A GENERALIZED ESTIMATING EQUATIONS APPROACH TO CAPTURE-RECAPTURE CLOSED POPULATION MODELS:

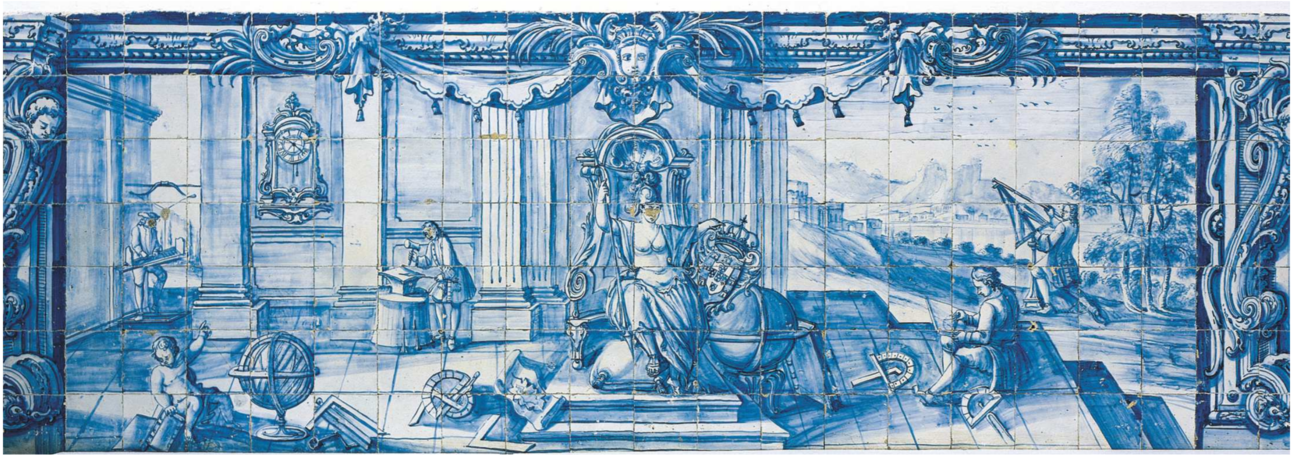
METHODS AND APPLICATIONS

Md. Abdus Salam Akanda

Thesis presented to the University of Évora
to obtain the Degree of Doctor in Mathematics
Area of Specialization: Statistics

SUPERVISOR: *Professor Russell Gerardo Alpizar Jara, PhD*

ÉVORA, JUNE, 2014



INSTITUTE FOR ADVANCED STUDIES AND RESEARCH

Abstract

Wildlife population parameters, such as capture or detection probabilities, and density or population size, can be estimated from capture-recapture data. These estimates are of particular interest to ecologists and biologists who rely on accurate inferences for management and conservation of the population of interest. However, there are many challenges to researchers for making accurate inferences on population parameters. For instance, capture-recapture data can be considered as binary longitudinal observations since repeated measurements are collected on the same individuals across successive points in times, and these observations are often correlated over time. If these correlations are not taken into account when estimating capture probabilities, then parameter estimates will be biased, possibly producing misleading results. Also, an estimator of population size is generally biased under the presence of heterogeneity in capture probabilities. The use of covariates (or auxiliary variables), when available, has been proposed as an alternative way to cope with the problem of heterogeneous capture probabilities. In this dissertation, we are interested in tackling these two main problems, (i) when capture probabilities are dependent among capture occasions in closed population capture-recapture models, and (ii) when capture probabilities are heterogeneous among individuals. Hence, the capture-recapture literature can be improved, if we could propose an approach to jointly account for these problems. In summary, this dissertation proposes: (i) a generalized estimating equations (GEE) approach to model possible effects in capture-recapture closed population studies due to correlation over time and individual heterogeneity; (ii) the corresponding estimating equations for each closed population capture-recapture model; (iii) a comprehensive analysis on various real capture-recapture data sets using classical, GEE and generalized linear mixed models (GLMM); (iv) an evaluation of the effect of accounting for correlation structures on capture-recapture model selection based on the 'Quasi-likelihood Information Criterion (QIC)'; (v) a comparison of the performance of population size estimators using GEE and GLMM approaches in the analysis of capture-recapture data. The performance of these approaches is evaluated by Monte Carlo (MC) simulation studies resembling real capture-recapture data. The proposed GEE approach provides a useful inference procedure for estimating population parameters, particularly when a large proportion of individuals are captured. For a low capture proportion, it is difficult to obtain reliable estimates for all approaches, but the GEE approach outperforms the other methods. Simulation results show that quasi-likelihood GEE provide lower standard error than partial likelihood based on generalized linear modelling (GLM) and GLMM approaches. The estimated population sizes vary on the nature of the existing correlation among capture occasions.

Keywords: Capture-recapture experiment; Correlation structure; Generalized estimating equations; Generalized linear mixed models; Heterogeneity; Population size estimation; Quasi-likelihood information criterion.

UMA ABORDAGEM DE EQUAÇÕES DE ESTIMAÇÃO GENERALIZADAS PARA MODELOS DE CAPTURA-RECAPTURA EM POPULAÇÃO FECHADAS: MÉTODOS E APLICAÇÕES

Resumo

Parâmetros populacionais em espécies de vida selvagens, como probabilidades de captura ou detecção, e abundância ou densidade da população, podem ser estimados a partir de dados de captura-recaptura. Estas estimativas são de particular interesse para ecologistas e biólogos que dependem de inferências precisas para a gestão e conservação das populações. No entanto, há muitos desafios para os investigadores fazer inferências precisas de parâmetros populacionais. Por exemplo, os dados de captura-recaptura podem ser considerados como observações longitudinais binárias uma vez que são medições repetidas coletadas nos mesmos indivíduos em pontos sucessivos no tempo, e muitas vezes correlacionadas. Se essas correlações não são levadas em conta ao estimar as probabilidades de captura, as estimativas dos parâmetros serão tendenciosas e possivelmente produzirão resultados enganosos. Também, um estimador do tamanho de uma população é geralmente enviesado na presença de heterogeneidade das probabilidades de captura. A utilização de co-variáveis (ou variáveis auxiliares), quando disponíveis, tem sido proposta como uma forma de lidar com o problema de probabilidades de captura heterogéneas. Nesta dissertação, estamos interessados em abordar dois problemas principais em mode1os de captura-recapturar para população fechadas, (i) quando as probabilidades de captura são dependentes entre ocasiões de captura, e (ii) quando as probabilidades de captura são heterogéneas entre os indivíduos. Assim, a literatura de captura-recaptura pode ser melhorada, se pudéssemos propor uma abordagem conjunta para estes problemas. Em resumo, nesta dissertação propõe-se: (i) uma abordagem de estimação de equações generalizadas (GEE) para modelar possíveis efeitos de correlação temporal e heterogeneidade individual nas probabilidades de captura; (ii) as correspondentes equações de estimação generalizadas para cada modelo de captura-recaptura em população fechadas; (iii) uma análise sobre vários conjuntos de dodas reais de captura-recaptura usando a abordagem clássica, GEE e modelos linear generalizados mistos (GLMM); (iv) uma avaliação do efeito das estruturas de correlação na selecção de modelos de captura-recaptura com base no 'critério de informação da Quasi-verossimilhança (QIC)'; (v) uma comparação da performance das estimativas do tamanho da população usando GEE e GLMM. O desempenho destas abordagens é avaliado usando simulações Monte Carlo (MC) que se assemelham a dados de captura-recapture reais. A abordagem GEE proposto é um procedimento de inferência útil para estimar parâmetros populacionais, especialmente quando uma grande proporção de indivíduos é capturada. Para uma proporção baixa de capturas, é difícil obter estimativas fiáveis para todas as abordagens aplicadas, mas GEE supera os outros métodos. Os resultados das simulações mostram que o método da quase-verossimilhança do GEE fornece estimativas do erro padrão menor do que o método da verossimilhana parcial dos modelos lineares generalizados (GLM) e GLMM. As estimstivas do tamanho da população variam de acordo com a natureza da correlação existente entre as ocasiões de captura.

Palavras-chave: Captura-recaptura experiência; Estrutura de correlação; Equações de estimação generalizadas; Modelos lineares generalizados mistos; Heterogeneidade; Estimativa de tamanho da população; Critério de informação quasi- verossimilhança.

Dedicated to

Shadiq Muntaqim Akanda (Maahin)

Maksuda Khanam (Shiuli)

Wasif Tahsin Akanda

Acknowledgement

First and foremost I would like to express my heartfelt gratitude to my supervisor Professor Russell Gerardo Alpizar Jara who gave me the opportunity to work with him. His insights and knowledge have been extremely valuable. He motivated me how to write a paper step by step. His patience, guidance, active support, continuous encouragement, amiability and generosity turned a difficult task into an immersing and enjoyable experience. During these years I have learned so much academic and non-academic things in a hospitable style. His door was always open for me right from the beginning. He is not only my supervisor, but also one of my closest best friends. It was my privilege to work with him. Thanks a lot for all the discussions and for the exhaustive text revisions indeed. I am profoundly grateful to him.

I express my sincere gratitude to the European Union for awarding me the Erasmus Mundus Mobility with Asia-West (EMMA-WEST) scholarship. I am also very grateful to my employer University of Dhaka, Bangladesh who facilitated my enrolment in this program. I feel that I have greatly enhanced my skills and broadened my outlook. The investment was worthwhile and I eagerly look to apply back my skills and contribute to my country. I am thankful to the Research Centre of Mathematics and Application (CIMA-UE) at the University of Évora and the Fundação Nacional para a Ciência e Tecnologia (FCT), Portugal for providing some financial support to participate in conferences under the project PEst-OE/MAT/UI0117/2011. I am also greatfully acknowledge the comments and suggestions made by jury members.

I am utmost grateful to EMMA-WEST coordinators Professor Md. Shafiqul Islam of University of Dhaka, Professor Imme van den Berg and Professor José Carlos Tiago de Oliveira of University of Évora for their cordial support during my study period. Much appreciation is also due to the Rector at the University of Évora for nurturing a wonderful environment that made life here such a stimulating and enjoyable experience. My thanks goes to everybody in the Mobility and International Relations Office, Department of Mathematics, Academic Services and Social Services at the University of Évora for their active support.

Much of the credit of the dissertation goes to my beloved sons Shadiq Muntaqim Akanda (Maahin) & Wasif Tahsin Akanda and my heartiest wife Maksuda Khanam (Shiuli). They supported me in innumerable ways and continues to inspire and encourage me along in life. I would not be here without their love and support. My faith makes all things possible, but their love and support make all things easy. The list of sacrifices that they has made during this research would not fit on the pages of this dissertation. I can not put into words the gratitude I have toward them.

Last but not least, I would like to say what I believe,

“A new idea is worthless until it’s practical approach succeeds.”

Md. Abdus Salam Akanda

Contents

List of Tables

List of Figures

Abbreviations and Symbols

The following abbreviations will be used in this work:

CR	Capture-recapture
ML	Maximum likelihood
GLM	Generalized linear models
GEE	Generalized estimating equations
GLMM	Generalized linear mixed models
i.i.d.	Independent and identically distribution
OR	Odds ratio
ME	Marginal effect
MC	Monte Carlo
AVE	Average
SE	Standard error
PRB	Percentage relative bias
CV	Coefficient of variation
CI	Confidence interval
COV	Coverage of confidence intervals
RMSE	Root mean square error
AIC	Akaike's information criterion
QIC	Quasi-likelihood information criterion
GCV	Generalized cross validation

The following symbols will be used in this work:

i	Index referring to individuals
j	Index referring to occasion
Y_{ij}	Indicator variable of the i^{th} individual being caught on the j^{th} occasion
Y_i	Vector of capture history of the i^{th} individual
T_i	Number of times that the i^{th} individual has been caught in a experiment
t_i	Time of the i^{th} individual is first captured
N	Population size
m	Number of capture occasions
n_j	Number of individuals captured on the j^{th} occasions
u_j	Number of unmarked individuals captured on the j^{th} occasions
v_j	Number of individuals captured for the last time on occasion j
n	Number of distinct individuals captured in m occasions
f_k	Number of individuals captured exactly k times in m occasions
m_j	Total number of marked individuals caught on capture occasion j
M_j	Number of marked individuals in the population at the time of capture occasion j
P_c	Probability of capture
P_r	Probability of recapture
$\bar{p}$	Mean capture probability
π_i	Probability of individual i capture at least once
α	Population correlation coefficient
r	Sample correlation coefficient
R_i	Correlation structure of individual i
P_{ij}	Capture probability of the i^{th} individual on the j^{th} occasion
M_0	Constant probability model

Continued on next page

M_{t}	Temporal variation model
M_{b}	Behavioral response model
M_{h}	Individual heterogeneity model
M_{tb}	Temporal variation and behavioral response model
M_{th}	Temporal variation and individual heterogeneity model
M_{bh}	Behavioral response and individual heterogeneity model
$\mathrm{M}_{\mathrm{tbh}}$	Temporal variation, behavioral response and individual heterogeneity model
g(.)	Link function
L(.)	Likelihood function
LL(.)	Log-likelihood function
QL(.)	Quasi-likelihood function
PL(.)	Partial likelihood function
β	Vector of regression parameters
β_{sex}	Regression coefficient for the effect of sex
β_{age}	Regression coefficient for the effect age
β_{wt}	Regression coefficient for the effect weight
β_t	Regression coefficient for the effect time
p	Number of parameters in model
$\hat{N}$	Population size estimator

Chapter 1

Introduction

This chapter introduces some important concepts (time dependence and heterogeneity) that are useful for the application of generalized estimating equations and mixed effects models to capture-recapture studies. Some capture-recapture data sets published in the literature are also presented and used for illustrative purposes. Finally, research goals and an overview of the following chapters is also given.

1.1 Clustering: A Source of Dependence

Often in applied statistics, after collecting some empirical data given a study design, the purpose of the analysis is to construct a statistical model to estimate population parameters. That means, we are interested in situations where the primary aim is to explain how a response variable of particular interest is related to a set of explanatory covariates. Classically, a single observation on the response variable is obtained for each observational unit and one of the fundamental hypotheses of standard statistical modelling in this case is independence among observations. Many types of studies, however, have designs which imply gathering data in response groups or clusters. Familiar examples of clusters are animal litters, families or schools. In each of these examples, a cluster is a collection of subunits on which observations are made. Another usual form of clustering arises when data are measured repeatedly on the same unit. In both cases the elements of a cluster share some common characteristics. Therefore, observations within a cluster tend to be more correlated than observations from different clusters. Thus, unlike in the classical setting where there exists a single source of variation between observational units, the heterogeneity among clusters introduces an additional source of variation and complicates the analysis. When this variation cannot be explained by measured covariates, we require advanced statistical methods for analysis which account explicitly for the clustering in the data.

1.2 Time Dependence: A Challenge in Capture-Recapture Studies

Capture-recapture data are collected on the same individuals across successive points in time; that is, the capture events are measured on the same individual at several time points. Thus, the capture history of each individual can be considered as a cluster of binary longitudinal observations that tend to be correlated. Hwang and Huggins (2007) also state that the assumption of independence among capture occasions is often violated in practice, but the authors still rely on that assumption. Observations within the same individual are generally positively correlated for any type of clustering (Fitzmaurice, 1995). If these correlations are not taken into account when estimating capture probabilities, then the parameter estimates will be biased, possibly producing misleading inferences. Assuming independence of correlated repeated observations within an individual can lead to a considerable loss of efficiency in estimating the regression parameters (Fitzmaurice, 1995). Hence one should drop the standard assumption of independence among observations within the same individual, and model the correlation structure explicitly.

1.3 Effect of Ignoring Time Dependence

A very common question that arise is ‘what are the impacts of ignoring time dependence when we use the conventional regression method assuming independence among capture occasions?’ From a statistical point of view, there are at least two consequences as a result of ignoring time dependence: incorrect inferences about regression coefficients and inefficient estimation of regression coefficients (Diggle et al., 2002). Now we discuss these two issues in detail through a simple model for correlated data.

Let Y_{ij} be an indicator variable of the i^{th} individual (cluster) on the j^{th} capture occasion , $i = 1, 2, ..., n, j = 1, 2, ..., m$. Here n is the number of cluster (individual) and m the cluster size (number of capture occasions). We assume that for each i,

$$\begin{aligned} &E(Y_{ij}) = \beta_0 + \beta_1 x_{ij}; \mathrm{Var}(Y_{ij}) = \sigma^2; \\ &\mathrm{Cov}(Y_{ij}, Y_{ik}) = \sigma^2\alpha; j \neq k = 1, 2, ..., m. \end{aligned} \tag{1.1}$$

The model also assume that the expected value of Y is a simple linear function of a covariate, x, and that the correlation, as a measure by correlation coefficient, between each pair of responses from the same cluster has the same value, α (say). Our interest is on the inference for β_1 when ignoring the correlation among capture occasions. The ordinary least square estimator of β_1 is given by

$$\hat{\beta}_1 = \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \bar{y}_i)(x_{ij} - \bar{x}_i) \Big/ \sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \bar{x}_i)^2,$$

with $\bar{y}_i = \sum\limits_{j=1}^{m} y_{ij}/m$ and $\bar{x}_i = \sum\limits_{j=1}^{m} x_{ij}/m$ be the sample means of the Y_i’s and x_i’s from the i^{th} individual.

1.3.1 Incorrect Variance

According to Diggle et al. (2002) and Liang (1999), the variance, V_1 of least square estimate $\hat{\beta}_1$, ignoring the time association is given by

$$V_1 = \sigma^2 \Big/ \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - \bar{x}_i)^2 = \sigma^2 / V_T,$$

where $\bar{x} = \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} x_{ij} / (mn)$, $\sigma^2 = \text{Var}(Y_{ij})$, and $V_T = \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} (x_{ij} - \bar{x}_i)^2$. The correct variance, V_2 of $\hat{\beta}_1$ considering time association has the following form

$$V_2 = V_1 \big\{ 1 + \alpha(m\phi - 1) \big\},$$

where $\phi = n \sum\limits_{i=1}^{n} (\bar{x}_i - \bar{x})^2 / V_T$ is the fraction of the total variation in the x's explained by the between cluster variation in $\bar{x}_i$'s and $\alpha = \text{Corr}(Y_{ij}, Y_{ik}), j \neq k = 1, 2, ..., m$. Two important cases of ϕ deserve special attention. One is when $\phi = 0$, i.e., $\bar{x}_1 = ... = \bar{x}_n$. This occurs commonly in longitudinal studies in which every individual's response is measured at the same set of times. In this case, β_1 is estimated mainly by using the time changes in Y. The other extreme cases is when $\phi = 1$, i.e., $x_{i1} = x_{i2} = ... = x_{im}$ for $i = 1, 2, ..., n$. In either case, invalid scientific conclusions, which could be either false positive or negative, may be drawn if V_1 is used as the variance estimate of $\hat{\beta}_1$ (see for details Diggle et al. 2002).

1.3.2 Loss of Efficiency

When the ordinary least square estimate $\hat{\beta}_1$ is an unbiased estimator of β_1, the well known Gauss-Markov theorem suggests that the uncertainty in estimating β_1 may be reduced by using the weighted least square estimator, $\tilde{\beta}_1$ (say), which properly accounts for the time association (Liang, 1999). Under the specified correlation structure in (1.1), Diggle et al. (2002) and Liang (1999) showed that $\tilde{\beta}_1$ has the variance of the following form

$$V_3 = V_1(1-\alpha)\big\{1+(m-1)\alpha\big\}\big/\big\{1-\alpha+m\alpha(1-\phi)\big\}.$$

Interestingly, no efficiency loss occurs when ϕ is equal to 0 or 1, irrespective of α and m. On the other hand, a great deal of efficiency loss by using $\hat{\beta}_1$ may result when ϕ approaches 0.5. Such phenomenon is more apparent with increased α but less so for m. The main message is that ignoring correlation could lead to a substantial loss of statistical power to estimate β_1, and that it is important to examine and incorporate the time association structure such as to help improve upon the efficiency for the β inference (see for details Diggle et al. 2002).

1.4 Measures of Time Dependence

The pattern of time dependence is very important in several regards:

(i) it may help to characterize the relationship between the capture probabilities and covariates, and

(ii) it is also important to estimate the population parameters accurately in capture-recapture studies.

Following Liang (1999), we briefly discuss how one measures the time association and its patterns. The correlation coefficient as a measure of association is less useful for binary responses because the range of α is narrowed considerably due to the constraint that

$$\max(0, \gamma_j + \gamma_k - 1) < \Pr(Y_j = Y_k = 1) < \min(\gamma_j, \gamma_k)$$

where $\gamma_j = \Pr(Y_j = 1)$ and $\gamma_k = \Pr(Y_k = 1)$ and the degree of constraint depends on values of the γ's.

As an alternative, one may consider the use of the odds ratio (OR)

$$\mathrm{OR}(Y_j, Y_k) = \frac{\Pr(Y_j = Y_k = 1)\Pr(Y_j = Y_k = 0)}{\Pr(Y_j = 1, Y_k = 0)\Pr(Y_j = 0, Y_k = 1)},$$

as a measure of association between a pair of responses, Y_j and Y_k. For this quantity, no constraint is induced other than the fact that it must be positive. A positive association results if OR is greater than one and a negative association may be claimed if OR is less than one. For a more detailed discussion on the use of odds ratios for within-cluster (time) associations, see Heagerty and Zeger (1998) for longitudinal data. At last, different measures of association between discrete covariates have been suggested in the literature; see Goodman and Kruskal (1979).

For continuous responses, the most commonly used measure of dependence between a pair of capture occasions from the same individual is the correlation coefficient, which is defined as

$$\alpha(Y_j, Y_k) = \frac{\mathrm{Cov}(Y_j, Y_k)}{\left\{\mathrm{Var}(Y_j)\mathrm{Var}(Y_k)\right\}^{1/2}}; j \neq k = 1, 2, ..., m.$$

This quantity (α) takes values between -1 and 1 inclusively. There exist very weak dependence between Y_j and Y_k if α is close to zero, and a strong association between Y_j and Y_k, if α close to 1 or -1. Furthermore, a positive association between Y_j and Y_k, i.e., $\alpha > 0$, means Y_j tends to be larger than expected if Y_k is and vice versa.

For longitudinal studies, it is a common belief that the correlation coefficient between two observations adjacent in times is likely to be larger than that when two observations are far apart in times. To capture this phenomenon, one may model

$$\alpha(Y_j, Y_k) = \alpha^{|t_j - t_k|},$$

where t_j and t_k are times at which Y_j and Y_k are observed, respectively. This pattern is known as the AR-1 (auto-regressive model of order one) model. For more detailed discussion on the patterns of within-individuals (time) association for longitudinal data, we refer the readers to the work by Diggle (1988), Diggle, Heagerty, Liang and Zeger (2002) and the references therein.

1.5 Estimation Methods for Correlated Data

An important consideration in the statistical modelling of correlated data concerns the type of response. Methods for continuous (read ‘normally distributed’) data are undeniably the best developed and the linear mixed models (Laird and Ware, 1982; Verbeke and Molenberghs, 1997, 2000) have played a prominent role in extending the general linear model to handle correlated continuous data. Owing to the elegant properties of the multivariate normal distribution, its theory and implementation are greatly simplified.

When the response variable is discrete (e.g. counts) or categorical (nominal or ordinal data), a first issue arises which is the lack of a discrete analogue to the multivariate normal distribution. Complete specification of the joint distribution of the response vector becomes more problematic and fully likelihood-based methods are generally awkward.

Another issue arises by this type of responses is that the researcher must distinguish among three broad model families: marginal models, random effects models or conditional models. For simplicity, let us concentrate on the special case of correlated binary responses. A marginal model is one in which marginal probabilities of response are directly modelled. There is an extensive statistical literature on marginal modelling of correlated binary responses. For example, Bahadur (1961) and Zhao and Prentice (1990), describe maximum likelihood estimation (MLE) where marginal correlations are used to account for the association among responses. Alternatively, the within-cluster (time) association can be parameterized in terms of marginal odds ratios, as shown by Dale (1986), Liang, Zeger and Qaqish (1992), Lang and Agresti (1994), Molenberghs and Lesaffre (1994), and Glonek and McCullagh (1995) for instance.

Since few joint probability models for multivariate categorical data permit tractable modelling of marginal probabilities, alternative methods have been in demand. Thus, Liang and Zeger (1986) and Zeger and Liang (1986) proposed the so-called generalized estimating equations (GEE), which do not require assumptions about the complete joint distribution of the response vector. Their approach relies on estimating functions and provides a natural extension of quasi-likelihood (Wedderburn, 1974) to the multivariate response setting. Standard generalized estimating equations require only correct specification of the univariate marginal probabilities while adopting some working assumptions about the association structure.

Drawing on direct analogies with linear models for continuous responses, another way to model the joint distribution of the response vector is to postulate the existence of unobserved latent variables, often called random effects. These can be thought of as representing various features shared by the subunits of a cluster and hence introduce correlation among observations. Such cluster-specific effects are usually assumed to be independent and identically distributed according to a certain mixing distribution. An additional assumption that is frequently used is that the observations within a cluster are conditionally independent given the random effects. When the mixing distribution is assumed Gaussian, the families of linear mixed models and generalized linear models (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989) can be combined to form the class of generalized linear mixed models (GLMM). These models have been studied, among others, by Stiratelli, Laird and Ware (1984), Anderson and Aitkin (1985), Gilmour, Anderson and Rae (1985), Zeger, Liang and Albert (1988), Breslow and Clayton (1993), Wolfinger and O'Connell (1993) and Goldstein and Rasbach (1996). Some authors have advocated different distributional assumptions for the mixing distribution (e.g. Lee and Nelder, 1996). Also, cluster-specific approaches are not limited to mixed models, as demonstrated by the popular beta-binomial model (Williams, 1975).

There are some important distinctions between the two model families described so far. In marginal models, parameters may be interpreted with respect to the marginal or population-averaged distribution; therefore, such models are often referred to as population-averaged models. In random-effects models, on the other hand, parameters have cluster-specific effects and these models are consequently also called cluster-specific. This distinction is, in effect, irrelevant for normal responses since parameters have both population-averaged and cluster-specific interpretations in this case, but it becomes critical with categorical data. Zeger, Liang and Albert (1988), for example, discuss these two approaches to modelling of longitudinal data using GEE to estimate model parameters.

The third class of models that is commonly used to model correlated data is that of (response) conditional models. In a conditional model the parameters describe a feature (e.g. response probability) of a set of responses conditionally on the other responses. Due to the popularity of marginal and random-effects models, these have received relatively little attention, especially in the case of clustered data. An example in the specific context of clustered binary responses was given by Molenberghs and Ryan (1999). Conditional models have been much more popular in the context of longitudinal data, where they have been termed transition models (Diggle, Heagerty, Liang and Zeger, 2002). The (response) conditional approach, however, is usually criticized because of the conditional interpretation of the parameters on other responses, and on cluster sizes.

The debate continues about the relative merits of the different approaches. For several years, it seemed that marginal models, particularly GEE, were the most popular. More recently, there has been a renewed interest in random-effects models. An advantage of using GLMM for the estimation of capture probabilities is to accommodate not only the heterogeneity due to measurable individual characteristics, but also the uncertainty that the measurable individual characteristics cannot explain. A comparative advantage of GEE over random-effects models relates to the ability of GEE to allow specific correlation structures to be assumed within capture occasions. However, there are merits and disadvantages to all three model families. Recently, Bayesian inference is also becoming population for modelling correlated binary data, but our focus is based on frequentist approaches. Arguably, model choice will depend not only on the application of interest, but also on the specific analysis goals.

1.6 Individual Heterogeneity: The Bane of Abundance Estimation

Individual heterogeneity refers to the variation among individual animals in their probability of capture. Some capture-recapture models assume that capture probability is constant across individuals within a group. When individuals vary in their capture probabilities, the most catchable animals are likely to be caught first and more often. Behavioral differences, age, sex, and other innate differences can make an individual more or less likely to be captured. There are two types of heterogeneity of capture probability, (i) heterogeneity due to measurable individual characteristics, and (ii) heterogeneity due to unmeasurable individual characteristics. Heterogeneity due to measurable individual characteristics refers to situations when the factors causing the differences can be identified. For examples, (i) Males and females may have different capture probabilities. Differences may occur because males are more active than females during the time of the survey, e.g., males are bringing food to offspring while the females brood offspring and so forth. (ii) Condition of individuals may affect capture probability because animals in poor condition are more easily attracted to traps. Heterogeneity due to unmeasurable individual characteristics refers to situations when the factors causing differences in capture probability cannot be identified. This could simply mean that we have absolutely no clue what might cause differences, but are willing to accept that (even though we are clueless) there might be differences that we cannot measure. This could also be situations where we cannot observe the difference at the time of capture. For example, in some cases we cannot identify the sex of individuals at the time of capture. Age may also cause differences in capture probabilities but probably could not be determined in the field.

Modelling heterogeneity in capture probabilities is a serious problem in estimating animal abundance in capture-recapture studies. Failure to account for such heterogeneity has long been known to cause substantial bias (Cormack, 1968; Manly, 1970; Otis et al., 1978; Chao and Lee, 1992; Lee and Chao, 1994). Moreover, Link (2003) shows that without strong assumptions on the underlying distribution, estimates of population size under individual heterogeneity model is fundamentally non-identifiable. The effect of ignoring heterogeneous probabilities of capture leads to capture probability being over estimated and abundance being underestimated (Hwang and Huggins, 2005). This can be partially overcome by modelling the heterogeneity of capture probabilities through individual covariates such as age, and sex, and environmental conditions such as temperature, rainfall, and location (Pollock et al., 1984; Huggins, 1989; Alho, 1990; Pollock, 2002). Huggins (1989) and Alho (1990) developed methods for incorporating measurable covariates causing heterogeneity in estimation, while Burnham (1972), Agresti (1994), Norris and Pollock (1996), Pledger (2000) and Dorazio and Royle (2003) developed methods for accommodating heterogeneity due to unmeasured covariates.

1.7 Motivating Data Sets

In this section we give a brief description of the motivating capture-recapture data sets considered in this dissertation. Some details on the captured species are also included.

1.7.1 Deer Mice Data

The Deer mice data considered in this dissertation were originally collected by V. Reid and distributed with the program CAPTURE of Otis et al. (1978) which is extracted from Appendix 1 from Huggins (1991). V. Reid reported the results of live-trapping Deer mice (*Peromyscus maniculatus*) (see Figure 1.1) in a drainage

Figure 1.1: A photograph of *Peromyscus maniculatus* (Deer mice)

Photo credit Phil Myers, collected from
http://sciblogs.co.nz/misc-ience/tag/sperm/

bottom of sagebrush, gambel oak, and serviceberry with pinyon pine and juniper on the uplands. The data were taken in a 6 consecutive night's live-trapping study ($m = 6$) in the summer of 1975 near Wet Swizer Creek, Rio Blanco County at East Stuart Gulch Colorado. A rectangular grid of 9×11 traps was used, with 50-foot (15.2-m) trap spacing. One Sherman live trap (for small mammals) was placed at each grid point and trapping was done twice daily (morning and night) (Otis et al., 1978). Totally 38 distinct mice were caught and associated with three covariates: age (young, semi-adult or adult), sex (male or female), and weights in grams. These data are well known and have analysed in numerous capture-recapture literature, for example see Huggins (1991), Huggins and Yip (1997) and Stanley and Richards (2005). The data are reported in Appendix C.

1.7.2 Least Chipmunk Data

In the summer of 1975, V. Reid laid out a 9 × 11 livetrapping grid with traps spaced 50 feet (15.2 m) apart at a Colorado location in a bottom area dominated by sagebrush and snowberry and peripherally by gambel oak, serviceberry, and juniper. Least chipmunk (*Eutamias minimus*) (see Figure 1.2) were trapped once a day, for 6 consecutive days ($m = 6$) (Otis et al., 1978). The data were recorded with the sex (male and female) and capture history of each captured individual. Based on the discriminant procedure described in Otis et al. (1978), the temporal variation model ($\mathrm{M_t}$) was indicated as the most appropriate model. Recently, Wang et al. (2007) and Xu et al. (2013) reanalyzed and discussed this data set. The data are reproduced in Appendix C.

Figure 1.2: A photograph of *Eutamias minimus* (Least chipmunk)

Photo credit Michael Patrikeev, collected from
http://www.wildnatureimages.org/id12.html

1.7.3 House Mouse Data

The House mouse data also considered in this dissertation were originally collected by Coulombe in his unpublished master's thesis in 1965, University of California, Los Angeles, and available as one of the sets of data accompanying their program CAPTURE (Rexstad and Burnham, 1991). He conducted a live trapping study on an outbreak of feral House mice (*Mus musculus*) (see Figure 1.3) in mid-December 1962 as the result of a population ecology study on salt marsh rodents at Ballana Creek, Los Angeles County, California.

Figure 1.3: A photograph of *Mus musculus* (House mouse)

Photo credit Josef Lubomir Hlasek, collected from
http://www.guildfordpestcontrol.com/RatandMice.html

A square 10×10 grid was used with traps spaced 3m apart and trapping was done twice daily, morning and evening, for 5 days (Otis et al., 1978, pp.62-64). Hence, there are 10 trapping occasions ($m = 10$). The data recorded associated with two covariates: age (juvenile, semi-adult or adult) and sex (male or female). This data set has previously analyzed and discussed by Otis et al. (1978), Huggins (1989), Huggins and Yip (2001), Hwang and Chao (2002). The data are also presented in Appendix C.

1.8 Research Goals and Outline of Subsequent Chapters

This research aims at establishing methodology to estimate population size in closed capture-recapture experiment taking into account individual heterogeneity to capture probabilities and dependency among capture occasions. For this purpose, we have reviewed the concerning methodologies and available estimators. Ultimately, we propose alternative analytical approach to estimate population size and make comparison with currently used methodologies.

This thesis consists of five more chapters. A brief outline is given below:

Chapter 2 presents a summary of the relevant methodologies that enhance to build up our proposed Generalized Estimating Equations (GEE) approach in capture-recapture closed population models to estimate the population parameters.

Chapter 3 presents the proposed quasi-likelihood based generalized estimating equations approach in capture-recapture closed population models to estimate capture probability and population size. Here GEE versions of all closed population capture-recapture models, and their corresponding estimating equations are also presented. Quasi-likelihood Information Criterion (QIC) is used to evaluate the effect of accounting for different correlation structures among capture occasions on capture-recapture model selection. Finally, we fit the behavioral response and heterogeneity among individuals model ($\mathrm{M_{bh}}$) using our proposed approach to real capture-recapture data and conduct a simulation study to examine the performance of the estimators and their standard errors.

In Chapter 4, we apply our proposed GEE approach in the temporal variation and heterogeneity among individuals model ($\mathrm{M_{th}}$) of capture-recapture closed population to estimate population parameters. Our approach is illustrated on three real capture-recapture data sets, and a simulation study is also conducted.

In Chapter 5, estimation of closed population size is considered under heterogeneity model ($\mathrm{M_h}$) through individual covariates. We consider a partial likelihood approach which is constructed from a conditional likelihood. This allows for extra modelling flexibility using techniques, such as GLM and GLMM. Here we compare estimates of population size and corresponding standard errors based on the three estimating methodologies: GLM, GLMM and our propose GEE. Three real data sets are used for illustration purposes. A simulation study is also conducted to show the performance of the estimation procedures.

Chapter 6 provides a general conclusion and further related research topics for future work.

Chapter 2

Methodological Reviews

This chapter begins with an introduction to capture-recapture methods followed by a brief literature review on founding capture-recapture closed population models. Other common statistical methodologies considered throughout this dissertation are also provided for the reader with general details and references. The methods presented in this chapter are not new and are simply intended to give a clearer exposition. Notation and techniques described here are presented in their simplest form, and will be generalized to the capture-recapture context in subsequent chapters.

2.1 Capture-Recapture Methods

The estimation of animal abundance is important for decision making and planning in both the theoretical and applied biological sciences. In recent years, capture-recapture methods have experienced important theoretical developments. New application areas have been added to their spectrum, in turn supporting new developments on the methodological side. Conventional capture-recapture models are either closed or open models. For example, closed population models arise when the time period between capture occasions is short, and it is assumed the population is closed to immigration, emigration, birth and death. Open population models arise when any of these assumptions is relaxed. It is also possible to have models that combine these assumptions. For example, the so called Robust Design assumes that there are series of capture occasions for which the population is closed interspersed with periods for which the population is assumed open (Pollock, 1982).

Capture-recapture has its origin in the Biological/Ecological Sciences with the work of Lincoln (1930) and Petersen (1896). More than one hundred years ago Petersen published his landmark paper suggesting what later became known as the Lincoln-Petersen estimator, still in use by numerous practitioners. The Lincoln-Petersen estimator and the original capture-recapture models of Schnabel (1938) assume a closed population where the primary aim is to estimate the size of the population and capture probabilities are treated as nuisance parameters. The pioneering work of Darroch (1958) provided a mathematical foundation for the theory. Seber (1982, 1986, 1992), Schwarz and Seber (1999), Chao (2001), Williams et al. (2002) and Amstrup et al. (2005) provided excellent and comprehensive reviews of models for estimating animal abundance in closed capture-recapture studies. Other important general references on closed models and applications include Otis, Burnham, White, and Anderson (1978), White, Anderson, Burnham, and Otis (1982), Pollock (1991), Hook and Regal (1995), and the two reviews by the International Working Group for Disease Monitoring and Forecasting (IWGDMF) (1995a,b). Recent encyclopedia articles include Cormack and Buckland (1997) and Chao (1998).

Capture-recapture techniques are often much less expensive and may be more informative than hypothesis testing (classical) approaches to estimate population parameters. Researchers face limited budgets in both the developing and the industrialized world. Those interested in the size of difficult-to-identify populations will undoubtedly find estimation procedures based on these methods appealing.

Capture-recapture methods can be seen to be applied in three major sciences:

(i) Biological sciences, where the size of animal populations and their diversity are of importance.

(ii) The life and medical sciences, where we often want to know the size of the population having hidden disease, such as depression or drug use.

(iii) The social sciences, where we may be interested in the amount of illegal activities, such as illegal immigration.

Since every estimation method is based on a set of assumptions, hence the first step is to decide what assumptions must hold if the estimated population size provide a suitable estimate of true population. The key general assumptions for capture-recapture methods are listed below.

(a) ***The population is closed either geographically and demographically.***

A population is assumed to be closed both geographically and demographically (i.e., no additions or deletions) when there is no change in population size (N). No births, deaths, or migration in or out of the population occur among the sampling periods. This assumption can be weakened. If only deletions occur during the experiment, N will be an estimator of the population size at the time of the first sample whereas if only additions occur N will be an estimator of the population size at the time of the last sample (Pollock et al., 1990).

(b) ***Animals do not lose their marks during the trapping occasions.***

Marks loss represents one form of violation of the closure assumption. The population size N will be overestimated if marks are lost. Carrying out short-term studies may sometimes reduce the loss of marks. Alternatively, double marking is a technique used to estimate mark loss (Seber, 1982).

(c) ***All marks are recorded accurately and are neither overlook nor misread, and are reported on each trapping occasion j.***

This assumption is violated when observers do not see tags that have been previously put on some of the marked animals. This problem is particularly common in fisheries. Sometimes missing tags can be avoided or reduced by careful planning of the study. A training program for the observers and pilot studies in the field are helpful to identify any problem with the marking method (Otis et al., 1978) which might cause marks to either be lost or overlooked.

(d) ***All animals are equally likely to be captured on each trapping occasion. This also implies that capture and marking do not affect catchability of the animal.***

This assumption is very difficult to achieve in practice because capture probabilities vary with time, behavioral (or trap) response, and inherent differences among individuals in the population (heterogeneity). Capture probabilities can change during the period of study depending on weather conditions (time variation). For example, a cold rainy day during the study might reduce activity of the animals and reduce the probability of capture. The trapping method used can also affect the trap response, and consequently the capture probability. Baiting traps, for example, is likely to lead to a trap happy response where marked animals are more likely to be caught than others due to differences in species, sex, age, social dominance, number and placement of traps or other inherent characteristics (homogeneity) (Pollock et al., 1990).

We are interested to estimate population size of capture-recapture closed population models. Hence all the models discussed in this thesis assume population closure. A closed population is one that experiences neither losses nor gains between sampling occasions. Therefore, the parameter we wish to estimate is population size N, which is assumed to be constant in the course of the capture-recapture study. Moreover, because the same individual animals compose the population on each trapping occasion, $j = 1, 2, ..., m$, we can conceive of the individuals as being numbered $i = 1, 2, ..., N$. The basic capture data are conveniently expressed in matrix form as

$$\underline{\mathrm{Y}} = \begin{pmatrix} Y_{11} & Y_{12} & ... & Y_{1m} \\ Y_{21} & Y_{22} & ... & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{N,1} & Y_{N,2} & ... & Y_{N,m} \end{pmatrix}$$

where the subscript i for matrix element Y_{ij} denotes an individual animal ($i = 1, 2, ..., N$) and j denotes the capture occasion ($j = 1, 2, ..., m$) and

$$Y_{ij} = \begin{cases} 1, & \text{if individual } i^{th} \text{ animal is caught on the } j^{th} \text{ capture occasion} \\ 0, & \text{otherwise.} \end{cases}$$

The $\underline{\mathrm{Y}}$ matrix is a simple way to record the capture or noncapture of each animal in the population on each trapping occasion. Row i gives the trapping results for individual i, while column j gives the results of the j^{th} trapping occasion. Capture-recapture estimators are based on probabilistic models of events giving rise to the data in $\underline{\mathrm{Y}}$. In models of closed populations, the relevant model parameters include capture probabilities at each of the capture occasions.

In many capture-recapture experiments, covariates are collected when the individuals are caught. These covariates could be individual-specific, such as sex, length or weight, or they could be common to all individuals caught in that sample, for example, air temperature.

2.1.1 Closed Population Models

The key parameter of interest in closed population models is the size of the population, denoted N. When studying animal populations, scientists and biologists are usually interested in sampling a population several ($m > 2$) times. In this situation, it is necessary to keep a complete capture history of those animals that have been captured at least once. Population size estimates could be affected by chances in capture probabilities. Capture probabilities vary due to factors such as time, behavioral response to capture, and other individual inherent factors (sex, age, etc.). Inherent heterogeneity among animals in the population causes bias on estimates of N while behavioral responses may cause positive or negative bias on estimates of N. A 'trap happy' response produces a negative bias, and consequently underestimates the population size N, whereas a 'trap shy' response causes positive bias overestimating population size N. In order to estimate N, we write the likelihood in terms of capture probabilities

$$\mathbf{P} = \begin{pmatrix} P_{11} & P_{12} & ... & P_{1m} \\ P_{21} & P_{22} & ... & P_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N,1} & P_{N,2} & ... & P_{N,m} \end{pmatrix}$$

where P_{ij} is the probability of individual i being caught on occasion j. The challenge is to recognize and model the patterns in these probabilities and to use the corresponding probability models to develop estimates of population size. We use information from the individuals observed, as well as, assumptions about $\mathbf{P}$ to estimate the number of individuals with capture history $(0, ..., 0)$ that were available for capture but never seen. The standard likelihood is

$$L(\mathbf{P}, N; \underline{Y}) \propto \frac{N!}{n!(N-n!)} \prod_{i=1}^{n} \prod_{j=1}^{m} P_{ij}^{Y_{ij}} (1 - P_{ij})^{1-Y_{ij}} \prod_{i=n+1}^{N} \prod_{j=1}^{m} (1 - P_{ij}). \tag{2.1}$$

This is the most general likelihood with assumptions about **P** leading to different classes of model where n is the total number of individuals were captured during the study period. A series of specific models for capture data derived by Otis et al. (1978) defining the following general structural model: P_{ij} = the capture probability of the i^{th} individual in the population being captured on the j^{th} trapping occasion, where $i = 1, 2, ..., N$, and $j = 1, 2, ..., m$. Following Otis et al. (1978), the model can incorporate up to three sources of variation among capture probabilities: a temporal effect (subscript t), a heterogeneity effect among individuals (subscript h) and a behavioral effect in response to the trap (subscript b). They give details about 8 classes of model: $\mathrm{M_0}$, $\mathrm{M_t}$, $\mathrm{M_b}$, $\mathrm{M_h}$, $\mathrm{M_{tb}}$, $\mathrm{M_{th}}$, $\mathrm{M_{bh}}$ and $\mathrm{M_{tbh}}$ assuming independence among capture occasions and individuals. Estimators are available of these models, and a computer program called CAPTURE (Otis et al., 1978; Rexstad and Burnham, 1991) has been written in order to help biologists and scientists to choose the most suitable model for a particular set of data. CAPTURE uses numerical methods to obtain the MLE of N when $m > 2$. A list of these models with a summary of their basic characteristics and assumptions follows:

• Constant Capture Probability- Model $\mathrm{M_0}$

The simplest of all models under consideration results from the assumption that all members of the population are equally at risk to capture on every trapping occasion. Moreover, the occasions themselves do not affect capture probabilities. We thus have a model in which there is no heterogeneity of capture probability, no behavioral response to capture, and no variation in the experimental situation over time. This model is designated Model $\mathrm{M_0}$, and involves only 2 parameters: N, the population size, and P, the probability that an animal is captured on any given capture occasion.

