Estimation and inference with complex count data from fisheries surveys, including over-dispersion, many nuisance parameters, and correlation.

by

©Shijia Wang

A thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirement for the Degree of

Master of Science

Department of Mathematics and Statistics

Memorial University

St. John's Newfoundland and Labrador, Canada April 2015

Abstract

We study methods to estimate regression and variance parameters for over-dispersed and correlated count data from highly stratified surveys. A challenge with such data is the large number of nuisance parameters which leads to computational issues and biased statistical inferences. We develop a profile generalized estimating equation (GEE) method that is more computationally efficient and compare it to marginal maximum likelihood (MLE) and restricted MLE (REML) methods. We use REML to address bias and inaccurate confidence intervals because of many nuisance parameters. The marginal MLE and REML approaches involve intractable integrals and we used a new R package that is designed for estimating complex nonlinear models that may include random effects. We conduct simulation analyses and conclude that the REML method is the better approach among the three methods we investigate.

Our applications involve counts of fish catches from highly-stratified research surveys. In the first application, we estimate the day and night (diel) effect for three species from bottom trawl research surveys. In the second application, we estimate the diel and vessel effects of two different snow crab surveys.

Acknowledgements

First of all, I want to express my gratitude to my supervisor Dr. Noel Cadigan. Due to the very limited space of this section, I briefly thank him for introducing me to this interesting area of statistics, for his practical guidance, insights, and support with my research. Also thanks for his exceptional generosity with his time and energy in my PhD application.

Thanks to the professors in MUN. They have given me excellent lectures and all kinds of help and support. Particularly, I would like to thank my co-supervisor Dr. Hong Wang, providing suggestions and inviting me to many holiday parties in his house. I am grateful to Dr. JC Loredo-osti and Dr. Zhaozhi Fan, for giving me suggestions and help in PhD application. Thanks to Dr. Asokan Variyath, for teaching me very interesting computing methodologies.

Further more I would like to thank CFER (Center for fisheries ecosystems research), for creating a friendly and comfortable working environment. Also, thanks to Statistical Society of Canada for offering me student travel award. Thanks to Hugues Benoît at the Department of Fisheries and Oceans in Moncton, New Brunswick, for providing the data for used in this thesis and for helpful discussions about my research.

I appreciate my parents for their support and encouragement. Most of all, I would like to show my appreciation to my wife Shufei, for her support and sense of humour.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Negative Binomial (NB) distribution is commonly used for analyzing biological count data with Poisson over-dispersion (e.g. Ross and Preece, 1985 [36]). The NB distribution can be generated from a gamma mixture of Poisson random variables, and this often motivates its use when there is between-individual heterogeneity in the Poisson means. The variance is equal to the mean for the Poisson distribution, but this is not necessarily so for the NB distribution. If $Y \sim$ NB with mean μ then $Var(Y) = \mu + k^{-1}\mu^2$, where k is called the dispersion parameter. The condition $Var(Y) > \mu$ is referred to as over-dispersion. In this thesis we use a Poisson-double-Gamma (PdG) mixture model for count data, where the mixing component is based on two gamma random variables to account for different sources of over-dispersion and correlation in the data. The resulting marginal distribution of a single observation is not NB in form but the mean and variance are the same as those of NB random variables.

Our objective is statistical inference about regression-type parameters based on highly stratified count data; in particular, counts of fish caught in bottom-trawl surveys. These research surveys provide important information for the assessment and management of many fish stocks worldwide. The sampling unit is defined as the area over the bottom covered by a trawl of specified width towed at a targeted fixed speed and distance. The NB distribution is often suggested to be appropriate for modelling catches from this type of survey (e.g. Gunderson, 1993 [19]; Kimura and Somerton, 2006 [22]), other types of survey fishing gear (e.g. Power and Moser, 1999 [32]), and commercial fisheries (e.g. Baum and Myers, 2004 [3]), although so-called delta distributions (e.g. Stefánsson, 1996 [42]), where zero values are treated separately and positive values are assumed to follow a lognormal distribution, are sometimes used. Other approaches have been proposed, such as the Log Gaussian Cox Process (LGCP) (e.g. Lewy and Kristensen, 2009 [25]), which is a mixture of Poisson-distributed observations with mean densities following a multivariate lognormal distribution.

Most trawl surveys in the Northwest Atlantic use a stratified survey design (e.g. Doubleday, 1981 [17]), where strata are based on contiguous spatial areas with similar

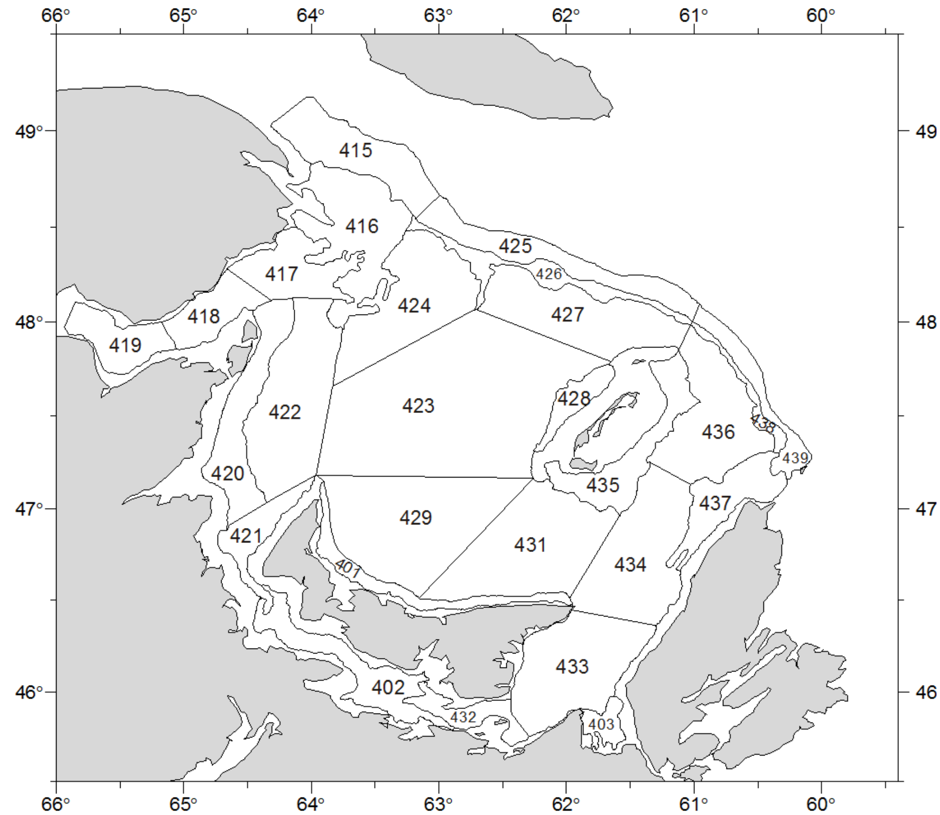

Figure 1.1: Stratified survey design for the southern Gulf of St. Lawrence off the coasts of New Brunswick, Nova Scotia, and Prince Edward Island.

bottom depths (e.g. Figure 1.1). Strata are constructed so that in many cases it is reasonable to assume that fish densities are homogeneous (i.e. identically distributed) within strata. Strata are usually relatively small to account for complex patterns of species occurrence related to bottom topography and sediment type (i.e. mud, sand, rocks), ocean currents and water temperatures. Hence, most surveys have many strata (25-200) and not many (≤ 10) samples per stratum.

We assume data are sampled from H strata with n_h sites per stratum. At a particular site more than one observation (i.e. n_{hi}; $i = 1, ..., n_h$) can occur with possibly different covariates. For example, in a fisheries survey there may be two vessels with somewhat different gears used for sampling and these vessels may fish at the same site to compare catch rates of fish. This is often referred to as comparative fishing. If the vessels always fish in different strata then potential differences in the fishing efficiency (i.e. ρ) of the vessels/gears will be confounded with differences in fish density between strata. Sometimes both vessels are used in the same strata which gives some information about differences in ρ between the two vessels/gears, given the assumption of within stratum homogeneity. If both vessels fish at the same site (i.e. paired tows) then this gives even better information on differences in ρ.

The model we propose for this type of data accommodates these sampling features. Let Y_{hij} be a random variable for the j'th observation in stratum h ($h = 1, ..., H$) and site i. We assume there is a stratum effect (μ_h), a site effect (γ_{hi}) and a replicate effect (γ_{hij}) at site i. Y_{hij} is assumed to be conditionally Poisson distributed with mean $E(Y_{hij}|\gamma_{hi}, \gamma_{hij}) = \mu_h \gamma_{hi} \gamma_{hij} \eta_{hij}$, and variance $Var(Y_{hij}|\gamma_{hi}, \gamma_{hij}) = \mu_h \gamma_{hi} \gamma_{hij} \eta_{hij}$, where η_{hij} is a function of a small number of regression parameters, denoted as β_k and covariates x_{hijk}, $k = 1, ..., p$. For example, $\eta_{hij} = \exp(\sum_{k=1}^{p} \beta_k x_{hijk})$ and x_{hijk} could be an indicator variable for vessel in which case $p = 2$. The μ_h's are treated as fixed parameters to estimate. The γ_{hi}'s are assumed to be independent and identically distributed (iid) gamma RV's with mean 1 and variance $1/k_s$, and the replicate effects are assumed to be iid gamma RV's with mean 1 and variance $1/k_c$. We expect $Var(\gamma_{hi}) > Var(\gamma_{hij}) \Rightarrow k_s < k_c$ since we expect the between-site variability to be greater than the within-site variability during repeated tows.

When the focus is on $\boldsymbol{\beta}$ then the μ_h's can be considered as nuisance parameters. However, k_c and k_s are not really nuisance parameters because they are important for statistical inferences (i.e. confidence intervals) for $\boldsymbol{\beta}$. It is well known that when H is large the resulting large number of nuisance parameters can cause bias when estimating $\boldsymbol{\beta}$, k_c and k_s (e.g. Barndorff-Nielsen and Cox, 1994 [2]). We use an example of the normal linear regression model to illustrate this. Let $\boldsymbol{y}$ be a $n \times 1$ vector of sample responses, $\boldsymbol{\beta}$ be a $p \times 1$ parameter vector and $\boldsymbol{X}$ be a $n \times p$ covariance matrix. We assume a linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}).$$

The ML estimators for $\boldsymbol{\beta}$ and σ^2 are

$$\hat{\boldsymbol{\beta}}_{\text{ML}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

$$\hat{\sigma}^2_{\text{ML}} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{ML}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{ML}})/n.$$

$\hat{\sigma}^2_{\text{ML}}$ is biased since

$$E(\hat{\sigma}^2_{\text{ML}}) = \frac{n-p}{n}\sigma^2,$$

and when p is large (i.e. $p = n/2$) the bias can be substantial. In Chapter 4 we study a procedure for producing an unbiased estimator of σ^2.

For the stratified model setting, Sartori (2003) [37] and Bellio and Sartori (2006) [5] showed that standard likelihood inferences may not be accurate unless $n_h > H$ on average. Bellio and Sartori (2006) [5] found that the Maximum Likelihood Estimation (MLE) relative bias for k_c was over 50% for a highly parameterized NB model. They suggested the MLE adjustment proposed by Severini (1998) [40] based on the modified profile likelihood is convenient to use for stratified count data, and demonstrated that this estimator had substantially lower bias than the MLE itself. Cadigan and Tobin (2010) [12] examined bias and mean squared error for several estimators of k_c. They proposed an adjusted double extended quasi-likelihood estimator of k_c that gave much improved performance compared to the MLE. In this thesis we use restricted maximum likelihood estimation (REML) to deal with this bias problem, and we show how this can be easily implemented with the software we use.

The marginal (with respect to replicate effects) distribution of Y_{hij} is NB, conditional on site effects. This is shown in Appendix A, and more information is available in Cameron and Trivedi (2013), who give a detailed description of Poisson random effects models. The conditional mean is $E(Y_{hij}|\gamma_{hi}) = \mu_{hij} = \mu_h \gamma_{hi} \eta_{hij}$ (see Appendix A) and the variance is $Var(Y_{hij}|\gamma_{hi}) = \mu_{hij} + \mu_{hij}^2/k_c$, where k_c is the NB dispersion parameter. The marginal distribution of Y_{hij} with respect to the random site effects (γ_{hi}) is not NB. For sites with no replicates (i.e. $n_{hi} = 1$), the marginal distribution of Y_{hi1} is

$$f(Y_{hi1} = y) = \frac{k_s^{k_s} k_c^{k_c} \mu_{hi1}^y \Gamma(y + k_c)}{\Gamma(k_s)\Gamma(k_c)\Gamma(y + 1)} \int_0^\infty \frac{t^{y+k_s-1} e^{-k_s t}}{(\mu_{hi1} t + k_c)^{y+k_c}} \mathrm{d}t,$$

with $E(Y_{hi1}) = \mu_{hi1} = \mu_h \eta_{hi1}$ and $Var(Y_{hi1}) = \mu_{hi1} + \mu_{hi1}^2/k_t$, where $k_t = k_s \cdot k_c/(1 + k_s + k_c)$ (see Appendix A). If there are multiple observations at a site then there will be marginal correlations in these $Y_{hi1}, ..., Y_{hin_{hi}}$ because there is a common γ_{hi} in their distribution. For example, if there are two replicates at a site then the marginal distribution of Y_{hi1} and Y_{hi2} is

$$\begin{aligned} f(Y_{hi1} = y_1, Y_{hi2} = y_2) = & \frac{k_s^{k_s} k_c^{2 \cdot k_c} \mu_{hi1}^{y_1} \mu_{hi2}^{y_2} \Gamma(y_1 + k_c)\Gamma(y_2 + k_c)}{\Gamma(k_s)\Gamma^2(k_c)\Gamma(y_1 + 1)\Gamma(y_2 + 1)} \\ & \cdot \int_0^\infty \frac{t^{y_1+y_2+k_s-1} e^{-k_s t}}{(\mu_{hi1} t + k_c)^{y_1+k_c} (\mu_{hi2} t + k_c)^{y_2+k_c}} \mathrm{d}t, \end{aligned}$$

(See Appendix A for the derivation). The mean of Y_{hij} is $E(Y_{hij}) = \mu_{hij} = \mu_h \eta_{hij}$, the marginal variance is $Var(Y_{hij}) = \mu_{hij} + (\mu_{hij})^2/k_t$, and the $Cov(Y_{hi1}, Y_{hi2}) = \mu_{hi1} \cdot \mu_{hi2}/k_s$. The distribution for more than two replicates can be derived similarly, and the forms of marginal variance and covariance are the same.

The model involves regression parameters $\boldsymbol{\beta}$, variance parameters k_s and k_t, and a large number of nuisance parameters μ_h. There are two main challenges for estimation and statistical inferences about $\boldsymbol{\beta}$. The first challenge is the difficulty in calculating the marginal likelihood function which involves intractable integration. The second challenge is the large number of nuisance parameters which cause bias in the estimation of variance parameters k_s and k_c, regression parameters $\boldsymbol{\beta}$ and their confidence intervals. A biased confidence interval for $\boldsymbol{\beta}$ means the probability that $\boldsymbol{\beta}$ falls in its $(1-\alpha)\%$ confidence interval is not equal to $(1-\alpha)\%$. Such bias is often caused by the biased estimation of the regression and variance parameters.

In the Chapter 2 we use generalized estimating equations (GEEs) to estimate the model parameters. This approach is commonly used for correlated count data (e.g. Paul and Zhang, 2014 [31]). Moreover, we propose a profile GEE approach that is more computationally efficient than the usual approach, especially when there are a large number of nuisance parameters. When we first started this research, GEE seemed like the most promising approach. However, we then learned of new software that made MLE more practical. In Chapter 3, we show how the model can be estimated by marginal MLE. This involves integrating the γ_{hi}'s out of the joint likelihood using a state-of-the-art software package called TMB (e.g. Kristensen, 2013 [21]) that used the Laplace approximation for the marginal likelihood. Compared with the GEE approach, MLE using TMB is easy to implement and the computational speed is much faster. These approaches are two ways that approximations are used to deal with the intractable integration involved in the marginal likelihood function. Neither approach addresses the bias problem caused by many nuisance parameters.

In Chapter 4, we use the REML approach to address the bias in variance parameter estimators and inaccurate confidence intervals for regression parameters because of a large number of nuisance parameters. REML is often considered to be an impractical method for complex non-linear and non-normal estimation problems; however, we can implement it easily with TMB. In Chapter 5 we use a simulation study to compare these three methods: GEE, MLE and REML. We also investigate the impact of different data characteristics (i.e. sample size, number of strata, etc) on the estimation of $\boldsymbol{\beta}$, k_c and k_s. We use ANOVA to help summarize the simulation results.

Chapter 6 involves two applications. In the first application, we estimate the day and night (diel) effect of trawling on three species using GEE, MLE and REML methods. The data were obtained from bottom trawl research surveys. We also compare our results with those obtained in a previously published study. In the second application we estimate the diel (day and night) and vessel effects of two different snow crab surveys conducted in the southern Gulf of St. Lawrence during 2003-2014.

Chapter 2

Generalized estimating equation method

2.1 Introduction

The generalized estimating equation (GEE) method is an extension of generalized linear model (GLM) to correlated (e.g. longitudinal) data (e.g. Liang and Zeger, 1986 [26]), and has origins from the quasi-likelihood methods introduced by Wedderburn (1974) [48] and Nelder and Wedderburn (1972) [29]. In this section we review the GEE method and apply it to our stratified model in the next section.

In the general model framework, we assume there are N clusters observed in a cluster sampling design. For a specific cluster i, we use $\mathbf{y_i} = (y_{i1}, y_{i2}, \cdots, y_{in_i})'$ to denote the vector of responses, and $\mathbf{x}_i$, a $p \times n_i$ matrix to denote the corresponding covariates. The marginal expectation of y_{ij} is μ_{ij}, and is assumed to be a function of the covariates, which can be expressed as $h(\mathbf{x}_{ij}, \boldsymbol{\Theta})$, where h is a known function and $\boldsymbol{\Theta}$ is a $p \times 1$ vector of regression parameter. Observations between clusters are assumed to be independent, but within clusters they are assumed to be correlated with each other.

The GEE functions proposed in Liang and Zeger (1986) [26] for regression parameter $\boldsymbol{\Theta}$ can be written in vector form

$$U(\boldsymbol{\Theta}) = \sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu'}_i(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Theta})), \tag{2.1}$$

where $U(\boldsymbol{\Theta})$ is a $p \times 1$ vector, $\boldsymbol{\mu}_i(\boldsymbol{\Theta}) = (\mu_{i1}(\boldsymbol{\Theta}), \mu_{i2}(\boldsymbol{\Theta}), \cdots \mu_{in_i}(\boldsymbol{\Theta}))'$ is the marginal mean vector for the response of cluster i with $\mu_{ij}(\boldsymbol{\Theta}) = h(\boldsymbol{x}_{ij}, \boldsymbol{\Theta})$, and $\boldsymbol{V}_i$ is the covariance matrix of cluster i. The GEE estimators of $\boldsymbol{\Theta}$ is derived by solving Eq.(2.1) equals to $\mathbf{0}$. This solution can be obtained via the Newton-Raphson method. We start with initial value $\boldsymbol{\Theta}^{(0)}$. The updating algorithm we use to estimate $\boldsymbol{\Theta}$ is

$$\boldsymbol{\Theta}^{(j+1)} = \boldsymbol{\Theta}^{(j)} - \left\{ \left. \frac{\partial U(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}'} \right|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(j)}} \right\}^{-1} U(\boldsymbol{\Theta}^{(j)}).$$

There are two advantages with using the GEE method. Firstly, the GEE method doesn't involve the marginal likelihood, which is often impossible to obtain analytically because of intractable integrals. These integrals can be difficult to compute numerically and this may also lead to estimation problems when using nonlinear optimization methods because some numerical integration methods can introduce sharp irregularities in the likelihood surface. The second advantage is that GEE estimators of regression parameters are consistent when the mean structure ($\boldsymbol{\mu_i}$, $i = 1, 2, \cdots, N$) is correctly specified even if the covariance matrix ($\boldsymbol{V_i}$, $i = 1, 2, \cdots, N$) is mis-specified (Wang and Carey, 2004 [45]). However, a disadvantage of GEE is that it does require calculation of the marginal mean and covariance which may be difficult in some cases.

In this chapter, we develop a GEE method to estimate model parameters due to the challenge of deriving the marginal likelihood function for the Poisson-double-Gamma (PdG) mixture model. We develop a profile GEE method that is computationally more efficient than the standard GEE method.

2.2 Profile Generalized Estimating Equation

We develop a GEE method to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ for the stratified count data model. Recall from Chapter 1 that we use $\boldsymbol{y_{hi}} = (y_{hi1}, y_{hi2}, \cdots y_{hin_{hi}})'$ to denote the vector of responses at site i in stratum h, and $\boldsymbol{x_{hi}}$ to denote the corresponding covariates, which is a $p \times n_{hi}$ matrix. The marginal expectation of y_{hij} is $\mu_{hij} = \exp(\alpha_h + \sum_{k=1}^{p} \beta_k x_{hijk})$ where $\alpha_h = \log(\mu_h)$. Let $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_H)'$, $\Theta = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ and $\boldsymbol{\mu_{hi}} = (\mu_{hi1}, \mu_{hi2}, \cdots, \mu_{hin_{hi}})'$. The GEE functions for Θ can be written in vector form,

$$U(\Theta) = \sum_{h,i} \frac{\partial \boldsymbol{\mu_{hi}}'(\Theta)}{\partial \Theta} \boldsymbol{V}_{hi}^{-1} \{\boldsymbol{y_{hi}} - \boldsymbol{\mu_{hi}}(\Theta)\}, \tag{2.2}$$

where $U(\Theta)$ is a $(H+p) \times 1$ vector. $\boldsymbol{V}_{hi}$ is the covariance matrix of $\boldsymbol{y}_{hi}$ in which the m'th diagonal element $\boldsymbol{V}_{hi}(m,m)$ is the variance of Y_{him} and the m,n'th element is the covariance between Y_{him} and Y_{hin}.

$$\begin{aligned} \boldsymbol{V}_{hi}(m,m) &= \mu_{him} + \frac{\mu_{him}^2}{\hat{k}_t} \\ &= \exp(\sum_{k=1}^{p} \beta_k x_{himk} + \alpha_h) + \frac{\exp(2\sum_{k=1}^{p} \beta_k x_{himk} + 2\alpha_h)}{\hat{k}_t}, \end{aligned} \tag{2.3}$$

$$\begin{aligned} \boldsymbol{V}_{hi}(m,n) &= \frac{\mu_{him}\mu_{hin}}{\hat{k}_s} \\ &= \frac{\exp(\sum_{k=1}^{p} \beta_k x_{himk} + \sum_{k=1}^{p} \beta_k x_{hink} + 2\alpha_h)}{\hat{k}_s}, \end{aligned} \tag{2.4}$$

where $\hat{k}_s$ and $\hat{k}_t$ are estimates of the variance and correlation parameters k_t and k_s (see Section 2.3). We solve Eq.(2.2) equals to $\mathbf{0}$ via the Newton-Raphson method to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Starting with initial value $\Theta^{(0)}$, the updating algorithm that can be used to find this solution is

$$\Theta^{(j+1)} = \Theta^{(j)} - \left\{ \left. \frac{\partial U(\Theta)}{\partial \Theta'} \right|_{\Theta=\Theta^{(j)}} \right\}^{-1} \cdot U(\Theta^{(j)}).$$

The GEE function in Eq.(2.2) is inefficient when Θ is large. Since $\boldsymbol{\alpha}$ is not of immediate interest, we treat them as nuisance parameters. In likelihood-based estimation we often use the profile likelihood approach to get an approximate likelihood function just for the parameter of interest, in which we replace the nuisance parameters with their maximum likelihood estimators when the main parameters of interest are temporally assumed to be known. Let θ denote the parameter of interest, λ denote the nuisance parameter and S denote the data. The the profile likelihood for θ is

$$\hat{L}(\theta|S) = \sup_{\lambda} L(\theta, \lambda|S).$$

We use the same idea with the GEE method to deal with the nuisance parameters. We replace $\boldsymbol{\alpha}$ in the estimating function by its estimator conditional on $\boldsymbol{\beta}$, which we denote as $\bar{\boldsymbol{\alpha}} = \bar{\boldsymbol{\alpha}}(\boldsymbol{\beta})$. This leads to the profile generalized estimating equation for $\boldsymbol{\beta}$,

$$U(\boldsymbol{\beta}) = \sum_{h,i} \frac{\partial \boldsymbol{\mu_{hi}}'\{\boldsymbol{\beta}; \bar{\boldsymbol{\alpha}}(\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}} \boldsymbol{V}_{hi}^{-1}[\boldsymbol{y_{hi}} - \boldsymbol{\mu_{hi}}'\{\boldsymbol{\beta}; \bar{\boldsymbol{\alpha}}(\boldsymbol{\beta})\}].$$

This is a $p \times 1$ estimating equation whereas Equation (2.2) is $(H + p) \times 1$ and when H is large the difference in the number of estimating equations to solve is large.

Let $\boldsymbol{\beta}^{(0)}$ be the initial value of $\boldsymbol{\beta}$. The algorithm for estimating the regression parameters is to iterate between the following steps until convergence is achieved:

***Step* 1**. Treating $\boldsymbol{\beta}^{(k)}$ as fixed and known, estimate $\bar{\boldsymbol{\alpha}}(\boldsymbol{\beta}^{(k)})$ by solving $U(\boldsymbol{\alpha}; \boldsymbol{\beta}^{(k)}) = \mathbf{0}$

for $\boldsymbol{\alpha}$, where $U(\boldsymbol{\alpha};\boldsymbol{\beta}^{(k)}) = [U(\alpha_1;\boldsymbol{\beta}^{(k)}), U(\alpha_2;\boldsymbol{\beta}^{(k)}), \cdots U(\alpha_H;\boldsymbol{\beta}^{(k)})]'$ and

$$U(\alpha_h;\boldsymbol{\beta}^{(k)}) = \sum_i \frac{\partial \boldsymbol{\mu'}_{hi}(\boldsymbol{\beta}^{(k)};\alpha_h)}{\partial \alpha_h} \boldsymbol{V}_{hi}^{-1}\{\boldsymbol{y}_{hi} - \boldsymbol{\mu}_{hi}(\boldsymbol{\beta}^{(k)};\alpha_h)\}. \tag{2.5}$$

***Step* 2**. Estimate $\boldsymbol{\beta}^{(k+1)}$ by solving

$$U(\boldsymbol{\beta}) = \sum_{h,i} \frac{\partial \boldsymbol{\mu'}_{hi}\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}} \boldsymbol{V}_{hi}^{-1}[\boldsymbol{y}_{hi} - \boldsymbol{\mu}_{hi}\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\}] = \mathbf{0}. \tag{2.6}$$

Note that $\partial \boldsymbol{\mu'}_{hi}(\alpha_h;\boldsymbol{\beta})/\partial \alpha_h = \boldsymbol{\mu'}_{hi}(\alpha_h;\boldsymbol{\beta})$ for all h and $\boldsymbol{\beta}$ so that Eq.(2.5) is

$$U(\alpha_h;\boldsymbol{\beta}^{(k)}) = \sum_i \boldsymbol{\mu'}_{hi}(\alpha_h;\boldsymbol{\beta}^{(k)}) \boldsymbol{V}_{hi}^{-1}\{\boldsymbol{y}_{hi} - \boldsymbol{\mu}_{hi}(\alpha_h;\boldsymbol{\beta}^{(k)})\}. \tag{2.7}$$

We use Eq.(2.7) and $U\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\} = 0$ to derive $\partial\bar{\alpha}_h(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$. Differentiating both sides with respect to $\boldsymbol{\beta}$ and using the chain rule, we obtain

$$\frac{\partial U\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}} = \mathbf{0} = \left.\frac{\partial U(\boldsymbol{\beta};\alpha_h)}{\partial \boldsymbol{\beta}}\right|_{\alpha_h=\bar{\alpha}_h(\boldsymbol{\beta})} + \left.\frac{\partial U(\boldsymbol{\beta};\alpha_h)}{\partial \alpha_h}\right|_{\alpha_h=\bar{\alpha}_h(\boldsymbol{\beta})} \cdot \frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Rightarrow$$
$$\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\left\{\left.\frac{\partial U(\boldsymbol{\beta};\alpha_h)}{\partial \alpha_h}\right|_{\alpha_h=\bar{\alpha}_h(\boldsymbol{\beta})}\right\}^{-1} \cdot \left.\frac{\partial U(\boldsymbol{\beta};\alpha_h)}{\partial \boldsymbol{\beta}}\right|_{\alpha_h=\bar{\alpha}_h(\boldsymbol{\beta})}$$

Hence, the term $\partial \boldsymbol{\mu'}_{hi}\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\}/\partial\boldsymbol{\beta}$ of the estimating equation of $\boldsymbol{\beta}$ in Eq.(2.6) is

$$\begin{aligned}\frac{\partial \boldsymbol{\mu'}_{hi}\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}} = \quad & \left[\mu_{hi1}\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\} \cdot \left\{\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hi1}\right\}, \mu_{hi2}\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\} \cdot \left\{\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hi2}\right\},\right. \\ & \left.\cdots, \mu_{hin_{hi}}\{\boldsymbol{\beta};\bar{\alpha}_h(\boldsymbol{\beta})\} \cdot \left\{\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hin_{hi}}\right\}\right].\end{aligned}$$

The algorithm for $\boldsymbol{\beta}$ in ***Step* 2** (Eq. 2.6) we use is

$$\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(j)} - \left\{\left.\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(j)}}\right\}^{-1} \cdot U(\boldsymbol{\beta}^{(j)}).$$

(See Appendix B.4 for the initial value $\boldsymbol{\beta}^{(0)}$.) The algorithm for $\boldsymbol{\alpha}$ in ***Step* 1** (Eq. 2.5) is

$$\boldsymbol{\alpha}^{(j+1)} = \boldsymbol{\alpha}^{(j)} - \left\{ \frac{\partial U(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}'} \bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(j)}} \right\}^{-1} \cdot U(\boldsymbol{\alpha}^{(j)}).$$

A more detailed description of the profile GEE is given in the Appendix B.1.

The profile GEE approach is more efficient to compute than the general GEE approach. ***Step* 2** in the profile GEE procedure often took more than 10 iterations to solve for $\boldsymbol{\beta}$, while ***Step* 1** took at most 4 iterations to solve for $\boldsymbol{\alpha}$ with the same convergence accuracy. Compared to the GEE in Eq.(2.2), the profile GEE was much more efficient because solving the $H + p$ dimensional GEE takes $(10 \times H + 10 \times p)$ steps or more, whereas for profile GEE it takes $(4 \times H + 10 \times p)$. When H is really large this makes a big difference.

2.3 Covariance parameters estimation

In the PdG mixture model we have two covariance parameters k_c and k_s (see Eq.(2.3) and Eq.(2.4)). Reliable estimation of the variance parameter k_c and the correlation parameter k_s is fairly important since the efficiency of the GEE estimator depends on how closely the estimated covariance structure approximates the true covariance structure (e.g. Crowder, 1995 [16]). We estimate $k_t = k_s \cdot k_c/(1 + k_s + k_c)$ instead of

k_c for simplicity, since k_t is the leading variance parameter and is a combination of k_s and k_c. Direct estimation of k_c is more complicated.

