Kent Academic Repository

Zhou, Ming (2018) *Statistical Development of Ecological Removal Models.* Doctor of Philosophy (PhD) thesis, University of Kent,.

Statistical Development of Ecological Removal Models

Ming Zhou

School of Mathematics, Statistics and Actuarial Science
University of Kent

This dissertation is submitted for the degree of
Doctor of Philosophy in Statistics

November 2018

Acknowledgements

I would like to express my gratitude to my supervisors, Dr. Rachel McCrea, Dr. Eleni Matechou and Dr. Diana Cole, for their support and patience during this PhD. They have encouraged me to think and work independently with their helpful advice. I thank them for their support for the written elements of the thesis and their advice to help me present my work to a wider audience. I am also grateful to the School of Mathematics, Statistics and Actuarial Science at the University of Kent for funding my PhD research. I also would like to thank all the administrative staff as well as teaching staff in the School of Mathematics, Statistics and Actuarial Science (SMSAS), for everything! Thank you for having me and it has been great experience! Thanks to my friends for making my life great during this time. I finally would like to thank my family, for their support, encouragement and understanding.

Abstract

Removal sampling is commonly used to estimate abundance of populations in which captured individuals are permanently removed from a study area. The classic removal model (Moran, 1951) assumes a constant capture probability and all animals are available for detection throughout the study, which results in a simple geometric decline of counts of removed individuals over time. However, the real data collected from some species exhibit unexpected fluctuations in the number of captured animals. The work in this thesis is driven by real data on common lizards, *Zootoca vivipara* and great crested newts, *Triturus cristatus*, where existing approaches may give rise to misleading conclusions.

When modelling removal data it is crucial to account for imperfect availability in the population, as individuals could sometimes temporarily become undetectable at study area, or emerge from an area outside the study. This thesis deals with three aspects of removal modelling: (i) We develop a robust design multievent removal modelling (RMER framework) which allows considerable flexibility in estimating temporary emigration as well as capture probability and the size of populations. We also consider the effect of sparse data and investigate the use of modelling different sources of data in conjunction with the removal data (Besbeas et al, 2002). (ii) The estimation of temporary emigration or population renewal for removal data relies on the use of the robust design (Zhou et al. 2018). However, there are many removal data which lack the robust design structure. Motivated by the analysis of a data set of common lizards collected under standard sampling protocol, we develop and evaluate the use of penalised maximum likelihood estimation to allow populations to be open to new individuals via birth/arrival for data sets without the robust design. (iii) We use four criteria to explore study design aspects of removal data with the robust design, including the trade-off in survey effort allocation between primary periods and secondary periods for a fixed level of total sampling effort. The models we propose can account for temporary emigration or new arrivals of individuals during removal sampling and represent a step forward with respect to current modelling approaches and will guide wildlife management.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Monitoring the abundance of protected species is a crucial task for conservation management. Various statistical models have been widely used to analyse datasets collected by ecologists with an aim of estimation of demographic parameters for animal populations. One of the fundamental parameters of interest in statistical ecology is the abundance or population size, which is the total number of animals in the study site. Removal sampling can be used to estimate the abundance of a population, which involves sequential removals of individuals from a defined study area. In Chapter 1 we provide a brief overview of removal sampling in Section 1.1 and discuss its relationship with another common sampling technique called capture-recapture sampling in Section 1.2. The structure of the thesis is given in Section 1.3. General methodology used in the thesis is described in Section 1.4

1.1 Background on removal sampling

Removal sampling is a sampling technique in which captured animals are removed from a defined study area. The sampling area is generally visited daily over a short period of time. During the study, ecologists permanently remove the captured individuals from the study site and relocate them to another new habitat. The removal data that arise from removal sampling is a vector of counts of individuals removed at each sampling occasion. The classic removal model was proposed by Moran (1951) and Zippin (1956, 1958), and the objective of removal modelling is to estimate the number of animals that remain uncaptured in the study area.

Removal sampling is commonly used for fish populations to measure abundance in stock assessment over time, see e.g. Otis et al. (1978); Carle and Strub (1978); Wang (1999) and Peterson et al. (2004). In recent years, removal sampling has been used as a conservation management tool called mitigation translocation to prevent the extinction of species. Species on a study site are often threatened by land development activities. Therefore human-wildlife conflict can be addressed by human-mediated translocations of endangered populations from the study area to a new habitat. Over the past few decades, mitigation translocation has become a popular and widespread animal management tool due to an exponential increase in the number of land development projects worldwide (Seddon and Maloney, 2007; IUCN/SSC, 2013; Germano et al., 2015). In addition, the use of such mitigation translocations is predicted to rise more rapidly if it is adopted for climate-change mitigation (IUCN/SSC, 2013). The motivations may differ from historical usage of removal sampling, but the sampling techniques used are similar, where translocations of protected species involve capture, relocation and release of species prior to the proposed building starting (IUCN/SSC, 2013).

A wide range of species have benefited greatly through mitigation translocations. Thousands of threatened land snails were relocated on the west coast of New Zealand since 2000 because of coal mining on public lands (Morris, 2010). In addition, amphibians and reptiles are frequently found at development sites in the UK, resulting in hundreds of such translocation projects annually (Germano et al., 2015). Such translocations were also carried out for burrowing owls in North America (Sarno et al., 2012), black rhinoceros in South Africa (Linklater et al., 2011) and kangaroo rats in California (Tennant and Germano, 2017) before the land clearing begun.

Although millions of pounds have been spent on removing protected animals out of the pathway of land development annually in the UK, such translocations may not meet the objective of preserving the target population as intended by legislation, and numerous reviews mention poor rates of success from these projects (Griffiths et al., 1989; Linnell et al., 1997; Germano et al., 2015). Failure of such projects can be a result of insufficient survey effort, resulting in too few animals being captured to establish a viable population elsewhere. Equally, many animals may go undetected at the removal site, resulting in the loss of the majority of the population when the site is developed. This raises questions concerning the amount of survey effort required to remove a significant proportion of the population. However, few publications highlight the need for improved statistical removal models to evaluate whether the number of translocated individuals represent a substantial proportion of total population size in the study site (Griffiths et al., 2015).

The classic removal model relies on a constant probability of capture and it assumes that there is no birth, emigration, immigration and death in the population across the study (Moran, 1951; Zippin, 1956, 1958). The estimation of population size can be especially problematic for species that are hard to capture and the current abundance modelling approaches for removal data may give rise to misleading conclusions, as the assumptions of the existing models can be violated in various ways. We discuss recent statistical developments of removal models in Section 2.5.

1.2 Relationship with capture-recapture sampling

A more commonly used method for monitoring wildlife populations is capture-recapture experiments. The purpose of capture recapture studies is often similar to removal sampling, i.e. to estimate the total number of individuals in the population. In capture-recapture studies, we capture and mark each individual upon the first capture with a unique identification tag, then release the marked individual back to the population in order to be able to identify it if it is caught on the subsequent sampling occasions.

Capture-recapture samplings represent perhaps the most popular technique for studying animal populations. A huge variety of statistical methods have been developed for capture-recapture data to estimate demographic parameters for wild populations (Cormack, 1964; Jolly, 1965; Otis et al., 1978; Lebreton et al., 1992; Schwarz and Arnason, 1996). Both Williams et al. (2002) and McCrea and Morgan (2014) give a comprehensive overview of the existing capture-recapture models, also see King (2014) for a review of recent advances.

The data obtained from capture-recapture sampling for each individual is a vector describing its capture history during the study. It is common to use 1s and 0s to indicate whether the individual is captured or not respectively in the capture history. For example, “011”’ indicates that the individual is caught at the second and third occasions but not at the first sampling occasion. If the study is carried out in the removal sampling protocol, then an equivalent removal record would be “010”’ as the captured individual will be permanently removed from the study area. In removal studies, the collected data are the number of unmarked individuals captured at each sampling occasion. Once the animals are captured, they are removed permanently from the population, therefore there is no recapture of individuals. The most common method of removal is physical removal; for instance, individuals are transplanted to another isolated pond in a mitigation translocation study or the animals are held in another area until the study is completed.

Traditionally removal sampling was conducted for assessing the exploitation of fish populations in fisheries industry Cowx (1983); Wyatt (2002); Rosenberger and

Dunham (2005). Removal models have been used to estimate the size of fish stocks and fishery scientists use the estimated population size to manage and restrict commercial fishing activities so that resources are not overfished. However, capture-recapture sampling is rarely used in the commercial fisheries industry as such fishing activities are profit-driven and fishermen are not interested in marking and discarding captured individuals.

Removal sampling has also been used for the removals of invasive species (Mills et al., 2004; Sullivan and Sullivan, 2013; Guillera-Arroita et al., 2014). Invasive species are one of the key causes of biodiversity loss and can cost more than \$1.4 trillion annually (Burgiel and Muir, 2010). Invasive species can pose severe threats to native species and their habitats when there is a lack of natural predators. Capture-recapture sampling is illogical for such problems as releasing the captured invasive individuals back into the ecological system is not a practical management strategy for invasive species. Therefore removal sampling has been used to reduce the abundance of invasive species, as physical removal is the only viable control for invasive populations.

In addition, the length of capture-recapture and removal studies is often different. Capture-recapture sampling is usually conducted every week or month and could last as long as a few decades for monitoring purposes, as long-term capture-recapture studies are often required for observing changes in the demographic parameters (Williams et al., 2002). By contrast, removal sampling is typically carried out for up to a couple of months. Therefore data collection for capture-recapture can be time-consuming and expensive compared with removal sampling.

1.3 Structure of the thesis

In this section, we briefly introduce the content of each chapter. In Chapter 2 we first show how to estimate the population size by two removal samples based on the assumptions made in the classic removal model. We then present the classic removal model with more than two removal samples and discuss its assumptions in detail. We also give the likelihood with time-varying capture probability in terms of covariates. The models discussed in Chapter 1 are fitted to a great crested newt data set for both males and females separately. Recent developments for the classic removal model are discussed at the end of Chapter 2.

In Chapter 3, we propose new removal models accounting for temporary emigration. The development is motivated by amphibian and reptile populations as they often temporarily emigrate to an area out of sight of ecologists. In capture-recapture studies, the sampling scheme and corresponding modelling approach that allows the estimation of temporary emigration of a population is called the robust design (Pollock, 1982). We develop novel removal models that bring together both the robust design (Pollock, 1982; Kendall et al., 1995) and a multievent structure (Pradel, 2005) where the underlying movement pattern of individuals between the study area and an area outside of the study is modelled as a partially hidden process. We use the machinery of hidden Markov models for the computation of the proposed likelihood function. We explore the benefits of the proposed approach compared with models without robust design. We also investigate estimates calculated from the classic removal model for populations exhibiting temporary emigration by simulations. We apply the new removal models to juvenile and adult common lizard data using integrated removal models with robust design and multievent framework.

In Chapter 4, we investigate which parameters in the models proposed in Chapter 3 can be estimated, i.e. whether or not a model is parameter redundant. Symbolic algebra is used to investigate parameter redundancy and to find the estimable combinations of parameters if the proposed model is parameter redundant. We demonstrate that the robust design can enable us to estimate otherwise confounded parameters individually and produce tables summarising for which models it is possible to estimate all the parameters. The methods presented in Chapter 4 provide almost identical findings to Chapter 3, without intensive simulations.

Although we show the benefits of the use of robust design in Chapters 3 and 4, a lot of removal data are obtained from a standard sampling protocol in which there is only one survey within each sampling day. In Chapter 5, we propose a penalised maximum likelihood approach to improve the estimation of open removal models where individuals could enter the study area during the removal study. We investigate three types of penalty term which are used in linear regression for our proposed models. We then illustrate the utility of the models by fitting them to common lizard data collected from a translocation project in Sandwich, UK.

The estimation of temporary emigration from removal data relies on the use of the robust design as demonstrated in Chapters 3 and 4, where there are at least two secondary occasions between primary occasions. Furthermore, the use of penalised maximum likelihood estimation discussed in Chapter 5 for removal data collected under the standard sampling protocol can be time-consuming due to the need to employ cross-validation. Furthermore, its performance relies on fairly large sample sizes. As the robust design is highly preferred for removal sampling, we look into the design of removal sampling at a single site while accounting for the availability of individuals under the robust design in Chapter 6. The aim is to consider how to allocate the total sampling effort for the removal sampling under the robust design sampling protocol. We use both simulations and analytic forms to investigate different sampling approaches where constant parameters are assumed.

1.4 Preliminary remarks on model implementation

Maximum likelihood estimation is used to estimate parameters throughout the thesis, unless otherwise stated. The optimisations are carried out in `R` and `Matlab`. For models in Chapters 2 and 3, we write the likelihood functions in `R`. In addition, we use a symbolic package `LinearAlgebra` in `Maple` to investigate the identifiability of proposed models in Chapter 4. As results in Chapter 4 are too long to display, we show `Maple` code used in this chapter in the electronic appendix. Detail on names of the files is given in the corresponding Chapter. All the likelihood functions in Chapters 5 were written in `Matlab` and optimised using its high performance computing facilities. In chapter 6, we use the `Symbolic Math Toolbox` in `Matlab` for calculating the analytical results and use `R` for a simulation study. We initialise the optimisation algorithm in each case using different starting values for both simulated and real data, in order to ensure the global maximum is obtained.

The parameters are transformed during likelihood optimisations in `R` and `Matlab`. For parameters that are nonnegative (i.e. constrained to $[0,\infty)$) such as population size, the log transformation $\theta = \log(N)$ is used in order to carry out unconstrained optimisation with respect to θ. Once the optmisaton is done, we use exponential function $N = \exp(\theta)$ to obtain N. Similarity, the logit transformation $p = 1/(1+e^{-\theta})$ is applied to those probability parameters that are constrained to $[0,1]$, then optimising with respect to θ which can take any value in $(-\infty,\infty)$. The maximum likelihood estimates are obtained by transforming those unconstrained estimates back to the original scales using the inverse functions of log or logit.

The standard errors and confidence intervals in Chapters 2, 3 and 5 are obtained by bootstrapping individual removal records because some parameters reached the boundaries of the parameter spaces when analysing real data. Non-parametric bootstrap samples are obtained by sampling with replacement from the original data. We describe the procedure below.

- Step 1: rewrite the removal data in terms of individual removal records for all sampling occasions. If we removed n_k individuals at the kth sampling occasion, we rewrite this with n_k "k"s individual records. For example, if the first three observations of a removal data set are $2,3,2$. We rewrite them as two "1", three

"2"s and two "3"s, i.e. $1,1,2,2,2,3,3$ and repeat the same process for all of the observations.

- Step 2: sample D individual removal records with replacement and rewrite them in terms of removal count data by calculating the resulting frequency table, where D is equal to the total number of individuals we observed in the original data.

- Step 3: repeat Step 2 a large number of times (e.g. 500 resamples) to produce bootstrap samples.

For example, Figure 1.1 shows 500 bootstrap samples for a juvenile common lizard data set that is analysed in Chapter 3. The model is fitted to all the samples and the 500 parameter estimates that are computed can be used to construct confidence intervals and standard errors. For simulations conducted in Chapter 6, we calculate the standard errors of the transformed parameters by obtaining the square roots of the diagonal elements of the inverse of the Hessian matrix. Standard errors of the parameters on the original scales are derived from those calculated on the transformed scale using the delta method (Morgan, 2008).

Model selection is conducted using the Akaike Information Crierion (AIC). The AIC of a model is defined as

$$\mathrm{AIC} = -2l + 2h$$

where l is maximised log-likelihood and h is the number of estimated parameters in the model (Akaike, 1973; Morgan, 2008). AIC is useful for both nested and non-nested models. The second term $2h$ can be interpreted as a penalty term for increasing the number of parameters in the model, so AIC penalises models with large number of parameters.

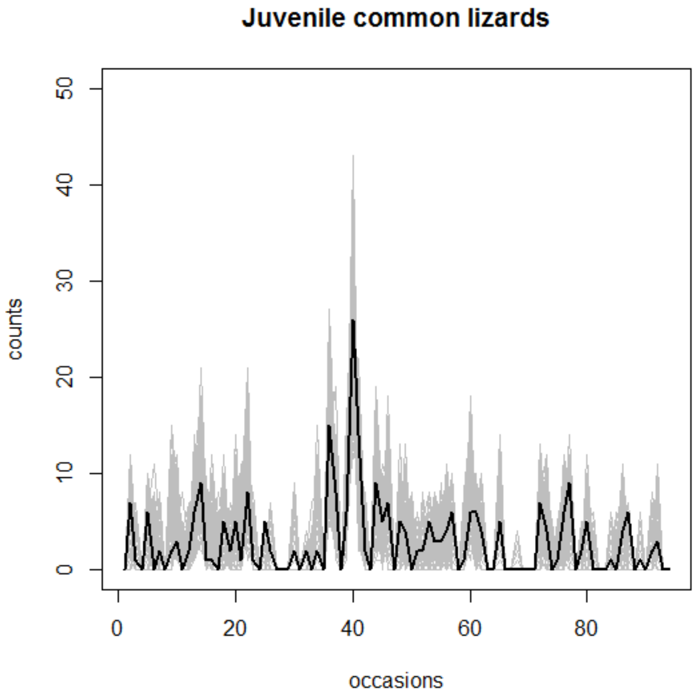


Fig. 1.1 500 bootstrap samples (gray lines) for juvenile common lizard data (black line).

Chapter 2

Classic Removal Model

2.1 Introduction

In Chapter 2 we review the classic removal model. In Section 2.1.1 and 2.1.2 we present early research for just one and two removal sampling occasions respectively. The formulation of the classic removal model with a constant capture probability will be presented in Section 2.2.1. Then we discuss possible violations of assumptions in the classic removal model in Section 2.2.2 that motivate the work presented in the thesis. In Section 2.3 we present a removal model with time-varying capture probability in terms of covariates. Two examples of data analysis of great crested newts is demonstrated in Section 2.4. We conclude Chapter 2 with a review of recent development for removal models in Section 2.5.

2.1.1 One removal sample

Removal sampling involves successive samples in which animals are captured and removed from the study area. The key question we are interested in is how many individuals are there in the population. Suppose the population size is denoted by N, then the estimate of population size is a function of the number of individuals removed at the first sampling occasion and the probability of an individual being captured. Suppose that we observe n_1 animals at the first sampling occasion, then the estimate of the population size can be expressed as

$$\hat{N} = \frac{n_1}{p} \tag{2.1}$$

for $0 < p \leq 1$, where the denominator p is the probability of an individual being captured and removed, which is often called the capture probability in the capture-recapture literature. The capture probability in this example is unknown as there is no information in the removal data to estimate p for removal sampling with just one sampling occasion. Therefore, the estimate of the population size cannot be calculated as Equation (2.1) contains the unknown capture probability p In Section 2.1.2 we explain how to estimate both N and p using just two removal samples.

2.1.2 Two removal samples

Seber and Whale (1970) derived the maximum likelihood estimator analytically for a removal model with two removal samples based on the assumption that the proportions of individuals being removed are equal for two sampling occasions in a closed population. This particular case of two removal samples was also briefly considered by Nees et al. (1957), Zippin (1958) and Seber and Le Cren (1967).

A closed population model, as the name implies, assumes that no birth/death and no immigration/emigration occur between sampling occasions. White et al. (1982) give a good overview of the closure assumption. The assumption of closure is usually valid for data collected in a relatively short period of time during a nonbreeding season for a population.

The idea is similar to the Lincoln-Petersen estimator for a closed population using capture-recapture data (Petersen, 1894; Lincoln, 1930). Note that the proportion of animals being removed at the first removal sample is $p = n_1/N$ which is obtained by rearranging Equation (2.1) for the capture probability. Before taking the second removal, there will be $N - n_1$ individuals available for capture in the study area. Let n_2 denote the number of individuals being removed at the second removal sample. The proportion of individuals being removed is $p = n_2/(N - n_1)$. As we assume that the proportion of removed individuals in the first sample is equal to the proportion of removed individuals in the second sample, the following is formulated

$$\frac{n_1}{N} = \frac{n_2}{N - n_1}. \tag{2.2}$$

The estimate of population size can be calculated by rearranging Equation (2.2) as below

$$\hat{N} = \frac{n_1^2}{n_1 - n_2} \tag{2.3}$$

where $n_1 > n_2$.

Therefore, the estimated capture probability is calculated as shown in Equation (2.4) by equating Equation (2.3) and Equation (2.1),

$$\hat{p} = 1 - \frac{n_2}{n_1} \tag{2.4}$$

where $n_1 > n_2$.

Note that the above maximum likelihood estimators only hold when $n_1 > n_2$ because we expect the number of individuals being removed to decrease over time under the assumptions mentioned before. If $n_1 < n_2$, then the estimates of the population size and the capture probability as shown in (2.3) and (2.4) respectively would be negative. If $n_1 = n_2$, then $\hat{p} = 1 - n_2/n_1 = 1 - 1 = 0$ and $\hat{N} = n_1^2/(n_1 - n_2) = n_1^2/0 = \infty$.

2.2 Classic removal model

In Section 2.1 we have introduced the classic removal model for one and two removal samples. In Section 2.2 we review the classic removal model for multiple removal samples from a closed population. The classic removal model proposed by Moran (1951) and Zippin (1956, 1958) is closely related to the M_b model for capture-recapture studies as discussed in Otis et al. (1978, pages 44-50). The M_b model assumes that capture probabilities vary by behavioral responses to first capture. It was designed to deal with situations where individuals become trap happy or trap shy in the subsequent sampling occasions after the first capture. Carothers (1973) termed this behavioral response as a “contagion of catchability” in the population. It implies that the behavior of animals tends to change after the initial capture. For example, if an individual was alarmed or was hurt in its initial capture, it is more likely that it will not enter a trap again, therefore it is expected that the recapture probability becomes lower as it is harder to capture that individual. The classic removal model is a special case of the M_b model in which the recapture probability becomes zero, as there is no recapture of animals in removal studies.