The model $\mathrm{M_0}$ has one capture probability that is constant across individual and time, mathematically,

$$P_{ij} = P;\ i = 1, 2, ..., N,\ \ j = 1, 2, ..., m.$$

It is straightforward to show that the joint probability distribution for the data under model $\mathrm{M_0}$ can be written as

$$\Pr(\{Y_\omega\} \mid N, P) = \frac{N!}{[\prod_\omega Y_\omega!](N-n)!} P^{n\cdot}(1-P)^{mN-n\cdot} \tag{2.2}$$

where $n. = \sum\limits_{j=1}^{m} n_j$, is the total number of captures where n_j is the number of individuals caught on each sampling occasion, and $\{Y_\omega\}$ denotes the number of animals exhibiting each capture history ω.

Although this model is very unlikely to hold in practice, it provides a base for generalization to more complex models, and has been used as the basis of tests of assumptions. In the case of several sampling occasions ($m > 2$), the only way to estimate N and its standard error is by using an iterative computer procedure. Simulation studies have shown that assumption violations, especially unequal catchability, cause serious bias, and for low capture probabilities (P) and small population size (N), the estimator of N is not precise (White et al., 1982).

- **Temporal Variation- Model** $\mathrm{M_t}$

The model $\mathrm{M_t}$ was originally developed by Schnabel (1938), and it takes into account that capture probabilities may vary over time. The model assume that all members of the population are equally at risk to be captured on the j^{th} trapping occasion. Thus, all animals have the same probability of capture on any particular trapping occasion, but that probability can change from one occasion to the next. For example, a cold rainy day may reduce the activity of the animals, and reduce their capture probabilities. On the other hand, animals may be more active on a warm sunny day and thus their capture probabilities increase.

Mathematically,

$$P_{ij} = P_j; \;\; i = 1, 2, ..., N; \;\; j = 1, 2, ..., m.$$

The set of $(m + 1)$ parameters involved in this model contains N, the population size, and the $P_j, j = 1, 2, ..., m$, where P_j is the probability of capture on the j^{th} occasion. The joint probability distribution for the data under model $\mathrm{M_t}$ can be written as

$$\Pr(\{Y_\omega\} \mid N, \underline{P}) = \frac{N!}{[\prod_\omega Y_\omega!](N-n)!} \times \prod_{j=1}^{m} P_j^{n_j}(1 - P_j)^{N-n_j} \tag{2.3}$$

where $\underline{P}$ is the vector of capture probabilities, $P_1, P_2, ..., P_m$. Thus, the statistics needed for estimation are simply the number of animals caught on each sampling occasion (n_j) and the total number of individuals captured (n).

Pollock et al. (1990) recommend the use of CAPTURE to find the maximum likelihood estimation (MLE) for N in this model. They also point out that a detailed capture history is not required for this model because of its simplicity. Yip (1991) came up with a estimate of N and its variance based on a method of moments for martingales. Bayesian methods have also been used for model $\mathrm{M_t}$. Some references are Castledine (1981), Gazey and Staley (1986), Smith (1988, 1991).

• Behavioral Response- Model $\mathrm{M_b}$

This model deals with the failure of the assumption that initial capture does not affect the probability of capture on subsequent occasions. That is, the model allows an animal to exhibit a behavioral response to capture and become either 'trap happy' or 'trap shy'. For example, if food is provided in a trap, the animal may be 'happy' after capture and try to enter another trap afterwards. Then its recapture probability increases. However, if an animal is frightened or hurt during initial capture or marking, it will try to avoid entering another trap, and thus its capture probability decreases. Formally, model $\mathrm{M_b}$ assumes that on any given trapping occasion, all unmarked animals have one probability of capture, and all marked animals have another probability of capture. It is assumed that capture probabilities do not vary with time. An implication of those assumptions is that all members of the population have the same probability of capture at the beginning of the experiment. Also note that the assumption is made that an animal's capture probability is altered only once, after first capture. Although one might think it is more realistic to allow the capture probability to be changed more than once (e.g., after both first and second capture) this more general assumption turns out to have no effect on the estimation of population size N. Therefore, for simplicity of presentation the assumption is made that all marked animals, regardless of the number of times they might have been captured, have the same capture probability.

The assumptions of model $\mathrm{M_b}$ result in 3 model parameters: N, population size; P_c if an individual has never been caught; and P_r if an individual has been caught before. Mathematically,

$$P_{ij} = \begin{cases} P_c, & \text{if } i \text{ has never been caught} \\ P_r, & \text{if } i \text{ has been caught before} \end{cases} \quad ; i = 1, 2, ...N; \;\; j = 1, 2, ..., m.$$

This model allows for the study to influence the behaviour of the animals. If the capture process is a bad experience for the animals, we get a trap-shy response with $P_c > P_r$. However, if the capture process is a good experience for the animals, we get a trap-happy response with $P_c < P_r$. Note that these responses can also be induced by the study design used. For example, if our sampling targets areas where animals were previously caught, we are likely to observe a trap-happy response.

To describe the likelihood function for this model, let q_j be the total number of marked animals caught on capture occasion j, with Q_j the number of marked animals in the population at the time of capture occasion j. The probability distribution for model $\mathrm{M_b}$ can be described in terms of the total number of recaptures

$$q_{\cdot} = \sum_{j=2}^{m} q_j,$$

during the study, the total number of marked individuals at the completion of the study (n), and the sum (over all occasions) of the number of marked animals available for capture at each capture occasion

$$Q_{\cdot} = \sum_{j=2}^{m} Q_j.$$

The corresponding probability density function as Williams et al. (2002) is given by

$$\begin{aligned}\Pr\bigl(\{Y_\omega\} \mid N, P_c, P_r\bigr) &= \frac{N!}{[\prod_\omega Y_\omega!](N-n)!} {P_c}^n \\ &\times \ (1-P_c)^{mN-n-Q_\cdot} \times {P_r}^{q_\cdot} \\ &\times \ (1-P_r)^{Q_\cdot - q_\cdot}. \end{aligned} \tag{2.4}$$

The trap response model ($\mathrm{M_b}$) is statistically similar to the removal model (Seber, 1982), but animals are not being physically removed from the population, instead they are considered removed from the unmarked population after the initial capture. A linear regression model to estimate N and P is also mentioned in the literature (Seber, 1982). Since little difference has been found between the linear regression method and the trap response model, Pollock et al. (1990) recommended the use of the maximum likelihood estimation available in the program CAPTURE for the trap response model. Maximum likelihood estimation has theoretical advantages.

- **Heterogeneity among Individuals- Model** $\mathrm{M_h}$

This model assumes that there is no difference between trapping occasions and no behavioral response to capture, but that there is heterogeneity among the capture probabilities of individuals. Different home range size, social dominance, age, gender or weight may cause individual heterogeneity of capture. For example, assume that male birds are used to finding food and female birds are used to guarding the nest; then the capture probability of males is higher than that of females. Model $\mathrm{M_h}$ has a different capture probability for each individual in the population which is constant through time, i.e.,

$$P_{ij} = P_i; \;\; i = 1, 2, ..., N; \;\; j = 1, 2, ..., m.$$

Conceptually, model $\mathrm{M_h}$ involves $N + 1$ parameters: the population size N and the set of capture probabilities $P_i, i = 1, 2, ..., N$, where P_i is the probability of capture of the i^{th} animal on any trapping occasion. This more general formulation of model $\mathrm{M_h}$ does not allow estimation of population size N because of the presence of too large a number of nuisance parameters (Otis et al., 1978).

The historically popular approach to estimate N for model $\mathrm{M_h}$ is to use the nonparametric jackknife of Burnham and Overton (1978). Chao (1988) proposed a moment estimator which is generally less biased than the jackknife estimator, especially when capture probabilities are small (Pollock et al., 1990). Stratification technique can also be used to reduce heterogeneity of capture probabilities (for example, separate analyses by sex or age or other variables which might cause heterogeneity in capture probabilities); however, there are disadvantages of stratification when the sample sizes are too small in the strata. Lee and Chao (1991), and Norris and Pollock (1996) have proposed nonparametric approaches to derive estimators of population size for this model. Alternatives include using finite mixtures for $(\mathbf{P})$ (Pledger, 2000) or assuming the capture probabilities $(\mathbf{P})$ are drawn from a probability density function with support on $[0, 1]$. Holzmann et al. (2006)

showed that the population size N is identifiable in model $\mathrm{M_h}$ conditional on a model for $(\mathbf{P})$. However, Link (2003) showed that different probability density functions for $(\mathbf{P})$ can lead to identical sufficient statistics with different values of N.

The approach of Burnham and Overton (1978) was to conceptualize the vector of capture probabilities $\{P_i\}$ as a random sample of size N from some probability distribution $F(P)$ defined on the interval $[0,1]$. The corresponding statistical model can be described in terms of the number f_j of animals caught on exactly j occasions:

$$\Pr(f_1, f_2, ..., f_m \mid F) = \frac{N!}{\big[\prod\limits_{j=1}^{m} f_j!\big]\big(N-n\big)!}{\eta_0}^{N-n}\big[\prod_{j=1}^{m}\eta_j^{f_j}\big] \tag{2.5}$$

where $\eta_j = \int_0^1 \frac{m!}{(m-j)!j!}P^j(1-P)^{m-j}\,dF(P)$. The cell probability η_j in Eq. (2.5) can be viewed as the average probability that an individual is caught exactly j times.

If the investigator can identify covariates (e.g., a size variable) associated with variation in capture probability among individuals, it is possible to use this additional information in estimation under a special case of model $\mathrm{M_h}$. Pollock et al. (1984), Huggins (1989, 1991), and Alho (1990) all consider modelling capture probability as a linear-logistic function of individual covariates, e.g., as

$$P_i = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}},$$

where β_0 and β_1 are parameters to be estimated and x_i is the covariate value for individual i.

• Temporal Variation and Behavioral Response- Model $\mathrm{M_{tb}}$

This model assumes a behavioral response to capture and also permits temporal variation in both initial capture and recapture probabilities. The model contains

$2m$ parameters: population size N, a vector $\underline{P}'_c = \{P_{c1}, P_{c2}, ..., P_{cm}\}$ of initial capture probabilities, and a vector $\underline{P}'_r = \{P_{r2}, P_{r3}, ..., P_{rm}\}$ of recapture probabilities. Mathematically,

$$P_{ij} = \begin{cases} P_{cj}, & \text{for any first capture,} \quad j = 1, 2, ..., m \\ P_{rj}, & \text{for any recapture,} \qquad j = 2, 3, ..., m \end{cases}$$

The corresponding joint probability distribution for the data can be written in several ways (Otis et al., 1978), including

$$\begin{aligned} \Pr(\{Y_\omega\} \mid N, P_c, P_r) &= \frac{N!}{\left[\prod_\omega Y_\omega!\right](N-n)!} \\ &\times \left[\prod_{j=1}^{m} P_{cj}^{u_j}(1-P_{cj})^{(N-Q_{j+1})}\right] \\ &\times \left[\prod_{j=2}^{m} P_{rj}^{q_j}(1-P_{rj})^{Q_j-q_j}\right] \end{aligned} \tag{2.6}$$

where u_j and q_j are the numbers of unmarked and marked animals, respectively, that are caught at time j, and Q_j is the number of marked animals present in the population at the time of occasion j.

Burnham (1972) has come up with a estimator for this model which assumes that capture probabilities vary with time and with behavioral effects (trap happiness, trap shyness). The model assumes that recapture probability is some function of initial capture probability, $C = \sqrt[\theta]{P}$, as otherwise there are too many parameters and the model is nonidentifiable. Parameter estimates for this model, such as population size, initial capture probabilities, and θ, have been implemented in the version of CAPTURE. The program also provides a variance-covariance matrix of the parameter estimates and goodness-of-fit test (Rexstad and Burnham, 1991). Again this estimator does not have good precision properties unless P and C are quite high. Lloyd (1994) developed a martingale estimator for this model and studied his efficiency compared to that of the $\mathrm{M_t}$ model.

• Temporal Variation and Individual Heterogeneity- Model $\mathrm{M_{th}}$

This model permits variation in capture probabilities P_{ij} both over time, $j = 1, 2, ..., m$, and for individual animals, $i = 1, 2, ..., N$. If, on the j^{th} trapping occasion, the i^{th} animal has a capture probability of $P_{ij} = P_i \times P_j$ where $0 \leq P_iP_j \leq 1$ that is independent of its capture history (i.e., there is no behavioral response to capture), then model $\mathrm{M_{th}}$ is the appropriate probability model for a capture-recapture experiment on such a population. Notice that the structure of P_{ij} implies that variation in capture probabilities due to time is independent of the variation caused by individual heterogeneity. Otis et al. (1978) viewed P_i as a random sample from some probability distribution $F(P)$ and described the probability distribution of the observed sample $\{Y_{ij}\}$ as

$$\mathrm{Pr}[Y_{ij}] = \mathrm{Pr}[Y_{ij}|n]\mathrm{Pr}[n],$$

with

$$\mathrm{Pr}[Y_{ij}|n] = \left[\prod_{j=1}^{m} P_j^{n_j}\right]\left[\prod_{i=1}^{n}\int_0^1 P^{Y_i}\,[\prod_{j=1}^{m}(1 - PP_j)^{1-Y_{ij}}]\,dF(P)\right],$$

where Y_i is the number of times animal i is captured and $\mathrm{Pr}[n]$ is the probability distribution of the number of animals (n) caught in the study, depending on the parameters $N, P_1, ..., P_j$, and the distribution $F(P)$.

This is useful as a conceptual model of how time and heterogeneity might operate as a simple product. Chao et al. (1992) has developed an estimator for this model. The proposed estimator has been implemented in the version of program CAPTURE. The program provides parameter estimates for population size and average capture probabilities for each occasion (Rexstad and Burnham, 1991).

The estimator may not have good precision unless capture probabilities are high on average.

If capture probability can be modelled using individual covariates, then the linear-logistic modelling approach of Pollock et al. (1984), Huggins (1989, 1991), and Alho (1990) can be implemented for this special case of $\mathrm{M_{th}}$. In particular, capture probability P_{ij} for individual i at time j can be modelled as,

$$P_{ij} = \frac{e^{\beta_{0j}+\beta_1 x_i}}{1+e^{\beta_{0j}+\beta_1 x_i}},$$

where β_0 and β_1 are parameters to be estimated and x_i is the covariate value for individual i.

• Behavioral Response and Individual Heterogeneity- Model $\mathrm{M_{bh}}$

The model assumes that every member of the population has a specific pair of capture probabilities: P_i, the probability that the i^{th} animal is caught on any trapping occasion given that it has not been previously captured; and C_i, the probability that the i^{th} animal is recaptured given that it has been captured at least once previously. Mathematically,

$$P_{ij} = \begin{cases} P_i, \ \text{for first capture}, \quad j = 1,2,...,m \\ C_i, \ \text{for all recapture}, \quad j = 2,3,...,m \end{cases}$$

for $i = 1,2,...,N$.

Thus, the model allows both behavioral response to first capture and individual heterogeneity of capture probabilities. An important and appealing characteristic of the model is that it allows the behavioral response to capture to vary with the animal, i.e., all members of the population do not exhibit an identical response to first capture. The most general formulation of the model involves $2N + 1$ total parameters consisting of N (population size), and 2 capture probabilities for each of the N members of the population. The density function for this model is available in Otis et al. (1978).

The model $\mathrm{M_{bh}}$, first considered by Pollock (1974) and later by Otis et al. (1974), is a generalization of the removal method (see Otis et al., 1978 and Pollock et al., 1990 for details). Otis et al. (1978) proposed an estimator termed by the general removal estimator while Pollock and Otto (1983) proposed an alternative estimator based on the jackknife method which is sometimes better than the generalized removal estimator. Lee and Chao (1991), and Norris and Pollock (1996) have proposed nonparametric approaches to derive estimators of population size for this model. If capture probabilities can be modelled using individual covariates, then the logistic models of Pollock et al. (1984), Huggins (1989, 1991), and Alho (1990) can be used with $\mathrm{M_{bh}}$.

• Temporal Variation, Behavioral Response and Individual Heterogeneity- Model $\mathrm{M_{tbh}}$

The most conceptually useful model is $\mathrm{M_{tbh}}$ because it represents the case in which all three sources of variation operate. The model contains an initial capture probability for each individual in each capture occasion ($P_{ij}, i = 1, 2, ..., N; j = 1, 2, ..., m$) as well as a recapture probability for each individual in each capture occasion after the first ($C_{ij}, i = 1, 2, ..., N; j = 2, 3, ..., m$). That is, for $i = 1, 2, ..., N$,

$$P_{ij} = \begin{cases} P_{ij}, & \text{for first capture,} \quad j = 1, 2, ..., m \\ C_{ij}, & \text{for all recapture,} \quad j = 2, 3, ..., m. \end{cases}$$

It is the most realistic model for a closed population, but the model has no practical use in estimating the size of the population because this model has $(2m-1)N+1$

parameters, which clearly are not estimable with the available data. Pledger (2000) considered estimation for model $\mathrm{M}_{\mathrm{tbh}}$ using linear-logistic modelling in conjunction with her finite mixtures approach. She considered time, behavioral response, and heterogeneity as main effects and incorporated interaction terms as well. Though her 'fully interactive' model with all interactions (including the three-way interaction) requires some constraints on parameters, she also considered partially interactive models incorporating most or all of the two way interactions.

2.1.2 Covariates in Capture-Recapture Models

In capture-recapture experiments field researchers frequently measure covariates in addition to capture records that might be associated with the capture probability of the animals in the population. The types of covariates considered include group or environmental covariates (e.g. temperature, humidity and rainfall) that may vary by occasion, but are constant over animals, and individual animal covariates (e.g. age, sex, body weight etc.) that are usually assumed to be constant over time. The covariates can help to explain capture probabilities and improve estimation of population size (N). Statisticians may think that the primary purpose of using covariates is to get a more robust model and avoid bias due to heterogeneity (in the closed models). Biologists may be drawn to covariate modelling to explore important ecological hypotheses on the relationship of capture probabilities to auxiliary variables such as animal weight. However, there are some very important reasons to model covariates in capture-recapture study.

(i) it improves the estimation of the model and thus precision of all parameter estimates is increased.

(ii) the models provide a clear description of the sources of heterogeneity, and the each covariate effect can be assessed.

(iii) model comparisons and model selection can be performed under a unified framework.

(iv) there may be inherent ecological importance in understanding the nature of the relationships between the parameters and specific environmental and individual animal covariates. Hence, it helps to investigate biologically interesting relationships. In this case the effect of the covariates improves knowledge about the population dynamics of the species.

Surprisingly, little attention was given to incorporating covariates in early capture-recapture development. One of the first papers to consider using covariate data in models of closed populations was written by Pollock et al. (1984). Pollock, Hines and Nichols (1984) proposed an estimation procedure that accommodates individual characteristics to model the capture probabilities of the animals. They only considered categorical covariates. Another difficulty arises in the model of Pollock et al. (1984) for individual covariates is that no information is available for the individuals in the population that were never captured. Huggins (1989, 1991) and Alho (1990) extended the case of continuous covariates. They developed a conditional likelihood model to estimate capture probabilities in terms of observable characteristics related to the capture individuals. However, one assumption is that the heterogeneity effects are fully determined by the observable covariates. The use of covariates does not take into account the heterogeneity due to unobservable innate characteristics of the individuals. If some important covariates were not recorded, the models might workless well than those without using covariates.

2.1.3 Horvitz-Thompson Estimator

The estimator proposed by Horvitz-Thompson (1952) is a general estimator for a population size, which can be used for any probability sampling plan. This includes both sampling with and without replacement. Let P_i^*, for $i = 1, 2, ..., N$, be the probability that the i^{th} individual of the population is included in the sample (inclusion probability). The Horvitz-Thompson (H-T) estimator is unbiased to the size of the population, and is given by

$$\hat{N} = \sum_{i \epsilon s_n} \frac{1}{P_i^*}, \tag{2.7}$$

where s_n is the sample composed only by n distinct sample animals, the value n is sometimes referred to as the 'effective' sample size. Therefore, this estimator is independent on the number of times selected each animal, since each animal is used only once. The variance of this estimator is given by

$$\text{Var}[\hat{N}] = \sum_{i=1}^{N} (\frac{1 - P_i^*}{P_i^*}) + \sum_{i=1}^{N} \sum_{j \neq i} (\frac{P_{ij}^* - P_i^* P_j^*}{P_i^* P_j^*}), \tag{2.8}$$

where, P_{ij}^* is the joint inclusion probability of units i and j.

An unbiased estimator of variance is given by

$$\widehat{\text{Var}}[\hat{N}] = \sum_{i \epsilon s_n} (\frac{1 - P_i^*}{P_i^{*2}}) + \sum_{i \epsilon s_n}^{N} \sum_{j \neq i} (\frac{P_{ij}^* - P_i^* P_j^*}{P_i^* P_j^*}) \frac{1}{P_{ij}^*}, \tag{2.9}$$

where the extra P_i^* in the denominator of the first term and the P_{ij}^* in the denominator of the second term can be attributed to the use of n sample units instead of the N population units in the theoretical variance, if all the joint inclusion probabilities are not zero. Note that this variance estimators may sometimes give negative results, then, alternatives unbiased approximations should be selected. The inclusion probabilities must be estimated when these probabilities are unknown, then the estimator is called Horvitz-Thompson type. This estimator is not centred, and there is no analytic expression for its variance, hence, it has to be approximated by the delta method or methods for re-sampling.

2.2 Generalized Linear Models

The combination of many statistical methods was presented by Nelder and Wedderburn (1972) exploring the idea of a generalized linear models (GLM). In statistics, the GLM is a versatile overview of ordinary linear regression that enables for response variables that have other than a normal distribution like logistic, Poisson, Binomial, Gamma (McCullah and Nelder, 1989). The GLM generalizes linear regression by allowing the linear regression model to be associated with the response variable via a link function, and by permitting the magnitude of its variance of each measurement to be a function of its predicted value. The GLM is a typical tool in statistics and are easily applied to ecological studies. The GLM approach is attractive because

(i) it provides a general theoretical framework for many commonly encountered statistical models;

(ii) it simplifies the implementation of these different models in statistical software, since essentially the same algorithm can be used for estimation, inference and assessing model adequacy for all GLM.

2.2.1 The Linear Model

Suppose Y is an $n \times 1$ column vector containing the response variable whose components are independently distributed with $n \times 1$ column vector of means μ, X is an $n \times p$ model or design matrix of explanatory covariates with rank p and a leading column vector of ones, β is a $p \times 1$ column vector of estimated coefficients, and ϵ is an $n \times 1$ column vector of disturbances. Then the multiple regression linear model in matrix notation is given by

$$Y = X\beta + \epsilon. \tag{2.10}$$

On the right-hand side, $X\beta$ is called the systematic component and ϵ is called the stochastic component. The linear model requires a relatively strict set of assumptions. The Gauss-Markov theorem states that if:

(i) the relationship between each explanatory covariate and the response variable is linear in structure,

(ii) the residuals are independent with mean zero and constant variance,

(iii) there is no correlation between any regressor and residual,

then the solution produced by selecting coefficient values that minimize the sum of the squared residuals is unbiased and has the lowest total variance among unbiased linear alternatives. The first two assumptions are eliminated with the basic generalized linear modelling approach and the third can be relaxed with more advanced forms. However, the dependence of the variance on the mean function must be known (except in the extension based on quasi-likelihood functions). Generalized linear models provide a way to analyze the effects of explanatory covariates in a way that closely resembles that of analyzing covariates in a standard linear model, except that the assumptions are far less confining. The key is the specification of a link function which relates the systematic component of the linear model ($X\beta$) with a wider class response variables and residual forms.

2.2.2 Specification of Generalized Linear Models

Generalized linear models are an extension of classical linear models (McCullah and Nelder, 1989). To simplify the transition to generalized linear models, the generalization of the linear model produces the following three-part specification:

(i) **The stochastic component:** the component of Y is the random or stochastic which remains distributed i.i.d. according to a specific exponential family distribution, with mean $\mu = E(Y)$. This component is sometimes also called the 'error structure', or 'response distribution'.

(ii) **The systematic component:** $\eta = X\beta$ is the systematic component producing the linear predictor, $X\beta$. This linear predictor describe how the location of the response variable changes with the explanatory covariates (Lindsey, 2007). Hence, the explanatory covariates, X, affect the observed response variable, Y, only through the functional form of the $g()$ function.

(iii) **The link function:** the stochastic component and the systematic component are linked by a function of η which is exactly the canonical link function summarized in Table 2.1. The link function connects the stochastic component which describes some response variable from a wide variety of forms to all of the standard normal theory supporting the systematic component through the mean function,

$$\begin{aligned} g(\mu) &= \eta = X\beta, \\ g^{-1}(g(\mu)) &= g^{-1}(\eta) = g^{-1}(X\beta) = \mu = E(Y). \end{aligned} \tag{2.11}$$

Hence the inverse link function ensures that $X\hat{\beta}$, here $\hat{\beta}$, the estimated coefficient vector, maintains the Gauss-Markov assumptions for linear models and all of the standard theory applies even though the response variable takes on a variety of nonnormal forms (Gill, 2001). The link function connects the linear predictor, the systematic component (η), to the expected value of the specified exponential family form (μ). It is essential that the link function chosen is differentiable and monotonic, such that β can be estimated, and that each value of $X\beta$ has only one corresponding μ value. The most important advantage of the link function is that it can be chosen independently of the distribution. In classical linear models the mean and therefore the linear predictor are identical, and the identity link is plausible in that both η and μ can take any value on the real line. The foremost common link function to use, is the canonical link function (McCullah and Nelder, 1989).

2.2.3 Properties of Generalized Linear Models

Assuming that the distribution of each component of Y follows an exponential family, taking the following form

$$f_Y(y_i;\theta_i,\phi) = \exp\left\{\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\} \tag{2.12}$$

for some specific functions $a(.), b(.)$ and $c(.)$, where θ_i is the canonical form of the location parameter, some function of the mean, μ_i. If ϕ is known, this follows an exponential-family model with canonical parameter θ_i. If ϕ is unknown, it may or may not be a two-parameter exponential family (McCullagh and Nelder, 1989). The most important distributions of the form (2.12) are summarized in Table 2.1.

- **First Two Moments of GLM**

The exponential dispersion family displayed in (2.12) does not explicitly reveal how the moments of Y_i, in particular the first two moments, may be related to θ_i and ϕ. By algebraic simplification of (2.12), we get

$$\begin{aligned} \mu_i &= b'(\theta_i), \\ \mathrm{Var}(Y_i) &= a(\phi)b''(\theta_i). \end{aligned} \tag{2.13}$$

The expression for the variance of Y_i explains the rationale behind the phrase 'dispersion parameters' for ϕ. Finally, given the relationship between θ_i and β through (2.12) and (2.13) above, $f(Y;\theta,\phi)$ can also be expressed as $f(Y;\beta,\phi)$ (McCullagh and Nelder, 1989).

- **Likelihood and Score Function**

To make inferences about the unknown parameters β and ϕ, we would like to develop the likelihood and score functions for (2.12). We assume that the Y_i's are independent and represent a random sample from the targeted population, the likelihood function for β and ϕ is simply proportional to

$$L(\beta,\phi) \propto \prod_{i=1}^{n} f(y_i;\beta,\phi) = \exp\Big\{\frac{\sum_{i=1}^{n}\theta_i(\beta)y_i - b(\theta_i(\beta))}{a(\phi)} + c(y_i,\phi)\Big\}.$$

Since the likelihood function and log likelihood function have identical modal points, it is convenient to work with the log likelihood function which is as follows,

$$l(\beta,\phi) = \frac{\sum_{i=1}^{n}\theta_i(\beta)y_i - b(\theta_i(\beta))}{a(\phi)} + c(y_i,\phi). \tag{2.14}$$

Consequently, the score function for β can be derived as

$$S_\beta(\beta,\phi) = \partial l(\beta,\phi)/\partial\beta = \sum_{i=1}^{n}\big(\frac{\partial\mu_i}{\partial\beta}\big)'\big(\mathrm{Var}(Y_i;\beta,\phi)\big)^{-1}\big(y_i-\mu_i(\beta)\big). \tag{2.15}$$

It is important to mention that the score function for β depends only on the first two moments of the Y_i's despite the full specification of $f(.)$ through (2.12). Moreover, given that $a(\phi)$ appears as a proportional factor in $\mathrm{Var}(Y_i)$, no knowledge on ϕ is needed to derive the MLE of β, $\hat{\beta}$, by solving $S_\beta(\beta,\phi)=0$, although the asymptotic variances of $\hat{\beta}$, $\sum(\beta,\phi)$, does depend on ϕ, where

$$\Sigma^{-1}(\beta,\phi) = \lim_{n\to\infty}\sum_{i=1}^{n}\big(\frac{\partial\mu_i}{\partial\beta}\big)'\big(\mathrm{Var}(Y_i;\beta,\phi)\big)^{-1}\big(\frac{\partial\mu_i}{\partial\beta}\big)/n. \tag{2.16}$$

2.2.4 Estimation of Parameters

In the generalized linear models, the unknown parameter β is typically estimated with the maximum likelihood method or the quasi-likelihood technique.

• Maximum Likelihood Estimates

Estimation of β is conducted via an iterative procedure known as iterative reweighted least squares (IRLS). The estimation procedure is based on maximizing the log-likelihood

$$l(\beta) = \sum_{i=1}^{n} \log\{f_{Y_i}(y_i;\theta)\}$$

with respect to β where θ is independent on β. This yields a set of estimating equations and an iterative procedure is then constructed. GLM share similar theoretical properties as linear models, hence GLM properties will not be covered here, for further details on GLM see Wood (2006).

• Quasi-likelihood Estimates

One needs to specify the random component for the response variables to implement the generalized linear modelling approach. In several situations, some details of the distribution governing the data is known, however the distribution may be unable to be specified correctly. This may be due to the fact that the underlying biological theory is not completely developed, and/or no substantial (empirical) knowledge of similar data from previous studies is available. This precludes the use of maximum likelihood method, which requires exact specification of the distribution in order to construct the likelihood function. Wedderburn (1974) first proposed the idea of Quasi-likelihood methods for regression analysis that requires few assumptions about the distribution of the response variable. To describe a quasi-likelihood function, only the relationship between the mean and variance needs to be specified through the variance function (Wedderburn, 1974).

Suppose the response y_i has expectation μ_i and variance function, $V(\mu_i)$, where V is some known function. The parameter of interest, β, is related to μ_i with inverse link function of $g(X_i\beta) = \mu_i$. Then for each observation the quasi-likelihood is given by

$$Q(y_i;\mu_i) = \int_{y_i}^{\mu_i} \frac{(y_i - w)}{V(w)} dw \tag{2.17}$$

or

$$\frac{\partial Q(y_i;\mu_i)}{\partial \mu_i} = \frac{(y_i - \mu_i)}{V(\mu_i)}.$$

To estimate β, we maximize the quasi-likelihood function by differentiating with respect to β, which yields

$$U(\beta) = D'V(\mu)^{-1}(Y - \mu) = 0, \tag{2.18}$$

where D is the matrix of partial derivatives $D_{ij} = \partial\mu_i/\partial\beta_j$ and $V(\mu)$ is the diagonal matrix with elements $(V(\mu_1), ..., V(\mu_n))$. For $V(\mu) = 1$, the quasi likelihood reduces to the least square method. This quasi-likelihood method is used to fit generalized linear models and generalized estimating equations for estimating regression coefficients.

Table 2.1: Characteristics of some common univariate exponential family distributions

	Normal	Poisson	Binomial	Gamma	Inverse Gaussian
Notation	$\mathcal{N}(\mu,\sigma^2)$	$P(\mu)$	$Bin(m,\pi)/m$	$G(\mu,v)$	$IG(\mu,\sigma^2)$
Range of y	$(-\infty,\infty)$	$0(1)\infty$	$\frac{0(1)m}{m}$	$(0,\infty)$	$(0,\infty)$
Dispersion parameter: ϕ	$\phi=\sigma^2$	1	$\frac{1}{m}$	$\phi=v^{-1}$	$\phi=\sigma^2$
Cumulant function: $b(\theta)$	$\theta^2/2$	$\exp(\theta)$	$\log(1+e^{\theta})$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$c(y;\phi)$	$-\frac{1}{2}(\frac{y^2}{\phi}+\log(2\pi\phi))$	$-\log(y!)$	$\log\binom{m}{my}$	$v\log(vy)-\log y-\log\Gamma(v)$	$-\frac{1}{2}\{\log(2\pi\phi y^3)+\frac{1}{\phi y}\}$
Canonical link: $\theta=g(\mu)$	μ	$\log(\mu)$	logit: $\log(\frac{\mu}{1-\mu})$ probit: $\Phi^{-1}(\mu)$ cloglog: $\log(-\log(1-\mu))$	$-\frac{1}{\mu}$	$1/\mu^2$
Inverse link: $\mu=g^{-1}(\theta)$	θ	$\exp(\theta)$	logit: $e^{\theta}/(1+e^{\theta})$ probit: $\Phi(\theta)$ cloglog: $1-\exp(-\exp(\theta))$	$-1/\theta$	$(-2\theta)^{-1/2}$
Variance function $V(y)$	1	μ	$\mu(1-\mu)$	μ^2	μ^3

The mean value parameter is denoted by μ, or by π for the binomial distribution.

The parameterization of the gamma distribution is such that its variance is μ^2/v.

The canonical parameter, denoted by θ, is defined by 2.12. The relationship between μ and θ is also given.

$\Phi(.)$ is the standard normal c.d.f.

Table 2.1 is adapted from McCullah and Nelder (1989).

2.3 Generalized Estimating Equations

The class of generalized linear models (GLM) plays a central role in regression problems which have discrete or continuous response variables. However they are based on the classical assumption that observations within a data set are independent. GLM were extended by Liang and Zeger (1986) such that longitudinal or correlated data could be analyzed, and this approach is known as the Generalized Estimating Equations (GEE) method. The method is called the GEE method because estimates are solutions of generalized estimating equations.

The GEE are called a population average or marginal method because GEE produce the average value of the individual regression lines for the regression coefficients. Hence, the regression coefficients estimated with GEE are called 'population average' (Diggle et al., 2002; Zeger and Liang, 1986). In marginal models, the primary interest of the analysis is to model the marginal expectation of the response variable given the covariates. In other words, for every one unit increase in a covariate across the population, the GEE tells us how much the average response would change (Zorn, 2001). For the binary covariates, the GEE calculates average response as the discrete change in the explanatory covariate from 0 to 1. The correlation, or more generally, the association between the response variables is modelled separately and is regarded as a nuisance parameter (Ziegler et al., 1996).

This method has received wide use in biological and medical applications such as epidemiology, gerontology, ecology, and biology (Ballinger, 2004; Akanda and Khanam, 2011), and is becoming increasingly popular in other disciplines such as organizational and psychological research. Much of the appeal of GEE is due to their broad capabilities, including: modelling correlated responses; allowing for time-varying covariates; and facilitating regression analysis on dependent variables that are not normally distributed (McCullagh and Nelder, 1989).

2.3.1 Notation

Consider a model of observations on a dependent variable Y_{ij} and a $p \times 1$ vector of covariates X_{ij}, observed at times $j = 1, 2, ..., m$ for individuals $i = 1, 2, ..., N$. Thus the number of clusters or individuals is N. Let $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{im})'$ be the $m \times 1$ vector of outcome values and $X_i = (X_{i1}, X_{i2}, ..., X_{im})'$ be the $m \times p$ matrix of covariate values for the i^{th} individual ($i = 1, 2, ..., N$). In the case of capture-recapture data, to correspond with the notation described above, the following notation is applied:

(i) Each animal forms one unit or cluster. Therefore, if only one animal is examined, $N = 1$. If two animals are examined, then $N = 2$. Thus, $i = 1$ for one animal and $i = 1, 2$ for two animals.

(ii) The response variable, Y_{ij}, is the capture history (0 or 1) where, 1 indicates capture and 0 otherwise. Thus, if one animal is examined, the response variable becomes Y_{1j}. If two animals are examined, there are two response vectors of Y_{1j} and Y_{2j}.

(iii) The observed time, j, corresponds with the time values at which the capture history is measured. For example $j = 1, 2, 3$ would correspond with capture-recapture experiments taken at time point 1, time point 2, and time point 3.

(iv) The equal number of time points m for all animals is considered.

2.3.2 Assumptions

There are four assumptions about the use of GEE to model correlated data that need to be articulated.

(i) The most crucial assumption is that the following conditional expectation needs to be specified correctly,

$$\mu_{ij} = E[Y_{ij} \mid X_{ij}] = E[Y_{ij} \mid X_i] \tag{2.19}$$

Equation (2.19) implies the conditional mean μ_{ij} of Y_{ij}, given the explanatory covariate X_i, measured at all possible time points m, is equal to a set of the same point specific explanatory variables X_{ij} (Dahmen and Ziegler, 2003).

(ii) The second assumption is that the response variable Y_{ij} should have a mean and variance which are characterised by a GLM (Equation 2.12).

(iii) It is further assumed that a true conditional $m \times m$ covariance matrix exists (Dahmen and Ziegler, 2003).

(iv) Finally, it is imperative that any missing data is missing completely at random (MCAR), otherwise results become inconsistent (Dobson and Williams, 2003).

2.3.3 Generalized Estimating Equations Modelling

To model the relation between the response and covariates, one can use a regression model similar to the generalized linear models, $g(\mu_i) = X_i\beta$, where $\mu_i = E(Y_i|X_i)$, g is a specified link function, and $\beta = (\beta_1, \beta_2, ..., \beta_p)'$ is a vector of unknown regression coefficients to be estimated.

The GEE approach estimates β through solving the following estimating equations (Liang and Zeger, 1986):

$$U(\beta) = \sum_{i=1}^{N} D_i' V_i^{-1}(Y_i - \mu_i) = 0 \tag{2.20}$$

where $D_i = \partial\mu_i(\beta)/\partial\beta'$ and V_i is a working variance-covariance matrix of Y_i. V_i can be expressed in terms of a working correlation matrix $R = R(\alpha)$ which represents the average dependency among capture occasions, $V_i = A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$, where A_i is a diagonal matrix with elements $\mathrm{Var}(Y_{ij}) = V(\mu_{ij})$, which is specified as a function of the mean μ_{ij}. The α may be some unknown parameters involved in the working correlation structure, which can be estimated through the methods of moments or another set of estimating equations. GEE permits a set of working correlation structures. A broad range of options available for specifying the correlation structure is an advantage for using the GEE approach. Some details (working correlation matrix, estimation of α) are discussed in the next chapter.

To estimate β, the GEE estimator (Equation 2.20) is rearranged to obtain the following,

$$\hat{\beta} = \sum_{i=1}^{N} (D_i' \hat{V}_i^{-1} D_i')^{-1} \sum_{i=1}^{N} (D_i' \hat{V}_i^{-1} Y_i). \tag{2.21}$$

As generalized estimating equations are not a likelihood-based method of estimation, computations based on likelihoods are not possible. Thus, in order to find a solution for Equation (2.21), estimation may be accomplished either via generalized weighted least-squares, or through an iterative process (Zorn, 2001).

2.3.4 Fitting GEE Models

One should fulfil some requirements when fitting a GEE model. Details are given below to accurately specify these conditions step by step.

Step 1 & 2: Linear predictor and best link function

To model the expected value of the marginal response for the population $\mu_i = E(Y_i)$ to a linear combination of the covariates, one must specify a link transformation function that will allow the response variable to be expressed as a vector of parameter estimates (β) in the form of an additive model (McCullagh and Nelder, 1989). The choices available for the link function depend primarily on the distribution specified, and a list of these available with GEE models can be seen in Table (2.2). This table gives the distributions and corresponding link functions currently available with GEE models in most statistical packages.

Table 2.2: Distribution choices and link functions in generalized estimating equations (GEE)

Distribution	Link Functions	Brief Description
Normal	Identity Link	This fits the same model as the glm
	Power Link	Any power transformation
	Reciprocal Link	Links using reciprocal of response variable
Binomial	Logit Link	Fits logistic regression models
	Probit Link	Fits cumulative probability functions
	Power Link	Any power transformation
	Reciprocal Link	Links using reciprocal of response variable
Poisson	Log Link	
	Power Link	Any power transformation
	Reciprocal Link	Links using reciprocal of response variable
Negative Binomial	Power Link	Any power transformation
Gamma	Power Link	Any power transformation
	Reciprocal Link	Links using reciprocal of response
Multinomial	Cumulative Logit Link	

Table 2.2 is adapted from Ballinger (2004).

Step 3: Distribution of the response variable

The next step involves specifying the distribution of the outcome variable such that the variance might be calculated as a function of the mean response calculated in step 1 and 2 (Hardin and Hilbe, 2003). GEE, like GLM, permit the specification of distributions from the exponential family of distributions, including the Normal, inverse Normal, binomial, Poisson, negative binomial, and gamma distributions. The GEE methodology is applied to the correlated capture-recapture data. Misspecification of the variance function, and thus the response distribution, can have important consequences and lead to incorrect statistical conclusions.

In fitting a GEE (or any GLM), one should make every reasonable effort to correctly specify the distribution of the response variable so that the variance can be efficiently calculated as a function of the mean and the regression coefficients can be properly interpreted (Ballinger, 2004). It is usual for the user to have some prior knowledge of the distribution of the response variable.

Step 4: Structure of the correlation within the response variable

The final step involves the specification of the form of the correlation of responses within units or nested within a group in the sample. Even though GEE models are generally robust to misspecification of the correlation structure (Liang and Zeger, 1986), it is still important that the user takes precautions in specifying this structure. This is because a structure that does not incorporate all of the information on the correlation of measurements within the cluster may result in inefficient estimators. Some correlation structures described in the next chapter.

Step 5: Fitting the model and diagnostics

A GEE model can be fitted to the data, however this usually takes considerable time and effort. Finally, and often most importantly, the model should be checked to see if it is adequate and justifiable, using numerous diagnostic techniques.