Some GEE methods have been proposed to estimate the covariance parameters. Given the regression parameter estimates $\hat{\boldsymbol{\beta}}$, the GEE method proposed by Prentice (1988) [33] is

$$U(\boldsymbol{\theta}) = \sum_{h,i} \frac{\partial \boldsymbol{\tau_{h,i}}'}{\partial \boldsymbol{\theta}} cov^{-1}(\boldsymbol{w_{h,i}})(\boldsymbol{w_{h,i}} - \boldsymbol{\tau_{h,i}}), \tag{2.8}$$

where $\boldsymbol{\theta} = (k_t, k_s)'$, $\boldsymbol{w_{hi}} = (r^2_{hi1}, r^2_{hi2}, \cdots, r^2_{hin_{hi}}, r_{hi1}r_{hi2}, \cdots r_{hi1}r_{hin_{hi}}, r_{hi2}r_{hi3}, \cdots r_{hi(n_{hi}-1)}r_{hin_{hi}})'$, $r_{hik} = y_{hik} - \mu_{hik}(\hat{\boldsymbol{\beta}})$ and $\boldsymbol{\tau_{hi}} = E(\boldsymbol{w_{hi}})$. The GEE estimator of $\boldsymbol{\theta}$ is derived by solving $U(\boldsymbol{\theta}) = \mathbf{0}$. This solution can be obtained via the Newton-Raphson Method. Extensions are the GEE1 (Liang *el al.*, 1992 [27]) and GEE2 (Zhao and Prentice, 1990 [50]) methods. Both GEE1 and GEE2 estimate regression and covariance parameters simultaneously. The GEE1 is

$$U_1(\boldsymbol{\beta};\boldsymbol{\theta}) = \sum_{h,i} \begin{pmatrix} \frac{\partial \boldsymbol{\mu}'_{hi}}{\partial \beta} & 0 \\ 0 & \frac{\partial \boldsymbol{\tau}'_{hi}}{\partial \boldsymbol{\theta}} \end{pmatrix} \begin{pmatrix} cov(\boldsymbol{y_{hi}}) & 0 \\ 0 & cov(\boldsymbol{w_{hi}}) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{y_{hi}} - \boldsymbol{\mu_{hi}} \\ \boldsymbol{w_{hi}} - \boldsymbol{\tau_{hi}} \end{pmatrix}.$$

where $\boldsymbol{\beta}$ is the regression parameter and $\boldsymbol{\mu}_{hi}$ is the marginal expectation of $\boldsymbol{y}_{hi}$, $\boldsymbol{\theta}$, $\boldsymbol{w_{hi}}$ and $\boldsymbol{\tau_{hi}}$ are the same as defined in Eq.(2.8). The GEE2 is

$$U_2(\boldsymbol{\beta};\boldsymbol{\theta}) = \sum_{h,i} \begin{pmatrix} \frac{\partial \boldsymbol{\mu}'_{hi}}{\partial \beta} & 0 \\ \frac{\partial \boldsymbol{\tau}'_{hi}}{\partial \beta} & \frac{\partial \boldsymbol{\tau}'_{hi}}{\partial \boldsymbol{\theta}} \end{pmatrix} \begin{pmatrix} cov(\boldsymbol{y_{hi}}) & cov(\boldsymbol{y_{hi}}, \boldsymbol{w_{hi}}) \\ cov(\boldsymbol{w_{hi}}, \boldsymbol{y_{hi}}) & cov(\boldsymbol{w_{hi}}) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{y_{hi}} - \boldsymbol{\mu_{hi}} \\ \boldsymbol{w_{hi}} - \boldsymbol{\tau_{hi}} \end{pmatrix}.$$

GEE1 and GEE2 estimators can be obtained via the Newton-Raphson method. If the mean and covariance structures are correctly specified, then GEE2 estimators are more efficient than GEE1 estimators and are nearly as efficient as MLE's (Liang, Zeger and Quash, 1992 [27]). However, both GEE1 and GEE2 are difficult to solve for the PdG model, due to the difficulty in constructing the covariance matrix (Sutradhar, 2003 [43]) and the computational difficulties when there are many replicates in one site. For example, if there are 6 replicates at one site, $cov(\boldsymbol{w_{hi}})$ would be a 15×15 matrix which involves mixed moments of order four (e.g. $E(r_{hi1}r_{hi2}r_{hi3}r_{hi4})$).

Many other methods have been proposed to estimate the covariance parameters. Chaganty (1997) [14] proposed a quasi-least squares (QLS) method to estimate the correlation parameter k_s. However, these correlation parameter estimators of this method are always biased (e.g. Wang and Carey, 2004 [45]). Wang and Carey (2004) [45] proposed a pseudo-likelihood method to estimate k_s and this method corrects the bias of QLS estimator. Moreover, Wang and Zhao (2007) [47] proposed a modified pseudo-likelihood approach to estimate the variance parameter k_t. The pseudo-likelihood approaches are preferable to us than the GEE1 and GEE2 methods since they don't involve the third and fourth order moments of response and the mixed moments of response.

Zhang and Paul (2013) [49] studied a GEE method for variance parameter estimation based on the squared residual regression method (i.e. Crowder, 1995 [16]). They showed that this estimator is at least as efficient as the modified pseudo-likelihood estimator. Hence, we use this method to estimate the variance parameter $\tau = 1/k_t$. Given the regression parameter estimates $\hat{\boldsymbol{\beta}}$, the GEE function for the variance parameter of our model is

$$U(\tau) = \sum_{h,i} \frac{\partial \nu\{\tau, \boldsymbol{\mu}_{hi}(\hat{\boldsymbol{\beta}})\}}{\partial \tau}^T \boldsymbol{V_{hi}}^{-1}(\tau)[\boldsymbol{r_{hi}}^2 - \nu\{\tau, \boldsymbol{\mu}_{hi}(\hat{\boldsymbol{\beta}})\}] \tag{2.9}$$

where $\boldsymbol{r_{hi}} = \boldsymbol{y_{hi}} - \boldsymbol{\mu}_{hi}(\hat{\boldsymbol{\beta}})$, $\nu\{\tau, \boldsymbol{\mu}_{hi}(\hat{\boldsymbol{\beta}})\} = \boldsymbol{\mu}_{hi}(\hat{\boldsymbol{\beta}}) + \tau\boldsymbol{\mu}^2_{hi}(\hat{\boldsymbol{\beta}})$ and $\boldsymbol{V_{hi}}(\tau)$ is a diagonal matrix which equals to $var(\boldsymbol{r_{ij}}^2)$. Third and fourth order moments of $\boldsymbol{y_{hi}}$ are required to calculate $var(\boldsymbol{r_{ij}}^2)$. We approximate these moments using NB moments (See Appendix B.2). Hence, the consistency of the estimate of τ should not be affected (e.g. Zhang and Paul, 2013 [49]). The estimator of τ is derived by solving Eq.(2.9) equals to 0 via Newton-Raphson method (See Appendix B.2). This GEE approach is preferable to us than the GEE1 and GEE2 methods since it doesn't involve the mixed moments.

This GEE method (e.g. Zhang and Paul, 2013 [49]) can't be used to estimate the correlation parameter k_s. However, we can still use pseudo-likelihood to estimate k_s. We could have used the pseudo-likelihood approach for k_t; however, in preliminary investigation we found that this approach was not as good as the GEE proposed by Zhang and Paul (2013) [49]. This is why we use the 2 different approaches for k_t and k_s. Wang and Carey (2004) [45] demonstrated good accuracy of pseudo-likelihood estimators of variance and correlation parameters for incomplete Gaussian measurements, clustered lognomal, and clustered Poisson data. We use this method to estimate the correlation parameter. The covariance matrix $\boldsymbol{V}_{hi}$ in Eq.(2.2) can be decomposed as

$$\boldsymbol{V}_{hi} = \boldsymbol{A}_{hi}^{\frac{1}{2}}(k_t)\boldsymbol{R}_{hi}(k_t, k)\boldsymbol{A}_{hi}^{\frac{1}{2}}(k_t) \tag{2.10}$$

(See Appendix B.2 for the decomposition). Let $\xi = 1/k_s$. Given estimates of regression parameters $\hat{\boldsymbol{\beta}}$ and variance parameter $\hat{\tau}$, the pseudo-likelihood function for ξ is

$$U(\xi) = \sum_{h,i} \text{trace}[\boldsymbol{P}_{hi}(\hat{\tau}, \xi)\{\boldsymbol{\epsilon_{hi}}(\hat{\tau})\boldsymbol{\epsilon_{hi}}'(\hat{\tau}) - \boldsymbol{R_{hi}}(\hat{\tau}, \xi)\}] \tag{2.11}$$

where $\boldsymbol{\epsilon_{hi}}(\hat{\tau}) = \boldsymbol{A_{hi}}^{-1/2}(\hat{\tau})\{\boldsymbol{y_{hi}} - \boldsymbol{\mu_{hi}}(\hat{\boldsymbol{\beta}})\}$ and

$\boldsymbol{P_{hi}}(\hat{\tau}, \xi) = \boldsymbol{R_{hi}}^{-1}(\hat{\tau}, \xi)\,\{\partial \boldsymbol{R_{hi}}(\hat{\tau}, \xi)/\partial\xi\}\,\boldsymbol{R_{hi}}^{-1}(\hat{\tau}, \xi)$ (See Appendix B.3). The estimator of ξ is derived by solving Eq.(2.11) equals to 0 via Newton-Raphson method.

The algorithm to estimate $\boldsymbol{\beta}$, τ and ξ is:

1. Start with the initial value $\boldsymbol{\beta}^{(0)}$, $\tau^{(0)}$ and $\xi^{(0)}$.
2. Given $\boldsymbol{\beta}^{(j)}$, $\tau^{(j)}$ and $\xi^{(j)}$, obtain an estimate $\tau^{(j+1)}$ using Eq.(2.9);
3. Use $\boldsymbol{\beta}^{(j)}$, $\xi^{(j)}$ and the updated $\tau^{(j+1)}$ to obtain an estimate $\xi^{(j+1)}$ using Eq.(2.11);
4. Use the updated $\tau^{(j+1)}$ and $\xi^{(j+1)}$ to obtain an estimate $\boldsymbol{\beta}^{(j+1)}$ using Eq.(2.5) and

(2.6);

5. Repeat the process until convergence is achieved.

2.4 Variance estimate

Under mild conditions, the GEE regression estimators are consistent and asymptotically normally distributed, that is $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, V_{\hat{\boldsymbol{\beta}}})$ asymptotically. The variance of $\hat{\boldsymbol{\beta}}$ can be estimated via the 'sandwich estimator' proposed by Liang and Zeger (1986) [26]. The sandwich estimator has the following form

$$V_{\hat{\boldsymbol{\beta}}} = M_0^{-1} M_1 M_0^{-1}, \tag{2.12}$$

where

$$M_0 = \sum_{h,i} D'_{hi}(\hat{\boldsymbol{\beta}}) V_{hi}(\hat{\boldsymbol{\beta}})^{-1} D_{hi}(\hat{\boldsymbol{\beta}}),$$

$$M_1 = \sum_{h,i} D'_{hi}(\hat{\boldsymbol{\beta}}) V_{hi}(\hat{\boldsymbol{\beta}})^{-1} (\boldsymbol{y}_{hi} - \boldsymbol{\mu}_{hi}(\hat{\boldsymbol{\beta}}))(\boldsymbol{y}_{hi} - \boldsymbol{\mu}_{hi}(\hat{\boldsymbol{\beta}}))' V_{hi}(\hat{\boldsymbol{\beta}})^{-1} D_{hi}(\hat{\boldsymbol{\beta}}),$$

and $D_{hi}(\boldsymbol{\beta}) = \partial \boldsymbol{\mu_{hi}(\beta)} / \partial \boldsymbol{\beta}$, $\boldsymbol{\mu}_{hi}$ and V_{hi} are the marginal mean and variance of $\boldsymbol{y}_{hi}$.

Unfortunately the profile GEE estimators for $\boldsymbol{\beta}$, k_t, and k_s are fairly complicated to implement and not simple to modify. In the next chapter we will decribe an approach that is much easier to implement and modify.

Chapter 3

Marginal maximum likelihood using TMB

In Chapter 2 we used a GEE approach to estimate the regression and variance parameters of the PdG mixture model. The main benefit of the GEE approach is that it doesn't involve the marginal likelihood, which may involve some intractable integration. However, the GEE method is difficult for the PdG model due to the following issues:

1. The profile GEE involves difficult calculations and tedious programming.

2. The simulation speed is too slow (See Table 5.9 in Chapter 5).

In this chapter we review the penalized quasi-likelihood approach and the marginal maximum likelihood approach using TMB to estimate the PdG model. We do not implement penalized quasi-likelihood estimators but we do use the approach to motivate our marginal maximum likelihood approach.

3.1 Penalized Quasi-likelihood

Breslow and Clayton (1993) [9] proposed the penalized quasi-likelihood (PQL) method to estimate generalized linear mixed models (GLMM) with normal random effects. They used the Laplace approximation (see Section 3.2.2) to derive the marginal quasi-likelihood.

Assume there are N clusters observed in a cluster sampling design. We use $\boldsymbol{y} = (y_1, y_2, \cdots, y_N)'$ to denote the responses, $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N)'$ to denote the corresponding covariates, and $\boldsymbol{\beta}$ to denote the regression parameters. We assume a vector $\boldsymbol{b}$ of normal random effects $\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{D}(\boldsymbol{\theta}))$, where $\boldsymbol{\theta}$ is the variance parameters. The conditional expectation of $\boldsymbol{y}$ given $\boldsymbol{b}$ is assumed to be

$$E(\boldsymbol{y}|\boldsymbol{b}) = \boldsymbol{\mu}^{\boldsymbol{b}} = h(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}),$$

where $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N)'$ is the corresponding covariates. The conditional variance of $\boldsymbol{y}$ given $\boldsymbol{b}$ is assumed to be

$$Var(\boldsymbol{y}|\boldsymbol{b}) = \phi \cdot v(\boldsymbol{\mu^b}).$$

where ϕ is a dispersion parameter which could be known or not known and $v(.)$ is a known variance function.

The integrated quasi-likelihood function for $(\boldsymbol{\beta}, \boldsymbol{\theta})$ is defined by

$$e^{ql(\boldsymbol{\beta},\boldsymbol{\theta})} \propto |\boldsymbol{D}|^{-1/2} \int \exp\{\boldsymbol{\kappa(b)}\} d\boldsymbol{b}, \tag{3.1}$$

where

$$\boldsymbol{\kappa(b)} = -\frac{1}{2\phi} \sum_{i=1}^{N} d_i(y_i; \mu_i^b) - \frac{1}{2} \boldsymbol{b^t D^{-1} b},$$

and

$$d_i(y_i; \mu_i) = -2 \int_{y_i}^{\mu_i} \frac{y_i - \mu}{v(\mu)} d\mu$$

denotes the deviance measure of fit. In Eq.(3.1), $ql(\boldsymbol{\beta}, \boldsymbol{\theta})$ denotes the log quasi-likelihood for the data. The main difficulty lies in the integration of the random effects in the deviance function.

Breslow and Clayton (1993) [9] used Laplace approximation for the integral in Eq.(3.1) which yields

$$ql(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\boldsymbol{D}| - \frac{1}{2} \log |\boldsymbol{\kappa''(\tilde{b})}| - \boldsymbol{\kappa(\tilde{b})}, \tag{3.2}$$

where $\tilde{\boldsymbol{b}} = \tilde{\boldsymbol{b}}(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the solution to

$$\boldsymbol{\kappa}'(\boldsymbol{b}) = -\sum_{i=1}^{N} \frac{(y_i - \mu_i^{\boldsymbol{b}})\boldsymbol{z_i}}{\phi v(\mu_i^{\boldsymbol{b}}) g'(\mu_i^{\boldsymbol{b}})} + \boldsymbol{D}^{-1}\boldsymbol{b} = \boldsymbol{0}$$

that minimize $\boldsymbol{\kappa}(\boldsymbol{b})$, where $g = h^{-1}$.

$$\begin{aligned} \boldsymbol{\kappa}''(\boldsymbol{b}) &= \boldsymbol{Z}^t \boldsymbol{W} \boldsymbol{Z} + \boldsymbol{D}^{-1} + \boldsymbol{R} \\ &\approx \boldsymbol{Z}^t \boldsymbol{W} \boldsymbol{Z} + \boldsymbol{D}^{-1}, \end{aligned} \tag{3.3}$$

where $\boldsymbol{W}$ is a $N \times N$ diagonal matrix with diagonal terms $w_i = \{\phi v(\mu_i^b)[\frac{\partial h^{-1}(\mu)}{\partial \mu}|_{\mu=\mu_i^b}]^2\}^{-1}$.

The remainder term

$$\boldsymbol{R} = -\sum_{i=1}^{N} (y_i - \mu_i^{\boldsymbol{b}})\boldsymbol{z_i} \frac{\partial}{\partial \boldsymbol{b}} \left[\frac{1}{\phi v(\mu_i^{\boldsymbol{b}}) g'(\mu_i^{\boldsymbol{b}})} \right]$$

has expectation 0 and is of lower order than the two leading terms in Eq.(3.3). Combining (3.1)-(3.3) and ignoring $\boldsymbol{R}$ leads to

$$ql(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx -\frac{1}{2} \log |\boldsymbol{I} + \boldsymbol{Z^t W Z D}| - \frac{1}{2\phi} \sum_{i=1}^{N} d_i(y_i; \mu_i^{\tilde{\boldsymbol{b}}}) - \frac{1}{2} \tilde{\boldsymbol{b}}^{\boldsymbol{t}} \boldsymbol{D^{-1}} \tilde{\boldsymbol{b}}. \tag{3.4}$$

In the estimation procedure, we first estimate $\tilde{\boldsymbol{b}} = \tilde{\boldsymbol{b}}(\boldsymbol{\beta}, \boldsymbol{\theta})$ for fixed $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ by solving $\boldsymbol{\kappa}'(\tilde{\boldsymbol{b}}) = \boldsymbol{0}$, then we obtain $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ for fixed $\boldsymbol{\theta}$ by maximizing the last two terms of Eq.(3.4)

$$-\frac{1}{2\phi} \sum_{i=1}^{N} d_i(y_i; \mu_i^{\tilde{\boldsymbol{b}}}) - \frac{1}{2} \tilde{\boldsymbol{b}}^{\boldsymbol{t}} \boldsymbol{D^{-1}} \tilde{\boldsymbol{b}}.$$

We obtain $\hat{\boldsymbol{\theta}}$ by maximizing $ql(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta})$ or $ql_1(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta})$ (see Eq.(4.1) in Chapter 4). The PQL estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$.

The PQL method is not directly applicable to the PdG model because the random effects of PQL are assumed to be normally distributed while the random effects of PdG model are gamma distributed. We could define a normal random effect $b \sim N(0,1)$, a gamma random effect $\gamma \sim \Gamma(\alpha_1, \alpha_2)$ and use F_1 and F_2, the corresponding CDFs, to model γ as a function of b. We can then modify the PQL for gamma random effects by replacing b with γ in the model, and use the random effect link function

$$F_2^-[F_1(b)] \sim \Gamma(\alpha_1, \alpha_2).$$

(e.g. Robert and Casella, 2013 [35]):

If X is a continuous random variable with CDF F, then the random variable $U = F(X)$ follows a uniform distribution on $[0,1]$. In converse, if U has a uniform distribution on $[0,1]$, F^- is the generalized inverse of F, then $F^-(U)$ has distribution F.

However, in our case the PQL involves difficult calculations (such as $\tilde{\boldsymbol{b}}$) and is difficult to implement. In the next Section, we will investigate a marginal maximum likelihood approach to estimate the PdG model. This approach involves integrating the gamma random effects out of the joint likelihood function using the Laplace approximation similar to PQL; however, we use TMB, a new software package, that is developed for such situations.

3.2 TMB: Automatic differentiation and Laplace approximation

Template Model Builder (TMB; e.g. Thorson *et al.*, 2014 [44]) is a free and open source R package (e.g. R Core Team, 2014 [34]) that is designed for estimating complex nonlinear models that may include random effects. The user only has to define the joint log-likelihood function of the data and (i.e. conditional on) the random effects as a C++ template function. Other operations such as integration and calculation of the marginal score function are done in R.

3.2.1 Automatic differentiation

Automatic Differentiation (AD; e.g. Fournier *et al.*, 2012 [18]), also known as Computational Differentiation or Algorithmic Differentiation, is a set of techniques that numerically differentiates a function, which frees us from calculating and incorporating the derivatives. Two methods, "source transformation" and "operator overloading" are commonly used to implement automatic differentiation. CppAD (e.g. Bell, 2012 [4]) implements the operator overloading approach which is easier to implemente and use compared with "source transformation". The TMB R package uses CppAD to provide up to third order derivatives of the joint log-likelihood function that the user writes in the C++ template (see Appendix C). These derivatives are required for the Laplace approximation of the marginal likelihood.

3.2.2 Laplace Approximation

The Laplace approximation (e.g. Skaug and Fournier, 2006 [41]) is used to approximate the intractable integral in the marginal likelihood (Eq. (3.5)). Let $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)'$ be the vector of response variables, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_i)'$ be the vector of latent random effects, and let $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)'$ be the vector of parameters (fixed effects). Let $f_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{\lambda})$ denote the conditional probability density function of $\boldsymbol{y}$ given $\boldsymbol{\lambda}$, and let $g_{\boldsymbol{\theta}}(\boldsymbol{\lambda})$ denote the marginal probability density function of the random effects $\boldsymbol{\lambda}$. The marginal likelihood function for $\boldsymbol{\theta}$ is defined by integrating out the random effects $\boldsymbol{\lambda}$ from $f_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{\lambda})g_{\boldsymbol{\theta}}(\boldsymbol{\lambda})$,

$$L(\boldsymbol{\theta}) = \int f_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{\lambda})g_{\boldsymbol{\theta}}(\boldsymbol{\lambda})d\boldsymbol{\lambda} = \int \exp\{h(\boldsymbol{\lambda}, \boldsymbol{\theta})\}d\boldsymbol{\lambda}, \tag{3.5}$$

where

$$h(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \log\{f_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{\lambda})\} + \log\{g_{\boldsymbol{\theta}}(\boldsymbol{\lambda})\}$$

is the joint penalized log-likelihood of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. The main computational challenge is in computing the integral in Eq.(3.5) when there is no analytical solution. TMB uses the Laplace approximation in Eq.(3.5), which yields the marginal likelihood approximation

$$L^*(\boldsymbol{\theta}) = \det\{H(\boldsymbol{\theta})\}^{-1/2} \exp[h\{\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta}), \boldsymbol{\theta}\}], \tag{3.6}$$

where

$$\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta}) = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}}\, h\{\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta}), \boldsymbol{\theta}\},$$

$$H(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial \boldsymbol{\lambda}^2} h(\boldsymbol{\lambda}, \boldsymbol{\theta})|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta})},$$

and $\det\{H(\boldsymbol{\theta})\}$ denotes the determinant of $H(\boldsymbol{\theta})$. The term $\exp[h\{\hat{\boldsymbol{\lambda}}(\boldsymbol{\theta}), \boldsymbol{\theta}\}]$ in Eq.(3.6) is a profile likelihood, which treats the random effects $\boldsymbol{\lambda}$ as nuisance parameters and $\boldsymbol{\theta}$ as the parameters of interest. The hessian, H, is evaluated by CppAD. Using the AD and Laplace approximation greatly simplifies the parameter estimation of hierarchical models. The TMB user just needs to specify the joint log-likelihood function $h(\boldsymbol{\lambda}, \boldsymbol{\theta})$. TMB uses the Cholesky decomposition of $H(\boldsymbol{\theta})$; therefore, the Laplace approximation is well defined only if $H(\boldsymbol{\theta})$ is positive definite.

In an R session, we read the data, dynamically link the C++ function template, set up the initial values for $\boldsymbol{\theta}$, specify the random effects, and optimize the objective function. TMB automatically provides a standard error report for $\hat{\boldsymbol{\theta}}$, and also any differentiable function of $\boldsymbol{\theta}$, $\phi(\boldsymbol{\theta})$ that the user specifies, by using the δ-method

$$Var(\phi(\hat{\boldsymbol{\theta}})) = -\left\{ \frac{\partial \phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \left[\frac{\partial^2 \{\log L^*(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} \frac{\partial \phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{3.7}$$

3.2.3 Model implementation

In this section we describe the PdG model implementation in TMB. Recall from Chapter 1 that the conditional distribution of the response variable Y_{hij} given random effects γ_{hi} is Negative Binomial distributed

$$f(Y_{hij} = y|\gamma_{hi}) = \frac{\Gamma(y + k_c)}{\Gamma(k_c)\Gamma(y + 1)} \left(\frac{\mu_{hij} \cdot \gamma_{hi}}{\mu_{hij} \cdot \gamma_{hi} + k_c} \right)^y \left(\frac{k_c}{\mu_{hij} \cdot \gamma_{hi} + k_c} \right)^{k_c}$$

where $\mu_{hij} = \exp(\alpha_h + \boldsymbol{x}'_{hij} \cdot \boldsymbol{\beta})$. The random effects γ_{hi} are Gamma distributed with density function

$$f(\gamma_{hi} = t) = \frac{1}{\Gamma(k_s)\frac{1}{k_s^{k_s}}} t^{k_s - 1} e^{-k_s t}$$

For convenience we estimate the logarithm of k_s and k_c which are $(-\infty, \infty)$ whereas k_s and k_c are $(0, \infty)$. The fixed effects model parameters are $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \log(k_s), \log(k_c))$, and the vector of latent random effects are $\boldsymbol{\lambda} = \log(\boldsymbol{\gamma})$.

We first specify the joint likelihood function in the C++ template (see Appendix C). TMB then calculates the marginal likelihood function using the Laplace approximation. The final step is to optimize this objective function in R.

Below is the operations we use in an R session:

```
[1]     library(TMB)

[2]     compile("ML.cpp")

[3]     dyn.load("ML")
```

```
 parameters <- list(log_kc=1,
                    log_k=0,
                    log_eta=rep(0,tmb.data$nstratum),
                    beta=rep(0,5),
                    log_site=rep(0,tmb.data$nsitep))

parameters.U <- list(
                   beta=rep(Inf,5),
                   log_k=2.3,
                   log_kc=3.4,
                   log_eta=rep(5,tmb.data$nstratum))

parameters.L <- list(
                   beta=rep(-Inf,5),
                   log_k=-1,
                   log_kc=-1,
                   log_eta=rep(-10,tmb.data$nstratum))
```

```
[23]  lower = unlist(parameters.L)

[24]  upper = unlist(parameters.U)


[26]  obj <-MakeADFun(tmb.data,parameters, random="log_site",DLL="ML")


[28]  system.time(opt<-nlminb(obj$par,obj$fn,obj$gr,lower=lower,upper=upper))

[29]  rep<-sdreport(obj)

[30]  summary(rep,"fixed")

[31]  summary(rep,"report")
```

The first line loads the TMB package. The second line compiles the C++ template and the third line links to that. The fifth to ninth line includes the initial values for the parameters, both fixed effects and random effects. Notice that the names of parameters should correspond to those in the C++ template. Next we set up the upper and lower bounds for the regression and variance parameter estimates, as well as for the nuisance parameters. Line 26 defines 'obj' containing the data, parameters, also specified the random effects. The last four lines optimize the objective function and generate a standard report.

Chapter 4

Restricted Maximum Likelihood Method

The maximum likelihood (ML) method does not take the degrees of freedom of fixed effects into account when estimating variance parameters. Hence, the estimators of the variance parameters k_s and k_t may be very biased and inefficient, when there are a large number of nuisance parameters. Cadigan and Tobin (2010) [12] demonstrated this for k_t in a fixed-effects model (i.e. no replicates or random effects) for highly-stratified NB data. In this chapter we use restricted maximum likelihood (REML) estimation, also known as residual maximum likelihood estimation, to address this issue.

The REML method was first proposed by Anderson and Bancroft (1952) [1] for balanced data and was extended by Patterson and Thompson (1971) [30] to the estimation of variance components in normal linear mixed models (see Section 4.1). The basic idea of REML is to maximize the part of the likelihood which is invariant to the fixed effects. The REML method was extended by Schall (1991) [38] and Breslow and Clayton (1993) [9] to generalized linear mixed models (GLMM) when normal random effects were introduced. Breslow and Clayton (1993) [9] use a term $-\frac{1}{2}\log|\boldsymbol{X^t V^{-1} X}|$ in $ql(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta})$ (see Eq.(3.4) in Chapter 3) to make the degrees of freedom adjustment. This REML function is

$$ql_1(\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) \approx -\frac{1}{2}\log|\boldsymbol{V}| - \frac{1}{2}\log|\boldsymbol{X^t V^{-1} X}| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^t \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}), \quad (4.1)$$

Lee and Nelder (2001) [24] proposed a new REML method for generalized linear mixed models (GLMM) with non-normal random effects based on the double extended quasi-likelihood (DEQL).

4.1 REML for normal linear mixed models

In this section we describe the REML method for linear mixed effects model with normal random effects. We do this to illustrate the technique and motivate the approach for complex survey count data. For normal linear mixed models, REML estimators are based on choosing a linear transformation of the response variable so that the distribution of the transformed response only involves the variance parameters. REML is based on residuals calculated after fitting the fixed effects (e.g. Searle *et al.*, 2009 [39]).

We assume a normal linear mixed effects model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \sum_{i=1}^{r} \boldsymbol{Z_i u_i} + \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I_n}),$$

where $\boldsymbol{y}$ is a $n \times 1$ vector of sample responses; $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects; $\boldsymbol{X}$ is a $n \times p$ covariance matrix; $\boldsymbol{u_i}$ is a $q_i \times 1$ vector of random effects, with $\boldsymbol{u_i} \sim N(\boldsymbol{0}, \sigma_i^2 \boldsymbol{I_{q_i}})$, $Cov(\boldsymbol{u_i}, \boldsymbol{u_j}) = \boldsymbol{0}$; $\boldsymbol{Z_i}$ is a $n \times q_i$ matrix.

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{V}),$$

where $\boldsymbol{V} = \sum_{i=1}^{r} \sigma_i^2 \boldsymbol{Z_i Z_i'} + \sigma_e^2 \boldsymbol{I_n}$.

The REML function is derived from a linear transformation of $\boldsymbol{y}$, denoted as $\boldsymbol{k'y}$, so that $\boldsymbol{k'y}$ contains no fixed effects; that is, for any $\boldsymbol{\beta}$

$$\boldsymbol{k'X\beta} = \boldsymbol{0} \Rightarrow \boldsymbol{k'X} = \boldsymbol{0}.$$

The form of $\boldsymbol{k}$ must be $\boldsymbol{k'} = \boldsymbol{c'}(\boldsymbol{I} - \boldsymbol{XX^-})$ or $\boldsymbol{k'} = \boldsymbol{c'}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^- \boldsymbol{X'})$ for any $\boldsymbol{c'}$ (See Searle *et al.*, 1998 [39] for the derivation), where $\boldsymbol{X^-}$ denotes the generalized inverse of $\boldsymbol{X}$.

$$\boldsymbol{k'y} \sim N(\boldsymbol{0}, \boldsymbol{k'Vk}).$$

The REML equation is

$$L(\boldsymbol{V}|\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{V}|^{1/2}} \exp\{-\frac{1}{2}(\boldsymbol{k'y})'\boldsymbol{V}^{-1}(\boldsymbol{k'y})\}. \tag{4.2}$$

The REML estimator is derived by maximizing $L(\boldsymbol{V}|\boldsymbol{y})$ in Eq.(4.2). We use an example to show that the REML method can correct the bias in ML estimator of variance parameter.