2.2.1 Formulation

Assumptions and Notation

Assume that animals are removed from the population of interest on K sampling occasions. The aim is to estimate the number of individuals we failed to capture within the study area. The following assumptions are made by the model, as described in Moran (1951):

- The population is closed to birth, emigration, immigration and death across the study.

- The probability of capture is constant across all individuals over sampling occasions.

As described before, the interpretation of the assumption of closure is that there is no birth, no death and no migration. In other words, all individuals are present and available for capture and removal throughout the study.

In order to construct the likelihood function, we introduce the following notation:

- n_k is the number of individuals being removed from the area at the kth sampling occasion, where $(k = 1, 2, \ldots, K)$.
- x_k is the cumulative number of removals conducted prior to the kth sampling occasion, i.e. $x_k = \sum_{j=1}^{k-1} n_j$, where $k = 2, 3, \ldots, K+1$. Because no individuals are removed before the first sampling occasion, x_1 is equal to zero.
- N denotes the population size.
- n_0 is the number of animals which fail to be captured and removed from the population, i.e. $n_0 = N - x_{K+1}$ by the end of the study.
- p is the probability of an individual being removed on a sampling occasion, assumed to be the same between individuals and constant for the duration of the study.

Likelihood Function

The probability of an individual being removed at the kth sampling occasions is

$$L_k(p) = (1-p)^{k-1} p,$$

which is the probability mass function of a geometric distribution with the probability of success being p. Therefore, the expected number of animals being removed in the classic removal model decreases geometrically over sampling time.

If we assume that the number of individuals being removed, n_k, given the total number of previous removals, x_{k-1} , is sampled from a binomial distribution with the capture probability p and $N - x_{k-1}$ as the number of trials, the vector of observed removal data $(n_1, n_2, \cdots, n_k)$ can be described by a multinomial distribution. The full multinomial likelihood function can be formed as a function of n_0 and p,

$$L(n_0, p) = \frac{N!}{n_0! \prod_{k=1}^{K} n_k!} \left\{ \prod_{k=1}^{K} \left[(1-p)^{k-1} p \right]^{n_k} \right\} (1-p)^{n_0 K},$$

where $N = n_0 + \sum_{k=1}^{K} n_k$, for $k = 1, 2, \ldots, K$ as discussed in Zippin (1958).

The above likelihood function simplifies to give,

$$L(n_0, p) = \frac{N!}{n_0! \prod_{k=1}^{K} n_k!} p^{x_{K+1}} (1-p)^{KN - \sum_{k=1}^{K+1} x_k}. \tag{2.5}$$

2.2.2 Violation of model assumptions

In ideal situations, the classic model results in a geometric decline of the expected number of individuals within the closed area as removals take place. Figure 2.1 shows two simulated removal data sets resulting from removal sampling with different capture probability ($p = 0.3$ or 0.6) for a population of 500 individuals over 20 sampling occasions. For both simulated data sets we observe all of the individuals in the population, while when the capture probability is 0.6 there are more zero observations at the end of the study as individuals are removed more quickly than those in the removal study with $p = 0.3$. The decreasing pattern of removal data is used to infer both the total population size and the capture probability. However, the assumptions of classical removal models can be easily violated in numerous ways.

Firstly, it is known that the assumption of constant capture probability is often violated. The classic removal model is often not appropriate for cryptic species, such as reptiles and amphibians, because trends in catchability often change over time as animals are fairly sensitive to changes in environmental conditions. All animals may not have the same probability of capture throughout the whole study. Different cohorts of individuals with respect to size, age or sex may have different susceptibilities to capture, causing heterogeneity in detectability among individuals. In such cases, the more catchable individuals are more likely to be removed at the early stages of the study, leaving the less catchable animals to comprise the remaining population.

Secondly, the estimates can be biased due to violation of the closure assumption of the population. For example, a population can be susceptible to losses of population between occasions, as well as gaining new individuals arriving in the study area. Populations experiencing fluctuations in terms of gains or losses between sampling occasions would lead to biased estimates if the classic model is fitted. For example, for reptiles and amphibians some of them are likely to become less active and remain undetected underground for a period of time. The assumption that an individual present within the study area is available for detection throughout the study will be violated if

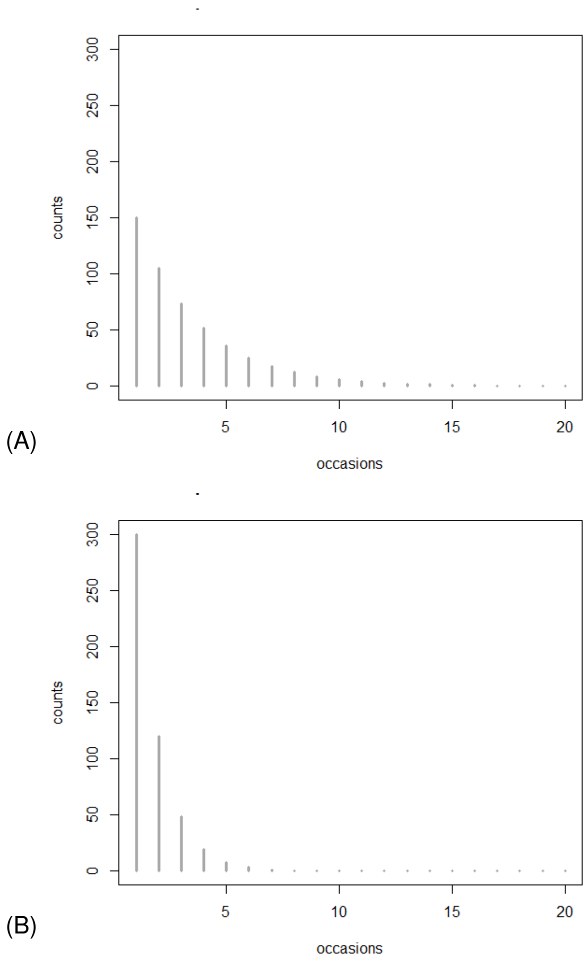


Fig. 2.1 Simulated counts of removed individuals obtained from the classic removal model with $K = 20$ sampling occasions where capture probability is 0.3 (A) or 0.6 (B).

the individual temporarily emigrates outside the study area and becomes undetectable for a period of time.

Furthermore, translocation projects for reptile and amphibian populations are often conducted from summer until autumn (Germano and Bishop, 2009) as mating takes place in the spring after hibernation in the winter and they often give birth from late summer (Edgar et al., 2010; Rafferty, 2011). As a result we could expect the new births to become available for capture during the study. Therefore a closed population is not a suitable assumption under these circumstances.

Simulations in Chapters 3 and 5 show how estimates are biased if the classic removal model is wrongly used for populations that are not closed.

Overall, the assumptions in the classic removal model can be violated in various ways. Therefore, more complicated models are needed to take account of those ecological behaviors and realistic.

2.3 Covariates in detection probability

In practice, it is expected that capture probability varies in terms of external effects. Especially for reptiles and amphibians, capture probability is likely to vary with weather covariates as discussed in Section 2.2.2. In this section, we model time-dependent capture probability in terms of covariates. Let p_k be the capture probability at time k $(k = 1, \ldots, K)$, then the time-varying capture probability p_k can be modelled using a logistic regression with additive effects for covariates,

$$\text{logit}(p_k) = \log\left(\frac{p_k}{1-p_k}\right) = \alpha + \beta_1 z_1 + \cdots + \beta_u z_u \tag{2.6}$$

where $u \leq K-2$, and α and β_q for $q = 1, \cdots, u$ are the coefficients to be estimated. Note that the number of covariates considered should not exceed $K-2$, otherwise there will be more parameters than the number of data points in the model. It is usually necessary to conduct an assessment of all available combinations of covariates in order to choose the best subset of covariates for a given dataset.

The likelihood function can be expressed as,

$$L(n_0, \alpha, \beta) = \frac{N!}{n_0! \prod_{k=1}^{K} n_k!} \prod_{k=1}^{K} p_k^{n_k} \prod_{k=1}^{K-1} (1-p_k)^{D-x_{k+1}} \prod_{k=1}^{K} (1-p_k)^{n_0} \tag{2.7}$$

where D is the total number of individuals being observed, $D = \sum_{k=1}^{K} n_k$.

The log-likelihood can be written as,

$$\begin{aligned} l(n_0, \alpha, \beta) = \log[\Gamma(N+1)] - \textstyle\sum_{k=1}^{K} \log[\Gamma(n_k+1)] - \log[\Gamma(n_0+1)] + \textstyle\sum_{k=1}^{K} n_k \log(p_k) + \\ \textstyle\sum_{k=1}^{K-1} (D - x_{k+1}) \log(1-p_k) + n_0 \textstyle\sum_{k=1}^{K} \log(1-p_k) \end{aligned} \tag{2.8}$$

where $\Gamma(Y+1) = Y!$.

2.4 Great crested newts examples

In Section 2.3 we have extended the classic removal model to consider time-dependent capture probability in terms of covariates. In Sections 2.4.1 and 2.4.2, we demonstrate how the removal model with time-varying capture probability p_k can be used in practice for a real data set. We also compare the results obtained from the classic removal model with a constant capture probability.

We analyse a removal data set of great crested newts. Great crested newts were collected on over 93 sampling occasions from March 2010 to June 2010 with 742 and 699 captured male and female individuals respectively. Four covariates are recorded for each sampling visit, which are: minimum air temperature (MinAT), maximum air temperature (MaxAT), minimum ground temperature (MinGT) and maximum ground temperature (MaxGT).

2.4.1 Modelling climatic covariates

In this section, removal models with time variations in capture probability p_k in the form of Equation (2.7) are fitted to the female and male populations separately, where p_k is modelled using a logistic regression as shown in Equation 2.6.

Results of all the fitted models ranked by AIC are shown in Table 2.1, where “cov” denotes an additive effect in terms of covariates and “c” denotes a constant parameter. We observe that all of the top eight performing models ranked by with AIC value for each gender incorporate MinAT as the covariate in the capture probability using a logistic regression of the form (2.6). Additionally, for each gender, 14 out of 15 models that considered covariates for the capture probability fit the data better than the geometric model with constant capture probability. It seems that the covariate MinGT alone cannot explain the variations in the capture probability well and the model with MinGT only was ranked last for both males and females as shown in Table 2.1.

In Figure 2.2, the model with lowest AIC, shows a better fit compared to the classic removal model. This is true for both males and females. Furthermore, the estimated parameters obtained from the top model with the lowest AIC and the classic removal model are shown in Table 2.2, where standard errors are computed using 500 bootstrap samples. We note that the estimates of population size are equal to the observed number of individuals for both female and male newts, which means that the model predicts that no individuals were left behind by the end of the study.

In Figure 2.3 we plot 95% confidence intervals calculated empirically from the 500 bootstrap samples for the top model for each gender. We observe that the variability in the confidence intervals is hardly noticeable in Figure 2.3. The narrowness of the confidence intervals is because the size of observed individuals (i.e. 742 and 699 for male and female respectively) is relatively large compared with the number of parameters in the model, and hence the estimated parameters listed in Table 2.2 have quite small standard errors. Therefore the confidence intervals are narrow for both male and female populations.

Table 2.1 Results of fitting the geometric removal model with time variations in p_t and constant p. Minimum and maximum air temperature are denoted by MinAT and MaxAT respectively. Minimum and maximum ground temperature are denoted by MinGT and MaxGT respectively. † in the cell indicates that the covariate of that column is used in the model for p. h is the number of parameters and ML is the value of the maximised loglikelihood.

Gender	Model	MinAT	MaxAT	MinGT	MaxGT	h	ML	AIC
Male	p(cov)	†		†	†	5	-218.64	228.64
	p(cov)	†	†	†	†	6	-218.65	230.65
	p(cov)	†		†		4	-228.90	236.90
	p(cov)	†	†	†		5	-228.50	238.50
	p(cov)	†			†	4	-232.78	240.78
	p(cov)	†	†		†	5	-232.72	242.72
	p(cov)	†				3	-240.14	246.14
	p(cov)	†	†			4	-240.06	248.06
	p(cov)		†	†	†	5	-324.03	334.03
	p(cov)		†		†	4	-328.05	336.05
	p(cov)				†	3	-336.47	342.47
	p(cov)			†	†	4	-334.87	342.87
	p(cov)		†	†		4	-340.15	348.15
	p(cov)		†			3	-342.83	348.83
	p(c)					2	-357.35	361.35
	p(cov)			†		3	-357.03	363.03
Female	p(cov)	†	†	†	†	6	-206.86	218.86
	p(cov)	†		†	†	5	-209.96	219.96
	p(cov)	†		†		4	-219.23	227.23
	p(cov)	†	†	†		5	-217.83	227.83
	p(cov)	†	†		†	5	-218.14	228.14
	p(cov)	†			†	4	-221.23	229.34
	p(cov)	†				3	-227.55	233.55
	p(cov)	†	†			4	-235.81	233.81
	p(cov)				†	3	-322.39	328.39
	p(cov)		†		†	4	-320.87	328.87
	p(cov)		†	†	†	5	-319.68	329.68
	p(cov)			†	†	4	-321.69	329.69
	p(cov)		†			3	-336.18	342.18
	p(cov)		†	†		4	-335.91	343.91
	p(c)					2	-340.20	344.20
	p(cov)			†		3	-340.10	346.10

Table 2.2 Estimated parameters from the model with lowest AIC and the classic removal model p for each gender in Table 2.1. SEs are calculated by bootstrapping.

Model	Estimate (SE)	Male	Female
p(cov)	$\hat{\alpha}$ (SE)	-2.29 (0.07)	-2.28 (0.90)
	$\hat{\beta}_{\text{MinAT}}$ (SE)	0.17 (0.01)	0.18 (0.01)
	$\hat{\beta}_{\text{MaxAT}}$ (SE)	-	0.02 (0.01)
	$\hat{\beta}_{\text{MinGT}}$ (SE)	-0.14 (0.03)	-0.12 (0.03)
	$\hat{\beta}_{\text{MaxGT}}$ (SE)	0.14 (0.02)	0.14 (0.03)
	$\hat{N}$ (SE)	742.00 (0.02)	699.00 (0.01)
p(c)	$\hat{p}$ (SE)	0.13 (0.06)	0.12 (0.05)
	$\hat{N}$ (SE)	742.00 (1.79×10^{-8})	699.00 (1.12×10^{-8})

In this section we have demonstrated how to consider time-varying p_t in terms of covariates for a single population (i.e. male or female) in removal models. We have shown that the results obtained from the classic removal model with a constant p show a poor fit to each data set, where the classic model was ranked the second-last for each gender. This application has demonstrated the importance of accounting for time-variation within the model set.

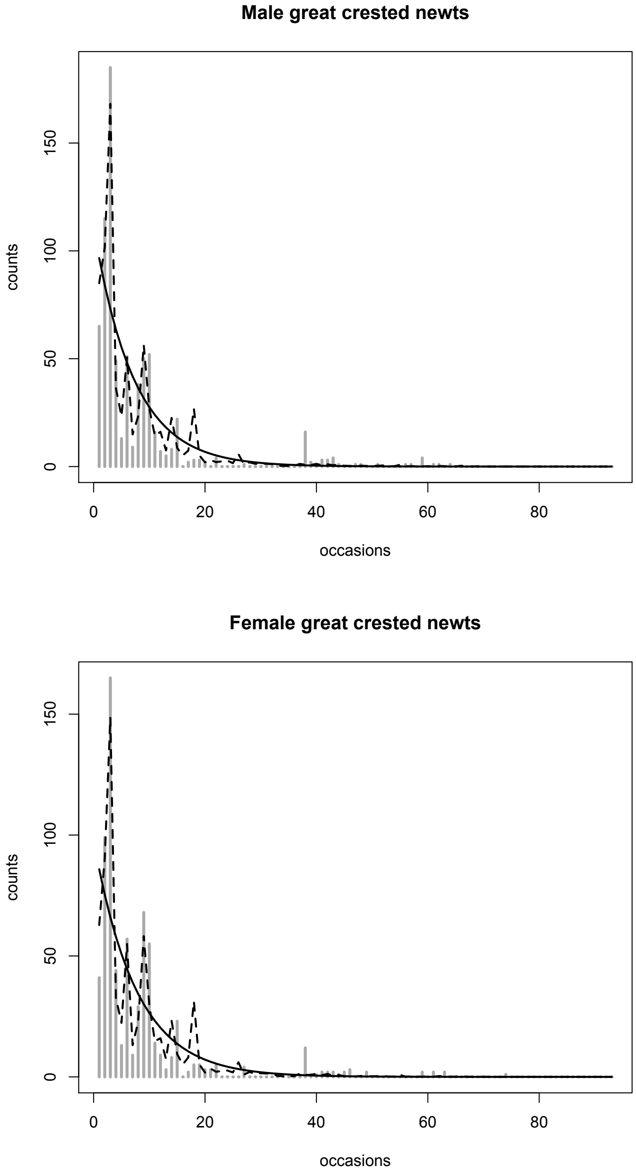


Fig. 2.2 Observed counts are plotted by vertical bars, together with fitted values obtained by the removal model with time variations in capture probability (dashed lines) and classical geometric model (solid lines) for males (top) and females (bottom).

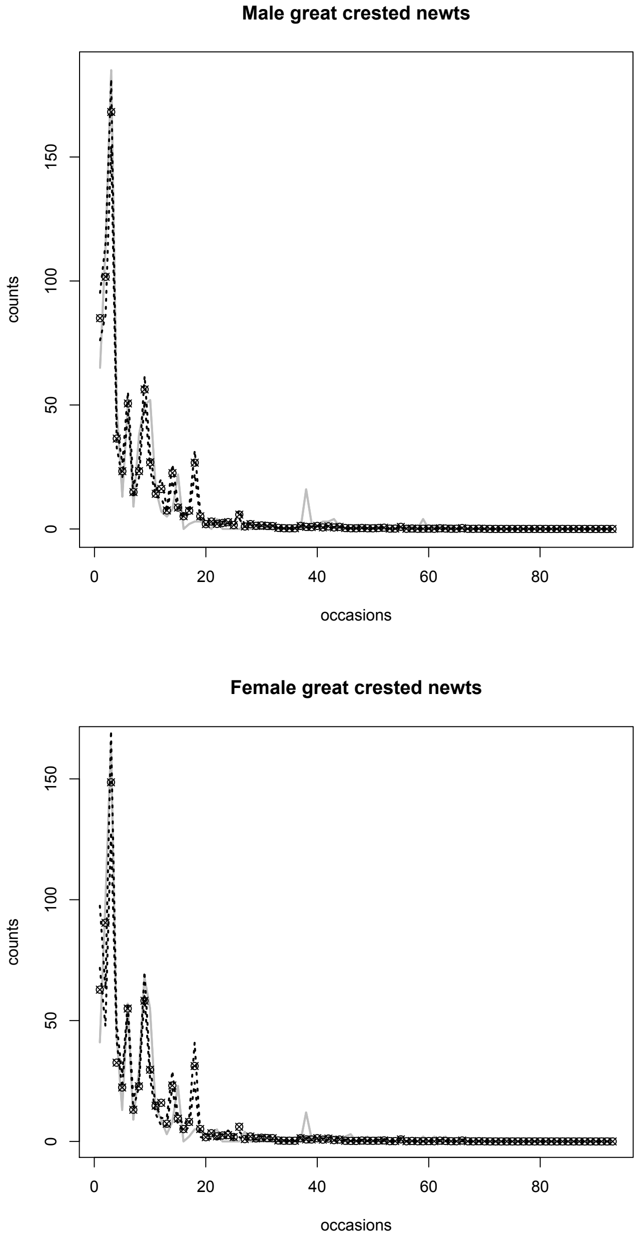


Fig. 2.3 95% bootstrap confidence intervals (dashed lines) of the fitted values (cross circles) obtained from the model with the lowest AIC value in Table 2.1 for males (top) and females (bottom). The original data are plotted using gray solid lines.

2.4.2 Modelling climatic covariates and gender simultaneously

In this section, we consider a removal model that considers time variations in capture probability p_k using both climatic covariates (i.e. MinAT, MaxAT, MinGT and MaxGT) and gender in the logistic regression as shown in Equation 2.6. We code male and female individuals as “0” and “1” respectively. The capture probability p_k in this model is

$$\text{logit}(p_k) = \log\left(\frac{p_k}{1-p_k}\right) = \alpha + \beta_0 g + \beta_1 z_1 + \cdots + \beta_u z_u \tag{2.9}$$

where $u \leq K-2$, and α and β_q for $q = 1, \cdots, u$ are the coefficients to be estimated. The gender covariate is $g = 0$ (male) or 1 (female).

We adapt the likelihood function (2.7) to accommodate different gender groups using an integrated population modelling approach (Besbeas et al., 2002). Suppose we conduct a removal experiment on W categories for a single species with the individual likelihoods defined as $L_1, L_2, \ldots, L_W$ which are in the form of Equation (2.7). If we assume there are W groups of individuals that are captured independently, the full likelihood function L can be written as the product of individual likelihoods, i.e $L = L_1 \times L_2 \times \cdots \times L_W$. Therefore for the newts data set analysed in this section, if we assume male and female individuals are removed independently the full likelihood is $L = L_{\text{male}} \times L_{\text{female}}$.