2.3.5 Remarks on GEE

The major, and most obvious advantage of GEE is they can be used to model non-Normal, correlated longitudinal data. This advantage is further strengthened by the broad range of options available that help specify the correlation between observations through the working correlation matrix. Another advantage is that even if an incorrect working correlation matrix is specified, it is still possible to obtain consistent parameter estimates for β that are asymptotically Normally distributed, provided the mean μ_i has been correctly specified as a function of all possible explanatory covariates X_i (Dahmen and Ziegler, 2003). GEE are gaining popularity, however there is some evidence that the use of an incorrect dependence structure within the GEE approach can produce worse results than if using an independent structure to model correlated data (Sutradhar and Das, 1999, Crowder, 1995). Horton and Lipsitz (1999) suggest that the GEE robust variance estimate should only be used when there are more than 20 units or clusters, that is, N should be greater than 20. If a data set contains fewer than 20 units, the model-based or naive approach to estimating the variance should be used, as it gives better estimates for the variance of β. Recently the GEE algorithm has been incorporated into many major statistical packages, including SAS, STATA, HLM, LIMDEP, GAUSS, SUDANN, R, and S-Plus.

2.4 Generalized Linear Mixed Models

Generalized linear mixed models (GLMM) (Breslow and Clayton, 1993) is an extension of the classes of linear mixed models (LMM) (Laird and Ware, 1982), and generalized linear models (GLM) (McCullagh and Nelder, 1982) in which the linear predictor contains random effects in addition to the fixed effects. The class of GLMM is a broad class of flexible statistical models that simultaneously enable (i) the modelling of non-normally distributed data, (ii) the specification of a linear or even non-linear link function between the mean of the response, and the predictor, and (iii) the accommodation of over-dispersion (McGilchrist, 1994) and correlation (Davis et al., 2000) by incorporating random effects.

2.4.1 Modelling Structure

Consider a sample of n independent random multivariate response $Y_i = (Y_{i1}, ..., Y_{im})'$, $i = 1, 2, ..., n$, where Y_{ij} is the j^{th} response ($j = 1, 2, ..., m$) to the i^{th} individual. We shall assume that Y_{ij} depends on a $p \times 1$ vector of fixed covariates x_{ij} associated with a vector of fixed effects $\beta = (\beta_1, ..., \beta_p)'$ and on a $q \times 1$ vector of fixed covariate z_{ij} associated with the multivariate $q \times 1$ random effect b_i. The GLMM satisfy the following conditions.

- Given b_i, the outcome variables $Y_{i1}, ..., Y_{im}$ are mutually independent with a density function belonging to the exponential family

$$f(Y_{ij}|b_i, \beta) = \exp\left\{ \frac{Y_{ij}\theta_{ij} - \alpha(\theta_{ij})}{d_{ij}(\phi)} + c(Y_{ij}, \phi) \right\} \tag{2.22}$$

  where θ_{ij} is the canonical parameter and ϕ is the scale parameter. The functions d_{ij} and c are specific to each distribution.

- The conditional mean and the conditional variance of Y_{ij} are given by

$$\begin{aligned} E(Y_{ij}|b_i) &= \mu_{ij}^{b_i} = h^{-1}(x'_{ij}\beta + z'_{ij}b_i), \\ \mathrm{Var}(Y_{ij}|b_i) &= v(\mu_{ij}^{b_i})d_{ij}(\phi), \end{aligned} \tag{2.23}$$

  where h and v are, respectively, the link and the variance function.

- The random effects $b_1, ..., b_n$ are mutually independent with a common underlying distribution G which depends on the unknown parameters α.

The likelihood can be considered as the joint distribution of the response and the random effects. To estimate β and α, the marginal likelihood of the response is obtained by integrating out the random effects. The integral can be approximated by penalized quasi-likelihood (Breslow and Clayon, 1993) which enables parameter estimation via an iterative procedure. GLMM are also very commonly used in ecological studies and they have well developed theoretical properties.

2.4.2 Modelling Correlation and Over-dispersion

The major aspect of random effects in GLMM is to accommodate the correlation existing among observations in the non-normally distributed data. In practice, correlation arises in a variety of contexts with dependent data. These type of data are always encountered in repeated measures designs where responses are measured repeatedly from the same experimental unit; in longitudinal studies where repeated measurements are made on the same individual across time; in hierarchical designs where clustering occurs on more than one level and the clusters are hierarchically nested; and in a wide range of studies employing cluster sampling. Depending on the design of dependent data collected, the correlation modelled by the random effects have variant interpretations. For example in cluster sampling, the random effects model the correlation between observations on the associated individuals within a cluster; while in repeated measure design, the random effects are responsible for describing the dependence among repeated observations on the same experimental unit. In the case that correlated random effects are included in GLMM, the association among the observations as well as among the random effects in the linear structure can be accommodated simultaneously.

By the term ‘over-dispersion’, it means that the actual variance of the observed response is larger than that explained by the nominal variance of the model. This phenomenon is not uncommon in practice. This often occurs when the unobserved heterogeneity, which arises from some unobservable covariate effects varying from one individual to another, has not been taken into account by the linear predictor. Sometimes, this could also be due to the inadequate specification of the model such as some relevant explanatory covariates have been omitted, or the link function is misspecified. Another possible situation of over-dispersion is that the data collected contain a lot of zeros. This phenomenon may not be easily incorporated in standard regression models unless some parameters can be specified to accommodate the situation. With the use of random effects, GLMM are capable to describe such situations in the sense that random effects act as the subject-specific variables to model or attribute over-dispersion to a particular source, and hence GLMM can be widely applicable to analyze non-normally distributed data with over-dispersion, particularly for those distributed according to the binomial or Poisson distributions.

To conclude, the random effects in the linear predictor of GLMM are useful in accounting for the correlation, over-dispersion, dependence, subject-specific inference, heteroscedasticity or any combinations of the above observed in the data. With their wide applicability and practical importance, GLMM became an extremely popular modelling tool and received considerable attention for analyzing the correlated and over-dispersed data in recent years.

2.5 Partial and Conditional Likelihood

Partial likelihood (Cox, 1975; Wong, 1986) arises from partitioning the full likelihood into products of conditional densities and only considering some of these product terms, isolating nuisance parameters. This partial likelihood may yield a much simpler form than the full likelihood (and still retain many of its properties due to the usual conditions on the conditional densities), which allows for more flexible and convenient inference.

Consider a vector y of observations represented by a random variable Y having density $f_Y(y;\theta)$ and suppose that Y is transformed to new random variables

(V, W) by a transformation not depending on the unknown parameter. We call $f_Y(v;\theta)$ the marginal likelihood based on V and $f_{W|V}(w|v;\theta)$ the conditional likelihood based on W given $V = v$, both being considered as functions of θ; the usual notation for probability densities is used. These definitions are both special cases of the following definition of partial likelihood. Suppose that the random variable Y, can be transformed into the sequence

$$(U_1, S_1, U_2, S_2, ..., U_m, S_m), \tag{2.24}$$

where the components may themselves be vectors. The number of pairs of terms may in some cases be random; alternatively we can imagine dummy variables added to complete the sequence up to the maximum conceivable m. The full likelihood of the sequence (2.24) is given by

$$\begin{aligned} f_Y(y;\theta) = & \prod_{j=1}^{m} f_{S_j|U^{(j)},S^{(j-1)}}\left(s_j|u^{(j)}, s^{(j-1)};\theta\right) \\ & \times \prod_{j=1}^{m} f_{U_j|U^{(j-1)},S^{(j-1)}}\left(u_j|u^{(j-1)}, s^{(j-1)};\theta\right), \end{aligned} \tag{2.25}$$

where $u^{(j)} = (u_1, ..., u_j)$ and $s^{(j)} = (s_1, ..., s_j)$. The first product of (2.25) is called the partial likelihood based on S in the sequence $\{U_j, S_j\}$. Partial likelihood was originally applied to survival data, however this approach is also appropriate for capture-recapture studies. The conditional likelihood based on W given $V = v$ corresponds to the special case

$$S_1 = W, \;\; U_1 = V$$

and the marginal likelihood of V to the special case in which U_1 is a known constant, $S_1 = V$ and $U_2 = W$. Both marginal and conditional likelihoods are in a natural sense ordinary likelihoods for derived experiments, but the same is not true in general for partial likelihood; this is because of the way the conditioning events change (Cox, 1975).

If the full likelihood contains nuisance parameters as well as the parameters of interest, then a conditional likelihood (Kalbfleisch and Sprott, 1970, 1973) may be considered, such that the nuisance parameters are eliminated, and inference about the parameters of interests can be made. Suppose now that y has density $f_Y(y;\theta,\psi)$ where θ is the parameter of interest and ψ is the nuisance parameter. Both θ and ψ may be considered as vectors, but for simplicity consider the one dimensional case. As in Kalbfleisch and Sprott (1973), suppose that for given value of θ, $T_\theta = (y_1, ..., y_n, \theta)'$ is sufficient for ψ, then the full likelihood is given by

$$f_Y(y;\theta,\psi) \propto f_Y(y;\theta|T_\theta) \times f_{T_\theta}(T_\theta;\theta,\psi).$$

The first product is called the conditional likelihood of θ. Note that it does not involve ψ, furthermore the second product contains no available information concerning θ in the absence of knowledge of ψ. The conditional likelihood is readily used in practice and can be applied to common model forms, such as GLMM (Sartori and Severini, 2004).

Chapter 3

A Generalized Estimating Equations Approach for Capture-Recapture Closed Population Models

---

Part of this chapter is in the form of publication and conference presented paper:

Akanda, M.A.S. and Alpizar-Jara, R. (2014). ‘A Generalized Estimating Equation Approach for Capture-Recapture Closed Population Models.’ *Environmental and Ecological Statistics*, DOI: 10.1007/s10651-014-0274-7.

Akanda, M.A.S. and Alpizar-Jara, R. (2012). ‘A Generalized Estimating Equation Approach for Estimating Population Size by Using Capture- Recapture Data.’ *Abstract of International Statistical Ecology Conference 2012*, 3–6 July 2012, Sundvolden Hotel, Oslo, Norway, Abstract no. 125.

The estimation of population parameters, such as capture probability, population size, or population density, is an important issue in many ecological applications. In this chapter, we propose a generalized estimating equations (GEE) approach to account for correlation over time instead of assuming independence as in the traditional closed population capture-recapture studies. Current modelling strategies involve a logit link-function of the covariates through a generalized linear modelling (GLM) framework. We also account for heterogeneity due to measurable individual covariates and over-dispersion, modelling capture probabilities as a function of individual covariates. The GEE versions of all closed population capture-recapture models and their corresponding estimating equations are proposed. We evaluate the effect of accounting for correlation structures on capture-recapture model selection based on the Quasi-likelihood Information Criterion (QIC). An example is used for an illustrative application and for comparison to currently used methodology. A Horvitz-Thompson-like estimator is used to obtain estimates of population size based on conditional arguments. A simulation study is conducted to evaluate the performance of the GEE approach in capture-recapture studies.

3.1 Background

There is a vast body of literature on the statistical methodology of capture-recapture studies, with much of it involved with building realistic models of the sampling processes. Otis et al. (1978) presented a detailed account of traditional closed population capture-recapture models. Seber (1986, 1992, 2002), Williams et al. (2002), Chao and Huggins (2005) discussed many capture-recapture techniques and associated statistical models. Pollock et al. (1984) suggested linear logistic regression modelling of capture probabilities using measurable time dependent environmental and other auxiliary variables (such as effort) without considering dependencies among capture occasions. Their approach is very straightforward, and it is now widely used to model capture probabilities in both closed and open population models. Huggins (1989) also uses data from capture-recapture experiments to estimate the size of a closed population when capture probabilities are heterogeneous, and modelled in terms of measurable covariates such as age, sex, weight, rainfall, etc. The modelling is based on a likelihood approach, conditional on the captured individuals, to estimate the associated parameters. Huggins (1991) extended his previous work, providing more details about model structures and considering capture probabilities in a linear logistic regression framework. Hwang and Huggins (2005) and Zhang (2012) introduced an estimating equation approach to examine the effect of heterogeneity on the estimation of population size using capture-recapture data by solving estimating equations assuming independence of capture occasions.

One may view capture-recapture data as binary longitudinal or repeated measurements since these are typically collected on the individuals across successive points in time (Huggins and Yip, 2001). These repeated observations are often correlated over time. Failure to account for this correlation provides biased parameter estimates. Hwang and Huggins (2007) also state that the assumption of independence among capture occasions is often violated in practice, but the authors still rely on that assumption. Bayesian methods are becoming popular in capture-recapture studies. An extensive Bayesian literature on capture-recapture closed population models includes Castledine (1981), Smith (1991), George and Robert (1992), Madigan and York (1997), Basu and Ebrahimi (2001), Ghosh and Norris (2005), King and Brooks (2008), Gosky and Ghosh (2009, 2011). Bayesian statistical modelling requires the development of a likelihood function for the observed data, given a set of parameters, as well as the joint prior distribution of all model parameters. Bayesian methods allow for estimation of the heterogeneity due to measurable and unmeasurable individual characteristics, but the performance of their estimates often depends on the chosen prior distributions. The method of selecting prior distributions is often too subjective (Lee et al., 2003).

The generalized estimating equations (GEE) method is an extension of the generalized linear modelling (GLM) framework of Nelder and Wedderburn (1972). The GEE approach is considered as one of the most important and widely applicable techniques in analyzing correlated binary outcomes. This approach, developed by Liang and Zeger (1986) and Zeger and Liang (1986), facilitates analysis of data collected in longitudinal, or repeated measures designs. The method has been widely used in medical and life sciences, such as epidemiology, ecology, gerontology, and biology. Generalized estimating equations use a generalized linear modelling framework to estimate more efficient and unbiased regression parameters relative to ordinary least squares regression. Moreover, GEE permit specification of a working correlation structure that accounts for a form of correlation among capture occasions.

Here we propose capture probabilities to be estimated using a GEE modelling approach, accounting for heterogeneity due to measurable individual characteristics and correlation over time instead of assuming independence. The GEE approach has also the ability to account for over-dispersion. The modelling strategies involve a logit link-function of the covariates through a generalized linear modelling (GLM) framework. A comparative advantage of GEE over Bayesian methods relates to the ability of GEE to allow specific correlation structures to be assumed within capture occasions. Eight standard closed population models exist, allowing for time, behavioral, and heterogeneity effects (Otis et al., 1978). The GEE versions of these models and their corresponding estimating equations are presented when capture probabilities depend on covariates. We also evaluate the effect of accounting for correlation structures on capture-recapture model selection based on the Quasi-likelihood Information Criterion (QIC) (Pan, 2001).

In Section 3.2 we describe GEE models considered in this work. Section 3.3 presents different types of capture-recapture closed population models within the

GEE framework. Model selection procedures are described and considered in Section 3.4. In Section 3.5, we fit GEE in capture-recapture methodology for different correlation structures and the estimation of the population size is examined. A simulation study is given in Section 3.6 to evaluate the performance of GEE in capture-recapture methodology. Finally, we provide some concluding remarks in Section 3.7.

3.2 Notation and Models

Suppose that the total number of individuals in the study population is N, and the possible number of capture occasions is $m(\geq 2)$. Let Y_{ij} be the indicator variable of the i^{th} individual being caught on the j^{th} occasion, that is, $Y_{ij} = 1$ if the i^{th} individual is captured on the j^{th} occasion, and Y_{ij}=0, otherwise. Let T_i be the number of occasions in which the i^{th} individual has been caught during the course of a experiment, then $T_i = \sum\limits_{i=1}^{m} Y_{ij}$. Let $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{im})'$ be the $m \times 1$ random vector to record the capture history of the i^{th} individual for the m capture occasions. Covariates z_i, x_{ij} and v_{ij} are associated with the i^{th} individual and the j^{th} occasion, for $i = 1, 2, ..., N$ and $j = 1, 2, ..., m$. Let z_i be an individual measurable covariate, such as age, sex or weight, etc. For simplicity, we consider z_i a single covariate, but the model can be easily generalized for z_i to be considered a vector of covariates. We also consider x_{ij} to be an measurable occasion related covariate for the i^{th} individual, and v_{ij} be an indicator of the capture of individual i^{th} prior to occasion j^{th}, which depends on occasions as well as individual, i.e.,

$$v_{ij} = \begin{cases} 1, & \text{if individual } i \text{ has been captured before } j \\ 0, & \text{elsewhere.} \end{cases}$$

Suppose that, P_{ij}, the probability that the i^{th} individual is captured on the j^{th} occasion is given by

$$P_{ij} = \Pr(Y_{ij} = 1 | z_i, x_{ij}, v_{ij}) = h(\beta_0 + \beta_1 z_i + \beta_{j+1} x_{ij} + \beta_b v_{ij}) \tag{3.1}$$

for $i = 1, 2, ..., N; j = 1, 2, ..., m$, where $h(u) = \exp(u)/(1 + \exp(u)) = (1 + \exp(-u))^{-1}$ is the logistic function and

$$X_i = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_i & z_i & \dots & z_i \\ x_{i1} & 0 & \dots & 0 \\ 0 & x_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_{im} \\ v_{i1} & v_{i2} & \dots & v_{im} \end{bmatrix}'$$

is the design matrix. The logit transform, $\ln[P_{ij}/(1-P_{ij})] = \beta_0 + \beta_1 z_i + \beta_{j+1} x_{ij} + \beta_b v_{ij}$ is a linear function of covariates corresponding to both, individual and environmental characteristics. This type of linear logistic regression model has been previously used by Pollock et al. (1984) and has been extensively used in the statistical literature (Cox, 1970). Note that model (3.1) is a more general version than those presented in Hwang and Huggins (2005) and Zhang (2012). This is because it is extended to include all possible effects in closed capture-recapture studies.

If $x_{i_1 j} = x_{i_2 j} = x_j$ for all $i_1 \neq i_2$ that is, individuals are captured assuming the same environmental characteristics (occasion related) for a particular occasion, those may vary only occasion to occasion, but do not depend on individual or other characteristics such as behavior. Hence, we could simplify (3.1) to

$$P_{ij} = \Pr(Y_{ij} = 1|z_i, x_j, v_{ij}) = h(\beta_0 + \beta_1 z_i + \beta_2 x_j + \beta_b v_{ij}) \tag{3.2}$$

for $i = 1, 2, ..., N; j = 1, 2, ..., m$.

In this case, the design matrix X_i has the following form

$$X_i = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_i & z_i & \dots & z_i \\ x_1 & x_2 & \dots & x_m \\ v_{i1} & v_{i2} & \dots & v_{im} \end{bmatrix}' .$$

The mean vector of Y_i is,

$$\mu_i = \begin{bmatrix} E(Y_{i1}) \\ E(Y_{i2}) \\ \vdots \\ E(Y_{im}) \end{bmatrix} = \begin{bmatrix} P_{i1} \\ P_{i2} \\ \vdots \\ P_{im} \end{bmatrix} = P_i,$$

where, $P_{ij} = \mu_{ij} = \Pr(Y_{ij} = 1|z_i, x_j, v_{ij}); j = 1, 2, ..., m; i = 1, 2, ..., N$. The probability of not capturing the i^{th} individual on the j^{th} occasion is $(1 - P_{ij})$, and the variance of Y_{ij} is $P_{ij}(1 - P_{ij}) = \mu_{ij}(1 - \mu_{ij})$.

Thus, the $m \times m$ variance-covariance matrix of Y_i is given by

$$\text{Var}(Y_i) = \begin{bmatrix} \text{Var}(Y_{i1}) & \text{Cov}(Y_{i1}, Y_{i2}) & \dots & \text{Cov}(Y_{i1}, Y_{im}) \\ \text{Cov}(Y_{i1}, Y_{i2}) & \text{Var}(Y_{i2}) & \dots & \text{Cov}(Y_{i2}, Y_{im}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(Y_{i1}, Y_{im}) & \text{Cov}(Y_{i2}, Y_{im}) & \dots & \text{Var}(Y_{im}) \end{bmatrix} .$$

In addition to the mean and covariance of the vector of responses, Y_{ij}, Liang and Zeger (1986) suggested to use a $m \times m$ working correlation matrix for each Y_i. The working correlation matrix (denoted by $R_i(\alpha)$) is an approximation to the actual correlation matrix of Y_i. While $R_i(\alpha)$ is assumed to be fully specified by the vector of unknown parameters α, its structure is determined by the investigator based on the study design, and it is often assumed to be constant across individuals. This working covariance matrix can be expressed in the following form:

$$V_i = A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}} \tag{3.3}$$

where, $A_i = \text{diag}[\text{Var}(Y_{i1}), \text{Var}(Y_{i2}), ..., \text{Var}(Y_{im})]$ is a $m \times m$ diagonal matrix with

$$A_i = \begin{bmatrix} \text{Var}(Y_{i1}) & 0 & \cdots & 0 \\ 0 & \text{Var}(Y_{i2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \text{Var}(Y_{im}) \end{bmatrix} = \phi \begin{bmatrix} \frac{\exp(u_{i1})}{(1+\exp(u_{i1}))^2} & 0 & \cdots & 0 \\ 0 & \frac{\exp(u_{i2})}{(1+\exp(u_{i2}))^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\exp(u_{im})}{(1+\exp(u_{im}))^2} \end{bmatrix},$$

ϕ is a known or estimated variance parameter (depending on the response distribution) to allow for over-dispersion, such that $\text{Var}(Y_{ij}) = \phi\text{Var}(P_{ij})$. In this model setting, $\phi = 1$ since the response variable follows a Bernoulli distribution. Moreover, we use logit link with continuous covariates (Carruthers et al., 2008; Hojsgaard and Halekoh, 2005). Hence, the model automatically accounts for over-dispersion.

Let the vector of parameters, $\beta = (\beta_0, \beta_1, \beta_2, \beta_b)'$ and $R_i(\alpha) = \text{Corr}(Y_i)$ is a $m \times m$ working correlation structure among $Y_{i1}, Y_{i2}, ..., Y_{im}$, which describes the average dependency of individuals being captured from occasion to occasion. For example, if individual captures are independent from occasion to occasion, then the $R_i(\alpha)$ has the form: $R_i(\alpha) = I$, where I is a $m \times m$ identity matrix.

Liang and Zeger (1986) showed that if the correlation matrices $R_i(\alpha)$ are correctly specified, the estimator $\hat{\beta}$ is consistent and asymptotically Normal. Furthermore, $\hat{\beta}$ is fairly robust against misspecification of $R_i(\alpha)$. They used the term working correlation matrix for $R_i(\alpha)$, and suggested that knowledge of the study design and results from exploratory analyses should be used to select a plausible form. Preferably, $R_i(\alpha)$ should depend on only a small number of parameters, using assumptions such as equicorrelation or autocorrelation. The most common structures used to model the working correlation matrix are the independent, exchangeable, autoregressive, stationary, nonstationary, unstructured, and fixed correlation structures. The broad range of options available for specifying the correlation structure is another advantage for using the GEE approach. Some of these structures are examined in more detail below.

(i) **Independent Structure:**

The independent structure is the simplest form that the working correlation matrix can take, as it assumes that no correlation actually exists and observations within the series are independent. With this structure, the working correlation matrix becomes the identity matrix, $R_i(\alpha) = I$, and the resulting GEE is then called the Independent Estimating Equations (Dahmen and Ziegler, 2003). No estimation of α is required, since no correlation is assumed to exist. This structure does not simply produce the algorithm used for a GLM, as it still involves the 'working' correlation matrix, which a GLM does not. For the independent structure, $R_i(\alpha)$ is defined as,

$$R_{j,k} = \begin{cases} 1, \ j = k \\ 0, \ \text{otherwise.} \end{cases}$$

In matrix notation this becomes,

$$R_i(\alpha) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

(ii) **Exchangeable Structure:**

The exchangeable structure assumes that there is a common correlation within observations. Thus, all of the correlations in $R_i(\alpha)$ are equal (Hedeker, 2005). An exchangeable correlation may be used when each pair of observations within a time frame has approximately the same correlation. For the exchangeable structure, $R_i(\alpha)$ is defined as,

$$R_{j,k} = \begin{cases} 1, \; j = k \\ \alpha, \text{ otherwise.} \end{cases}$$

In matrix notation this becomes,

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \dots & 1 \end{bmatrix}.$$

(iii) **Autoregressive Structure:**

For data that are correlated within cluster over time, an autoregressive correlation structure is specified to set the time correlations as an exponential function of this lag period, which is determined by the user (Ballinger, 2004). The autoregressive structure assumes time dependence for the association between observations and considers each time series to be an AR(m) process. The most difficult task for this structure is determining the correct order of the autoregressive process (Hardin and Hilbe, 2003). It is common to choose an AR(1) structure, which is defined as,

$$R_{j,k} = \begin{cases} 1, \; j = k \\ \alpha^{|j-k|}, \text{ otherwise.} \end{cases}$$

In matrix notation this becomes,

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{n-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n-1} & \alpha^{n-2} & \alpha^{n-3} & \dots & 1 \end{bmatrix}.$$

(iv) **Unstructured structure:**

The unstructured form of the working correlation matrix is the most general of all of the correlations discussed in this dissertation as no structure is imposed on the correlation matrix. This form requires all $m(m-1)/N$ correlations of $R_i(\alpha)$ to be estimated, and thus when there are many time points this structure becomes very computationally burdensome.

An unstructured correlation matrix is used when there is no logical ordering for the observations in the cluster, and is recommended if the number of observations is small in a balanced and complete design (Horton and Lipsitz, 1999). This correlation matrix is the most efficient structure, but is only useful when there are relatively few observations as its estimate is not guaranteed to be a positive number and there is often a problem with inverting $R_i(\alpha)$ (Hedeker, 2005). For the unstructured structure, $R_i(\alpha)$ is defined as,

$$R_{j,k} = \begin{cases} 1, \ j = k \\ \alpha_{jk}, \ \text{otherwise.} \end{cases}$$

In matrix notation this becomes,

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha_{12} & \dots & \alpha_{1m} \\ \alpha_{21} & 1 & \dots & \alpha_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \dots & 1 \end{bmatrix}.$$

Following the notation in Dobson (2010), let D_i be the matrix of derivatives $\partial\mu_i/\partial\beta'$. If the z_i, x_j and v_{ij} are observable covariates for each captured individual, the vector of parameters $\beta = (\beta_0, \beta_1, \beta_2, \beta_b)'$ for model (3.2) could be estimated by solving the following generalized estimation equations:

$$U(\beta) = \sum_{i=1}^{N} D_i' V_i^{-1}(Y_i - \mu_i) = 0. \tag{3.4}$$

However, the number of total individuals, N, is unknown and our purpose is to estimate it. Also, the covariates z_i, x_j and v_{ij} are not known for individuals that are never captured. To accommodate these situations, equation (3.4) has to be modified as in Zhang (2012). Let Ψ be a set of distinct individuals captured in at least one occasion that are indexed by $i = 1, 2, ..., n$ and the uncaptured individuals would be indexed by $i = n + 1, ..., N$ without lost of generality. Thus, Y_{ij} is conditional on the captured individuals (n) (i.e., $T_i \geq 1$) like Huggins (1989) and Zhang (2012).

Therefore, an estimator of β, the vector of parameters defined by model (3.2), can be obtained by solving the following generalized estimating equations:

$$U(\beta) \equiv \sum_{i=1}^{n} D_i' V_i^{-1}\big(Y_i - \mu_i\big) = 0. \tag{3.5}$$

We may call these estimating equations as a transformed GEE for the capture-recapture models (GEECR). To solve equation (3.5), we consider the mean vector,

$\mu_i = (\mu_{i1}, ..., \mu_{im})'$ and the variance-covariance matrix V_i. The variance-covariance matrix varies according to the nature of correlation structures $R_i(\alpha)$, as described earlier. We used iterative procedures, e.g., Fisher scoring (McCullagh and Nelder, 1989) to handle convergence problems associated with the fully iterated GEE. After solving for β, the standardized residuals are calculated to estimate α.

Specifically, if we consider the independent working correlation structure ($R(\alpha) = I$), this is assuming that $Y_{i1}, Y_{i2}, ..., Y_{im}$ are independent and letting $D_i = A_i X_i$, then the working covariance matrix (3.3) becomes

$$V_i = A_i^{\frac{1}{2}} I A_i^{\frac{1}{2}} = A = \begin{bmatrix} \mathrm{Var}(Y_{i1}) & 0 & \dots & 0 \\ 0 & \mathrm{Var}(Y_{i2}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathrm{Var}(Y_{im}) \end{bmatrix}.$$

Hence, the vector of parameters, β can be estimated by solving the following GEECR equations,

$$\sum_{i=1}^{n} X_i' \big(Y_i - \mu_i\big) = 0. \tag{3.6}$$

These estimating equations are the same that are obtained by assuming that the observations of each individual are independent of each other (Hwang and Huggins, 2005).

We can estimate the vector of parameters β by solving the following GEECR equations for the exchangeable and autoregressive correlation structures

$$\sum_{i=1}^{n} X_i' A_i V_i^{-1} \big(Y_i - \mu_i\big) = 0 \tag{3.7}$$

where, the covariance matrix (V_i) for Y_i changes according to the assumption of exchangeable and autoregressive working correlation structures within the repeated responses.

Equations (3.7) can be solved by an iterative procedure, and at each iteration, the correlation parameter α can be estimated from the current Pearson's residuals defined by

$$\hat{r}_{ij} = \frac{Y_{ij} - \hat{\mu}_{ij}}{[\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})]^{1/2}},$$

where $\hat{\mu}_{ij}$ depends upon the current value for β. Liang and Zeger (1986) estimated α as following,

$$\hat{\alpha} = \sum_{i=1}^{n} \sum_{j>l}^{m} \hat{r}_{ij}\hat{r}_{il} \Big/ \left\{ \frac{1}{2} nm(m-1) - p \right\}$$

where p is the number of parameters in the model.

The vector of parameters, β can be estimated by solving the following GEECR equations under pairwise correlation structure,

$$\sum_{i=1}^{n} X_i' A_i \left(A_i^{\frac{1}{2}} \hat{R}(\alpha) A_i^{\frac{1}{2}} \right)^{-1} \left(Y_i - \mu_i \right) = 0. \tag{3.8}$$

where, according to Liang and Zeger (1986), $R(\alpha)$ can be estimated by

$$\hat{R}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} A_i^{-1/2} \big(Y_i - \hat{\mu}_i\big) \big(Y_i - \hat{\mu}_i\big)' A_i^{-1/2}.$$

3.2.1 Asymptotic Properties of $\hat{\beta}$

The GEE model outlined here has a number of attractive properties for applied researchers. Because the first two terms of (3.5) do not depend on Y_i, the score equations converge to zero (and thus have consistent roots) so long as $E(Y_i - \mu_i) = 0$. Assuming that the model for μ is correctly specified, GEE estimates of β (here, $\hat{\beta}_{GEE}$) will be consistent in n. Moreover, $n^{1/2}(\hat{\beta}_{GEE} - \beta)$ is asymptotically multivariate normal, and the covariance matrix of the estimates can be consistently estimated by the inverse of the derivative of the quasi-score function, evaluated at $\hat{\alpha}$ and $\hat{\beta}$ following the theorem given by Liang and Zeger (1986). The proof of the theorem is reported in Appendix A.

Theorem 1: Under mild regularity conditions and given that:

1). $\hat{\alpha}$ is $n^{\frac{1}{2}}$-consistent given β and ϕ;

2). $\hat{\phi}$ is $n^{\frac{1}{2}}$-consistent given β; and

3). $\mid \partial\hat{\alpha}(\beta,\phi)/\partial\phi \mid \leq H(Y,\beta)$ which is $O_p(1)$, then $n^{1/2}(\hat{\beta}_{GEE}-\beta)$ is asymptotically multivariate Gaussian with zero mean and covariance matrix V_{GEE} given by

$$\mathrm{Var}(\hat{\beta}_{GEE}) = \lim_{n\to\infty} n\Big(\sum_{i=1}^{n} D_i' V_i^{-1} D_i\Big)^{-1}\Big[\sum_{i=1}^{n} D_i' V_i^{-1}\mathrm{Cov}(Y_i) V_i^{-1} D_i\Big]\Big(\sum_{i=1}^{n} D_i' V_i^{-1} D_i\Big)^{-1}.$$

Most important, the asymptotic consistency of $\hat{\beta}_{GEE}$ holds even in the presence of misspecification of the 'working' correlation structure α; thus, GEE offer the potential of providing asymptotically unbiased estimates of the parameters of primary interest even in cases where the exact nature of the intracluster dependence is unknown. Note, however, that the consistency of the variance estimate for $\hat{\beta}$ does depend on proper specification of the working correlation structure; misspecification of the working correlation structure yields estimates of $\hat{\beta}$ which are still asymptotically normal, but for which $\mathrm{Var}(\hat{\beta})$ does not equal $(D'V^{-1}D)^{-1}$. In such cases, Liang and Zeger (1986) propose a 'robust' estimate of the variance-covariance matrix of $\hat{\beta}$:

$$\mathrm{Var}(\hat{\beta}_{GEE}) = n\Big(\sum_{i=1}^{n} \hat{D}_i' \hat{V}_i^{-1} \hat{D}_i\Big)^{-1}\Big[\sum_{i=1}^{n} \hat{D}_i' \hat{V}_i^{-1} S_i \hat{V}_i^{-1} \hat{D}_i\Big]\Big(\sum_{i=1}^{n} \hat{D}_i' \hat{V}_i^{-1} \hat{D}_i\Big)^{-1} \qquad (3.9)$$

where $S_i = (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)'$ is a simple empirical covariance estimate. This 'robust' (or 'empirically-corrected') estimate is analogous to that derived by White (1982) and is consistent even under misspecification of the correlation matrix. In practice, this robust estimator is nearly always used, since a key reason for using GEE in the first place is the belief that the observations in the data are dependent.

3.2.2 Estimation of Population Size and their Variance

Derived GEE estimates for β, can be used to estimate the capture probabilities, P_{ij}. To estimate the population size we follow the method of Huggins (1989) which is based on a Horvitz-Thompson estimator (1952). Let $\pi_i(\beta) = \Pr(T_i \geq 1|X_i) = 1 - \prod_{j=1}^{m}(1-P_{ij})$ be the probability of being captured at least once given the covariates in the course of the trapping experiment, where $T_i = \sum_{j=1}^{m} Y_{ij}$, as defined earlier. An unbiased estimate of the population size N is

$$\hat{N}(\beta) = \sum_{i=1}^{n} \pi_i(\beta)^{-1}, \tag{3.10}$$

and an estimate of the variance of $\hat{N}$ is given by

$$\widehat{\mathrm{Var}}(\hat{N}) = \sum_{i=1}^{n} \pi_i(\beta)^{-2}\big(1 - \pi_i(\beta)\big) + \Delta(\beta)' \mathit{\Gamma}(\beta)^{-1} \Delta(\beta) \tag{3.11}$$

where $\mathit{\Gamma}(\beta)$ represents the conditional information matrix for β and $\Delta(\beta)$ is the vector of $\sum_{i=1}^{n} \pi_i(\beta)^{-2} \partial \pi_i(\beta)/\partial \beta$ with all quantities evaluated at $\hat{\beta}$.

3.3 Different Kinds of Capture-Recapture Closed Models

The most general closed population capture-recapture model is denoted by $\mathrm{M_{tbh}}$, considered by Otis et al. (1978). The subscripts t, h, and b refer to time, heterogeneity, and behavioral effects, respectively, indicating which effects are present in the modelling of capture probabilities. The specification of a closed population capture-recapture model is given by the restrictions imposed on the capture probability parameters. These capture probabilities depend on the vector of parameters β. The general estimating equations of closed population models are equations (3.7) for exchangeable and autoregressive correlation structures, and equations (3.8) for pairwise correlation structure. As in Huggins (1991), we could consider a class of models related to those of Otis et al. (1978) when appropriate, but rather than the usual likelihood approach and AIC, we use GEE estimation and quasi-likelihood methods considering variance structure under independence. The GEE versions of the class of models and their corresponding estimating equations are presented below. To our knowledge, with the exception of estimating equations for models $\mathrm{M_0}$ and $\mathrm{M_h}$, none of them have been previously proposed.

Model 1. ($\mathrm{M_0}$)

If the individual capture probability does not vary with time, the individual's prior capture history, or any covariates, model (3.1) simplifies to $\ln[P_{ij}/(1-P_{ij})] = \beta_0; j = 1, 2, ..., m, i = 1, 2, ..., N$. This model assumes that all the individuals are equally at risk of capture on every trapping occasion. This model is a re-parameterization of model ($\mathrm{M_0}$) in Otis et al. (1978). Suppose, $P_{ij} = h(\beta_0) = P_0$, then equations (3.5) becomes the result of the homogeneity assumption (Hwang and Huggins, 2005; Zhang, 2012), and is given by:

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{m} Y_{ij} - \frac{mP_0}{1-(1-P_0)^m}\right) = 0. \tag{3.12}$$

Model 2. ($\mathrm{M_b}$)

$$\ln[P_{ij}/(1-P_{ij})] = \beta_0 + \beta_b v_{ij}; j = 1, 2, ..., m, i = 1, 2, ..., N.$$

This model allows the capture probabilities to vary only according to an individual's prior capture history and is a re-parameterization of model $\mathrm{M_b}$ of Otis et al. (1978). Let, $P_{ij} = h(\beta_0 + \beta_{ij} v_{ij})$ then equations (3.5) can be written as the following classes of estimating equations,

$$\begin{cases} \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}(Y_{ij} - \mu_{ij}) = 0 \\ \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}(Y_{ij} - \mu_{ij})v_{ij} = 0. \end{cases} \tag{3.13}$$

Model 3. ($\mathrm{M_t}$)

$$\ln[P_{ij}/(1-P_{ij})] = \beta_0 + \beta_{j+1}x_j; j = 1, 2, ..., m, i = 1, 2, ..., N,$$

where $\beta_{m+1} = 0$ for identifiability. This model allows that all animals have the same probability of capture on any particular trapping occasion, but that probability can change from one occasion to the next, that is, capture probabilities vary only by time. In this case, $P_{ij} = h(\beta_0 + \beta_{j+1}x_j) = P_j$, then equations (3.5) can be written as the following estimating equations,

$$\begin{cases} \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}(Y_{ij} - \mu_{ij}) = 0 \\ \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}(Y_{ij} - \mu_{ij})x_j = 0. \end{cases} \tag{3.14}$$

Model 4. (special case of $\mathrm{M_h}$)

$$\ln[P_{ij}/(1-P_{ij})] = \beta_0 + \beta_1 z_i; j = 1, 2, ..., m, i = 1, 2, ..., N,$$

where β_1 is the effect of individual covariate. This model accounts for heterogeneity among the capture probabilities of individuals only, but there is no difference in capture probabilities among trapping occasions and no behavioural response to capture. There is no equivalent model of Otis et al. (1978) although this model is a restricted version of their model $\mathrm{M_h}$. If $P_{ij} = h(\beta_0 + \beta_1 z_i) = P_i$, then equations (3.5) can be written as

$$\begin{cases} \sum\limits_{i=1}^{n}\left(T_i - \frac{mP_i}{1-(1-P_i)^m}\right) = 0 \\ \sum\limits_{i=1}^{n}\left(T_i - \frac{mP_i}{1-(1-P_i)^m}\right)z_i = 0. \end{cases} \tag{3.15}$$

Equation (3.15) has been previously considered by Hwang and Huggins (2005).

Model 5. ($\mathrm{M_{tb}}$)

$$\ln[P_{ij}/(1-P_{ij})] = \beta_0 + \beta_{j+1}x_j + \beta_b v_{ij}; j = 1, 2, ..., m, i = 1, 2, ..., N,$$

where $\beta_{m+1} = 0$ and β_b is the effect of individual's prior capture history. This model allows that the probability of capture changes after initial capture, and that temporal changes also affect capture probabilities. Suppose, $P_{ij} = h(\beta_0 + \beta_{j+1}x_j + \beta_b v_{ij})$ then the set of estimating equations from equations (3.5) are given below:

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij}) = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})x_j = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})v_{ij} = 0. \end{cases} \tag{3.16}$$

Model 6. (special case of $\mathrm{M_{bh}}$)

$$\ln[P_{ij}/(1 - P_{ij})] = \beta_0 + \beta_1 z_i + \beta_b v_{ij}; j = 1, 2, ..., m, i = 1, 2, ..., N.$$

In this case, the capture probabilities depend on both behavioural response to first capture and individual heterogeneity. There is no equivalent model of Otis et al. (1978), although this model is a restricted version of their model $\mathrm{M_{bh}}$. Let, $P_{ij} = h(\beta_0 + \beta_1 z_i + \beta_b v_{ij})$ then equations (3.5) can be written as

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij}) = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})z_i = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})v_{ij} = 0. \end{cases} \tag{3.17}$$

Model 7. (special case of $\mathrm{M_{th}}$)

$$\ln[P_{ij}/(1 - P_{ij})] = \beta_0 + \beta_1 z_i + \beta_{j+1} x_j; j = 1, 2, ..., m, i = 1, 2, ..., N,$$

where $\beta_{m+1} = 0$. This model accounts for heterogeneity resulting from individual covariates and time, but does not depend on individual's prior capture history.

There is no equivalent model of Otis et al. (1978) although this model is a restricted version of their model $\mathrm{M_{th}}$. Here, $P_{ij} = h(\beta_0 + \beta_1 z_i + \beta_{j+1} x_j)$, then the corresponding estimating equations are the followings,

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij}) = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})z_i = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})x_j = 0. \end{cases} \tag{3.18}$$

Model 8. (special case of $\mathrm{M_{tbh}}$)

$$\ln[P_{ij}/(1 - P_{ij})] = \beta_0 + \beta_1 z_i + \beta_{j+1} x_j + \beta_b v_{ij}; j = 1, 2, ..., m, i = 1, 2, ..., N.$$

This is Model 6 with probabilities allowed to vary over time. Again this is a special case of the model $\mathrm{M_{tbh}}$ of Otis et al. (1978), but there is not equivalent to their model. Suppose $P_{ij} = h(\beta_0 + \beta_1 z_i + \beta_{j+1} x_j + \beta_b v_{ij})$, then we get the following estimating equations from GEECR:

$$\begin{cases} \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij}) = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})z_i = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})x_j = 0 \\ \sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - \mu_{ij})v_{ij} = 0. \end{cases} \tag{3.19}$$

3.4 Methods of Model Selection

The GLM methodology relies on maximum likelihood theory for independent observations (McCullagh and Nelder, 1989) whereas the GEE methodology relies on the quasi-likelihood theory (Wedderburn, 1974) for analyzing longitudinal data, and no assumption is made regarding the distribution of response observations.

Therefore, model selection criteria derived under likelihood theory in GLM, as an example, Akaike's information criterion (AIC) (Akaike, 1973) is not appropriate to GEE directly. However, Pan (2001) proposed a modification of AIC, called 'quasi-likelihood information criterion' (QIC). The quasi-likelihood in QIC is constructed on the basis of the independence working correlation assumption. This criterion can also be applied to select the appropriate working correlation structure and therefore the best subset of covariates that is nearest to the true model from a set of potential candidate models in GEE analyses.