Example: Recall from chapter 1 that the ML estimator of σ^2 of the linear regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}), \tag{4.3}$$

is $\hat{\sigma}^2_{\text{ML}} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{ML}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{ML}})/n$, where $\hat{\boldsymbol{\beta}}_{\text{ML}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$. $\hat{\sigma}^2_{\text{ML}}$ is biased because

$$E(\hat{\sigma}^2_{\text{ML}}) = \frac{n-p}{n}\sigma^2.$$

The REML function is

$$f(\sigma^2|\boldsymbol{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}|\boldsymbol{k'k}|^{1/2}} \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{k'y})'(\boldsymbol{k'k})^{-1}(\boldsymbol{k'y})\},$$

and the REML estimator of σ^2 is

$$\hat{\sigma}^2_{\text{REML}} = \frac{(\boldsymbol{k'y})'(\boldsymbol{k'k})^{-1}(\boldsymbol{k'y})}{n}.$$

$\hat{\sigma}^2_{\text{REML}}$ is an unbiased estimator of σ^2 since

$$E(\hat{\sigma}^2_{\text{REML}}) = \sigma^2, \tag{4.4}$$

(See Appendix D for the derivation).

4.2 Integrated REML

From a Bayesian perspective, REML can be viewed as maximizing a marginal likelihood for a hierarchical model (e.g. Searle *et al.*, 1998 [39]). Let $\boldsymbol{\theta}_r = (\boldsymbol{\beta}, \boldsymbol{\mu})$ denote the regression model fixed effects and let $\boldsymbol{\theta}_v = (k_s, k_c)$ denote the variance parameters. REML estimates of $\boldsymbol{\theta}_v$ can be derived by integrating Eq. (3.5) over $\boldsymbol{\theta}_r$ using a non-informative prior. This means that the "density" $f(\boldsymbol{\theta_r}) = 1$. The REML likelihood function for $\boldsymbol{\theta}_v$ is

$$L(\boldsymbol{\theta}_v|\boldsymbol{y}) = \int\int f(\boldsymbol{y}, \boldsymbol{\theta}_v|\boldsymbol{\theta}_r, \boldsymbol{\gamma}) f(\boldsymbol{\gamma}) \partial\boldsymbol{\theta}_r \partial\boldsymbol{\gamma}.$$

We will use an example to show the efficacy of this integrated REML.

4.2.1 Integrated REML to estimate σ^2 of a linear regression model

In this example we use the integrated REML to estimate σ^2 of Eq.(4.3). Recall that the ML estimators of $\boldsymbol{\beta}$ and σ^2 are

$$\hat{\boldsymbol{\beta}}_{\text{ML}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

$$\hat{\sigma}^2_{\text{ML}} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{ML}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{ML}})/n.$$

The likelihood function of $\boldsymbol{\beta}$ and σ^2 is

$$L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\}.$$

We integrate the likelihood function over $\boldsymbol{\beta}$ and obtain the REML function

$$\begin{aligned} L(\sigma^2|\boldsymbol{y}) &= \int L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) d\boldsymbol{\beta} \\ &= \int \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\} d\boldsymbol{\beta} \\ &= \frac{|\boldsymbol{X}'\boldsymbol{X}|^{-1/2}}{(\sqrt{2\pi\sigma^2})^{n-p}} \exp\left\{ -\frac{1}{2\sigma^2}\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}\right\}, \end{aligned}$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. The REML estimator of σ^2 is

$$\hat{\sigma}^2_{\text{REML}} = \frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}}{n-p},$$

which is unbiased since

$$E(\hat{\sigma}^2_{\text{REML}}) = \sigma^2.$$

(See Appendix D for the derivation).

4.2.2 Integrated REML in TMB

The integrated REML is often considered to be impractical for a mixture model with non-Gaussian random effects due to the intractable integration over the fixed effects. However, the Laplace approximation of the integral over the fixed effects can be easily implemented in TMB since we only need to specify both $\boldsymbol{\theta}_r$ and $\boldsymbol{\gamma}$ as "random effects" in the R session, whereas other operations are the same as the ML method. Hence, the C++ template function of REML is the same as that of the ML method and the only difference is we treat both $\boldsymbol{\beta}$, $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ as random effects in R session as follows:

```
obj <- MakeADFun(tmb.data,parameters,random=c("log_eta","beta","log_site"),
DLL="ML")
```

The estimation procedure of REML method is to:

1. Estimate the variance parameters k_s and k_c by using the REML method;

2. Using these variance parameter estimates $\hat{k}_s$ and $\hat{k}_c$, use the marginal ML method to estimate the fixed effects $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$.

Chapter 5

Simulation Study on PdG Mixture Model

5.1 Introduction

5.1.1 Simulation factors

In this chapter we present a simulation study to compare the three methods we have studied: GEE, ML using TMB, and REML. We investigate the effects that different study design factors (i.e. sample size, number of strata, etc.) may have on the reliability of estimates and confidence intervals. Our preliminary investigation showed that the important factors are: the total number of strata H; the number of sites per stratum n_h, $h = 1, 2, \cdots H$; the nuisance parameters μ_h and the covariance parameters k_s and k_c.

Therefore, we choose these simulation factors:

1. three "levels" of nuisance parameters, $\mu_h = 1, 5, 10$, each the same for all H strata;
2. low and median between-site over-dispersion, $k_s = 1$ and 3, and within-site over-dispersion $k_c = 5 \cdot k_s$;
3. small and large number of strata, $H = 25, 100$;
4. number of sites per stratum, $n_h = 5, 15, 30$ when $H = 25$, and $n_h = 5, 15$ when $H = 100$;
5. $p = 5$ regression parameters, $\boldsymbol{\beta} = (-1, -0.25, 0, 0.25, 1)$, whereby the values for $\boldsymbol{\beta}$ cover a range of effects, from small to large.

In fishery surveys, most of the sites only contain one observation, while some may have replicates. Hence, we generate observations consistent with the fishery surveys by generating only one observation at most of the sites and replicates at a smaller number of sites. For example, when $n_h = 5$, three of the sites had one observation and the other two sites had 2 or 3 observations. The replicates we generate in each site for different levels of n_h are listed in Table 5.1. Note that we did not use $n_h = 30$ for $H = 100$ since it is not a realistic scenario and the GEE estimator was slow for this case. The three estimation methods can also be treated as three different factors, therefore, we have a total of 90 factors $(3\mu_h \times 2k \times (2n_h + 3n_h) \times 3method = 90)$.

Table 5.1: Number of replicate sites in the simulation study for each level of n_h.

	Replicates per site				
n_h	1	2	3	5	8
5	3	1	1	-	-
15	11	2	1	1	-
30	20	6	2	1	1

5.1.2 Simulation Setup

We used R software to generate the random numbers. Recall from Chapter 1 that site effects $\gamma_{hi} \sim \text{Gamma}(k_s, \frac{1}{k_s})$, replicates effects $\gamma_{hij} \sim \text{Gamma}(k_c, \frac{1}{k_c})$, and response variables $Y_{hij}|\gamma_{hi}\gamma_{hij} \sim \text{Poisson}(\mu_h\gamma_{hi}\gamma_{hij}\eta_{hij})$, where $\eta_{hij} = \exp(\sum_{k=1}^{p} \beta_k x_{hijk})$.

The steps we use to generate random simulated data are

1. generate normal random numbers $x_{hijk} \sim N(0, 1.5^2)$, and compute $\eta_{hij} = \exp(\sum_{k=1}^{5} \beta_k x_{hijk})$;

2. generate gamma random numbers $\gamma_{hi} \sim \text{Gamma}(k_s, \frac{1}{k_s})$ and $\gamma_{hij} \sim \text{Gamma}(k_c, \frac{1}{k_c})$;

3. generate response $y_{hij} \sim \text{Poisson}(\lambda_{hij})$, where $\lambda_{hij} = \mu_h \gamma_{hi} \gamma_{hij} \eta_{hij}$.

In R we can't generate a PdG random number directly.

In the model estimation, we set upper and lower estimation bounds for k_s, k_c, and μ_h:

- the upper bound for $\log(\mu_h)$ is 5 and lower bound is -10;
- the upper bound for $\log(k_s)$ is 2.3 and lower bound is -2.3;
- the upper bound for $\log(k_c)$ is 2.3 when the true value of $k_s = 1$, and is 3.4 when $k_s = 3$, the lower bound is always -2.3.

Initial values are an important issue since all optimization methods we use are derivative based. The R procedure 'glm' was first used to obtain initial values of $\boldsymbol{\beta}$ (see Appendix B.4) for the GEE method, and we used the upper bound of k_s and the lower bound of k_t ($k_t = k_s * k_c/(1 + k_s + k_c)$) to be the corresponding initial values. The initial values of $\boldsymbol{\beta}$, $\log(\mu)$, $\log(k_s)$ and $\log(k_c)$ were all set at zero for the ML and REML methods.

5.1.3 Analysis Methods

We conducted $N = 2000$ simulations for each of the 90 factor combinations. We choose 2000 simulations because we will examine the accuracy of confidence intervals (see below). Hence, there are in total 180,000 simulations. We focused on the bias, standard bias (sbias) and root mean square error (RMSE) for $\boldsymbol{\beta}$, the bias for k_s and k_c, and the coverage of standard linear-approximation confidence intervals for $\boldsymbol{\beta}$. The bias of $\hat{\boldsymbol{\beta}}$, $\hat{k}_t$ and $\hat{k}_s$ is defined as the simulation average minus true value. The standardized bias (sbias) of $\hat{\boldsymbol{\beta}}$ is

$$\text{sbias}(\hat{\boldsymbol{\beta}}) = \frac{1}{N}\sum_{j=1}^{N}\frac{(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta})}{sd(\hat{\boldsymbol{\beta}}_j)},$$

where $sd(\hat{\boldsymbol{\beta}}_j)$ is the standard deviation of $\hat{\boldsymbol{\beta}}_j$, which is a vector and depend on the method being used to estimate. The root mean square error (RMSE) of $\hat{\boldsymbol{\beta}}$ is defined as

$$\text{RMSE}(\hat{\boldsymbol{\beta}}) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta})^2}.$$

We focused on three aspects of coverage. The 95% confidence intervals of $\boldsymbol{\beta}$ were defined as

$$[\hat{\boldsymbol{\beta}} - 1.96 \cdot sd(\hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}} + 1.96 \cdot sd(\hat{\boldsymbol{\beta}})],$$

where $sd(\hat{\boldsymbol{\beta}})$ is a vector and depend on the method being used to estimated, "sandwich estimator" is used for GEE method and for ML and REML, $sd(\hat{\boldsymbol{\beta}})$ are provided by a standard error report.

We computed the percent of simulations where:

CI. $\boldsymbol{\beta}$ were outside their 95% confidence intervals, LC. $\boldsymbol{\beta}$ were less than their lower intervals,

UC. $\boldsymbol{\beta}$ were greater than their upper intervals.

If α is the nominal probability of the upper confidence limit then the standard deviation of the simulation estimate of this probability is $\sqrt{\alpha(1-\alpha)/N}$ where N is the number of simulations. If $\alpha = 0.025$ and $N = 2000$ then the simulation standard deviation is 0.0035 and the width of the simulation confidence interval for α is $2 \times 2 \times 0.0035 = 1.4\%$ which is adequate for our purposes.

Convergence of the parameter estimates for REML and ML was very good; usually 100% and never less than 99%. For the GEE method convergence was usually greater than 98%, sometimes between 95% and 98%, and worse in three cases: 1) $k = 1$, $H = 25$, $n_h = 5$, 2) $\mu_h = 5, 10$ and $k = 3$, $H = 25$, and 3) $n_h = 5$ and $\mu_h = 10$. In the worst case convergence was 89%. Note that our analyses are based on the converged simulations.

5.2 Simulation Analysis

Simulation results of bias, standard bias (sbias) and root mean square error (RMSE) for $\boldsymbol{\beta}$, the bias for k_s and k_c, 95% confidence interval (C.I.) coverage for $\boldsymbol{\beta}$ and coverage for the lower (C.L.) and upper limits (U.L.) for the three estimation procedures we investigated are presented in Table (E.1)-Table (E.18) in Appendix E. However, results for each performance measure (e.g. bias) are based on three large tables and it is very difficult to summarize the estimation results directly from these tables. Hence, we use ANOVA to help summarize the simulation results. We treat the estimation results as responses (e.g. bias, sbias), the simulation factors (e.g. μ_h, H) as covariates, and use R procedures 'glm' and 'anova' to summarize the results. For example, for standardized bias, the R code we used is

```
>sum.sbias.beta<-glm(sbias~method+beta+mu+k+H+nh, family=gaussian,
data=sim.beta.sbias)
>summary(sum.sbias.beta)
>anova(sum.sbias.beta)
```

We only present results for the factors (and 2nd order interactions) that explained most of the variation in bias, standard bias, etc. The values in Table 5.2 is the percent of total deviance explained by the factors and their interactions; larger values mean the factor or the interaction is more important in determining the simulation results (bias, sbias, RMSE, CI, LC, UC). Hence, we used ANOVA to determine the most important factors impacting the simulation results.

Table 5.2: ANOVA results for bias, standard bias (sbias), root mean square error (RMSE), 95% confidence interval coverage (CI) and the upper (UC) and lower CI coverage (LC) for $\boldsymbol{\beta}$. Values are the percent of total deviance explained by the factors and their interactions.

factor& interaction	df	bias	sbias	RMSE	CI	LC	UC
method	2	0.24	0.37	0.02	56.70	56.25	55.52
$\boldsymbol{\beta}$	4	7.69	9.58	0.01	0.13	0.40	0.62
μ_h	2	1.98	1.34	19.73	0.08	0.08	0.08
k	1	3.49	5.94	7.91	1.19	1.04	1.30
H	1	0.04	0.02	16.56	0.37	0.27	0.48
n_h	2	0.11	0.09	46.36	19.63	19.15	19.52
method$\times\boldsymbol{\beta}$	8	42.01	36.17	0.00	0.16	0.37	0.32
method$\times\mu_h$	4	5.98	7.76	0.05	0.25	0.29	0.42
method$\times k$	2	0.15	0.11	0.01	2.05	1.81	2.24
method$\times H$	2	0.01	0.03	0.00	0.26	0.33	0.19
method$\times n_h$	4	0.46	0.80	0.00	17.64	17.64	17.17
$\boldsymbol{\beta} \times \mu_h$	8	3.31	6.10	0.00	0.07	0.31	0.21
$\boldsymbol{\beta} \times k$	4	2.45	4.76	0.00	0.07	0.13	0.13
$\boldsymbol{\beta} \times H$	4	4.90	7.39	0.00	0.05	0.08	0.17
$\boldsymbol{\beta} \times n_h$	8	19.06	11.22	0.01	0.18	0.29	0.28
$\mu_h \times k$	2	1.79	2.62	0.00	0.11	0.17	0.20
$\mu_h \times H$	2	2.31	2.79	1.14	0.01	0.08	0.11
$\mu_h \times n_h$	4	2.89	2.08	3.53	0.14	0.15	0.18
$k \times H$	1	0.00	0.66	0.47	0.14	0.43	0.01
$k \times n_h$	2	1.13	0.11	1.35	0.55	0.44	0.67
$H \times n_h$	1	0.00	0.05	2.83	0.24	0.29	0.18

5.2.1 Bias of $\boldsymbol{\beta}$

The ANOVA indicates that the most important factors affecting bias were the true values of $\boldsymbol{\beta}$, the estimation method$\times\boldsymbol{\beta}$, and $\boldsymbol{\beta} \times n_h$ (see Table 5.2). Table 5.3 shows the simulated biases after combining over the factors that are insignificant (i.e. H,

μ_h, and k).

Table 5.3: bias∗1000 of $\boldsymbol{\beta}$ for factors method, n_h and $\boldsymbol{\beta}$

Method	n_h	β_1	β_2	β_3	β_4	β_5
GEE	5	2.630	-2.059	0.901	-0.107	-1.359
ML	5	-0.277	-0.582	2.766	-0.875	-1.859
REML	5	-8.916	-2.477	2.706	1.296	6.616
GEE	15	-0.112	0.612	0.333	-0.044	-0.374
ML	15	1.149	-0.679	-0.426	0.532	-1.836
REML	15	-0.848	-1.205	-0.457	1.140	0.307
GEE	30	-0.793	0.222	0.405	-0.173	-0.101
ML	30	0.803	1.262	-0.009	1.148	-2.082
REML	30	0.029	1.060	-0.008	1.324	-1.262

For these factors the worst bias is for the REML-based estimator of β_1 when $n_h = 5$ (Table 5.3). The true value is -1 and the average simulation estimate is -1.009. All three estimators of $\boldsymbol{\beta}$ are basically unbiased. Patterns in average simulation bias are unclear and may simply be caused by simulation error.

5.2.2 Standardized bias of $\boldsymbol{\beta}$

The ANOVA indicates that the most important factors affecting standardized bias are the true values of $\boldsymbol{\beta}$, the estimation method×$\boldsymbol{\beta}$, and $\boldsymbol{\beta} \times n_h$ (see Table 5.2). Table 5.4 shows the simulated standardized biases after combining over the factors that are insignificant (i.e. H, μ_h, and k).

Table 5.4: standardized bias (sbias)*1000 of $\boldsymbol{\beta}$ for factors method, n_h and $\boldsymbol{\beta}$

Method	n_h	β_1	β_2	β_3	β_4	β_5
GEE	5	26.463	-6.679	10.245	-2.503	-17.322
ML	5	6.182	-5.085	12.612	-4.990	-15.432
REML	5	-40.880	-16.147	11.552	7.080	31.991
GEE	15	3.195	2.138	7.206	-0.778	-6.693
ML	15	18.482	-4.602	-4.056	5.494	-21.901
REML	15	-1.638	-9.714	-4.450	11.381	-0.660
GEE	30	-5.898	0.966	3.908	0.260	-2.474
ML	30	6.939	16.382	-1.486	11.565	-23.234
REML	30	-1.502	14.247	-1.435	13.490	-14.378

For these factors the worst standardized bias is for the REML-based estimator of β_1 when $n_h = 5$ (Table 5.4). For this case the standardized bias was -0.041.

5.2.3 Root mean square error of $\boldsymbol{\beta}$

The most important factors that impact RMSE are n_h, μ_h and H (see Table 5.2). We combine and average over the insignificant factors (see Table 5.5). RMSE decreases when μ_h, H and n_h increases. This makes sense for n_h and H, as increasing these factors leads to an increase in the total sample size. It also makes sense that μ_h has a similar effect. If the data were Poisson distributed then a good approach for inferences about $\boldsymbol{\beta}$ would be to condition on the total catch in all strata. This total catch would usually be larger when μ_h is larger and this gives some intuition why this parameter acts like a sample size effect. The three estimators of $\boldsymbol{\beta}$ have similar

RMSE because the method factor explains very little of the variation in RMSE (see Table 5.2).

Table 5.5: Root mean square error of $\boldsymbol{\beta}$ for factors μ_h, n_h and H.

	$H = 25$			$H = 100$	
μ_h	$n_h = 5$	$n_h = 15$	$n_h = 30$	$n_h = 5$	$n_h = 15$
1	0.341	0.187	0.120	0.164	0.092
5	0.217	0.121	0.077	0.105	0.060
10	0.188	0.107	0.069	0.092	0.053

5.2.4 Confidence Interval of $\boldsymbol{\beta}$

The most important factors that impact the accuracy of confidence intervals are the estimation method and n_h, with a large interaction between these two factors (see Table 5.2). Table 5.6 gives the simulation results after combining and averaging over the insignificant factors, which indicates that among the three methods we investigate:

- REML confidence intervals are the most accurate with total and one-sided coverage probabilities close to the 0.05 or 0.025 nominal values;
- GEE intervals are too narrow particularly when $n_h = 5$, since both the total and one-sided coverage probabilities are more different than their nominal values.

The REML confidence intervals are more accurate because the estimates of k_t and k_s are more accurate (see below).

Table 5.6: 95% confidence interval (C.I.) coverage for $\boldsymbol{\beta}$ and coverage for the lower (C.L) and upper limits (U.L).

		n_h		
	method	5	15	30
	GEE	11.25	7.38	6.43
C.I	ML	6.29	5.20	4.92
	REML	5.05	5.00	4.95
	GEE	5.59	3.72	3.21
C.L	ML	3.19	2.63	2.48
	REML	2.52	2.52	2.50
	GEE	5.66	3.66	3.23
C.U	ML	3.10	2.57	2.44
	REML	2.52	2.48	2.45

5.2.5 ANOVA for variance parameters

The ANOVA (Table 5.7) indicates that

- the most important factors affecting the bias of k_s are μ_h, k_s and n_h, with an interaction between method and μ_h;
- the most important factors affecting the bias of k_t are method, μ_h and n_h and k_s.

Table 5.7: ANOVA results for bias in estimates of k_s and k_t.

Factor	df	k_s	k_t	Interaction	df	k_s	k_t
method	2	3.65	11.52	$k_s \times n_h$	2	0.09	5.75
μ_h	2	52.18	17.86	method$\times n_h$	4	2.41	8.28
k_s	1	7.17	14.22	$\mu_h \times n_h$	4	6.78	6.15
H	1	0.62	0.55	method$\times k$	2	0.29	4.04
n_h	2	14.56	21.64	method$\times \mu_h$	4	7.35	2.81

5.2.6 Bias of k_t and k_s

Table 5.8 gives the bias of k_t and k_s after combining and averaging over insignificant factors. The bias of k_s estimates (Table 5.8) indicate that:

- the bias of k_s decreases when μ_h or n_h increases. The bias is large when $\mu_h = 1$;
- REML has the lowest bias for k_s except when $\mu_h = 1$ in which case the GEE bias is lower.

We also calculated the percent of times in simulations that the k_s estimates hit the bounds for the three different methods: 15.8%, 1.07% and 0.28% of the GEE k_s estimates hit the upper bound when $\mu_h = 1, 5$ and 10 respectively; 46.6%, 0.81% and 0.07% of the ML k_s estimates hit the upper bound when $\mu_h = 1, 5$ and 10 respectively; 29.1%, 0.06% and 0% of the REML k_s estimates hit the upper bound; very few k_s estimates hit the lower bound.

The bias of k_t estimates (Table 5.8) indicates that:

- when μ_h or n_h are larger, the bias of k_t tends to be lower;
- the REML estimates of k_t has the lowest bias.

It is well-known that MLE's of the NB dispersion parameter are less precise when μ_h is small and this is what our simulation results also demonstrated.

Table 5.8: Mean bias in estimates of k_s and k_t for factors μ, n_h, k_s and method.

		μ_h=1		μ_h=5		μ_h=10	
n_h	method	$k_s = 1$	$k_s = 3$	$k_s = 1$	$k_s = 3$	$k_s = 1$	$k_s = 3$
		bias of k_s estimates					
5	GEE	2.81	5.65	1.17	2.70	0.96	1.84
	ML	6.38	6.84	0.56	2.22	0.41	1.40
	REML	3.68	5.03	0.10	0.29	0.06	0.11
15	GEE	0.45	1.81	0.24	0.62	0.12	0.36
	ML	1.91	5.96	0.24	0.65	0.17	0.42
	REML	1.59	4.52	0.09	0.17	0.05	0.06
30	GEE	0.20	0.70	-0.00	0.31	-0.18	0.21
	ML	0.85	3.44	0.16	0.36	0.10	0.22
	REML	0.67	2.56	0.08	0.14	0.04	0.05
		bias of k_t estimates					
5	GEE	0.98	3.29	0.60	1.50	0.51	1.02
	ML	0.89	3.64	0.30	1.22	0.23	0.88
	REML	0.24	0.75	0.06	0.13	0.04	0.07
15	GEE	0.25	0.82	0.14	0.40	0.07	0.23
	ML	0.27	1.16	0.09	0.35	0.07	0.25
	REML	0.16	0.42	0.04	0.08	0.03	0.04
30	GEE	0.12	0.34	-0.00	0.21	-0.12	0.14
	ML	0.18	0.69	0.06	0.19	0.04	0.13
	REML	0.15	0.39	0.04	0.07	0.02	0.03

5.2.7 Simulation time

We also investigated the computing time for each simulation. Method, H and n_h are the most important factors affecting the computing time. Table 5.9 indicates that: the computing speed of REML is the fastest, especially when the sample size is large; the GEE method is much slower than ML and REML.

Table 5.9: Time for one simulation (second) for method, H and n_h

	$H = 25$			$H = 100$		
	$n_h = 5$	$n_h = 15$	$n_h = 30$	$n_h = 5$	$n_h = 15$	total
ML	0.78	1.34	2.43	3.59	9.56	17.7
REML	0.93	1.36	2.31	2.95	6.87	14.42
GEE	2.88	5.46	13.96	13.49	30.89	66.68

5.3 Summary

Our simulation results show that:

1. The regression parameter estimates were almost unbiased for all three methods.
2. The REML estimates of variance parameters had the lowest bias and were almost unbiased when the strata sample sizes were large except when the Poisson mean was low (i.e. $\mu_h = 1$).
3. The GEE method tended to underestimate the variance of regression estimates which led to a less accurate confidence interval.
4. The simulation speed for ML and REML is much faster than the GEE method.

In summary, the REML is the most preferable method for this mixture model among the three methods we discussed.

Chapter 6

Applications

6.1 Application 1: Diel effects for three species from a bottom trawl survey of the southern Gulf of St. Lawrence

6.1.1 Background

Fisheries and Oceans Canada has conducted bottom trawl surveys of the southern Gulf of St. Lawrence annually since 1971. The main objective of these surveys is to estimate the abundance of multiple species and how this changes from year to year. However, changes in fish abundance may not be reflected directly by changes in the average catch of fish species in the survey. Changes in vessel and gear may also have an impact on the fish catches. In addition, fish catches in trawl surveys can differ between day and night due to diel behaviour, such as vertical migrations and burrowing in sediments (e.g. Benoît and Swain (2003) [7]). Accounting for these diel differences can improve the precision of abundance indices estimated from the data and can help to eliminate biases if data collected only during the day are combined with those collected under 24 hr sampling. Prior to 1984, surveys in this area occurred only during the day and it is therefore necessary to adjust for diel variations in more recent data sets to allow for large scale temporal comparisons across years using day-only surveys and 24 hour surveys.

Benoît and Swain (2003) [7] estimated the relative catchability during day and night for a large number of marine fish species off the east coast of Canada using count data collected in the bottom-trawl research survey. The data included trawl hauls conducted at day and night at 67 sites (pairs) in 1988, as well as day and night hauls that were not paired at the same sites but occurred in common strata over the course of seven years of surveys (1985-1991). These authors analyzed the paired and unpaired data in separate analyses (see Benoît and Swan, 2003 [7]). We will use some of these data jointly which should provide more reliable statistical inferences (see Table 6.1).

Here we focus on the data for three species, white hake (*Urophycis tenuis*), thorny skate (*Amblyraja radiata*) and yellowtail flounder (*Limanda ferruginea*) from the 67 paired hauls and 19 unpaired hauls that were made in 1988 in 26 strata. We excluded sets or pairs of sets when there was no catch of a species for all sets within a stratum (see Figure 6.1 and Table 6.1).

Table 6.1: Frequency of tows at sites.

Number of tows	1	2
white hake	13	56
yellowtail flounder	6	38
thorny skate	19	64

The fishing vessel was unchanged so there is no vessel effect. However, there may exist a diel effect since the survey was conducted for 24 hours per day (see Table 6.2). We define a night tow as occurring within the interval of [19:00 hrs, 07:00 hrs], and a day tow otherwise. Figure 6.2 and Figure 6.3 indicate diel effects may exist, especially for yellowtail flounder.

Table 6.2: Frequency of day tows and night tows.

	Day tow	Night tow
white hake	67	58
yellowtail flounder	43	39
thorny skate	79	68

We analyze these data using the three methods discussed in this thesis that jointly

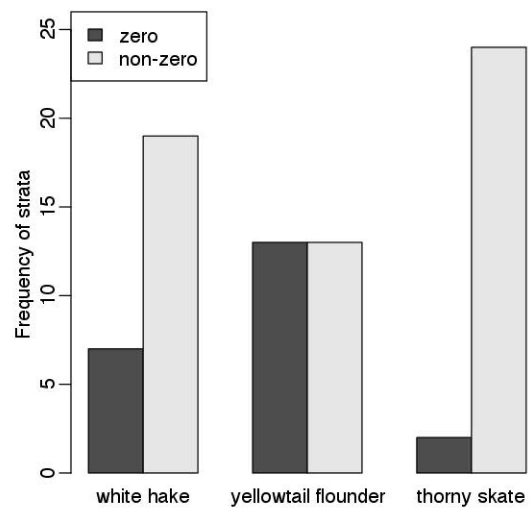

Figure 6.1: Frequency of strata in which the total catch from all sets was zero or non-zero, for the three species.

treat the paired and unpaired data to determine if the catch rates differed between day and night. We define a day/night indicator covariate x that is

$$x = \begin{cases} 1 & \text{if it is a night tow,} \\ 0 & \text{if it is a day tow.} \end{cases}$$

We use the PdG mixture model outlined in the thesis with $p = 1$ and β is the logarithm of the night effect.

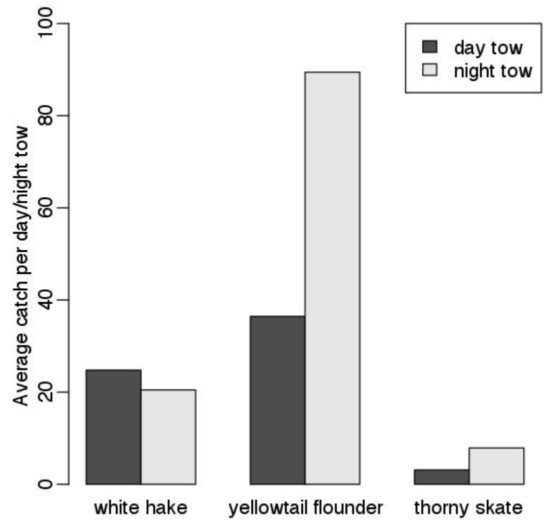

Figure 6.2: Average catch per day/night tow for the three species

6.1.2 Data Analysis

The MLE and REML estimates of β were more similar than the GEE estimates (Table 6.3). The ML and GEE standard errors for $\hat{\beta}$ were smaller than the REML-based ones, and the estimates of k_s and k_t were larger. However, all three methods lead to the same conclusion that there is no diurnal effect for white hake (*Urophycis tenuis*) but that there is for thorny skate (*Amblyraja radiata*) and yellowtail flounder (*Limanda ferruginea*). The parameter estimates and conclusions are comparable to those obtained by Benoit and Swain (2003) [7] using the paired data only (Table 6.3), though our estimate for yellowtail flounder is higher, likely as a result of the added unpaired hauls in our analysis. A notable difference between our data analysis results and those of Benoît and Swan (2003) [7] was that despite including more data, the standard errors on β were larger in our analysis. However, the higher variance estimates may be appropriate since Benoît and Swan (2003) [7], and Casey and Myers (1998) [15] concluded that standard errors for estimates of the β parameter from linear models with extra-Poisson and extra-Binomial variability were too small. This led to a higher frequency across species of nominally statistically significant results compared to what was obtained using randomization tests. This is consistent with Cadigan and Dowden (2010) [11] and Cadigan and Bataineh (2012) [10] who found that simulated confidence intervals about β from an over-dispersed binomial model based on paired-catches with within-site Poisson over-dispersion were much too narrow. Better results were found using GLMMs with random effects that more closely matched those in their simulated populations, and more closely resembled what may occur in real surveys.