Results of the fitted models ranked by AIC scores are displayed in Table 2.3. The top five models include MinAT as a covariate for capture probability. Similar results are obtained in Table 2.1, where male and female populations are modelled individually. Additionally, the estimates and standard errors obtained from the top model with time-varying capture probabilities in terms of MinAT and MinGT are shown in Table 2.4. Results are similar with those shown in Table 2.1, where both tables suggest we have captured all individuals in the populations by the end of the study.

To investigate possible differences between male and female populations, we explore contingency table in Table 2.5 and use chi-square test for association to compare males and females. Because both the observed and expected frequencies are small after the 16th sampling occasion, we present those frequencies as one group (i.e. ≥ 16) in Table 2.5. The null hypothesis H_0 assumes that there is no association between male and female populations, while the alternative hypothesis H_a claims that an association does exist. The expected counts of male and female individuals are shown in the columns E_k^{m} and E_k^{fm} respectively in Table 2.5. The chi-square test statistic is calculated as

$$\chi^2 = \sum_{k=1}^{16} \frac{(O_k - E_k)^2}{E_k}$$

where the square of the differences between the observed and expected counts divided by the expected count (i.e. $(O_k^{\mathrm{m}} - E_k^{\mathrm{m}})^2 / E_k^{\mathrm{m}}$ and $(O_k^{\mathrm{fm}} - E_k^{\mathrm{fm}})^2 / E_k^{\mathrm{fm}}$) are shown across of the cells in Table 2.5 for both male and female individuals.

The test statistic χ^2 for this case is calculated as the sum of the last two columns in Table 2.5, and we obtain $\chi^2 = 252.77$. The degrees of freedom are equal to $(16-1)(2-1) = 15$. So, the p-value for the chi-square is $\mathrm{P}(\chi^2 \leq 252.77) = 3.34 \times 10^{-45}$, giving strong evidence against the null hypothesis of no association. Therefore the difference between observed and expected counts of individuals is not negligible between gender groups.

In this section we have shown how to consider time-varying p_t in terms of both weather covariates and gender simultaneously in the logistic regression. We also compared males and females using chi-square test for association where grouping is used to avoid small expected frequencies after the 16th sampling occasion. This example is useful for demonstrating how to handle gender as a covariate along with time-varying capture probability.

Table 2.3 Results of fitting the geometric removal model with time and gender variations in p_t. Minimum and maximum air temperature are denoted by MinAT and MaxAT respectively. Minimum and maximum ground temperature are denoted by MinGT and MaxGT respectively. † in the cell indicates the covariate is used. h is the number of parameters and ML is the value of the maximised loglikelihood.

MinAT	MaxAT	MinGT	MaxGT	h	ML	AIC
†		†		6	896.42	908.42
†	†	†		7	896.14	910.14
†	†		†	7	904.24	918.24
†				5	935.37	945.37
†	†			6	934.37	946.37
	†	†	†	7	1290.92	1304.92
†	†	†	†	8	1290.92	1306.92
	†	†		6	1300.39	1312.39
		†	†	6	1313.37	1325.37
	†		†	6	1313.38	1325.38
			†	5	1317.85	1327.85
†			†	6	1356.16	1368.16
	†			5	1360.78	1370.78
		†		5	1394.73	1404.73

Table 2.4 Estimated parameters from the model with lowest AIC in Table 2.3 when modelling climatic covariates and gender simultaneously. SEs are calculated by 500 bootstrapped samples.

Parameter	Estimate (SE)
$\hat{\alpha}$	-2.37 (0.20)
$\hat{\beta}_{\text{gender}}$	-0.10 (0.01)
$\hat{\beta}_{\text{MinAT}}$	0.18 (0.01)
$\hat{\beta}_{\text{MinGT}}$	-0.11 (0.03)
$\hat{N}_{\text{male}}$	742.00 (0.02)
$\hat{N}_{\text{female}}$	699.00 (0.01)

Table 2.5 Grouped observed (O_k^{m} and O_k^{fm}) and expected (E_k^{m} and E_k^{fm}) counts of individuals are between second and fifth columns. Chi-square test statistics contributions are shown in the last two columns. The superscripts m and fm represent male and female respectively. *k* represents the *k*th sampling occasion.

k	O_k^{m}	O_k^{fm}	E_k^{m}	E_k^{fm}	$(O_k^{\mathrm{m}} - E_k^{\mathrm{m}})^2/E_k^{\mathrm{m}}$	$(O_k^{\mathrm{fm}} - E_k^{\mathrm{fm}})^2/E_k^{\mathrm{fm}}$
1	65	41	80.18	69.98	2.87	12.00
2	115	99	88.43	78.04	7.98	5.63
3	185	165	163.89	148.08	2.72	1.93
4	49	44	44.51	40.58	0.45	0.29
5	13	13	28.21	25.88	8.20	6.41
6	51	57	49.64	46.09	0.04	2.58
7	9	9	20.17	18.84	6.19	5.14
8	38	29	29.98	28.26	2.14	0.02
9	50	68	66.61	64.27	4.14	0.22
10	52	55	24.05	23.47	32.48	42.34
11	16	14	14.05	13.82	0.27	0.00
12	7	9	8.62	8.53	0.31	0.03
13	5	3	8.34	8.29	1.34	3.38
14	8	8	29.66	30.13	15.82	16.25
15	22	23	7.27	7.44	29.82	32.55
≥ 16	57	62	78.33	78.33	5.81	3.41

2.5 Recent developments

Removal models are commonly used to estimate abundance for populations (Pollock, 1991; Hilborn and Walters, 1992), and they have recently been presented as a class of hierarchical models. Dorazio and Howard (2005) present a hierarchical removal model where the sampling sites are assumed to have several distinct sub-sites located spatially. Additionally, Laplanche (2010) compares the performance of 12 hierarchical models using removal sampling data obtained at multiple locations. As an alternative to the logistic model presented in Chapter 2, Clair et al. (2013) proposed a hierarchical removal model using catch effort to account for time variations in capture probability when removal sampling effort (e.g. number of traps/hunters) is recorded on each sampling occasion at different survey locations. Chandler et al. (2011) developed a spatially explicit temporary emigration model permitting the estimation of population density for point count data such as removal sampling, double-observer sampling, and distance sampling. However, their model cannot be applied to removal data when spatial information is unavailable. More recently, Matechou et al. (2016) developed a Bayesian approach for removal data observed at a single site which allows for population renewal through birth/immigration as well as for population depletion through death/emigration in addition to the removal process. In the subsequent chapters we propose new removal models on a single site motivated by real data and violations of the assumptions in the classic removal model as described in Section 2.2.2.

Within this thesis we show that the estimation of temporary emigration from removal data relies on the use of the robust design as demonstrated in Chapters 3 and 4, where there are at least two secondary occasions within a primary occasion. Furthermore, we develop new penalised maximum likelihood estimation in Chapter 5 for removal data collected under the standard sampling protocol. However, this approach can be time-consuming due to the need to employ cross-validation and its performance relies on fairly large sample sizes. As the robust design is highly preferred for removal sampling, we look into the design of removal sampling at a single site while accounting for the availability of individuals under the robust design in Chapter 6. The aim is to consider how to allocate the total sampling effort for the removal sampling under the robust design sampling protocol. We use both simulations and analytic forms to investigate different sampling approaches where constant parameters are assumed.

Chapter 3

Removal Models Accounting for Temporary Emigration

3.1 Introduction

Ecologists working on mitigation projects have devoted a large amount of time to removal models that assume immigration and emigration are non-existent. However, these effects are generally quite important, and immigration and emigration often occurs which results in population dynamics in the population. Models ignoring immigration and emigration of individuals could produce biased estimates of total population size if populations of interest do exhibit these ecological behaviors. Therefore, in Chapter 3 we develop a new removal model that enables the estimation of emigration and immigration within a population. The work has been published in Zhou et al. (2018).

All existing removal models assume that any emigration from the population is permanent during the study, but this assumption will be violated if individuals temporarily emigrate to an area outside the sampling area. Amphibians and reptiles exhibit this behavior in practice as they can temporarily hide in shelters resulting in zero detection probabilities during some sampling occasions for part of the population (Edgar et al., 2010; Rafferty, 2011). For example, slow-worms, *Anguis fragilis*, a legless lizard, primarily live underground or underneath objects lying on the ground and although they may be detected basking on the ground, most activity takes place out of the sight of ecologists (Edgar et al., 2010). Such temporary emigration can be modelled as a partially hidden process between two states that describes the underlying movement pattern of individuals between the study area and an area outside of the study.

The multievent framework, formulated by Pradel (2005), accommodates uncertainty of states for capture-recapture data. However, no approach currently exists for modelling temporary emigration for removal studies and ignoring such ecological features of species may result in a biased estimate of the number of individuals left behind after the end of removal projects and subsequently risk translocation failure if only existing modelling approaches are used.

Pollock (1982) was the first paper in capture recapture studies to propose an approach called robust design which allows the estimation of important demographic parameters including estimates of immigration and emigration as well as population size. The robust design is a combination of the open Cormack-Jolly-Seber model (Cormack, 1964; Jolly, 1965; Seber, 1965) and closed population models (Otis et al., 1978). It accommodates multiple secondary sampling occasions within each primary sampling period, and enables the estimation of temporary emigration from the study site (Kendall et al., 1995, 1997; Kendall and Bjorkland, 2001). The population is assumed to be open for temporary emigration between primary occasions but closed within each primary sampling period. Such emigration can be modelled as a first-order Markov process with different transition probabilities for individuals depending on which state they currently reside. Motivated by real data and ecological features of amphibians and reptiles, we develop novel removal models that bring together both the robust design (Pollock, 1982; Kendall et al., 1995) and a multievent structure (Pradel, 2005) for removal data using maximum likelihood inference. The objective is to provide an unbiased estimate of the number of animals remaining at the site at the end of the removal project.

This chapter is structured as follows: in Section 3.2 we describe the parameters, the robust design multievent removal modelling (RMER) framework and the formulation of integrated RMER (IRMER) for modelling multiple populations simultaneously. We also discuss ways of constraining the parameters of RMER and IRMER models under various ecological scenarios. Section 3.3 presents simulations for the proposed removal models and explores the benefits of the new modelling approach compared with multievent removal models (MER) without robust design. We also demonstrate by simulation that the existing geometric removal models (denoted as GRM) overestimate the number of individuals remaining at the site at the end of the study for a population exhibiting temporary emigration. We also investigate the use of a constant survival probability as an extra parameter in our proposed model in Section 3.4. Within Section 3.5 we present the results obtained from fitting IRMER models to juvenile and adult data of common lizards, *Zootoca vivipara*. Conclusions are discussed in Section 3.6.

3.2 Formulation

3.2.1 Notation

Consider a removal experiment conducted at a site on a population of N individuals, where N is the total number of animals that become exposed to sampling efforts at least once during the study. Individuals are permanently removed from the study area once captured during the study period. Suppose there are two states in the model: individuals in state 1 are present and available for removal, while individuals in state 2 are absent from the study site and hence unavailable for capture. We assume the removal study is conducted within a robust design framework comprising T primary periods and k_i secondary sampling occasions within the ith primary period, $i = 1, \ldots, T$. Transitions between two states are only allowed between primary sampling periods and the population is assumed to be closed within a primary session. We denote the total number of sampling occasions by $K(= \sum_{i=1}^{T} k_i)$. The removal data set that arises is a matrix with entry $n_{i,j}$ representing the number of individuals removed at the jth secondary sample within the ith primary period, where $i = 1, \ldots, T$ and $j = 1, \ldots, k_i$. The total number of individuals removed is denoted by D, where $D = \sum_{i=1}^{T} \sum_{j=1}^{k_i} n_{i,j}$. The parameters in the model are defined below:

- n_0: the number of animals that have not been removed by the end of the study, where $n_0 = N - D$.

- $\boldsymbol{\pi}$: the initial state matrix, defined as a row vector, $(\pi, 1-\pi)$, where π represents the proportion of individuals in state 1 and the complement of π, $1-\pi$, the proportion of individuals in state 2, at the start of the study.

- $\boldsymbol{\Phi}_i$: the state transition matrix, where ϕ_i^{12} and ϕ_i^{21} are transition probabilities from state 1 to state 2, and transition probabilities from state 2 to state 1 respectively, between the ith and $(i+1)$th primary period, where $i = 1, \ldots, T-1$.

$$\boldsymbol{\Phi}_i = \begin{pmatrix} 1-\phi_i^{12} & \phi_i^{12} \\ \phi_i^{21} & 1-\phi_i^{21} \end{pmatrix}$$

- $\boldsymbol{B}_{i,j}$: the state-event matrix, where events are "Removed" and "Not Removed" in the first and second column respectively. States 1 and 2 are in the first and second row respectively. $p_{i,j}$ is the probability that an individual is captured at the jth secondary sample within the ith primary period, where $i = 1, \ldots, T$ and

$j = 1, \ldots, k_i$.

$$\boldsymbol{B}_{i,j} = \begin{pmatrix} p_{i,j} & 1 - p_{i,j} \\ 0 & 1 \end{pmatrix}$$

We define $\boldsymbol{P}_{i,j}$ to be the detection probability matrix at the ith primary period and the corresponding jth secondary sample as a diagonal matrix with diagonal elements equal to the first column vector of $\boldsymbol{B}_{i,j}$ corresponding to event "Removed".

$$\boldsymbol{P}_{i,j} = \text{Diag}\{\boldsymbol{B}_{i,j}(\cdot, 1)\} = \begin{pmatrix} p_{i,j} & 0 \\ 0 & 0 \end{pmatrix}$$

Similarly, $\boldsymbol{Q}_{i,j}$ is the diagonal probability matrix of not detecting individuals with diagonal elements equal to the second column of $\boldsymbol{B}_{i,j}$.

$$\boldsymbol{Q}_{i,j} = \text{Diag}\{\boldsymbol{B}_{i,j}(\cdot, 2)\} = \begin{pmatrix} 1 - p_{i,j} & 0 \\ 0 & 1 \end{pmatrix}$$

Note that constant parameters are designated by the absence of the subscript from the corresponding time-specific parameters, e.g. p denotes a constant capture probability over time.

3.2.2 The likelihood formulation

We adopt a multievent approach (Pradel, 2005) taking into account the robust design framework (Kendall et al., 1995, 1997; Kendall and Bjorkland, 2001) for the computation of the likelihood function. Consider a removal model with robust design and multievent framework for a study with T primary sampling periods and k_i secondary sampling occasions within the ith primary period, where $i = 1, \ldots, T$.

The probability of an individual being removed at the 1st sample within the 1st primary period is,

$$L_{1,1} = \boldsymbol{\pi} \boldsymbol{Q}_{1,0} \boldsymbol{P}_{1,1} \mathbf{1}_2$$

where $\boldsymbol{Q}_{1,0} = I_2$ is the 2×2 identity matrix and $\mathbf{1}_2$ is the column vector of two ones (and thereafter).

The probability of an individual being captured and removed at the jth sample within the 1st primary period is,

$$L_{1,j} = \boldsymbol{\pi} \boldsymbol{Q}_{1,0} \boldsymbol{Q}_{1,1} \cdots \boldsymbol{Q}_{1,j-1} \boldsymbol{P}_{1,j} \mathbf{1}_2$$

where $j = 1, \ldots, k_1$.

Then, the probability of an individual being removed at the 1st sample within the ith primary period is,

$$L_{i,1} = \boldsymbol{\pi}\boldsymbol{Q}_{1,0}\boldsymbol{Q}_{1,1}\cdots\boldsymbol{Q}_{1,k_1}\boldsymbol{\Phi}_1\boldsymbol{Q}_{2,0}\cdots\boldsymbol{Q}_{2,k_2}\boldsymbol{\Phi}_2\cdots\boldsymbol{\Phi}_{i-1}\boldsymbol{Q}_{i,0}\boldsymbol{P}_{i,1}\mathbf{1}_2$$

where $i = 2,\dots,T$, $j = 1,\dots,k_i$ and $\boldsymbol{Q}_{i,0} = I_2$ is the 2×2 identity matrix (and thereafter).

The probability of an individual being removed at the jth sample within the ith primary period is,

$$L_{i,j} = \boldsymbol{\pi}\boldsymbol{Q}_{1,0}\boldsymbol{Q}_{1,1}\dots\boldsymbol{Q}_{1,k_1}\boldsymbol{\Phi}_1\boldsymbol{Q}_{2,0}\cdots\boldsymbol{Q}_{2,k_2}\boldsymbol{\Phi}_2\dots\boldsymbol{\Phi}_{i-1}\boldsymbol{Q}_{i,0}\cdots\boldsymbol{Q}_{i,j-1}\boldsymbol{P}_{i,j}\mathbf{1}_2$$

where $i = 2,\dots,T$, $j = 1,\dots,k_i$.

The probability of not being removed by the end of the study is given by

$$L_0 = \boldsymbol{\pi}\boldsymbol{Q}_{1,0}\boldsymbol{Q}_{1,1}\cdots\boldsymbol{Q}_{1,k_1}\boldsymbol{\Phi}_1\boldsymbol{Q}_{2,0}\cdots\boldsymbol{Q}_{2,k_2}\boldsymbol{\Phi}_2\cdots\boldsymbol{\Phi}_{T-1}\boldsymbol{Q}_{T,0}\cdots\boldsymbol{Q}_{T,k_T}\mathbf{1}_2.$$

The full product multinomial likelihood is given by

$$\boldsymbol{L}(\boldsymbol{\pi},\boldsymbol{\phi}_i^{12},\boldsymbol{\phi}_i^{21},\boldsymbol{p}_{i,j},\boldsymbol{n}_0|\boldsymbol{n}_{i,j}) = \frac{N!}{n_{1,1}!n_{1,2}!\dots n_{T,k_T}!n_0!}\Big(\prod_{i=1}^{T}\prod_{j=1}^{k_i}L_{i,j}^{n_{i,j}}\Big)L_0^{n_0}. \tag{3.1}$$

We note that the likelihood function in (3.1) can be easily adapted to accommodate multiple species or different age/sex groups for a single species using an integrated population modelling approach (Besbeas et al., 2002; McCrea and Morgan, 2014, Chapter 12). Consider a removal experiment conducted on W species (or W categories for a single species) with the individual likelihoods defined as $L_1, L_2, \dots, L_W$ which are of the form shown in Equation (3.1). Assuming the W groups of individuals are removed independently, the full likelihood L can be written as the product of individual likelihoods, i.e $L = L_1 \times L_2 \times \cdots \times L_W$.

The model belongs to the family of hidden Markov models (Pradel, 2005; Zucchini et al., 2016), so standard errors can be obtained from the Hessian. When some parameters lie on the boundary, non-parametric bootstrap can be used instead to compute standard errors and confidence intervals (Zucchini et al., 2016, Section 3.6).

3.2.3 Constraints

The model that assumes fully time-dependent parameters (π, $p_{i,j}$, ϕ_i^{12}, ϕ_i^{21}, n_0) has $K+2T$ parameters and the number of observations is K. Such a model is termed as parameter redundant (Cole et al., 2010) as it can be reparameterised in terms of a smaller set of parameters, and therefore we cannot estimate all of the parameters individually without further constraints. We discuss parameter redundancy of various models in Chapter 4.

One natural way of enabling the estimation of parameters involves assuming parameters are constant over time (Sandland and Kirkwood, 1981). Furthermore, we consider a list of alternative forms of constraints for our proposed models, described below:

- Constraints related to detection probability.
    - The time-dependent capture probability $p_{i,j}$ can be modelled using a logistic regression (North and Morgan, 1979) in terms of covariate $z_{i,j}$ at the jth secondary occasion within the ith primary period, i.e. $\text{logit}(p_{i,j}) = \log\{p_{i,j}/(1-p_{i,j})\} = \alpha + \beta z_{i,j}$. We label this constraint “Z”.
    - If the capture probability is constant over time, then we denote it by “C”.
- Constraints related to the initial state parameter.
    - $\pi = \{1/(T-1)\}\sum_{i=1}^{T-1}\{\phi_i^{21}/(\phi_i^{12}+\phi_i^{21})\}$, the initial state parameter π can be constrained using this expression if we assume the population is initially allocated to two states according to the mean of the stationary distributions of the transition matrices across time which is $\{\phi_i^{21}/(\phi_i^{12}+\phi_i^{21}), \phi_i^{12}/(\phi_i^{21}+\phi_i^{21})\}$, $i = 1,\dots,T-1$. We name this constraint “S” to represent that the stationary distribution is being assumed for π.
- Constraints related to transition probability.

  The superscript “t" in the following constraints denotes fully time-dependent transition probabilities and the absence of the superscript indicates that constant transition parameters are assumed.
    - $\phi_i^{12} + \phi_i^{21} = 1$. This constraint is equivalent to the random emigration/-movement model for capture-recapture data sampled with robust design as described in Kendall et al. (1995, 1997), where they set $\gamma_i^{'} = \gamma_i^{''}$ which are ϕ_i^{12} and $1-\phi_i^{21}$ respectively in our notation. It implies that the probability of being in the unobservable state between the ith and $(i+1)$th primary session is the same for individuals in and individuals outside the study area. In this case, we treat the transition probability ϕ_i^{12} as a free parameter to

be estimated in the model, and the transition probability ϕ_i^{21} is reparameterized using the constraint, i.e. $\phi_i^{21} = 1 - \phi_i^{12}$. This reparameterization can be simplified as $\phi^{12} + \phi^{21} = 1$ if we assume constant transition probabilities between primary periods. This constraint is labelled "R" and "R^t" to denote constant and time-dependent random emigration respectively.