We already defined, Y_i to be the vector of capture history of the i^{th} individual and X_i to be the design matrix. Hence under the framework of GLM, $g(\mu_i) = X_i\beta$, where $g(\cdot)$ is the link function and $\mu_i = E(Y_i)$. Then according to Akaike (1973), the AIC is given by

$$\mathrm{AIC} = -2LL + 2p,$$

denoting LL as the log-likelihood and p is the number of parameters in the model. Pan (2001) modified the above formula and created an adjustment for the penalty term $2p$ for GEE, leading to the subsequent formula for QIC as

$$\mathrm{QIC} = -2Q(\hat{\mu}; I) + 2\mathrm{trace}(\hat{\Omega}_I^{-1}\hat{V}_R), \tag{3.20}$$

where I symbolizes the covariance structure under independence used to calculate the quasi-likelihood function. The coefficient estimates $\hat{\beta}$ and robust variance estimator $\hat{V}_R$ are obtained from a general working correlation structure $R(\alpha)$. Another variance estimator $\hat{\Omega}_I$ is obtained under the independence correlation structure assumption. Here $\hat{\mu_i} = g^{-1}(X_i\hat{\beta})$ and $g^{-1}(\cdot)$ is the inverse link function in the GEE model. When $\mathrm{trace}(\hat{\Omega}_I^{-1}\hat{V}_R) \approx \mathrm{trace}(I) = p$, then the simplified version of QIC, denoted by $\mathrm{QIC_u}$ (Pan, 2001) is given by,

$$\mathrm{QIC_u} = -2Q(\hat{\mu}; I) + 2p. \tag{3.21}$$

Table 3.1: Variance and quasi-likelihood functions for commonly used distributions in the exponential family

Distribution	$V(\mu)$	$Q(\mu)$
Bernoulli	$\mu(1-\mu)$	$yln(\frac{\mu}{1-\mu}) + ln(1-\mu)$
Normal	1	$-\frac{1}{2}\sum(y-\mu)^2$
Poisson	μ	$yln(\mu) - \mu$
Gamma	μ^2	$-(\frac{y}{\mu} + ln(\mu))$
Negative Binomial	$\mu + \mu^2$	$y(ln(\mu)) - 2ln(\mu + 1)$
Inverse Gaussian	μ^3	$-\frac{y}{2\mu^2} + \frac{1}{\mu}$

However, $\mathrm{QIC_u}$ cannot be used to select the best working covariance structure because of the assumption of asymptotic equivalence of $\hat{\Omega}_I$and $\hat{V}_R$ (Cui, 2007). Therefore, we use QIC to select an optimal correlation structure. Based on this best correlation structure we can further select the best subset of explanatory variables. The model with the smallest QIC value will be chosen as the most parsimonious model with the best correlation structure (Pan, 2001; Cui, 2007). Other statistics, such as the Wald χ^2 and deviance, cannot be used for comparing GEE models because they do not take into account the number of parameters in the model. Therefore, they may give misleading decisions for model selection in GEE analyses.

The general form of the quasi-likelihood model in (3.20) and (3.21) (Wedderburn, 1974) is given by

$$Q(\mu) = \int_{y}^{\mu} \frac{y-w}{V(w)} dw. \tag{3.22}$$

$V(\mu)$ is the variance of the response observations which is a function of the mean μ. The value of $V(\mu)$ is given in Table 3.1 for some exponential family distributions. We can compute the value of the quasi-likelihood $Q(\mu)$ by substituting with the corresponding value of $V(\mu)$ in (3.22).

3.5 An Illustrative Example

For illustration purposes, we considered the Deer mice (*Peromyscus maniculatus*) data set extracted from Appendix 1 of Huggins (1991). This data set, collected by V. Reid at East Stuart Gulch Colorado, was also distributed with program CAPTURE (Otis et al., 1978). We select this data set since it has capture history with corresponding individual covariates of *Peromyscus maniculatus* for the application of GEE in capture-recapture methodology and for comparison to currently used methodology. The data set to be analyzed consists of 38 distinct deer mice of which 21 males and 17 females that were caught on 6 trapping occasions. The averaged observed weight is 14.53 grams, with sample standard deviation 4.84. There were 24 youngs, 3 semi-adults and 11 adults *Peromyscus maniculatus.* The semi-adults were considered as adults in our analysis. In the previous discussion of this data set, Otis et al. (1978, p. 32) using their techniques, determined that the most suitable model of those available at that time was model $\mathrm{M_b}$, which allows for capture probabilities to vary in response to prior capture. Huggins (1991) applied this data set on seven models and use AIC to select among them. Model 5, a special case of model $\mathrm{M_{bh}}$ of Otis et al. (1978) was then chosen as the 'best' model. We use this model as a real example for illustration of GEE in capture-recapture methodology.

For model $\mathrm{M_{bh}}$, the following equation is formulated according to the covariates of the data set,

$\ln(\frac{P_{ij}}{1-P_{ij}}) = \beta_0 + \beta_{sex} \times sex_{(i)} + \beta_{age} \times age_{(i)} + \beta_{wt} \times weight_{(i)} + \beta_b \times v_{ij}; j = 1, 2, ..., m; i = 1, 2, ..., n$, where, β_{sex}, β_{age}, β_{wt} and β_b represent the sex, age, weight and behavior effect respectively, and v_{ij} is the indicator for capture or not capture before occasion j. The capture probabilities depend on the individual covariates and its past capture history. The covariates are assumed to remain the same during the course of trapping experiment.

Here we estimate the parameters assuming independence, exchangeable, unstruc-

Table 3.2: GEE estimates (standard error in parenthesis) for *Peromyscus maniculatus*'s capture-recapture data assuming various correlation structure models

	Independent Correlation					Exchangeable Correlation				
Covariate	Coff. (S.E.)	P-value	95% CI	O.R.	M.E.	Coff. (S.E.)	P-value	95% CI	O.R.	M.E.
sex	0.702 (.289)	.015	(.136, 1.27)	2.02	.173	0.709 (.316)	.025	(.091, 1.33)	2.03	.175
age	-1.536 (.529)	.004	(-2.57, -.498)	.215	-.366	-1.573 (.578)	.007	(-2.71, -.440)	0.21	-.374
weight	0.127 (.052)	.015	(.024, .229)	1.14	.032	0.129 (.057)	.024	(.017, .241)	1.14	.032
v_{ij}	0.670 (.289)	.020	(.104, 1.24)	1.95	.166	0.528 (.290)	.069	(-.041, 1.096)	1.70	.131
constant	-1.935 (.641)	.003	(-3.19, -.678)	⋯	⋯	-1.880 (.699)	.007	(-3.25, -.509)	⋯	⋯
$\hat{N}$	40.67 (2.18)	⋯	⋯	⋯	⋯	39.88 (1.79)	⋯	⋯	⋯	⋯
	Autoregressive Correlation					Pairwise Correlation				
sex	0.691 (.280)	.013	(.143, 1.234)	2.00	.171	0.544 (.296)	.066	(-.036, 1.12)	1.72	.135
age	-1.52 (.513)	.003	(-2.53, -.519)	.218	-.363	-1.55 (.544)	.004	(-2.62, -.485)	.212	-.370
weight	0.126 (.051)	.013	(.027, .225)	1.13	.031	0.130 (.053)	.015	(.026, .235)	1.14	.032
v_{ij}	0.743 (.283)	.009	(.188, 1.30)	2.10	.184	0.593 (.250)	.018	(.102, 1.08)	1.81	.147
constant	-1.97 (.621)	.002	(-3.16, -.750)	⋯	⋯	-1.82 (.651)	.005	(-3.09, -.540)	⋯	⋯
$\hat{N}$	39.64 (1.34)	⋯	⋯	⋯	⋯	39.17 (1.13)	⋯	⋯	⋯	⋯

tured and autoregressive working correlation structures among capture occasions to estimate capture probabilities. The resulting capture probabilities is used to estimate population size. Table 3.2 provides the point estimates and standard error of the parameters, P-value, 95% confidence interval, odds ratio (O.R.) and marginal effects (M.E.). Odds ratio estimates the chances of a particular event capturing in one group (for instance male) in relation to its rate of capture in another group (female), which describes the strength of association or dependence between two groups. For continuous explanatory covariates, the odds ratio tells us what the odds (the ratio of the probability of an event capturing to the probability of its not capturing) are for a unit change in the explanatory covariate. Marginal effects measure the expected instantaneous change in the response variable as a function of a change in a certain explanatory covariate while keeping all the other covariates constant. The marginal effects are computed at their means for the continuous explanatory covariates. For the binary covariates, however, the marginal effects are calculated as the discrete change in the explanatory covariate form 0 to 1. The estimation is done by the statistical software stata 11 (Stata, 2010) and R (R Development Core Team, 2013). We observe that the covariates age, sex, weight and v_{ij} are all significant for a nominal significant level of 0.07 (see P-values in Table 3.2). The probability of capture appears to be significantly higher for males than for females, assuming all other covariates remain fixed. For instance, the probability of capturing males is about twice the probability of capturing females except for pairwise correlation structure, where 1.72 times. Males are more likely to increase recapture by about 17% points than females except when pairwise correlation structure is used. In that case there is about 14% increase. In terms of odds ratio, the probability of capturing individuals in the young age group are approximately 4.65 times more than the probability of capturing individuals in the adult age group. The marginal effects also indicate that for any given correlation structure, individuals in the adult age group are less likely to be captured than the ones in the young age group. There is a 14% increase in the odds of capturing for one unit increase of weight, but slightly lower for autoregressive correlation structure. Marginal effects indicate that individuals are 3.2% more likely to increase recapture for one unit increase of weight. The individuals that are captured before a given capture time are 1.95, 1.70, 2.10 and 1.81 times more likely to increase recapture in terms of odds ratio, and are more likely to increase recapture by 16.6%, 13.1%, 18.4% and 14.7% in terms of marginal effect for independence, exchangeable, autoregressive and pairwise correlation structure models respectively than the individuals that are not captured before a given capture time. All the β's are significant and the positive estimate of β_b shows that the animals are trap happy for this data set which support the findings of Otis et al. (1978). Considering the assumption of various working correlation structures, the estimates of population size are given in Table 3.2, those are quite close to the estimates given by Otis et al. (1978) using their model $\mathrm{M_b}$ ($\hat{N} = 41, \mathrm{SE}(\hat{N}) = 3.01$).

However, the estimated standard errors of different parameters are not very similar for various correlation structures (see Table 3.2). We observed that estimates obtained under the assumption of exchangeable correlation within the responses having greater standard errors than the others. We also observed that time correlation play an important role when analyzing this data. The question is what sort of correlation structure to use. One strategy may be selecting a model that shows better relative efficiency of the estimated coefficients as GEE provide unbiased estimates. Table 3.3 compares the relative efficiencies of the parameters obtained from different methods with respect to parameters obtained under the assumption of independence correlation structure. It is shown that for almost all covariates the estimates obtained under autoregressive correlation are more efficient estimators when compared to the others but the relative efficiency of the estimated coefficient of v_{ij} is 1.021 under autoregressive correlation structure and 1.156 under pairwise correlation structure. Although the differences are minimal, the QIC model selection criteria (Pan, 2001) suggests that a model with pairwise correlation structure should be chosen as it has the smallest QIC (297.693) (see Table 3.3). There are small differences among $\hat{N}$, and QIC for the different correlation structures. This is generally true particularly for two cases, when (i) all individuals have the same number of capture occasions (m), and (ii) the levels of the covariates do not vary within an individual. In the example, the values of all covariates sex, age, and weight for an individual are similar for all capture occasions.

Table 3.3: QIC for the various working correlation structure models and relative efficiencies of the estimated coefficients under independent correlation structure model

Covariate	Independence	Exchangeable	Autoregressive	Pairwise
QIC	298.386	298.938	298.229	297.693
$\mathrm{QIC_u}$	297.303	297.546	297.369	297.776
sex	1.000	0.915	1.032	0.976
age	1.000	0.915	1.031	0.972
weight	1.000	0.912	1.020	0.981
v_{ij}	1.000	0.997	1.021	1.156

3.6 A Simulation Study

Since we found that covariates in autoregressive correlation structure were generally more efficient for the applied data set, a Monte Carlo (MC) simulation study was carried out considering this approach to evaluate the performance of GEE in capture-recapture studies. The program was written in R (R Development Core Team, 2013) and simulations were run on Core i3-2310M CPU computer.

Program inputs were the number of Monte Carlo replicates (B), population size (N), mean capture probability ($\bar{p}$), number of trapping occasions (m), correlation coefficient (α), behavioral effect, and individual covariates. Factors used in the simulations were population size, $N = 100, 200$, and 500; mean capture probability, $\bar{p} = 0.1$, 0.3, and 0.5; number of trapping occasions, $m = 6$ and 8; and correlation coefficient, $\alpha = -0.5$, -0.3, -0.1, 0, 0.1, 0.3, 0.5 for the autoregressive correlation structure. Based on results from a real data analysis, we generated two individual level covariates, one was considered as continuous, for example, *weight* that follows $\mathcal{N}(15, 2^2)$ and another was discrete for example, *sex* that follows $Bin(1, 0.5)$ distribution. We generated correlated capture history Y_{ij} following the method of Qaqish (2003), and behavioral effects were generated from the capture history as an indicator variable assuming the value 1 if individual i has been captured before j, and 0 otherwise. One thousand repetitions were completed for each combination of factors.

To assess estimators performances we computed average estimate of population size, AVE($\hat{N}$); standard error of the population estimates, SE($\hat{N}$); average captured individuals ($\bar{n}$); percentage relative bias, PRB $= 100 \times (E[\hat{N}] - N) \div N$, where $E(\hat{N})$ is estimated by AVE($\hat{N}$); percent coefficient of variation, CV $= 100 \times \text{SE}(\hat{N}) \div E[\hat{N}]$ and normal-based 95% confidence intervals (CI), i.e., $E(\hat{N}) \pm 1.96 \times \text{SE}(\hat{N})$. Estimator performance is based on comparisons of root mean square error, RMSE $= \sqrt{\widehat{\text{Var}}(\hat{N}) + \text{bias}^2}$.

Simulation results in Table 3.4 and Table 3.5 show that estimation of average animal abundance (AVE($\hat{N}$)) is almost the same as true population size (N) when there is no correlation ($\alpha = 0$) for the trapping occasions 6 and 8 for the average capture probability 0.5 ($\bar{p} = 0.5$). For lower average capture probability ($\bar{p} = 0.1$ or 0.3), the GEE estimator in capture-recapture studies is not stable. For low average capture probability, the estimated population size and its standard error vary depending on the strength of linear correlation among capture occasions and related factors. For a fixed average capture probability and trapping occasions, the estimated population size and its standard error is higher for negative correlation comparatively to the same strength of positive correlation. The performance of estimators for 8 trapping occasions is better than for 6 trapping occasions considering lower CV, absolute value of PRB and RMSE. Overall, the simulation study clearly shows that estimator performance of population size and its standard error depends on average capture probability, correlation among capture occasions and number of trapping occasions. Finally, we observe that an inverse relationship exists between estimated population size and the linear correlation among capture occasions for fixed average capture probability and trapping occasions.

Table 3.4: Under model M_{bh} -simulation results (1000 repetitions) of $m = 6$ trapping occasions for the proposed GEE approach assuming autoregressive correlation structure

$\bar{p}$	α	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	95%CI	PRB	CV	RMSE
0.10	-0.1	100	53	70.96	5.73	59.73-82.19	-29.04	8.07	30.08
0.10	0.0	100	53	64.05	5.04	54.17-73.93	-35.95	7.87	36.69
0.10	0.1	100	42	57.90	4.43	49.22-66.58	-42.10	7.65	42.63
0.10	0.3	100	42	45.71	3.18	39.48-51.94	-54.29	6.97	54.60
0.10	0.5	100	17	34.90	2.12	30.74-39.06	-65.10	6.07	65.32
0.30	-0.3	100	97	104.71	3.41	98.03-111.39	4.71	3.23	5.58
0.30	-0.1	100	91	100.03	3.18	93.80-106.26	0.03	3.18	3.45
0.30	0.0	100	90	96.85	3.05	90.87-102.83	-3.15	3.15	4.97
0.30	0.1	100	84	93.19	2.86	87.58-98.80	-6.81	3.07	8.01
0.30	0.3	100	74	83.76	2.41	79.04-88.48	-16.24	2.88	16.84
0.30	0.5	100	66	71.55	1.78	68.06-75.04	-28.45	2.49	28.89
0.50	-0.5	100	99	101.58	1.28	99.07-104.09	1.58	1.26	1.62
0.50	-0.3	100	99	101.41	1.28	98.90-103.92	1.41	1.26	1.52
0.50	-0.1	100	99	100.69	1.26	98.22-103.16	0.69	1.25	1.23
0.50	0.0	100	98	100.00	1.23	97.59-102.41	0.00	1.23	1.30
0.50	0.1	100	97	99.18	1.21	96.81-101.55	-0.82	1.22	1.78
0.50	0.3	100	94	95.72	1.08	93.60-97.84	-4.28	1.13	4.84
0.50	0.5	100	88	90.06	0.86	88.37-91.75	-9.94	0.95	10.42
0.10	-0.1	200	93	141.28	8.03	125.54-157.02	-29.36	5.68	59.71
0.10	0.0	200	91	127.48	7.03	113.70-141.26	-36.26	5.51	73.22
0.10	0.1	200	87	115.32	6.15	103.27-127.37	-42.34	5.34	85.26
0.10	0.3	200	64	91.47	4.43	82.79-100.15	-54.26	4.84	108.86
0.10	0.5	200	63	70.32	2.93	64.58-76.06	-64.84	4.16	129.89
0.30	-0.3	200	190	209.52	4.78	200.15-218.89	4.76	2.28	10.34
0.30	-0.1	200	185	199.84	4.48	191.06-208.62	-0.08	2.24	4.90
0.30	0.0	200	182	193.43	4.26	185.08-201.78	-3.28	2.20	8.46
0.30	0.1	200	179	186.22	4.01	178.36-194.08	-6.89	2.16	14.85
0.30	0.3	200	148	167.92	3.37	161.31-174.53	-16.04	2.01	32.67
0.30	0.5	200	140	143.34	2.45	138.54-148.14	-28.33	1.71	57.09
0.50	-0.5	200	199	203.12	1.81	199.57-206.67	1.56	0.89	3.15
0.50	-0.3	200	199	202.64	1.80	199.11-206.17	1.32	0.89	2.79
0.50	-0.1	200	199	201.31	1.76	197.86-204.76	0.66	0.87	1.94
0.50	0.0	200	198	199.98	1.73	196.59-203.37	-0.01	0.86	1.86
0.50	0.1	200	195	198.04	1.67	194.77-201.31	-0.98	0.84	3.03
0.50	0.3	200	191	191.50	1.48	188.60-194.40	-4.25	0.77	9.10
0.50	0.5	200	177	179.79	1.19	177.46-182.12	-10.11	0.65	20.71
0.10	-0.1	500	257	352.12	12.59	327.44-376.80	-29.58	3.58	148.87
0.10	0.0	500	248	319.80	11.10	298.04-341.56	-36.04	3.47	180.95
0.10	0.1	500	211	287.44	9.63	268.57-306.31	-42.51	3.35	213.14
0.10	0.3	500	212	228.25	6.91	214.71-241.79	-54.35	3.03	272.06
0.10	0.5	500	156	173.78	4.45	165.06-182.50	-65.24	2.56	326.43
0.30	-0.3	500	477	523.27	7.55	508.47-538.07	4.65	1.44	24.11
0.30	-0.1	500	448	498.92	7.02	485.16-512.68	-0.22	1.41	7.81
0.30	0.0	500	447	483.64	6.72	470.47-496.81	-3.27	1.39	18.37
0.30	0.1	500	425	465.19	6.30	452.84-477.54	-6.96	1.35	35.97
0.30	0.3	500	396	418.34	5.24	408.07-428.61	-16.33	1.25	82.34
0.30	0.5	500	346	357.31	3.82	349.82-364.80	-28.54	1.07	143.09
0.50	-0.5	500	499	507.81	2.85	502.22-513.40	1.56	0.56	7.85
0.50	-0.3	500	498	506.54	2.83	500.99-512.09	1.31	0.56	6.67
0.50	-0.1	500	492	503.22	2.77	497.79-508.65	0.65	0.55	3.93
0.50	0.0	500	494	499.73	2.70	494.44-505.02	-0.05	0.54	2.87
0.50	0.1	500	491	495.02	2.61	489.90-500.14	-1.00	0.53	6.15
0.50	0.3	500	472	478.24	2.32	473.69-482.79	-4.35	0.48	22.39
0.50	0.5	500	453	448.80	1.82	445.23-452.37	-10.24	0.41	51.68

Table 3.5: Under model $\mathrm{M_{bh}}$ -simulation results (1000 repetitions) of $m = 8$ trapping occasions for the proposed GEE approach assuming autoregressive correlation structure

$\bar{p}$	α	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	95%CI	PRB	CV	RMSE
0.10	-0.1	100	66	81.66	5.56	70.76-92.56	-18.34	6.81	19.78
0.10	0.0	100	58	74.90	4.98	65.14-84.66	-25.10	6.65	26.10
0.10	0.1	100	57	68.18	4.44	59.48-76.88	-31.82	6.51	32.51
0.10	0.3	100	39	54.52	3.27	48.11-60.93	-45.48	6.00	45.93
0.10	0.5	100	40	42.05	2.16	37.82-46.28	-57.95	5.14	58.21
0.30	-0.3	100	98	103.44	2.40	98.74-108.14	3.44	2.32	3.83
0.30	-0.1	100	97	101.11	2.34	96.52-105.70	1.11	2.31	2.53
0.30	0.0	100	96	99.42	2.28	94.95-103.89	-0.58	2.29	2.60
0.30	0.1	100	90	97.20	2.19	92.91-101.49	-2.80	2.26	4.08
0.30	0.3	100	92	90.51	1.92	86.75-94.27	-9.49	2.12	10.25
0.30	0.5	100	86	80.60	1.52	77.62-83.58	-19.40	1.88	19.88
0.50	-0.5	100	99	100.38	0.62	99.16-101.60	0.38	0.62	0.39
0.50	-0.3	100	99	100.38	0.63	99.15-101.61	0.38	0.63	0.42
0.50	-0.1	100	98	100.25	0.64	99.00-101.50	0.25	0.64	0.49
0.50	0.0	100	97	100.08	0.64	98.83-101.33	0.08	0.64	0.59
0.50	0.1	100	98	99.71	0.64	98.46-100.96	-0.29	0.64	0.91
0.50	0.3	100	97	98.05	0.62	96.83-99.27	-1.95	0.63	2.49
0.50	0.5	100	96	93.97	0.54	92.91-95.03	-6.03	0.57	6.53
0.10	-0.1	200	119	163.49	7.84	148.12-178.86	-18.26	4.79	37.93
0.10	0.0	200	118	149.12	6.98	135.44-162.8	-25.44	4.68	51.83
0.10	0.1	200	114	136.50	6.22	124.31-148.69	-31.75	4.56	64.26
0.10	0.3	200	104	109.49	4.53	100.61-118.37	-45.26	4.14	90.91
0.10	0.5	200	77	83.78	2.99	77.92-89.64	-58.11	3.56	116.47
0.30	-0.3	200	197	206.71	3.35	200.14-213.28	3.35	1.62	7.09
0.30	-0.1	200	185	202.11	3.24	195.76-208.46	1.06	1.60	3.78
0.30	0.0	200	190	198.75	3.18	192.52-204.98	-0.62	1.60	3.88
0.30	0.1	200	178	194.65	3.07	188.63-200.67	-2.67	1.58	6.68
0.30	0.3	200	177	181.59	2.72	176.26-186.92	-9.20	1.50	19.04
0.30	0.5	200	163	160.80	2.08	156.72-164.88	-19.60	1.30	39.71
0.50	-0.5	200	199	200.76	0.87	199.05-202.47	0.38	0.44	0.77
0.50	-0.3	200	199	200.72	0.89	198.98-202.46	0.36	0.44	0.77
0.50	-0.1	200	198	200.46	0.89	198.72-202.20	0.23	0.45	0.74
0.50	0.0	200	197	200.10	0.89	198.36-201.84	0.05	0.45	0.88
0.50	0.1	200	198	199.39	0.89	197.65-201.13	-0.31	0.45	1.34
0.50	0.3	200	197	196.23	0.84	194.58-197.88	-1.88	0.43	4.33
0.50	0.5	200	188	188.24	0.72	186.83-189.65	-5.88	0.39	12.20
0.10	-0.1	500	296	407.08	12.28	383.01-431.15	-18.59	3.02	94.46
0.10	0.0	500	280	372.94	10.99	351.40-394.48	-25.41	2.94	128.11
0.10	0.1	500	279	338.94	9.67	319.99-357.89	-32.21	2.85	161.80
0.10	0.3	500	225	271.21	7.04	257.41-285.01	-45.76	2.60	229.22
0.10	0.5	500	195	207.43	4.56	198.49-216.37	-58.51	2.20	292.81
0.30	-0.3	500	482	516.50	5.33	506.05-526.95	3.30	1.03	16.93
0.30	-0.1	500	479	504.63	5.12	494.59-514.67	0.93	1.02	6.82
0.30	0.0	500	472	496.07	4.98	486.31-505.83	-0.79	1.00	6.91
0.30	0.1	500	459	484.83	4.78	475.46-494.20	-3.04	0.99	16.45
0.30	0.3	500	429	452.17	4.18	443.98-460.36	-9.57	0.92	48.54
0.30	0.5	500	387	401.20	3.23	394.87-407.53	-19.76	0.81	99.25
0.50	-0.5	500	499	501.88	1.38	499.18-504.58	0.38	0.27	1.89
0.50	-0.3	500	499	501.81	1.39	499.09-504.53	0.36	0.28	1.86
0.50	-0.1	500	499	501.10	1.40	498.36-503.84	0.22	0.28	1.46
0.50	0.0	500	498	500.06	1.39	497.34-502.78	0.01	0.28	1.38
0.50	0.1	500	494	498.38	1.39	495.66-501.10	-0.32	0.28	2.53
0.50	0.3	500	495	490.36	1.31	487.79-492.93	-1.93	0.27	10.21
0.50	0.5	500	472	470.18	1.11	468.00-472.36	-5.96	0.24	30.34

3.7 Discussion

In this chapter, we proposed a generalized estimating equations approach to model capture probabilities of heterogeneous populations, and to evaluate the effect of accounting for correlation structures on capture-recapture model selection. The GEE approach is an important tool for getting unbiased estimates in the analysis of correlated capture-recapture data when capture probabilities are modelled as a function of covariates. To evaluate the pattern of time dependency is important in several regards: (i) it may help to characterize the relationship between the capture probability and covariates, and (ii) it may also be important to estimate the population parameters accurately in capture-recapture studies. A common question arise is 'what happens if one ignores the time dependency and uses the traditional regression methodology assuming independence among capture occasions?' From a theoretical point of view, there are at least two consequences as results of ignoring time dependency: incorrect assessment of the regression estimates and inefficient estimation of regression coefficients. Therefore, estimated capture probabilities may be incorrect and consequently population size may not be accurately estimated. We account for heterogeneity due to measurable individual characteristics modelling capture probabilities as a function of covariates. The GEE versions of all closed population capture-recapture models considered by Otis et al. (1978) and Huggins (1991), and their corresponding estimating equations are also presented. A real application is presented for the various working correlation structures that are permitted in GEE instead of assuming independence of capture occasions. The GEE approach in capture-recapture studies provides a useful inference procedure for estimating population size, particularly when the capture probability is high but the estimates become unstable for low capture probabilities. However, other existing methods in the capture-recapture literature allowing for heterogeneity have similar problems (Hilborn et al., 1976; Nichols and Pollock, 1983; Nichols, 1986). Estimation results agree with Otis et al. (1978), but our model jointly accounts for the heterogeneous capture probabilities, time dependence and over-dispersion. According to QIC, it is shown that the GEE method for pairwise correlation structure, is more efficient than the other working correlation structures for the data. To the best of our knowledge, this is the first time that QIC has been applied to capture-recapture studies. The simulation study shows that the estimated population size varies on the nature of existing correlation among capture occasions. Hence, it is important to consider the type of correlation structure among capture occasions when estimating animal population parameters in capture-recapture studies.

Chapter 4

A Generalized Estimating Equations Approach and Model Selection Allowing for Heterogeneity and Time Dependence

---

Part of this chapter is in the form of publication and conference presented paper:

Akanda, M.A.S. and Alpizar-Jara, R. (To be submitted). 'A Generalized Estimating Equations Approach and Model Selection Allowing for Heterogeneity and Time Dependence.'

Akanda, M.A.S. and Alpizar-Jara, R. (2013). 'Estimation of Capture Probabilities Using a GEE Approach.' *Atas do XX Congresso Anual da Sociedade Portuguesa de Estatistica* (Maia, M., Campos, P. and Silva, P. D., eds.), pp. 27–35.

Akanda, M.A.S. and Alpizar-Jara, R. (2012). 'Estimation of Capture Probabilities and Model Selection Using a Generalized Estimating Equation Approach.' *Resumos de XX Congresso Anual da Sociedade Portuguesa de Estatistica*, September 26–29, 2012, Porto, Portugal, pp.595–596.

In this chapter, the proposed generalized estimating equations (GEE) approach is applied to a capture-recapture closed population $\mathrm{M_{th}}$ model to data sets from various capture-recapture experiments. The GEE approach accounts for temporal variation, heterogeneity that can be explained by measurable individual characteristics, and correlation among capture occasions. Here, we model capture probabilities as a function of individual and environmental covariates. We argued that heterogeneity of capture probability and correlation among capture occasions should be accounted for. Quasi-likelihood information criterion (QIC) is used for selecting best fitted model. The estimates of capture probabilities may then be used to estimate population size. A simulation study is also conducted to evaluate the performance of our proposed methodology.

4.1 Background

The primary aim is to estimate the population size in capture-recapture experiments. There are many challenges to statisticians for making accurate inferences, but we are interested in, if (i) the capture probabilities are heterogeneous between individuals, and (ii) capture probabilities are dependent among capture occasions in capture-recapture studies. Importantly, ignoring heterogeneity arising from individuals or clusters of individuals may induce biases in parameter estimates (Barry et al., 2003; Hwang and Huggins, 2005). Modelling dependency among capture occasions is another important issue in capture-recapture studies. Failure to account for this dependency also provides biased estimates. Other authors have addressed the problem of time dependence in capture probabilities using multilevel analyses in open population models (Choquet et al., 2013), log-linear models in epidemiological indications (Zwane and Heijden, 2005), and some sort of dependencies deal through the modelling of behavioral effects (Pradel and Sanz-Aguilar, 2012). Our interest is in estimating the size of a closed population using a sub-model of the type M_{th}, where heterogeneity and time effect are considered, there is no behavioral response to capture, and the capture probabilities depend on covariates. There are various advantages for models incorporating covariates such as, (i) the models provide a clear explanation of the sources of heterogeneity, and each covariate effect can be assessed; and (ii) if all relevant covariates are included, then these models generally yield better estimators with respect to bias and precision.

A wide variety of approaches have been considered when fitting capture-recapture closed population M_{th} model, including martingale methods (Lloyd and Yip, 1991), sample coverage models (Chao et al., 1992), log-linear and latent class models (Agresti, 1994), finite mixture models (Pledger, 2000), the use of individual covariates in GLM (Huggins, 1989) and robust P-spline approach (Stoklosa and Huggins, 2012). Similar to Huggins (1989, 1991), we assume that the capture probabilities can be modeled through a logit-link function of the covariates, but in this chapter, we use a generalized estimating equations (GEE) (Liang and Zeger, 1986; Zeger and Liang, 1986) approach which accounts for correlation over time instead of assuming independence of capture occasions. A quasi-likelihood procedure is used in this approach to estimate the regression parameters associated to the capture probabilities and population size. The term ‘quasi’ in this context indicates that the model for the mean response depends only on the covariates of interest and not on any random effects.

The models considered here are briefly described in Section 4.2, and they are fitted to the different data sets in Section 4.3. In section 4.4, we discuss a proposed model selection procedure based on quasi-likelihood information criterion (QIC). A simulation study is conducted to show the performance of our proposed approach in Section 4.5. Finally, in Section 4.6 we conclude with some discussion.

4.2 Model Formulation

We follow the same notation described in the previous chapters to build-up modelling structure. Let, the probability P_{ij} that the i^{th} animal is captured on the j^{th} capture occasion is given by

$$P_{ij} = \Pr(Y_{ij} = 1|z_i, x_{ij}) = h(\beta_0 + \beta_1 z_i + \beta_{j+1} x_{ij}) \tag{4.1}$$

for $i = 1, 2, ..., N; j = 1, 2, ..., m$ where $h(u) = \exp(u)/\{1 + \exp(u)\}$ is the logistic function and

$$X_i = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_i & z_i & \dots & z_i \\ x_{i1} & 0 & \dots & 0 \\ 0 & x_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_{im} \end{bmatrix}'$$

is the design matrix. This design matrix is useful to construct a capture-recapture closed population model of the type $\mathrm{M_{th}}$ where variation among individuals, time, and correlation among capture occasions may be explained by the individual covariate z_i and environmental covariate x_{ij}.

If the environmental covariate x_{ij} only depends on the j^{th} capture occasion for the i^{th} individual, then we could simplify equation (4.1) to

$$P_{ij} = \Pr(Y_{ij} = 1|z_i, x_j) = h(\beta_0 + \beta_1 z_i + \beta_2 x_j) \tag{4.2}$$

for $i = 1, 2, ..., N; j = 1, 2, ..., m$.

In this case, the design matrix X_i have the following form

$$X_i = \begin{bmatrix} 1 & 1 & ... & 1 \\ z_i & z_i & ... & z_i \\ x_1 & x_2 & ... & x_m \end{bmatrix}'.$$

Therefore, the vector of parameters, $\beta = (\beta_0, \beta_1, \beta_2)'$ for model (4.2) can be obtained by solving the following GEECR following the previous Section 3.2:

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - \mu_i) = 0. \tag{4.3}$$

Let $\hat{\beta}$ be the resulting estimator of β, then the Horvitz-Thompson estimator (1952), as defined in (3.10), may be used to estimate population size as in Huggins (1989). We can estimate the variance of the estimated population size $\hat{N}$ by using the formula (3.11).

4.3 Illustrative Examples

In this section, we apply the techniques discussed in Section 4.2 to real data sets.

4.3.1 Example 1: Deer mice data

Our first example concerns the captures of 38 distinct Deer mice (*Peromyscus maniculatus*) for $m = 6$ capture occasions from Section 1.7.1. Basic capture-recapture frequency statistics in Table 4.1 are computed by Rcapture package (Baillargeon and Rivest, 2007). It displays, for $j = 1, 2, ..., m$, the number of animals captured exactly j times (f_j), the number of animals captured for the first time on occasion j (u_j), the number of animals captured for the last time on occasion j (v_j) and the number of animals captured on occasion j (n_j). Table 4.1 shows that

9 out of 38 animals (23.68%) were captured only once. An interesting tool to explore heterogeneity in the capture probabilities are the graphs of $\log\big(f_j/\binom{m}{j}\big)$ and $\log(u_j)$ versus j generated by the Rcapture package (Baillargeon and Rivest, 2007). Figure 4.1 shows that there is some evidence of heterogeneity in capture probabilities. We estimated the population size, standard error and AIC for model selection by applying capture-recapture estimators available in Rcapture package (see Table 4.2). The notation and models in Table 4.2 are described in Baillargeon and Rivest (2007) and references therein.

Table 4.1: Frequency statistics for the capture-recapture occasions of Deer mice data

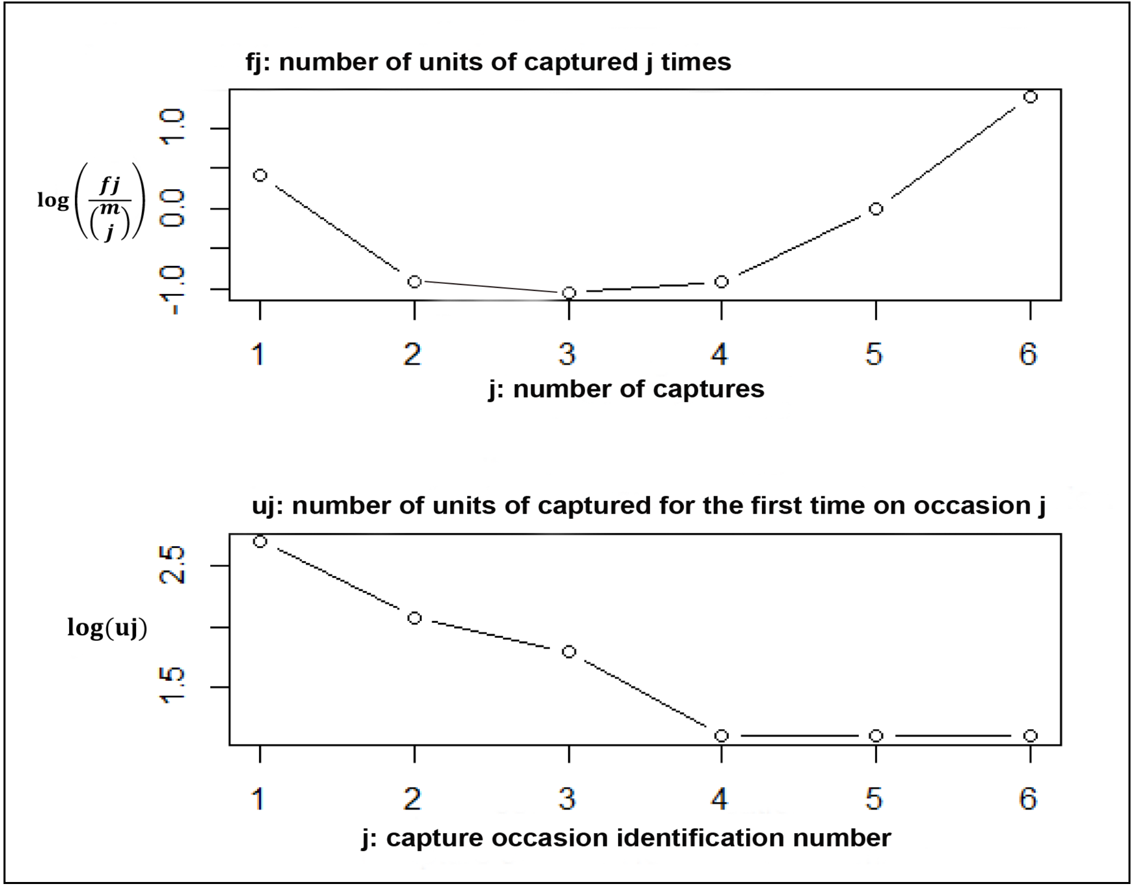
	f_j	u_j	v_j	n_j
$j = 1$	9	15	0	15
$j = 2$	6	8	0	20
$j = 3$	7	6	2	16
$j = 4$	6	3	2	19
$j = 5$	6	3	9	25
$j = 6$	4	3	25	25

We use this data to apply our proposed generalized estimating equations approach in the capture-recapture $\mathrm{M_{th}}$ model using covariates. According to the model formulation of $\mathrm{M_{th}}$, the following equation is used for the Deer mice data,

$$\ln\Big(\frac{P_{ij}}{1-P_{ij}}\Big) = \beta_0 + \beta_{age} \times age_{(i)} + \beta_{sex} \times sex_{(i)} + \beta_{wt} \times weight_{(i)} + \beta_t \times time_{(j)}; \quad (4.4)$$

$j = 1, 2, ..., m$; $i = 1, 2, ..., n$, where, β_{age}, β_{sex}, β_{wt} and β_t represent the age, sex, weight and time effect respectively. The estimated model parameters, their standard errors, P-values, and odds ratios (OR) assuming various correlation structures within the repeated measures for the capture events at different occasions are reported in Table 4.3. The odds ratios indicate that for any given correlation structure, individuals in the young age group are more likely to be captured than

Figure 4.1: Exploratory heterogeneity graph of $\log\left(f_j/\binom{m}{j}\right)$ and $\log(u_j)$ versus j generated by the Rcapture function for the Deer mice data



the ones in the adult age group, assuming all other covariates remain fixed. The odds of capturing individuals in the young age group are about (1/.171) 5.85 times higher than the ones in the adult age group, but (1/.146) 6.85 times higher for pairwise correlation structure model. The probability of capturing males is about twice the probability of capturing females. The probability of capture appears to be significantly higher for males than for females at a nominal significance level of 0.05, except for the pairwise correlation structure (see P-values in Table 4.3). There is also about a 15% increase in the odds of capturing for one unit increase of weight except when pairwise correlation structure is used. In that case there is about 18% increase. There is a 26% increase in the risk of capturing for changing one occasion to another occasion, but slightly lower for pairwise correlation

Table 4.2: Abundance estimations and model fits of Deer mice data using capture-recapture methods

Models	Abundance	stderr	Deviance	df	AIC
M_0	38.5	0.7	84.542	61	143.512
M_t	38.4	0.7	74.839	56	143.809
M_h Chao (LB)	43.6	5.0	64.754	57	131.724
M_h Poisson2	40.2	1.8	67.896	60	128.866
M_h Darroch	45.7	5.8	65.127	60	126.097
M_h Gamma3.5	60.7	19.7	64.979	60	125.949
M_th Chao (LB)	43.3	4.8	52.676	52	129.646
M_th Poisson2	39.9	1.7	56.234	55	127.204
M_th Darroch	45.7	5.7	53.056	55	124.026
M_th Gamma3.5	62.7	21.3	52.918	55	123.888
M_b	42.3	3.8	70.866	60	131.836
M_bh	44.9	17.9	70.264	59	133.234

Note: 1 eta parameter has been set to zero in the M_h Chao model

structure model. This outcome may indicate that animals are trap happy as the capturing risk of individuals increases from one occasion to another occasion.

Table 4.3: GEE estimates for *Peromyscus maniculatus*'s capture-recapture data under the various working correlation structures

	Independent Correlation				Exchangeable Correlation			
Cov.	Coff.	Std. Err.	P-value	O.R.	Coff.	Std. Err.	P-value	O.R.
age	-1.767	.531	0.001	0.171	-1.775	.654	0.007	0.170
sex	0.753	.291	0.010	2.124	0.759	.358	0.034	2.136
weight	0.143	.053	0.007	1.154	0.144	.065	0.028	1.154
time	0.233	.085	0.006	1.262	0.233	.080	0.004	1.262
cons.	-2.532	.722	0.000	···	-2.539	.853	0.003	···
	Autoregressive Correlation				Pairwise Correlation			
age	-1.760	.551	0.001	0.172	-1.926	.628	0.002	0.146
sex	0.761	.302	0.012	2.140	0.632	.341	0.064	1.881
weight	0.142	.055	0.010	1.153	0.164	.062	0.009	1.178
time	0.233	.087	0.007	1.262	0.220	.070	0.002	1.246
cons.	-2.530	.748	0.001	···	-2.625	.812	0.001	···

4.3.2 Example 2: Least chipmunk data

For our second example we consider the Least chipmunk (*Eutamias minimus*) data from Section 1.7.2. The numbers of animals caught for $m = 6$ occasions (n_1 to n_6) were 7, 15, 16, 24, 19, 7 and $\sum n_k = 88$. Out of these 88 captures, $n = 45$ distinct animals were captured where 22 males and 23 females. The recorded capture frequencies (f_1 to f_6) were 21, 12, 7, 3, 2, 0. The average capture frequencies for male and female were 1.86 and 2.04 respectively. Hence, we might guess some heterogeneity of capture probabilities for individual covariates. For these data, Otis et al. (1978) applied a procedure of model selection and found that there was significant time variation and model $\mathrm{M_t}$ was selected as the most appropriate model for the data. The f_j plot in Figure 4.2 shows that there exist little heterogeneity in the capture probabilities and suggests that $\mathrm{M_h}$ or $\mathrm{M_{th}}$ may be considered for this data set (Rivest, 2008). Model $\mathrm{M_{th}}$ is more general than $\mathrm{M_h}$ model. Therefore, combining of the above mentioned two issues, we would like to use model $\mathrm{M_{th}}$ to apply the GEE approach in capture-recapture methodology for this data. The estimated population size, standard error and AIC for model selection using capture-recapture methods are given in Table 4.4.