Table 6.3: Data analysis for three fish species: white hake, thorny skate, and yellowtail flounder. ([a]B&S denote the parameter estimates obtained by [7])

	white hake				thorny skate			
	REML	ML	GEE	B&S^a	REML	ML	GEE	B&S^a
β	0.136	0.146	0.275	0.135	0.715	0.729	0.765	0.850
$se(\beta)$	0.223	0.219	0.201	0.160	0.160	0.155	0.120	0.142
k_s	0.850	1.350	2.451	–	2.742	5.889	10.000	–
k_t	0.438	0.614	0.991	–	1.119	1.634	1.911	–

	yellowtail flounder			
	REML	ML	GEE	B&S^a
β	1.326	1.291	1.184	0.818
$se(\beta)$	0.274	0.269	0.173	0.148
k_s	0.228	0.368	0.629	–
k_t	0.133	0.212	0.467	–

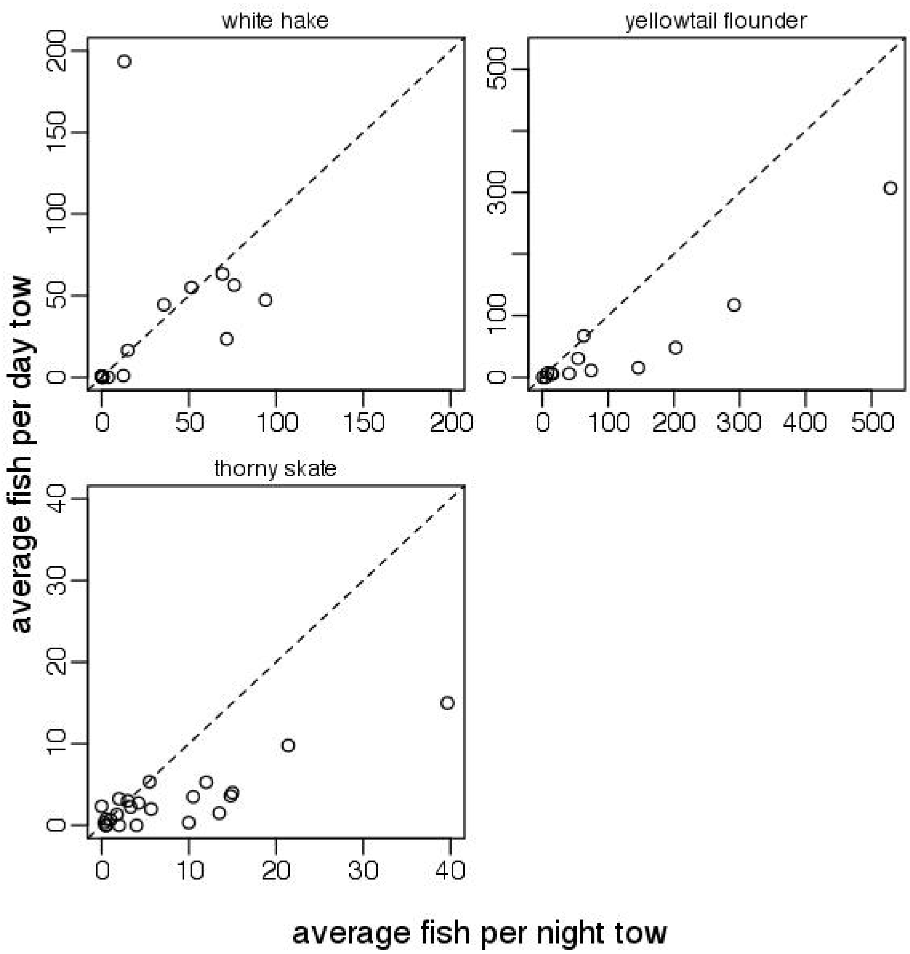

Figure 6.3: Average catch per day vs night tow for each stratum for the three species. 44.4%, 92.3% and 79.2% of the points are below the reference line for white hake, yellowtail flounder and thorny skate respectively.

6.2 Application 2: Diel and vessel effects for snow crab surveys of the southern Gulf of St. Lawrence.

6.2.1 Background

In this application, we investigate the abundance of snow crab in the southern Gulf of St. Lawrence (sGSL; see Figure 1.1). Two different surveys provide insights into snow crab abundance. The multi-species research vessel survey (RVS) has been conducted annually since 1971 and has provided information about snow crab in the catches since 1980. The other source of information is called the crab survey (CS), which has been conducted annually since 1988 and is focused purely on snow crab. Here we focus on the data from 2003-2014 collected during the RVS and CS (see Table 6.4).

Inconsistencies among surveys due to different vessels are present. The vessels used in the RVS were: *CCGS Wilfred Templeman* in 2003; *CCGS Alfred Needler* from 2004-2005; *CCGS Teleost* from 2004-present. Before changing the vessel from *CCGS Alfred Needler* to *CCGS Teleost*, paired tows were used to estimate the relative catchability (see Table 6.4). The vessels used in the CS were: *Marco-Michel* from 2003-2012 (SCS1); *Jean-Mathieu* from 2013-2014 (SCS2). No comparative fishing was conducted for these two vessels.

Table 6.4: Frequency of strata and sites sampled for the RVS and the CS. Numbers of parentheses indicate sites with paired-tows.

	RVS		CS	
	strata	sites	strata	sites
2003	22	78	20	317
2004	24	176 (36)	20	347
2005	24	145 (86)	20	355
2006	24	165	20	354
2007	24	163	20	355
2008	24	177	20	355
2009	24	148	20	355
2010	24	137	20	354
2011	24	126	20	353
2012	24	142	22	321
2013	24	122	21	351
2014	24	156	21	351

There may exist a diel effect in the RVS since the fishing time is conducted for 24 hours per day (see Table 6.5). We define a night tow as occurring within the interval of [19:00 hrs, 07:00 hrs], and a day tow otherwise. Figure 6.4- Figure 6.6 show that the diel assumption for the RVS is reasonable. There is no diel effect in the CS since it is only conducted during the day.

6.2.2 Model Setup

The model contains a number of parameters that account for the catchability of crab. First, we define a parameter that accounts for the diel effect (δ_j) in catchability in

Table 6.5: Frequency of tows for the RVS (day/night).

RVS	day tow	night tow
2003	39	39
2004	108	104
2005	107	124
2006	83	82
2007	86	77
2008	85	92
2009	74	74
2010	73	64
2011	68	58
2012	72	70
2013	65	57
2014	78	78
Total	938	919

RVS

$$\delta_j = \begin{cases} 1 & j \text{ is a day tow,} \\ \delta & j \text{ is a night tow.} \end{cases}$$

Second, we define a set of parameters, the vessel effects (q_v), to account for catchability differences between vessels.

$$q_v = \begin{cases} 1 & v = Teleost\ , \quad \text{2004-2014,} \\ q_{WT \to T}, & v = Wilfred\ Templeman\ , \quad \text{2003,} \\ q_{AN \to T}, & v = Alfred\ \ Needler\ , \quad \text{2004-2005,} \\ q_{SCS1 \to T}, & v\, in\, SCS1\ , \quad \text{2003-2012,} \\ q_{SCS2 \to T}, & v\, in\, SCS2\ , \quad \text{2013-2014.} \end{cases}$$

Table 6.6: Average catch of snow crab (number) per tow the CS.

Year	Average catch
2003	8.46
2004	9.57
2005	6.84
2006	6.50
2007	5.38
2008	4.03
2009	2.54
2010	2.99
2011	5.58
2012	6.52
2013	5.37
2014	5.37

The notation $q_{a \to b}$ indicates the catchability of vessel a relative to vessel b. The catchability of the $CCGS\ Teleost$ is the reference vessel and is fixed at one. Note that by combining information from both surveys in one model it is possible to estimate the catchability of both vessels in the CS and the relative catchability of these vessels is the ratio of their q_v's, whereas it is not possible to estimate this using only data from the CS survey because there was no comparative fishing between these vessels and they were not used in surveys for the same year.

We use the PdG model outlined in this thesis. The model is based on the counts of commercial sized crabs (males $\geq$ 95mm). We define:

- Y_{hyvij} as the catch from the j'th tow in stratum h and year y for survey vessel

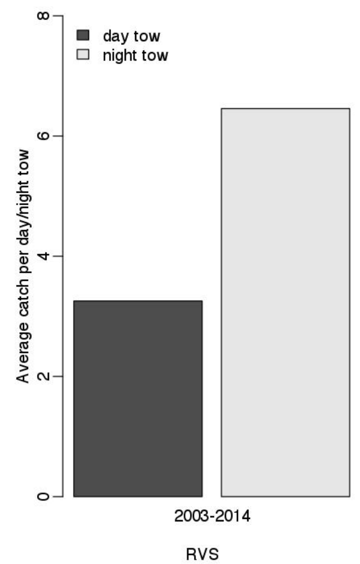

Figure 6.4: Average catch of snow crab (units) per day/night tow for the RVS in 2003-2014

v at site i;

- random site effects, $\gamma_{hyi} \sim \text{Gamma}(k_s, \frac{1}{k_s})$;

- random repeat tow effects at site i, $\gamma_{hyij} \sim \text{Gamma}(k_c, \frac{1}{k_c})$;

- μ_{hy} as the density of crab for stratum h and year y; in the RVS and CS, for any stratum that produced zero catch for all sets, we assigned $\exp(-10)$ to the density μ_{hy} because the density must be greater than zero to calculate the log-likelihood, otherwise the MLE of μ_{hy} in this case is zero which is an infeasible value;

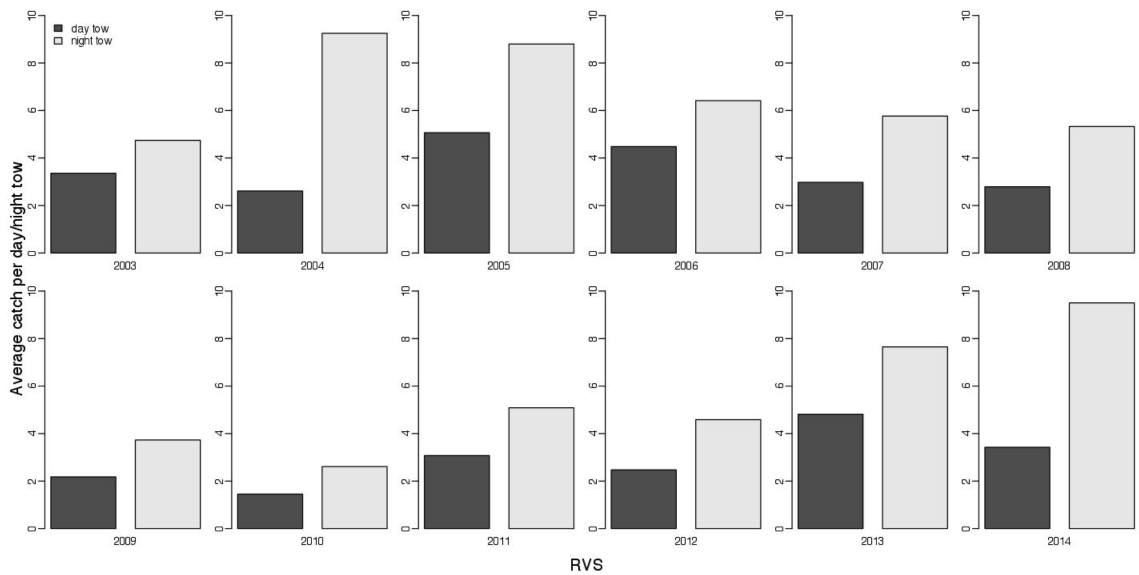

Figure 6.5: Annual average catch of snow crab (units) per day/night tow for the RVS in 2003-2014

- Z_{hyvij} is an offset to make adjustments to account for subsampling of catches (ratio), to standardize for variations in tow distance and standardize μ_{hy} to number per Km^2,

$$Z_{hyvij} = \log(d_{hyvij}/d_0) + \log(\text{ratio}_{hyvij}) + \log(0.0405028),$$

where d_{hyvij} is the actual tow distance for the j'th tow and d_0 is the nominal tow distance, 0.0405028 Km^2 refers to the swept area of a standard $CCGS$ $Teleost$ tow (e.g. Benoît and Cadigan, 2014 [6]);

- the marginal expectation of Y_{hyvij} as μ_{hyvij} and

$$\log(\mu_{hyvij}) = \log(q_v) + \log(\delta_j) + \log(\mu_{hy}) + Z_{hyvij}. \tag{6.1}$$

As stated previously in Chapter 1, if $Y_{hyvij}|\gamma_{hyi}\gamma_{hyij} \sim \text{Poisson}(\mu_{hyvij})$ then the marginal distribution is difficult to specify but the marginal mean and variance are the same as those of NB random variables, $E(Y_{hyvij}) = \mu_{hyvij}$ and $Var(Y_{hyvij}) = \mu_{hyvij} + \mu^2_{hyvij}/k_t$, where $1/k_t = 1/k_s + 1/k_c + 1/k_s k_c$. For survey sites with paired tows, the marginal covariance is $Cov(Y_{hyvi1}, Y_{hyvi2}) = \mu_{hyvi1}\mu_{hyvi2}/k_s$. We use the three methods discussed in this thesis to estimate this model.

In the estimation procedure, the stratum effects were treated as nuisance parameters for the purposes of estimating δ and q_v. However, an important goal of these snow crab surveys is to estimate the stratum size-weighted average density of snow crab (number per Km^2). Let W_h be the size of stratum h, and μ_y be the average density of snow crab in year y:

$$\mu_y = \frac{\sum_{h\in H} W_h \mu_{hy}}{\sum_{h\in H} W_h}.$$

We estimate $\hat{\mu}_y$ via

$$\hat{\mu}_y = \frac{\sum_{h\in H} W_h \hat{\mu}_{hy}}{\sum_{h\in H} W_h}.$$

The standard deviations of $\log(\hat{\mu}_y)$ for the ML and REML methods are automatically provided by the standard error report of TMB. The standard deviations of $\log(\hat{\mu}_y)$

for the GEE method are estimated via the 'sandwich estimator' (e.g. Eq.(2.12)) and δ-method (e.g. Eq.(3.7)). The 95% confidence interval of μ_y is

$$\Big[\exp\{\log(\hat{\mu}_y) - 1.96 \cdot sd[\log(\hat{\mu}_y)]\}, \exp\{\log(\hat{\mu}_y) + 1.96 \cdot sd[\log(\hat{\mu}_y)]\}\Big].$$

6.2.3 Model Approximation

For sites with only one observation, the marginal density function of Y_{hyvi1} is

$$f(Y_{hyvi1} = y) = \frac{k_s^{k_s} k_c^{k_c} \mu_{hyvij}^{y} \Gamma(y + k_c)}{\Gamma(k_s)\Gamma(k_c)\Gamma(y+1)} \int_0^\infty \frac{t^{y+k_s-1} e^{-k_s t}}{(\mu_{hyvi1} t + k_c)^{y+k_c}} dt, \tag{6.2}$$

where μ_{hyvi1} is defined as Eq.(6.1). In Figure (F.1)- Figure (F.4) in Appendix F we compare the probability mass function in Eq.(6.2) with the Negative Binomial mass function

$$f_h(Y_{hyvij} = y) = \frac{\Gamma(y + k_t)}{\Gamma(k_t)\Gamma(y+1)} \left(\frac{\mu_{hyvij}}{\mu_{hyvij} + k_t}\right)^y \left(\frac{k_t}{\mu_{hyvij} + k_t}\right)^{k_t}, \tag{6.3}$$

where $k_t = k_s \cdot k_c/(k_s + k_c + 1)$. Figure (F.5)- Figure (F.8) in Appendix F are the ratio of cumulative mass function for Eq.(6.2) and Eq.(6.3). Figure (F.1)- Figure (F.8) indicate that the NB distribution with a probability mass function Eq.(6.3) is a good approximation to the distribution with probability mass function Eq.(6.2). Considering that most sites just contain one observation (see Table 6.4), we use Eq.(6.3) to approximate the marginal likelihood function of the PdG model with only one observation in a site. Using this approximation reduced the number of random effects to integrate out, from 5903 to 122. This greatly improves the computational speed of our estimators.

6.2.4 Estimation Results

Table (6.7) and Figures (6.7)-Figure (6.9) show the estimates and corresponding 95% confidence intervals of some snow crab model parameters using the GEE, the ML and the REML methods. Results indicate that

- The estimates of δ and q_v for the three methods are similar, but the GEE confidence intervals are more narrow than those of the ML and REML methods. The estimates of k_s and k_c for the ML and REML are more similar than those of the GEE method.
- The three estimates of δ all indicate that for the RVS, adult snow crabs are more catchable at night.
- The three estimates of q_v all indicate that the estimated relative catchability of *Marco-Michel* (SCS1) compared to the *CCGS Teleost* is the largest among the 4 vessels used for sampling since 2003, while the estimated relative catchability of WT is the smallest. The new vessel *Jean-Mathieu* (SCS2), used from 2013,

has a lower relative catchability than *Marco-Michel* (SCS1).

Table 6.7: Estimates (mean, 95% confidence intervals) of some snow crab model parameters for data from 2003-2014

	ML			REML			GEE		
	Estimate	Lower	upper	Estimate	Lower	upper	Estimate	Lower	upper
$q_{AN\to T}$	0.925	0.781	1.094	0.924	0.779	1.096	0.976	0.819	1.163
$q_{WT\to T}$	0.620	0.435	0.884	0.619	0.430	0.890	0.618	0.466	0.820
$q_{SCS1\to T}$	20.308	18.126	22.753	20.283	18.068	22.770	19.814	19.117	20.537
$q_{SCS2\to T}$	14.969	12.257	18.282	14.969	12.185	18.390	14.592	13.265	16.052
δ	1.728	1.515	1.971	1.726	1.509	1.975	1.655	1.500	1.825
k_s	1.166	0.986	1.381	1.098	0.931	1.293	1.250	–	–
k_c	4.797	3.104	7.412	4.716	3.065	7.256	5.300	–	–

Figure (6.10) shows the 4T snow crab abundance estimates (μ_y) and 95% confidence intervals from 2003-2014 using the three methods discussed above. The snow crab abundance decreased during 2003-2009 but has increased since 2010 and was unchanged in 2013-14.

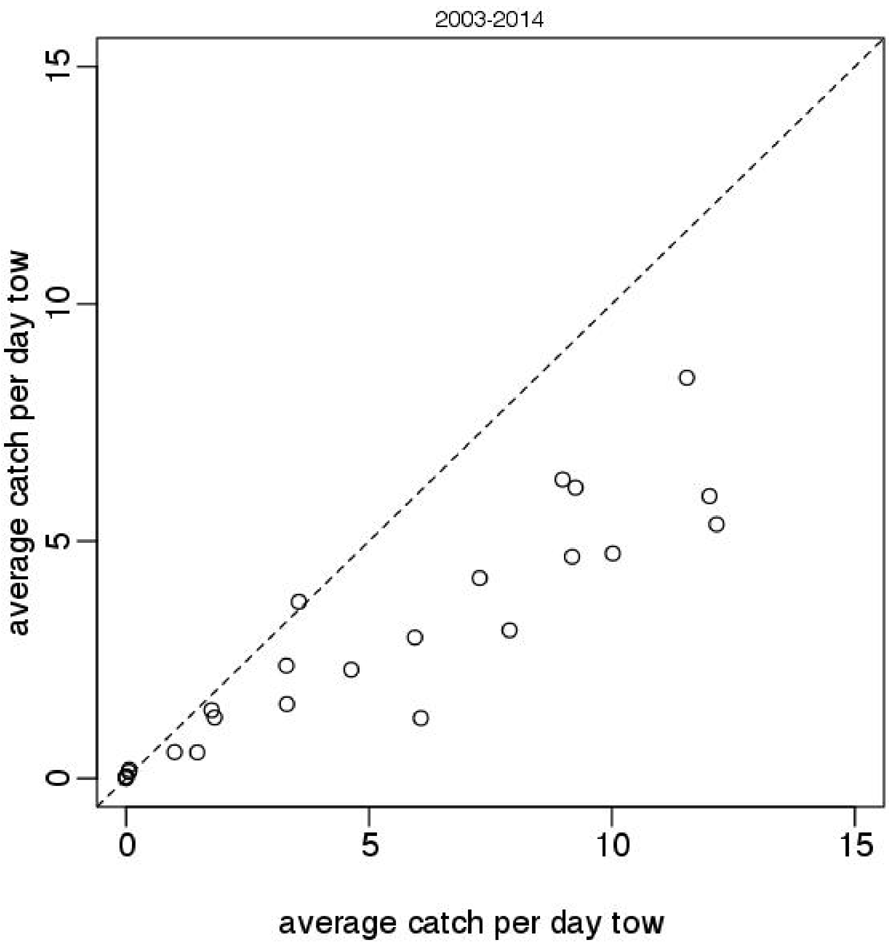

Figure 6.6: Average catch of snow crab per day vs night tow for each stratum for the RVS in 2003-2014, 67% of the points are off the reference line.

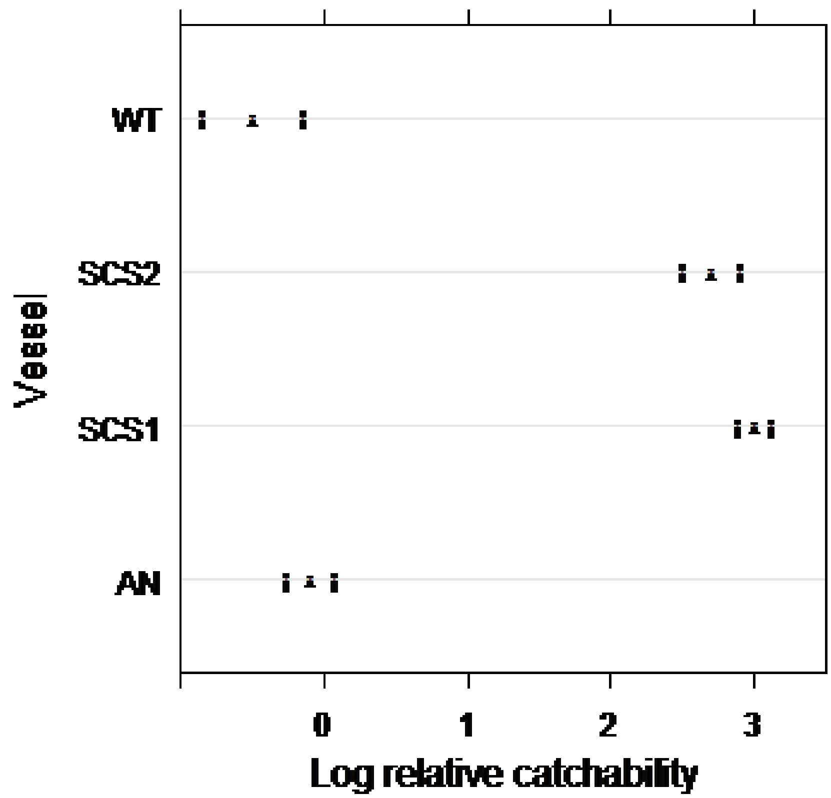

Figure 6.7: ML estimates (middle points) of survey vessel/gear catchabilities, $\log(q_v)$ with 95% confidence intervals. WT is CCGS Wilfred Templeman→CCGS Teleost, AN is CCGS Alfred Needler→CCGS Teleost. The entries SCS are for the catchability of the snow crab survey vessel/gear, relative to the Teleost: *Marco-Michel* (SCS1) for 2003-2012, *Jean-Mathieu* (SCS2) for 2013-2014

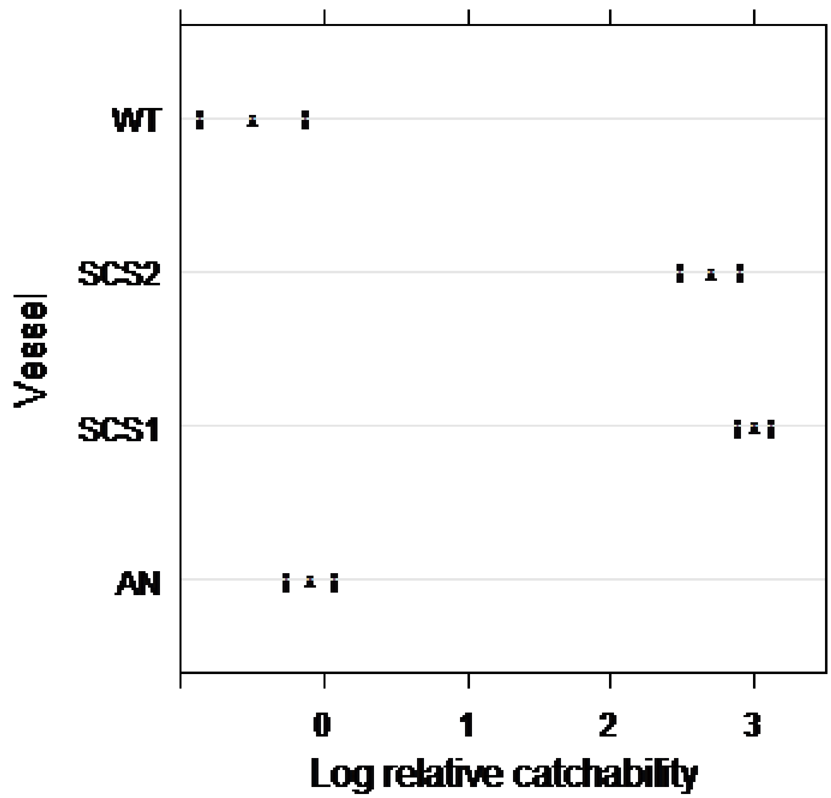

Figure 6.8: REML estimates (middle points) of survey vessel/gear catchabilities, $\log(q_v)$ with 95% confidence intervals. WT is CCGS Wilfred Templeman→CCGS Teleost, AN is CCGS Alfred Needler→CCGS Teleost. The entries SCS are for the catchability of the snow crab survey vessel/gear, relative to the Teleost: *Marco-Michel* (SCS1) for 2003-2012, *Jean-Mathieu* (SCS2) for 2013-2014

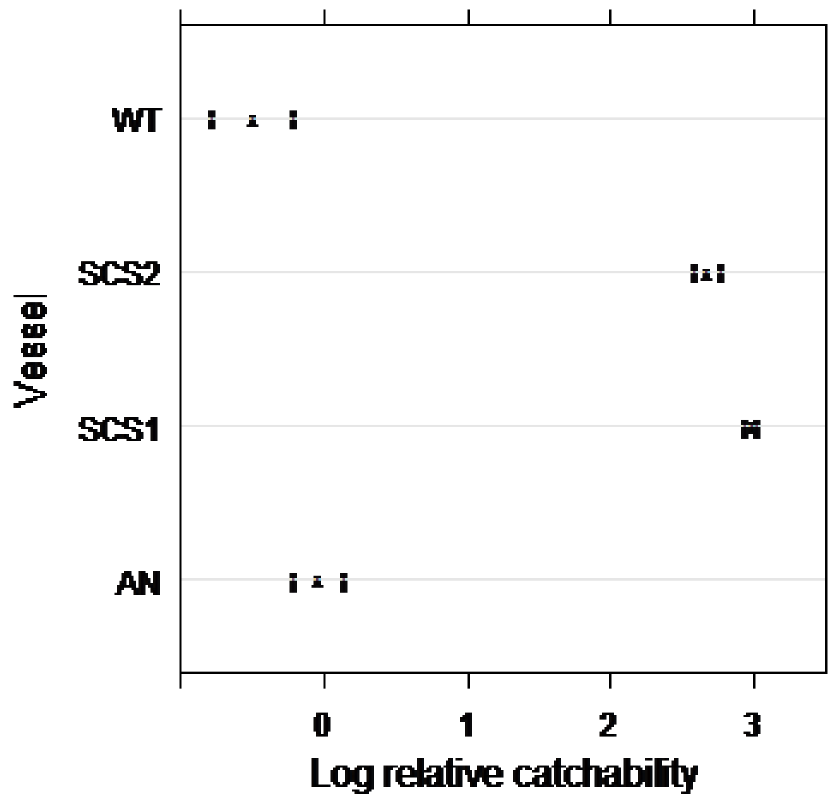

Figure 6.9: GEE estimates (middle points) of survey vessel/gear catchabilities, $\log(q_v)$ with 95% confidence intervals. WT is CCGS Wilfred Templeman→CCGS Teleost, AN is CCGS Alfred Needler→CCGS Teleost. The entries SCS are for the catchability of the snow crab survey vessel/gear, relative to the Teleost: *Marco-Michel* (SCS1) for 2003-2012, *Jean-Mathieu* (SCS2) for 2013-2014

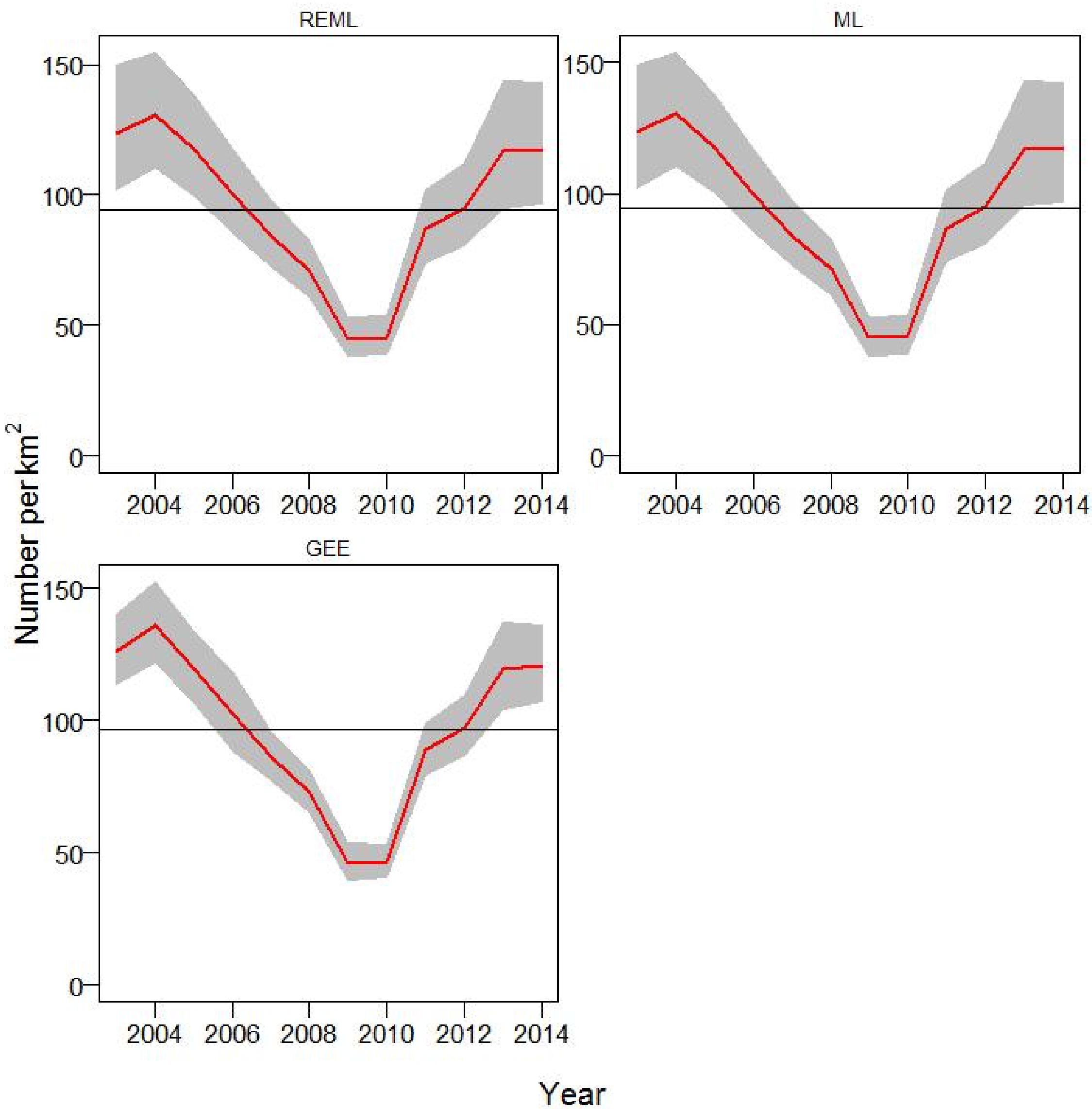

Figure 6.10: Estimates of 4T snow crab abundance from 2003-2014 using three methods. The shaded region indicates 95% confidence intervals. The horizontal line indicates the series average.