- $\phi_i^{12} + \phi_i^{21} = v$ for $v \in (0,2)$ where v is an additional free parameter to be estimated. This constraint is motivated by the constraint of $\phi_i^{12} + \phi_i^{21} = 1$, but relaxes the assumption of the sum of two transition probabilities at primary period i being equal to 1. The free parameters are fully time-dependent ϕ_i^{12} and the constant v. Note that the limits of possible values of ϕ_i^{12} are conditional on v. The upper and lower bounds for ϕ_i^{12} are shown in the following,

  $$\begin{cases} \phi_i^{12} \in [v-1,1], & \text{for} \quad v \in [1,2] \\ \phi_i^{12} \in [0,v], & \text{for} \quad v \in [0,1] \end{cases}$$

  We label this constraint "V^t".
- $\phi_i^{12} = \phi_i^{21}$. This is an "even flow" model (Kendall et al., 1997), where the probability of transitioning from the study area to an unobservable state is the same as the probability of moving back to the study area between the ith and $(i+1)$th primary period. This is denoted by "E^t" and "E" for time-varying and constant transition parameters respectively.
- $\phi_T^{12} = \phi_{T-1}^{12}$ and $\phi_T^{21} = \phi_{T-1}^{21}$; the penultimate and final transition probabilities are assumed to be equal in this case. It is commonly used to overcome identifiability of the parameters in the first-order Markovian robust design model for capture-recapture data, as suggested in Kendall et al. (1997). We label this constraint by "2" as the transition probabilities are equal for the last two transitions. If constraint "2" is combined with either constraint "R^t" or "E^t", then "R$_2^t$" or "E$_2^t$" is used respectively.
- Suppose we have W populations of interest indexed as $w = 1, \ldots, W$. Time-dependent transition probabilities for population w, for example $\phi_{i,w}^{12}$, can be modeled using $\text{logit}(\phi_{i,w}^{12}) = \log\{\phi_{i,w}^{12}/(1-\phi_{i,w}^{12})\} = \eta_{1,i} + \gamma_w$, where $\eta_{1,i}$ is the logit of transition probabilities for a baseline population (numbered 1) and γ_w represents an additive effect of group w. We denote this constraint by "R$_a^t$" or "E$_a^t$" to represent the additive effect for the time-varying transition probabilities. We demonstrate that our IRMER model for two populations must be combined together with either "R^t" or "E^t", otherwise there will

> be more parameters in the model than the number of observations. If we have more than two populations, we can use "$\mathrm{R}^t_{a,w}$" or "$\mathrm{E}^t_{a,w}$" to represent the additive effect for $\phi^{12}_{i,w}$ for population w. In addition, for the IRMER models the absence of the subscript in "R^t_a" and "E^t_a" (i.e "R^t" and "E^t") denotes that time-dependent transition probabilities are equal for both of the populations accounting for the same constraint (i.e "R^t" or "E^t").

We propose a model name that is composed of both the model structure and the different combinations of constraints for all models that we consider. We employ the structure of "MODEL.XYZ" in Tables 3.2, 3.4, 3.5 and 3.6, where "MODEL" represents the model we considered (e.g. MER, RMER or IRMER). We also indicate the constraints used after a dot at the end of the model name. "X" and "Y" respectively represent the constraint used for the initial state parameter (e.g. "S") and the transition probabilities (e.g. "R^t_a"). For the IRMER models, we use the population number as the subscript in "S" to denote which population is subject to the constraint "S", for example, "$\mathrm{S}_{1,2}$" indicates that both initial state distributions for population 1 and 2 are assumed to be stationary. In addition, "N" suggests that no constraint has been considered. Furthermore, the last letter "Z" indicates whether the capture probability is constant over time ("C") or time-dependent in term of covariates ("Z"). For clarity, IRMER.NE_2C denotes the IRMER model with no constraint for the initial state parameter, the constraint "E^t_2" for the time-varying transition probabilities and a constant capture probability.

3.3 Simulation study

The aim of these simulations is to examine the bias and precision of maximum likelihood estimators for RMER, MER and IRMER within the likely range of ecological applications of the model. Three simulation settings, denoted by *Setting 3.3.1*, *Setting 3.3.2* and *Setting 3.3.3* are investigated, where RMER and MER models under *Setting 3.3.1* have constant transition probabilities, RMER models under *Setting 3.3.2* have time-varing transition probabilities and *Setting 3.3.3* presents results obtained from IRMER models with time-dependent transition probabilities. Simulations for MER models are used for comparison with corresponding RMER models with the same constraint without the use of the robust design in *Setting 3.3.1*. For *Setting 3.3.1* and *Setting 3.3.2*, 500 simulations are conducted for a study with $N = 500$ individuals, $K = 10$ or $K = 20$, with $T = 5$ or $T = 10$ primary periods and 2 secondary sampling occasions within each primary period (i.e. $k_1 = k_2 = \cdots = k_T = 2$). For *Setting 3.3.3*, we conduct simulations for IRMER, where 500 replicates are simulated for a removal study with two populations (numbered 1 and 2) where the population sizes are $N = 300$ and $M = 200$. We consider eight scenarios because the performance of the models depends on the relationship between ϕ_i^{12} and ϕ_i^{21} and on capture probability. Models with constraint "R" or "R^t" are investigated under Scenarios 1-4, while simulations are conducted for constraint "E" or "E^t" under Scenarios 5-8.

- Scenario 1: high capture probability and individuals tend to stay offsite ($\phi^{12} > \phi^{21}$),
- Scenario 2: high capture probability and individuals tend to stay onsite ($\phi^{12} < \phi^{21}$),
- Scenario 3: low capture probability and individuals tend to stay offsite ($\phi^{12} > \phi^{21}$),
- Scenario 4: low capture probability and individuals tend to stay onsite ($\phi^{12} < \phi^{21}$),
- Scenario 5: high capture probability and high mobility ($\phi^{12} = \phi^{21} > 0.5$),
- Scenario 6: high capture probability and low mobility ($\phi^{12} = \phi^{21} < 0.5$),
- Scenario 7: low capture probability and high mobility ($\phi^{12} = \phi^{21} > 0.5$),
- Scenario 8: low capture probability and low mobility ($\phi^{12} = \phi^{21} < 0.5$).

Under *Setting 3.3.2* and *Setting 3.3.3*, $\phi_i^{12} > \phi_i^{21}$ are set for the majority of the transition times for Scenarios 1 and 3, while $\phi_i^{12} < \phi_i^{21}$ for the majority of transition times for Scenarios 2 and 4. The true values of parameters used in the simulations are presented in the subsequent sections (*Setting 3.3.1*, *Setting 3.3.2* and *Setting 3.3.3*). In addition, the true values of capture probabilities p and transition probabilities ϕ_i^{12} for simulating the data under Scenarios 5-8 for the constraint "E" or "E^t" are the same with those under Scenarios 1-4 described as below.

3.3.1 Setting 3.3.1 RMER/MER with constant transition probabilities

Results related to constraint "R"

We show results from RMER.NNC, RMER.SNC, RMER.NRC, RMER.SRC and MER.SRC. We are interested in the precision of the estimators for the constraints used/not used for the initial state parameter and the transition probability for the RMER models. In addition, we show the results for the MER.SRC model where both "S" and "R" are taken into account but without the robust design for comparison.

The true value of constant capture probability is 0.3 under both Scenarios 3 and 4, and is 0.7 under Scenarios 1 and 2. In addition, we use $\phi^{12} = 0.8$, $\phi^{21} = 0.2$ in Scenarios 1 and 3, when individuals tend to move to the unobservable state, while in Scenarios 2 and 4 we define $\phi^{12} = 0.4$, $\phi^{21} = 0.6$ so that individuals tend to move to the observable state. The true value of the initial state parameter π is defined as the first element of the stationary distribution of the corresponding transition matrix.

Under Scenario 2 we captured nearly all of the individuals in the population so the estimates of population size are on the boundary of the parameter space as shown in Figure 3.1(A). As shown in Figures 3.1 and 3.2, it is clear that estimation of population size N is reliable for all models when $K = 20$, although long positive tails are recognised under Scenarios 1 and 3 when individuals tend to emigrate offsite and capturing them becomes impossible. When $K = 10$, longer positive tails are observed and the estimates of population size are biased downwards for models RMER.NNC and RMER.NRC under Scenarios 1 and 3. The results for detection probability p show that the use of the robust design considerably improves the performance in terms of bias, compared with the MER.SRC model that exhibits large bias for p in all cases. The bias in estimating ϕ_{12} is modest for RMER.SRC with both constraints "S" and "R" for all cases even when we only have $K = 10$ sampling occasions. In contrast, estimation of ϕ_{12} for MER.SRC without the robust design is not reliable for any cases. In addition, RMER.NNC yields biased estimates for ϕ^{12} due to near parameter redundancy. Overall, we conclude that RMER.SRC performs the best.

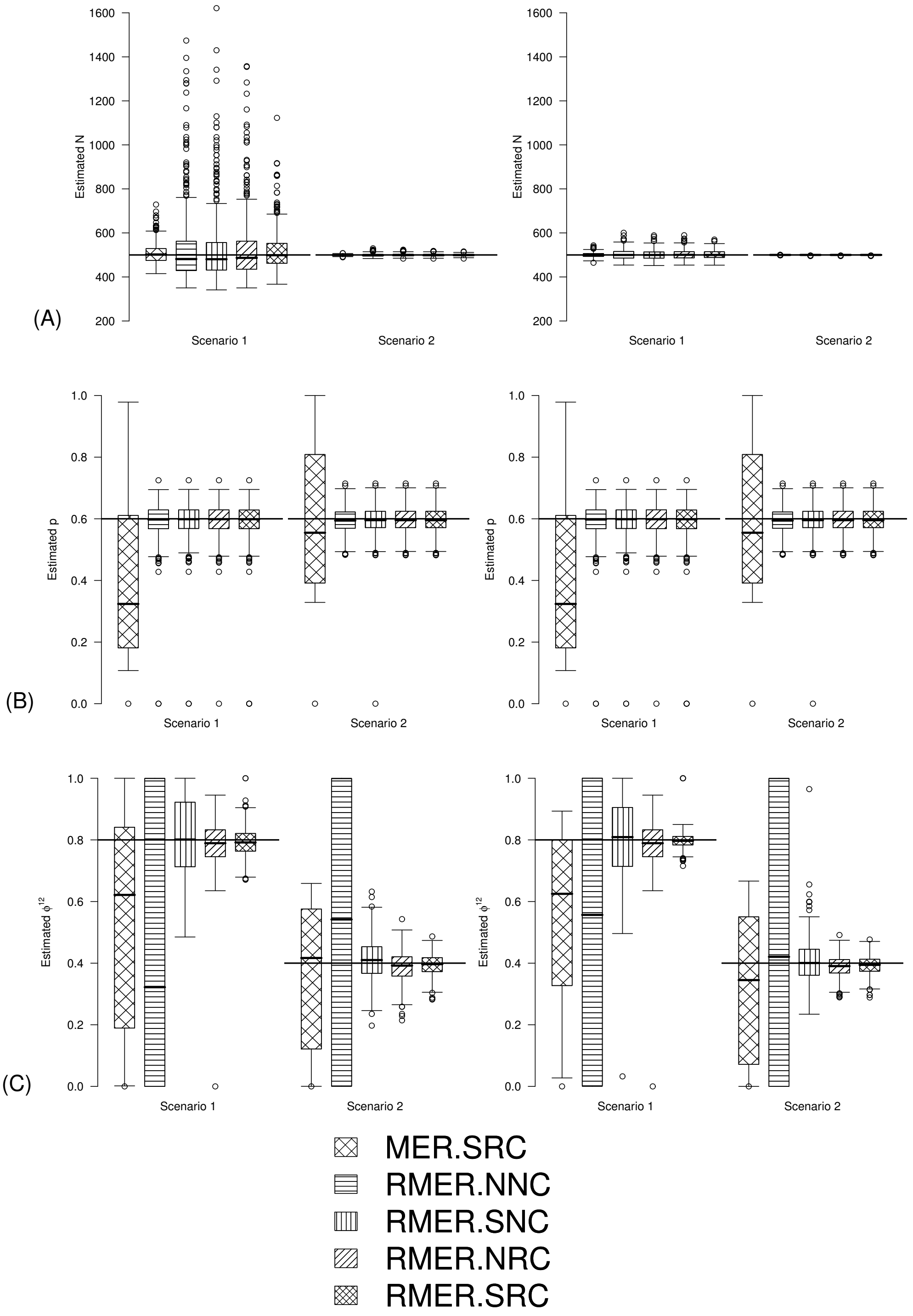


Fig. 3.1 Estimated population size N (A), capture probability p (B) and transition probability ϕ^{12} (C) for simulations with $K = 10$ and $K = 20$ sampling occasions displayed in the first and second columns respectively under Scenarios 1 and 2.

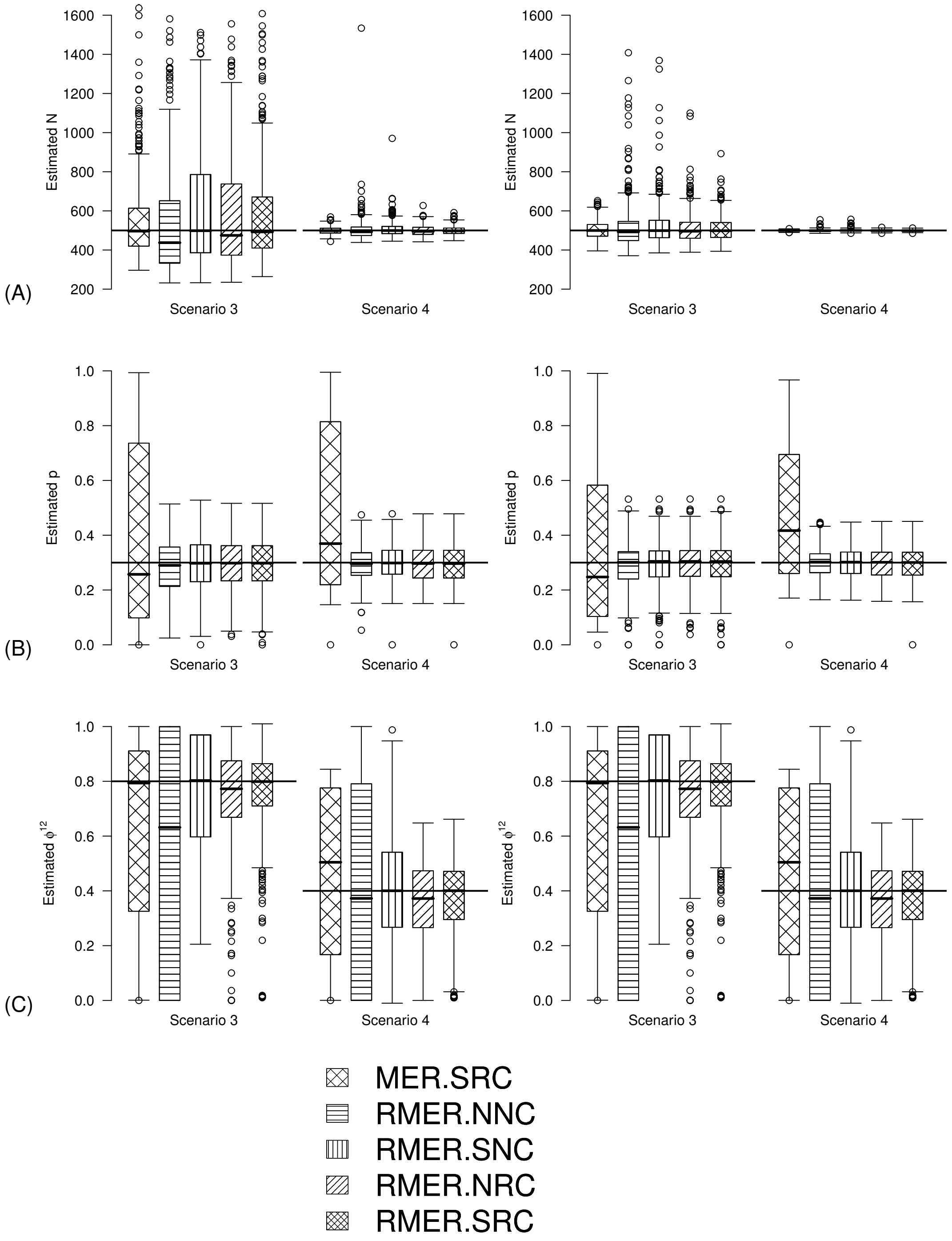


Fig. 3.2 Estimated population size N (A), capture probability p (B) and transition probability ϕ^{12} (C) for simulations with $K = 10$ and $K = 20$ sampling occasions displayed in the first and second columns respectively under Scenarios 3 and 4.

Results related to constraint "E"

In this Section, we show simulation results for models with constraint "E" under Scenarios 5-8. When the population has high mobility under Scenarios 5 and 7, individuals are exposed to the study area more frequently, so nearly all of the individuals are removed from the population by the end of the study. Therefore, we observe boundary estimates of population size as shown in Figure 3.3(A). In Figure 3.3(B) the results of estimated capture probabilities are unbiased apart from those obtained from the MER.NEC model. In Figure 3.3(C), unbiased estimates of ϕ^{12} are only obtained from the RMER.SEC model when we have $K = 20$ sampling occasions. However, the results in Section 3.3.2 suggest that biased estimated transition probabilities are obtained from the models that incorporate the constraint "E^t" when time-dependent transition probabilities are considered.

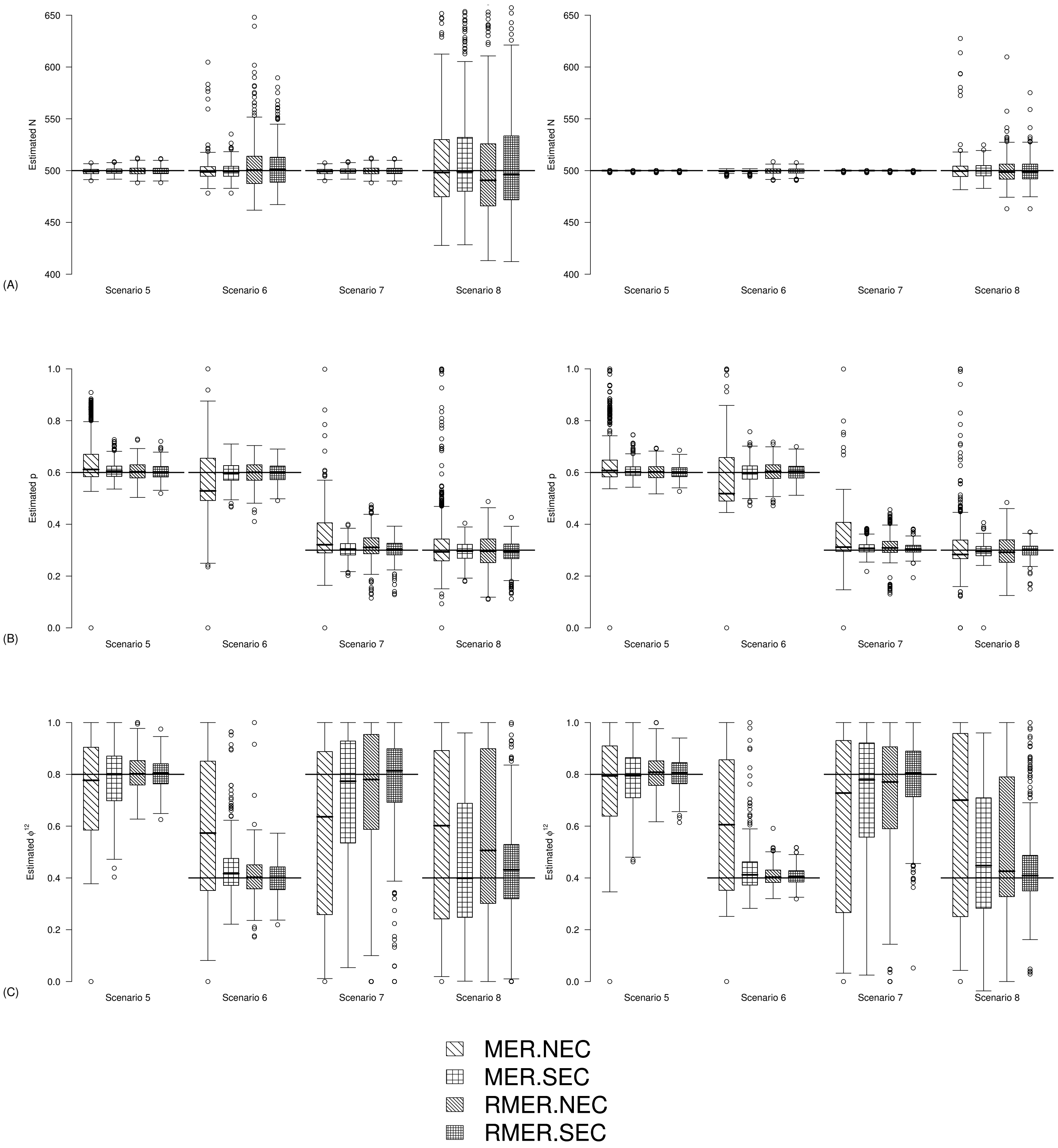


Fig. 3.3 Estimated population size N (A), capture probability p (B) and transition probability ϕ^{12} (C) for simulations with $K = 10$ and $K = 20$ sampling occasions displayed in the first and second columns respectively.