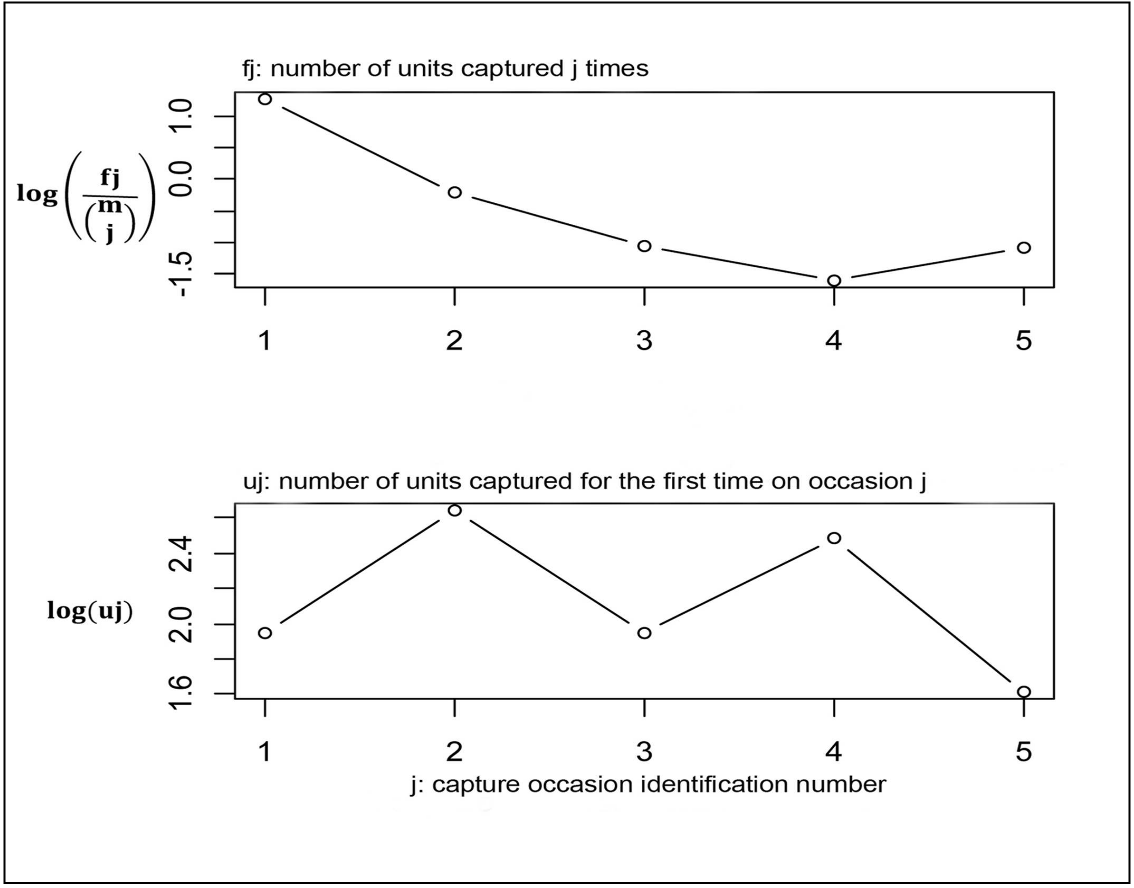
The following equation can be used to applicable the GEE approach for this data,

$$\ln\left(\frac{P_{ij}}{1-P_{ij}}\right) = \beta_0 + \beta_{sex} \times sex_{(i)} + \beta_t \times time_{(j)}; \tag{4.5}$$

$j = 1, 2, ..., m$; $i = 1, 2, ..., n$, where, β_{sex} and β_t represent the effect of sex and time respectively.

The estimation results are reported in Table 4.5 applying the GEE approach in the closed population capture-recapture $\mathrm{M_{th}}$ model for this data. The odds ratios indicate that for any given correlation structure, the probability of capture appears to be higher for females than for males, assuming all other covariates remain fixed. The odds of capturing females are about (1/.872) 1.15 times higher than those of

Figure 4.2: Exploratory heterogeneity graph of $\log\big(f_j/\binom{m}{j}\big)$ and $\log(u_j)$ versus j generated by the Rcapture function of the Least chipmunk data



males but (1/.775) 1.29 times higher for pairwise correlation structure model. The probability of capturing increases for changing one occasion to another. There is a 6% increase in the risk of capturing for changing one occasion to another occasion, but slightly lower for pairwise correlation structure model. In such case there is only 1% increase. This finding may suggest that animals are trap happy for this data set.

Table 4.4: Abundance estimations and model fits of Least chipmunk data using capture-recapture methods

Models	Abundance	stderr	Deviance	df	AIC
M_o	52.2	3.7	69.021	61	129.770
M_t	51.2	3.3	46.662	56	117.412
M_h Chao (LB)	60.3	8.9	63.739	58	130.488
M_h Poisson2	56.2	5.6	65.650	60	128.400
M_h Darroch	67.4	13.7	64.096	60	126.845
M_h Gamma3.5	84.5	31.6	63.910	60	126.660
M_{th} Chao (LB)	59.2	8.3	39.685	53	116.435
M_{th} Poisson2	55.2	5.2	42.062	55	114.812
M_{th} Darroch	67.9	14.1	40.080	55	112.829
M_{th} Gamma3.5	89.8	36.0	39.899	55	112.649
M_b	60.9	12.7	67.316	60	130.066
M_{bh}	50.2	12.8	61.283	59	126.033

Note: 1 eta parameter has been set to zero in the M_h Chao model

Table 4.5: GEE estimates for *Eutamias minimus's* capture-recapture data under the various working correlation structures

	Independent Correlation				Exchangeable Correlation			
Cov.	Coff.	Std. Err.	P-value	O.R.	Coff.	Std. Err.	P-value	O.R.
sex	-.137	.260	0.599	0.872	-.137	.256	0.592	0.872
time	.058	.076	0.447	1.059	.058	.077	0.449	1.060
cons.	-.865	.325	0.008	⋯	-.865	.324	0.008	⋯
	Autoregressive Correlation				Pairwise Correlation			
sex	-.137	.260	0.599	0.872	-.256	.223	0.252	0.775
time	.058	.076	0.445	1.060	.009	.069	0.896	1.009
cons.	-.865	.325	0.008	⋯	-.609	.280	0.029	⋯

4.3.3 Example 3: House mice data

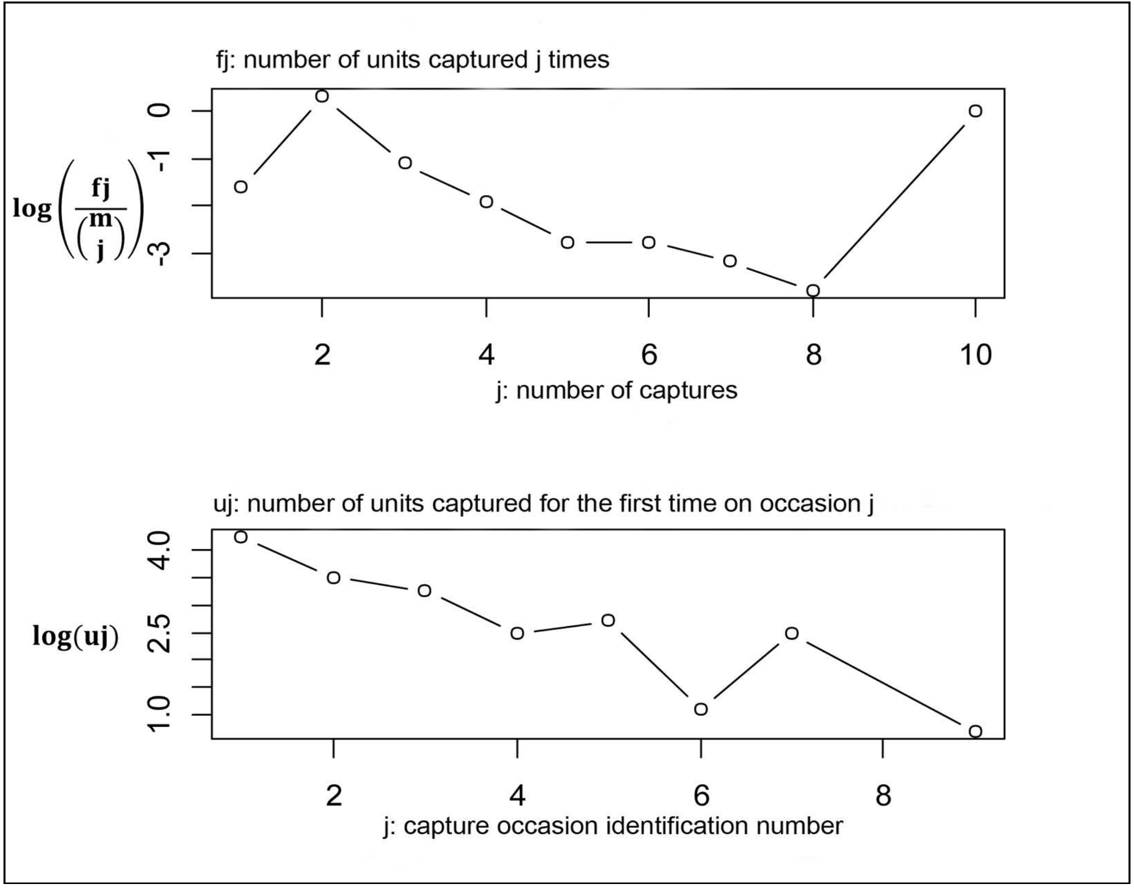
Our third example concerns the captures of the House mice (*Mus musculus*), see Section 1.7.3. For this data set, a total of 173 individuals were captured. The data were recorded including two covariates: age (juvenile, semi-adult or adult) and sex (male or female). Detailed capture information was given as an example in program CAPTURE (Rexstad and Burnham, 1991). We excluded two records in the analysis because the covariates for the two mice were missing. Therefore, $n = 171$ distinct House mice were captured at least once. We considered the groups of juveniles and semi-adults into a 'non-adult' class since there were only 8 juveniles. The data consist of 77 non-adults (45 males, 32 females) and 94 adults (41 males, and 53 females). The numbers of house mice caught for $m = 10$ occasions (n_1 to n_{10}) were 68, 60, 62, 52, 73, 41, 76, 35, 76, 38 and $\sum n_k = 581$. Out of these 581 captures, $n = 171$ distinct animals were captured. The recorded capture frequencies (f_1 to f_{10}) were 2, 62, 40, 31, 16, 13, 5, 1, 0, 1. On average, the capture frequencies for males and females were 3.08 and 3.72, respectively; the capture frequencies for adults and non-adults were 3.81 and 2.90, respectively. The estimated population size, standard error and AIC for model selection using capture-recapture methods is given in Table 4.6. For this data set, Otis et al. (1978) selected model $\mathrm{M_{th}}$ as the most parsimonious model. The f_j plot in Figure 4.3 also suggest that there exist heterogeneity among capture probabilities. Hence, we consider the $\mathrm{M_{th}}$ model for the application of GEE approach in capture-recapture methodology based on this data set.

According to the model formulation of $\mathrm{M_{th}}$, the following equation is useful for the available data,

$$\ln\left(\frac{P_{ij}}{1 - P_{ij}}\right) = \beta_0 + \beta_{age} \times age_{(i)} + \beta_{sex} \times sex_{(i)} + \beta_t \times time_{(j)}; \qquad (4.6)$$

$j = 1, 2, ..., m;\ i = 1, 2, ..., n$, where, β_{sex}, β_{age} and β_t represent the sex, age and

Figure 4.3: Exploratory heterogeneity graph of $\log\bigl(f_j/\binom{m}{j}\bigr)$ and $\log(u_j)$ versus j generated by the Rcapture function of the House mice data



time effect respectively.

The estimation results using GEE approach are summarized in Table 4.7. We observe that the covariates age, sex and time are all significant for a nominal significant level 0.05 (see P-values in Table 4.7). The odds ratios indicate that for any given correlation structure, individuals in the adult age group are more likely to be captured than the ones in the young age group, assuming all other covariates remain fixed. There is about a 46% increase in the odds of capturing for the individual in adult age group than the individual in young age group but slightly lower for autoregressive and pairwise correlation structures. In those cases, there are about 45% increase. The probability of capture appears to be significantly lower for males than for females. According to the odds ratios, the

Table 4.6: Abundance estimations and model fits of House mice data using capture-recapture methods

Models	Abundance	stderr	Deviance	df	AIC
M_0	174.0	1.8	567.562	1021	848.791
M_t	173.7	1.7	507.888	1012	807.117
M_h Chao (LB)	175.0	2.2	552.801	1019	838.030
M_h Poisson2	174.9	2.1	556.666	1020	839.895
M_h Darroch	175.7	2.7	565.209	1020	848.439
M_h Gamma3.5	174.6	2.6	567.386	1020	850.615
M_{th} Chao (LB)	174.8	2.1	491.010	1010	794.239
M_{th} Poisson2	174.6	2.0	495.296	1011	796.525
M_{th} Darroch	175.7	2.7	504.444	1011	805.673
M_{th} Gamma3.5	174.8	2.7	507.371	1011	808.600
M_b	173.0	1.6	566.198	1020	849.427
M_{bh}	173.7	18.8	564.773	1019	850.002

Note: 1 eta parameter has been set to zero in the M_h Chao model

Table 4.7: GEE estimates for *Mus musculus's* capture-recapture data under the various working correlation structures

	Independent Correlation				Exchangeable Correlation			
Cov.	Coff.	Std. Err.	P-value	O.R.	Coff.	Std. Err.	P-value	O.R.
age	.379	.105	0.000	1.461	.379	.102	0.000	1.461
sex	-.232	.104	0.026	0.793	-.232	.100	0.021	0.793
time	-.041	.018	0.024	0.960	-.041	.018	0.024	0.960
cons.	-.545	.138	0.000	···	-.544	.136	0.000	···
	Autoregressive Correlation				Pairwise Correlation			
age	.374	.099	0.000	1.454	.375	.101	0.000	1.454
sex	-.230	.098	0.019	0.794	-.223	.099	0.025	0.801
time	-.039	.017	0.024	0.962	-.044	.018	0.016	0.957
cons.	-.551	.131	0.000	···	-.529	.139	0.000	···

odds of capturing females are about (1/.793) 1.26 times higher than those of males. The risk of capturing decreases for changing one occasion to another occasion that is capture probabilities depend on time which support the findings of Otis et al. (1978). This finding may also suggest that animals are trap shy for this data.

4.4 Model Selection

We have seen that time correlation play an important role when analyzing data sets using GEE approach. Hence, we need to select a correlation structure that builds up the most parsimonious model in GEE analysis. We found that the estimated standard errors of different parameters are not very similar. One may select the best fitting model observing the relative efficiency of the estimated coefficients as GEE approach provides unbiased estimates. Quasi-likelihood information criteria (QIC) is also applicable for selecting best fitting model when GEE approach is used in capture-recapture studies that has been described in the previous chapter. QIC is a modified version of the usual AIC, which allows comparisons of GEE models and selection of a correlation structure (Pan, 2001). Table 4.8 compares the relative efficiencies of the parameters obtained from different models with respect to the ones obtained under the assumption of independence working correlation structure. We also provide the QIC for various working correlation structure models for each data set previously analyzed.

For the Deer mice data, it is shown that for almost all covariates the estimates obtained under independence correlation structure are more efficient as compared to the others except for capture time. The relative efficiency of the estimated coefficient (under independent correlation structure) of capture time is 1.063 in exchangeable correlation structure and 1.214 in pairwise correlation structure. The QIC suggests that pairwise correlation structure model has the smallest QIC (297.81) and thus is chosen as the best fitting model for this data set. Under the pairwise correlation structure, the estimated population is 39.17 with standard error 1.13. Otis et al. (1978) estimated population size for this data set was 41 with standard error 3.05 and Huggins (1991) estimated population size 42.26 with a standard error of 3.75 under model M_0. Huggins (1991) also used his M_{bh} model using covariates and estimated population size 47.144 with standard error

Table 4.8: Quasi-likelihood information criterion (QIC) under various correlation structure models and relative efficiencies of the coefficients under independence correlation structure model for the various data sets

Data	Covariates	Independence	Exchangeable	Autoregressive	Pairwise
Deer mice	QIC	298.75	298.73	298.85	297.81
	QIC_{u}	294.91	294.91	294.91	294.31
	age	1.000	0.812	0.964	0.846
	sex	1.000	0.813	0.964	0.853
	weight	1.000	0.815	0.964	0.855
	time	1.000	1.063	0.977	1.214
L. chip.	QIC	354.54	345.54	345.54	345.10
	QIC_{u}	346.03	346.03	346.03	346.69
	sex	1.000	1.016	1.000	1.166
	time	1.000	0.987	1.000	1.101
House mice	QIC	2173.59	2173.59	2172.61	2173.49
	QIC_{u}	2173.92	2173.92	2172.93	2173.96
	age	1.000	1.029	1.061	1.040
	sex	1.000	1.040	1.061	1.051
	time	1.000	1.000	1.059	1.000

7.18. In the Least chipmunk data, Table 4.8 clearly shows that pairwise correlation structure model is the best fitting model considering relative efficiencies and QIC. We estimated population size 52.27 with standard error 2.87 under pairwise correlation structure model for the Least chipmunk data. Wang et al. (2007) applied a Bayesian approach and estimated population size of 50 with a standard error of 3.14 for the same data set. All the covariates are more efficient under autoregressive correlation structure according to the relative efficiencies for the House mice data. The model selection criterion QIC also suggest that autoregressive correlation structure model may be the best choice for this data set. The estimated population size is 175.08 with standard error 2.07 for the House mice data considering autoregressive correlation structure. Huggins (1989) examined this data set and modelled the individual heterogeneity as a function of the sex and age category of the individuals. He estimated the population size 176.9 with standard error 2.01 from this data set. The discrete-time sample coverage method for model $\mathrm{M_{th}}$ (Chao et al., 1992) yields a population size estimate of 174 with an estimated standard error 1.0; a similar continuous-time approach (Chao and Lee, 1993) gives an estimate of 172 with an estimated standard error 3.3. All these estimation procedures imply that almost no or only few individuals were missed in the capture-recapture experiment. Moreover, the GEE estimation results of the three examples agree with Otis et al.(1978), but in addition to the heterogeneous capture probabilities, our applied GEE approach also accounts for the time dependence.

4.5 Simulation Study

A Monte Carlo (MC) simulation study is carried out to investigate performance of the estimators. We considered the $\mathrm{M_{th}}$ model to evaluate the performance of the GEE approach. We use R (R Development Core Team, 2013) to write a program and run on an Intel(R) Core(TM) i5-3320M CPU computer. We used number of Monte Carlo replicates (B), population size (N), mean capture probability ($\bar{p}$), number of capture occasions (m), correlation coefficient (α), individual and environmental covariates as simulation inputs. Factors used in the simulations were population size, $N = 100$, 200, and 500; mean capture probability, $\bar{p} = 0.1$, 0.3, and 0.5; number of capture occasions, $m = 6$ and 10; and correlation coefficient, $\alpha = -0.5$, -0.3, -0.1, 0, 0.1, 0.3, 0.5. We generated correlated capture history Y_{ij} following the method of Qaqish (2003) considering autoregressive correlation structure. Individuals capture probabilities depended upon the sex (SEX) and weight (WT) of the individual and allowed an environmental covariate (ENV) for each occasion. Following the real data analysis, the simulated individuals were assigned their sex with probability 0.5 and the weights were normally distributed with mean 15 and variance 4. The environmental covariates were normal with mean 2 and variance 1. All the covariates were recomputed for each simulation. One thousand repetitions were completed for each combination of factors.

Statistics computed were average estimate of population size, AVE($\hat{N}$); standard error of the population estimates, SE($\hat{N}$); average captured individuals ($\bar{n}$); percentage relative bias, PRB $= 100 \times (E[\hat{N}] - N) \div N$, where $E[\hat{N}]$ is estimated by AVE($\hat{N}$); percentage coefficient of variation, CV $= 100 \times \text{SE}(\hat{N}) \div E[\hat{N}]$ and confidence intervals (CI). Confidence intervals were computed at the nominal 95% level based on $E[\hat{N}] \pm 1.96 \times \text{SE}(\hat{N})$. The 'best' estimator was chosen by comparing the bias, and root mean square error (RMSE $= \sqrt{\widehat{\text{Var}}(\hat{N}) + \text{bias}^2}$) of all estimators.

The main simulation results are summarized in Table 4.9 and Table 4.10. The GEE approach in capture-recapture studies performs well for estimating population size (N) when there is no linear correlation ($\alpha = 0$) among capture occasions and for the high average capture probability 0.5 ($\bar{p} = 0.5$). For those cases, there is low standard error, absolute value of PRB, coefficient of variation and RMSE. The performance of GEE estimator is poor, and it is very difficult to obtain reliable estimates when the average capture probabilities ($\bar{p} = 0.1$ or 0.3) are low. The estimated population size and its standard error vary depending on the number of capture occasions and strength of linear correlation among capture occasions. For a fixed average capture probability and capture occasions, the estimated population size is higher, but its standard error is lower for negative correlation comparatively to the same strength of positive correlation. The simulation results also show that the GEE approach overestimate the population size for negative linear correlation, and underestimate for positive linear correlation among capture occasions at the high average capture probability ($\bar{p} = 0.5$). There exists an inverse relationship between estimated population size and the linear correlation among capture occasions for the fixed average capture probability and capture occasions. The performance of estimators for $m = 10$ capture occasions is better than for $m = 6$ capture occasions yielding lower CV, absolute value of PRB and RMSE. In general, the simulation study clearly shows that estimator performance of population size and its standard error depends on number of capture occasions, average capture probability, and correlation among capture occasions.

Table 4.9: Under model $\mathrm{M_{th}}$ - simulation results (1000 repetitions) of $m = 6$ capture occasions for the proposed GEE approach assuming autoregressive correlation structure

$\bar{p}$	α	N	$\hat{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	95%CI	PRB	CV	RMSE
0.10	-0.1	100	49.7	67.30	6.78	54.01-80.60	-32.70	7.26	33.39
0.10	0.0	100	46.9	61.77	6.86	48.33-75.22	-38.23	7.17	38.84
0.10	0.1	100	43.7	55.77	6.49	43.06-68.48	-44.23	7.04	44.70
0.10	0.3	100	37.3	44.60	5.74	33.36-55.84	-55.40	6.65	55.70
0.10	0.5	100	30.1	33.74	5.19	23.56-43.92	-66.27	5.95	66.47
0.30	-0.3	100	94.1	104.92	3.07	98.91-110.92	4.92	3.31	5.79
0.30	-0.1	100	90.7	99.88	3.39	93.23-106.53	-0.12	3.18	3.40
0.30	0.0	100	88.6	96.92	3.73	89.61-104.23	-3.08	3.12	4.84
0.30	0.1	100	85.9	93.30	4.18	85.10-101.49	-6.71	3.03	7.90
0.30	0.3	100	79.2	84.24	4.44	75.54-92.94	-15.76	2.74	16.37
0.30	0.5	100	70.6	73.37	4.84	63.88-82.87	-26.63	2.29	27.06
0.50	-0.5	100	99.9	101.62	0.34	100.96-102.28	1.62	1.28	1.66
0.50	-0.3	100	99.8	101.37	0.59	100.22-102.53	1.37	1.26	1.49
0.50	-0.1	100	99.1	100.66	1.03	98.64-102.67	0.66	1.24	1.22
0.50	0.0	100	98.6	100.11	1.29	97.58-102.64	0.11	1.24	1.30
0.50	0.1	100	97.6	99.06	1.60	95.91-102.20	-0.94	1.20	1.86
0.50	0.3	100	94.5	95.63	2.34	91.04-100.21	-4.37	1.10	4.96
0.50	0.5	100	89.3	90.03	3.00	84.15-95.91	-9.97	0.93	10.41
0.10	-0.1	200	99.7	134.78	9.81	115.56-154.01	-32.61	5.11	65.95
0.10	0.0	200	93.9	123.21	9.53	104.52-141.89	-38.40	5.04	77.39
0.10	0.1	200	87.6	111.47	9.25	93.34-129.60	-44.27	4.95	89.01
0.10	0.3	200	74.9	89.19	8.26	73.01-105.37	-55.41	4.63	111.12
0.10	0.5	200	60.7	67.34	7.17	53.28-81.40	-66.33	4.02	132.85
0.30	-0.3	200	188.2	209.51	4.26	201.15-217.87	4.75	2.32	10.42
0.30	-0.1	200	181.6	199.90	4.85	190.40-209.40	-0.05	2.24	4.85
0.30	0.0	200	176.8	193.27	5.27	182.94-203.60	-3.36	2.19	8.55
0.30	0.1	200	171.9	186.19	5.69	175.05-197.33	-6.91	2.11	14.94
0.30	0.3	200	158.8	168.51	6.11	156.54-180.48	-15.75	1.91	32.08
0.30	0.5	200	141.2	146.41	6.72	133.24-159.58	-26.80	1.58	54.01
0.50	-0.5	200	199.9	203.19	0.46	202.28-204.09	1.59	0.90	3.22
0.50	-0.3	200	199.5	202.69	0.87	200.99-204.39	1.35	0.89	2.83
0.50	-0.1	200	198.3	201.38	1.39	198.66-204.10	0.69	0.87	1.96
0.50	0.0	200	197.1	200.06	1.81	196.51-203.61	0.03	0.86	1.82
0.50	0.1	200	195.3	198.05	2.26	193.63-202.47	-0.98	0.84	2.98
0.50	0.3	200	189.4	191.55	3.24	185.20-197.90	-4.23	0.76	9.05
0.50	0.5	200	178.6	179.88	4.44	171.18-188.58	-10.06	0.63	20.60
0.10	-0.1	500	248.9	336.38	15.85	305.30-367.45	-32.72	3.23	164.39
0.10	0.0	500	234.6	307.51	15.05	278.01-337.01	-38.50	3.18	193.08
0.10	0.1	500	219.3	278.60	14.63	249.93-307.26	-44.28	3.12	221.89
0.10	0.3	500	187.2	221.95	13.30	195.87-248.03	-55.61	2.89	278.37
0.10	0.5	500	151.9	167.64	11.07	145.94-189.34	-66.47	2.49	332.54
0.30	-0.3	500	470.5	523.52	6.70	510.39-536.65	4.70	1.47	24.45
0.30	-0.1	500	453.4	498.60	7.74	483.43-513.77	-0.28	1.42	7.87
0.30	0.0	500	442.1	482.75	8.42	466.25-499.25	-3.45	1.38	19.20
0.30	0.1	500	429.8	464.98	8.85	447.63-482.33	-7.00	1.33	36.12
0.30	0.3	500	397.4	421.29	9.90	401.89-440.70	-15.74	1.20	79.33
0.30	0.5	500	352.5	365.01	10.93	343.59-386.43	-27.00	0.99	135.43
0.50	-0.5	500	499.8	507.89	0.78	506.35-509.42	1.58	0.57	7.92
0.50	-0.3	500	498.7	506.66	1.34	504.04-509.27	1.33	0.56	6.79
0.50	-0.1	500	495.6	503.14	2.29	498.65-507.62	0.63	0.55	3.88
0.50	0.0	500	492.9	500.08	2.83	494.54-505.63	0.02	0.54	2.83
0.50	0.1	500	488.3	494.98	3.59	487.95-502.02	-1.00	0.53	6.17
0.50	0.3	500	473.4	478.55	5.24	468.29-488.81	-4.29	0.48	22.08
0.50	0.5	500	446.1	449.17	7.06	435.34-463.00	-10.17	0.39	51.32

Table 4.10: Under model $\mathrm{M_{th}}$ - simulation results (1000 repetitions) of $m = 10$ capture occasions for the proposed GEE approach assuming autoregressive correlation structure

$\bar{p}$	α	N	$\hat{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	95%CI	PRB	CV	RMSE
0.10	-0.1	100	66.5	86.62	6.30	74.28-98.96	-13.38	5.53	14.79
0.10	0.0	100	62.9	80.27	5.97	68.57-91.97	-19.73	5.44	20.62
0.10	0.1	100	59.5	74.49	5.88	62.96-86.02	-25.51	5.33	26.18
0.10	0.3	100	51.1	61.33	5.81	49.94-72.72	-38.67	5.02	39.11
0.10	0.5	100	41.6	47.62	5.40	37.05-58.20	-52.38	4.36	52.65
0.30	-0.3	100	97.2	102.18	1.05	100.12-104.23	2.18	1.71	2.42
0.30	-0.1	100	96.2	100.98	1.48	98.08-103.88	0.98	1.68	1.77
0.30	0.0	100	95.4	100.06	1.70	96.72-103.40	0.06	1.65	1.71
0.30	0.1	100	94.1	98.58	2.06	94.56-102.61	-1.42	1.60	2.50
0.30	0.3	100	90.7	94.60	2.67	89.36-99.83	-5.40	1.47	6.03
0.30	0.5	100	83.9	87.10	3.61	80.02-94.18	-12.90	1.23	13.39
0.50	-0.5	100	98.0	100.11	0.02	100.07-100.15	0.11	0.33	0.11
0.50	-0.3	100	98.0	100.11	0.05	100.00-100.21	0.11	0.33	0.12
0.50	-0.1	100	98.0	100.08	0.18	99.72-100.43	0.08	0.33	0.20
0.50	0.0	100	97.9	100.03	0.28	99.49-100.58	0.03	0.32	0.28
0.50	0.1	100	97.8	99.93	0.43	99.09-100.76	-0.07	0.32	0.43
0.50	0.3	100	97.2	99.29	0.92	97.49-101.09	-0.71	0.31	1.16
0.50	0.5	100	95.0	97.07	1.64	93.86-100.28	-2.93	0.27	3.36
0.10	-0.1	200	135.0	173.00	8.77	155.80-190.19	-13.50	3.90	28.40
0.10	0.0	200	128.7	161.38	8.68	144.36-178.39	-19.31	3.82	39.59
0.10	0.1	200	121.0	148.61	8.58	131.80-165.43	-25.69	3.75	52.10
0.10	0.3	200	104.3	121.89	8.17	105.87-137.90	-39.06	3.48	78.54
0.10	0.5	200	84.8	94.22	7.74	79.04-109.40	-52.89	3.01	106.07
0.30	-0.3	200	196.3	204.24	1.50	201.30-207.17	2.12	1.21	4.49
0.30	-0.1	200	194.3	201.73	2.10	197.62-205.83	0.86	1.17	2.71
0.30	0.0	200	192.7	199.80	2.34	195.22-204.39	-0.10	1.15	2.35
0.30	0.1	200	190.5	197.20	2.90	191.52-202.88	-1.40	1.12	4.03
0.30	0.3	200	183.1	188.79	3.91	181.14-196.45	-5.61	1.02	11.87
0.30	0.5	200	170.1	174.35	5.03	164.49-184.21	-12.83	0.85	26.14
0.50	-0.5	200	198.0	200.22	0.03	200.15-200.28	0.11	0.23	0.22
0.50	-0.3	200	198.0	200.21	0.07	200.07-200.35	0.10	0.23	0.22
0.50	-0.1	200	197.9	200.15	0.24	199.69-200.61	0.08	0.23	0.28
0.50	0.0	200	197.8	200.04	0.40	199.26-200.82	0.02	0.22	0.40
0.50	0.1	200	197.6	199.82	0.63	198.58-201.06	-0.09	0.22	0.66
0.50	0.3	200	196.3	198.52	1.34	195.90-201.14	-0.74	0.21	2.00
0.50	0.5	200	192.0	194.12	2.36	189.49-198.75	-2.94	0.18	6.34
0.10	-0.1	500	340.5	431.38	13.71	404.51-458.24	-13.73	2.45	69.98
0.10	0.0	500	323.9	402.25	13.42	375.94-428.55	-19.55	2.41	98.67
0.10	0.1	500	305.8	371.16	14.00	343.72-398.60	-25.77	2.36	129.60
0.10	0.3	500	264.1	304.75	12.95	279.38-330.13	-39.05	2.19	195.68
0.10	0.5	500	214.5	234.28	12.29	210.19-258.38	-53.14	1.87	266.00
0.30	-0.3	500	493.9	510.55	2.25	506.14-514.97	2.11	0.76	10.79
0.30	-0.1	500	489.0	504.44	3.29	497.99-510.89	0.89	0.74	5.53
0.30	0.0	500	484.7	499.33	3.90	491.68-506.98	-0.13	0.72	3.96
0.30	0.1	500	479.0	492.58	4.70	483.37-501.79	-1.49	0.70	8.78
0.30	0.3	500	460.5	471.31	6.30	458.96-483.66	-5.74	0.63	29.38
0.30	0.5	500	428.1	435.24	7.72	420.11-450.37	-12.95	0.52	65.22
0.50	-0.5	500	498.0	500.53	0.05	500.43-500.63	0.11	0.15	0.53
0.50	-0.3	500	498.0	500.50	0.14	500.23-500.77	0.10	0.14	0.52
0.50	-0.1	500	497.9	500.35	0.40	499.56-501.14	0.07	0.14	0.54
0.50	0.0	500	497.6	500.10	0.59	498.94-501.25	0.02	0.14	0.60
0.50	0.1	500	497.1	499.54	0.96	497.65-501.42	-0.09	0.14	1.07
0.50	0.3	500	493.8	496.24	2.09	492.15-500.32	-0.75	0.13	4.30
0.50	0.5	500	483.2	485.47	3.87	477.89-493.04	-2.91	0.11	15.04

4.6 Discussion

Generalized estimating equations (GEE) approach is an important tool for getting unbiased estimates in the analysis of capture-recapture correlated data when capture probabilities are modelled as a function of covariates. The GEE approach has been applied for adjusting capture probabilities of heterogeneous population and accounting for correlation structures on capture occasions. Estimation results and model selection criteria (QIC) show that the best GEE model for a given working correlation structure in capture-recapture $\mathrm{M_{th}}$ model depends on the data set. The GEE estimator performs well when the capture probabilities are high but the estimates seem to be unreliable for low capture probabilities. Simulation results also reveal that estimated population parameters vary on the on the nature of existing correlation among capture occasions, the number of capture occasions and average capture probability. We agree with Hwang and Huggins (2005) suggestion that estimators dealing with heterogeneity of capture probabilities should be considered. Moreover, it is also important to consider the correlation structure among capture occasions.

Chapter 5

Estimation of Capture Probabilities Using GEE and Mixed Effect Approaches

Part of this chapter is in the form of publication and conference presented paper:

Akanda, M.A.S. and Alpizar-Jara, R. (2014). 'Estimation of Capture Probabilities Using Generalized Estimating Equations and Mixed Effects Approaches.' *Ecology and Evolution*, 4(7), pp.1158–1165.

Akanda, M.A.S. and Alpizar-Jara, R. (2013). 'Generalized Estimating Equations and Mixed Effect Modeling Approaches in Capture-Recapture Closed Population Models.' *Resumos de XXI Congresso da Sociedade Portuguesa de Estatistica*, November 29 – December 2, 2013, Aveiro, Portugal, pp.233–235.

Akanda, M.A.S. and Alpizar-Jara, R. (2013). 'Estimation of Capture Probabilities Using GEE and Mixed Effect Approaches.' *Abstract of EURING 2013 Analytical Meeting & Workshop*, April 28–May 4, 2013, Athens, Georgia, USA, pp.44.

In this chapter, a proposed generalized estimating equations (GEE) and generalized linear mixed modelling (GLMM) approaches are used to estimate capture probabilities and population size for closed population models. We account for individual heterogeneity in capture probabilities, modelling capture probabilities as a function of individual covariates. Generalized linear models (GLM) assumes independence, GEE account for dependency among capture occasions and GLMM are useful to accommodate heterogeneity due to unmeasurable individual characteristics. The purpose of this chapter is to compare the results of estimating population size and its standard error by using these methodologies in the analysis of a capture-recapture data. For illustrative purposes, we analyze several data sets that has already been discussed in the literature, and compare results with currently used methodology. We build-up various modelling structures using individual covariates in our analysis. Conditional arguments are used to obtain a Horvitz-Thompson-like estimator for estimating population size. A simulation study is also conducted to show the performance of the estimation procedures.

5.1 Background

The estimation of the size of a closed wildlife population in capture-recapture experiment is a vital issue in ecological statistics. The population is assumed to be closed over the course of the trapping experiment, implying that there are no births, deaths, and migrations during the study period. In this chapter, we are interested in estimating the population size and standard error of a sub-model of the type $\mathrm{M_h}$, where individual heterogeneity can be modelled as a function of covariates. Development of capture-recapture models dealing with individual heterogeneity in capture probabilities has been one of the most challenging tasks. Failure to account for such heterogeneity has long been known to cause substantial bias in population estimates (Otis et al., 1978; Lee and Chao, 1994; Hwang and

Huggins, 2005). Moreover, Link (2003) shows that without strong assumptions on the underlying distribution, estimates of population size under model $\mathrm{M_h}$ is fundamentally non-identifiable.

The use of covariates (or auxiliary variables), if available, has been proposed as an alternative way to partially cope with the problem of heterogeneous capture probabilities (Pollock et al., 1984; Huggins, 1989; Alho, 1990; Pollock, 2002). The idea is to model capture probabilities as a function of individual (i.e., age, sex, and weight) and environmental (i.e., temperature, rainfall, and location) covariates, using a generalized linear modeling (GLM) approach, such as logistic regression. The method of Huggins (1989, 1991), based on a conditional likelihood to estimate population size, has become very popular, but it assumes independence among capture occasions (Huggins and Hwang, 2011). Hwang and Huggins (2007) also state that the assumption of independence among capture occasions is often violated in practice but the authors still rely on that assumption. Some sort of dependencies among capture occasions deal through the modelling of behavioral effects such as trap happiness and trap shyness being special cases in the capture-recapture literature (Yang and Chao, 2005; Pradel and Sanz-Aguilar, 2012).

One alternative approach is to use a generalized estimating equations (GEE) to account for a working correlation structure over capture occasions (Liang and Zeger, 1986), and use individual measurable characteristics to model heterogeneity in capture probabilities. A mixed effects modelling approach may also be used to model heterogeneity of individual measurable and unmeasurable characteristics in capture-recapture experiments motivating the use of generalized linear mixed models (GLMM) (Pinheiro and Bates, 2000). The GLMM can be used for estimating capture probabilities, which has a linear predictor consisting of fixed effects relating to measurable covariates and random effects relating to unmeasurable covariates. The random effects structure consisting of a random intercept is used to account for correlation among different observations on the same individual or among observations on different individuals (e.g., Breslow and Clayton, 1993 and references therein). The need to consider random effects beside any fixed effect in capture-recapture models is well emphasized by Barry et al. (2003) and Gimenez and Choquet (2010). Some authors have previously introduced the use of GLMM (logit models with normal random effects) (e.g., Coull and Agresti, 1999, 2000; Stoklosa et al., 2011). Both techniques are suitable for the analysis of capture-recapture correlated data since in GEE and GLMM methods a correction is made for the dependency among capture occasions and taking into account the individual heterogeneity of capture probabilities. An advantage of using GLMM for the estimation of capture probabilities is to accommodate not only the heterogeneity attributed to individual characteristics, but also the heterogeneity that cannot be explained by the observed individual characteristics. A possible advantage of GEE over random-effects models and Bayesian methods relates to the ability of GEE to allow specific correlation structures to be assumed between capture occasions. Here we propose a GEE approach for estimating capture probabilities and population size in capture-recapture closed population studies. We also compare the results of population size estimates and their standard errors, when using the two estimation methodologies (i.e., GEE and GLMM). A two-fold estimation procedure is build-up. The first step estimates regression parameters associated with capture probabilities depending on individual covariates. The second step estimates population size using a Horvitz-Thompson-like estimator. In the next section, we describe the notation and models that are used to estimate capture probabilities and population size. Section 5.3, illustrates the methodologies for several real data sets. A simulation study is presented in Section 5.4. We conclude with some discussions in Section 5.5.

5.2 Notation and Models

Consider a population consisting of N animals in a capture-recapture experiment over m capture occasions, $j = 1, 2, ..., m$. Let Y_{ij} be a binary outcome, equalling 1 if the i^{th} animal is being caught on the j^{th} capture occasion and 0 otherwise. Let $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{im})'$ be a random vector with the capture history of individual i. Let $T_i = \sum_{j=1}^{m} Y_{ij}$ be the number of times the i^{th} animal has been caught in the course of the trapping closed population study. Let t_i be the time the i^{th} individual is first captured. Heterogeneity in captured probabilities is often explained by individual measurable covariate z_i, such as age, sex, weight, etc. For simplicity, we consider z_i a single covariate, but the model can be easily generalized for z_i to be considered a vector of covariates. Let the probability that the i^{th} animal is captured on any trapping occasion j, be

$$P_i(\beta) = \Pr(Y_{ij} = 1|X_i) = h(X_i\beta); i = 1, 2, ..., N, j = 1, 2, ..., m \qquad (5.1)$$

where

$$X_i = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_i & z_i & \dots & z_i \end{bmatrix}' ; (i = 1, 2, ..., N)$$

is the design matrix, $\beta = (\beta_0, \beta_1)'$ is the vector of parameters associated with the covariates, and $h(u) = (1 + \exp(-u))^{-1}$ is the logistic function. This is an $\mathrm{M_h}$ model where variation in capture probabilities among individuals is explained by the covariate z_i. The probability of not capturing the i^{th} individual on the j^{th} occasion is $(1 - P_i(\beta))$ and the variance of Y_{ij} is $P_i(\beta)(1 - P_i(\beta))$ (Liang and Zeger, 1986). Then $T_i \sim Bin(m, P_i(\beta))$ and $\pi_i(\beta) = 1 - (1 - P_i(\beta))^m$ is the probability of individual i being captured at least once, given the covariate z_i. Let the set of distinct individuals captured at least in one occasion be indexed by $i = 1, 2, ..., n$ and uncaptured individuals would be indexed by $i = n + 1, ..., N$ without loss of generality. To estimate the population size, once an estimator $\hat{\beta}$ of β is available, the Horvitz-Thompson estimator $\hat{N} = \sum_{i=1}^{n} \dfrac{1}{\pi_i(\hat{\beta})}$ may be used as in Huggins (1989).

5.2.1 Methods Based on Partial Likelihood

The full likelihood is proportional to

$$\prod_{i=1}^{n} \frac{P_i(\beta)^{T_i}\{1 - P_i(\beta)\}^{m-T_i}}{\pi_i(\beta)} \prod_{i=1}^{n} \pi_i(\beta) \prod_{i=n+1}^{N} \{1 - \pi_i(\beta)\}. \tag{5.2}$$

Since the number of total individuals, N, is unknown and the covariates are not known for individuals that are never captured, this likelihood cannot be directly evaluated. The conditional likelihood (Huggins, 1989) is the first product component of (5.2) and it can be formulated as a GLM (Huggins and Hwang, 2011) for the positive Binomial distribution (Patil, 1962). It may be rewritten as

$$\prod_{i=1}^{n} P_i(\beta)^{T_i-1}\{1 - P_i(\beta)\}^{m-t_i-(T_i-1)} \prod_{i=1}^{n} \left[\frac{\{1 - P_i(\beta)\}^{t_i-1} P_i(\beta)}{\pi_i(\beta)}\right]. \tag{5.3}$$

When the full likelihood is partitioned into a product of conditional densities, then a partial likelihood (Cox, 1975) may arise considering some of the product terms, but it involves only the parameters of interest, isolating the nuisance parameters. Therefore, the partial likelihood, PL(β) is the first product of the equation (5.3), which is the likelihood of the number of recaptures after the first capture (Stoklosa et al., 2011). For a given t_i, $(T_i - 1)|t_i \sim Bin(m - t_i, P_i(\beta))$, which is used to estimate the parameters β.