Chapter 7

Discussion

In this thesis we developed a profile GEE method to estimate the regression and variance parameters for a Poisson-double-Gamma mixture model where the mixing components are two gamma random variables. The context for this was over-dispersed and correlated count data from highly stratified surveys. The challenges we addressed were 1) computational issues and 2) bias in variance parameter estimates and inaccurate confidence intervals for regression parameters because of a large number of nuisance parameters. We compared the GEE method to MLE and REML methods. In simulations the regression parameter estimates were almost unbiased for all three methods. The REML estimator of variance parameters had the lowest bias and was almost unbiased when the strata sample sizes were large except when the Poisson mean was low (i.e. $\mu = 1$). The GEE method, as expected, tended to over-estimate the NB variance parameters which led to an underestimation of the variance of regression estimates and less accurate confidence intervals. In summary, among the three methods we investigated, REML is the most preferred for statistical inference with this mixture model.

Parameter estimation for our over-dispersed and correlated count data model is fairly simple using TMB. The user just needs to specify the joint log-likelihood function $h(\boldsymbol{\lambda}, \boldsymbol{\theta})$; TMB then provides the marginal likelihood and its gradient function automatically. The analytical gradient greatly improves the speed and accuracy of marginal MLE's using a gradient-based optimization method. Our investigation also shows that TMB would be very appropriate for a model with non-linear random effects.

We treated the stratum effects as nuisance parameters for the purposes of estimating $\boldsymbol{\beta}$. However, an important goal of most fisheries surveys is to estimate the stratum size-weighted average effect (e.g. snow crab survey). If N_h is the total possible number of sampling sites in stratum h then the size-weighted average is $\bar{\mu} = \sum_{h=1}^{H} N_h \mu_h / \sum_{h=1}^{H} N_h$. Usually a time-series of surveys is available and estimates of annual trends in $\bar{\mu}$ are used to indicate trends in fish stock size. It is important to adjust for confounding variables such as changes in vessels or other sampling protocols. In future research we will extend results to this objective and establish some theoretical properties of $\bar{\mu}$ estimators. Particularly for the longitudinal survey case it may be desirable to also treat the stratum effects as random with a spatial distribution that evolves smoothly from year to year.

Appendices

Appendix A

PdG Mixture Model

In Chapter 1, we introduced the a Poisson-Gamma mixture model, where the mixing components are two different gamma random variables to account for different sources of correlation and overdispersion. We use the same notation as Chapter 1. Let Y_{hij} be a random variable (RV) for the j'th observation in stratum h $(h = 1, \cdots, H)$ and site i. Conditional on the stratum effect (μ_h), a random site effect (γ_{hi}) and a replicate effect γ_{hij} at site i, Y_{hij} is assumed to be Poisson distributed with mean $E(Y_{hij}|\gamma_{hi}\gamma_{hij}) = \mu_h\gamma_{hi}\gamma_{hij}\eta_{hij}$, and variance $Var(Y_{hij}|\gamma_{hi}\gamma_{hij}) = \mu_h\gamma_{hi}\gamma_{hij}\eta_{hij}$. We define $\mu_{hij} = \mu_h\eta_{hij}$, the Probability mass function of Y_{hij} conditional on the random effects γ_{hi} and γ_{hij} is

$$f(Y_{hij}|\gamma_{hi}\gamma_{hij}) = \frac{(\mu_{hij}\gamma_{hi}\gamma_{hij})^{y_{hij}}}{y_{hij}!} \exp\{-\mu_{hij}\gamma_{hi}\gamma_{hij}\}. \tag{A.1}$$

The random effects γ_{hi} is assume to be gamma distributed with mean 1 and variance $1/k_s$, with the probability density function

$$f(\gamma_{hi}) = \frac{k_s^{k_s}}{\Gamma(k_s)} \gamma_{hi}^{k_s-1} e^{-k_s\gamma_{hi}}. \tag{A.2}$$

The random effects γ_{hij} is assume to be gamma distributed with mean 1 and variance $1/k_c$, the probability density function is

$$f(\gamma_{hij}) = \frac{k_c^{k_c}}{\Gamma(k_c)} \gamma_{hij}^{k_c-1} e^{-k_c\gamma_{hij}}. \tag{A.3}$$

The marginal (with respect to γ_{hij}) probability density function of Y_{hij} conditional on γ_{hi} is

$$\begin{aligned} f(Y_{hij}|\gamma_{hi}) &= \int_0^\infty f(Y_{hij}|\gamma_{hi}\gamma_{hij}) \cdot f(\gamma_{hij}) d\gamma_{hij} \\ &= \int_0^\infty \frac{(\mu_{hij}\gamma_{hi}\gamma_{hij})^{y_{hij}}}{y_{hij}!} \exp\{-\mu_{hij}\gamma_{hi}\gamma_{hij}\} \frac{k_c^{k_c}}{\Gamma(k_c)} \gamma_{hij}^{k_c-1} e^{-k_c\gamma_{hij}} d\gamma_{hij} \\ &= \frac{\Gamma(y_{hij}+k_c)}{\Gamma(k_c)\Gamma(y_{hij}+1)} \frac{k_c^{k_c}(\mu_{hij}\gamma_{hi})^{y_{hij}}}{(\mu_{hij}\gamma_{hi}+k_c)^{y_{hij}+k_c}} \cdot \\ & \quad \int_0^\infty \frac{(\mu_{hij}\gamma_{hi}+k_c)^{y_{hij}+k_c}}{\Gamma((y_{hij}+k_c)} \cdot \gamma_{hij}^{(y_{hij}+k_c-1)} e^{-(\mu_{hij}\gamma_{hi}+k_c)\gamma_{hij}} d\gamma_{hij} \\ &= \frac{\Gamma(y_{hij}+k_c)}{\Gamma(k_c)\Gamma(y_{hij}+1)} (\frac{k_c}{\mu_{hij}\gamma_{hi}+k_c})^{k_c} (\frac{\mu_{hij}\gamma_{hi}}{\mu_{hij}\gamma_{hi}+k_c})^{y_{hij}}. \end{aligned} \tag{A.4}$$

However, the marginal distribution of Y_{hij} with respect to the random sites effects (γ_{hi}) is not NB. For sites with no replicates (i.e. $n_{hi} = 1$), the marginal distribution of Y_{hi1} is

$$\begin{aligned} f(Y_{hi1}=y) &= \int_0^\infty f(Y_{hi1}|\gamma_{hi})\cdot f(\gamma_{hi})d\gamma_{hi} \\ &= \int_0^\infty \frac{\Gamma(y+k_c)}{\Gamma(k_c)\Gamma(y+1)}(\frac{k_c}{\mu_{hi1}\gamma_{hi}+k_c})^{k_c}(\frac{\mu_{hi1}\gamma_{hi}}{\mu_{hi1}\gamma_{hi}+k_c})^y\frac{k_s^{k_s}}{\Gamma(k)}\gamma_{hi}^{k_s-1}e^{-k_s\gamma_{hi}}\mathrm{d}\gamma_{hi} \\ &= \frac{k_s^{k_s}k_c^{k_c}\mu_{hi1}^y\Gamma(y+k_c)}{\Gamma(k_s)\Gamma(k_c)\Gamma(y+1)}\int_0^\infty \frac{t^{y+k_s-1}e^{-k_st}}{(\mu_{hi1}t+k_c)^{y+k_c}}\mathrm{d}t. \end{aligned} \tag{A.5}$$

The marginal mean and variance of Y_{hi1} are

$$E(Y_{hij}) = E(E(Y_{hij}|\gamma_{hi})) = E(\mu_{hi1}\gamma_{hi}) = \mu_{hi1}, \tag{A.6}$$

$$\begin{aligned} Var(Y_{hi1}) &= Var(E(Y_{hi1}|\gamma_{hi})) + E(Var(Y_{hi1}|\gamma_{hi})) \\ &= Var(\mu_{hi1}\gamma_{hi}) + E(\mu_{hi1}\gamma_{hi} + \frac{\mu_{hi1}^2\gamma_{hi}^2}{k_c}) \\ &= \frac{\mu_{hi1}^2}{k_s} + \mu_{hi1} + \mu_{hi1}^2(1+\frac{1}{k_s})\frac{1}{k_c} \\ &= \mu_{hi1} + \frac{\mu_{hi1}^2}{k_t}, \end{aligned} \tag{A.7}$$

where $k_t = \frac{k_s \cdot k_c}{1+k_s+k_c}$.

If there are two replicates in one site, the marginal likelihood function is

$$
\begin{aligned}
& f(Y_{hi1} = y_1, Y_{hi2} = y_2) \\
= & \int_0^\infty f(Y_{hi1}|\gamma_{hi}) \cdot f(Y_{hi2}|\gamma_{hi}) f(\gamma_{hi}) d\gamma_{hi} \\
= & \int_0^\infty \frac{\Gamma(y_1 + k_c)}{\Gamma(k_c)\Gamma(y_1 + 1)} (\frac{k_c}{\mu_{hi1}\gamma_{hi} + k_c})^{k_c} (\frac{\mu_{hi1}\gamma_{hi}}{\mu_{hi1}\gamma_{hi} + k_c})^{y_1} \frac{k_s^{k_s}}{\Gamma(k_s)} \gamma_{hi}^{k_s - 1} e^{-k_s \gamma_{hi}} \cdot \\
& \frac{\Gamma(y_2 + k_c)}{\Gamma(k_c)\Gamma(y_2 + 1)} (\frac{k_c}{\mu_{hi2}\gamma_{hi} + k_c})^{k_c} (\frac{\mu_{hi2}\gamma_{hi}}{\mu_{hi2}\gamma_{hi} + k_c})^{y_2} \mathrm{d}\gamma_{hi} \\
= & \frac{k_s^{k_s} k_c^{2 \cdot k_c} \mu_{hi1}^{y_1} \mu_{hi2}^{y_2} \Gamma(y_1 + k_c)\Gamma(y_2 + k_c)}{\Gamma(k_s)\Gamma^2(k_c)\Gamma(y_1 + 1)\Gamma(y_2 + 1)} \cdot \int_0^\infty \frac{t^{y_1 + y_2 + k_s - 1} e^{-k_s t}}{(\mu_{hi1} t + k_c)^{y_1 + k_c} (\mu_{hi2} t + k_c)^{y_2 + k_c}} \mathrm{d}t.
\end{aligned} \tag{A.8}
$$

The marginal mean and variance can be derived similiarly to single observation case. There will be marginal correlation in the two observation at a site because there is a common random site effect. The covariance of Y_{hi1} and Y_{hi2} is

$$
\begin{aligned}
Cov(Y_{hi1}, Y_{hi2}) &= E(Y_{hi1}Y_{hi2}) - E(Y_{hi1})E(Y_{hi2}) \\
&= E(E(Y_{hi1}Y_{hi2}|\gamma_{hi})) - E(E(Y_{hi1}|\gamma_{hi}))E(E(Y_{hi2}|\gamma_{hi})) \\
&= \mu_{hi1}\mu_{hi2}E(\gamma_{hi}^2) - \mu_{hi1}\mu_{hi2} \\
&= \frac{\mu_{hi1}\mu_{hi2}}{k_s}.
\end{aligned}
$$

The correlation of Y_{hi1} and Y_{hi2} is

$$\begin{aligned} Corr(Y_{hi1}, Y_{hi2}) &= \frac{Cov(Y_{hi1}, Y_{hi2})}{\sqrt{Var(Y_{hi1})} \cdot \sqrt{Var(Y_{hi1})}} \\ &= \frac{\frac{\mu_{hi1}\mu_{hi2}}{k_s}}{\sqrt{\mu_{hi1} + \frac{\mu_{hi1}^2}{k_t}} \cdot \sqrt{\mu_{hi2} + \frac{\mu_{hi2}^2}{k_t}}}. \end{aligned}$$

Appendix B

Some details for the Generalized estimating equation approach

B.1 GEE with nuisance parameter

B.1.1 Derivation of $\partial\bar{\alpha}_h(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ and $\partial^2\bar{\alpha}_h(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$

The GEE function for the nuisance parameter α_h is

$$U(\alpha_h) = \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\alpha_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\alpha_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\alpha_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix},$$

where $Z_{hik}(\boldsymbol{\beta}) = \boldsymbol{x}'_{hik}\boldsymbol{\beta}$, the q'th diagonal element of V_{hi} is

$$V_{hi}(q,q) = e^{Z_{hiq}(\boldsymbol{\beta})} + \frac{e^{2Z_{hiq}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t},$$

the q,p'th element is

$$V_{hi}(q,p) = \frac{e^{Z_{hip}(\boldsymbol{\beta})} \cdot e^{Z_{hiq}(\boldsymbol{\beta})} \cdot e^{\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s},$$

where $\hat{k}_s$ and $\hat{k}_t$ are the estimates of covariance parameters. The GEE estimator $\hat{\alpha}_h$ follows

$$U(\bar{\alpha}_h) = \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix} = 0. \quad \text{(B.1)}$$

Take derivatives to $\boldsymbol{\beta}$ on both sides of equation (B.1)

$$\frac{\partial U(\bar{\alpha}_h)}{\partial \boldsymbol{\beta}} = \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})} \cdot \boldsymbol{x_{hi1}}, e^{Z_{hi2}(\boldsymbol{\beta})} \cdot \boldsymbol{x_{hi2}}, \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot \boldsymbol{x_{hij}}] \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$- \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \cdot der.V_{hi} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$- \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})}(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hi1}) \\ e^{\bar{\eta}_{syv}(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})}(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hi2}) \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})}(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hij}) \end{bmatrix} = \mathbf{0},$$

where the m'th diagonal element in $der.V_{hi}$ is

$$der.V_{hi}(m,m) = e^{Z_{him}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{him} + \frac{e^{2Z_{him}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + 2\boldsymbol{x}_{him}),$$

the m, n'th element is

$$der.V_{hi}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{him} + \boldsymbol{x}_{hin}),$$

After some simplification, we can obtain

$$\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{A_h}{B_h},$$

where

$$\begin{aligned}
A_h = \quad & \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hi1}, e^{Z_{hi2}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hi2}, \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hij}] \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix} \\
& - \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \cdot der.V_{hi1} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix} \\
& - \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{syvi1}(\boldsymbol{\beta})} \boldsymbol{x}_{hi1} \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \boldsymbol{x}_{hi2} \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \boldsymbol{x}_{hij} \end{bmatrix},
\end{aligned}$$

$$B_h = \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$+ \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \cdot der.V_{hi2} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix},$$

the m'th diagonal element of $der.V_{hi2}$ is

$$der.V_{hi2}(m,m) = \frac{e^{2Z_{him}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t},$$

the m,n'th element of $der.V_{hi2}$ is

$$der.V_{hi2}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s},$$

the m'th diagonal element of $der.V_{hi1}$ is

$$der.V_{hi1}(m,m) = e^{Z_{him}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{him} + \frac{e^{2Z_{him}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t} \cdot 2\boldsymbol{x}_{hmk},$$

the m,n'th element of $der.V_{hi1}$ is

$$der.V_{hi1}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s} \cdot (\boldsymbol{x}_{hmk} + \boldsymbol{x}_{hnk}).$$

Furthermore, we can obtain

$$\frac{\partial^2 U(\bar{\alpha}_h)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \frac{A1_h}{B1_h},$$

where

$$B1_h = \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$+ \sum_i [e^{Z_{hi1}(\beta)}, e^{Z_{hi2}(\beta)}, \cdots e^{Z_{hij}(\beta)}] \cdot V_{hi}^{-1} \cdot derr.V_{hi2} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix},$$

$$A1_h = \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hi1} \cdot \boldsymbol{x}'_{hi1}, \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hi1} \cdot \boldsymbol{x}'_{hi1}] \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$- \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hi1}, \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hij}] \cdot V_{hi}^{-1} \cdot der.V'_{hi} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$- \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hi1}, \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{hij}] \cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi1}(\boldsymbol{\beta})}\big(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hi1}\big) \\ e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi2}(\boldsymbol{\beta})}\big(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hi2}\big) \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hij}(\boldsymbol{\beta})}\big(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hij}\big) \end{bmatrix}$$

$$- \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})} \cdot \boldsymbol{x}'_{hi1}, \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot \boldsymbol{x}'_{hij}] \cdot V_{hi}^{-1} \cdot der.V'_{hi} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$+\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \cdot der.V'_{hi} \cdot V_{hi}^{-1} \cdot der.V_{hi} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$-\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \cdot derr.V_{hi1} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\eta}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$+\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \cdot der.V_{hi} \cdot V_{hi}^{-1} \cdot der.V'_{hi} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$+\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \cdot der.V_{hi} \cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})}(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hi1}) \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})}(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hi2}) \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})}(\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hij}) \end{bmatrix}$$

$$-\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}\cdot \boldsymbol{x}'_{hi1}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}\cdot \boldsymbol{x}'_{hij}]\cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})}(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hi1}) \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})}(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hi2}) \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})}(\frac{\partial\bar{\eta}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hij}) \end{bmatrix}$$

$$+\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}]\cdot V_{hi}^{-1}\cdot der.V'_{hi}\cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})}(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hi1}) \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})}(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hi2}) \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})}(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hij}) \end{bmatrix}$$

$$-\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}]\cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})}(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hi1})(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}+\boldsymbol{x}'_{hi1}) \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})}(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hi2})(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}+\boldsymbol{x}'_{hi2}) \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})}(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}+\boldsymbol{x}_{hij})(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}+\boldsymbol{x}'_{hij}) \end{bmatrix},$$

where the m'th diagonal element of $der.V'_{hi}$ is

$$der.V'_{hi}(m,m) = e^{Z_{him}(\boldsymbol{\beta})}\cdot \boldsymbol{x}'_{him} + \frac{e^{2Z_{him}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t}\cdot(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}+2\boldsymbol{x}'_{him}),$$

the m,n'th element of $der.V'_{hi}$ is

$$der.V'_{hi}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k_s}}\cdot(\frac{\partial\bar{\alpha}_h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}+\boldsymbol{x}'_{him}+\boldsymbol{x}'_{hin}),$$

the m'th diagonal element of $derr.V_{hi1}$ is

$$derr.V_{hi1}(m,m) = e^{Z_{him}(\boldsymbol{\beta})} \cdot \boldsymbol{x}_{him}\boldsymbol{x}'_{him} + \frac{e^{2Z_{him}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + 2\boldsymbol{x}_{him}) \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + 2\boldsymbol{x}'_{him}),$$

the m, n'th element of $derr.V_{hi1}$ is

$$derr.V_{hi1}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hmk} + \boldsymbol{x}_{hnk}) \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta}')}{\partial \boldsymbol{\beta}} + \boldsymbol{x}'_{hmk} + \boldsymbol{x}'_{hnk}),$$

the m'th diagonal element of $derr.V_{hi2}$ is

$$derr.V_{hi2}(m,m) = \frac{e^{2Z_{him}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t},$$

the m, n'th element of $derr.V_{hi2}$ is

$$derr.V_{hi2}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s}.$$

B.1.2 Updating algorithm for $\boldsymbol{\beta}$

The estimating equation of $\boldsymbol{\beta}$ is

$$U(\boldsymbol{\beta}) = \sum_{h,i} [e^{Z_{hi1}(\boldsymbol{\beta})} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hi1}), \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hij})] \cdot V_{hi}^{-1} \cdot \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix},$$

where the m'th diagonal element of V_{hi} is

$$V_{hi}(m,m) = e^{Z_{him}(\boldsymbol{\beta})} + \frac{e^{2Z_{him}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t},$$

the m, n'th element of V_{hi} is

$$V_{hi}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s}.$$

Then

$$\frac{\partial U(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \sum_{h,i} [e^{Z_{hi1}(\boldsymbol{\beta})} \cdot ((\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hi1}) \cdot \boldsymbol{x}'_{syvi1} + \frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}), \cdots$$

$$e^{Z_{hij}(\boldsymbol{\beta})} \cdot ((\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hij}) \cdot \boldsymbol{x}'_{syvij} + \frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'})] \cdot V_{hi}^{-1} \cdot \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$-[e^{Z_{hi1}(\boldsymbol{\beta})} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hi1}), \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hij})]$$

$$\cdot V_{hi}^{-1} \cdot der.V'_{hi} \cdot V_{hi}^{-1} \cdot \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$-[e^{Z_{hi1}(\boldsymbol{\beta})} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hi1}), \cdots e^{Z_{hij}(\boldsymbol{\beta})} \cdot (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \boldsymbol{x}_{hij})] \cdot V_{hi}^{-1}$$

$$\cdot \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi1}(\boldsymbol{\beta})} (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hi1}) \\ e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hi2}(\boldsymbol{\beta})} (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hi2}) \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta}) + Z_{hij}(\boldsymbol{\beta})} (\frac{\partial \bar{\alpha}_h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{x}'_{hij}) \end{bmatrix},$$

where the m'th diagonal element of $der.V'_{hi}$ is

$$der.V'_{hi}(m,m) = e^{Z_{him}(\boldsymbol{\beta})} \cdot \boldsymbol{x}'_{him} + \frac{e^{2Z_{him}(\boldsymbol{\beta}) + \bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t} \cdot \boldsymbol{x}'_{him},$$

the m, n'th element of $der.V'_{hi}$ is

$$der.V'_{hi}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s} \cdot (\boldsymbol{x}'_{him} + \boldsymbol{x}'_{hin}),$$

The algorithm (2.5) for $\boldsymbol{\beta}$ in ***Step* 1** can be written in form

$$\boldsymbol{\beta}^{(j+1)} = \boldsymbol{\beta}^{(j)} - \left\{ \sum_{h,i} \frac{\partial U_{hi}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(j)}} \right\}^{-1} \cdot \left\{ \sum_{h,i} U_{hi}(\boldsymbol{\beta}^{(j)}) \right\}.$$

B.1.3 Updating algorithm for $\boldsymbol{\alpha}$

Recall that the h'th row in the estimating equation of $U(\boldsymbol{\alpha})$ is

$$U(\alpha_h) = \sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix} = 0$$

The h'th diagonal element in $\partial U(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}'$ is

$$\frac{\partial U(\alpha_h)}{\partial \alpha_h} = -\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \cdot der.V_{hi} \cdot V_{hi}^{-1} \begin{bmatrix} y_{hi1} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ y_{hi2} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ y_{hij} - e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix}$$

$$-\sum_i [e^{Z_{hi1}(\boldsymbol{\beta})}, e^{Z_{hi2}(\boldsymbol{\beta})}, \cdots e^{Z_{hij}(\boldsymbol{\beta})}] \cdot V_{hi}^{-1} \begin{bmatrix} e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi1}(\boldsymbol{\beta})} \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hi2}(\boldsymbol{\beta})} \\ \vdots \\ e^{\bar{\alpha}_h(\boldsymbol{\beta})+Z_{hij}(\boldsymbol{\beta})} \end{bmatrix},$$

where the m'th diagonal element and the off-diagonal element in m'th row, n'th column of $der.V_{hi}$ are

$$der.V_{hi}(m,m) = \frac{e^{2Z_{him}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_t},$$

$$der.V_{hi}(m,n) = \frac{e^{Z_{him}(\boldsymbol{\beta})+Z_{hin}(\boldsymbol{\beta})+\bar{\alpha}_h(\boldsymbol{\beta})}}{\hat{k}_s}.$$

The off-diagonal elements in $\partial U(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}'$ are all equal to zero. The algorithm (2.5) for $\boldsymbol{\alpha}$ in ***Step* 1** is

$$\boldsymbol{\alpha}^{(j+1)} = \boldsymbol{\alpha}^{(j)} - \left\{ \frac{\partial U(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}'}|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(j)}} \right\}^{-1} \cdot \{U(\boldsymbol{\alpha}^{(j)})\}.$$

B.2 Variance parameter estimation

The covariance matrix $\boldsymbol{V}_{hi}$ in Eq.(2.2) can be decomposed as

$$\boldsymbol{V}_{hi} = \boldsymbol{A}_{hi}^{\frac{1}{2}}(k_t)\boldsymbol{R}_{hi}(k_t, k)\boldsymbol{A}_{hi}^{\frac{1}{2}}(k_t)$$

where $\boldsymbol{A}_{hi}(k_t) = diag(\sqrt{\upsilon(k_t, \boldsymbol{\mu_{hi}})})$, υ is the variance function and

$$\upsilon(k_t, \boldsymbol{\mu_{hi}}) = \boldsymbol{\mu_{hi}} + \frac{\boldsymbol{\mu_{hi}}^2}{k_t},$$

$\boldsymbol{R}_{hi}(k_t, k)$ is the correlation matrix, the diagonal elements of $\boldsymbol{R}_{hi}(k_t, k)$ are all equal to 1, the m, n'th element of $\boldsymbol{R}_{hi}(k_t, k)$ is equal to the correlation between Y_{him} and Y_{hin}

$$\boldsymbol{R}_{hi}(m, n) = \frac{\frac{\mu_{hi1}\mu_{hi2}}{k_s}}{\sqrt{\mu_{hi1} + \frac{\mu_{hi1}^2}{k_t}} \cdot \sqrt{\mu_{hi2} + \frac{\mu_{hi2}^2}{k_t}}}.$$

Given the regression parameter estimates $\hat{\boldsymbol{\beta}}$, the GEE function proposed by Zhang and Paul (2013) [49] for the variance parameter of our model has the following form

$$U(\tau) = \sum_{h,i}[\frac{\partial\nu(\tau, \mu_{hi1}(\hat{\boldsymbol{\beta}}))}{\partial\tau}, \cdots, \frac{\partial\nu(\tau, \mu_{hij}(\hat{\boldsymbol{\beta}}))}{\partial\gamma}]\boldsymbol{V}_{hi}^{-1}(\tau)\begin{bmatrix} r_{hi1}^2 - \nu(\tau, \mu_{hi1}(\hat{\boldsymbol{\beta}})) \\ r_{hi2}^2 - \nu(\tau, \mu_{hi2}(\hat{\boldsymbol{\beta}})) \\ \vdots \\ r_{hij}^2 - \nu(\tau, \mu_{hij}(\hat{\boldsymbol{\beta}})) \end{bmatrix},$$

where $\nu(\tau, \mu_{hij}(\hat{\boldsymbol{\beta}})) = \mu_{hij}(\hat{\boldsymbol{\beta}}) + \tau\mu_{hij}^2(\hat{\boldsymbol{\beta}})$, $r_{hij} = y_{hij} - \mu_{hij}(\hat{\boldsymbol{\beta}})$, and

$\boldsymbol{V}_{hi} = \sqrt{var(\boldsymbol{r}^2_{hi})}corr(\boldsymbol{r}^2_{hi})\sqrt{var(\boldsymbol{r}^2_{hi})}$. Consider the difficulty in modeling the structure of $corr(\boldsymbol{r}^2_{hi})$, we use the identity matrix for $corr(\boldsymbol{r}^2_{hi})$ since it doesn't affect the consistency of the estimate of τ (Zhang and Paul, 2013 [49]). Furthermore, we use NB moments to approximate $var(\boldsymbol{r}^2_{hi})$,

$$var(r^2_{hij}) = E(r^4_{hij}) - [E(r^2_{hij})]^2,$$

where $[E(r^2_{hij})]^2 = \mu_{hij} + \mu^2_{hij} \cdot \tau$, and the fourth central moment is

$$E(r^4_{hij}) = E(y^4_{hij}) - 4E(y^3_{hij})E(y_{hij}) + 6E(y^2_{hij})[E(y_{hij})]^2 - 3[E(y_{hij})]^4,$$

where the sample moments is

$$E(y^2_{hij}) = \frac{r \cdot (1-p_{hij})}{p_{hij}} + \frac{r \cdot (r+1) \cdot (1-p_{hij})^2}{p^2_{hij}},$$

$$E(y^3_{hij}) = \frac{r \cdot (1-p_{hij})}{p_{hij}} + \frac{3 \cdot r \cdot (r+1) \cdot (1-p_{hij})^2}{p^2_{hij}} + \frac{r \cdot (r+1) \cdot (r+2) \cdot (1-p_{hij})^3}{p^3_{hij}},$$

$$E(y^4_{hij}) = \frac{r \cdot (1-p_{hij})}{p_{hij}} + \frac{7 \cdot r \cdot (r+1) \cdot (1-p_{hij})^2}{p^2_{hij}} + \frac{6 \cdot r \cdot (r+1) \cdot (r+2) \cdot (1-p_{hij})^3}{p^3_{hij}},$$

$$+ \frac{r \cdot (r+1) \cdot (r+2) \cdot (r+3) \cdot (1-p_{hij})^4}{p^4_{hij}},$$

where $r = 1/\tau$ and $p_{hij} = 1/(1 + \mu_{hij} \cdot \tau)$.

The estimating equation of τ is

$$U(\tau) = \sum_{h,i} \sum_{k=1}^{j} \frac{\mu^2_{hik} \cdot [r^2_{hik} - (\mu_{hik} + \mu^2_{hik} \cdot \tau)]}{var(r^2_{hik})},$$

further we obtain

$$\frac{\partial U(\tau)}{\partial \tau} = -\sum_{h,i}\sum_{k=1}^{j}\frac{\mu_{hij}^4}{var(r_{hij}^2)},$$

We start with initial value $\tau^{(0)}$, use algorithm

$$\tau^{(j+1)} = \tau^{(j)} - \left\{\frac{\partial U(\tau)}{\partial \tau}|_{\tau=\tau^{(j)}}\right\}^{-1} \cdot U(\tau^{(j)}),$$

to estimate τ.