3.3.2 Setting 3.3.2 RMER with time-varying transition probabilities

Results related to constraint "R^t"

RMER models with constant transition probabilities may be not realistic for real data as individuals may tend to stay in one state at times. Here we investigate models RMER.NR^{t}C, RMER.SR^{t}C, RMER.SR$^t_2$C by simulation under Scenarios 1-4.

The true value of constant capture probability p is 0.55 for Scenarios 1 and 2 and 0.3 for both Scenarios 3 and 4. In addition, the true transition probabilities ϕ_i^{12} for simulating the data for RMER models under Scenario 3 when $K = 20$ are (0.8, 0.7, 0.8, 0.3, 0.6, 0.7, 0.8, 0.6, 0.6) where individuals tend to stay in the area outside the study for the majority of times which is more realistic for real data. Furthermore, the vector of true ϕ_i^{12} is defined as (0.4, 0.4, 0.8, 0.4, 0.4, 0.8, 0.4, 0.4, 0.4) when individuals are more likely to be present onsite. For a study with $K = 10$ occasions, we specify the true ϕ_i^{12} as (0.7, 0.2, 0.7, 0.7) for Scenario 3, and (0.3, 0.8, 0.3, 0.3) for Scenario 4. The value of the initial state parameter π is set to be the mean of the first element of the stationary distributions of transition matrices across time.

Figure 3.4 illustrates that the bias in the estimation of population size N is negative for the RMER.NR^{t}C model across all scenarios. The estimates of population size N are slightly biased downwards for RMER.SR^{t}C under Scenario 3 and unbiased for other Scenarios. The results of the estimated transition probabilities in Figure 3.5 suggest that unbiased estimates are obtained from RMER.SR^{t}C with both constraints "S" and "R^t". We also observe that the RMER.NR^{t}C model underestimates most of the transition probabilities in Figure 3.5. Furthermore, the estimated transition probabilities ϕ_{12} between the last two primary periods obtained from the RMER.NR^{t}C model are on the boundary across all Scenarios.

We also observe biased estimates for the RMER models with constraint "E^t", so none of the other of RMER models yields unbiased estimates of time-dependent transition probabilities apart from RMER.SR$^t_2$C which additionally incorporates the constraint "R$^t_2$".Therefore we conclude that when modelling time-dependent transition probabilities, we need to consider at least the combination of the constraint "S" and "R^t".

Results related to constraint "E^t"

In this Section, we present simulation results for the RMER.NE^{t}C, RMER.SE^{t}C and RMER.SE$^t_2$C models.

The boundary estimates of population size when $K = 20$ shown in Figure 3.6 is because we capture almost all of the individuals in the population during the study of length $K = 20$. The results in Figure 3.7 suggest that time-dependent transition probabilities obtained from the models incorporating the constraint “E^t” are not reliable. Therefore, the use of constraint “E^t” is not recommended if we want to model time-dependent transition probabilities.

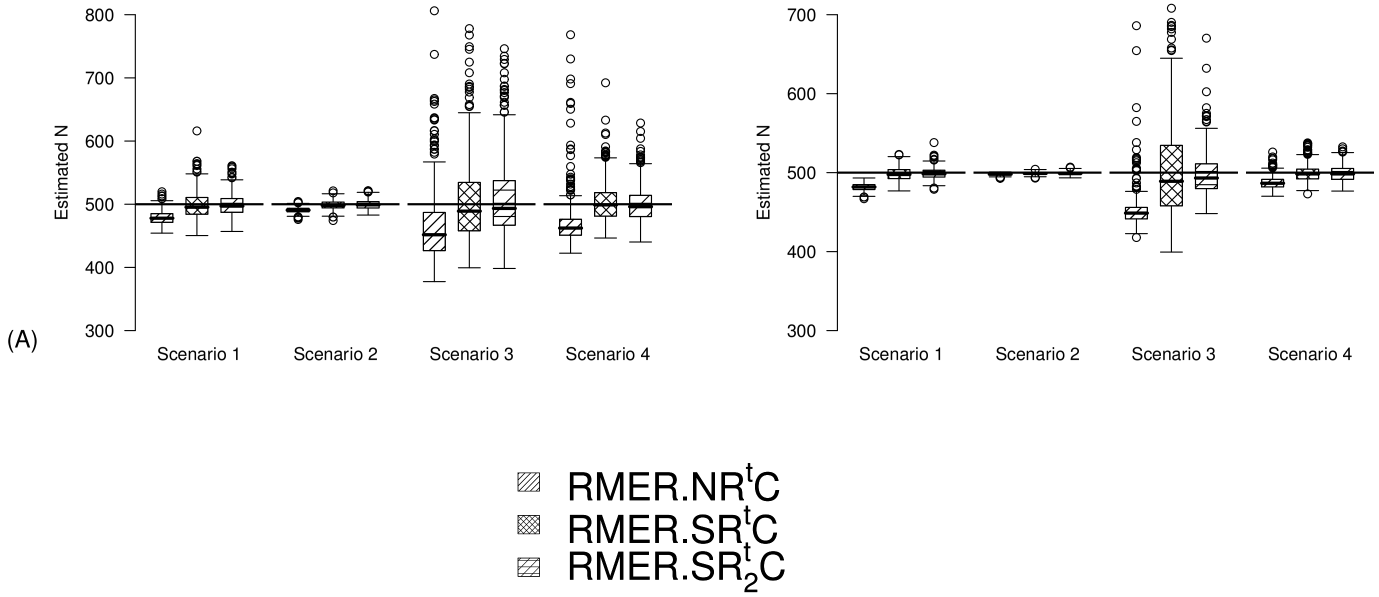


Fig. 3.4 Results of estimated population sizes. Simulations with $K = 10$ and $K = 20$ sampling occasions are displayed on the first and second columns respectively.

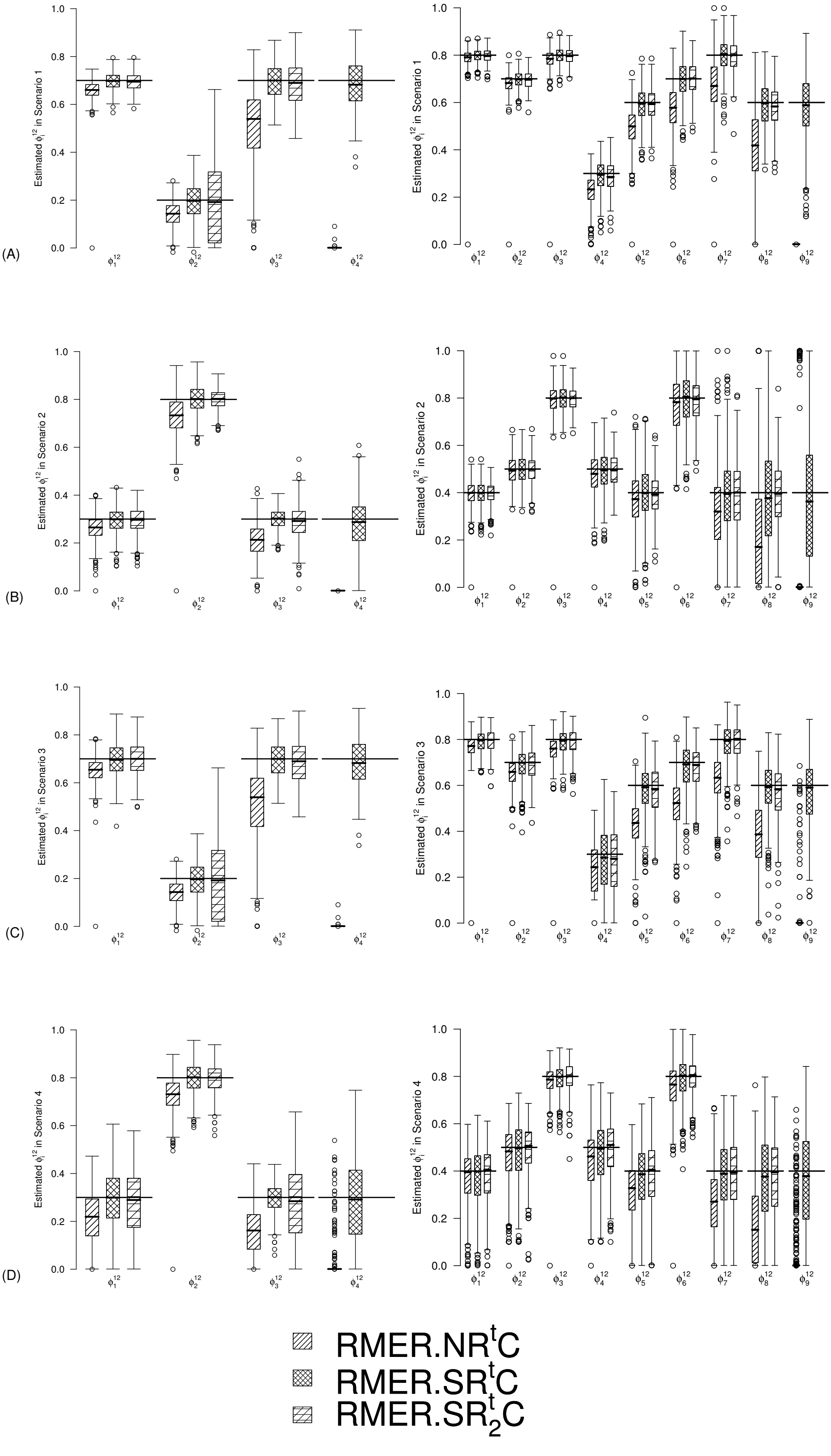


Fig. 3.5 Estimated time-dependent transition probability ϕ_i^{12} for simulations with $K = 10$ and $K = 20$ sampling occasions displayed in the first and last columns respectively.

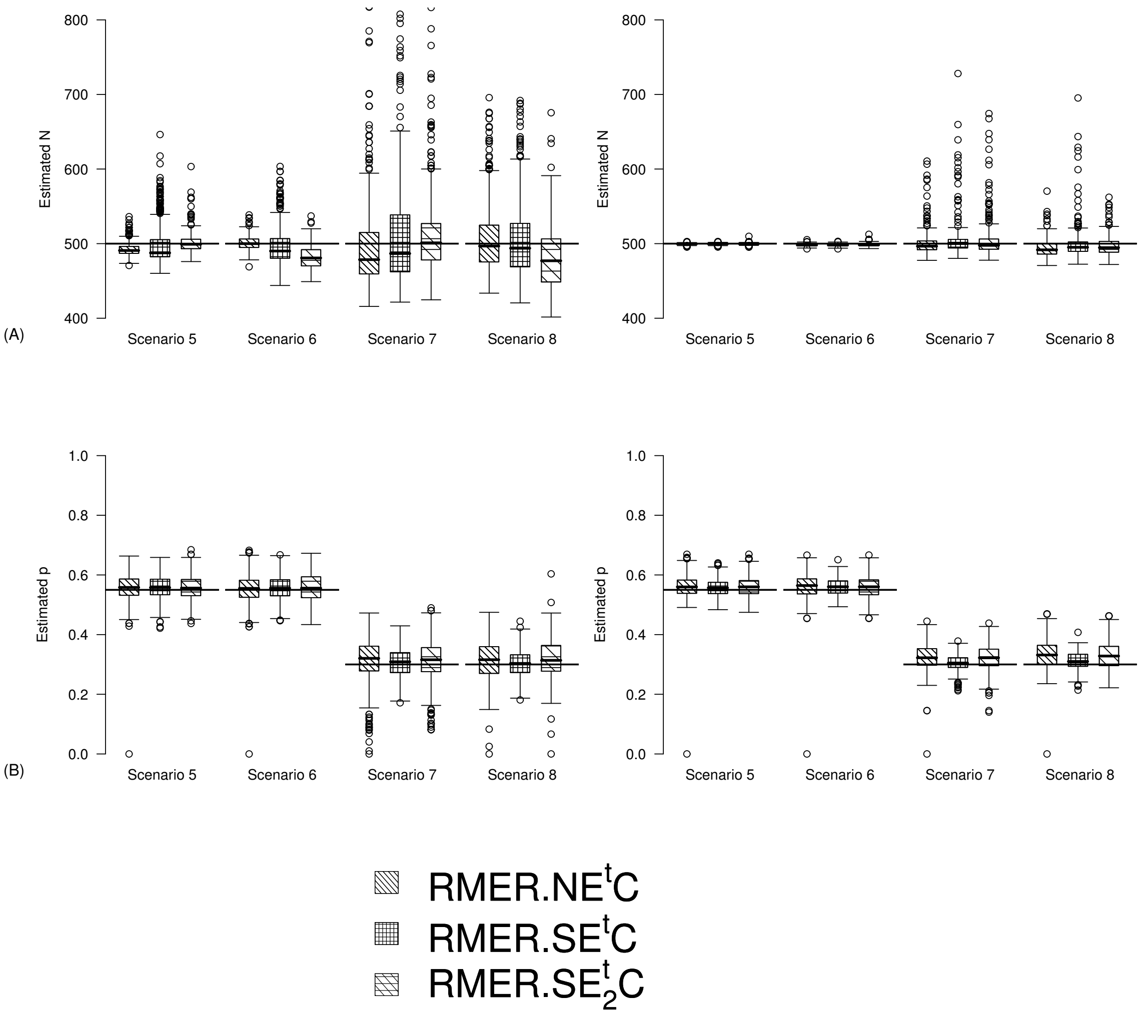


Fig. 3.6 Estimated population sizes N (A) and capture probabilities p (B) for simulations with $K = 10$ and $K = 20$ sampling occasions displayed in the first and last columns respectively.

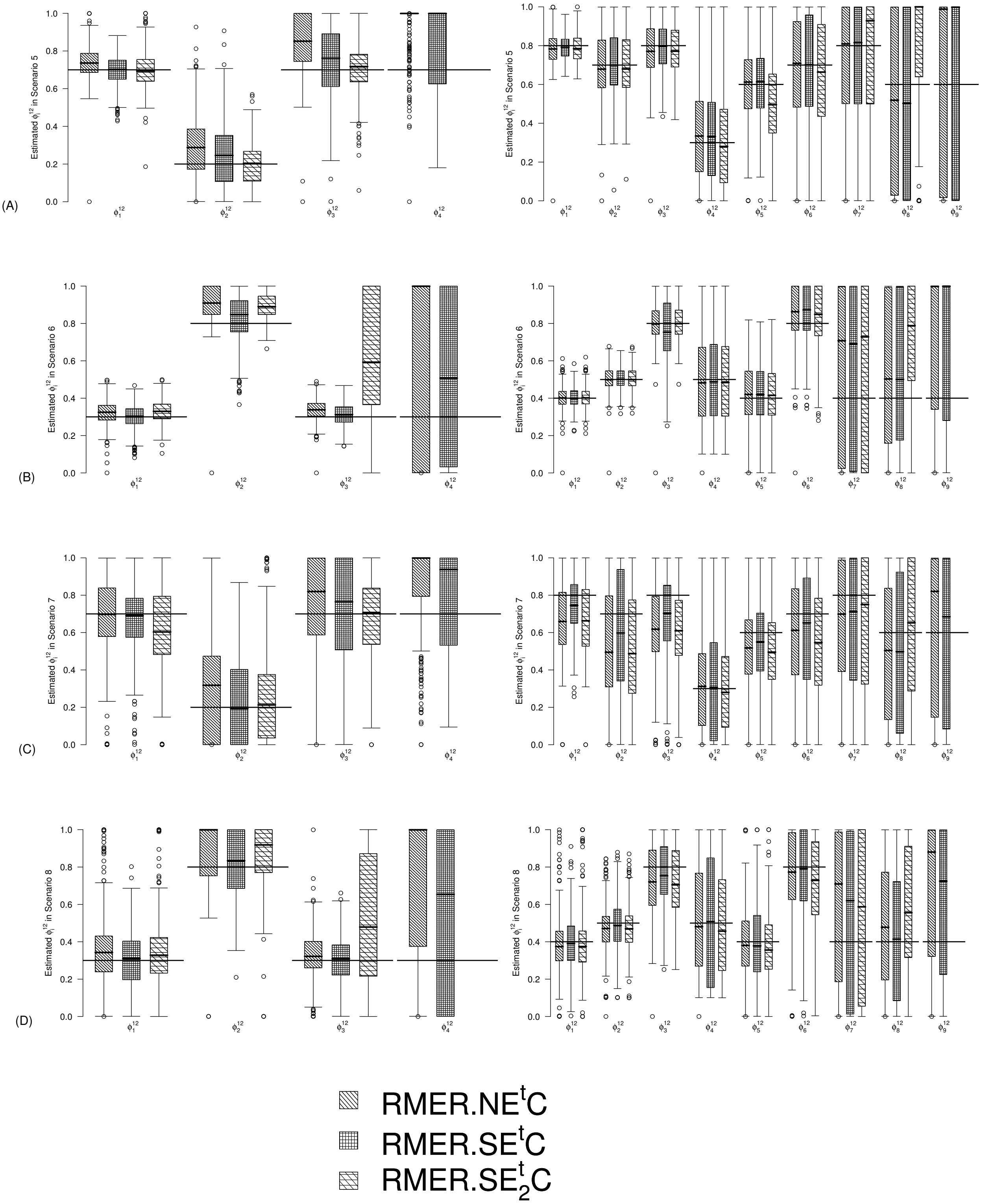


Fig. 3.7 Estimated time-dependent transition probability ϕ_i^{12} for simulations with $K = 10$ and $K = 20$ sampling occasions displayed in the first and last columns respectively. (A) Estimated ϕ_i^{12} under Scenario 5. (B) Estimated ϕ_i^{12} under Scenario 6. (C) Estimated ϕ_i^{12} under Scenario 7. (D) Estimated ϕ_i^{12} under Scenario 8.

Results related to constraint "V^t"

Here we investigate the generalization of the constraint "R^t" ($\phi_i^{12} + \phi_i^{21} = 1$) when $K = 10$. Simulations are displayed in Figures 3.8 and 3.9. Figure 3.8 show that unbiased estimates of capture probability are obtained for all values of v we tried ($v = 0.4, 1 or 1.4$) under all scenarios. We found that when $v = 1.4$, all of the estimates are unbiased under Scenario 2 and Scenario 4. However, the transition probabilities are overestimated under Scenario 1 and Scenario 3 as shown in Figure 3.9. When $v \leq 1$ the estimated v is biased high for all scenarios and the model overestimates the majority of time-dependent transition probabilities. This constraint may be useful when individuals have high mobility.

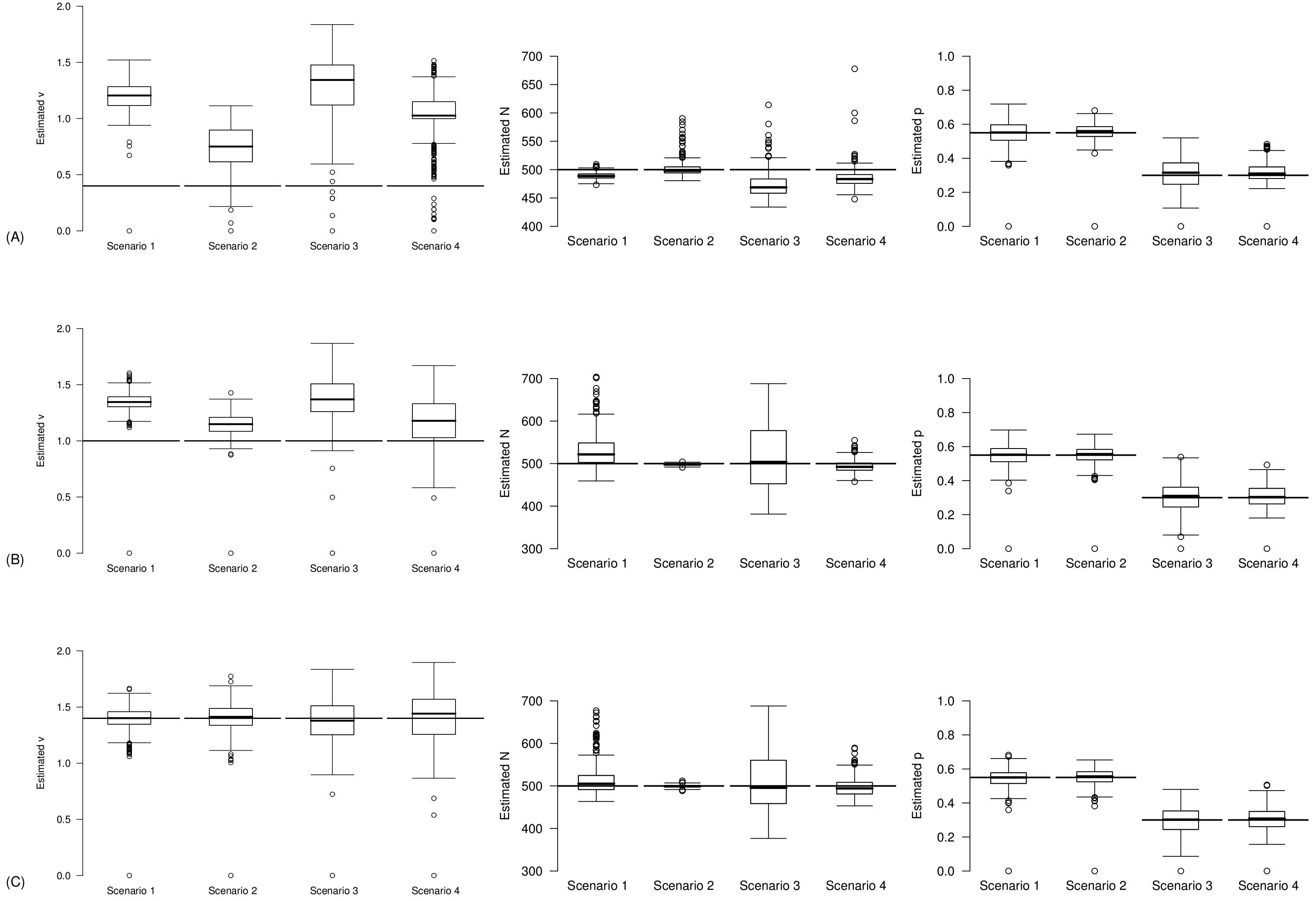


Fig. 3.8 Simulation results of ν (left), N (middle) and p (right) for RMER.NV^{t}C model with the constraint of $\phi_i^{12} + \phi_i^{21} = \nu$ where $\nu = 0.4$ (A), $\nu = 1$ (B) and $\nu = 1.4$ (C).