• Generalized Linear Models

To apply GLM we suppose that $P_i(\beta) = h(X_i\beta)$ for a vector of parameters β, where X_i is the design matrix for fixed measurable covariates as defined earlier. Hence, we can use partial likelihood conditioning on the first capture time. Estimation of β is conducted via an iterative procedure known as iterative re-weighted least squares (IRLS). The estimation procedure is based on maximizing the penalized partial log-likelihood.

Let $\hat{\beta}$ be the estimate of β using partial likelihood and $P_i(\hat{\beta}) = h(X_i\hat{\beta})$ then we get an estimate of the variance of $\hat{N}$ is given by

$$\widehat{\mathrm{Var}}(\hat{N}) = \sum_{i=1}^{n} \pi_i(\beta)^{-2}\big(1 - \pi_i(\beta)\big) + \Delta(\beta)' \mathit{\Gamma}(\beta)^{-1}\Delta(\beta) \tag{5.4}$$

where $\mathit{\Gamma}(\beta)$ represents the conditional information matrix for β and $\Delta(\beta)$ is the vector of $\sum\limits_{i=1}^{n} \pi_i(\beta)^{-2}\partial\pi_i(\beta)/\partial\beta$ with all quantities evaluated at $\hat{\beta}$.

• Generalized Linear Mixed Models

To utilize a simple GLMM with a random effect, we suppose that $P_i(\beta) = h(X_i\beta + \sigma_b\nu_i)$ where ν_i is a realization of the standard normal random variable $Z_i \sim \mathcal{N}(0, 1)$, with $\sigma_b > 0$. One can think of Z_i as an unmeasured covariate, as a way to model heterogeneity, or as a way to model correlated data. The use of random effects reflects the belief that there is heterogeneity that can not be explained by covariates. This type of GLMM model deserves special attention for several reasons. First allows modeling of the time variation (or occasion-to-occasion variation) separately for each individual. Second we model the individual variation by postulating a distribution. The partial likelihood can be considered as the joint distribution of the response and the random effects. To estimate β and σ_b, the marginal likelihood of the response is obtained by integrating out the random effects. The integration can be approximated by penalized quasi-likelihood (Breslow and Clayon, 1993), which enables parameter estimation via an iterative procedure. The variance of $\hat{N}$ for a smoothing parameter λ may be estimated according to Stoklosa et al. (2011) by using the following formula, $\widehat{\mathrm{Var}}(\hat{N}, \lambda) = \sum\limits_{i=1}^{n} \dfrac{1 - \pi_i(\beta)}{\pi_i(\beta)^2} + \{X_i\eta(\beta)\}'\mathrm{Var}(\beta)\{X_i\eta(\beta)\}$, where $\eta(\beta)$ is a vector with $\eta_i(\beta) = \pi_i(\beta)^{-2} m P_i(\beta)\{1 - \pi_i(\beta)\}$, and all quantities are evaluated at $\hat{\beta}$. The smoothing parameter λ, which is part of the quasi-likelihood procedure, controls the degree of roughness of the estimated functions. To obtain an optimal value for λ we used generalized cross validation (GCV) technique (Wood, 2006).

5.2.2 Generalized Estimating Equations Approach

Let $V_i = A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$ be the covariance matrix of Y_i, where, $A_i = \text{diag}[\text{Var}(Y_{i1}), \text{Var}(Y_{i2}), ..., \text{Var}(Y_{im})]$ is a $m \times m$ diagonal matrix and $R_i(\alpha)$ is known as the working correlation structure among $Y_{i1}, Y_{i2}, ..., Y_{im}$ to describe the average dependency of individuals being captured from occasion to occasion. A GEE approach permits several types of working correlation structure $R_i(\alpha)$ (for details, see Diggle et al., 2002). For the description that follows and for simplicity, we consider an independence working correlation structure, $R_i(\alpha) = \text{I}$ where I is an identity matrix. The covariate z_i is never known for the individuals that have not been captured. Therefore, Y_{ij} is conditional on the captured individuals (n) (i.e., $T_i \geq 1$) with the corresponding observed individual covariates like Huggins (1989) and Zhang (2012). The probability that the i^{th} individual is captured on the j^{th} occasion (P_{ij}) given that the i^{th} individual is observed at least once is, $\Pr(Y_{ij} = 1 | T_i \geq 1) = P_{ij} / \big(1 - \prod_{k=1}^{m} (1 - P_{ik})\big)$. Let $\mu_{ij} = \text{E}(Y_{ij} | T_i \geq 1) = P_{ij} / \big(1 - \prod_{k=1}^{m} (1 - P_{ik})\big)$, and D_i be the matrix of derivatives $\partial \mu_i / \partial \beta'$, where, $\mu_i = (\mu_{i1}, \mu_{i2}, ..., \mu_{im})'$, hence $D_i = A_i X_i$. The variance v_{ij} of Y_{ij} given $T_i \geq 1$ is $v_{ij} = \text{Var}(Y_{ij} / T_i \geq 1) = P_{ij} \big(1 - P_{ij} - \prod_{k=1}^{m} (1 - P_{ik})\big) / \big[1 - \prod_{k=1}^{m} (1 - P_{ik})\big]^2$. Considering, $V_i = \text{diag}(v_{ij})$, an estimator of β can be obtained by solving the following generalized estimating equations:

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1} (Y_i - \mu_i) = 0. \qquad (5.5)$$

For a given $\hat{\beta}$, then $\hat{\pi}_i(\hat{\beta}) = 1 - (1 - P_i(\hat{\beta}))^m$ and we can estimate the variance of $\hat{N}$ using equation (5.4). If the individual capture probability does not depend on time, individual prior capture history and any covariates, the model (5.1) simplifies to $P_i(\beta) = h(\beta_0) = P_0$, which is a re-parameterization of model M_0 of Otis et al. (1978) (see Huggins, 1991, Hwang and Huggins, 2005). This model assumes all the individuals have equal capture probabilities. Then the estimating equations for β_0 is simplified to (derivation is given in Appendix A)

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{m} Y_{ij} - \frac{mP_0}{1-(1-P_0)^m}\right) = 0. \tag{5.6}$$

Let $\hat{\beta}_0$ be the resulting estimator of β_0 then $\hat{\pi}_0 = 1 - (1 - \hat{P}_0)^m$ where $\hat{P}_0 = h(\hat{\beta}_0)$.

5.3 Applications

We apply the techniques discussed in the previous Section to real life data sets.

5.3.1 Example 1: Deer mice data

The data set from Section 1.7.1 are considered for our first example. For these data, $n = 38$ distinct Deer mice *(Peromyscus maniculatus)* were captured at least once over $m = 6$ capture occasions. The estimation results for the applied models are summarized in Table 5.1. Based on the Akaike information criterion (AIC), the model constitute by age, sex, and weight (Model 10) may be preferred because of lower AIC (98.03) within all partial likelihood GLM models (Model 1, Model 4, Model 5, Model 6, Model 7 and Model 9 produced AIC 108.02, 107.57, 99.33, 109.94, 100.67, 100.99 and 102.59 respectively). Here, we get the estimated population size 38.87 with standard error 1.16. The quasi-likelihood GEE and partial likelihood GLMM provide similar results except for two cases in GLMM, (i) Model 19 for sex and random variable, and (ii) Model 22 for sex, weight and random variables. In both cases the estimated population size is 47.95 with standard error 5.48 for model 19 and 5.50 for model 22. Estimation results clearly show that the quasi-likelihood GEE models provide consistent results yielding lower standard error compared to the results of GLM and GLMM models for this data set, see Table 5.1.

5.3.2 Example 2: Least chipmunk data

Our second example concerns the captures of Least chipmunks (*Eutamias minimus*) from Section 1.7.2. Over the $m = 6$ capture occasions, $n = 45$ distinct animals were captured at least once. The estimation results are summarized in Table 5.2. The inclusion of the covariate sex does not necessary improve our estimates of population size which are very similar, except when the random effect is considered in the GLMM, which is based on partial likelihood estimation. This may indicated that there is unmodeled individual heterogeneity in capture probabilities that is not being accounted for with the other models (GLM and GEE). The population estimate, in this case, is of approximately 74 individuals with a standard error of 12. Both values are quite high when compared to the values obtained with the other estimation strategies. Although, GLMM accounts for heterogeneity due to unmeasured individual characteristics, it may also be overestimating population size at the expenses of greater loss in precision, possibly due to the increase in the number of model parameters that are estimated. In contrast, quasi-likelihood GEE methodology provided lower standard errors, when compared to results from the Bayesian approach of Wang et al. (2007) for the same data set. The latter authors estimated population size of 50 with a standard error of 3.14. The GEE estimation results also agree with Otis et al. (1978), but our model jointly takes into account heterogeneity in capture probabilities and correlation among capture occasions.

Table 5.1: Comparison of parameter estimates (standard error in parenthesis) for Deer mice data after fitting models with and without covariates. A realization of the standard normal random variable $Z_i \sim \mathcal{N}(0,1)$ is ν_i. QL= quasi-likelihood and PL= partial likelihood. Numbers in this table are rounded to two decimal places; therefore 0.00 does not mean zero

Model No.						logit$\{P_i(\beta)\}$					$\hat{N}$
	Intercept-only models										
1.	PL GLM	0.46									38.13
		(0.18)									(0.37)
2.	QL GEE	0.11									38.43
		(0.18)									(0.66)
3.	PL GLMM	0.03							+	0.00 ν_i	38.56
		(0.27)								(0.75)	(0.88)
	Linear covariates models										
4.	PL GLM	0.65	-	0.60 age							38.23
		(0.22)		(0.38)							(0.52)
5.	PL GLM	-0.22			+	1.19 sex					38.52
		(0.27)				(0.37)					(0.83)
6.	PL GLM	0.30					+	0.01 weight			38.13
		(0.58)						(0.04)			(0.37)
7.	PL GLM	-0.08	-	0.33 age	+	1.12 sex					38.59
		(0.33)		(0.40)		(0.38)					(0.90)
8.	PL GLM	-0.56			+	1.21 sex	+	0.02 weight			38.55
		(0.65)				(0.37)		(0.04)			(0.86)
9.	PL GLM	-1.40	-	2.03 age			+	0.18 weight			38.06
		(0.82)		(0.69)				(0.07)			(0.94)
10.	PL GLM	-1.71	-	1.57 age	+	0.99 sex	+	0.15 weight			38.87
		(0.85)		(0.71)		(0.39)		(0.07)			(1.16)
11.	QL GEE	0.35	-	0.77 age							38.83
		(0.17)		(0.28)							(0.93)
12.	QL GEE	-0.49			+	1.01 sex					39.09
		(0.20)				(0.27)					(1.08)
13.	QL GEE	0.21					-	0.01 weight			38.48
		(0.43)						(0.03)			(0.69)
14.	QL GEE	-0.22	-	0.64 age	+	0.92 sex					39.46
		(0.24)		(0.29)		(0.28)					(1.27)
15.	QL GEE	-1.69	-	2.26 age			+	0.18 weight			39.62
		(0.61)		(0.51)				(0.05)			1.37
16.	QL GEE	-0.45			+	1.01 sex	-	0.00 weight			39.10
		(0.47)				(0.27)		(0.03)			(1.08)
17.	QL GEE	-1.95	-	1.90 age	+	0.81 sex	+	0.16 weight			39.85
		(0.14)		(0.15)		(2.24)		(1.17)			(1.45)
18.	PL GLMM	0.27	-	0.70 age					+	0.00 ν_i	38.89
		(0.30)		(0.50)						(0.71)	(1.20)
19.	PL GLMM	-0.46			+	0.85 sex			+	1.44 ν_i	47.95
		(0.36)				(0.48)				(0.00)	(5.48)
20.	PL GLMM	0.16					-	0.01 weight	+	7.62 ν_i	38.53
		(0.92)						(0.06)		(0.72)	(0.87)
21.	PL GLMM	-0.20	-	0.57 age	+	0.76 sex			+	0.00 ν_i	39.28
		(0.43)		(0.52)		(0.50)				(0.70)	(1.57)
22.	PL GLMM	-0.44			+	0.85 sex	-	0.00 weight	+	1.44 ν_i	47.95
		(0.84)				(0.49)		(0.05)		(0.00)	(5.50)
23.	PL GLMM	-1.57	-	2.02 age			+	0.16 weight	+	0.00 ν_i	39.49
		(1.09)		(0.92)				(0.09)		(0.70)	(1.77)
24.	PL GLMM	-1.65	-	1.76 age	+	0.69 sex	+	0.14 weight	+	6.10 ν_i	39.49
		(1.20)		(1.04)		(0.58)		(0.10)		(0.66)	(1.80)

Table 5.2: Comparison of parameter estimates (standard error in parenthesis) for Least chipmunk data after fitting models with and without covariates. A realization of the standard normal random variable $Z_i \sim \mathcal{N}(0,1)$ is ν_i. QL= quasi-likelihood and PL= partial likelihood. Numbers in this table are rounded to two decimal places; therefore 0.00 does not mean zero

Model No.		logit$\{P_i(\beta)\}$					$\hat{N}$
	Intercept-only models						
1.	PL GLM	-0.82					50.72
		(0.18)					(3.33)
2.	QL GEE	-0.73					49.66
		(0.13)					(2.27)
3.	PL GLMM	-0.85			+	0.00 ν_i	51.03
		(0.26)				(0.73)	(4.10)
	Linear covariates models						
4.	PL GLM	-0.81	-	0.03 sex			50.73
		(0.25)		(0.37)			(3.35)
5.	QL GEE	-0.84	-	0.21 sex			52.40
		(0.18)		(0.26)			(2.94)
6.	PL GLMM	-0.83	-	0.14 sex	+	1.59 ν_i	74.16
		(0.34)		(0.49)		(0.00)	(12.06)

5.3.3 Example 3: House mice data

Our third example due to the House mice data set of Coulombe (1965), concerns the captures of 173 *Mus musculus*, see Section 1.7.3. The data recorded associated with two covariates: age (juvenile, semi-adult or adult) and sex (male or female). Detailed capture information was given as an example in program CAPTURE (Rexstad and Burnham, 1991). Two records are excluded in the analysis because the covariates for the two mice were missing. Hence, for this data set, $n = 171$ distinct House mice were captured at least once over $m = 10$ capture occasions. We estimated parameters using GLM, GEE and GLMM approaches in Table 5.3 for intercept only models and linear covariates models. If there were no available information of individual covariate, the partial likelihood GLMM gives very slightly lower population size estimates with lower standard error in comparison to the partial likelihood GLM. The linear covariate model of age and sex (Model 6) is the more parsimonious within all partial likelihood GLM models because it produced lower AIC (547.93) in comparison of other models, which estimate the population size is 175.82 with a standard error of 2.50. Note that Model 1 (intercept only model), Model 4 (covariate model of age) and Model 5 (covariate model of sex) produced AIC 563.17, 549.86 and 559.19 respectively. The population size estimates are almost same for the quasi-likelihood GEE and partial likelihood GLMM for all combination of covariate models. This may be due to the fact that random effect had a negligible effect on the heterogeneity across individuals in GLMM. Table 5.3 also shows that the quasi-likelihood based GEE approach provides lower standard error in comparison with other applied approaches. Huggins (1989) considered a closed population with heterogeneous capture probabilities modelled using covariates and estimated the population size as 176.9 with standard error 2.01 for this dataset. If model $\mathrm{M_h}$ (capture probabilities vary by individual) is assumed, the jackknife procedures of Burnham and Overton (1978) estimated the population size as 175 with standard error 43.4, and the methods of Chao (1988) as implemented in CAPTURE estimated the population size as 173 with standard error 0.18.

Table 5.3: Comparison of parameter estimates (standard error in parenthesis) for House mice data after fitting different models with and without covariates. A realization of the standard normal random variable $Z_i \sim \mathcal{N}(0,1)$ is ν_i. QL= quasi-likelihood and PL= partial likelihood

Model No.		$\text{logit}\{P_i(\beta)\}$							$\hat{N}$
	Intercept-only models								
1.	PL GLM	-0.72							174.34
		(0.06)							(1.96)
2.	QL GEE	-0.66							173.73
		(0.05)							(1.67)
3.	PL GLMM	0.06					-	$0.68\,\nu_i$	173.92
		(0.40)						(0.07)	(1.84)
	Linear covariates models								
4.	PL GLM	-1.01	+	0.48 age					175.47
		(0.10)		(0.12)					(2.38)
5.	PL GLM	-0.58			-	0.30 sex			174.77
		(0.08)				(0.12)			(2.13)
6.	PL GLM	-0.87	+	0.45 age	-	0.24 sex			175.82
		(0.12)		(0.13)		(0.12)			(2.50)
7.	QL GEE	-0.95	+	0.45 age					174.89
		(0.08)		(0.10)					(2.00)
8.	QL GEE	-0.54			-	0.31 sex			174.40
		(0.07)				(0.10)			(1.87)
9.	QL GEE	-0.81	+	0.42 age	-	0.25 sex			175.18
		(0.10)		(0.11)		(0.10)			(2.09)
10.	PL GLMM	-0.91	+	0.41 age			+	$0.03\,\nu_i$	174.52
		(0.10)		(0.13)				(0.37)	(2.07)
11.	PL GLMM	-0.54			-	0.29 sex	+	$0.03\,\nu_i$	174.20
		(0.09)				(0.13)		(0.37)	(1.95)
12.	PL GLMM	-0.77	+	0.37 age	-	0.24 sex	+	$0.03\,\nu_i$	174.72
		(0.12)		(0.13)		(0.13)		(0.36)	(2.14)

5.4 Simulation Study

A simulation study was conducted in order to evaluate the performance of the estimators. The effect of heterogeneity among observed individuals was modelled using two covariates, sex ($male = 1$ and $female = 0$) and $weight$. Four levels of population sizes $N = 50,\ 100,\ 200,$ and 500 and two levels of capture occasions $m = 6$ and 10 were considered. For each individual, we assigned sex with probability 0.5 from a Bernoulli distribution and $weight$ from a normal distribution with mean 15 and variance 4. These values are based on the previous data analysis. Individual capture probabilities were modelled with a logistic regression, such that

$$P_i = \frac{e^{\beta_0 + \beta_1 \times sex_i + \beta_2 \times weight_i}}{1 + e^{\beta_0 + \beta_1 \times sex_i + \beta_2 \times weight_i}}, \tag{5.7}$$

where β_0 is the constant term, β_1 and β_2 represents the sex and $weight$ effects respectively. A positive β_1 implies that the sex taking value 1 is more catchable, and a possitive β_2 means that the catchability increases with $weight$. We considered three different simulation scenarios for capture probabilities, (a) high capture probabilities ($\beta_0 = -3.5$); (b) medium capture probabilities ($\beta_0 = -4.0$); (c) low capture probabilities ($\beta_0 = -4.5$) and their averaged are presented in Table 2. In addition, a Gaussian random effect with mean 0 and $\sigma_b = 0.1$ was included as an unmeasured covariate to ensure the existence of heterogeneity due to unmeasured individual characteristics. For each simulation scenario, GLM, GEE and GLMM approaches were used for data analyses, and to assess estimators performances. The simulation study was carried out with 1,000 Monte Carlo replicates.

We present the average of the resulting 1,000 estimated population size, AVE($\hat{N}$) and the average number of distinct individuals that are captured ($\bar{n}$) in the experiments. To evaluate estimators' performance, we present the standard errors of the estimated population size, SE($\hat{N}$); percentage relative bias, PRB $= 100 \times (\mathrm{E}(\hat{N}) - N) \div N$, where E($\hat{N}$) is estimated by AVE($\hat{N}$); root mean square

Table 5.4: Simulated capture probability scenarios for the capture probability model, $\text{logit}(P_i) = \beta_0 + \beta_1 \times sex + \beta_2 \times weight$. $\bar{p}$ represents average capture probability when $weight = 15$ and π_i represents the average probability of an individual being captured at least once in the whole experiment

Simulation Scenarios	Effects of covariates			$\bar{p}$		π_i			
						$m = 6$		$m = 10$	
	β_0	β_1	β_2	*male*	*female*	*male*	*female*	*male*	*female*
(a) High	-3.5	0.1	0.2	0.40	0.38	0.95	0.94	0.98	0.98
(b) Medium	-4.0	0.1	0.2	0.29	0.27	0.87	0.85	0.94	0.92
(c) Low	-4.5	0.1	0.2	0.20	0.18	0.73	0.70	0.83	0.80

error, RMSE $= \sqrt{\text{Var}(\hat{N}) + \text{Bias}^2}$; percentage coefficient of variation, CV $= 100 \times \text{SE}(\hat{N}) \div \text{E}(\hat{N})$ and confidence interval coverage (%) (COV) for the estimates of population size.

The simulation results for 6 capture occasions are given in Table 5.5. We noticed that all estimation procedures for scenario (a) perform well. There is little bias, low standard errors, low coefficient of variation for $\hat{N}$. In this scenario, confidence interval coverage for all estimators are very good (93-96%), considering a nominal level of 95%. As in our example, the exception is with GLMM that tends to overestimate population size. Overestimation is particularly severe when capture probabilities are low, see for instance, results of scenarios (b) and (c). Confidence interval coverage for GLMM is also poor (77-90%) in these scenarios. For all scenarios, the GEE approach performs well for estimating population size. This approach also consistently provides lower standard errors and lower RMSE when compared to GLM and GLMM estimators, although the differences are minimal for GEE-GLM comparisons. Therefore, our simulation results indicate that the general performance of estimators obtained from GEE are better than GLM and GLMM. The GEE approach may overcome the effect of random effects due to its ability accounting for the correlation structure among capture occasions. The simulation results for 10 capture occasions are presented in Table 5.6. The performance of estimators for 10 capture occasions is better than for 6 capture occasions yielding lower CV, absolute value of PRB, RMSE but higher COV. This is generally true because the average capture probability is higher for 10 capture occasions than for 6 capture occasions. Therefore, simulation results reveal that the general performance of estimator is good if a large proportion of individuals are captured. For cases where only a small proportion of individuals are captured, the estimates become unstable but the GEE approach outperforms the other methods.

Table 5.5: Simulation results (1000 repetitions) considering $m = 6$ capture occasions. Averages of the numbers of captured individuals, ($\bar{n}$); the estimates of population size, AVE($\hat{N}$); standard errors of the estimated population size, SE($\hat{N}$); percentage relative bias, PRB $= 100 \times (\mathrm{E}(\hat{N}) - N) \div N$, where E($\hat{N}$) is estimated by AVE($\hat{N}$); root mean square error, RMSE $= \sqrt{\mathrm{Var}(\hat{N}) + \mathrm{Bias}^2}$; percentage coefficient of variation, CV $= 100 \times \mathrm{SE}(\hat{N}) \div \mathrm{E}(\hat{N})$ and confidence interval coverage (%), COV

(a) High	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV
PL GLM	50	45	50.72	2.92	1.44	5.76	3.01	95.4	200	180	200.46	5.12	0.23	2.55	5.14	94.2
QL GEE	50	45	50.60	2.17	1.20	4.29	2.25	95.9	200	180	200.49	4.02	0.25	2.01	4.04	95.1
PL GLMM	50	45	51.29	3.27	2.59	6.38	3.52	94.6	200	180	202.98	5.95	1.49	2.93	6.65	93.7
PL GLM	100	92	100.63	3.77	0.63	3.75	3.82	94.5	500	460	500.65	7.97	0.13	1.59	8.00	93.2
QL GEE	100	92	100.66	2.90	0.66	2.88	2.97	95.8	500	460	500.87	6.28	0.17	1.25	6.34	95.3
PL GLMM	100	92	101.81	4.30	1.81	4.22	4.67	95.9	500	460	506.56	9.07	1.31	1.79	11.20	93.1
(b) Medium	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV
PL GLM	50	41	51.95	5.85	3.90	11.30	6.17	94.8	200	162	201.71	9.79	0.86	4.85	9.94	94.9
QL GEE	50	41	51.71	3.70	3.42	7.16	4.08	94.6	200	162	201.54	6.59	0.77	3.27	6.77	94.6
PL GLMM	50	41	55.83	7.50	11.66	13.40	9.50	88.5	200	162	211.52	12.28	5.76	5.81	16.80	86.2
PL GLM	100	84	101.56	7.16	1.56	7.05	7.33	94.3	500	421	501.74	14.89	0.35	2.97	15.00	94.6
QL GEE	100	84	101.51	4.82	1.51	4.75	5.05	95.2	500	421	501.92	10.31	0.38	2.05	10.50	95.2
PL GLMM	100	84	106.58	9.06	6.58	8.50	11.20	89.1	500	421	526.33	18.90	5.27	3.59	32.40	83.3
(c) Low	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV
PL GLM	50	36	56.79	14.13	13.58	24.90	15.70	96.0	200	146	204.25	17.78	2.13	8.71	18.30	93.8
QL GEE	50	36	55.59	11.46	11.18	20.60	12.80	95.7	200	146	204.56	10.21	2.28	4.99	11.20	94.3
PL GLMM	50	36	81.26	23.84	62.52	29.30	39.30	78.3	200	146	245.53	26.45	22.77	10.80	52.70	78.5
PL GLM	100	69	104.61	14.01	4.61	13.40	14.80	95.7	500	356	504.24	26.68	0.85	5.29	27.00	95.0
QL GEE	100	69	103.53	11.48	3.53	11.10	12.00	94.6	500	356	503.86	15.45	0.77	3.07	15.90	94.5
PL GLMM	100	69	131.07	21.14	31.07	16.10	37.60	77.2	500	356	576.72	37.06	15.34	6.43	85.20	77.4

Table 5.6: Simulation results (1000 repetitions) considering $m = 10$ capture occasions. Averages of the numbers of captured individuals, ($\bar{n}$); the estimates of population size, AVE($\hat{N}$); standard errors of the estimated population size, SE($\hat{N}$); percentage relative bias, PRB $= 100 \times (\mathrm{E}(\hat{N}) - N) \div N$, where E($\hat{N}$) is estimated by AVE($\hat{N}$); root mean square error, RMSE $= \sqrt{\mathrm{Var}(\hat{N}) + \mathrm{Bias}^2}$; percentage coefficient of variation, CV $= 100 \times \mathrm{SE}(\hat{N}) \div \mathrm{E}(\hat{N})$ and confidence interval coverage (%), COV

(a) High	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV
PL GLM	50	48	50.15	1.07	0.30	2.13	1.08	96.8	200	192	200.25	2.04	0.13	1.02	2.06	96.7
QL GEE	50	48	50.13	0.95	0.26	1.90	0.96	96.4	200	192	200.08	1.83	0.04	0.91	1.83	96.3
PL GLMM	50	48	50.16	1.10	0.32	2.19	1.11	94.6	200	192	200.18	2.05	0.09	1.02	2.06	94.7
PL GLM	100	98	100.11	1.43	0.11	1.43	1.43	94.3	500	492	500.20	3.11	0.04	0.62	3.11	95.1
QL GEE	100	98	100.14	1.36	0.14	1.35	1.36	96.3	500	492	500.18	3.03	0.04	0.61	3.03	96.2
PL GLMM	100	98	100.15	1.45	0.15	1.44	1.45	94.6	500	492	500.28	3.19	0.06	0.64	3.20	94.9
(b) Medium	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV
PL GLM	50	44	50.39	2.34	0.78	4.64	2.37	95.9	200	178	200.24	4.28	0.12	2.14	4.29	95.4
QL GEE	50	44	50.43	1.88	0.86	3.73	1.93	95.8	200	178	200.37	3.57	0.19	1.78	3.59	95.0
PL GLMM	50	44	50.72	2.54	1.44	5.01	2.64	94.2	200	178	201.16	4.67	0.58	2.32	4.81	94.5
PL GLM	100	95	100.47	3.14	0.47	3.12	3.17	95.2	500	473	500.76	6.71	0.15	1.34	6.75	94.6
QL GEE	100	95	100.42	2.98	0.42	2.97	3.01	96.5	500	473	500.66	6.35	0.13	1.27	6.38	96.1
PL GLMM	100	95	100.92	3.32	0.92	3.29	3.45	93.4	500	473	502.03	7.20	0.41	1.43	7.48	94.1
(c) Low	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV	N	$\bar{n}$	AVE($\hat{N}$)	SE($\hat{N}$)	PRB	CV	RMSE	COV
PL GLM	50	38	51.13	4.77	2.26	9.33	4.90	95.1	200	152	201.07	8.54	0.53	4.25	8.61	95.4
QL GEE	50	38	51.27	3.34	2.54	6.51	3.57	95.4	200	152	201.02	6.14	0.51	3.05	6.22	94.8
PL GLMM	50	38	53.01	5.58	6.02	10.53	6.34	87.9	200	152	207.42	10.06	3.71	4.85	12.5	88.3
PL GLM	100	86	101.25	6.18	1.25	6.11	6.31	96.4	500	431	500.98	13.04	0.20	2.60	13.08	95.0
QL GEE	100	86	101.31	5.97	1.31	5.89	6.11	94.2	500	431	500.65	12.57	0.13	2.51	12.58	95.4
PL GLMM	100	86	104.71	7.35	4.71	7.02	8.73	88.6	500	431	512.15	15.21	2.43	2.97	19.46	88.7

5.5 Discussion

Individual heterogeneity and time dependence are fundamentally important in real life applications of capture-recapture studies. This work allows to permit capture-recapture data analysis using individual covariates, that account for heterogeneity in capture probabilities and dependency among capture occasions. The main purpose of this chapter was to compare the performances of estimating population size and its standard error using statistical techniques such as, quasi-likelihood for GEE, and partial likelihood for GLM and GLMM. Furthermore, estimation results of different model structures using covariates are compared with each others.

There are several issues related to the applied techniques and estimation procedures.

First, it is not straightforward to interpret the magnitude of the regression coefficients obtained from either GEE or GLMM analysis. Fundamentally both methods produce ‘pooled’ regression coefficients accounting for time and individual heterogeneity relationships. This has the following implications for the interpretation of the regression coefficients: suppose for a particular individual the probability of capture is relatively high at each of the capture occasions and little changes over time. If for that particular individual the weight is also relatively high at each of the capture occasions, this indicates an individual heterogeneity relationship between probability of capture and the weight. Suppose that for another individual the probability of capture increases rapidly along the capturing periods. Suppose the similar pattern is also found for the weight in the same individual. This indicates a time relationship between the probability of capture and weight. These two relationships should be taken into account in the capture-recapture analysis as they are part of the capture-recapture relationship. Basically GEE or GLMM approach combines the two possible relationships into one estimating regression coefficient.

Second, in GEE and GLMM methods a correction is made for the dependency of the record of capture history within one individual, hence both techniques are suitable to analyse capture-recapture data. The question may arise: which one is the best method? Unfortunately, it is quite difficult to give a clear answer. Statistical results, and as a result of biological or management implications, sometimes may differ between GEE and GLMM. One possible reason is that likelihood methods are different for estimating parameters in GEE and GLMM. Selection of parameter estimates is of essential importance and should be supported for appropriate biological or management action.

Third, we could not compare relative fit of GEE versus GLM or GLMM directly mainly for two reasons.

(i) GEE method is based on the quasi-likelihood theory, while GLM and GLMM typically use a maximum-likelihood theory for model estimation. Comparative measures such as Akaike's Information Criterion (AIC) (Akaike, 1974) could be used for GLM or GLMM, whereas the quasi-likelihood under the independence model information criterion (QIC) (Pan, 2001) could be used for GEE for evaluating relative fit of models, but there is no unique model selection criterion that can be used for both models.

(ii) Estimated regression parameters and their significance may be different because we estimate regression parameters based on partial likelihood method in GLM or GLMM, while quasi-likelihood method is used to estimate regression parameters in GEE. Hence their comparison would not be appropriate.

However, to get accurate capture probabilities, one should fit statistical model considering data type and research objectives. We have applied a set of generalized models (i.e., GLM, GEE and GLMM) to estimate population parameters using capture-recapture data. We have also made comparison of the results of estimating population size and its standard error for the different model structures using covariates. The quasi-likelihood GEE approach seems to perform better than GLM and GLMM approaches because the standard errors of the estimated population size are consistently lower. The estimators perform well when average capture probabilities are high, but it is difficult to obtain reliable estimates of GLMM approach for low capture probabilities. For cases where only a small proportion of individuals are captured, the GEE approach provides better RMSE and is robust to violation of the assumption of independence among capture occasions. This approach also provides means of exploring factors thought to be responsible for differences in capture probability among individuals. Hence it is important to account for correlation structure among capture occasions when estimating animal population parameters in capture-recapture studies.

Chapter 6

General Conclusion and Future Direction

This dissertation proposes a generalized estimating equations (GEE) approach in capture-recapture closed population models that accounts for heterogeneity due to measurable individual characteristics and dependency among capture occasions. The proposed approach can also account for over-dispersion, modelling capture probabilities as a logit link function of continuous explanatory covariates, and/or considering Bernoulli response covariate through a generalized linear modelling framework. The generalized estimating equations versions of the class of models related to those of Otis et al. (1978), when appropriate, are considered and their corresponding estimating equations are also developed. We proposed to use 'quasi-likelihood information criterion (QIC)' to analyse correlated capture-recapture data and for model selection.

We have discussed several examples where the generalized estimating equations approach is used to estimate population parameters (such as, capture probabilities, and population size) from wildlife capture-recapture data. We also use a generalized linear mixed modelling (GLMM) approach to model heterogeneity of individual measurable and unmeasurable characteristics in capture-recapture experiments where heterogeneity due to unmeasurable individual characteristics is modelled as a random effect. Moreover, comparisons of population size estimates has been made using generalized estimating equations and mixed effects approaches. In particular, the proposed generalized estimating equations approach has been applied into closed population capture-recapture models taking into account dependency among capture occasions, these include: behavioral response and individual heterogeneity model ($\mathrm{M_{bh}}$) in Chapter 3; temporal variation and individual heterogeneity model ($\mathrm{M_{th}}$) in Chapter 4; and individual heterogeneity model ($\mathrm{M_h}$) and a comparison with GLMM approach in capture-recapture studies in Chapter 5.

The use of auxiliary information in the form of covariates was considered throughout the dissertation. If these covariates are biologically meaningful, then such information can greatly improve inference of population parameters. Our proposed approach performs well to estimate unknown population size in comparison with other currently used methodologies for data analysis. The estimation results of generalized estimating equations approach also ensure that time correlation play an important role when analyzing capture-recapture correlated data.

The simulation studies of the proposed generalized estimating equations approach for different closed population capture-recapture models show that estimator performance is good if a large proportion of individuals are captured. For a low capture proportion, the estimates seem to be unstable, but the generalized estimating equations approach outperforms the other approaches (i.e., GLM and GLMM). The simulation studies also indicate that the estimator performance depends on average capture probability, number of capture occasions and correlation among capture occasions.

Hence, our analyses underpin the importance of considering heterogeneity in capture probabilities and correlation among capture occasions when estimating population parameters in capture-recapture studies.

With the increasing development of computational power and efficient use of statistical tools, we see a fruitful future for the proposed generalized estimating equations approach in capture-recapture studies, and hope the presented technique will become standard practice to users. Even though our interests focused on ecological applications we also envisage the presented techniques to be used in related areas such as epidemiological applications. A list of future studies are suggested below:

(i) One potential future avenue for research is to develop a model that include heterogeneity due to unmeasured individual covariates into our proposed generalized estimating equations approach in capture-recapture closed population models considering that the generalized estimating equations approach is population average or marginal model. Further work needs to be done to increase the utility of the modelling approach when applied to often complex, different species that have been monitored as a capture-recapture study. This can help to better understand their biological characteristics and behavior.

(ii) Extend our proposed approach to open population capture-recapture models that will simultaneously take into account the heterogeneity due to measurable and unmeasurable individual characteristics of capture probabilities and dependency among capture occasions.

(iii) One may combine Bayesian destination with our proposed approach. In this case, researchers also need to modify the model selection criterion, QIC for selecting the best model.

(iv) A user-friendly program needs to be developed. It should be able to simulate data or to analyse real dataset as in CAPTURE or MARK programs. This can make the method of analysis more appealing and assessable to researchers to estimate the population parameters.

Appendix A

Proof and Derivation

Proof of Theorem 1:

Write $\alpha^*(\beta) = \hat{\alpha}\{\beta, \hat{\phi}(\beta)\}$ and under some regularity conditions $n^{\frac{1}{2}}(\hat{\beta}_{\text{GEE}} - \beta)$ can be approximated by

$$\left[\sum_{i=1}^{n} -\frac{\delta}{\delta\beta} U_i\{\beta, \alpha^*(\beta)\}/n\right]^{-1} \left[U_i\{\beta, \alpha^*(\beta)\}/n^{\frac{1}{2}}\right],$$

where

$$\begin{aligned}\delta U_i\{\beta, \alpha^*(\beta)\}/\delta\beta &= \partial U_i\{\beta, \alpha^*(\beta)\}/\partial\beta + [\partial U_i\{\beta, \alpha^*(\beta)\}/\partial\alpha^*]\{\partial\alpha^*(\beta)/\partial\beta\} \\ &= A_i + B_i C.\end{aligned}$$

Let β be fixed and Taylor expansion gives

$$\begin{aligned}\frac{\sum\limits_{i=1}^{n} U_i\{\beta, \alpha^*(\beta)\}}{n^{\frac{1}{2}}} &= \frac{\sum\limits_{i=1}^{n} U_i\{\beta, \alpha\}}{n^{\frac{1}{2}}} + \frac{\sum\limits_{i=1}^{n} \delta/\delta\alpha U_i\{\beta, \alpha\}}{n} n^{\frac{1}{2}}(\alpha^* - \alpha) + o_{prob}(1) \\ &= A^* + B^* C^* + o_{prob}(1),\end{aligned}$$

where the sums are over $i = 1, 2, ..., n$. Now, $B^* = o_{prob}(1)$, since $\partial U_i(\beta, \alpha)/\partial\alpha$ are linear functions of S_i's whose means are zero, and the properties of exponential family give

$$\begin{aligned}C^* &= n^{\frac{1}{2}}\left[\hat{\alpha}\{\beta, \hat{\phi}(\beta)\} - \hat{\alpha}(\beta, \phi) + \hat{\alpha}(\beta, \phi) - \alpha\right] \\ &= n^{\frac{1}{2}}\left\{\frac{\partial \hat{\alpha}}{\partial \phi}(\beta, \phi^*)(\hat{\phi} - \phi) + \hat{\alpha}(\beta, \phi) - \alpha\right\} = O_{prob}(1).\end{aligned}$$

Consequently, $\sum_{i=1}^{n} U_i\{\beta, \alpha^*(\beta)\}/n^{\frac{1}{2}}$ is asymptotically equivalent to A^* whose asymptotic distribution is multivariate Gaussian with zero mean and covariance matrix

$$\lim_{n\to\infty}\left\{\sum_{i=1}^{n} D_i' V_i^{-1} \mathrm{Cov}(Y_i) V_i^{-1} D_i/n\right\}.$$

Finally, it is easy to show that $\sum_{i=1}^{n} B_i = o_{prob}(n)$, $C = O_{prob}(1)$ and that $\sum_{i=1}^{n} A_i/n$ converges as $n \to \infty$ to $-\sum_{i=1}^{n} D_i' V_i^{-1} D_i/n$. This completes the proof.

Derivation of estimating equations (5.6):

We have,

$$U(\beta) = \frac{1}{n}\sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - \mu_i) = 0$$

$$\text{or, } \frac{1}{n}\sum_{i=1}^{n} (A_i X_i)' V_i^{-1}(Y_i - \mu_i) = 0 \text{ (since, } D_i = A_i X_i)$$

$$\text{or, } \frac{1}{n}\sum_{i=1}^{n} X_i' A_i A_i^{-1}\big(Y_i - \mu_i\big) = 0 \left(\text{since } \; V_i = A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}} = A_i^{\frac{1}{2}} I A_i^{\frac{1}{2}} = A_i\right)$$

$$\text{or, } \frac{1}{n}\sum_{i=1}^{n} X_i'\big(Y_i - \mu_i\big) = 0$$

$$\text{or, } \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \left(\begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{im} \end{pmatrix} - \begin{pmatrix} \mu_{i1} \\ \vdots \\ \mu_{im} \end{pmatrix} \right) = 0$$

$$\text{or, } \frac{1}{n}\sum_{i=1}^{n} \left(\sum_{j=1}^{m} Y_{ij} - \sum_{j=1}^{m} \mu_{ij} \right) = 0$$

$$\text{or, } \frac{1}{n}\sum_{i=1}^{n} \left(\sum_{j=1}^{m} Y_{ij} - \sum_{j=1}^{m} \frac{P_0}{1-(1-P_0)^m} \right) = 0$$

$$\left(\mu_{ij} = \frac{P_{ij}}{1 - \prod_{k=1}^{m}(1-P_{ik})} = \frac{P_0}{1 - \prod_{k=1}^{m}(1-P_0)} = \frac{P_0}{(1-P_0)^m}; \text{ since, } \; P_{ij} = h(\beta_0) = P_0 \right)$$

$$\text{Hence, } \frac{1}{n}\sum_{i=1}^{n} \left(\sum_{j=1}^{m} Y_{ij} - \frac{mP_0}{1-(1-P_0)^m} \right) = 0$$

Appendix B

Computational Code

Some base code implemented in R Project program, release 3.0.2 (http://www.r-project.org/).