B.3 Correlation parameter estimation

We use pseudo-likelihood approach to estimate the correlation parameter k_s, we estimate $\xi = 1/k_s$ for convenience. Given estimates of regression parameters $\hat{\boldsymbol{\beta}}$ and variance parameter $\hat{\tau}$, the pseudo-likelihood function for ξ is

$$U(\xi) = \sum_{h,i} \text{trace}[\boldsymbol{P}_{hi}(\hat{\tau},\xi)\{\boldsymbol{\epsilon_{hi}}(\hat{\tau})\boldsymbol{\epsilon_{hi}}'(\hat{\tau}) - \boldsymbol{R_{hi}}(\hat{\tau},\xi)\}], \quad \text{(B.2)}$$

where $\boldsymbol{\epsilon_{hi}}(\hat{\tau}) = \boldsymbol{A_{hi}}^{-1/2}(\hat{\tau})\{\boldsymbol{y_{hi}} - \boldsymbol{\mu_{hi}}(\hat{\boldsymbol{\beta}})\}$ and

$$\boldsymbol{\epsilon}'_{hi} = (\frac{y_{hi1} - \mu_{hi1}(\hat{\boldsymbol{\beta}})}{\sqrt{\mu_{hi1}(\hat{\boldsymbol{\beta}}) + \hat{\gamma}\cdot\mu_{hi1}^2(\hat{\boldsymbol{\beta}})}}, \frac{y_{hi2} - \mu_{hi2}(\hat{\boldsymbol{\beta}})}{\sqrt{\mu_{hi2}(\hat{\boldsymbol{\beta}}) + \hat{\tau}\cdot\mu_{hi2}^2(\hat{\boldsymbol{\beta}})}}, \cdots, \frac{y_{hij} - \mu_{hij}(\hat{\boldsymbol{\beta}})}{\sqrt{\mu_{hij}(\hat{\boldsymbol{\beta}}) + \hat{\tau}\cdot\mu_{hij}^2(\hat{\boldsymbol{\beta}})}}),$$

$\boldsymbol{P_{hi}}(\hat{\tau},\xi) = \boldsymbol{R_{hi}}^{-1}(\hat{\tau},\xi)\{\partial\boldsymbol{R_{hi}}(\hat{\tau},\xi)/\partial\xi\}\boldsymbol{R_{hi}}^{-1}(\hat{\tau},\xi)$, $\boldsymbol{A}_{hi}(\hat{\tau}) = diag(\sigma_{hij}^2)$, $\sigma_{hij}^2 = \mu_{hij} + \mu_{hij}^2\cdot\hat{\tau}$, $\boldsymbol{R}_{hi}(\hat{\tau},\xi) = \boldsymbol{A}_{hi}^{-1/2}(\hat{\tau})\boldsymbol{V}_{hi}(\hat{\tau},\xi)\boldsymbol{A}_{hi}^{-1/2}(\hat{\tau})$, where $\boldsymbol{V}_{hi}(\hat{\tau},\xi)$ is the covariance matrix of $\boldsymbol{y}_{hi}$. $\boldsymbol{R}_{hi}$ is a symmetric matrix with

$$\boldsymbol{R}_{hi}(m,m) = 1,$$

$$\boldsymbol{R}_{hi}(m,n) = \frac{\xi \cdot \mu_{him}(\hat{\boldsymbol{\beta}}) \cdot \mu_{hin}(\hat{\boldsymbol{\beta}})}{\sqrt{\mu_{him}(\hat{\boldsymbol{\beta}}) + \hat{\tau} \cdot \mu_{him}^2(\hat{\boldsymbol{\beta}})} \cdot \sqrt{\mu_{hin}(\hat{\boldsymbol{\beta}}) + \hat{\tau} \cdot \mu_{hin}^2(\hat{\boldsymbol{\beta}})}}.$$

$\frac{\partial \boldsymbol{R}_{hi}}{\partial \xi}$ is a symmetric matrix with

$$\frac{\partial \boldsymbol{R}_{hi}}{\partial \xi}(m,m) = 0,$$

$$\frac{\partial \boldsymbol{R}_{hi}}{\partial \xi}(m,n) = \frac{\mu_{him}(\hat{\boldsymbol{\beta}}) \cdot \mu_{hin}(\hat{\boldsymbol{\beta}})}{\sqrt{\mu_{him}(\hat{\boldsymbol{\beta}}) + \hat{\tau} \cdot \mu_{him}^2(\hat{\boldsymbol{\beta}})} \cdot \sqrt{\mu_{hin}(\hat{\boldsymbol{\beta}}) + \hat{\tau} \cdot \mu_{hin}^2(\hat{\boldsymbol{\beta}})}}.$$

Furthermore,

$$\frac{\partial U(\xi)}{\partial \xi} = \sum_{h,i}(-2\boldsymbol{\epsilon}_{hi}'\boldsymbol{P}_{hi}\boldsymbol{R}_{hi}\boldsymbol{P}_{hi}\boldsymbol{\epsilon}_{hi} + \text{trace}\{\boldsymbol{P}_{hi}\frac{\partial \boldsymbol{R}_{hi}}{\partial \xi}\}).$$

We start with initial value $\xi^{(0)}$, use update function

$$\xi^{(j+1)} = \xi^{(j)} - (\frac{\partial U(\xi)}{\partial \xi}|_{\xi=\xi^{(j)}})^{-1} \cdot U(\xi^{(j)}),$$

to estimate ξ.

B.4 Initial value of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ for GEE approach

Initial value is an important issue in Newton-Raphson method. In this section, we introduce the method of obtaining initial value for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Consider the linear fixed effects model

$$y_{hij} = \alpha_h + \boldsymbol{x_{hij}}' \cdot \boldsymbol{\beta} + \varepsilon_{hij}$$

where y_{hij} is the response variable, α_h is a measure of the effect of the h'th treatment, $\boldsymbol{\beta}$ is the regression coefficients, ε_{hij} is the random variable and $\boldsymbol{x_{hij}}$ is the covariates. We can rewrite this model in the form

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \cdots \alpha_H, \beta_1, \beta_2, \cdots \beta_p)'$,

$$\boldsymbol{Y} = (y_{111}, y_{112}, \cdots y_{11n_{11}}, y_{121}, \cdots y_{12n_{12}}, \cdots y_{1n_1n_{1n_1}}, y_{211}, \cdots y_{2n_2n_{2n_2}}, \cdots y_{Hn_Hn_{Hn_H}})'$$

$$\boldsymbol{\varepsilon} = (\varepsilon_{111}, \varepsilon_{112}, \cdots \varepsilon_{11n_{11}}, \varepsilon_{121}, \cdots \varepsilon_{12n_{12}}, \cdots \varepsilon_{1n_1n_{1n_1}}, \varepsilon_{211}, \cdots \varepsilon_{2n_2n_{2n_2}}, \cdots \varepsilon_{Hn_Hn_{Hn_H}})'$$

and

$$\boldsymbol{Z} = \begin{pmatrix}
1 & 0 & \cdots & 0 & x_{1111} & x_{1112} & \cdots & x_{111p} \\
1 & 0 & \cdots & 0 & x_{1121} & x_{1122} & \cdots & x_{112p} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
1 & 0 & \cdots & 0 & x_{11n_{11}1} & x_{11n_{11}2} & \cdots & x_{11n_{11}p} \\
1 & 0 & \cdots & 0 & x_{1211} & x_{1212} & \cdots & x_{121p} \\
1 & 0 & \cdots & 0 & x_{1221} & x_{1222} & \cdots & x_{122p} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
1 & 0 & \cdots & 0 & x_{1n_1n_{1n_1}1} & x_{1n_1n_{1n_1}2} & \cdots & x_{1n_1n_{1n_1}p} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 1 & x_{H111} & x_{H112} & \cdots & x_{H11p} \\
0 & 0 & \cdots & 1 & x_{H121} & x_{H122} & \cdots & x_{H12p} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 1 & x_{H1n_{Hn_{H1}}1} & x_{H1n_{Hn_{H1}}2} & \cdots & x_{H1n_{Hn_{H1}}p} \\
0 & 0 & \cdots & 1 & x_{H211} & x_{H212} & \cdots & x_{H21p} \\
0 & 0 & \cdots & 1 & x_{H221} & x_{H222} & \cdots & x_{H22p} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & 1 & x_{Hn_Hn_{Hn_H}1} & x_{Hn_Hn_{Hn_H}2} & \cdots & x_{Hn_Hn_{Hn_H}p}
\end{pmatrix}$$

Then, R command 'glm' could be used to solve this linear model, treating $\boldsymbol{\alpha}$ as 'factor'. The obtained estimates of $\boldsymbol{\beta}$ is the initial value of regression parameters for GEE method.

For sites with no replicates, we can use R command 'nlminb' to check the profile GEE method, since in this case the GEE estimating equation has same form as the score function of NB model.

Appendix C

TMB: C++ template function

The C++ template for this model is

```
#include <TMB.hpp>

template<class Type>

Type objective_function<Type>::operator() ()

{

  DATA_INTEGER(nstratum);

  DATA_INTEGER(nsite);

  DATA_INTEGER(nobs);

  DATA_IVECTOR(istratum);

  DATA_IVECTOR(isite);
```

```
DATA_MATRIX(Xmatrix);
DATA_VECTOR(y);
DATA_SCALAR(log_k);
DATA_SCALAR(log_kc);

PARAMETER_VECTOR(beta);
PARAMETER_VECTOR(log_eta);
PARAMETER_VECTOR(log_site);

vector<Type> eta(nstratum);
eta=exp(log_eta);
vector<Type> log_mu(nobs);
log_mu=log_eta(istratum-1)+Xmatrix*beta;

vector<Type> log_mui(nobs);
log_mui=log_mu+log_site(isite-1);
vector<Type> mui=exp(log_mui);

vector<Type> site=exp(log_site);
```

```
  Type kc=exp(log_kc);
  Type k=exp(log_k);
  Type nll=0.0;

   for(int i=0;i<nobs;i++){
   nll-= lgamma(y(i)+kc)-lgamma(kc)-lgamma(y(i)+Type(1.0))+y(i)*log_mui(i)-y(i)*
log(mui(i)+kc)+kc*(log_kc-log(mui(i)+kc));
   }

   vector<Type> lli;
   lli=k*log_k-lgamma(k)+k*log_site-k*site;

   nll-=sum(lli);

  ADREPORT(k)
  ADREPORT(kc)
  ADREPORT(eta)
```

```
  return nll;



}
```

Appendix D

Derivation for ML and REML estimator of σ^2

Let $\boldsymbol{y}$ be a $n \times 1$ vector of sample response, $\boldsymbol{\beta}$ be a $p \times 1$ parameter vector and $\boldsymbol{X}$ be a $n \times p$ covariance matrix, we assume a linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \text{where} \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}).$$

The likelihood function is

$$L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \right\}.$$

We could obtain the ML estimators for $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}}_{\text{ML}} = (\boldsymbol{X'}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

D.1 REML for normal linear mixed model

To start REML method, we choose $\boldsymbol{k'} = (\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})$ so that $\boldsymbol{k'}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$ and

$$\boldsymbol{k'}\boldsymbol{y} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{k'}\boldsymbol{k}).$$

The REML function is

$$f(\sigma^2|\boldsymbol{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}|\boldsymbol{k'}\boldsymbol{k}|^{1/2}} \exp\{-\frac{1}{2\sigma^2}(\boldsymbol{k'}\boldsymbol{y})'(\boldsymbol{k'}\boldsymbol{k})^{-1}(\boldsymbol{k'}\boldsymbol{y})\}$$

and the REML estimator of σ^2 is

$$\hat{\sigma}^2_{\mathrm{REML}} = \frac{(\boldsymbol{k'}\boldsymbol{y})'(\boldsymbol{k'}\boldsymbol{k})^{-1}(\boldsymbol{k'}\boldsymbol{y})}{n}$$

Since

$$\begin{aligned} E(\hat{\sigma}^2_{\mathrm{REML}}) &= \frac{E[(\boldsymbol{k'}\boldsymbol{y})'(\boldsymbol{k'}\boldsymbol{k})^{-1}(\boldsymbol{k'}\boldsymbol{y})]}{n} \\ &= \frac{tr[(\boldsymbol{k'}\boldsymbol{k})^{-1}E(\boldsymbol{k'}\boldsymbol{y}\boldsymbol{y'}\boldsymbol{k})]}{n} \\ &= \frac{tr[(\boldsymbol{k'}\boldsymbol{k})^{-1}\boldsymbol{k'}Var(\boldsymbol{y})\boldsymbol{k}]}{n} \\ &= \frac{\sigma^2 tr[(\boldsymbol{k'}\boldsymbol{k})^{-1}\boldsymbol{k'}\boldsymbol{k}]}{n} \\ &= \sigma^2. \end{aligned} \tag{D.1}$$

$\hat{\sigma}^2_{\mathrm{REML}}$ is an unbiased estimator of σ^2

D.2 Integrated REML

We use integrated REML to estimate σ^2. The REML function can be obtained by integrating the likelihood function over $\boldsymbol{\beta}$ using a non-informative density function

$$\begin{aligned} L(\sigma^2|\boldsymbol{y}) &= \int L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) d\boldsymbol{\beta} \\ &= \int \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\right\} d\boldsymbol{\beta}. \end{aligned}$$

Note that

$$(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta}) = \boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y} + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{\text{ML}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{\text{ML}}),$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.

$$\begin{aligned} L(\sigma^2|\boldsymbol{y}) &= \frac{|\boldsymbol{X}'\boldsymbol{X}|^{-1/2}}{(\sqrt{2\pi\sigma^2})^{n-p}} \exp\left\{ -\frac{1}{2\sigma^2}\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}\right\} \\ &\quad \cdot \int \frac{1}{(\sqrt{2\pi\sigma^2})^p|\boldsymbol{X}'\boldsymbol{X}|^{-1/2}} \exp\left\{ -\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{\text{ML}})'[(\boldsymbol{X}'\boldsymbol{X})^{-1}]^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{\text{ML}})\right\} d\boldsymbol{\beta} \\ &= \frac{|\boldsymbol{X}'\boldsymbol{X}|^{-1/2}}{(\sqrt{2\pi\sigma^2})^{n-p}} \exp\left\{ -\frac{1}{2\sigma^2}\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}\right\}. \end{aligned}$$

Then the REML estimator of σ^2 is

$$\hat{\sigma}^2_{\text{REML}} = \frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}}{n-p},$$

$$\begin{aligned} E(\hat{\sigma}^2_{\text{REML}}) &= \frac{E[\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}]}{n-p} \\ &= \frac{E[tr\{(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}\boldsymbol{y}'\}]}{n-p} \\ &= \frac{tr[(\boldsymbol{I}-\boldsymbol{H})E(\boldsymbol{y}\boldsymbol{y}')]}{n-p} \\ &= \frac{tr[(\boldsymbol{I}-\boldsymbol{H})Var(\boldsymbol{y}\boldsymbol{y}')]}{n-p} \\ &= \sigma^2. \end{aligned} \tag{D.2}$$

The integrated REML estimator of σ^2 is unbiased.

Appendix E

simulation table

Table E.1: Mean bias of regression and variance parameters for the GEE method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5	k_s	k_t
1	1	25	5	0.005	-0.002	0.004	-0.003	0.000	3.323	1.053
1	1	25	15	-0.006	0.003	-0.008	-0.001	-0.001	0.479	0.257
1	1	25	30	-0.003	0.003	-0.000	-0.007	-0.001	0.198	0.123
1	1	100	5	0.007	0.002	-0.001	0.004	-0.014	2.318	0.916
1	1	100	15	0.002	-0.002	0.002	0.001	-0.002	0.418	0.241
1	3	25	5	-0.003	-0.005	-0.001	-0.002	-0.001	5.157	3.071
1	3	25	15	0.002	0.002	0.001	0.002	0.002	2.088	0.852
1	3	25	30	0.001	0.002	0.002	0.003	-0.000	0.696	0.344
1	3	100	5	0.005	-0.001	0.002	0.002	-0.002	6.132	3.514
1	3	100	15	-0.000	-0.003	0.003	-0.004	0.000	1.535	0.780
5	1	25	5	0.010	0.002	0.011	0.002	0.007	1.237	0.613
5	1	25	15	0.005	0.008	0.004	-0.001	-0.000	0.237	0.135
5	1	25	30	-0.001	-0.002	0.002	-0.001	0.001	-0.005	-0.004
5	1	100	5	0.008	-0.006	0.002	-0.002	-0.003	1.110	0.589
5	1	100	15	-0.001	0.001	0.000	0.003	0.001	0.251	0.149
5	3	25	5	0.009	-0.004	-0.002	0.001	0.010	2.968	1.607
5	3	25	15	0.000	-0.001	0.001	-0.001	0.000	0.651	0.419
5	3	25	30	0.001	0.001	-0.001	0.002	-0.000	0.307	0.215
5	3	100	5	0.002	0.002	0.002	-0.001	-0.004	2.439	1.403
5	3	100	15	0.002	-0.000	0.000	0.000	0.001	0.580	0.380
10	1	25	5	-0.004	-0.014	-0.006	-0.001	-0.007	0.951	0.491
10	1	25	15	0.000	0.003	-0.003	-0.000	-0.000	0.115	0.062
10	1	25	30	-0.002	-0.003	0.000	0.002	-0.001	-0.183	-0.116
10	1	100	5	0.002	0.002	-0.000	-0.001	-0.004	0.974	0.528
10	1	100	15	-0.002	-0.001	-0.001	0.000	-0.003	0.123	0.080
10	3	25	5	0.002	0.001	0.005	-0.002	-0.003	1.956	1.029
10	3	25	15	0.000	-0.001	0.002	-0.001	-0.005	0.296	0.173
10	3	25	30	-0.001	0.000	-0.000	-0.001	0.002	0.205	0.144
10	3	100	5	0.001	0.002	0.002	-0.001	0.000	1.731	1.020
10	3	100	15	-0.001	0.000	0.000	-0.000	0.000	0.419	0.279

Table E.2: Mean bias of regression and variance parameters for the ML method

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5	k_s	k_t
1	1	25	5	0.011	0.013	0.003	-0.010	-0.009	6.062	1.015
1	1	25	15	0.001	-0.004	0.000	0.002	-0.000	2.289	0.296
1	1	25	30	0.007	-0.002	-0.001	0.003	-0.005	0.854	0.177
1	1	100	5	0.004	-0.004	0.003	-0.003	-0.004	6.706	0.767
1	1	100	15	0.004	-0.001	0.001	-0.001	-0.007	1.537	0.240
1	3	25	5	-0.012	-0.001	0.008	0.009	0.001	6.719	3.670
1	3	25	15	-0.001	0.001	-0.001	0.002	-0.002	5.654	1.321
1	3	25	30	-0.002	0.003	0.002	0.001	-0.002	3.441	0.694
1	3	100	5	0.004	-0.005	0.003	0.003	0.003	6.957	3.615
1	3	100	15	0.004	-0.004	0.001	-0.001	-0.002	6.273	1.004
5	1	25	5	-0.006	-0.004	0.004	-0.001	-0.007	0.607	0.327
5	1	25	15	0.003	-0.003	-0.003	0.003	-0.002	0.251	0.103
5	1	25	30	-0.001	0.004	0.001	0.001	-0.003	0.156	0.060
5	1	100	5	-0.000	-0.001	0.001	-0.007	0.004	0.521	0.275
5	1	100	15	0.001	-0.001	-0.001	-0.001	-0.001	0.237	0.086
5	3	25	5	-0.003	0.001	0.010	0.001	0.002	2.535	1.349
5	3	25	15	-0.002	0.003	-0.002	-0.002	-0.003	0.684	0.369
5	3	25	30	-0.001	-0.001	0.002	-0.000	0.002	0.358	0.193
5	3	100	5	-0.003	0.000	-0.003	-0.000	-0.002	1.913	1.094
5	3	100	15	0.002	0.000	-0.001	0.002	0.000	0.612	0.322
10	1	25	5	-0.004	-0.005	0.002	-0.003	-0.007	0.427	0.252
10	1	25	15	-0.001	-0.000	-0.000	0.001	-0.001	0.171	0.079
10	1	25	30	0.001	0.002	-0.002	0.001	-0.003	0.102	0.044
10	1	100	5	0.000	-0.003	0.001	0.004	-0.003	0.385	0.214
10	1	100	15	-0.000	-0.000	-0.001	0.002	-0.004	0.163	0.069
10	3	25	5	0.004	0.000	0.001	-0.003	0.001	1.527	0.958
10	3	25	15	0.003	0.001	0.000	-0.000	-0.002	0.437	0.271
10	3	25	30	0.000	0.002	-0.002	0.001	-0.001	0.223	0.130
10	3	100	5	0.002	0.001	0.001	0.000	-0.001	1.275	0.798
10	3	100	15	0.001	-0.000	0.001	0.001	0.001	0.399	0.230

Table E.3: Mean bias of regression and variance parameters for the REML method

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5	k_s	k_t
1	1	25	5	-0.012	0.011	0.003	-0.004	0.012	4.275	0.256
1	1	25	15	-0.004	-0.005	0.000	0.003	0.005	2.095	0.168
1	1	25	30	0.006	-0.002	-0.001	0.003	-0.004	0.670	0.150
1	1	100	5	-0.013	-0.007	0.003	0.000	0.014	3.080	0.232
1	1	100	15	-0.000	-0.002	0.001	-0.000	-0.003	1.086	0.149
1	3	25	5	-0.023	-0.003	0.008	0.011	0.010	4.790	0.866
1	3	25	15	-0.003	-0.000	-0.001	0.002	0.001	4.496	0.478
1	3	25	30	-0.003	0.003	0.002	0.002	-0.001	2.565	0.389
1	3	100	5	-0.003	-0.007	0.004	0.005	0.011	5.270	0.641
1	3	100	15	0.002	-0.004	0.001	-0.001	0.000	4.551	0.369
5	1	25	5	-0.015	-0.006	0.004	0.002	0.001	0.122	0.064
5	1	25	15	0.002	-0.003	-0.003	0.004	0.001	0.094	0.045
5	1	25	30	-0.002	0.004	0.001	0.001	-0.003	0.079	0.037
5	1	100	5	-0.007	-0.003	0.001	-0.005	0.012	0.085	0.052
5	1	100	15	-0.001	-0.002	-0.001	-0.001	0.001	0.082	0.038
5	3	25	5	-0.009	-0.000	0.011	0.002	0.007	0.417	0.161
5	3	25	15	-0.003	0.003	-0.002	-0.002	-0.001	0.200	0.083
5	3	25	30	-0.001	-0.001	0.002	-0.000	0.002	0.142	0.072
5	3	100	5	-0.008	-0.001	-0.004	0.001	0.003	0.155	0.089
5	3	100	15	0.001	0.000	-0.001	0.002	0.001	0.148	0.070
10	1	25	5	-0.011	-0.007	0.002	-0.001	-0.000	0.068	0.041
10	1	25	15	-0.003	-0.001	-0.001	0.001	0.000	0.051	0.027
10	1	25	30	0.001	0.002	-0.002	0.001	-0.002	0.042	0.021
10	1	100	5	-0.005	-0.005	0.001	0.006	0.002	0.048	0.032
10	1	100	15	-0.001	-0.001	-0.001	0.003	-0.003	0.045	0.024
10	3	25	5	-0.000	-0.001	0.001	-0.003	0.005	0.163	0.093
10	3	25	15	0.002	0.001	0.000	0.000	-0.000	0.080	0.045
10	3	25	30	-0.000	0.002	-0.002	0.001	-0.001	0.054	0.030
10	3	100	5	-0.001	-0.000	0.001	0.001	0.002	0.059	0.043
10	3	100	15	0.000	-0.000	0.001	0.001	0.002	0.050	0.027

Table E.4: Standard bias (sbias) of $\boldsymbol{\beta}$ for the GEE method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.026	-0.006	0.013	-0.003	-0.002
1	1	25	15	-0.025	0.016	-0.038	-0.008	-0.010
1	1	25	30	-0.013	0.025	-0.002	-0.052	-0.012
1	1	100	5	0.046	0.012	-0.005	0.030	-0.101
1	1	100	15	0.024	-0.019	0.018	0.005	-0.025
1	3	25	5	-0.025	-0.022	0.002	-0.014	0.012
1	3	25	15	0.006	0.011	0.005	0.014	0.015
1	3	25	30	0.011	0.014	0.015	0.030	-0.006
1	3	100	5	0.036	-0.006	0.012	0.016	-0.013
1	3	100	15	-0.006	-0.035	0.041	-0.047	0.005
5	1	25	5	0.053	0.026	0.054	0.010	0.022
5	1	25	15	0.053	0.062	0.034	-0.004	-0.005
5	1	25	30	-0.010	-0.016	0.022	-0.016	0.007
5	1	100	5	0.076	-0.057	0.019	-0.017	-0.033
5	1	100	15	-0.005	0.011	0.000	0.038	0.015
5	3	25	5	0.073	-0.030	-0.009	0.006	0.056
5	3	25	15	0.008	-0.006	0.015	-0.009	-0.004
5	3	25	30	0.010	0.009	-0.014	0.037	-0.013
5	3	100	5	0.036	0.036	0.023	-0.010	-0.054
5	3	100	15	0.046	-0.001	0.008	0.008	0.018
10	1	25	5	-0.012	-0.076	-0.026	-0.013	-0.053
10	1	25	15	0.012	0.021	-0.029	-0.002	-0.005
10	1	25	30	-0.018	-0.038	0.005	0.022	-0.021
10	1	100	5	0.029	0.015	0.000	-0.008	-0.048
10	1	100	15	-0.035	-0.012	-0.015	0.007	-0.045
10	3	25	5	0.013	0.005	0.049	-0.021	-0.029
10	3	25	15	0.008	-0.009	0.027	-0.013	-0.066
10	3	25	30	-0.016	0.002	-0.006	-0.027	0.029
10	3	100	5	0.026	0.026	0.035	-0.019	0.004
10	3	100	15	-0.015	0.006	0.006	-0.003	0.003

Table E.5: Standard bias (sbias) of $\boldsymbol{\beta}$ for the ML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.040	0.041	0.008	-0.035	-0.040
1	1	25	15	0.010	-0.018	-0.001	0.007	-0.010
1	1	25	30	0.057	-0.016	-0.011	0.021	-0.044
1	1	100	5	0.033	-0.019	0.016	-0.018	-0.034
1	1	100	15	0.045	-0.008	0.012	-0.011	-0.074
1	3	25	5	-0.030	0.005	0.030	0.028	-0.007
1	3	25	15	-0.002	0.006	-0.006	0.010	-0.016
1	3	25	30	-0.014	0.034	0.017	0.013	-0.021
1	3	100	5	0.037	-0.037	0.020	0.017	0.021
1	3	100	15	0.047	-0.048	0.016	-0.017	-0.027
5	1	25	5	-0.022	-0.013	0.018	-0.004	-0.036
5	1	25	15	0.022	-0.018	-0.018	0.022	-0.015
5	1	25	30	-0.012	0.042	0.011	0.014	-0.043
5	1	100	5	0.006	-0.012	0.010	-0.059	0.027
5	1	100	15	0.016	-0.015	-0.018	-0.018	-0.021
5	3	25	5	-0.011	0.009	0.060	0.008	0.001
5	3	25	15	-0.014	0.033	-0.014	-0.027	-0.028
5	3	25	30	-0.012	-0.022	0.036	-0.004	0.025
5	3	100	5	-0.040	0.003	-0.040	-0.003	-0.027
5	3	100	15	0.044	0.011	-0.024	0.038	0.003
10	1	25	5	-0.015	-0.024	0.010	-0.016	-0.037
10	1	25	15	-0.012	-0.002	-0.004	0.004	-0.011
10	1	25	30	0.017	0.025	-0.020	0.014	-0.036
10	1	100	5	0.006	-0.031	0.007	0.040	-0.033
10	1	100	15	-0.000	-0.005	-0.015	0.036	-0.062
10	3	25	5	0.032	0.007	0.002	-0.021	0.001
10	3	25	15	0.039	0.012	0.003	-0.004	-0.023
10	3	25	30	0.005	0.034	-0.041	0.011	-0.021
10	3	100	5	0.037	0.011	0.011	0.002	-0.022
10	3	100	15	0.027	-0.000	0.020	0.027	0.024

Table E.6: Standard bias (sbias) of $\boldsymbol{\beta}$ for the REML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	-0.020	0.031	0.005	-0.020	0.017
1	1	25	15	-0.013	-0.022	-0.001	0.013	0.014
1	1	25	30	0.047	-0.019	-0.011	0.024	-0.033
1	1	100	5	-0.057	-0.034	0.014	0.004	0.064
1	1	100	15	0.001	-0.019	0.011	-0.001	-0.032
1	3	25	5	-0.061	-0.002	0.028	0.034	0.021
1	3	25	15	-0.015	0.001	-0.006	0.014	-0.000
1	3	25	30	-0.023	0.032	0.016	0.015	-0.012
1	3	100	5	-0.011	-0.043	0.023	0.028	0.066
1	3	100	15	0.022	-0.054	0.014	-0.010	-0.003
5	1	25	5	-0.057	-0.024	0.017	0.008	0.001
5	1	25	15	0.012	-0.024	-0.019	0.025	0.002
5	1	25	30	-0.020	0.040	0.011	0.016	-0.033
5	1	100	5	-0.052	-0.028	0.008	-0.040	0.093
5	1	100	15	-0.010	-0.024	-0.018	-0.011	0.007
5	3	25	5	-0.044	0.001	0.059	0.014	0.034
5	3	25	15	-0.028	0.030	-0.015	-0.022	-0.013
5	3	25	30	-0.020	-0.024	0.036	-0.003	0.033
5	3	100	5	-0.092	-0.012	-0.043	0.016	0.027
5	3	100	15	0.019	0.006	-0.023	0.045	0.025
10	1	25	5	-0.043	-0.034	0.010	-0.005	-0.006
10	1	25	15	-0.023	-0.005	-0.004	0.008	0.002
10	1	25	30	0.009	0.023	-0.020	0.015	-0.029
10	1	100	5	-0.043	-0.045	0.006	0.051	0.015
10	1	100	15	-0.018	-0.010	-0.015	0.042	-0.044
10	3	25	5	0.001	-0.000	0.003	-0.016	0.029
10	3	25	15	0.026	0.010	0.003	-0.001	-0.010
10	3	25	30	-0.002	0.034	-0.041	0.013	-0.013
10	3	100	5	-0.011	-0.001	0.010	0.012	0.023
10	3	100	15	0.007	-0.005	0.019	0.034	0.044