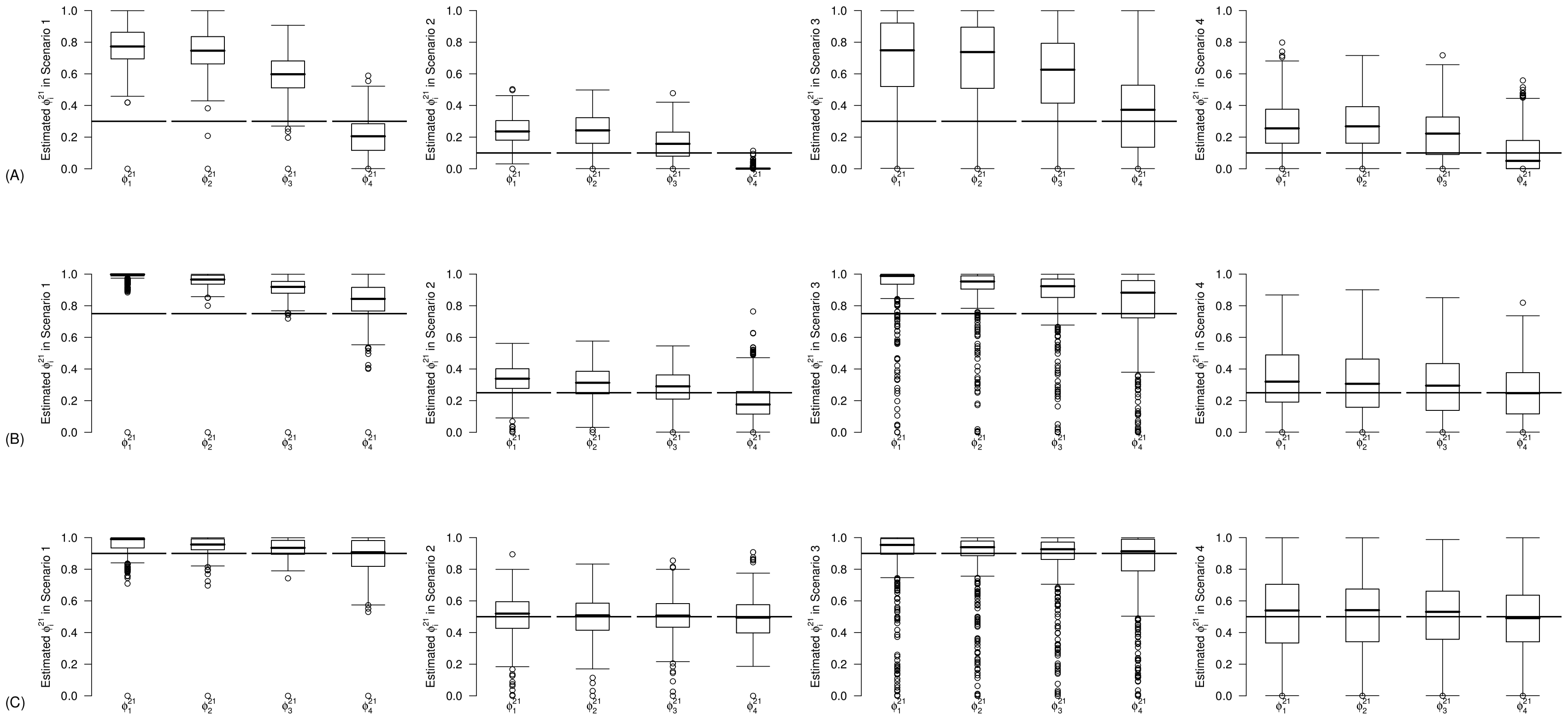


Fig. 3.9 Estimated time-dependent transition probabilities ϕ_i^{12} for RMER.NV^{t}C model with the constraint $\phi_i^{12} + \phi_i^{21} = v$ where $v = 0.4$ (A), $v = 1$ (B) and $v = 1.4$ (C).

3.3.3 Setting 3.3.3 IRMER with time-varying transition probabilities

We have demonstrated that we need to consider at least the combination of the constraint “R^t" and “S” in the model to get unbiased estimates of parameter in Section 3.3.2. Under this setting, we investigate the IRMER models IRMER.NR^{t}C, IRMER.S$_{1,2}$R^{t}C, IRMER.S$_{1,2}$R^t_aC and IRMER.S$_1$R^t_aC. These model are also investigated in Chapter 4 using symbolic algebra. The true values of parameters for population 1 are the same for those under *Setting 3.3.2*. We define the true value of the additive effect γ_2 for the population 2 to be -0.5 for IRMER.S$_{1,2}$R^t_aC and IRMER.S$_1$R^t_aC models when constraint “R^t_a” is used.

The estimates of population sizes for two populations and the additive effect γ_2 obtained from IRMER modelling are displayed in Figure 3.10. In Figures 3.10 the estimated population sizes N, M and transition probabilities are biased downwards for the IRMER.NR^{t}C model across all Scenarios. Therefore, we showed that the integrated modelling approach (Besbeas et al., 2002) cannot improve the biased estimates in the IRMER.NR^{t}C model with the constraint “R^t" only. Estimation of population size is unbiased for all models with the initial state constraint when $K = 20$. When the study has $K = 10$ sampling occasions, all of the IRMER.S$_{1,2}$R^{t}C, IRMER.S$_{1,2}$R^t_aC and IRMER.S$_1$R^t_aC models slightly underestimate population sizes under Scenarios 3 and 4. Moreover, the estimation of the additive effect γ_2 is unbiased under Scenarios 1 and 3 for both $K = 10$ and $K = 20$. However, when we have a small number of sampling occasions ($K = 10$) under Scenarios 2 and 4, γ_2 is slightly underestimated and it becomes unbiased for $K = 20$ sampling occasions. Figure 3.11 shows that the estimates of capture probability in IRMER.NR^{t}C are biased upwards, while they are unbiased for all other models. ϕ_i^{12} obtained from the IRMER.NR^{t}C model are underestimated as shown in Figure 3.12. Additionally, the estimated ϕ_i^{12} are unbiased initially from the start of the study but become slightly underestimated for the last few transitions. This is because removal sampling was conducted over a period of time and many individuals were removed from the study, so we expect fewer individuals to become available for capture by the end of the study which provides less information about transition probabilities.

As we have identified RMER and IRMER models that are able to provide reliable estimates of fully time-dependent transition probabilities, as discussed in *Setting 3.3.2* and *Setting 3.3.3*, we assume we can also model time-specific transition probabilities as a logistic function of covariates for RMER and IRMER models (denoted as constraint “Y” in Table 3.4).

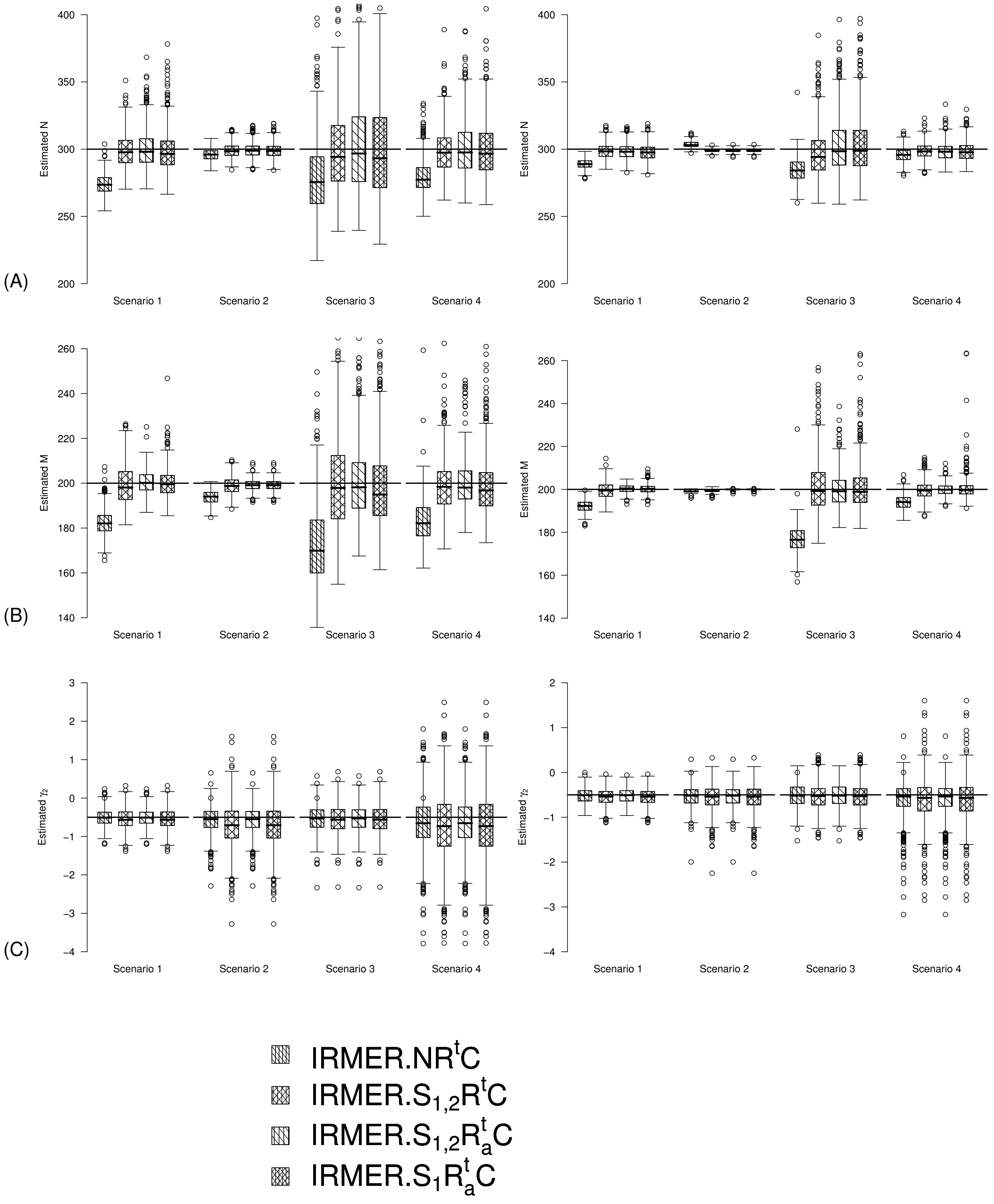


Fig. 3.10 Estimated population sizes N and M for simulations under Scenarios 1, 2, 3 and 4 with $K = 10$ and $K = 20$ sampling occasions displayed in the first and second columns respectively.

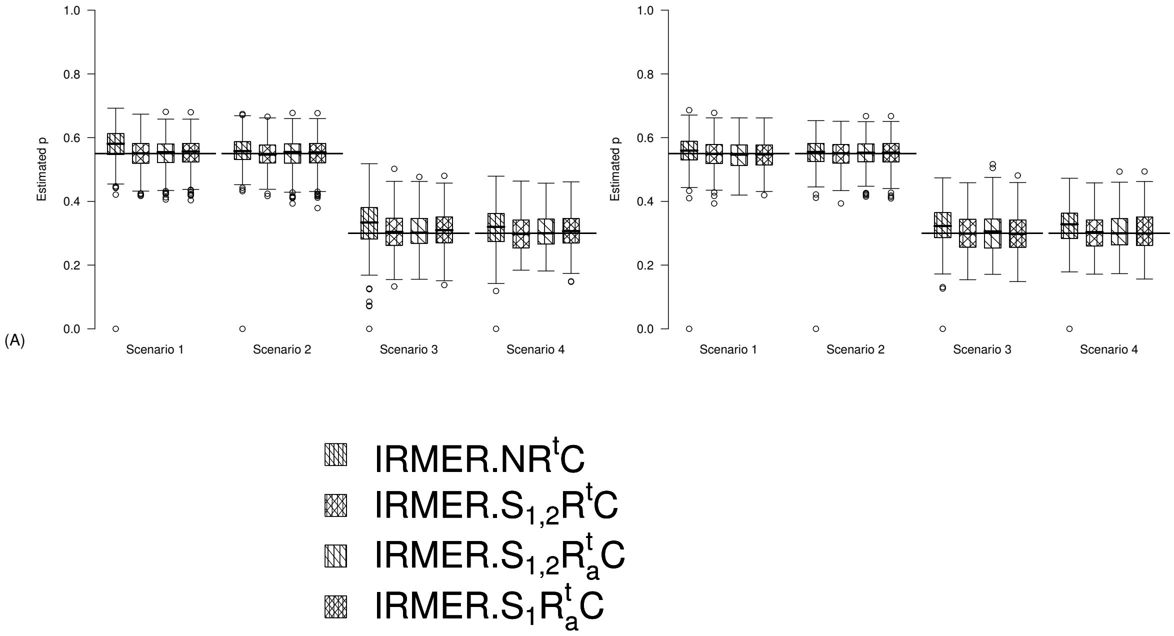


Fig. 3.11 Estimated capture probability p, initial state parameter π and additive effect γ_2 on transition probabilities for simulations with $K = 10$ and $K = 20$ sampling occasions displayed in the first and last columns respectively.

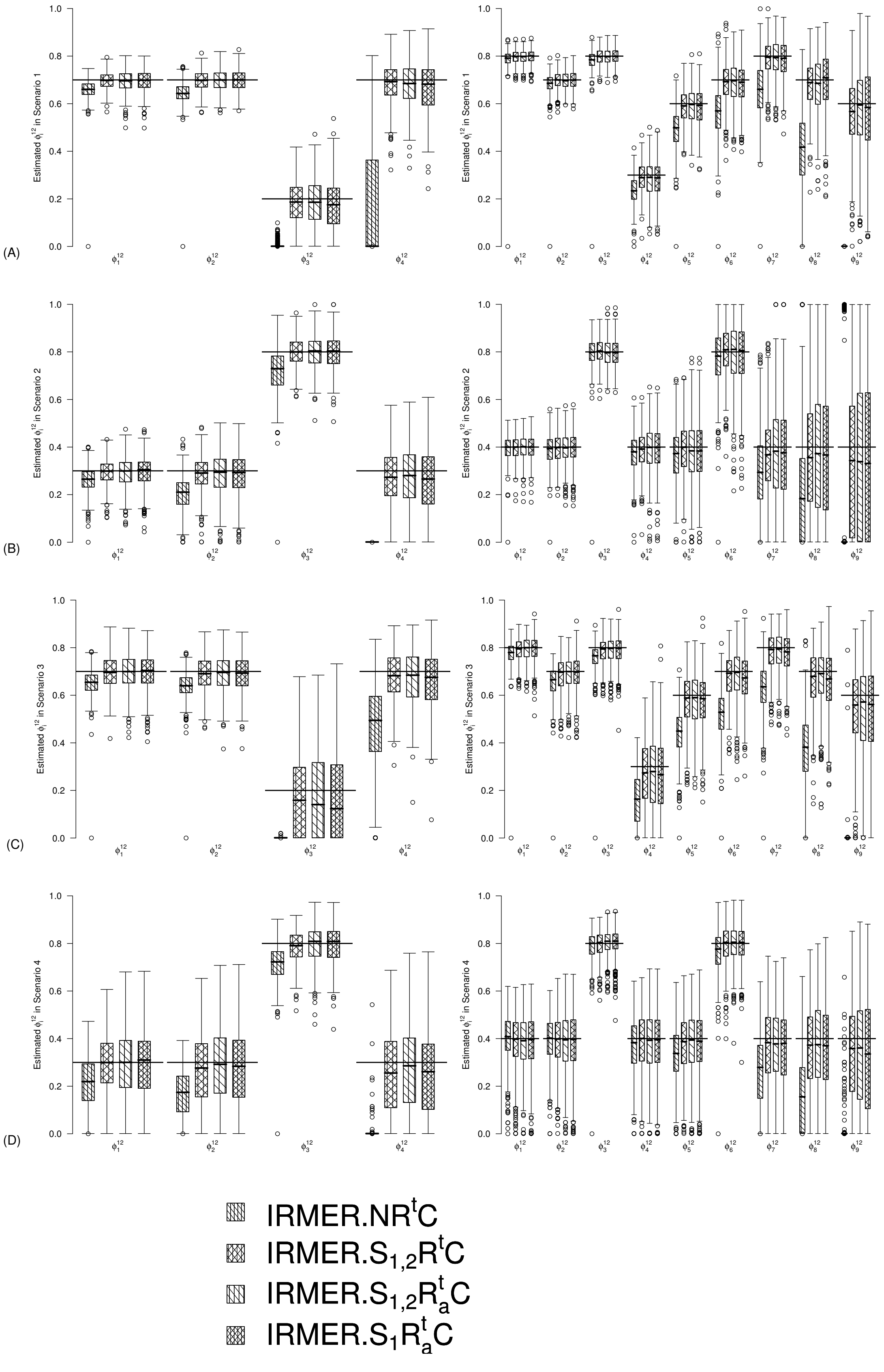


Fig. 3.12 Estimated transition probabilities ϕ_i^{12} for simulations with $K = 10$ and $K = 20$ sampling occasions displayed in the first and last columns respectively.

3.3.4 Results from the GRM model if ignoring temporary emigration

Setting	K	param	Scenario			
			Scenario 1	Scenario 2	Scenario 3	Scenario 4
3.3.1	10	N	3.850	43.127	2.915	10.979
	20	N	11.674	43.072	4.158	12.976
	10	p	−1.000	−0.997	−0.996	−0.994
	20	p	−0.997	−0.997	−0.997	−0.997
			Scenario 5	Scenario 6	Scenario 7	Scenario 8
3.3.2	10	N	56.918	26.973	9.069	4.984
	20	N	16.866	18.294	17.411	15.315
	10	p	−0.994	−1.000	−0.992	−0.998
	20	p	−0.996	−0.998	−0.996	−0.998
			Scenario 1	Scenario 2	Scenario 3	Scenario 4
3.3.3	10	N	3.894	3.990	3.186	3.866
	20	N	3.763	3.775	3.233	3.708
	10	M	9.728	9.226	4.714	6.531
	20	M	2.571	2.154	4.350	3.622
	10	p	−0.995	−0.994	−0.995	−0.995
	20	p	−0.996	−0.994	−0.995	−0.995

Table 3.1 Median relative bias in $\hat{N}$ ($\times 10^4$), $\hat{M}$ ($\times 10^4$) and $\hat{p}$ under each simulation setting. K is number of total sampling occasions.

If we ignored temporary immigration we would use the GRM model from Chapter 2. In this section we fit the GRM models to the data simulated under each Scenario. The median relative bias in the estimated population size and capture probability are presented in Table 3.1. We observe that under each simulation setting the GRM models overestimate population sizes and underestimate the capture probability for the simulated data exhibiting temporary emigration with the robust design sampling protocol.

3.4 Consider mortality in the RMER.SR^{t}C model

As permanent departure from the population is possible during the study, it may be useful to model mortality and temporary emigration simultaneously. In this section, we investigate whether we can estimate a constant survival probability ψ in the RMER.SRC model. We denote this new model by RMER.SR^t_dC as we consider an

Table 3.2 True parameters are listed in p_{true} and ψ_{true}. True transition probabilities used are the same (0.8,0.7,0.8,0.3,0.6,0.7,0.8,0.6,0.6) for every scenario.

Scenario	p_{true}	ψ_{true}
F1	0.3	0.95
F2	0.3	0.8
F3	0.3	0.4
G1	0.55	0.95
G2	0.55	0.8
G3	0.55	0.4

additional death state in the transition matrix, therefore there is a “d" in the subscript of R in the model notation.

We conducted a simulation study for a population of $N = 500$ individuals with $K = 20$ sampling occasions. We list the true values of parameters we used in Table 3.2. where ϕ_i^{12} are (0.8,0.7,0.8,0.3,0.6,0.7,0.8,0.6,0.6) for each of the scenarios considered, i.e. individuals are assumed to tend to stay off-site. Median relative bias are summarised in Table 3.3. The estimated population sizes N are negatively biased across all scenarios and the estimates of survival probability $\hat{\psi}$ are biased high apart from those under Scenario F1. The simulation results in Table 3.3 suggest the RMER.SR$_d^t$C model with a constant survival probability cannot be used in practice.