- **Estimation of model parameters applying Generalized Estimating Equations (GEE) approach to capture-recapture closed population models considering independence working correlation structure in GEE:**

```
estim<-function(data,initial){
flno<-data[,2]
cap<-y<-data[,3]
sex<-data[,4]
age<-data[,5]
weight<-data[,6]
behav<-data[,7]
count<-0
repeat{
        b0<-initial[1]
        b1<-initial[2]
        b2<-initial[3]
        b3<-initial[4]
        b4<-initial[5]
        m<-6; k<-0
        rij<-0
```

```
score<-c(0,0,0,0,0)
inf<-matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
ncol=5,byrow=T)
for(i in 1:length(flno)){
if(flno[i]==1&&flno[i+1]==2&&flno[i+2]==3&&flno[i+3]==4&&flno[i+4]==5
&&flno[i+5]==6){
        k<-k+1
        m1<-exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i])/
        (1+exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i]))
        m2<-exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1])/
        (1+exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1]))
        m3<-exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2])/
        (1+exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2]))
        m4<-exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3])/
        (1+exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3]))
        m5<-exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4])/
        (1+exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4]))
        m6<-exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5])/
        (1+exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5]))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
ai<-matrix(c(v[1],0,0,0,0,0,0,v[2],0,0,0,0,0,0,v[3],0,0,0,0,0,0,
v[4],0,0,0,0,0,0,v[5],0,0,0,0,0,0,v[6]),ncol=m,byrow=T)
ymi<-c(y[i]-m1,y[i+1]-m2,y[i+2]-m3,y[i+3]-m4,y[i+4]-m5,y[i+5]-m6)
xit<-matrix(c(1,1,1,1,1,1,age[i],age[i+1],age[i+2],age[i+3],age[i+4],
age[i+5],sex[i],sex[i+1],sex[i+2],sex[i+3],sex[i+4],sex[i+5],weight[i],
weight[i+1],weight[i+2],weight[i+3],weight[i+4],weight[i+5],behav[i],
behav[i+1],behav[i+2],behav[i+3],behav[i+4],behav[i+5]),ncol=m,byrow=T)
xi<-t(xit)
score<-score+xit%*%ymi
inf<-inf+xit%*%ai%*%xi
        }       }
cat(``The score vector is\n")
print(score)
infinv<-solve(inf)
cat(``The inverse information matrix is\n")
print(infinv)
est<-initial+infinv%*%score
cat(``The initial estimate is\n")
print(est)
conv<-abs(score)
if(conv[1]<=0.0001&&conv[2]<=0.0001&&conv[3]<=0.0001&&conv[4]<=0.0001
&&conv[5]<=0.0001)
break
initial<-est
```

```
pij<-exp(est[1]+age*est[2]+sex*est[3]+weight*est[4]+behav*est[5])/
(1+exp(est[1]+age*est[2]+sex*est[3]+weight*est[4]+behav*est[5]))
          }
cat(‘‘The required number of iteration is\n”)
print(count)
cat(‘‘The estimates of regression coefficients are\n”)
print(est)
p<-matrix(c(pij), ncol=m,byrow=T)
prod<-1
for(k in 1:m){
prod<-prod*(1-p[,k])}
pi<-1-prod
Nhat<-sum(1/pi)
varA<-sum((1-pi)/(pi)^2)
delta.beta<-xit%*%(pi^(-2)*(1-p)^m * m * p)
varB<-(t(delta.beta)%*%infinv)%*%delta.beta
varN<-varA+varB
se.Nhat<-sqrt(varN)
cat(‘‘The estimated population size is\n”)
print(Nhat)
cat(‘‘The standard error of the estimated population size is\n”)
print(se.Nhat)
wald0<-est[1]/sqrt(infinv[1,1])
cat(‘‘The Wald statistic for b0 is\n”)
print(wald0)
wald1<-est[2]/sqrt(infinv[2,2])
cat(‘‘The Wald statistic for b1 is\n”)
print(wald1)
wald2<-est[3]/sqrt(infinv[3,3])
cat(‘‘The Wald statistic for b2 is\n”)
print(wald2)
wald3<-est[4]/sqrt(infinv[4,4])
cat(‘‘The Wald statistic for b3 is\n”)
print(wald3)
wald4<-est[5]/sqrt(infinv[5,5])
cat(‘‘The Wald statistic for b4 is\n”)
print(wald4)
se_c<-c(sqrt(infinv[1,1]),sqrt(infinv[2,2]),sqrt(infinv[3,3]),
sqrt(infinv[4,4]), sqrt(infinv[5,5]))
cat(‘‘The standard errors of regression coefficients are\n”)
print(se_c)
or<-c(exp(est[1]),exp(est[2]),exp(est[3]),exp(est[4]),exp(est[5]))
cat(‘‘The odd ratios of regression coefficients are\n”)
print(or)}
```

• Estimation of model parameters applying GEE approach to capture-recapture closed population models considering exchangeable working correlation structure in GEE:

```
estim<-function(data,initial){
flno<-data[,2]
cap<-y<-data[,3]
sex<-data[,4]
age<-data[,5]
weight<-data[,6]
behav<-data[,7]
count<-0
repeat{
        b0<-initial[1]
        b1<-initial[2]
        b2<-initial[3]
        b3<-initial[4]
        b4<-initial[5]
        k<-0; m<-6
        rij<-0
score<-c(0,0,0,0,0)
inf<-matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
ncol=5,byrow=T)
for(i in 1: length(flno)){
if(flno[i]==1&&flno[i+1]==2&&flno[i+2]==3&&flno[i+3]==4&&flno[i+4]==5
&&flno[i+5]==6){
        k<-k+1
        m1<-exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i])/
        (1+exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i]))
        m2<-exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1])/
        (1+exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1]))
        m3<-exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2])/
        (1+exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2]))
        m4<-exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3])/
        (1+exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3]))
        m5<-exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4])/
        (1+exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4]))
        m6<-exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5])/
        (1+exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5]))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
ai<-matrix(c(v[1],0,0,0,0,0,0,v[2],0,0,0,0,0,0,v[3],0,0,0,0,0,0,
```

```
v[4],0,0,0,0,0,0,v[5],0,0,0,0,0,0,v[6]),ncol=m,byrow=T)
r<-c((y[i]-m1)/sqrt(v[1]),(y[i+1]-m2)/sqrt(v[2]),
(y[i+2]-m3)/sqrt(v[3]),(y[i+3]-m4)/sqrt(v[4]),
(y[i+4]-m5)/sqrt(v[5]),(y[i+5]-m6)/sqrt(v[6]))
rij<-rij+r[1]*r[2]+r[1]*r[3]+r[1]*r[4]+r[1]*r[5]+r[1]*r[6]+
r[2]*r[3]+r[2]*r[4]+r[2]*r[5]+r[2]*r[6]+r[3]*r[4]+r[3]*r[5]+
r[3]*r[6]+r[4]*r[5]+r[4]*r[6]+r[5]*r[6]
                  }
         }
count<-count+1
cat(‘‘The required number of iteration is\n”)
print(count)
alpha<-rij/(m*k-1)
cat(‘‘The value of alpha is\n”)
print(alpha)
ria<-matrix(c(1,alpha,alpha,alpha,alpha,alpha,alpha,1,alpha,
alpha,alpha,alpha,alpha,alpha,1,alpha,alpha,alpha,alpha,alpha,
alpha,1,alpha,alpha,alpha,alpha,alpha,alpha,1,alpha,alpha,alpha,
alpha,alpha,alpha,1),ncol=m,byrow=T)
cat(‘‘The correlation matrix is\n”)
print(ria)
for(i in 1:length(flno)){
if(flno[i]==1&&flno[i+1]==2&&flno[i+2]==3&&flno[i+3]==4&&
flno[i+4]==5&&flno[i+5]==6){
         k<-k+1
         m1<-exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i])/
         (1+exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i]))
         m2<-exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1])/
         (1+exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1]))
         m3<-exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2])/
         (1+exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2]))
         m4<-exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3])/
         (1+exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3]))
         m5<-exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4])/
         (1+exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4]))
         m6<-exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5])/
         (1+exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5]))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
ai<-matrix(c(v[1],0,0,0,0,0,0,v[2],0,0,0,0,0,0,v[3],0,0,0,0,0,0,
v[4],0,0,0,0,0,0,v[5],0,0,0,0,0,0,v[6]),ncol=m,byrow=T)
vij<-matrix(c(sqrt(v[1]),0,0,0,0,0,0,sqrt(v[2]),0,0,0,0,0,0,
sqrt(v[3]),0,0,0,0,0,0,sqrt(v[4]),0,0,0,0,0,0,sqrt(v[5]),0,0,
0,0,0,0,sqrt(v[6])),ncol=m,byrow=T)
```

```
var<-vij%*%ria%*%vij
inv<-solve(var)
ymi<-c(y[i]-m1,y[i+1]-m2,y[i+2]-m3,y[i+3]-m4,y[i+4]-m5,y[i+5]-m6)
xit<-matrix(c(1,1,1,1,1,1,age[i],age[i+1],age[i+2],age[i+3],
age[i+4],age[i+5],sex[i],sex[i+1],sex[i+2],sex[i+3],sex[i+4],
sex[i+5],weight[i],weight[i+1],weight[i+2],weight[i+3],weight[i+4],
weight[i+5]),behav[i], behav[i+1],behav[i+2],behav[i+3],behav[i+4],
behav[i+5],ncol=m,byrow=T)
dit<-xit%*%ai
di<-t(dit)
score<-score+dit%*%inv%*%ymi
inf<-inf+dit%*%inv%*%di
                        }
            }
cat(``The score vector is\n")
print(score)
infinv<-solve(inf)
cat(``The inverse information matrix is\n")
print(infinv)
est<-initial+infinv%*%score
cat(``The initial estimate is\n")
print(est)
conv<-abs(score)
if(conv[1]<=0.0001&&conv[2]<=0.0001&&conv[3]<=0.0001&&
conv[4]<=0.0001&&conv[5]<=0.0001)
break
initial<-est
pij<-exp(est[1]+age*est[2]+sex*est[3]+weight*est[4]+behav*est[5])/
(1+exp(est[1]+age*est[2]+sex*est[3]+weight*est[4]+behav*est[5]))
            }
cat(``The required number of iteration is\n")
print(count)
cat(``The estimates of regression coefficients are\n")
print(est)
p<-matrix(c(pij), ncol=m,byrow=T)
prod<-1
for(k in 1:m){
prod<-prod*(1-p[,k])}
pi<-1-prod
Nhat<-sum(1/pi)
varA<-sum((1-pi)/(pi)^2)
delta.beta<-xit%*%(pi^(-2)*(1-p)^m * m * p)
varB<-(t(delta.beta)%*%infinv)%*%delta.beta
varN<-varA+varB
```

```
se.Nhat<−sqrt(varN)
cat(‘‘The estimated population size is\n”)
print(Nhat)
cat(‘‘The standard error of the estimated population size is\n”)
print(se.Nhat)
wald0<−est[1]/sqrt(infinv[1,1])
cat(‘‘The Wald statistic for b0 is\n”)
print(wald0)
wald1<−est[2]/sqrt(infinv[2,2])
cat(‘‘The Wald statistic for b1 is\n”)
print(wald1)
wald2<−est[3]/sqrt(infinv[3,3])
cat(‘‘The Wald statistic for b2 is\n”)
print(wald2)
wald3<−est[4]/sqrt(infinv[4,4])
cat(‘‘The Wald statistic for b3 is\n”)
print(wald3)
wald4<−est[5]/sqrt(infinv[5,5])
cat(‘‘The Wald statistic for b4 is\n”)
print(wald4)
se_c<−c(sqrt(infinv[1,1]),sqrt(infinv[2,2]),sqrt(infinv[3,3]),
sqrt(infinv[4,4]), sqrt(infinv[5,5]))
cat(‘‘The standard errors of regression coefficients are\n”)
print(se_c)
or<−c(exp(est[1]),exp(est[2]),exp(est[3]),exp(est[4]),exp(est[5]))
cat(‘‘The odd ratios of regression coefficients are\n”)
print(or)
         }
```

• Estimation of model parameters applying GEE approach to capture-recapture closed population models considering autoregressive working correlation structure in GEE:

```
estim<-function(data,initial){
flno<-data[,2]
cap<-y<-data[,3]
sex<-data[,4]
age<-data[,5]
weight<-data[,6]
behav<-data[,7]
count<-0
repeat{
          b0<-initial[1]
          b1<-initial[2]
          b2<-initial[3]
          b3<-initial[4]
          b4<-initial[5]
          k<-0; m<-6
          rij<-0
score<-c(0,0,0,0,0)
inf<-matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
ncol=5,byrow=T)
for(i in 1: length(flno)){
if(flno[i]==1&&flno[i+1]==2&&flno[i+2]==3&&flno[i+3]==4&&flno[i+4]==5
&&flno[i+5]==6){
          k<-k+1
          m1<-exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i])/
          (1+exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i]))
          m2<-exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1])/
          (1+exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1]))
          m3<-exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2])/
          (1+exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2]))
          m4<-exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3])/
          (1+exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3]))
          m5<-exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4])/
          (1+exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4]))
          m6<-exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5])/
          (1+exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5]))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
ai<-matrix(c(v[1],0,0,0,0,0,0,v[2],0,0,0,0,0,0,v[3],0,0,0,0,0,0,
```

```
v[4],0,0,0,0,0,0,v[5],0,0,0,0,0,0,v[6]),ncol=m,byrow=T)
r<-c((y[i]-m1)/sqrt(v[1]),(y[i+1]-m2)/sqrt(v[2]),
(y[i+2]-m3)/sqrt(v[3]),(y[i+3]-m4)/sqrt(v[4]),
(y[i+4]-m5)/sqrt(v[5]),(y[i+5]-m6)/sqrt(v[6]))
rij<-rij+r[1]*r[2]+r[1]*r[3]+r[1]*r[4]+r[1]*r[5]+r[1]*r[6]+
r[2]*r[3]+r[2]*r[4]+r[2]*r[5]+r[2]*r[6]+r[3]*r[4]+r[3]*r[5]+
r[3]*r[6]+r[4]*r[5]+r[4]*r[6]+r[5]*r[6]
                    }
          }
count<-count+1
cat(‘‘The required number of iteration is\n”)
print(count)
alpha<-rij/(m*k-1)
cat(‘‘The value of alpha is\n”)
print(alpha)
ria<-matrix(c(1,alpha,alpha^2,alpha^3,alpha^4,alpha^5,alpha,1,
alpha,alpha^2,alpha^3,alpha^4,alpha^2,alpha,1,alpha,alpha^2,
alpha^3,alpha^3,alpha^2,alpha,1,alpha,alpha^2,alpha^4,alpha^3,
alpha^2,alpha,1,alpha,alpha^5,alpha^4,alpha^3,alpha^2,alpha,1),
ncol=m,byrow=T)
cat(‘‘The correlation matrix is\n”)
print(ria)
for(i in 1: length(flno)){
if(flno[i]==1&&flno[i+1]==2&&flno[i+2]==3&&flno[i+3]==4&&
flno[i+4]==5&&flno[i+5]==6){
          k<-k+1
          m1<-exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i])/
          (1+exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i]))
          m2<-exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1])/
          (1+exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1]))
          m3<-exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2])/
          (1+exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2]))
          m4<-exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3])/
          (1+exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3]))
          m5<-exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4])/
          (1+exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4]))
          m6<-exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5])/
          (1+exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5]))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
ai<-matrix(c(v[1],0,0,0,0,0,0,v[2],0,0,0,0,0,0,v[3],0,0,0,0,0,0,
v[4],0,0,0,0,0,0,v[5],0,0,0,0,0,0,v[6]),ncol=m,byrow=T)
vij<-matrix(c(sqrt(v[1]),0,0,0,0,0,0,sqrt(v[2]),0,0,0,0,0,0,
sqrt(v[3]),0,0,0,0,0,0,sqrt(v[4]),0,0,0,0,0,0,sqrt(v[5]),0,0,
```

```
0,0,0,0,sqrt(v[6])),ncol=m,byrow=T)
var<-vij%*%ria%*%vij
inv<-solve(var)
ymi<-c(y[i]-m1,y[i+1]-m2,y[i+2]-m3,y[i+3]-m4,y[i+4]-m5,y[i+5]-m6)
xit<-matrix(c(1,1,1,1,1,1,age[i],age[i+1],age[i+2],age[i+3],
age[i+4],age[i+5],sex[i],sex[i+1],sex[i+2],sex[i+3],sex[i+4],
sex[i+5],weight[i],weight[i+1],weight[i+2],weight[i+3],weight[i+4],
weight[i+5]),behav[i], behav[i+1],behav[i+2],behav[i+3],behav[i+4],
behav[i+5],ncol=m,byrow=T)
dit<-xit%*%ai
di<-t(dit)
score<-score+dit%*%inv%*%ymi
inf<-inf+dit%*%inv%*%di
                    }
          }
cat(‘‘The score vector is\n”)
print(score)
infinv<-solve(inf)
cat(‘‘The inverse information matrix is\n”)
print(infinv)
est<-initial+infinv%*%score
cat(‘‘The initial estimate is\n”)
print(est)
conv<-abs(score)
if(conv[1]<=0.0001&&conv[2]<=0.0001&&conv[3]<=0.0001&&
conv[4]<=0.0001&&conv[5]<=0.0001)
break
initial<-est
pij<-exp(est[1]+age*est[2]+sex*est[3]+weight*est[4]+behav*est[5])/
(1+exp(est[1]+age*est[2]+sex*est[3]+weight*est[4]+behav*est[5]))
          }
cat(‘‘The required number of iteration is\n”)
print(count)
cat(‘‘The estimates of regression coefficients are\n”)
print(est)
p<-matrix(c(pij), ncol=m,byrow=T)
prod<-1
for(k in 1:m){
prod<-prod*(1-p[,k])}
pi<-1-prod
Nhat<-sum(1/pi)
varA<-sum((1-pi)/(pi)^2)
delta.beta<-xit%*%(pi^(-2)*(1-p)^m * m * p)
varB<-(t(delta.beta)%*%infinv)%*%delta.beta
```

```
varN<-varA+varB
se.Nhat<-sqrt(varN)
cat(``The estimated population size is\n")
print(Nhat)
cat(``The standard error of the estimated population size is\n")
print(se.Nhat)
wald0<-est[1]/sqrt(infinv[1,1])
cat(``The Wald statistic for b0 is\n")
print(wald0)
wald1<-est[2]/sqrt(infinv[2,2])
cat(``The Wald statistic for b1 is\n")
print(wald1)
wald2<-est[3]/sqrt(infinv[3,3])
cat(``The Wald statistic for b2 is\n")
print(wald2)
wald3<-est[4]/sqrt(infinv[4,4])
cat(``The Wald statistic for b3 is\n")
print(wald3)
wald4<-est[5]/sqrt(infinv[5,5])
cat(``The Wald statistic for b4 is\n")
print(wald4)
se_c<-c(sqrt(infinv[1,1]),sqrt(infinv[2,2]),sqrt(infinv[3,3]),
sqrt(infinv[4,4]), sqrt(infinv[5,5]))
cat(``The standard errors of regression coefficients are\n")
print(se_c)
or<-c(exp(est[1]),exp(est[2]),exp(est[3]),exp(est[4]),exp(est[5]))
cat(``The odd ratios of regression coefficients are\n")
print(or)
         }
```

• Estimation of model parameters applying GEE approach to capture-recapture closed population models considering pairwise working correlation structure in GEE:

```
estim<-function(data, initial){
flno<-data[,2]
cap<-y<-data[,3]
sex<-data[,4]
age<-data[,5]
weight<-data[,6]
behav<-data[,7]
count<-0
repeat{
        b0<-initial[1]
        b1<-initial[2]
        b2<-initial[3]
        b3<-initial[4]
        b4<-initial[5]
        k<-0; m<-6
score<-c(0,0,0,0,0)
inf<-matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0),
ncol=5,byrow=T)
rij<-matrix(c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0),ncol=m, byrow=T)
for(i in 1: length(flno)){
if(flno[i]==1&&flno[i+1]==2&&flno[i+2]==3&&flno[i+3]==4&&
flno[i+4]==5&&flno[i+5]==6){
        k<-k+1
        m1<-exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i])/
        (1+exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i]))
        m2<-exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1])/
        (1+exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1]))
        m3<-exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2])/
        (1+exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2]))
        m4<-exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3])/
        (1+exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3]))
        m5<-exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4])/
        (1+exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4]))
        m6<-exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5])/
        (1+exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5]))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
```

```
ai<-matrix(c(v[1],0,0,0,0,0,0,v[2],0,0,0,0,0,0,v[3],0,0,0,0,
0,0,v[4],0,0,0,0,0,0,v[5],0,0,0,0,0,0,v[6]),ncol=m,byrow=T)
vij<-matrix(c(sqrt(v[1]),0,0,0,0,0,0,sqrt(v[2]),0,0,0,0,
0,0,sqrt(v[3]),0,0,0,0,0,0,sqrt(v[4]),0,0,0,0,0,0,
sqrt(v[5]),0,0,0,0,0,0,sqrt(v[6])),ncol=m,byrow=T)
ymi<-c(y[i]-m1,y[i+1]-m2,y[i+2]-m3,y[i+3]-m4,y[i+4]-m5,y[i+5]-m6)
ymit<-t(ymi)
vijinv<-solve(vij)
rij<-rij+vijinv%*%ymi%*%ymit%*%vijinv
          }
                   }
count<-count+1
cat(‘‘The number of iteration is\n”)
print(count)
ria<-rij/k
cat(‘‘The correlation matrix is\n”)
print(ria)
for(i in 1: length(flno)){
if(flno[i]==1&&flno[i+1]==2&&flno[i+2]==3&&flno[i+3]==4&&
flno[i+4]==5&&flno[i+5]==6){
          k<-k+1
          m1<-exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i])/
          (1+exp(b0+b1*age[i]+b2*sex[i]+b3*weight[i]+b4*behav[i]))
          m2<-exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1])/
          (1+exp(b0+b1*age[i+1]+b2*sex[i+1]+b3*weight[i+1]+b4*behav[i+1]))
          m3<-exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2])/
          (1+exp(b0+b1*age[i+2]+b2*sex[i+2]+b3*weight[i+2]+b4*behav[i+2]))
          m4<-exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3])/
          (1+exp(b0+b1*age[i+3]+b2*sex[i+3]+b3*weight[i+3]+b4*behav[i+3]))
          m5<-exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4])/
          (1+exp(b0+b1*age[i+4]+b2*sex[i+4]+b3*weight[i+4]+b4*behav[i+4]))
          m6<-exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5])/
          (1+exp(b0+b1*age[i+5]+b2*sex[i+5]+b3*weight[i+5]+b4*behav[i+5]))
v<-c(m1*(1-m1),m2*(1-m2),m3*(1-m3),m4*(1-m4),m5*(1-m5),m6*(1-m6))
ai<-matrix(c(v[1],0,0,0,0,0,0,v[2],0,0,0,0,0,0,v[3],0,0,0
,0,0,0,v[4],0,0,0,0,0,0,v[5],0,0,0,0,0,0,v[6]),ncol=m,byrow=T)
vij<-matrix(c(sqrt(v[1]),0,0,0,0,0,0,sqrt(v[2]),0,0,0,0,
0,0,sqrt(v[3]),0,0,0,0,0,0,sqrt(v[4]),0,0,0,0,0,0,
sqrt(v[5]),0,0,0,0,0,0,sqrt(v[6])),ncol=m,byrow=T)
var<-vij%*%ria%*%vij
inv<-solve(var)
ymi<-c(y[i]-m1,y[i+1]-m2,y[i+2]-m3,y[i+3]-m4,y[i+4]-m5,y[i+5]-m6)
xit<-matrix(c(1,1,1,1,1,1,age[i],age[i+1],age[i+2],age[i+3],
age[i+4],age[i+5],sex[i],sex[i+1],sex[i+2],sex[i+3],sex[i+4],
```

```
sex[i+5],weight[i],weight[i+1],weight[i+2],weight[i+3],
weight[i+4],weight[i+5],behav[i], behav[i+1],behav[i+2],
behav[i+3],behav[i+4],behav[i+5]),ncol=m,byrow=T)
dit<-xit%*%ai
di<-t(dit)
score<-score+dit%*%inv%*%ymi
inf<-inf+dit%*%inv%*%di
                    }
          }
cat(‘‘The score vector is\n”)
print(score)
infinv<-solve(inf)
cat(‘‘The inverse information matrix is\n”)
print(infinv)
est<-initial+infinv%*%score
cat(‘‘The initial estimate is\n”)
print(est)
conv<-abs(score)
if(conv[1]<=0.0001&&conv[2]<=0.0001&&conv[3]<=0.0001&&
conv[4]<=0.0001&&conv[5]<=0.0001)
break
initial<-est
pij<-exp(est[1]+age*est[2]+sex*est[3]+weight*est[4]+behav*est[5])/
(1+exp(est[1]+age*est[2]+sex*est[3]+weight*est[4]+behav*est[5]))
          }
cat(‘‘The required number of iteration is\n”)
print(count)
cat(‘‘The estimates of regression coefficients are\n”)
print(est)
p<-matrix(c(pij), ncol=m,byrow=T)
prod<-1
for(k in 1:m){
prod<-prod*(1-p[,k])}
pi<-1-prod
Nhat<-sum(1/pi)
varA<-sum((1-pi)/(pi)^2)
delta.beta<-xit%*%(pi^(-2)*(1-p)^m * m * p)
varB<-(t(delta.beta)%*%infinv)%*%delta.beta
varN<-varA+varB
se.Nhat<-sqrt(varN)
cat(‘‘The estimated population size is\n”)
print(Nhat)
cat(‘‘The standard error of the estimated population size is\n”)
print(se.Nhat)
```

```
wald0<-est[1]/sqrt(infinv[1,1])
cat(‘‘The Wald statistic for b0 is\n”)
print(wald0)
wald1<-est[2]/sqrt(infinv[2,2])
cat(‘‘The Wald statistic for b1 is\n”)
print(wald1)
wald2<-est[3]/sqrt(infinv[3,3])
cat(‘‘The Wald statistic for b2 is\n”)
print(wald2)
wald3<-est[4]/sqrt(infinv[4,4])
cat(‘‘The Wald statistic for b3 is\n”)
print(wald3)
wald4<-est[5]/sqrt(infinv[5,5])
cat(‘‘The Wald statistic for b4 is\n”)
print(wald4)
se_c<-c(sqrt(infinv[1,1]),sqrt(infinv[2,2]),sqrt(infinv[3,3]),
sqrt(infinv[4,4]), sqrt(infinv[5,5]))
cat(‘‘The standard errors of regression coefficients are\n”)
print(se_c)
or<-c(exp(est[1]),exp(est[2]),exp(est[3]),exp(est[4]),exp(est[5]))
cat(‘‘The odd ratios of regression coefficients are\n”)
print(or)
          }
```

• Estimation of Quasi-likelihood Information Criterion (QIC) for GEE approach in capture-recapture studies:

```
QIC = function(model.R){
      require(MASS)
      model.indep = update(model.R, corstr = ‘‘independence’’)
      mu.R = model.R$fitted.values # Quasi−likelihood
      y = model.R$y
      type = family(model.R)$family
      quasi.R = switch(type,
                  poisson = sum((y∗log(mu.R)) − mu.R),
                  gaussian = sum(((y − mu.R)ˆ2)/−2),
                  binomial = sum(y∗log(mu.R/(1 − mu.R)) + log(1 − mu.R)),
                  Gamma = sum(−y/(mu.R − log(mu.R))),
                  stop(‘‘Error: distribution not recognized”))
# Trace Term (penalty for model complexity)
omegaI = ginv(model.indep$geese$vbeta.naiv)
# Omega−hat(I) via Moore−Penrose generalized
inverse of a matrix in MASS package
AIinverse = solve(model.indep$geese$vbeta.naiv) #solve via indenity
Vr = model.R$geese$vbeta
trace.R = sum(diag(omegaI %∗% Vr))
p_x = length(mu.R) #number non−redunant columns in design matrix
QIC = 2∗(trace.R − quasi.R) #QIC
QICu = (−2)∗quasi.R + 2∗p_x
#Approximation assuming model structured correctly
output = c(QIC,QICu, quasi.R, trace.R, p_x)
names(output) = c(‘QIC’, ‘QICu’, ‘Quasi Lik’, ‘Trace’, ‘p_x’)
return(output) }
```

• STATA code for the estimation of QIC considering various correlation structures:

```
qic cap sex age weight behav, i(id) t(flno) family(binomial)
link(logit) corr(independent)
qic cap sex age weight behav, i(id) t(flno) family(binomial)
link(logit) corr(exchangeable)
qic cap sex age weight behav, i(id) t(flno) family(binomial)
link(logit) corr(AR 1)
qic cap sex age weight behav, i(id) t(flno) family(binomial)
link(logit) corr(unstructured)
```

• Simulation code for GEE approach in capture-recapture M_{bh} model:

```
# nsim: number of Monte Carlo replicates; N: population size;
m: capture occasions; r: correlation parameter; mu: mean
vector of capture probability
set.seed(999)
require(binarySimCLF)
sim<−function(nsim,N,m,r,mu){
         npar<−2
         para<−matrix(0,nsim,npar)
for (s in 1:nsim){
         R<−ar1(m,r) #correlation matrix
         natural.compat<−chkBinC(R,mu)
         V<−cor2var(R,mu) #covariance matrix from R and mu
         B<−allReg(V)
         clf.compat<−blrchk1(mu,B); #checks for CLF compatibility
         if (clf.compat){
         y<−mbsclf(N,mu,B);
                  }
         else {print(‘‘Not CLF compatible”)
                            }
c.hist<−y$y # Correlated capture histories for specified probability
mean vector and correlation matrix
c.mat<−c.hist[apply(c.hist,1,sum)>0,] #capture history for observed
individual's only
z1<−z2<−z3<−z4<−z5<−z6<−z7<−z8<−z9<−z10<−numeric(length(c.mat[,1]))
for (i in 1:length(c.mat[,1])){
if (c.mat[i,1]==1) z2[i] <− 1 else z2[i] <− 0
if (c.mat[i,2]==1||c.mat[i,1]==1) z3[i] <− 1 else z3[i] <− 0
if (c.mat[i,3]==1||c.mat[i,2]==1||c.mat[i,1]==1)z4[i]<−1 else z4[i]<−0
if (c.mat[i,4]==1||c.mat[i,3]==1||c.mat[i,2]==1||c.mat[i,1]==1)
z5[i] <− 1 else z5[i] <− 0
if (c.mat[i,5]==1||c.mat[i,4]==1||c.mat[i,3]==1||c.mat[i,2]==1||
c.mat[i,1]==1) z6[i] <− 1 else z6[i] <− 0
if (c.mat[i,6]==1||c.mat[i,5]==1||c.mat[i,4]==1||c.mat[i,3]==1||
c.mat[i,2]==1||c.mat[i,1]==1) z7[i] <− 1 else z7[i] <− 0
if (c.mat[i,7]==1||c.mat[i,6]==1||c.mat[i,5]==1||c.mat[i,4]==1||
c.mat[i,3]==1||c.mat[i,2]==1||c.mat[i,1]==1)z8[i]<−1 else z8[i]<−0
if (c.mat[i,8]==1||c.mat[i,7]==1||c.mat[i,6]==1||c.mat[i,5]==1||
c.mat[i,4]==1||c.mat[i,3]==1||c.mat[i,2]==1||
c.mat[i,1]==1) z9[i] <− 1 else z9[i] <− 0
if (c.mat[i,9]==1||c.mat[i,8]==1||c.mat[i,7]==1||c.mat[i,6]==1||
c.mat[i,5]==1||c.mat[i,4]==1||c.mat[i,3]==1||c.mat[i,2]==1||
```

```
c.mat[i,1]==1) z10[i]<-1 else z10[i]<-0
        }
zj<-cbind(z1,z2,z3,z4,z5,z6,z7,z8,z9,z10)
# behavioral response i.e., captured at the previous stage or not
zij<-c(t(zj))
nobs<-nrow(c.mat)      # number of individuals captured at least once
cap<-c(t(c.mat))
id <- gl(nobs, m, nobs*m)
flno<- rep(1:m, nobs)
sex<-rep(rbinom(nobs,1,0.5),each=m)    #individual discrete covariate
weight<-rep(rnorm(nobs,mean=15,sd=5),each=m)#individual continuous covariate
data <- data.frame(id,flno,cap,sex,weight,zij)
require(gee)
gee<-gee(cap~sex+weight+zij, id=id, data=data,family=binomial,
scale.fix =TRUE, scale.value = 1, corstr=‘‘AR-M”, Mv=1)
pij<-fitted(gee)
pi<-matrix(c(pij), ncol=m,byrow=T)
prod<-1
for(k in 1:m){
prod<-prod*(1-pi[,k])
        }
phi<-1-prod
nhat<-sum(1/phi)
para[s,1]<-nhat
para[s,2]<-nobs
        }
para
cind<-mean(para[,2])
cat(‘‘The average capture individuals is\n”)
print(cind)
cat(‘‘The average population size estimate is\n”)
print(mean(para[,1]))
se<-sqrt(var(para[,1]))
cat(‘‘The standard error of the population estimates is\n”)
print(se)
lcov<-mean(para[,1])-1.96*sqrt(var(para[,1]))
ucov<-mean(para[,1])+1.96*sqrt(var(para[,1]))
cat(‘‘The 95% confidence interval (95 %CI) is\n”)
print(c(lcov,ucov))
prb<-(mean(para[,1])-N)/N*100
cat(‘‘The percentage relative bias(PRB) of the estimated population is\n”)
print(prb)
cv<-sqrt(var(para[,1]))/mean(para[,1])*100
cat(‘‘The percentage cv of the estimated population is\n”)
```

```
print(cv)
rmse<-sqrt(var(para[,1])+(mean(para[,1])-N)^2)
cat(‘‘The root mean square error (RMSE) of the population estimate is\n”)
print(rmse)
        }
```

• Simulation code for GEE approach in capture-recapture $\mathrm{M_{th}}$ model:

```
# nsim: number of Monte Carlo replicates; N: population size;
m: capture occasions; r: correlation parameter; mu: mean
vector of capture probability
set.seed(999)
sim<-function(nsim,N,m,r,mu){
npar<-2
para<-matrix(0,nsim,npar)
for (s in 1:nsim){
require(binarySimCLF)
R<-ar1(m,r)     #correlation matrix
natural.compat<-chkBinC(R,mu)
V<-cor2var(R,mu)     #covariance matrix from R and mu
B<-allReg(V)
clf.compat<-blrchk1(mu,B); #checks for CLF compatibility
if (clf.compat){
y<-mbsclf(N,mu,B);
} else {print(‘‘Not CLF compatible”)}
c.hist<-y$y  #correlated capture histories for specified probability
mean vector and correlation matrix.
c.mat<-c.hist[apply(c.hist,1,sum)>0,]
#capture history for observed individual's only.
n<-nrow(c.mat)  # number of individuals captured at least once
cap<-c(t(c.mat))
id <- gl(n, m, n*m)
flno<- rep(1:m, n)
sex<-rep(rbinom(n,1,0.5),each=m)           # individual discrete covariate
weight<-rep(rnorm(n,mean=15,sd=2),each= m)# individual continuous covariate
ENV<-rep(rnorm(m,mean=2,sd=1), n)          # time varying covariate
data <- data.frame(id,flno,cap,sex,weight,ENV)
require(gee)
gee<-gee(cap~ sex+weight+ENV, id=id, data=data,family=binomial,
scale.fix =TRUE, scale.value = 1, corstr=‘‘AR-M”, Mv=1)
pij<-fitted(gee)
pi<-matrix(c(pij), ncol=m,byrow=T)
```

```
prod<-1
for(k in 1:m){
prod<-prod*(1-pi[,k])
        }
phi<-1-prod
nhat<-sum(1/phi)
para[s,1]<-nhat
para[s,2]<-n
        }
para
acind<-mean(para[,2])
cat(''The average capture individuals is\n")
print(acind)
mu<-mean(para[,1])
cat(''The average estimated population size is\n")
print(mu)
se<-sqrt(var(para[,1]))
cat(''The standard error of the estimated population is\n")
print(se)
lcov<-mu-1.96*se
ucov<-mu+1.96*se
cat(''The 95% confidence interval (95 %CI) is\n")
print(c(lcov,ucov))
prb<-(mu-N)/N*100
cat(''The percentage relative bias(PRB) of the estimated population is\n")
print(prb)
cv<-sqrt(var(para[,1]))/mean(para[,1])*100
cat(''The percentage cv of the estimated population is\n")
print(cv)
rmse<-sqrt(var(para[,1])+(mean(para[,1])-N)^2)
cat(''The root mean square error (RMSE) of the estimated population is\n")
print(rmse)
        }
```

• Estimation of model parameters under Generalized Linear Models (GLM):

```
# t1: time of first capture; Ti: number of times captured
library(foreign)
data<-read.dta(''cou.glmm.dta")
tau <- 10 # number of capture occasions
attach(data)
D <- nrow(data) # number of individuals captured at least once
```

```
h <- tau - t1   # number of trials
R <- (Ti- 1)/h  # observed proportions
R[is.na(R)] <- 0
require(MASS);require(nlme);require(PL.popN)
cou.glm<-glm(R~1,family=binomial,weight=h,data=data)
summary(cou.glm)
PL.popN(cou.glm,tau,D)
cou.glm<-glm(R~age,family=binomial,weight=h,data=data)
summary(cou.glm)
PL.popN(cou.glm,tau,D)
cou.glm<-glm(R~sex,family=binomial,weight=h,data=data)
summary(cou.glm)
PL.popN(cou.glm,tau,D)
cou.glm<-glm(R~age+sex,family=binomial,weight=h,data=data)
summary(cou.glm)
PL.popN(cou.glm,tau,D)
```

• Estimation of model parameters under Generalized Linear Mixed Effects Models (GLMM):

```
# t1: time of first capture; Ti: number of times captured
library(foreign)
data<-read.dta(``cou.glmm.dta")
tau <- 10 # number of capture occasions
attach(data)
D <- nrow(data) # number of individuals captured at least once
h <- tau - t1   # number of trials
R <- (Ti- 1)/h  # observed proportions
R[is.na(R)] <- 0
require(MASS);require(nlme);require(PL.popN)
nID <- 1 : D
cou.glmm<-glmmPQL(R~1,random=~1|nID,family=binomial,verbose=FALSE)
summary(cou.glmm)
PL.popN(cou.glmm,tau,D)
cou.glmm<-glmmPQL(R~age,random=~1|nID,family=binomial,verbose=FALSE)
summary(cou.glmm)
PL.popN(cou.glmm,tau,D)
cou.glmm<-glmmPQL(R~sex,random=~1|nID,family=binomial,verbose=FALSE)
summary(cou.glmm)
PL.popN(cou.glmm,tau,D)
cou.glmm<-glmmPQL(R~age+sex,random=~1|nID,family=binomial,verbose=FALSE)
summary(cou.glmm)
PL.popN(cou.glmm,tau,D)
```

• Simulation code under model $\mathrm{M_h}$ using GLM approach:

```
# nsim: number of Monte Carlo replicates; N: population size;
m: capture occasions; npar: number of parameters
set.seed(999)
sim<-function(nsim,N,m,b0){
npar<-5
para<-matrix(0,nsim,npar)
for (s in 1:nsim){
sex<-rbinom(N,1,0.5)           # individual discrete covariate
weight<-rnorm(N,mean=15,sd=2) # individual continuous covariate
zi<-rnorm(N,0,1)               # random effect
y <- p <- matrix(NA, m,N)
for (i in c(1:N)) {
p[, i] <- plogis(b0+0.1*sex[i]+0.2*weight[i]+0.1*zi[i])
y[, i] <- rbinom(m, 1, p[,i])
        }
cap<-t(y)     # population capture histories
Ti<-apply(cap,1,sum)    # number of times captured
z1<-z2<-z3<-z4<-z5<-z6<-numeric(length(cap[,1]))
for (i in 1:length(cap[,1])){
if (cap[i,1]==1)z1[i]<-1 else z1[i]<-0
if (cap[i,2]==1&cap[i,1]==0)z2[i]<-2 else z2[i]<-0
if (cap[i,3]==1&cap[i,1]==0&cap[i,2]==0)z3[i]<-3 else z3[i]<-0
if (cap[i,4]==1&cap[i,1]==0&cap[i,2]==0&cap[i,3]==0)z4[i]<-4
else z4[i]<-0
if (cap[i,5]==1&cap[i,1]==0&cap[i,2]==0&cap[i,3]==0&cap[i,4]==0)
z5[i]<-5 else z5[i]<-0
if (cap[i,6]==1&cap[i,1]==0&cap[i,2]==0&cap[i,3]==0&cap[i,4]==0&
cap[i,5]==0)z6[i]<-6 else z6[i]<-0
        }
t1<-z1+z2+z3+z4+z5+z6  # time of first capture
pop.data <- data.frame(cap,sex,weight,Ti,t1)
sex<-subset(sex,Ti>0)
weight<-subset(weight,Ti>0)
t1<-subset(t1,Ti>0)
Ti<-subset(Ti,Ti>0)
data<-data.frame(sex, weight, Ti, t1)
D <- nrow(data) # number of individuals captured at least once
h <- m - t1      # number of trials
R <- (Ti- 1)/h  # observed proportions
R[is.na(R)] <- 0
require(MASS);require(nlme);require(PL.popN)
```

```
glm <- glm(R ~ sex+weight, family = binomial, weight=h, data=data)
est<-PL.popN(glm, m, D)
para[s,1]<-est[4]$N.hat       # estimated population size
para[s,2]<-est[5]$N.hat.se    # standard error
para[s,3]<-D
para[s,1]- 1.96*para[s,2]->para[s,4]
para[s,1]+ 1.96*para[s,2]->para[s,5]
         }
para
acind<-mean(para[,3])
cat(``The average capture individuals is\n")
print(acind)
mu<-mean(para[,1])
cat(``The average estimated population size is\n")
print(mean(para[,1]))
se<-sqrt(var(para[,1]))
cat(``The standard error of the estimated population is\n")
print(se)
prb<-(mean(para[,1])-N)/N*100
cat(``The percentage relative bias(PRB) of the estimated population is\n")
print(prb)
cv<-se/mean(para[,1])*100
cat(``The percentage cv of the estimated population is\n")
print(cv)
rmse<-sqrt(var(para[,1])+(mean(para[,1])-N)^2)
cat(``The root mean square error (RMSE) of the estimated population is\n")
print(rmse)
require(Rlab)
count(N<para[,4]) -> bad.lower
count(para[,5]<N) -> bad.upper
civ<-(nsim-(bad.lower + bad.upper))/nsim*100
cat(``The 95% confidence interval coverage (COV) is\n")
print(civ)
         }
```

• Simulation code under model $\mathrm{M_h}$ using GLMM approach:

```
# nsim: number of Monte Carlo replicates; N: population size;
m: capture occasions; npar: number of parameters
library(foreign)
set.seed(999)
sim<-function(nsim,N,m,b0){
npar<-5
para<-matrix(0,nsim,npar)
for (s in 1:nsim){
sex<-rbinom(N,1,0.5)              # individual discrete covariate
weight<-rnorm(N,mean=15,sd=2)  # individual continuous covariate
zi<-rnorm(N,0,1)                  # random effect
y <- p <- matrix(NA, m,N)
for (i in c(1:N)) {
p[, i] <- plogis(b0+0.1*sex[i]+0.2*weight[i]+0.1*zi[i])
y[, i] <- rbinom(m, 1, p[,i])
}
cap<-t(y)   # population capture histories
Ti<-apply(cap,1,sum) # number of times captured
z1<-z2<-z3<-z4<-z5<-z6<-numeric(length(cap[,1]))
for (i in 1:length(cap[,1])){
if (cap[i,1]==1)z1[i]<-1 else z1[i]<-0
if (cap[i,2]==1&cap[i,1]==0)z2[i]<-2 else z2[i]<-0
if (cap[i,3]==1&cap[i,1]==0&cap[i,2]==0)z3[i]<-3 else z3[i]<-0
if (cap[i,4]==1&cap[i,1]==0&cap[i,2]==0&cap[i,3]==0)z4[i]<-4
else z4[i]<-0
if (cap[i,5]==1&cap[i,1]==0&cap[i,2]==0&cap[i,3]==0&cap[i,4]==0)
z5[i]<-5 else z5[i]<-0
if (cap[i,6]==1&cap[i,1]==0&cap[i,2]==0&cap[i,3]==0&cap[i,4]==0&
cap[i,5]==0)z6[i]<-6 else z6[i]<-0
          }
t1<-z1+z2+z3+z4+z5+z6     # time of first capture
pop.data <- data.frame(cap,sex,weight,Ti,t1)
data<- subset(pop.data, Ti>0)
sex<-subset(sex,Ti>0)
weight<-subset(weight,Ti>0)
t1<-subset(t1,Ti>0)
Ti<-subset(Ti,Ti>0)
D <- nrow(data)    # number of individuals captured at least once
h <- m - t1        # number of trials
R <- (Ti- 1)/h     # observed proportions
R[is.na(R)] <- 0
```

```
nID <- 1 : D
require(MASS);require(nlme);require(simex);require(sm);
require(PL.popN)
glmm<-glmmPQL(R~sex+weight,random=~1|nID,family=binomial,
data=data,verbose=FALSE)
est<-PL.popN(glmm, m, D)
para[s,1]<-est[4]$N.hat
para[s,2]<-est[5]$N.hat.se
para[s,3]<-D
para[s,1]- 1.96*para[s,2]->para[s,4]
para[s,1]+ 1.96*para[s,2]->para[s,5]
          }
para
acind<-mean(para[,3])
cat(‘‘The average capture individuals is\n”)
print(acind)
mu<-mean(para[,1])
cat(‘‘The average estimated population size is\n”)
print(mu)
se<-sqrt(var(para[,1]))
cat(‘‘The standard error of the estimated population is\n”)
print(se)
prb<-(mean(para[,1])-N)/N*100
cat(‘‘The percentage relative bias(PRB) of the estimated population is\n”)
print(prb)
cv<-se/mean(para[,1])*100
cat(‘‘The percentage cv of the estimated population is\n”)
print(cv)
rmse<-sqrt(var(para[,1])+(mean(para[,1])-N)^2)
cat(‘‘The root mean square error (RMSE) of the estimated population is\n”)
print(rmse)
require(Rlab)
count(N<para[,4]) -> bad.lower
count(para[,5]<N) -> bad.upper
civ<-(nsim-(bad.lower + bad.upper))/nsim*100
cat(‘‘The 95% confidence interval coverage (COV) is\n”)
print(civ)
          }
```