Table E.7: Root mean square error (RMSE) of $\boldsymbol{\beta}$ for the GEE method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.376	0.367	0.376	0.378	0.367
1	1	25	15	0.205	0.213	0.203	0.200	0.207
1	1	25	30	0.133	0.132	0.135	0.137	0.133
1	1	100	5	0.177	0.181	0.178	0.182	0.181
1	1	100	15	0.103	0.102	0.102	0.101	0.103
1	3	25	5	0.300	0.292	0.292	0.297	0.306
1	3	25	15	0.163	0.163	0.162	0.166	0.165
1	3	25	30	0.105	0.106	0.107	0.108	0.110
1	3	100	5	0.144	0.144	0.143	0.147	0.149
1	3	100	15	0.083	0.082	0.083	0.081	0.083
5	1	25	5	0.269	0.258	0.262	0.261	0.265
5	1	25	15	0.151	0.148	0.148	0.151	0.151
5	1	25	30	0.098	0.094	0.094	0.094	0.095
5	1	100	5	0.129	0.127	0.123	0.126	0.128
5	1	100	15	0.075	0.072	0.075	0.074	0.075
5	3	25	5	0.179	0.176	0.175	0.172	0.178
5	3	25	15	0.098	0.096	0.097	0.100	0.101
5	3	25	30	0.064	0.064	0.063	0.063	0.064
5	3	100	5	0.086	0.086	0.086	0.085	0.084
5	3	100	15	0.049	0.048	0.049	0.048	0.050
10	1	25	5	0.234	0.244	0.238	0.245	0.244
10	1	25	15	0.137	0.139	0.138	0.139	0.138
10	1	25	30	0.089	0.091	0.090	0.088	0.087
10	1	100	5	0.118	0.118	0.118	0.117	0.118
10	1	100	15	0.069	0.066	0.068	0.067	0.069
10	3	25	5	0.152	0.147	0.148	0.153	0.153
10	3	25	15	0.086	0.083	0.086	0.084	0.086
10	3	25	30	0.054	0.054	0.055	0.053	0.054
10	3	100	5	0.075	0.074	0.071	0.073	0.072
10	3	100	15	0.043	0.043	0.043	0.042	0.041

Table E.8: Root mean square error (RMSE) of $\boldsymbol{\beta}$ for the ML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.389	0.380	0.377	0.384	0.388
1	1	25	15	0.212	0.205	0.210	0.212	0.213
1	1	25	30	0.133	0.133	0.130	0.134	0.129
1	1	100	5	0.186	0.182	0.181	0.185	0.188
1	1	100	15	0.103	0.100	0.102	0.102	0.105
1	3	25	5	0.293	0.295	0.298	0.308	0.302
1	3	25	15	0.160	0.166	0.165	0.162	0.167
1	3	25	30	0.107	0.110	0.108	0.108	0.107
1	3	100	5	0.145	0.148	0.144	0.142	0.145
1	3	100	15	0.081	0.082	0.080	0.083	0.081
5	1	25	5	0.253	0.258	0.252	0.261	0.255
5	1	25	15	0.142	0.142	0.141	0.143	0.140
5	1	25	30	0.089	0.091	0.090	0.090	0.091
5	1	100	5	0.124	0.122	0.121	0.127	0.123
5	1	100	15	0.070	0.070	0.069	0.069	0.070
5	3	25	5	0.180	0.169	0.176	0.174	0.174
5	3	25	15	0.098	0.097	0.098	0.099	0.099
5	3	25	30	0.065	0.064	0.062	0.061	0.063
5	3	100	5	0.085	0.085	0.084	0.084	0.086
5	3	100	15	0.050	0.049	0.047	0.049	0.048
10	1	25	5	0.226	0.220	0.225	0.223	0.225
10	1	25	15	0.127	0.125	0.124	0.129	0.130
10	1	25	30	0.083	0.081	0.082	0.081	0.082
10	1	100	5	0.108	0.111	0.108	0.110	0.110
10	1	100	15	0.063	0.060	0.063	0.061	0.062
10	3	25	5	0.145	0.143	0.143	0.146	0.149
10	3	25	15	0.084	0.082	0.083	0.083	0.085
10	3	25	30	0.052	0.054	0.053	0.053	0.053
10	3	100	5	0.073	0.070	0.070	0.072	0.070
10	3	100	15	0.041	0.041	0.041	0.041	0.041

Table E.9: Root mean square error (RMSE) of $\boldsymbol{\beta}$ for the REML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.396	0.389	0.385	0.393	0.397
1	1	25	15	0.212	0.206	0.211	0.213	0.214
1	1	25	30	0.133	0.133	0.130	0.134	0.128
1	1	100	5	0.187	0.181	0.181	0.185	0.190
1	1	100	15	0.103	0.099	0.102	0.102	0.105
1	3	25	5	0.298	0.299	0.302	0.311	0.306
1	3	25	15	0.161	0.166	0.165	0.163	0.168
1	3	25	30	0.107	0.110	0.108	0.107	0.107
1	3	100	5	0.147	0.149	0.145	0.143	0.147
1	3	100	15	0.081	0.082	0.080	0.083	0.081
5	1	25	5	0.255	0.257	0.254	0.261	0.256
5	1	25	15	0.141	0.141	0.141	0.142	0.140
5	1	25	30	0.089	0.091	0.089	0.090	0.091
5	1	100	5	0.124	0.122	0.121	0.126	0.124
5	1	100	15	0.070	0.070	0.069	0.069	0.070
5	3	25	5	0.180	0.170	0.176	0.174	0.174
5	3	25	15	0.098	0.097	0.098	0.099	0.099
5	3	25	30	0.064	0.064	0.062	0.061	0.063
5	3	100	5	0.085	0.085	0.084	0.084	0.086
5	3	100	15	0.050	0.049	0.047	0.049	0.048
10	1	25	5	0.226	0.220	0.225	0.223	0.225
10	1	25	15	0.127	0.125	0.124	0.129	0.130
10	1	25	30	0.083	0.081	0.082	0.081	0.082
10	1	100	5	0.108	0.111	0.107	0.110	0.110
10	1	100	15	0.063	0.060	0.063	0.061	0.062
10	3	25	5	0.145	0.143	0.143	0.146	0.149
10	3	25	15	0.084	0.082	0.083	0.083	0.084
10	3	25	30	0.052	0.054	0.053	0.053	0.053
10	3	100	5	0.073	0.070	0.070	0.071	0.070
10	3	100	15	0.041	0.041	0.041	0.041	0.041

Table E.10: Coverage for the lower limits of 95% confidence interval (C.L) of $\boldsymbol{\beta}$ for the GEE method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.073	0.062	0.068	0.075	0.074
1	1	25	15	0.039	0.044	0.040	0.042	0.038
1	1	25	30	0.033	0.031	0.035	0.038	0.033
1	1	100	5	0.056	0.053	0.047	0.057	0.038
1	1	100	15	0.039	0.039	0.038	0.034	0.033
1	3	25	5	0.054	0.049	0.054	0.054	0.073
1	3	25	15	0.033	0.035	0.029	0.041	0.042
1	3	25	30	0.029	0.030	0.032	0.028	0.036
1	3	100	5	0.056	0.050	0.042	0.054	0.058
1	3	100	15	0.036	0.037	0.037	0.029	0.034
5	1	25	5	0.080	0.066	0.074	0.062	0.069
5	1	25	15	0.052	0.052	0.051	0.041	0.051
5	1	25	30	0.034	0.026	0.032	0.032	0.037
5	1	100	5	0.062	0.048	0.052	0.052	0.053
5	1	100	15	0.036	0.038	0.037	0.035	0.040
5	3	25	5	0.062	0.055	0.050	0.050	0.057
5	3	25	15	0.033	0.032	0.033	0.040	0.035
5	3	25	30	0.030	0.033	0.029	0.031	0.029
5	3	100	5	0.051	0.044	0.048	0.040	0.034
5	3	100	15	0.032	0.029	0.035	0.033	0.034
10	1	25	5	0.061	0.065	0.068	0.075	0.061
10	1	25	15	0.045	0.048	0.041	0.041	0.044
10	1	25	30	0.036	0.037	0.038	0.037	0.028
10	1	100	5	0.050	0.054	0.048	0.054	0.054
10	1	100	15	0.032	0.029	0.030	0.031	0.034
10	3	25	5	0.056	0.052	0.061	0.056	0.056
10	3	25	15	0.040	0.032	0.038	0.031	0.036
10	3	25	30	0.028	0.030	0.035	0.026	0.031
10	3	100	5	0.055	0.052	0.041	0.046	0.038
10	3	100	15	0.033	0.034	0.035	0.039	0.033

Table E.11: Coverage for the lower limits of 95% confidence interval (C.L) of $\boldsymbol{\beta}$ for the ML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.039	0.035	0.035	0.031	0.032
1	1	25	15	0.027	0.025	0.028	0.034	0.030
1	1	25	30	0.029	0.021	0.023	0.028	0.018
1	1	100	5	0.037	0.029	0.037	0.031	0.031
1	1	100	15	0.033	0.023	0.029	0.025	0.025
1	3	25	5	0.029	0.035	0.034	0.041	0.035
1	3	25	15	0.029	0.028	0.029	0.026	0.029
1	3	25	30	0.025	0.034	0.028	0.026	0.021
1	3	100	5	0.031	0.034	0.031	0.035	0.028
1	3	100	15	0.035	0.021	0.026	0.029	0.024
5	1	25	5	0.031	0.034	0.034	0.038	0.035
5	1	25	15	0.026	0.028	0.029	0.029	0.019
5	1	25	30	0.024	0.025	0.028	0.029	0.023
5	1	100	5	0.025	0.029	0.030	0.028	0.031
5	1	100	15	0.022	0.020	0.023	0.022	0.022
5	3	25	5	0.033	0.028	0.041	0.029	0.030
5	3	25	15	0.025	0.030	0.029	0.024	0.023
5	3	25	30	0.024	0.025	0.023	0.016	0.024
5	3	100	5	0.025	0.032	0.029	0.032	0.025
5	3	100	15	0.032	0.029	0.024	0.030	0.025
10	1	25	5	0.032	0.029	0.035	0.029	0.028
10	1	25	15	0.030	0.025	0.022	0.025	0.027
10	1	25	30	0.029	0.023	0.025	0.028	0.025
10	1	100	5	0.025	0.035	0.034	0.036	0.032
10	1	100	15	0.025	0.021	0.025	0.026	0.019
10	3	25	5	0.037	0.030	0.026	0.034	0.029
10	3	25	15	0.030	0.028	0.030	0.027	0.021
10	3	25	30	0.025	0.032	0.021	0.024	0.023
10	3	100	5	0.040	0.026	0.029	0.032	0.033
10	3	100	15	0.030	0.026	0.029	0.029	0.021

Table E.12: Coverage for the lower limits of 95% confidence interval (C.L) of $\boldsymbol{\beta}$ for the REML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.026	0.024	0.022	0.024	0.024
1	1	25	15	0.022	0.021	0.028	0.032	0.031
1	1	25	30	0.030	0.022	0.023	0.028	0.017
1	1	100	5	0.021	0.020	0.023	0.023	0.031
1	1	100	15	0.030	0.021	0.026	0.025	0.025
1	3	25	5	0.022	0.024	0.029	0.032	0.025
1	3	25	15	0.027	0.024	0.026	0.024	0.026
1	3	25	30	0.022	0.033	0.026	0.025	0.019
1	3	100	5	0.018	0.020	0.020	0.027	0.022
1	3	100	15	0.029	0.019	0.025	0.024	0.022
5	1	25	5	0.024	0.028	0.028	0.031	0.032
5	1	25	15	0.025	0.025	0.029	0.029	0.021
5	1	25	30	0.024	0.025	0.029	0.030	0.025
5	1	100	5	0.019	0.020	0.027	0.025	0.032
5	1	100	15	0.022	0.021	0.022	0.022	0.024
5	3	25	5	0.023	0.021	0.030	0.023	0.022
5	3	25	15	0.024	0.027	0.028	0.024	0.021
5	3	25	30	0.024	0.025	0.024	0.017	0.024
5	3	100	5	0.017	0.022	0.025	0.025	0.025
5	3	100	15	0.029	0.029	0.021	0.029	0.025
10	1	25	5	0.026	0.022	0.030	0.026	0.025
10	1	25	15	0.030	0.025	0.022	0.025	0.028
10	1	25	30	0.029	0.024	0.026	0.029	0.027
10	1	100	5	0.021	0.029	0.029	0.037	0.032
10	1	100	15	0.025	0.022	0.026	0.026	0.021
10	3	25	5	0.025	0.025	0.022	0.030	0.023
10	3	25	15	0.027	0.025	0.028	0.026	0.021
10	3	25	30	0.024	0.031	0.021	0.024	0.022
10	3	100	5	0.030	0.021	0.025	0.029	0.028
10	3	100	15	0.029	0.025	0.028	0.029	0.021

Table E.13: Coverage for the upper limits of 95% confidence interval (C.U) of $\boldsymbol{\beta}$ for the GEE method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.068	0.065	0.070	0.068	0.063
1	1	25	15	0.036	0.044	0.035	0.039	0.045
1	1	25	30	0.029	0.039	0.031	0.037	0.038
1	1	100	5	0.047	0.052	0.052	0.052	0.072
1	1	100	15	0.036	0.035	0.037	0.032	0.038
1	3	25	5	0.071	0.062	0.058	0.061	0.061
1	3	25	15	0.033	0.035	0.040	0.034	0.038
1	3	25	30	0.028	0.025	0.034	0.029	0.029
1	3	100	5	0.051	0.049	0.052	0.050	0.061
1	3	100	15	0.040	0.034	0.035	0.032	0.034
5	1	25	5	0.073	0.058	0.065	0.066	0.073
5	1	25	15	0.038	0.042	0.036	0.044	0.047
5	1	25	30	0.037	0.035	0.033	0.034	0.033
5	1	100	5	0.049	0.060	0.042	0.055	0.059
5	1	100	15	0.038	0.033	0.036	0.032	0.036
5	3	25	5	0.053	0.050	0.057	0.056	0.053
5	3	25	15	0.032	0.038	0.038	0.038	0.041
5	3	25	30	0.028	0.027	0.030	0.027	0.034
5	3	100	5	0.043	0.034	0.046	0.040	0.042
5	3	100	15	0.028	0.030	0.032	0.034	0.036
10	1	25	5	0.066	0.085	0.073	0.081	0.076
10	1	25	15	0.039	0.040	0.042	0.051	0.045
10	1	25	30	0.033	0.040	0.039	0.036	0.038
10	1	100	5	0.051	0.051	0.058	0.061	0.060
10	1	100	15	0.039	0.031	0.033	0.035	0.039
10	3	25	5	0.050	0.051	0.044	0.057	0.058
10	3	25	15	0.038	0.036	0.041	0.032	0.045
10	3	25	30	0.029	0.031	0.032	0.030	0.029
10	3	100	5	0.045	0.042	0.036	0.048	0.046
10	3	100	15	0.032	0.033	0.031	0.031	0.024

Table E.14: Coverage for the upper limits of 95% confidence interval (C.U) of $\boldsymbol{\beta}$ for the ML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.035	0.031	0.037	0.046	0.039
1	1	25	15	0.028	0.030	0.032	0.028	0.029
1	1	25	30	0.021	0.021	0.018	0.023	0.021
1	1	100	5	0.028	0.036	0.027	0.036	0.036
1	1	100	15	0.022	0.025	0.021	0.020	0.034
1	3	25	5	0.027	0.030	0.028	0.033	0.035
1	3	25	15	0.022	0.031	0.031	0.025	0.034
1	3	25	30	0.025	0.027	0.024	0.028	0.028
1	3	100	5	0.026	0.037	0.030	0.027	0.029
1	3	100	15	0.022	0.033	0.023	0.033	0.021
5	1	25	5	0.032	0.038	0.030	0.032	0.035
5	1	25	15	0.022	0.025	0.022	0.025	0.028
5	1	25	30	0.021	0.024	0.021	0.021	0.025
5	1	100	5	0.028	0.025	0.032	0.033	0.026
5	1	100	15	0.020	0.029	0.025	0.025	0.032
5	3	25	5	0.037	0.027	0.029	0.034	0.029
5	3	25	15	0.025	0.022	0.029	0.035	0.028
5	3	25	30	0.028	0.025	0.021	0.025	0.025
5	3	100	5	0.037	0.029	0.033	0.029	0.035
5	3	100	15	0.022	0.025	0.021	0.025	0.023
10	1	25	5	0.025	0.029	0.032	0.030	0.033
10	1	25	15	0.023	0.021	0.026	0.029	0.026
10	1	25	30	0.025	0.025	0.025	0.021	0.032
10	1	100	5	0.028	0.029	0.029	0.024	0.030
10	1	100	15	0.025	0.017	0.028	0.021	0.025
10	3	25	5	0.025	0.028	0.034	0.033	0.035
10	3	25	15	0.021	0.025	0.022	0.032	0.033
10	3	25	30	0.027	0.024	0.029	0.027	0.026
10	3	100	5	0.024	0.036	0.025	0.033	0.022
10	3	100	15	0.021	0.025	0.022	0.024	0.026

Table E.15: Coverage for the upper limits of 95% confidence interval (C.U) of $\boldsymbol{\beta}$ for the REML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.025	0.023	0.026	0.032	0.022
1	1	25	15	0.027	0.029	0.030	0.026	0.029
1	1	25	30	0.023	0.021	0.018	0.023	0.020
1	1	100	5	0.023	0.025	0.021	0.027	0.019
1	1	100	15	0.022	0.022	0.020	0.018	0.028
1	3	25	5	0.022	0.024	0.022	0.025	0.025
1	3	25	15	0.019	0.028	0.028	0.022	0.028
1	3	25	30	0.025	0.025	0.024	0.027	0.025
1	3	100	5	0.023	0.029	0.025	0.019	0.017
1	3	100	15	0.021	0.032	0.021	0.029	0.018
5	1	25	5	0.030	0.031	0.028	0.028	0.031
5	1	25	15	0.024	0.024	0.024	0.025	0.026
5	1	25	30	0.022	0.025	0.021	0.023	0.025
5	1	100	5	0.030	0.021	0.027	0.029	0.021
5	1	100	15	0.021	0.032	0.027	0.026	0.029
5	3	25	5	0.031	0.019	0.024	0.025	0.022
5	3	25	15	0.023	0.022	0.026	0.032	0.026
5	3	25	30	0.027	0.026	0.020	0.025	0.025
5	3	100	5	0.034	0.026	0.028	0.021	0.026
5	3	100	15	0.021	0.024	0.019	0.024	0.021
10	1	25	5	0.024	0.025	0.025	0.028	0.027
10	1	25	15	0.025	0.022	0.025	0.029	0.028
10	1	25	30	0.026	0.026	0.026	0.022	0.032
10	1	100	5	0.027	0.028	0.025	0.022	0.028
10	1	100	15	0.029	0.020	0.028	0.021	0.028
10	3	25	5	0.025	0.021	0.030	0.025	0.028
10	3	25	15	0.021	0.025	0.021	0.031	0.029
10	3	25	30	0.026	0.024	0.028	0.027	0.025
10	3	100	5	0.024	0.029	0.023	0.026	0.017
10	3	100	15	0.021	0.025	0.022	0.021	0.024

Table E.16: Coverage for 95% confidence interval (C.I) of $\boldsymbol{\beta}$ for the GEE method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.142	0.127	0.137	0.143	0.137
1	1	25	15	0.075	0.088	0.075	0.081	0.082
1	1	25	30	0.062	0.070	0.066	0.075	0.071
1	1	100	5	0.102	0.104	0.098	0.109	0.110
1	1	100	15	0.074	0.074	0.075	0.066	0.071
1	3	25	5	0.125	0.111	0.112	0.115	0.134
1	3	25	15	0.066	0.070	0.069	0.075	0.080
1	3	25	30	0.057	0.054	0.066	0.057	0.065
1	3	100	5	0.107	0.099	0.093	0.104	0.119
1	3	100	15	0.076	0.071	0.072	0.060	0.068
5	1	25	5	0.153	0.124	0.138	0.128	0.141
5	1	25	15	0.090	0.094	0.087	0.085	0.098
5	1	25	30	0.071	0.062	0.064	0.066	0.070
5	1	100	5	0.110	0.108	0.094	0.107	0.111
5	1	100	15	0.074	0.071	0.073	0.067	0.075
5	3	25	5	0.115	0.105	0.108	0.106	0.110
5	3	25	15	0.065	0.070	0.071	0.078	0.077
5	3	25	30	0.058	0.060	0.058	0.057	0.063
5	3	100	5	0.094	0.078	0.094	0.079	0.076
5	3	100	15	0.059	0.060	0.068	0.066	0.071
10	1	25	5	0.127	0.150	0.142	0.156	0.137
10	1	25	15	0.084	0.089	0.083	0.092	0.089
10	1	25	30	0.068	0.077	0.077	0.073	0.067
10	1	100	5	0.101	0.105	0.106	0.116	0.114
10	1	100	15	0.071	0.060	0.063	0.067	0.073
10	3	25	5	0.106	0.103	0.105	0.113	0.114
10	3	25	15	0.078	0.069	0.079	0.063	0.081
10	3	25	30	0.056	0.061	0.067	0.056	0.060
10	3	100	5	0.101	0.094	0.077	0.094	0.084
10	3	100	15	0.065	0.067	0.066	0.069	0.056

Table E.17: Coverage for 95% confidence interval (C.I) for $\boldsymbol{\beta}$ for the ML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.075	0.065	0.072	0.077	0.071
1	1	25	15	0.055	0.055	0.060	0.061	0.059
1	1	25	30	0.051	0.043	0.041	0.051	0.039
1	1	100	5	0.065	0.065	0.064	0.067	0.067
1	1	100	15	0.055	0.047	0.050	0.045	0.058
1	3	25	5	0.056	0.066	0.061	0.074	0.069
1	3	25	15	0.052	0.058	0.059	0.051	0.062
1	3	25	30	0.050	0.061	0.052	0.054	0.048
1	3	100	5	0.057	0.071	0.061	0.061	0.057
1	3	100	15	0.057	0.053	0.050	0.062	0.044
5	1	25	5	0.063	0.072	0.064	0.069	0.070
5	1	25	15	0.049	0.052	0.051	0.054	0.047
5	1	25	30	0.044	0.049	0.049	0.051	0.048
5	1	100	5	0.053	0.054	0.061	0.060	0.057
5	1	100	15	0.042	0.049	0.048	0.046	0.054
5	3	25	5	0.070	0.055	0.070	0.062	0.058
5	3	25	15	0.050	0.052	0.057	0.058	0.051
5	3	25	30	0.052	0.051	0.044	0.041	0.049
5	3	100	5	0.063	0.060	0.062	0.060	0.061
5	3	100	15	0.054	0.055	0.045	0.055	0.048
10	1	25	5	0.057	0.058	0.068	0.059	0.060
10	1	25	15	0.053	0.046	0.049	0.054	0.053
10	1	25	30	0.054	0.048	0.050	0.050	0.057
10	1	100	5	0.053	0.064	0.062	0.060	0.062
10	1	100	15	0.051	0.038	0.053	0.047	0.044
10	3	25	5	0.062	0.058	0.060	0.067	0.064
10	3	25	15	0.051	0.053	0.052	0.059	0.054
10	3	25	30	0.051	0.055	0.050	0.051	0.048
10	3	100	5	0.064	0.062	0.054	0.065	0.054
10	3	100	15	0.051	0.052	0.051	0.052	0.048

Table E.18: Coverage for 95% confidence interval (C.I) for $\boldsymbol{\beta}$ for the REML method.

μ_h	k_s	H	n_h	β_1	β_2	β_3	β_4	β_5
1	1	25	5	0.052	0.047	0.049	0.056	0.046
1	1	25	15	0.049	0.050	0.057	0.058	0.059
1	1	25	30	0.053	0.043	0.041	0.051	0.037
1	1	100	5	0.043	0.045	0.043	0.050	0.051
1	1	100	15	0.052	0.044	0.046	0.042	0.053
1	3	25	5	0.045	0.048	0.051	0.057	0.051
1	3	25	15	0.046	0.051	0.053	0.046	0.054
1	3	25	30	0.048	0.058	0.051	0.052	0.045
1	3	100	5	0.041	0.049	0.044	0.046	0.039
1	3	100	15	0.050	0.051	0.046	0.053	0.040
5	1	25	5	0.053	0.058	0.056	0.059	0.062
5	1	25	15	0.048	0.050	0.053	0.054	0.047
5	1	25	30	0.046	0.050	0.051	0.053	0.050
5	1	100	5	0.050	0.041	0.054	0.054	0.053
5	1	100	15	0.043	0.053	0.050	0.048	0.053
5	3	25	5	0.054	0.040	0.054	0.049	0.044
5	3	25	15	0.047	0.049	0.054	0.056	0.047
5	3	25	30	0.051	0.051	0.043	0.042	0.048
5	3	100	5	0.051	0.049	0.053	0.045	0.051
5	3	100	15	0.050	0.053	0.040	0.053	0.046
10	1	25	5	0.050	0.047	0.056	0.054	0.052
10	1	25	15	0.055	0.047	0.048	0.054	0.055
10	1	25	30	0.056	0.051	0.052	0.051	0.058
10	1	100	5	0.049	0.056	0.054	0.059	0.059
10	1	100	15	0.054	0.042	0.054	0.048	0.048
10	3	25	5	0.050	0.045	0.052	0.055	0.051
10	3	25	15	0.048	0.050	0.049	0.057	0.050
10	3	25	30	0.050	0.054	0.048	0.051	0.048
10	3	100	5	0.054	0.051	0.048	0.056	0.045
10	3	100	15	0.050	0.050	0.050	0.051	0.045

Appendix F

Comparison of NB distribution and PdG model with single observation

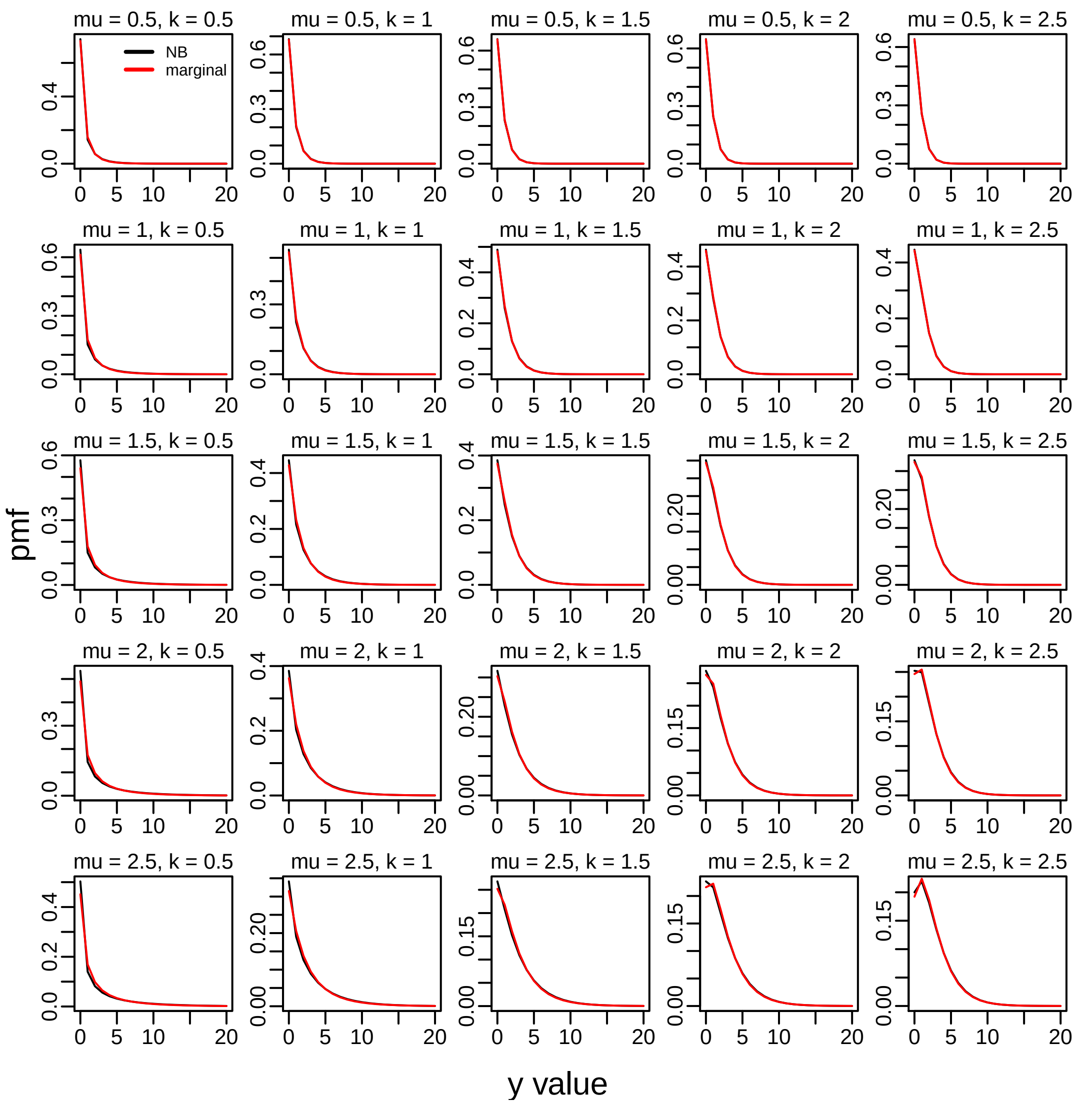

Figure F.1: Comparison of probability mass function in equation (6.2) with the Negative binomial mass function

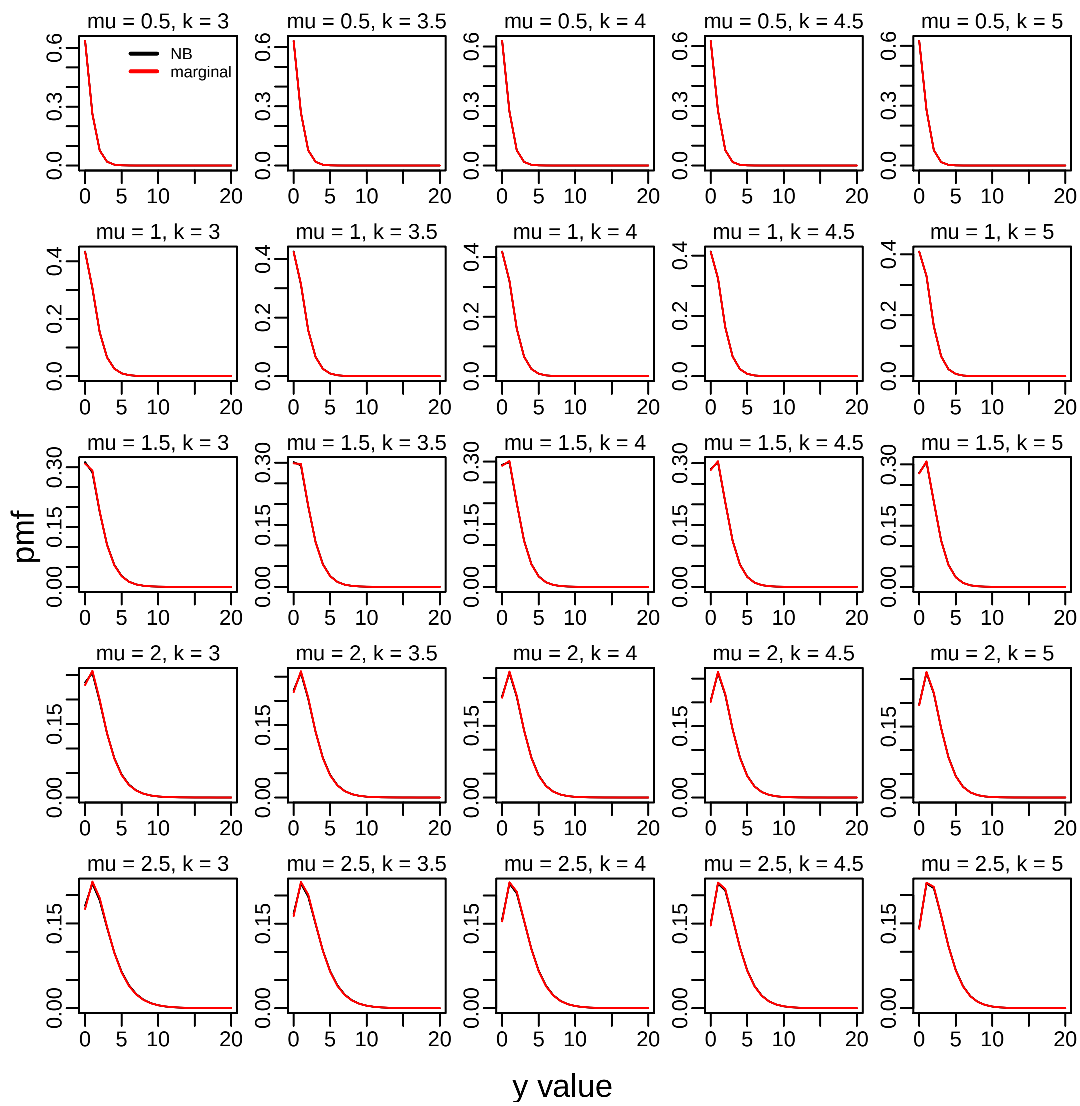

Figure F.2: Comparison of probability mass function in equation (6.2) with the Negative binomial mass function

