of estimators of $N$ in the binomial $N, p$ distribution. *Journal of the American Statistical Association* 89, 344-352.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320-340.

Henderson, C.R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant breeding*, National research Council Publication No. 892, pp 141-63. National Academy of Sciences.

Henderson, G.P. (1992). Solution number 1628. *Crux Mathematicorum* 18, 111-112.

Hinde, J.P. (1982). Compound Poisson regression models. *GLIM 82: Proceedings of the International Conference in Generalised Linear Models* Gilchrist, R. (editor) Springer-Verlag, Berlin, 109-121.

Huber, P.J. (1981). *Robust Statistics.* New York: Wiley.

Johnson, N.L. and Kotz, S. (1992). *Discrete Distributions.* Boston: Houghton Mifflin Company.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Distributions.* (2nd ed.) New York: John Wiley.

Kale, B.K. (1962). On the solution of likelihood equations by iterative processes - multiparameter case. *Biometrika* 49, 479-486.

Karim, M.R. and Zeger, S.L. (1992). Generalized linear models with random effects; Salamander mating revisited. *Biometrics* 48, 631-644.

Katti, S.K. and Gurland, J. (1962). Efficiency of certain methods of estimation for the negative binomial and the Neyman type A distributions. *Biometrika* 49, 215-226.

Kendall, M. and Stuart, A. (1977). *The Advanced Theory of Statistics* Vol 1. (4th ed.) London: Griffin.

Kimber, A.C. (1983). Trimming in gamma samples. *Applied Statistics* 32, 7-14.

Lawless, J.F. (1987a). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* 15, 209-225.

Lawless, J.F. (1987b). Regression methods for Poisson process data. *Journal of the American Statistical Association* 82, 808-815.

Levin, B. and Reeds, J. (1977). Compound multinomial likelihood functions are unimodal: Proof of a conjecture of I. J. Good. *The Annals of Statistics* 5, 79-87.

Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.

Lukacs, E. (1975). *Stochastic Convergence.* London: Academic Press.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics* 11, 59-67.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* (2nd ed.) London: Chapman and Hall.

McGilchrist, C.A. (1994). Estimation in generalized mixed models. *J. R. Statist. Soc. B* 56, 61-69.

McNeney, B. and Petkau, J. (1994). Overdispersed Poisson regression models for studies of air pollution and human health. *The Canadian Journal of Statistics* 22, 421-440.

Mihram, G.A. (1972). *SIMULATION Statistical Foundations and Methodology.* New York and London: Academic Press.

Nelder, J.A. and Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika* 74, 221-232.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *J. R. Statist. Soc. A* 135, 370-384.

Ogden, T.L. (1982) The reproducibility of asbestos counts. Health and Safety Executive, Research Paper 18: United Kingdom.

Paik, M.C. (1992). Parametric variance function estimation for nonnormal repeated measurement data. *Biometrics* 48, 19-30.

Patil, G.P. (1960). On the evaluation of the negative binomial distribution with examples. *Technometrics* 2, 501-505.

Payton, M.E., Young, L.J. and Young, J.H. (1989). Bounds for the difference between median and mean of beta and negative binomial distributions. *Metrika* 36, 347-354.

Pessin, V.(1961). Some asymptotic properties of the negative binomial distribution. *The Annals of Mathematical Statistics*, 32, 922-923. (abstract).

Piegorsch, W.W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics* 46, 863-867.

Pierce, D.A. and Schafer, D.W. (1986) Residuals in generalized linear models. *Journal of the American Statistical Association* 81, 977-986.

Pieters, E.P., Gates, C.E., Matis, J.H. and Sterling, W.L. (1977). Small sample comparison of different estimators of negative binomial parameters. *Biometrics* 33, 718-723.

Rao, C.R. and Kleffe, J. (1988) *Estimation of Variance Components and Applications.* North-Holland, Amsterdam.

Robinson, G.K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science* 6, 15-51.

Rocke, D.M. (1983). Robust statistical analysis of interlaboratory studies. *Biometrika* 70, 421-431.

Rocke, D.M. (1991). Robustness and balance in mixed models. *Biometrics* 47, 303-309.

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

Searle, S.R. (1995). An overview of variance components estimation. *Metrika* 42, 215-230.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components.* New York: John Wiley.

Sibuya, M., Yoshimura, I. and Shimura, R. (1964). Negative multinomial distribution. *Annals of the Institute of Statistical Mathematics, Tokyo* 16, 409-426.

Smyth, G.K. (1989). Generalized linear models with varying dispersion. *J. R. Statist. Soc. B* 51, 47-60.

Staudte, R.G. and Sheather, S.J. (1990) *Robust Estimation and Testing.* New York: John Wiley.

Stein, G.Z. (1988). Modelling counts in biological populations. *Math. Scientist* 13, 56-65.

Swallow, W.H. and Monahan, J.F. (1984) Monte carlo comparison of ANOVA, MINQUE, REML, and ML estimators of variance components. *Technometrics* 26, 47-57.

Teicher, H. (1955). An inequality on Poisson probabilities. *The Annals of Mathematical Statistics* 26, 147-149.

Thall, P.F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* 44, 197-209.

Tripathi, R.C. (1985). Negative Binomial Distribution. *Encyclopedia of Statistical Sciences* Vol 6, 169-177. New York: Wiley.

van de Ven, R. (1993). Estimating the shape parameter for the negative binomial distribution. *J. Statist. Comput. Simul.* 46, 111-123.

van de Ven, R. and Weber, N.C. (1993). Bounds for the median of the negative binomial distribution. *Metrika* 40, 185-189.

van de Ven, R. and Weber, N.C. (1995). Log-linear models for mean and dispersion in mixed Poisson regression models. *Austral. J. Statist.* 37, 205-216.

van de Ven, R. and Weber, N.C. (1996). Percentile related bounds and estimator for the mean of the negative binomial distribution. *Statistics*, 345-352.

van Duijn (1992). A mixed model approach to the analysis of count data. *Statistical Modelling* P.G.M. van der Heijden et al. (editors) Elsevier Science Publishers B.V., Amsterdam, 75-84.

Waters, W.E. (1959). A quantitative measure of aggregation in insects. *Journal of Economic Entomology* 52, 1180-1184.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61, 439-447.

Willson, L.J., Folks, J.L. and Young, J.H. (1984). Multistage estimation compared with fixed-sample-size estimation of the negative binomial parameter k. *Biometrics* 40, 109-117.

Windham, M.P. (1995). Robustifying model fitting. *J. R. Statist. Soc. B* 57, 599-609.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: A pseudo likelihood approach. *J. Statist. Comput. Simul.* 48, 233-243.

Zeger, S.L., Liang, K.Y. and Albert, P.S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* 44, 1049-1060.

Zeger, S.L. and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42, 121-130.