• Simulation code under model $\mathrm{M_h}$ using GEE approach:

```
# nsim: number of Monte Carlo replicates; N: population size;
m: capture occasions; npar: number of parameters
library(foreign)
set.seed(999)
sim<-function(nsim,N,m,initial,b_0){
npar<-5
para<-matrix(0,nsim,npar)
for (s in 1:nsim){
sex<-rbinom(N,1,0.5)          # individual discrete covariate
weight<-rnorm(N,mean=15,sd=2) # individual continuous covariate
zi<-rnorm(N,0,1)              # random effects
c <- p <- matrix(NA, m,N)
for (i in c(1:N)) {
p[, i]<-plogis(b_0+0.1*sex[i]+0.2*weight[i]+0.1*zi[i])
c[, i]<-rbinom(m, 1, p[,i])
        }
c.mat<-t(c)   # population capture histories
c.hist<-c.mat[apply(c.mat,1,sum)>0,] # observed capture histories
cap<-c(t(c.hist))
Ti<-apply(c.mat,1,sum)  # number of times captured
n<-nrow(c.hist)         # number of individuals captured at least once
id <- gl(n, m, n*m)     # individual id
flno<- rep(1:m, n)      # capture occasions
sex<-rep(subset(sex,Ti>0), each=m)
weight<-rep(subset(weight,Ti>0), each=m)
Ti<-rep(subset(Ti,Ti>0),each=m)
data<-data.frame(id, flno, cap,sex, weight, Ti)
flno<-data[,2]
cap<-y<-data[,3]
sex<-data[,4]
weight<-data[,5]
count<-0
repeat{
        b0<-initial[1]
        b1<-initial[2]
        b2<-initial[3]
        score<-c(0,0,0)
        inf<-matrix(c(0,0,0,0,0,0,0,0,0),ncol=3, byrow=T)
        rij<-0
        k<-0
for(i in 1: length(flno)){
```

```
if(flno[i]==1&&flno[i+1]==2&&flno[i+2]==3&&flno[i+3]==4&&flno[i+4]==5
&&flno[i+5]==6){
        k<-k+1
        m1<-exp(b0+b1*sex[i]+b2*weight[i])/
        (1+exp(b0+b1*sex[i]+b2*weight[i]))
        m2<-exp(b0+b1*sex[i+1]+b2*weight[i+1])/
        (1+exp(b0+b1*sex[i+1]+b2*weight[i+1]))
        m3<-exp(b0+b1*sex[i+2]+b2*weight[i+2])/
        (1+exp(b0+b1*sex[i+2]+b2*weight[i+2]))
        m4<-exp(b0+b1*sex[i+3]+b2*weight[i+3])/
        (1+exp(b0+b1*sex[i+3]+b2*weight[i+3]))
        m5<-exp(b0+b1*sex[i+4]+b2*weight[i+4])/
        (1+exp(b0+b1*sex[i+4]+b2*weight[i+4]))
        m6<-exp(b0+b1*sex[i+5]+b2*weight[i+5])/
        (1+exp(b0+b1*sex[i+5]+b2*weight[i+5]))
pi<-1-((1-m1)*(1-m2)*(1-m3)*(1-m4)*(1-m5)*(1-m6))
mu<-c(m1/pi,m2/pi,m3/pi,m4/pi,m5/pi,m6/pi)
v<-c(mu[1]*(1-mu[1]),mu[2]*(1-mu[2]),mu[3]*(1-mu[3]),mu[4]*(1-mu[4]),
mu[5]*(1-mu[5]),mu[6]*(1-mu[6]))
ai<-matrix(c(v[1],0,0,0,0,0,0,v[2],0,0,0,0,0,0,v[3],0,0,0,0,0,0,v[4],
0,0,0,0,0,0,v[5],0,0,0,0,0,0,v[6]),ncol=m,byrow=T)
ymi<-c(y[i]-mu[1],y[i+1]-mu[2],y[i+2]-mu[3],y[i+3]-mu[4],y[i+4]-mu[5],
y[i+5]-mu[6])
xit<-matrix(c(1,1,1,1,1,1,sex[i],sex[i+1],sex[i+2],sex[i+3],sex[i+4],
sex[i+5],weight[i],weight[i+1],weight[i+2],weight[i+3],weight[i+4],
weight[i+5]),ncol=m,byrow=T)
xi<-t(xit)
score<-score+xit%*%ymi
inf<-inf+xit%*%ai%*%xi
        }
                }
infinv<-solve(inf)
est<-initial+infinv%*%score
conv<-abs(score)
if(conv[1]<=0.0001&&conv[2]<=0.0001&&conv[3]<=0.0001)
break
initial<-est
pij<-exp(est[1]+sex*est[2]+weight*est[3])/
(1+exp(est[1]+sex*est[2]+weight*est[3]))
        }
p<-matrix(c(pij), ncol=m,byrow=T)
prod<-1
for(k in 1:m){
prod<-prod*(1-p[,k])}
```

```
pi<-1-prod
Nhat<-sum(1/pi)
varA<-sum((1-pi)/(pi)^2)
delta.beta<-xit%*%(pi^(-2)*(1-p)^m * m * p)
varB<-(t(delta.beta)%*%infinv)%*%delta.beta
varN<-varA+varB
se.Nhat<-sqrt(varN)
para[s,1]<-Nhat
para[s,2]<-se.Nhat
para[s,3]<-n
para[s,1]- 1.96*para[s,2]->para[s,4]
para[s,1]+ 1.96*para[s,2]->para[s,5]
        }
para
acind<-mean(para[,3])
cat(``The average capture individuals is\n")
print(acind)
mu<-mean(para[,1])
cat(``The average estimated population size is\n")
print(mean(para[,1]))
se<-sqrt(var(para[,1]))
cat(``The standard error of the population estimate\n")
print(se)
prb<-(mean(para[,1])-N)/N*100
cat(``The percentage relative bias(PRB) of the estimated population is\n")
print(prb)
cv<-se/mean(para[,1])*100
cat(``The percentage cv of the estimated population is\n")
print(cv)
rmse<-sqrt(var(para[,1])+(mean(para[,1])-N)^2)
cat(``The root mean square error (RMSE) of the estimated population is\n")
print(rmse)
require(Rlab)
count(N<para[,4]) -> bad.lower
count(para[,5]<N) -> bad.upper
civ<-(nsim-(bad.lower + bad.upper))/nsim*100
cat("The 95% confidence interval coverage (COV) is\n")
print(civ)
        }
```

Appendix C

Motivating Data

Table C1: Capture of Deer mice (*Peromyscus maniculatus*) collected by V. Reid at East Stuart Gulch, Colorado, USA. The columns represented the sex (m: male, f: female), the ages (y: young, sa: semi-adult, a: adult), the weights in grams, and the capture histories of 38 distinct individuals over 6 trapping occasions (1: captured, 0: not captured)

sex	age	weight	Capture occasion						sex	age	weight	Capture occasion					
m	y	12	1	1	1	1	1	1	m	y	13	0	1	1	0	1	0
f	y	15	1	0	0	1	1	1	f	y	5	0	1	0	1	0	1
m	y	15	1	1	0	0	1	1	f	a	20	0	1	0	0	0	1
m	y	15	1	1	0	1	1	1	m	y	12	0	1	0	0	1	1
m	y	13	1	1	1	1	1	1	f	y	6	0	0	1	0	0	0
m	a	21	1	1	0	1	1	1	f	a	22	0	0	1	1	1	1
m	y	11	1	1	1	1	1	0	f	y	10	0	0	1	0	1	1
m	sa	15	1	1	1	0	0	1	f	y	14	0	0	1	1	1	1
m	y	14	1	1	1	1	1	1	f	a	19	0	0	1	0	0	0
m*	y	14	1	1	0	1	1	1	f*	a	19	0	0	1	0	1	0
m	y	13	1	1	0	1	1	1	f	a	20	0	0	0	1	0	0
f	a	22	1	1	1	0	1	1	m	sa	16	0	0	0	1	1	1
m	y	14	1	1	1	1	1	1	f	y	11	0	0	0	1	1	0
m	y	11	1	0	1	1	1	0	m	y	14	0	0	0	0	1	0
f	y	10	1	0	0	1	0	0	f	y	11	0	0	0	0	1	0
f	a	23	0	1	0	0	1	0	m	a	24	0	0	0	0	1	0
f	y	7	0	1	1	0	0	1	m	y	9	0	0	0	0	0	1
m	y	8	0	1	0	0	0	1	m	sa	16	0	0	0	0	0	1
m	a	19	0	1	0	1	0	1	f	a	19	0	0	0	0	0	1

Note: The semi-adults (sa) were regrouped as adults (a) in the analysis, as in Huggins (1991). The two individuals marked with asterisks are missing from Appendix 1 of Huggins (1991).

Table C2: Capture of Least chipmunk (*Eutamias minimus*) collected by V. Reid at Colorado, USA. The columns represented the sex (m: male, f: female)and the capture histories of 45 distinct individuals over 6 trapping occasions (1: captured, 0: not captured)

sex	Capture occasion						sex	Capture occasion					
f	1	0	0	0	0	0	f	0	0	1	0	0	0
m	1	1	1	0	1	0	m	0	0	1	0	0	0
f	1	0	0	0	0	0	m	0	0	1	0	0	0
f	1	0	0	1	0	0	f	0	0	1	1	0	0
m	1	0	0	0	0	0	f	0	0	1	0	0	0
f	1	0	1	0	1	0	f	0	0	0	1	0	0
f	1	0	1	0	0	0	f	0	0	0	1	0	0
m	0	1	1	1	1	0	f	0	0	0	1	0	0
m	0	1	0	1	0	1	m	0	0	0	1	0	0
m	0	1	0	0	0	0	m	0	0	0	1	0	0
f	0	1	1	1	1	1	m	0	0	0	1	0	0
f	0	1	1	1	0	0	m	0	0	0	1	0	0
m	0	1	1	1	0	0	m	0	0	0	1	1	0
f	0	1	0	1	1	1	f	0	0	0	1	0	0
f	0	1	0	1	1	0	f	0	0	0	1	0	0
m	0	1	1	0	1	0	m	0	0	0	1	1	0
m	0	1	0	0	0	1	f	0	0	0	1	1	0
f	0	1	0	0	1	0	f	0	0	0	0	1	1
f	0	1	1	1	1	1	m	0	0	0	0	1	0
f	0	1	0	0	0	0	m	0	0	0	0	1	1
m	0	1	0	1	0	0	m	0	0	0	0	1	0
f	0	0	1	0	1	0	m	0	0	0	0	1	0
m	0	0	1	1	1	0							

Table C3: Capture of House mouse (*Mus musculus*) collected by Coulombe (1965) at Ballana Creek, Los Angeles County, California. The columns represented the sex (1: male, 0: female), the ages (0: young, 1: adult), and the capture histories of 171 distinct individuals over 10 trapping occasions (1: captured, 0: not captured)

age	sex	Capture occasion										age	sex	Capture occasion									
1	0	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	1	1	0	0	0	0
1	1	1	1	1	1	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
1	0	1	0	1	0	0	1	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
0	1	1	1	0	0	0	0	1	0	1	1	1	0	1	0	1	0	1	1	0	0	1	1
1	1	1	0	0	0	1	0	0	1	0	0	1	1	1	0	1	0	0	0	1	0	0	0
0	0	1	0	0	0	1	0	1	0	1	0	1	1	1	0	1	0	0	0	0	0	0	0
1	1	1	1	0	1	0	0	1	0	1	0	1	0	1	1	1	1	0	0	1	0	1	0
0	0	1	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	1	0	1	1	1	1
1	1	1	1	0	0	1	1	0	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1
1	0	1	0	1	1	0	1	0	0	1	1	0	1	1	0	1	1	1	0	1	0	1	0
1	0	1	0	0	1	1	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0
1	1	1	0	0	0	1	1	0	0	0	0	1	1	1	0	0	1	1	1	1	0	0	0
1	0	1	0	1	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	1
1	0	1	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	1	0	0	0
1	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	0	0
1	0	1	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0
1	1	1	0	0	1	0	0	1	0	0	0	1	0	1	0	1	0	1	0	0	0	0	0
1	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0
0	0	1	0	0	0	1	0	1	1	1	0	1	0	1	1	1	1	1	1	1	0	1	0
1	1	1	1	0	1	1	0	1	1	0	0	1	0	1	1	1	0	1	0	1	0	1	0
1	0	1	1	0	0	1	0	1	0	0	0	1	1	1	0	1	0	0	0	1	0	1	0
0	0	1	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	1	0
0	1	1	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1	1	1	0
0	1	1	0	0	1	0	0	0	0	0	0	1	0	1	1	1	0	1	1	0	0	1	1
0	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0	1	1	0
1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	0
1	0	1	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	0	1	0
1	0	1	1	0	1	1	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0
1	0	1	1	0	0	0	0	1	0	0	0	1	0	0	1	1	1	1	1	1	0	0	0
1	0	1	0	1	0	1	1	1	0	0	1	1	0	0	1	1	0	0	1	0	1	0	0
0	1	1	0	1	1	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	1	0	0
1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0
1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	1	1	0	1	1
0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	1	0	1	1
1	1	1	1	0	1	1	0	1	0	1	0	0	0	0	1	1	1	1	0	1	0	1	1
1	1	1	1	0	0	0	0	1	1	1	0	1	1	0	1	0	0	0	1	1	0	0	0
1	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1
0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0
0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	1	0
1	1	1	1	0	0	1	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1
1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0

Continued on next page

age	sex	Capture occasion									
0	0	0	1	0	0	1	1	0	0	0	0
0	0	0	1	1	0	1	1	1	0	0	0
0	0	0	1	1	0	1	1	0	0	0	0
0	1	0	1	1	0	0	0	1	0	1	0
0	1	0	1	0	0	0	0	0	1	1	0
0	1	0	1	1	1	0	1	0	0	1	0
1	0	0	1	0	1	1	0	0	0	1	1
0	1	0	1	0	0	0	0	1	0	0	0
1	0	0	1	0	1	0	0	0	0	0	1
1	0	0	1	0	0	1	0	0	0	1	1
0	0	0	1	0	1	0	0	0	1	0	0
0	1	0	1	0	0	0	0	1	0	0	0
1	0	0	1	0	0	1	0	0	0	0	0
0	1	0	1	0	0	0	0	1	0	1	1
0	1	0	1	0	0	1	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	1	0
0	1	0	0	1	0	0	0	0	0	1	0
0	1	0	0	1	0	1	1	0	0	0	0
1	1	0	0	1	1	0	0	0	0	1	0
0	1	0	0	1	1	0	0	0	0	0	0
0	0	0	0	1	0	1	0	1	0	1	1
1	1	0	0	1	0	1	0	0	0	0	0
0	1	0	0	1	0	0	0	0	1	0	0
1	0	0	0	1	0	1	0	0	1	0	1
1	0	0	0	1	0	0	0	0	0	1	0
0	1	0	0	1	0	0	0	1	0	1	0
1	0	0	0	1	0	0	0	0	0	1	0
0	1	0	0	1	0	0	0	1	0	0	1
1	0	0	0	1	0	1	0	0	0	0	1
1	0	0	0	1	0	1	0	1	1	0	0
0	1	0	0	1	0	0	0	0	0	1	0
1	1	0	0	1	0	0	1	0	0	1	0
1	0	0	0	1	1	0	0	0	0	0	0
0	1	0	0	1	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	1
1	1	0	0	1	0	0	0	0	1	0	0
0	1	0	0	1	0	1	0	0	0	0	1
0	0	0	0	1	0	1	0	1	1	0	0
1	0	0	0	1	0	0	0	1	0	0	0
1	0	0	0	1	0	1	0	0	0	1	0
1	0	0	0	1	0	1	0	1	0	1	1
0	1	0	0	0	1	0	0	1	0	1	0
0	1	0	0	0	1	0	1	1	1	0	0
0	1	0	0	0	1	0	1	0	1	0	0
1	1	0	0	0	1	0	1	0	0	0	0
1	1	0	0	0	1	0	0	0	1	0	0
0	1	0	0	0	1	0	1	0	0	0	0
0	0	0	0	0	1	0	0	1	0	0	0
1	0	0	0	0	1	0	1	1	0	1	0
0	1	0	0	0	1	0	0	0	0	1	0
0	1	0	0	0	1	0	0	1	0	0	1
0	1	0	0	0	1	0	0	0	0	1	0
1	0	0	0	0	1	0	0	0	0	1	0
1	1	0	0	0	0	1	0	0	1	0	0
0	1	0	0	0	0	1	0	1	0	0	0
0	1	0	0	0	0	1	0	0	1	0	1
0	0	0	0	0	0	1	0	0	0	0	1
1	1	0	0	0	0	1	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	1	0
1	1	0	0	0	0	1	0	0	1	0	0
0	0	0	0	0	0	1	0	1	0	1	0
1	0	0	0	0	0	1	0	0	0	1	0
1	0	0	0	0	0	1	0	1	0	1	0
1	1	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0
0	1	0	0	0	0	1	0	0	0	1	0
1	0	0	0	0	0	1	0	1	0	0	0
0	1	0	0	0	0	1	1	0	0	0	0
1	0	0	0	0	0	0	1	0	0	1	0
0	1	0	0	0	0	0	1	0	1	0	1
0	0	0	0	0	0	0	1	1	0	1	0
1	1	0	0	0	0	0	0	1	0	1	0
1	1	0	0	0	0	0	0	1	1	0	0
0	1	0	0	0	0	0	0	1	1	0	0
1	1	0	0	0	0	0	0	1	0	0	1
0	0	0	0	0	0	0	0	1	0	1	0
1	0	0	0	0	0	0	0	1	1	1	0
0	1	0	0	0	0	0	0	1	0	1	0
0	1	0	0	0	0	0	0	1	1	1	0
1	1	0	0	0	0	0	0	1	1	0	1
1	1	0	0	0	0	0	0	1	1	0	0
0	1	0	0	0	0	0	0	1	0	1	0
0	1	0	0	0	0	0	0	1	0	1	0
1	0	0	0	0	0	0	0	0	0	1	1
1	0	0	0	0	0	0	0	0	0	1	1

Bibliography

Agresti, A. (1994). 'Simple capture-recapture models permitting unequal catchability and variable sampling effort.' *Biometrics*, **50**, pp.494–500.

Akanda, M. A. S. and Khanam, M. (2011). *Goodness-of-fit tests for GEE model: Methods and applications.* VDM Verlag, Germany.

Akaike, H. (1973). 'Information theory and an extension of the maximum likelihood principle'. In: B.N. Petrov and F. Csaki (eds) *2nd International Symposium on Information Theory*, Akademiai Kiado, Budapest, pp.267–281.

Akaike, H. (1974). 'A new look at statistical model identification.' *IEEE Transactions on Automatic Control, AC-19*, pp.716–723.

Alho, J. (1990). 'Logistic regression in capture-recapture models.' *Biometrics*, **46**, pp.623–635.

Amstrup, S. C., McDonald, T. L. and Manly, B. F. J. (eds.). (2005). *Handbook of capture-recapture analysis.* Princeton University Press: Princeton, New Jersey.

Anderson, D. A. and Aitkin, M. (1985). 'Variance component models with binary response: interviewer variability.' *Journal of the Royal Statistical Society, Series B*, **47**, pp.203–210.

Bahadur, R. R. (1961). *A representation of the joint distribution of responses to n dichotomous items. In Studies in Item Analysis and Prediction.* H. Solomon (ed.), Stanford Mathematical Studies in the Social Sciences VI. Stanford: Stanford University Press, pp.158–168.

Baillargeon, S. and Rivest, L. P. (2007). 'Rcapture: Loglinear models for capture-recapture in R.' *Journal of Statistical Software*, **19(5)**, pp.1-31. http://www.jstatsoft.org/v19/i05.

Ballinger, G. (2004). 'Using generalized estimating equations for longitudinal data analysis.' *Organizational Research Methods*, **7(2)**, pp.127–150.

Barry, S. C., Brooks, S. P., Catchpole, E. A. and Morgan, B. J. T. (2003). 'The analysis of ring-recovery data using random effects.' *Biometrics*, **59**, pp.54–65.

Basu, S. and Ebrahimi, N. (2001). 'Bayesian capture-recapture methods for error detection and estimation of population size: heterogeneity and dependence.' *Biometrika*, **88**, pp.269–279.

Breslow, N. E. and Clayton, D. G. (1993). 'Approximate inference in generalized linear mixed models.' *Journal of the American Statistical Association*, **88 (421)**, pp.9–25.

Burnham, K. P. (1972). *Estimation of population size in multinomial capture-recapture studies when capture probabilities vary among animals.* PhD thesis, Oregon State University, Oregon.

Burnham, K. P. and Overton, W. S. (1978). 'Estimation of the size of a closed population when capture probabilities vary among animals.' *Biometrika*, **65**, pp.625–633.

Carruthers, E., Lewis, K., McCue, T. and Westley, P. (2008). *Generalized linear models: model selection, diagnostics, and overdispersion.* Memorial University of Newfoundland, http://www.mun.ca/biology/dschneider/b7932/B7932Final4Mar2008.pdf. Accessed 13 April 2013.

Castledine, B. (1981). 'A Bayesian analysis of multiple-recapture sampling for a closed population.' *Biometrika*, **67(1)**, pp.197–210.

Chao, A. (1988). 'Estimating animal abundance with capture frequency data.' *Journal of Wildlife Management*, **52**, pp.295–300.

Chao, A. (1998). 'Capture-recapture,' in *Encyclopedia of Biostatistics*, eds. Armitage, P. and Colton, T., New York: Wiley, pp.482–486.

Chao, A. (2001). An overview of closed capture-recapture models. *Journal of Agricultural, Biological and Environmental Statistics*, **6**, pp.158–75.

Chao, A. and Lee, S. M. (1992). 'Estimating the number of classes via sample coverage.' *Journal of the American Statistical Association*, **87(417)**, pp.210–217.

Chao, A. and Lee, S. M. (1993). 'Estimating population size for continuous time capture-recapture models via sample coverage.' *Biometrical Journal*, **35**, pp.29–45.

Chao, A., Lee, S. M. and Jeng, S. L. (1992). 'Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal.' *Biometrics*, **48**, pp. 201–216.

Chao, A. and Huggins, R. M. (2005). *Modern closed-population capture-recapture models.* In: Amstrup C,McDonald TL, Manly BFJ (eds) Handbook of capture-recapture analysis. Princeton University Press, Princeton, NJ, pp 58–87.

Choquet, R., Sanz-Aguilar, A., Doligez, B., Nogue, E., Pradel, R., Gustafsson, L. and Gimenez, O. (2013). ‘Estimating demographic parameters from capture-recapture data with dependence among individuals within clusters.’ *Methods in Ecology and Evolution*, **4**, pp.474–482.

Cormack, R. M. (1968). ‘The statistics of capture-recapture.’ *Oceanography and Marine Biology Annual Review*, **6**, pp.455–506.

Cormack, R. M. and Buckland, S. T. (1997). ‘Capture and recapture (Update)’, in *Encyclopedia of Statistical Sciences* (Update vol. 1), eds. Kotz, S., Read, C. B. and Banks,D. L., New York: Wiley, pp.79–84.

Coull, B. A. and Agresti, A. (1999). ‘The use of mixed logit models to reflect heterogeneity in capture-recapture studies.’ *Biometrics*, **55**, pp.294–301.

Coull, B. A. and Agresti, A. (2000). ‘Random effects modeling of multivariate binomial responses using the multivariate binomial logit-normal distribution.’ *Biometrics*, **56**, pp.73–80.

Cui, J. (2007). ‘QIC program and model selection in GEE analyses.’ *The Stata Journal*, **7(2)**, pp.209–220.

Crowder, M. (1995). ‘On the use of a working correlation matrix in using generalised linear models for repeated measures.’ *Biometrika*, **82(2)**, pp.407–410.

Cox, D. R. (1970). *Analysis of binary data.* London: Methuen.

Dahmen, G. and Ziegler, A. (2003). ‘Generalized estimating equations in controlled clinical trials: Hypothesis testing.’ *Biometrical Journal*, **46**, pp.214–232.

Dale, J. R. (1986). ‘Global cross-ratio models for bivariate discrete ordered responses.’ *Biometrics*, **42**, pp.909–917.

Darroch, J. N. (1958). ‘The multiple recapture census. I: Estimation of a closed population.’ *Biometrika*, **45**, pp.343–359.

Davis, R., Dunsmuir, W. and Wang, Y. (2000). ‘On autocorrelation in a Poisson regression model.’ *Biometrika*, **87**, pp.491–505.

Diggle, P. J. (1988). ‘An approach to the analysis of repeated measures.’ *Biometrics*, **44**, pp.959–971.

Diggle, P. J., Heagerty, P. J., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of longitudinal data.* Oxford University Press, New York.

Dobson, A. J. (2010). *An introduction to generalized linear models.* Second Edition, Chapman & Hall/CRC, New York.

Dobson, D. P. A. and Williams, G. (2003). *Application of generalized estimating equations to longitudinal data.* Longitudinal Studies, The University of Queensland.

Dorazio, R. M. and Royle, J. A. (2003). ‘Mixture models for estimating the size of a closed population when capture rates vary among individuals.’ *Biometrics*, **59**, pp.351–364.

Fitzmaurice, G. M. (1995). ‘A caveat concerning independence estimating equations with multiple multivariate binary data.’ *Biometrics*, **51**, pp.309–317.

Gazey, W. J. and Staley, M. J. (1986). ‘Population estimation from mark-recapture experiments using a sequential Bayes algorithm.’ *Ecology*, **67**, pp.941–951.

George, E. and Robert, C. (1992). ‘Capture-recapture sampling via gibbs sampling.’ *Biometrika*, **79**, pp.677–683.

Ghosh, S. K. and Norris, J. (2005). 'Bayesian capture-recapture analysis of a closed population allowing for heterogeneity between animals.' *Journal of Agricultural Biological and Environmental Statistics*, **10**, pp.35–49.

Gill, J. (2001). *Generalized linear models: A unified approach.* A SAGE University Press.

Gilmour, A. R., Anderson, R. D. and Rae, A. L. (1985). 'The analysis of binomial data by a generalized linear mixed model.' *Biometrika*, **72**, pp.593–599.

Gimenez, O. and Choquet, R. (2010). 'Incorporating individual heterogeneity in studies on marked animals using numerical integration: Capture-recapture mixed models.' *Ecology*, **91**, pp.951–957.

Glonek, G. F. V. and McCullagh, P. (1995). 'Multivariate logistic models.' *Journal of the Royal Statistical Society, Series B*, **57**, pp.533–546.

Goodman, L. A. and Kruskal, W. H. (1979). *Measures of association for cross-classifications.* Springer-Verlag: New York.

Goldstein, H. and Rasbash, J. (1996). 'Improved approximations for multilevel models with binary responses.' *Journal of the Royal Statistical Society, Series A*, **159**, pp.505–513.

Gosky, R. and Ghosh, S. K. (2009). 'A comparative study of Bayesian model selection criteria for capture-recapture models for closed populations.' *Journal of Modern Applied Statistical Methods*, **9**, pp.68–80.

Gosky, R. and Ghosh, S. K. (2011). 'A comparative study of Bayes estimators of closed population size from capture-recapture data.' *Journal of Statistical Theory and Practice*, **5**, pp.241–260.

Hardin, J. W. and Hilbe, J. M. (2003). *Generalized estimating equations.* Chapman and Hall, United States of America.

Heagerty, P. and Zeger, S. L. (1998). ‘Lorelogram: a regression approach to exploring dependence in longitudinal categorical responses.’ *Journal of the American Statistical Association*, **93**, pp.150–163.

Hedeker, D. (2005). *Longitudinal data analysis.* University of Illinois, Chicago, United States of America.

Hilborn, R., Redfield, J. A. and Krebs, C. J. (1976). ‘On the reliability of enumeration for mark and recapture census of voles.’ *Canadian Journal of Zoology*, **54**, pp.1019–1024.

Hojsgaard, S. and Halekoh, U. (2005). *Overdispersion.* Danish Institute of Agricultural Sciences, Copenhagen. http://gbi.agrsci.dk/statistics/courses. Accessed 13 April 2013.

Holzmann, H., Munk, A. and Zucchini, W. (2006). ‘On identifiability in capture-recapture models.’ *Biometrics*, **62**, pp.934–936.

Hook, E. B. and Regal, R. R. (1995). ‘Capture-recapture methods in epidemiology: methods and limitations.’ *Epidemiological Reviews*, **17**, pp.243–264.

Horton, N. J. and Lipstiz, S. R.(1999). ‘Review of software to fit generalized estimating equation regression models.’ *The American Statistican*, **53**, pp.160–169.

Horvitz, D. G. and Thompson, D. J. (1952). ‘A Generalization of sampling without replacement from a finite universe.’ *Journal of the American Statistical Association*, **47**, pp.663–685.

Huggins, R. M. (1989). ‘On the statistical analysis of capture experiments.’ *Biometrika*, **76**, pp.133–140.

Huggins, R. M. (1991). ‘Some practical aspects of a conditional likelihood approach to capture experiments.’ *Biometrics*, **47**, pp.725–732.

Huggins, R. M. and Hwang, W. H. (2011). 'A review of the use of conditional likelihood in capture-recapture experiments.' *International Statistical Review*, **73(3)**, pp.385–400.

Huggins, R. M. and Yip, P. S. (1997). 'Statistical analysis of removal experiments with the use of auxillary variables.' *Statistica Sinica*, **7**, pp.705–712.

Huggins, R. M. and Yip, P. S. (2001). 'A note on nonparametric inference for capture-recapture experiments with heterogeneous capture probabilities.' *Statistica Sinica*, **11**, pp.843–853.

Hwang, W. and Chao, A. (2002). 'Continuous-time capture-recapture models with covariates.' *Statistica Sinica*, **12**, pp.1115–1131.

Hwang, W. H. and Huggins, R. M. (2005). 'An examination of the effect of heterogeneity on the estimation of population size using capture-recapture data.' *Biometrika*, **92**, pp.229–233.

Hwang, W. H. and Huggins, R. M. (2007). 'Application of semiparametric regression models in the analysis of capture-recapture experiments.' *Australian and New Zealand Journal of Statistics*, **49**, pp.191–212.

International Working Group for Disease Monitoring and Forecasting.(1995a). 'Capture-recapture and multiple-record systems estimation. I: History and theoretical development.' *American Journal of Epidemiology*, **142**, pp.1047–1058.

International Society for Disease Monitoring and Forecasting (1995b). 'Mark-recapture and multiple-record systems estimation II Application in human disease.' *American Journal of Epidemiology*, **142** pp.1059–1068.

Jϕrgensen, B. (1987). 'Exponential dispersion models (with discussion).' *Journal of the Royal Statistical Society, Series B*, **49**, pp.127–162.

Kalbeisch, J. D. and Sprott, D. A. (1970). 'Application of likelihood methods to models involving large numbers of parameters.' *Journal of the Royal Statistical Society*, Series B **32**, pp.175–208.

Kalbeisch, J. D. and Sprott, D. A. (1973). 'Marginal and conditional likelihoods.' *Sankhya*, **35**, pp.311–328.

King, R. and Brooks, S. P. (2008). 'Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty.' *Biometrics*, **64**, pp.816–824.

Laird, N. M. and Ware, J. H. (1982). 'Random effects models for longitudinal data.' *Biometrics*, **38**, pp.963–974.

Lang, J. B. and Agresti, A. (1994). 'Simultaneously modeling joint and marginal distributions of multivariate categorical responses.' *Journal of the American Statistical Association*, **89**, pp.625–632.

Lee, S. M. and Chao, A. (1991). 'Estimating population size via sample coverage for closed capture-recapture models.' *Proceeding 1990 Taipei Symposium in Statistics*, Chao, M. T. and Cheng, P. E. (editors). Institute of Statistics Science, Academia Sinica.

Lee, S. and Chao, A. (1994). 'Estimating population size via sample coverage for closed capture-recapture models.' *Biometrics*, **50(1)**, pp.88–97.

Lee, S. M., Hwang, W. H. and Huang, L. H. (2003). 'Bayes estimation of population size from capture-recapture models with time variation and behavior response.' *Statistica Sinica*, **13**, pp.477–494.

Lee, Y. and Nelder, J. A. (1996). 'Hierarchical generalized linear models.' *Journal of the Royal Statistical Society, Series B*, **58**, pp.619–678.

Liang, K. Y. (1999). *Generalized linear models, estimating functions and multivariate extensions.* Johns Hopkins University.

Liang, K. Y., Zeger, S. L. and Qaqish, B. (1992). 'Multivariate regression analyses for categorical data.' *Journal of the Royal Statistical Society, Series B*, **54**, pp.3–40.

Liang, K. and Zeger, S. L. (1986). 'Longitudinal data analysis using generalized linear models.' *Biometrika*, **73(1)**, pp.13–22.

Liang, K. and Zeger, S. L. (1993). 'Regression analysis for correlated data.' *Annual Review of Public Health,* **14**, pp.43–68.

Lincoln, F. C. (1930). 'Calculating waterfowl abundance on the basis of banding returns.' *United States Department of Agriculture Circular*, **118**, 1–4.

Lindsey, J. K. (2007). *Applying generalized linear models.* Limburgs Universitair Centrum, Diepenbeek.

Link, W. A. (2003). 'Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities.' *Biometrics*, **59**, pp.1125–1132.

Lloyd, C. J. (1994). 'Efficiency of martingale methods in recapture studies.' *Biometrika*, **81**, pp.305–315.

Lloyd, C. J. and Yip, P. (1991). A unification of inference from capture-recapture studies through martingale estimating functions. In Estimating Functions, Ed. V. P. Godambe, pp. 65-88. Oxford: Clarendon Press.

Madigan, D. and York, J. C. (1997). 'Bayesian methods for estimation of the size of a closed population.' *Biometrika*, **84**, pp.19–31.

Manly, B .F. J. (1970). 'A simulation study of animal population estimation using the capture-recapture method.' *Journal of Applied Ecology*, **7**, pp.13–39.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models.* 2nd ed. London: Chapman & Hall.

McGilchrist, C. A. (1994). 'Estimation in Generalized Mixed Models.' *Journal of the Royal Statistical Society, Series B*, **56**, pp.61–69.

Molenberghs, G. and Lesaffre, E. (1994). 'Marginal modeling of correlated ordinal data using a multivariate Plackett distribution.' *Journal of the American Statistical Association*, **89**, pp.633–644.

Molenberghs, G. and Ryan, L. (1999). 'An exponential family model for clustered multivariate binary data.' *Environmetrics*, **10**, pp.279–300.

Nelder, J. A. and Wedderburn, R. W. M. (1972). 'Generalized linear models.' *Journals of the Royal Statistical Society, Series A*, **135**, pp.370–384.

Nichols, J. D. (1986). 'On the use of enumeration estimators for interspecific comparisons, with comments on a "trappability" estimator.' *Journal of Mammalogy*, **67**, pp.590–593.

Nichols, J. D. and Pollock, K. H. (1983). 'Estimation methodology in contemporary small mammal capture-recapture studies.' *Journal of Mammalogy*, **64**, pp.253–260.

Norris, J. L. and Pollock, K. H. (1996). 'Nonparametric MLE under two closed capture-recapture models with heterogeneity.' *Biometrics*, **52**, pp.639–649.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978), 'Statistical inference from capture data on closed animal populations.' *Wildlife Monograph*, **62**, pp.1–130.

Pan, W. (2001). 'Akaike's information criterion in generalized estimating equations.' *Biometrics*, **57**, pp.120–125.

Patil, G. P. (1962). 'Maximum likelihood estimation for generalized power series distributions and its application to a truncated binomial distribution.' *Biometrika*, **49(1&2)**, pp.227–237.

Petersen, C. G. J. (1896). 'The yearly immigration of young plaice into the Limfiord from the German Sea.' *Report of the Danish Biological Station (1895)*, **6**, 5–84.

Pinheiro, J. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* Statistics and Computing Series, Springer-Verlag, New York, NY.

Pledger, S. (2000). 'Unified maximum likelihood estimates for closed capture-recapture models using mixtures.' *Biometrics*, **56**, pp.434–442.

Pollock, K. H. (1974). *The assumption of equal catchabi1ity of animals in tag-recapture experiments.* Ph.D. Thesis, Cornell University, Ithaca, New York 14850.

Pollock, K. H. (1982). 'A capture recapture design robust to unequal probability of capture.' *Journal of Wildlife Management*, **46**, pp.752–757.

Pollock, K. H. (1991). 'Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future.' *Journal of the American Statistical Association.*, **86**, pp.225–238.

Pollock, K. H. (2002). 'The use of auxiliary variables in capture-recapture modeling: an overview.' *Journal of Applied Statistics*, **29**, pp.85–102.

Pollock, K. H., Hines, J. and Nichols, J. (1984). 'The use of auxiliary variables in capture-recapture and removal experiments.' *Biometrics*, **40**, pp.329–340.

Pollock, K. H., Nichols, J. D., Brownie, C. and Hines. J. E. (1990). *Statistical inference for capture-recapture experiments.* Wildlife Monographs. 107.

Pollock, K. H. and Otto, M. C. (1983). 'Robust estimation of population size in closed animal populations from capture-recapture experiments.' *Biometrics*, **39**, pp.1035–1049.

Pradel, R. and Sanz-Aguilar, A. (2012). 'Modeling trap-awareness and related phenomena in capture-recapture studies.' *PLoS ONE*, **7(3)**: e32666. doi:10.-1371/journal.pone.0032666.

Qaqish, B. F. (2003). 'A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations.' *Biometrika*, **90**, pp.455–463.

R Development Core Team, (2013). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rexstad, E. and Burnham, K. P. (1991). *User's guide for interactive program CAPTURE.* Colorado cooperative Fish and Wildlife Research Unit, Fort Collins.

Sartori, N. and Severini, T. A. (2004). 'Conditional likelihood inference in generalized linear mixed models.' *Statistica Sinica*, **14**, pp.349–360.

Schnabel, Z. E. (1938). 'The estimation of the total fish population of a lake.' *The American Mathematical Monthly*, **45(6)**, pp.348–352.

Schwarz, C. J. and Seber, G. A. F. (1999). 'Estimating animal abundance: Review III.' *Statistical Science*, **14(4)**, pp.427-456.

Seber, G. A. F. (1982). *The estimation of animal abundance.* 2nd edition. London: Griffin.

Seber, G. A. F. (1986). 'A review of estimating animal abundance.' *Biometrics* **42**, pp.267–292.

Seber, G. A. F. (1992). 'A review of estimating animal abundance II.' *International Statistical Review* **60**, pp.129–166.

Seber, G. A. F. (2002). *The estimation of animal abundance and related parameters.* London, Edward Arnold.

Smith, P. J. (1988). 'Bayesian methods for multiple capture-recapture surveys.' *Biometrics*, **44**, pp.1177–1189.

Smith, P. J. (1991). 'Bayesian analysis for a multiple capture-recapture model.' *Biometrika*, **78**, pp.399–407.

Stanley, T. R. and Richards, J. D.(2005). 'Software review: a program for testing capture-recapture data for closure.' *Wildlife Society Bulletin*, **33(2)**, pp.782–785.

StataCorp (2010). *Stata statistical software: Release 11*, College Station, TX: StataCorp.

Stiratelli, R., Laird, N. and Ware, J. H. (1984). 'Random-effects models for serial observations with binary response.' *Biometrics*, **40**, pp.961–971.

Stoklosa, J., Hwang, W., Wu, S. and Huggins, R. (2011). 'Heterogeneous capture-recapture models with covariates: A partial likelihood approach for closed populations.' *Biometrics*, **67**, pp.1659–1665.

Sutradhar, B. C. and Das, K. (1999). 'On the efficiency of regression estimators in generalised linear models for longitudinal data'. *Biometrika*, **86(2)**, pp.459–465.

Verbeke, G. and Molenberghs, G. (1997). *Linear mixed models in practice: A SAS-oriented approach.* New York: Springer.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer.

Wang, X., He, C. Z. and Sun, D. (2007). 'Bayesian population estimation for small sample capture-recapture data using noninformative priors.' *Journal of Statistical Planning and Inference*, **137**, pp.1099-1118.

Wedderburn, R. W. M. (1974). 'Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method.' *Biometrika*, **61(3)**, pp.439–477.

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

White, G. C., Anderson, D. R., Burnham, K. P., and Otis, D. L. (1982). *'Capture-recapture and removal methods for sampling closed populations.'* Los Alamos National Laboratory, LA-8787-NERP, Los Alamos, New Mexico, USA.

Williams, D. A. (1975). 'The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity.' *Biometrics*, **31**, pp.949–952.

Williams, B. K., Nichols, J. D. and Conroy, M. J. (2002). *Analysis and management of animal populations.* Academic Press, San Diego, California.

Wolfinger, R. D. and O'Connell, M. (1993). 'Generalized linear mixed models: a pseudo-likelihood approach.' *Journal of Statistical Computation and Simulation*, **48**, pp.233–243.

Wong, W. H. (1986). 'Theory of partial likelihood.' *The Annals of statistics*, **14**, pp.88–123.

Wood, S. N. (2006). *Generalized additive models: An introduction with R.* Boca Raton, Florida: Chapman and Hall/CRC.

Xu, C., Sun, D. and He, C. (2013). 'Objective Bayesian analysis for a capture-recapture model.' *Annals of the Institute of Statistical Mathematics*, DOI 10.1007/s10463-013-0413-1.

Yang, H. C. and Chao, A. (2005). 'Modeling animals behavioral response by Markov chain models for capture-recapture experiments.' *Biometrics*, **61(4)**, 1010-1017.

Yip, P. (1991). 'Conditional inference on a mixture model for the analysis of count data.' *Communications in Statistics - Theory and Methods*, **20(7)**, pp.2045–2057.

Zeger, S. L. and Liang, K.(1986). 'Longitudinal data analysis for discrete and continuous outcomes.' *Biometrika*, **42(1)**, pp.121–130.

Zeger, S. L., Liang, K. Y. and Albert, P. S. (1988). 'Models for longitudinal data: A generalized estimating equation approach.' *Biometrics*, **44(4)**, pp.1049–1060.

Zhang, S. (2012). 'A GEE approach for estimating size of hard-to-reach population by using capture recapture data.' *Statistics*, **46(2)**, pp.175–183.

Zhao, L. P. and Prentice, R. L. (1990). 'Correlated binary regression using a quadratic exponential model.' *Biometrika*, **77**, pp.642–648.

Ziegler, U. G. A., Kastner, C. and Blettner, M. (1996). *The generalized estimation equations in the past ten year: An overview and a biomedical application.*

Zorn, C. J. W. (2001). 'Generalized estimating equation models for correlated data: a review with applications.' *American Journal of Political Science*, **45(2)**, pp.470–490.

Zwane, E. and van der Heijden, P. (2005). ‘Population estimation using the multiple system estimator in the presence of continuous covariates.’ *Statistical Modelling*, **5**, pp.39–52.