Therefore, when analysing the real data set in Section 3.5 we assume that all individuals survive as transolocation studies are usually conducted over a relatively short period of time (up to months). Improved estimates of survival probability can be obtained from removal data by collecting ancillary information during removal sampling, for example, concurrent capture-recapture sampling as suggested in Gould and Pollock (1997) or a few capture-recapture sampling occasions prior to removal sampling which is a design we could investigate in the future.

Table 3.3 Median relative bias (MBias) in estimates $\hat{p}$, $\hat{\psi}$ and $\hat{N}$ from the R-SR^t_dC model under different Scenarios. Simulations are conducted for $N = 500$ simulated individuals with $K = 20$ sampling occasions where there two secondary samples within each primary period.

Scenario	$\hat{p}_{\text{MBias}}$	$\hat{\psi}_{\text{MBias}}$	$\hat{N}_{\text{MBias}}$
F1	5.53×10^{-3}	-1.41×10^{-1}	-2.04×10^{-1}
F2	2.14×10^{-1}	3.81×10^{-1}	-5.08×10^{-1}
F3	2.72×10^{-1}	5.52×10^{-1}	-3.92×10^{-1}
G1	7.03×10^{-3}	0.83×10^{-1}	-1.28×10^{-1}
G2	1.97×10^{-2}	6.73×10^{-2}	-3.78×10^{-1}
G3	7.64×10^{-2}	1.12	-6.40×10^{-1}

3.5 Data analysis

Removal of common lizards, *Zootoca vivipara*, was conducted daily in both the morning and afternoon from the 13th of September to the 29th of October 2010. There were 94 sampling occasions, with 13 missed visits. 334 common lizards were captured and permanently removed from the study site over the course of the removal experiment. The removals consisted of 274 juvenile (denoted as “ju” in the subscripts in Tables 3.4 and 3.5) and 60 adult (denoted as “ad” in the subscripts in Tables 3.4 and 3.5) individuals. Eight covariates: mean air temperature, maximum air temperature, minimum air temperature, precipitation, average humidity, max humidity, min humidity and season stage, were recorded daily.

Migration and dispersal of reptiles and amphibians are generally limited during the daytime (Edgar et al., 2010), therefore we used a robust design approach for our analysis, with days corresponding to primary periods and the repeated samples within days being the secondary sampling occasions. Hence, there are $T = 47$ primary occasions and the number of secondary samples within each primary period are $k_1 = k_2 = \cdots = k_{47} = 2$. We note that the two secondary samples within each primary period share the same weather covariates due to the daily recorded schedule.

Given the nature of the available data, we assume juveniles and adults are sampled independently and hypothesize that their transition probabilities may be related. This is ecologically sensible since the dynamics exhibited by the population are likely to be driven by external influences. We use the theory of integrated population modelling (McCrea et al., 2010) and define the global likelihood to be the product of individual likelihoods as described in Section 3.2.2, i.e. $L = L_{\text{ju}}L_{\text{ad}}$ where L_{ju} and L_{ad} are the

Table 3.4 Model selection for models with time-varying transition probabilities fitted to common lizard data.

Model code	Model	Covariate	h	ML	ΔAIC
	IRMER model				
$\text{IRMER.S}_{\text{ju,ad}}\ \text{R}_a^t\text{Z}$	$\phi_{\text{ju}}^{12}(\text{t})$, $\phi_{\text{ad}}^{12}(\text{t+}\gamma_{\text{ad}})$, p(cov)	precipitation	51	−259.53	0
$\text{IRMER.S}_{\text{ju}}\ \text{R}_a^t\ \text{Z}$	π_{ad}, $\phi_{\text{ju}}^{12}(\text{t})$, $\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}})$, p(cov)	precipitation	52	−259.85	2.64
$\text{IRMER.S}_{\text{ju,ad}}\text{R}^t\ \text{Z}$	$\phi_{\text{ju}}^{12}(\text{t})$, p(cov)	precipitation	50	−261.94	2.82
$\text{IRMER.S}_{\text{ju}}\text{R}_a^t\text{Z}$	$\pi_{\text{ad}},\phi_{\text{ju}}^{12}(t),\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}}),p(\text{cov})$	average humidity	52	−262.04	7.02
$\text{IRMER.S}_{\text{ju,ad}}\text{R}_a^t\text{C}$	$\phi_{\text{ju}}^{12}(t),\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}}),p(c)$	-	50	−269.86	17.46
$\text{IRMER.S}_{\text{ju,ad}}\text{R}_a^t\text{Z}$	$\phi_{\text{ju}}^{12}(t),\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}}),p(\text{cov})$	mean air temperature	51	−268.98	18.90
$\text{IRMER.S}_{\text{ju}}\text{R}_a^t\text{Z}$	$\pi_{\text{ad}},\phi_{\text{ju}}^{12}(t),\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}}),p(\text{cov})$	max air temperature	52	−268.09	19.12
$\text{IRMER.S}_{\text{ju}}\text{R}_a^t\text{C}$	$\pi_{\text{ad}},\phi_{\text{ju}}^{12}(t),\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}}),p(c)$	-	51	−269.73	20.39
$\text{IRMER.S}_{\text{ju,ad}}\text{R}_a^t\text{Z}$	$\phi_{\text{ju}}^{12}(t),\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}}),p(\text{cov})$	max air temperature	51	−270.71	22.36
$\text{IRMER.S}_{\text{ju,ad}}\text{R}_a^t\text{Z}$	$\phi_{\text{ju}}^{12}(t),\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}}),p(\text{cov})$	min air temperature	51	−271.21	23.36
$\text{IRMER.S}_{\text{ju,ad}}\ \text{R}^t\ \text{C}$	$\phi_{\text{ju}}^{12}(\text{t})$, p($c$)	-	49	−273.27	23.48
$\text{IRMER.S}_{\text{ju}}\ \text{R}_a^t\ \text{Z}$	π_{ad}, $\phi_{\text{ju}}^{12}(\text{t})$, $\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}})$, p(cov)	max humidity	52	−274.03	31.00
$\text{IRMER.S}_{\text{ju,ad}}\text{R}^t\ \text{Z}$	$\phi_{\text{ju}}^{12}(\text{t})$, p(cov)	max humidity	50	−276.55	32.04
$\text{IRMER.S}_{\text{ju,ad}}\text{R}^t\ \text{Z}$	$\phi_{\text{ju}}^{12}(\text{t})$, p(cov)	min air temperature	50	−277.12	33.18
$\text{IRMER.S}_{\text{ju,ad}}\text{R}^t\ \text{Z}$	$\phi_{\text{ju}}^{12}(\text{t})$, p(cov)	max air temperature	50	−277.28	33.50
$\text{IRMER.S}_{\text{ju,ad}}\ \text{R}_a^t\text{Z}$	$\phi_{\text{ju}}^{12}(\text{t})$, $\phi_{\text{ad}}^{12}(\text{t+}\gamma_{\text{ad}})$, p(cov)	average humidity	51	−276.37	33.68
$\text{IRMER.S}_{\text{ju,ad}}\ \text{R}_a^t\text{Z}$	$\phi_{\text{ju}}^{12}(\text{t})$, $\phi_{\text{ad}}^{12}(\text{t+}\gamma_{\text{ad}})$, p(cov)	min humidity	51	−276.72	34.38
$\text{IRMER.S}_{\text{ju,ad}}\text{R}^t\ \text{Z}$	$\phi_{\text{ju}}^{12}(\text{t})$, p(cov)	min humidity	50	−276.75	34.44
$\text{IRMER.S}_{\text{ju}}\ \text{R}_a^t\ \text{Z}$	π_{ad}, $\phi_{\text{ju}}^{12}(\text{t})$, $\phi_{\text{ad}}^{12}(t+\gamma_{\text{ad}})$, p(cov)	min humidity	52	−278.67	40.28

likelihood for juvenile and adult populations respectively, and are both of the form described in Equation (3.1).

We also considered incorporating the climatic covariates using a logistic regression (North and Morgan, 1979) to account for the time variation exhibited within the transition and capture probabilities. The results from performing model selection on the integrated data are displayed in Tables 3.4 and 3.5 with $\phi^{12}(t)$ and $\phi^{12}(\text{cov})$ respectively.

All of the models ranked by AIC in Table 3.4 include fully time-dependent transition probabilities ϕ_i^{12}. The $\text{IRMER.S}_{\text{ju,ad}}\text{R}_a^t\text{Z}$ model with the smallest AIC value considers the additive effect on transition probabilities for the adult population (constraint "R_a^t"), a logistic regression for time-varying capture probabilities $p_{i,j}$ in terms of precipitation covariates (constraint "Z"), where juveniles and adults share the same $p_{i,j}$, and constraint "S" is used for the initial state parameters for both juveniles and adults. We computed standard errors and confidence intervals empirically using a non-parametric bootstrap method (Buckland, 1980; Buckland and Garthwaite, 1991). 500 simulated resamples are obtained based on the original counts of removed individuals. Results of data analysis from the top three models with the lowest AIC are presented in Table 3.7. The estimate of the number of individuals not captured is 57.74

Table 3.5 Model selection for models with time-varying transition probabilities in terms of covariates using logistic regression fitted to common lizard data. ΔAIC is the AIC difference compared with the model with lowest AIC in Table 3.4.

Model code	Model	Covariate	h	ML	ΔAIC
IRMER.$S_{\text{ju,ad}}Y_a^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	min humidity	6	−374.37	139.68
IRMER.$S_{\text{ju,ad}}Y_a^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	average humidity	6	−374.29	139.52
IRMER.$S_{\text{ju,ad}}Y_a^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	precipitation	7	−373.68	140.30
IRMER.$S_{\text{ju}}Y_a^t$ Z	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	precipitation	8	−372.98	140.90
IRMER.$S_{\text{ju,ad}}Y_a^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	precipitation	6	−375.00	140.94
IRMER.$S_{\text{ju}}Y_a^t$ C	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	average humidity	7	−374.12	141.18
IRMER.$S_{\text{ju}}Y_a^t$ C	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	min humidity	7	−374.22	141.38
IRMER.$S_{\text{ju}}Y_a^t$ Z	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	average humidity	8	−373.30	141.53
IRMER.$S_{\text{ju,ad}}Y_a^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	average humidity	7	−373.68	141.54
IRMER.$S_{\text{ju}}Y_a^t$ Z	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	min humidity	8	−375.77	141.80
IRMER.$S_{\text{ju,ad}}Z_a^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	min humidity	7	−374.44	141.82
IRMER.$S_{\text{ju}}Y_a^t$ C	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	precipitation	7	−374.88	142.70
IRMER.$S_{\text{ju,ad}}Y_a^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	max air temperature	7	−375.53	144.00
IRMER.$S_{\text{ju}}Y_a^t$ Z	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	max air temperature	8	−374.83	144.60
IRMER.$S_{\text{ju,ad}}Y_a^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	max humidity	6	−376.86	144.66
IRMER.$S_{\text{ju,ad}}Y_a^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	mean air temperature	6	−377.02	144.98
IRMER.$S_{\text{ju}}Y_a^t$ C	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	max air temperature	7	−376.23	145.40
IRMER.$S_{\text{ju,ad}}Y_a^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	mean air temperature	7	−376.54	146.02
IRMER.$S_{\text{ju,ad}}Z_a^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	max humidity	7	−376.56	146.06
IRMER.$S_{\text{ju}}Y_a^t$ C	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	max humidity	7	−376.75	146.44
IRMER.$S_{\text{ju}}Y_a^t$ Z	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	max humidity	8	−373.43	146.48
IRMER.$S_{\text{ju,ad}}Y^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(\text{cov})$	precipitation	6	−377.94	146.82
IRMER.$S_{\text{ju}}Y_a^t$ Z	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	mean air temperature	8	−375.89	146.72
IRMER.$S_{\text{ju}}Y_a^t$ C	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	mean air temperature	7	−376.93	146.80
IRMER.$S_{\text{ju,ad}}Y_a^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	min air temperature	6	−377.94	146.82
IRMER.$S_{\text{ju,ad}}Z^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(c)$	precipitation	5	−379.10	147.14
IRMER.$S_{\text{ju,ad}}Y_a^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	max air temperature	6	−378.42	147.78
IRMER.$S_{\text{ju,ad}}Y_a^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	min air temperature	7	−377.83	148.60
IRMER.$S_{\text{ju}}Y_a^t$ C	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(c)$	min air temperature	7	−377.84	148.62
IRMER.$S_{\text{ju}}Y_a^t$ Z	$\pi_{\text{ad}}, \phi_{\text{ju}}^{12}(\text{cov}), \phi_{\text{ad}}^{12}(\text{cov}+\gamma_{\text{ad}}), \text{p}(\text{cov})$	min air temperature	8	−377.17	149.28
IRMER.$S_{\text{ju,ad}}Y^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(c)$	min humidity	5	−382.06	153.06
IRMER.$S_{\text{ju,ad}}Y^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(c)$	average humidity	5	−382.34	153.62
IRMER.$S_{\text{ju,ad}}Y^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(\text{cov})$	min humidity	6	−384.08	154.34
IRMER.$S_{\text{ju,ad}}Y^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(\text{cov})$	average humidity	6	−382.33	155.60
IRMER.$S_{\text{ju,ad}}Y^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(c)$	max humidity	5	−384.47	157.88
IRMER.$S_{\text{ju,ad}}Y^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(\text{cov})$	max air temperature	6	−383.94	158.82
IRMER.$S_{\text{ju,ad}}Y^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(c)$	mean air temperature	5	−385.02	158.98
IRMER.$S_{\text{ju,ad}}Y^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(\text{cov})$	max humidity	6	−384.08	159.10
IRMER.$S_{\text{ju,ad}}Y^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(\text{cov})$	mean air temperature	6	−384.80	160.54
IRMER.$S_{\text{ju,ad}}Y^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(c)$	min air temperature	5	−385.94	160.82
IRMER.$S_{\text{ju,ad}}Y^t$ C	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(c)$	max air temperature	5	−386.05	161.04
IRMER.$S_{\text{ju,ad}}Y^t$ Z	$\phi_{\text{ju}}^{12}(\text{cov}), \text{p}(\text{cov})$	min air temperature	6	−385.68	162.30

Table 3.6 Model selection for geometric removal models (GRM) fitted to common lizard data. ΔAIC is the AIC difference compared with the model with lowest AIC in Table 3.4.

Model code	Model	Covariate	h	ML	ΔAIC
GRM.Z	p(cov)	max air temperature	4	-382.43	151.80
GRM.Z	p(cov)	mean air temperature	4	-385.06	157.06
GRM.Z	p(cov)	precipitation	4	-385.78	158.50
GRM.Z	p(cov)	average humidity	4	-386.96	160.86
GRM.Z	p(cov)	min air temperature	4	-387.83	162.60
GRM.Z	p(cov)	max humidity	4	-388.03	162.98
GRM.Z	p(cov)	min humidity	4	-388.03	163.00
GRM.C	p(c)	-	3	-390.05	163.84

(SE 146.02, 95% bootstrap CI 31.75, 473.33) for juveniles and 3.36 (SE 49.76, 95% bootstrap CI 0.02, 180.11) for adults. The precision of both of these is low, as a result of the sparseness of the observed data. The estimated additive effect of transition probabilities for adults is -0.97 (SE 1.54, 95% bootstrap CI -1.84, 2.75). The estimate of the intercept of the logistic regression for capture probability is -1.98 (SE 0.60, 95% bootstrap CI -3.60, -1.46). The estimate of the slope of the logistic regression for capture probability is 1.61 (SE 0.37, 95% bootstrap CI 0.98, 2.32). The estimates and 95% CI for the transition probabilities are available in Figure 3.14. Standard errors for the time-specific transition probabilities are large for some primary sessions. The poor precision is likely due to the small sample sizes.

We also considered the classic geometric removal model in Table 3.6, where GRM.C and GRM.Z represent the classic removal model with constant and time-varying capture probabilities in terms of covariates respectively for both juveniles and adults, where the same capture probability at each sampling occasion for both populations is assumed. Because the geometric removal model does not take into account transitions, we skip the notations for the constraints of initial state parameter and transition probability in the GRM model.

The estimate of the population size obtained by the GRM.Z model is 161.12 (SE 52, 95% bootstrap CI 51.91, 230.41) for juveniles and 34.90 (SE 18.41 , 95% bootstrap CI 0.01, 71.65) for adults. Furthermore, the estimated population size obtained by the GRM.C is 131.82 (SE 36.37, 95% bootstrap CI 83.07, 222.14) and 28.47 (SE 7.96, 95% bootstrap CI 17.80, 48.25) for juveniles and adults respectively. We observe that the GRM models have larger estimates of the population sizes for both of the populations.

Table 3.7 Estimates, standard errors (SE) and 95% confidence interval (CI) for the top three models as identified in Table 3.4.

Model	IRMER.$S_{\text{ju,ad}}$ R^t_aZ	IRMER.S_{ju} R^t_aZ	IRMER.$S_{\text{ju,ad}}$ R^tZ
ΔAIC	0	2.62	2.82
$\hat{n}_{0,ju}$	57.74	48.29	212.12
SE (95% CI)	146.02 (31.75, 474.33)	152.06 (31.23, 480.07)	100.93 (79.39, 458.75)
$\hat{n}_{0,ad}$	3.34	1.97	45.89
SE (95% CI)	49.76 (0.02, 180.11)	59.39 (0.04, 191.06)	28.87 (18.32, 110.09)
$\hat{\alpha}$	-1.98	-1.87	-3.03
SE (95% CI)	0.60 (-3.60, -1.46)	0.65 (-3.63 -1.47)	0.32 (-3.53, -2.32)
$\hat{\beta}$	1.61	1.547	0.31
SE (95% CI)	0.37 (0.98, 2.32)	0.47 (-0.09, 2.41)	0.290 (1.37, 2.64)
$\hat{\gamma_{ad}}$	-0.97	-1.059	-
SE (95% CI)	1.525 (-1.84, 2.74)	1.57 (-1.87, 2.78)	-
$\hat{\pi_{ad}}$	-	0.37	-
SE (95% CI)	-	0.51 (0.51, 1.00)	-

The fitted counts of individuals removed at each occasion for both juvenile and adult population are displayed in Figure 3.13. A visual assessment of observed and expected numbers based on the non-parametric bootstrap provided no evidence of systematic lack of fit of the selected model. Common lizards are diurnal and are known to spend less time basking on the surface as they can flexibly operate at low temperatures as described in Edgar et al. (2010). Furthermore, juveniles exhibit more powers of dispersal than adults as they can rapidly colonize new habitats which often become available adjacent to already occupied sites. These characteristics are supported by the results from our top model, suggesting that the transition probabilities of juveniles are higher than for adults. As seen from Table 3.4, none of the available climatic covariates collected during the study adequately accounted for the time-dependent transition probabilities. However, the logistic regression of time-varying capture probabilities in terms of precipitation is supported by our top model.

3.6 Discussion

Translocation mitigation projects undertaken prior to land development have become increasingly popular for protected species. Removal models have considerable potential to inform the design and execution of translocations, but need to take into account temporary emigration to reduce the risk of biased estimates of the number of animals not captured. In Chapter 3, we have extended the classic removal model to accommodate a robust design sampling strategy and multi-event framework with one unobservable state to allow for individuals to become temporarily unavailable for capture. Simulations have demonstrated that the RMER models perform better than the MER models with standard data sampling. Our approaches can yield unbiased estimates of the number of individuals in the populations residing in the sampling area. However, simulation is time-consuming if we want to examine all possible model structures. In Chapter 4 we use symbolic approaches to determine whether parameters in a model can be estimated individually or not.

The adequate design of sampling protocols is fundamental at the data analysis stage. Imperfect exposure to sampling for the population of interest is generally hard to deal with when analyzing the data. The robust design is a well-known sampling technique for capture recapture data that enables us to estimate otherwise confounded parameters (Kendall and Bjorkland, 2001; Kendall et al., 1997, 1995; Pollock, 1982). We have demonstrated that the use of the robust design for removal data enables the estimation of transition probabilities between observable and unobservable states. In addition, RMER models result in estimators of population size and capture probability which have better properties than MER under the standard sampling protocol. Therefore, we would like to raise the awareness of good study design for removal experiments as in our experience only a small number of removal studies have repeated samplings conducted within a day, However if sampling strategies were simply altered to allow for multiple secondary samples, uncertainty in estimates of detection and transition probabilities would reduce considerably.

The general RMER model with fully time-specific parameters is not identifiable, because there are more parameters than the number of removal data points. Although the assumption of constant parameters across time is the most straightforward way of constraining models in order to enable estimation, using simulations we have demonstrated that the best performing models with least bias incorporate at least two constraints - constraint “R^t” which denotes random emigration (the fully time-dependent transition probabilities are constrained by $\phi_i^{12} + \phi_i^{21} = 1$) and constraint “S” (the initial state parameter π is constrained as the first element of the mean of the stationary distributions of the transition matrices across time).

Our proposed RMER model is general and can be extended to the IRMER modelling which permits the analysis of multiple data sources, exploiting the relationship between parameters expected between related populations. Simulations have shown IRMER models perform well for two populations. Furthermore, we have applied the IRMER model to two age groups (adults and juveniles) of common lizard data and the results align with our understanding of the natural history of the populations of this species.

Further adaptations of the model could be considered, for example, transition probabilities may depend on discrete states such as sex, age or breeding status, and it might be of interest to consider how individual covariates could be incorporated into the model.