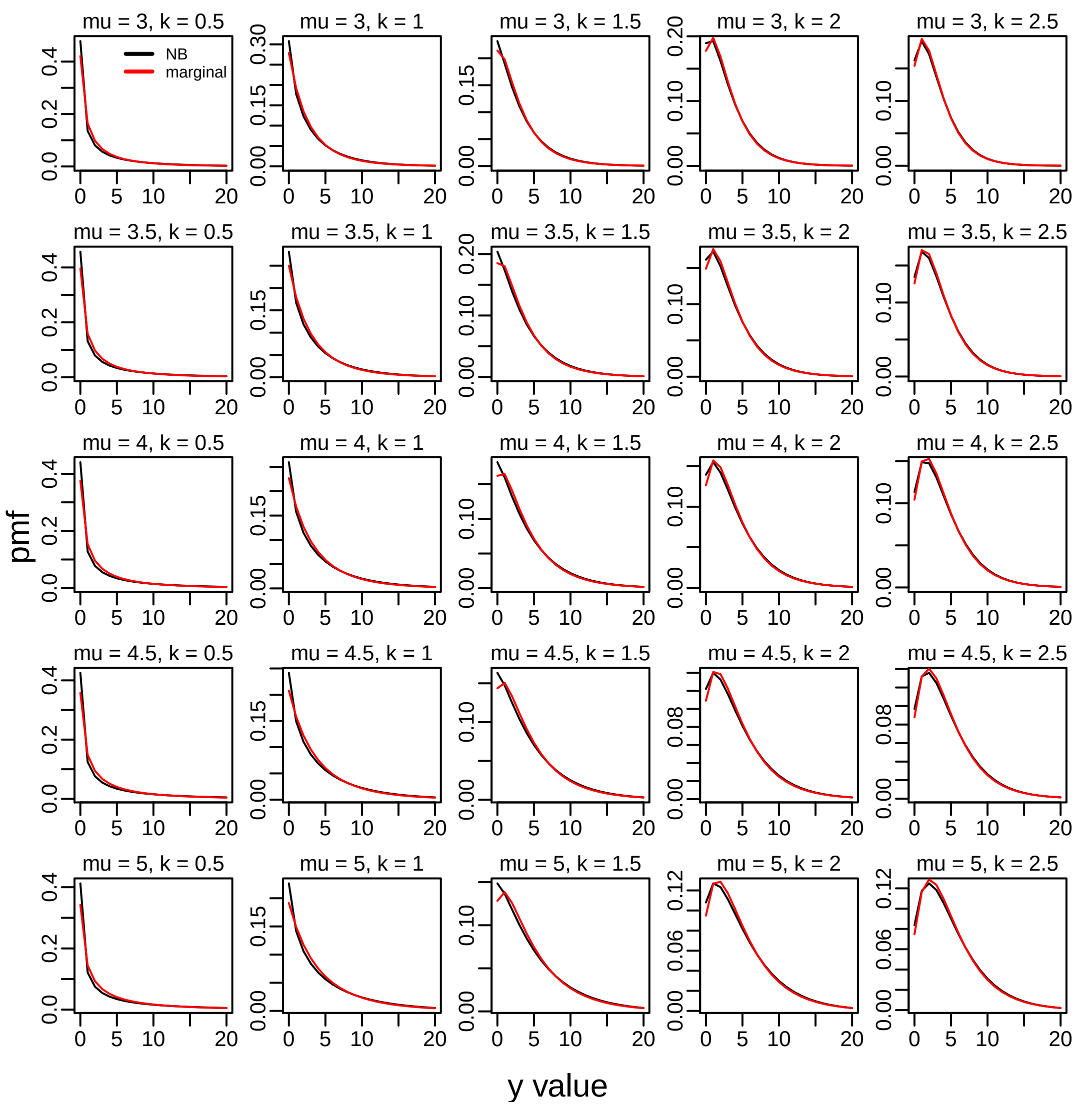

Figure F.3: Comparison of probability mass function in equation (6.2) with the Negative binomial mass function

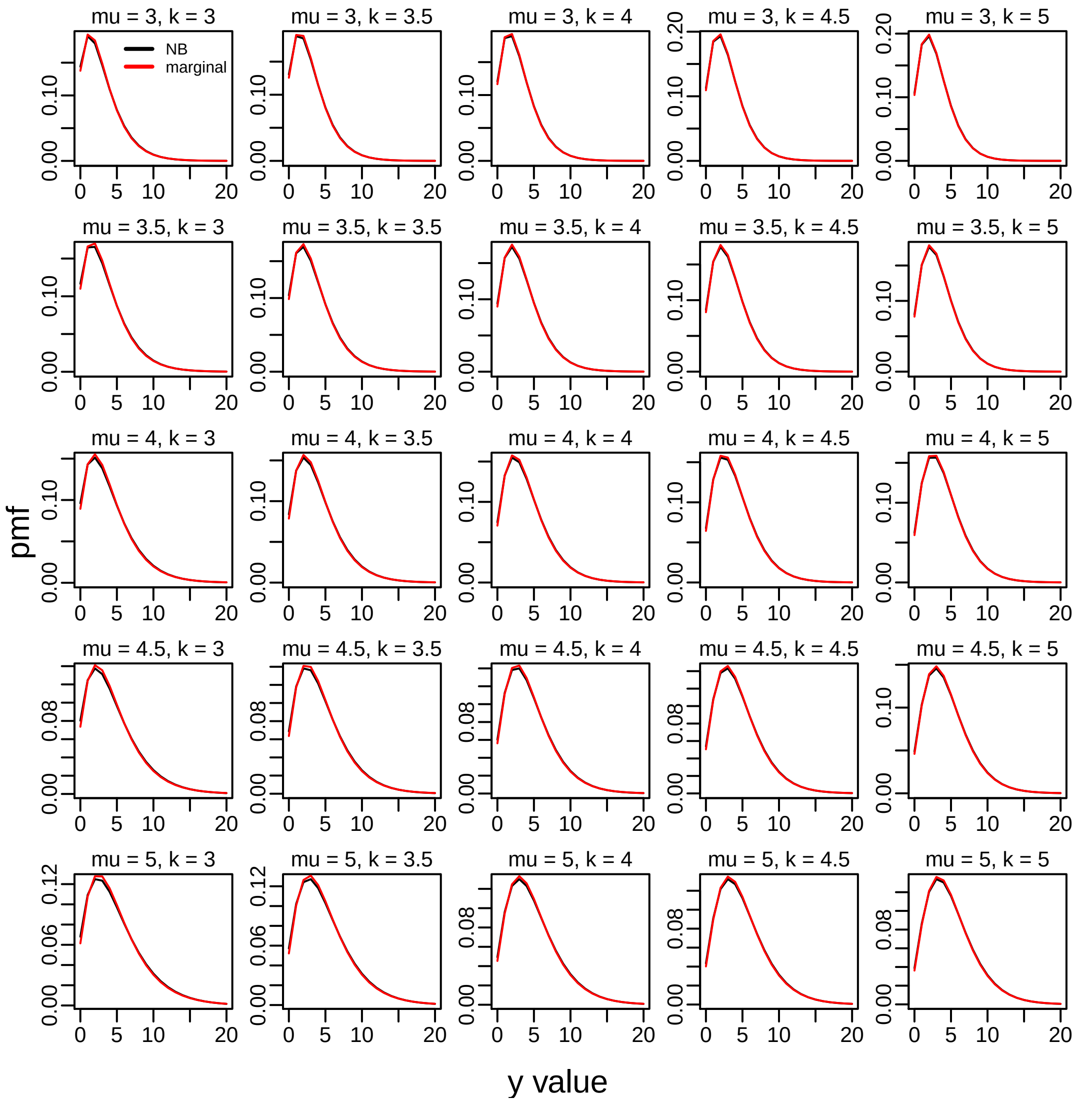

Figure F.4: Comparison of probability mass function in equation (6.2) with the Negative binomial mass function

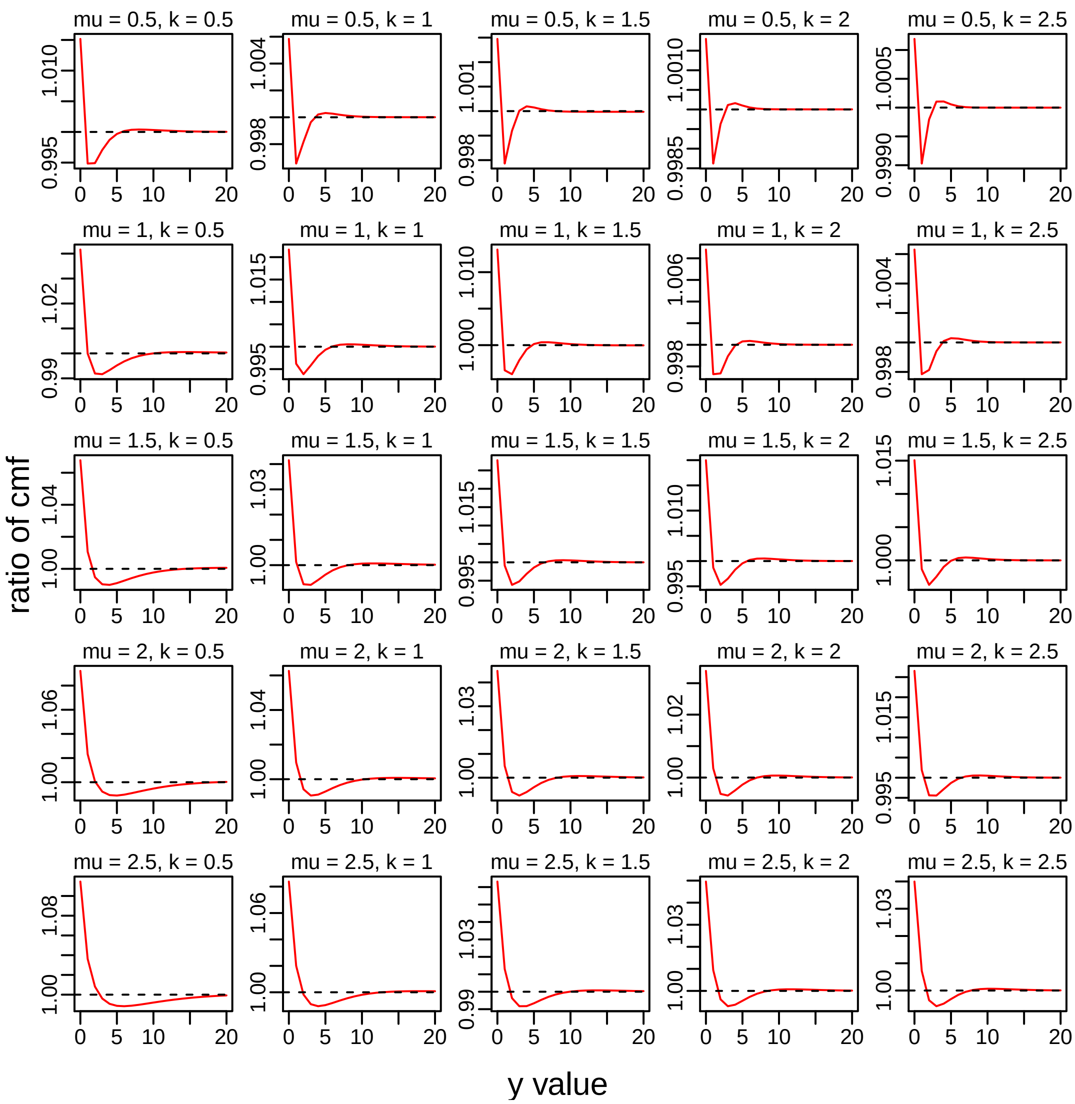

Figure F.5: Ratio of cumulative mass function for equation (6.2) and NB distribution

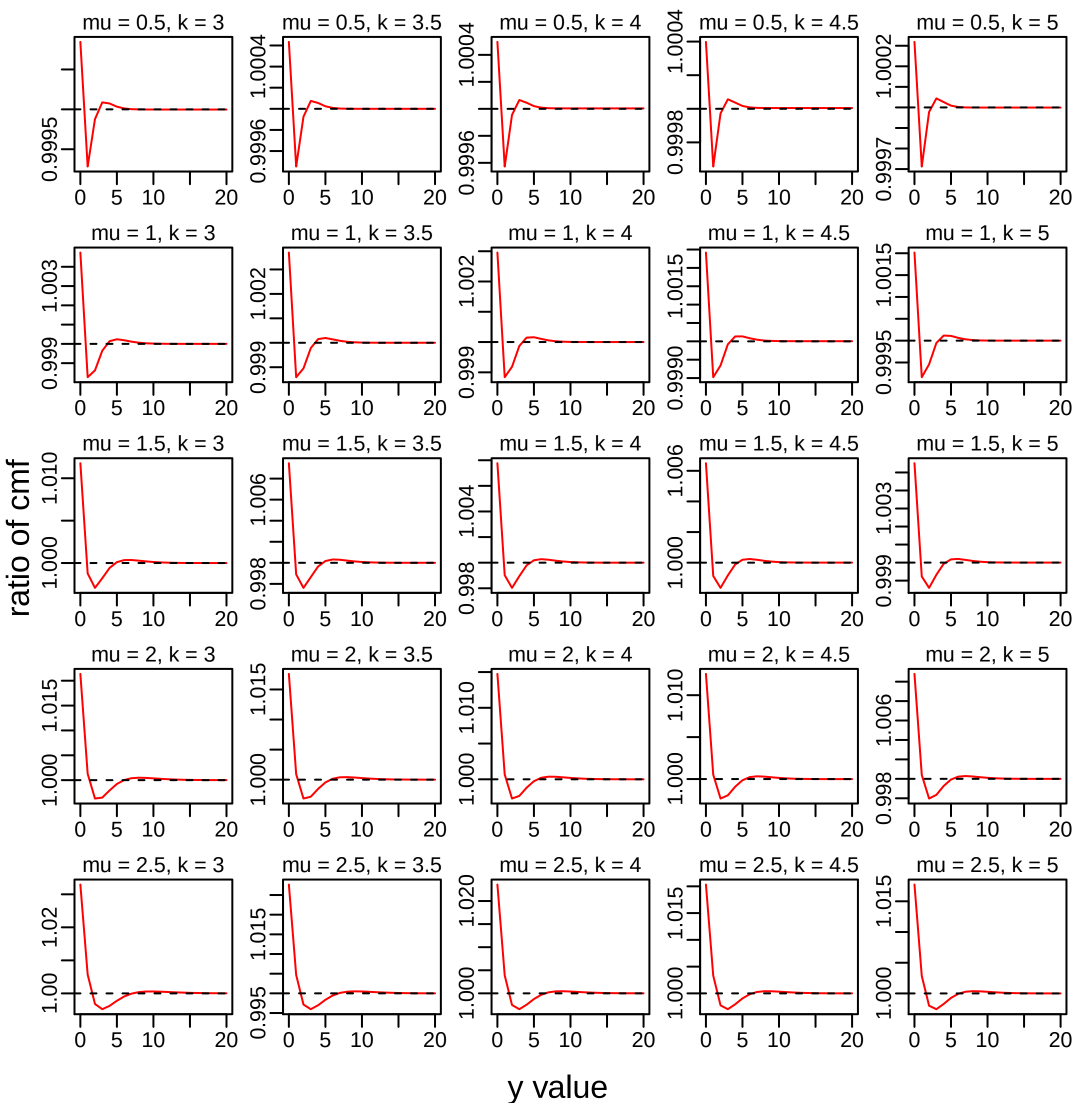

Figure F.6: Ratio of cumulative mass function for equation (6.2) and NB distribution

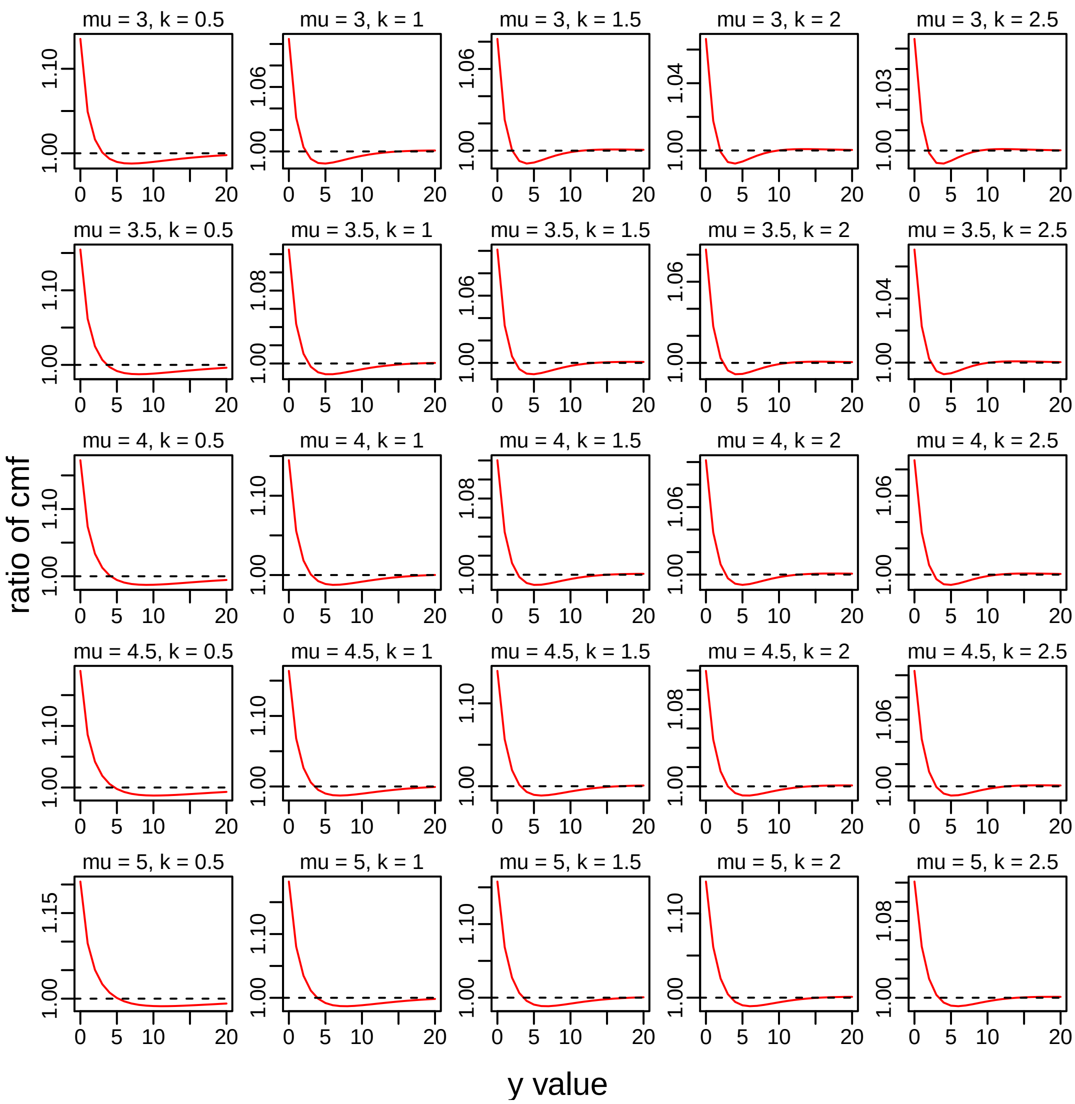

Figure F.7: Ratio of cumulative mass function for equation (6.2) and NB distribution

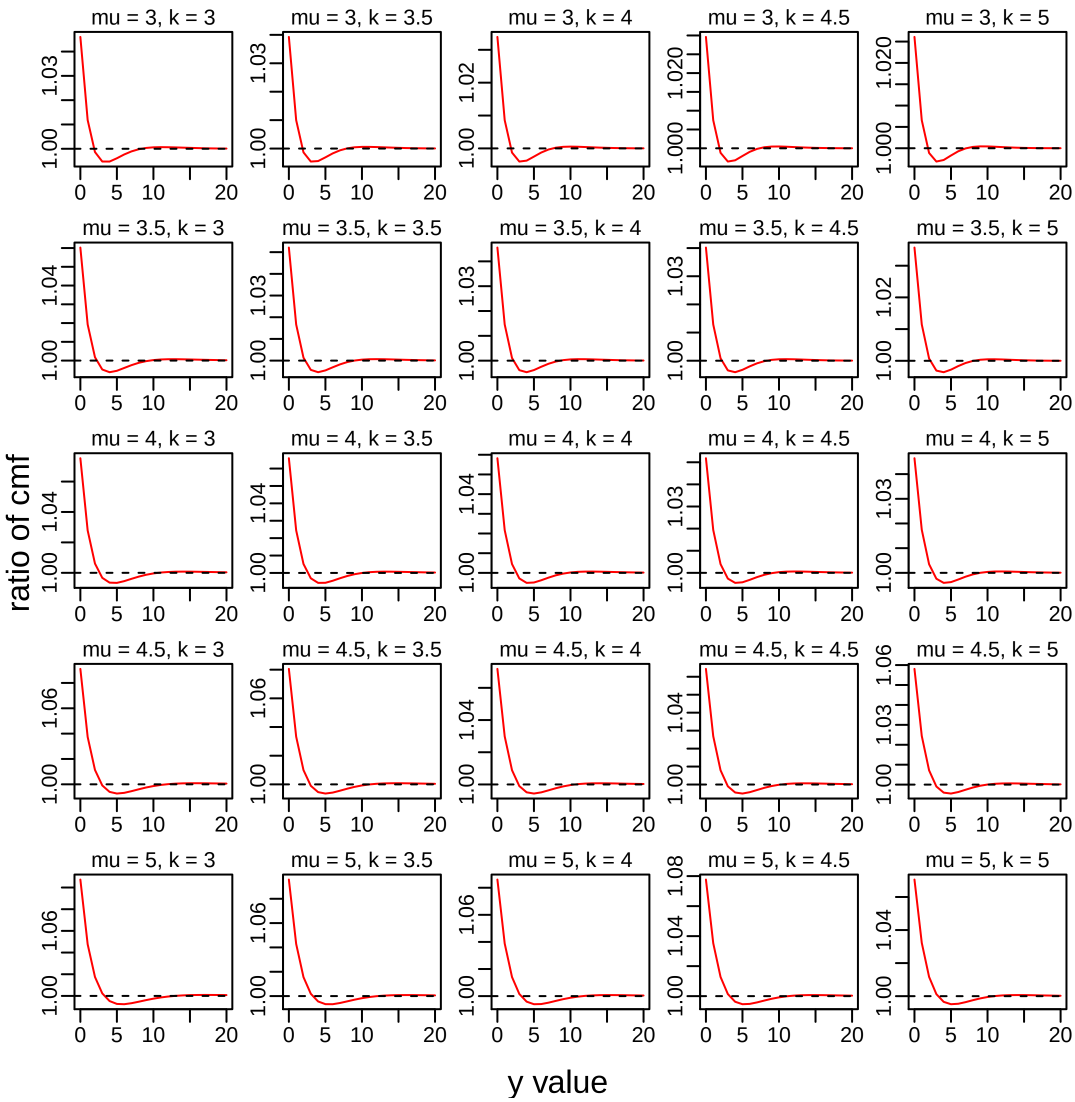

Figure F.8: Ratio of cumulative mass function for equation (6.2) and NB distribution

Bibliography

[1] Anderson, R. L., & Bancroft, T. A. (1952). Statistical theory in research.

[2] Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*, (London: Chapman and Hall).

[3] Baum, J. K. and Myers, R. A. (2004). Shifting baselines and the decline of pelagic sharks in the Gulf of Mexico.*Ecology Letters* **7,** 135–145.

[4] Bell, B. M. (2012). CppAD: a package for C++ algorithmic differentiation. *Computational Infrastructure for Operations Research.*

[5] Bellio, R. and Sartori, N. (2006). Practical use of modified maximum likelihoods for stratified data. *Biometrical Journal* **48,** 876–886.

[6] Benoît, H.P., & Cadigan, N. 2014. Model-based estimation of commercial-sized snow crab (*Chionoecetes opilio*) abundance in the southern Gulf of St. Lawrence, 1980-2013, using data from two bottom trawl surveys. *DFO Can. Sci. Advis. Sec. Res. Doc.* 2014/082. v + 24 p.

[7] Benoît, H. P., & Swain, D. P. (2003). Accounting for length-and depth-dependent diel variation in catchability of fish and invertebrates in an annual bottom-trawl survey. *ICES Journal of Marine Science: Journal du Conseil,* **60(6),** 1298-1317.

[8] Berger, J. O., Liseo, B., & Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science,* **14(1),** 1-28.

[9] Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88(421),** 9-25.

[10] Cadigan, N. G., & Bataineh, O. M. (2012). Inference about the ratio of means from Negative Binomial paired count data. *Environmental and Ecological Statistics,* **19(2),** 269-293.

[11] Cadigan, N. G., & Dowden, J. J. (2010). Statistical inference about the relative efficiency of a new survey protocol, based on paired-tow survey calibration data.

[12] Cadigan, N. G., & Tobin, J. (2010). Estimating the negative binomial dispersion parameter with highly stratified surveys. *Journal of Statistical Planning and Inference* **140,** 2138-2147.

[13] Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data (No. 53).* Cambridge university press.

[14] Chaganty, N. R. (1997). An alternative approach to the analysis of longitudinal data via generalized estimating equations. *Journal of Statistical Planning and Inference* **63(1),** 39-54.

[15] Casey, J. M., & Myers, R. A. (1998). Diel variation in trawl catchability: is it as clear as day and night?. *Canadian Journal of Fisheries and Aquatic Sciences,* **55(10),** 2329-2340.

[16] Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* **82(2),** 407-410.

[17] Doubleday, W.G. (1981). Manual on groundfish surveys in the Northwest Atlantic. *Northwest Atlantic Fisheries Organization Scientific Council Studies* **2,** 7-55.

[18] Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., ... & Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software* **27(2),** 233-249.

[19] Gunderson, D.R. (1993). *Surveys of fisheries resources.* New York: John Wiley.

[20] Hall, D. B., & Severini, T. A. (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association,* **93(444),** 1365-1375.

[21] Kristensen, K. (2013). TMB: General random effect model builder tool inspired by ADMB. *R package version 1.0.*

[22] Kimura, D. K., & Somerton, D. A. (2006). Review of statistical aspects of survey sampling for marine fisheries. *Reviews in Fisheries Science* **14,** 245-283.

[23] Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. Journal of the Royal Statistical Society. *Series B (Methodological),* 619-678.

[24] Lee, Y., & Nelder, J. A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* **88(4),** 987-1006.

[25] Lewy, P. and Kristensen, K. (2009). Modelling the distribution of fish accounting for spatial correlation and over-dispersion. *Canadian Journal of Fisheries and Aquatic Sciences* **66,** 1809-1820.

[26] Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73(1),** 13-22.

[27] Liang, K. Y., Zeger, S. L., & Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 3-40.

[28] Nelder, J. A., & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika,* **74(2),** 221-232.

[29] Nelder, J. A., and R. W. M. Wedderburn. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society Series B* **54,** 3-40.

[30] Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58(3),** 545-554.

[31] Paul, S., & Zhang, X. (2014). Small sample GEE estimation of regression parameters for longitudinal data. *Statistics in medicine* **33(22),** 3869-3881.

[32] Power, J. H. and Moser, E. B. (1999). Linear model analysis of net catch data using the negative binomial distribution. *Canadian Journal of Fisheries and Aquatic Sciences* **56,** 191-200.

[33] Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 1033-1048.

[34] R Core Team (2014). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria

[35] Robert, C., & Casella, G. (2013). *Monte Carlo statistical methods.* Springer Science & Business Media.

[36] Ross, G.J.S. and Preece, D.A. (1985). The negative binomial distribution. *The Statistician* **34,** 323-326.

[37] Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika* **90,** 533–549.

[38] Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78(4),** 719-727.

[39] Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance components* (Vol. 391). John Wiley & Sons.

[40] Severini, T. A. (1998). An approximation to the modified profile likelihood function. *Biometrika* **85,** 403–411.

[41] Skaug, H. J., & Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. *Computational Statistics & Data Analysis* **51(2),** 699-709.

[42] Stefánsson, G. (1996). Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES Journal of Marine Science: Journal du Conseil* **53,** 577-588.

[43] Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses. *Statistical Science,* **18(3),** 377-393.

[44] Thorson, J. T., Skaug, H. J., Kristensen, K., Shelton, A. O., Ward, E. J., Harms, J. H., & Benante, J. A. (2014). The importance of spatial models for estimating the strength of density dependence. *Ecology.*

[45] Wang, Y. G., & Carey, V. J. (2004). Unbiased estimating equations from working correlation models for irregularly timed repeated measures. *Journal of the American Statistical Association,* 99(467).

[46] Wang, M., & Long, Q. (2011). Modified robust variance estimator for generalized estimating equations with improved smallsample performance. *Statistics in medicine,* **30(11),** 1278-1291.

[47] Wang, Y. G., & Zhao, Y. (2007). A modified pseudolikelihood approach for analysis of longitudinal data. *Biometrics* **63(3),** 681-689.

[48] Wedderburn, R. W. M. (1974). Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika* **61,** 439-447.

[49] Zhang, X., & Paul, S. (2013). Variance function in regression analysis of longitudinal data using the generalized estimating equation approach. *Journal of Statistical Computation and Simulation* (ahead-of-print), 1-10.

[50] Zhao, L. P., & Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77(3),** 642-648.