All of the models presented in Chapter 3 are based on the assumption of first-order Markovian transitions between the observable and unobservable state, such that the transition from primary period i to $i+1$ depends only on the state of the individual at the ith primary period. Future work includes the investigation of modelling temporary emigration as a second-order Markov chain for removal data building on the work presented by Kendall and Nichols (2002) where they explored the estimability of state transition probabilities in a second-order Markov process for capture-recapture data.

Spatial information has been widely used in the capture recapture literature (Royle et al., 2013), however, there is no spatial information on sampling available for the real data. Transolcation projects are generally poorly documented globally. In the UK, less than 10% of submitted reports contain detailed population monitoring data and one-half of the cases on file lack any type of report (Germano et al., 2015). In order to optimize the success of translocation studies, we should not only design the study properly, but also record any informative component which may help evaluate the sampling methodologies.

We have demonstrated that the estimation of temporary emigration for removal data relies on the use of the robust design. However, there are many removal data sets which cannot be analyzed using either the RMER or IRMER model due to the lack of repeated samples within a primary period during data collection. We investigate alternative ways of modelling removal data without the robust design in Chapter 5.

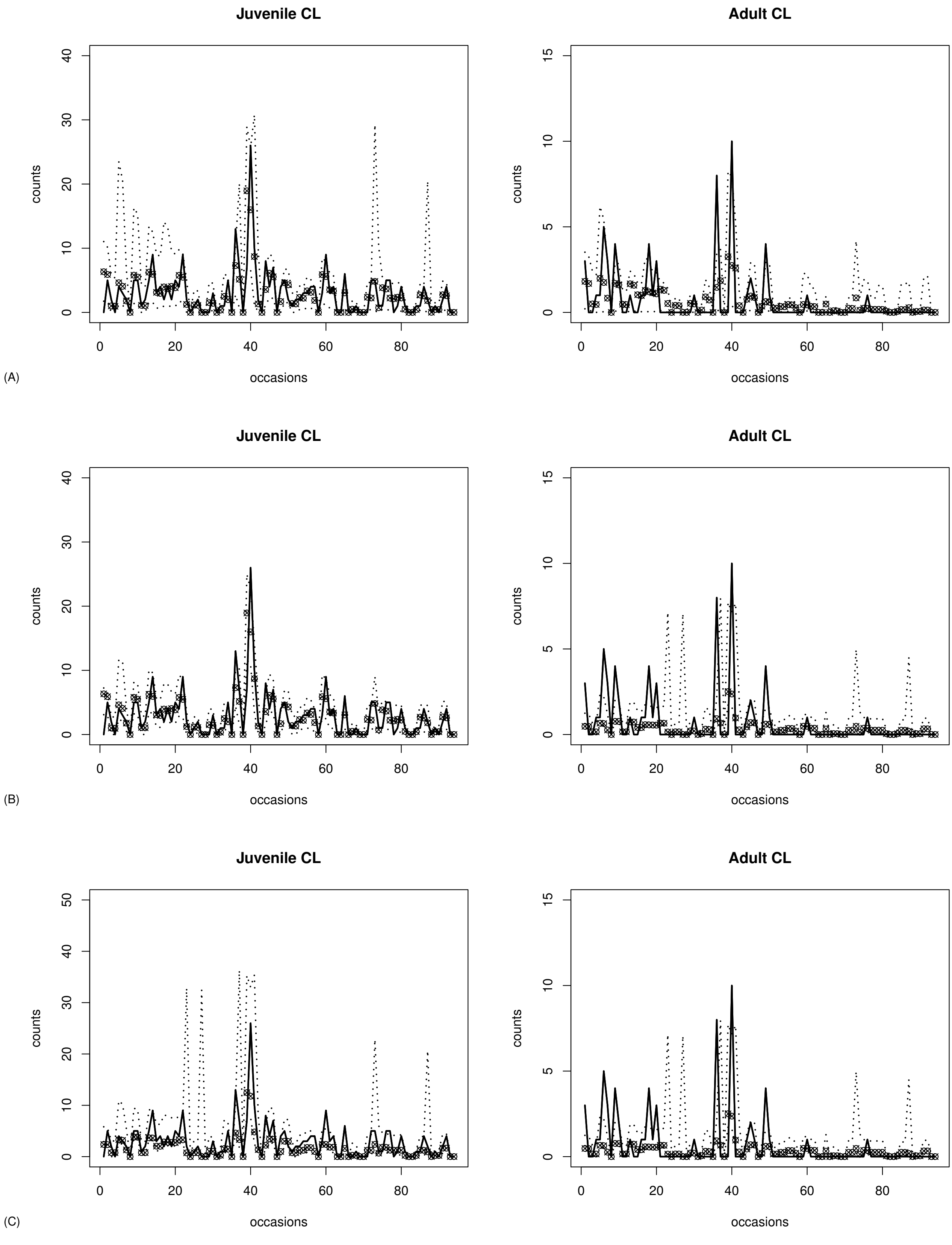


Fig. 3.13 Predicted counts of common lizards (CL) removed at each occasion, shown by the circle and cross dots. Black solid lines are the real data. 95% confidence intervals obtained from 500 bootstrap samples are indicated within dotted lines. (A) Results obtained from the IRMER.$S_{\mathrm{ju,ad}}R^t_aZ$. (B) Results obtained from the IRMER.$S_{\mathrm{ju}}R^t_aZ$ model with ΔAIC=2.62. (C) Results ordained from the IRMER.$S_{\mathrm{ju,ad}}R^tZ$ model with ΔAIC=2.82.

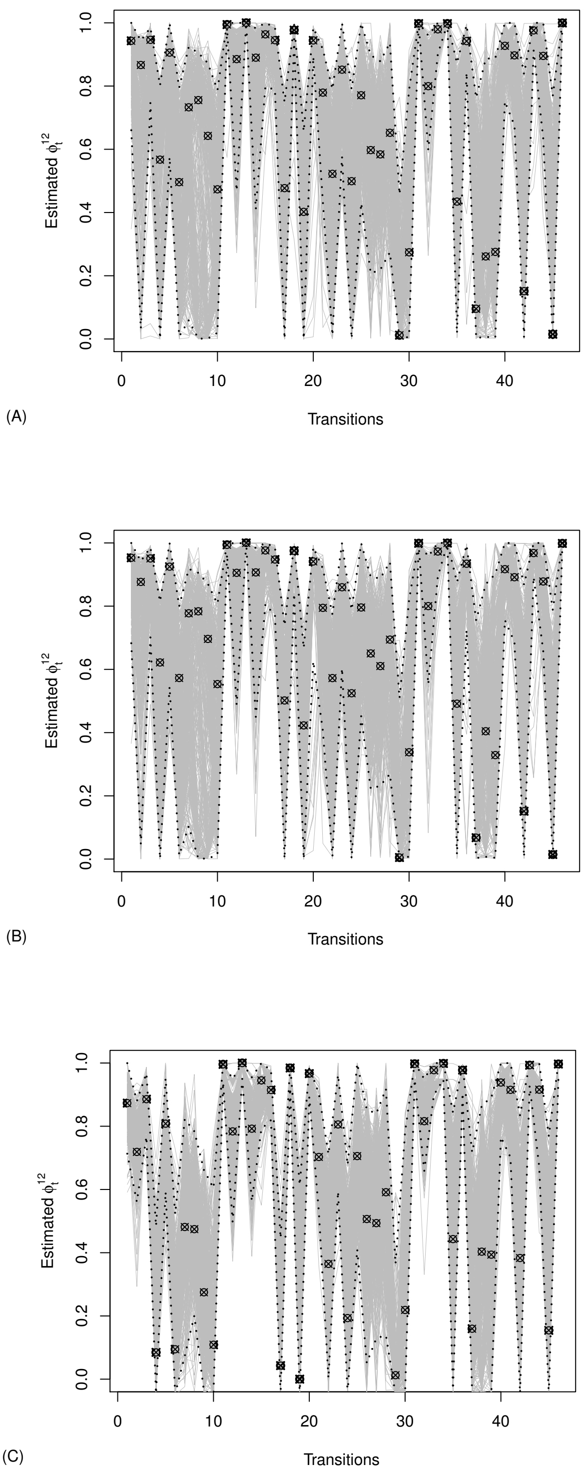


Fig. 3.14 Estimates of transition probabilities ϕ_i^{12} obtained from the top three models for common lizards, shown by the circle and cross dots. Grey lines are the estimated ϕ_i^{12} from 500 bootstrap samples. 95% confidence intervals are indicated within dotted lines. (A) Estimates of ϕ_i^{12} obtained from the IRMER.$\mathrm{S}_{\mathrm{ju,ad}}\mathrm{R}_a^t\mathrm{Z}$. (B) Estimates of ϕ_i^{12} obtained from the second top model, IRMER.$\mathrm{S}_{\mathrm{ju}}\mathrm{R}_a^t\mathrm{Z}$ model with ΔAIC=2.62. (C) Estimates of ϕ_i^{12} ordained from the IRMER.$\mathrm{S}_{\mathrm{ju,ad}}\mathrm{R}^t\mathrm{Z}$ model with ΔAIC=2.82.

Chapter 4

Parameter Redundancy for Removal Models

4.1 Introduction

As the complexity of ecological models increases, it is sometimes hard to identify whether we can estimate all the parameters in a model. Numerical simulation is a common way to investigate the performance of the estimates in a model (Gimenez et al., 2004). However, it can be time-consuming if we want to explore a diverse range of parameter spaces. So, efficient methods for examining parameter identifiability are in demand.

A model is termed parameter redundant if it is impossible to estimate all the parameters individually, because the model could be reparameterised in terms of a smaller number of parameters (Catchpole and Morgan, 1997). The techniques for detecting parameter redundancy have been developed for a wide range of applications including complex capture-recapture models (see, for example, Catchpole and Morgan, 1997; Cole et al., 2010, 2012; Cole, 2012; Cole et al., 2014; Hubbard et al., 2014). These provide an efficient and general way to assess whether parametric models are parameter redundant using symbolic algebra. In Chapter 4 we use these symbolic methods as tools to investigate the parameter redundancy status of removal models that are newly developed in this thesis. The work present in Chapters 3 and 4 are also available in Zhou et al. (2018). All the calculations in this chapter are carried out in `Maple` using the symbolic package `LinearAlgebra` and Maple procedures developed in Catchpole et al. (2002) and Cole et al. (2010). `Maple` code is available in the electronic appendix for all of the examples discussed.

We describe the methods for detecting parameter redundancy and demonstrate the use of these tools with illustrative examples for removal models in Section 4.2. From simulations in Chapter 3, we find that RMER and IRMER models employing the robust design framework and certain constraints can give unbiased estimates, however the estimates obtained from MER models without robust design are biased. In this Chapter, we explore the removal models presented in Chapter 3 symbolically. The results and their interpretation are summarised in Section 4.3. Chapter 4 finishes with a discussion in Section 4.4.

4.2 Methods for detecting parameter redundancy

4.2.1 Determining if a model is parameter redundant

Parameters are confounded in a parameter redundant model, so we cannot estimate all the parameters individually. Certain capture-recapture models can be formed based on a probability matrix that indicates probabilities of first capture and recaptures for each sampling occasion. Catchpole and Morgan (1997) showed that parameter redundancy in a capture-recapture model can be detected by forming a derivative matrix that is calculated by differentiating each of the non-zero elements in the probability matrix with respect to each parameter in the model. In this thesis we employ the methods of Cole et al. (2010) to assess whether a model is parameter redundant by forming a derivative matrix.

Cole et al. (2010) generalise the method of detecting parameter redundancy discussed in Catchpole and Morgan (1997) by forming the derivative matrix by differentiating an exhaustive summary with respect to the parameters. An exhaustive summary, $\kappa(\theta)$ is a vector of parameter combinations that uniquely represent the model. The probability matrix is an example exhaustive summary that could be used for a capture-recapture model. For removal models, a suitable exhaustive summary is a vector of probabilities of individuals being removed at the kth occasion, for $k = 1, \ldots, K$. The derivative matrix is then

$$\mathbf{D} = \left[\frac{\partial \kappa(\theta)}{\partial \theta} \right]$$

where θ is a vector of h parameters.

Once the matrix $\mathbf{D}$ is formed, the rank of the derivative matrix, r, which is the number of linearly independent rows of $\mathbf{D}$, is calculated. The deficiency of the model, d, is calculated as $h - r$. If $d > 0$, the model is parameter redundant. Otherwise, if

$d = 0$, then the model is termed full rank. For a full rank model, all the parameters can be estimated in principle.

`Maple` has a symbolic algebra computer package where we can obtain the rank of $\mathbf{D}$ symbolically (Catchpole et al., 2002). For example, we could load the `LinearAlgebra` package within `Maple` by calling the command `with(LinearAlgebra)`. In addition, the `Rank(D1)` function could calculate the rank of a derivative matrix $D1$ by performing Gaussian elimination on the rows of $D1$. However, there are situations where the symbolic approach fails because the terms in the derivative matrix are algebraically complicated or the number of terms is too large for the limited computational memory. As a result, the computation of the rank of the derivative matrix is not feasible for complex model structures or a large number of terms. Throughout the thesis the total number of elements in the exhaustive summary is no more than ten, as this is large enough to determine the structure of the model and small enough to compute the results algebraically for removal models considered.

Example 4.2.1 for the RMER.SR^{t}C model is shown to demonstrate the method for detecting parameter redundancy.

Example 4.2.1: RMER.SR^{t}C model

The RMER.SR^{t}C model, which incorporates both the constraint "S" and "R$_t$", is described in Chapter 3. Consider the model with $K = 8$ sampling occasions with 4 primary periods and 2 secondary samples, then the vector of parameters is $\theta = [p \quad \phi_1^{12} \quad \phi_2^{12} \quad \phi_3^{12}]$, where p represents the detection probability and ϕ_i^{12} denotes transition probability from state 1 to state 2 between the ith and $(i+1)$th primary period. The exhaustive summary is defined by a vector of probabilities of capturing individuals at the jth secondary period within the ith primary period.

$$\kappa(\theta) = \left[L_{i,j}\right] =$$

$$\begin{bmatrix} (1-\frac{1}{3}\phi_1^{12}-\frac{1}{3}\phi_2^{12}-\frac{1}{3}\phi_3^{12})p \\ (1-\frac{1}{3}\phi_1^{12}-\frac{1}{3}\phi_2^{12}-\frac{1}{3}\phi_3^{12})(1-p)p \\ \{(1-\frac{1}{3}\phi_1^{12}-\frac{1}{3}\phi_2^{12}-\frac{1}{3}\phi_3^{12})(1-p)^2(1-\phi_1^{12})+(\frac{1}{3}\phi_1^{12}-\frac{1}{3}\phi_2^{12}-\frac{1}{3}\phi_3^{12})(1-\phi_1^{12})\}p \\ \vdots \end{bmatrix}.$$

We remove the more complicated probability that an animal is never captured to make the exhaustive summary simpler. This is possible as we assume that we observe at least one individual at each occasion, and all the probabilities, including the probability that an animal is never captured, sum to one. Therefore, as the probability that an animal is never captured can be written as a function of the other exhaustive summary terms it can be removed. This is an identical argument for the one used in Catchpole and Morgan (1997) for capture-recapture models.

Then, the derivative matrix of the exhaustive summary for the parameters can be derived as

$$\mathbf{D} = \left[\frac{\partial \kappa(\theta)}{\partial \theta}\right] =$$

$$\begin{bmatrix} (1-\frac{1}{3}\phi_1^{12}-\frac{1}{3}\phi_2^{12}-\frac{1}{3}\phi_3^{12}) & (1-\frac{1}{3}\phi_1^{12}-\frac{1}{3}\phi_2^{12}-\frac{1}{3}\phi_3^{12})(1-2p) & \cdots \\ -\frac{1}{3}p & -\frac{1}{3}(1-p)p & \cdots \\ -\frac{1}{3}p & -\frac{1}{3}(1-p)p & \cdots \\ -\frac{1}{3}p & -\frac{1}{3}(1-p)p & \cdots \end{bmatrix}$$

Due to the complexity of terms in the exhaustive summary and the derivative matrix, we do not show the rest of the terms in this example, the full results are available in `Maple` file `ex4.2.1.mw` instead. We find that the rank of the derivative matrix is $r = 4$, which is equal to the number of parameters $h = 4$ in the model. We conclude that the RMER.SR^{t}C model has deficiency of zero as $d = h - r = 0$. Therefore the model is determined to be full rank when there are 8 sampling occasions. Hence, theoretically it is possible to estimate all the parameters individually.

4.2.2 Combinations of estimable parameters

As the derivative matrix is a unique representation of a model, it provides more information than the rank of $\mathbf{D}$, which only tells us the number of independent parameter combinations that can be estimated. For a parameter redundant model, it is possible to show if any of the original parameters are estimable.

Let us suppose we obtain a parameter redundant model with deficiency $d > 0$, then the derivative matrix $\mathbf{D}$ can be used to determine which original parameters and combinations of parameters are estimable (Catchpole et al., 1998; Cole et al., 2010). We can determine if any of the original parameters are independently estimable by solving for the left null space α of derivative matrix $\mathbf{D}$, i.e. $\alpha^T\mathbf{D} = 0$, for a vector $\alpha(\theta)$. As the model has a deficiency of d, there will be d linearly independent solutions to $\alpha^T\mathbf{D} = 0$. If we denote entries of the d solutions by $\alpha_{i,j}$, for $i = 1,\ldots,h$ and $j = 1,\ldots,d$, then for any i for which $\alpha_{i,j} = 0$ for all j, the ith parameter in the original parameter vector θ is estimable and is also the ith parameter which respect to which we are differentiating to form the matrix $\mathbf{D}$.

To find other estimable parameter combinations, we solve the system of linear first-order partial differential equations (Catchpole et al., 1998; Chappell and Gunn, 1998; Evans and Chappell, 2000; Cole et al., 2010),

$$\sum_{i=1}^{h} \alpha_{ij} \frac{\partial f}{\partial \theta_i} = 0, \quad j = 1, \ldots, d$$

which are known as Lagrange equations (Cox and Miller, 1965).

We demonstrate how this method works using the illustrative Example 4.2.2 below.

Example 4.2.2: MER.NNC model

The MER.NNC model in Chapter 3 considers a multievent framework allowing for temporary emigration but does not have a robust design structure. It contains a constant detection probability, constant transition probabilities and an initial state parameter. Therefore, the parameters in this case are $\theta = [p \quad \phi^{12} \quad \phi^{21} \quad \pi]$, where p is the detection probability, ϕ^{12} and ϕ^{21} represent transition probabilities between state 1 and state 2, and π is the probability of being in state 1 initially. We consider the model with 8 occasions of removal, and we define the exhaustive summary as a vector of probabilities of an individual being captured at the the kth occasion as the specification determines the model. We only show the first three terms in the probability matrix below as the probabilities of removal dramatically get more complicated as the number of sampling times increases. Full results are available in the `Maple` code in the file `ex4.2.2.mw`.

$$\kappa(\theta) = \left[L_{k,1}\right] =$$

$$\begin{bmatrix} \pi p \\ \{\pi(1-p)(1-\phi^{12}) + (1-\pi)\phi^{21}\}p \\ \left[\{\pi(1-p)(1-\phi^{12}) + (1-\pi)\phi^{12}\}(1-p)(1-\phi^{12}) + \{\pi(1-p)\phi^{12} + (1-\pi)(1-\phi^{21}\}\phi^{21})\right]p \\ \vdots \end{bmatrix} \tag{4.1}$$

where $L_{k,1}$ is the probability of animals being captured at the kth sampling occasion.

The derivative matrix is given by,

$$\mathbf{D} = \left[\frac{\partial \kappa(\theta)}{\partial \theta}\right] = \begin{bmatrix} \pi & -\pi(1-\phi^{12})p + \pi(1-p)(1-\phi^{12}) + (1-\pi)\phi^{21} & \cdots \\ 0 & -\pi(1-p)p & \cdots \\ 0 & (1-\pi)p & \cdots \\ p & \{(1-p)(1-\phi^{12}) - \phi^{21}\}p & \cdots \end{bmatrix}.$$

and is found to have symbolic rank 3. However, there are 4 parameters, therefore the model is parameter redundant with deficiency 1. In order to find if any of the original parameter can be estimated, we solve $\alpha^T \mathbf{D} = 0$, which gives,

$$\alpha = \begin{bmatrix} -\frac{p}{\pi} & \frac{p\phi^{12} - p + \phi^{21}}{\pi(p-1)} & \frac{\phi^{21}}{\pi} & 1 \end{bmatrix}.$$

As there are no zeros in α, none of the original parameters are individually estimable. To find the estimable parameter combinations, we solve the partial differential equation,

$$-\frac{\partial f}{\partial p}\frac{p}{\pi} + \frac{\partial f}{\partial \phi^{12}}\frac{p\phi^{12} - p + \phi^{21}}{\pi(p-1)} + \frac{\partial f}{\partial \phi^{21}}\frac{\phi^{21}}{\pi} + \frac{\partial f}{\partial \pi} = 0,$$

which shows we can estimate $p\phi^{21}$, $p\pi$, and $p(\phi^{12} - 1) - \phi^{12} - \phi^{21}$.

4.2.3 Generalization of parameter redundancy results

In Sections 4.2.1 and 4.2.2, we demonstrate how to determine the results of parameter redundancy for a fixed number of K sampling occasions. It is possible to generalize the results of parameter redundancy to any number of samples as described in Catchpole and Morgan (1997) and Cole et al. (2010). In this section, we show how to obtain general parameter redundancy results for our removal models with two secondary occasions within each primary period. Note that we could theoretically generalize the results to any number of secondary occasions. We only show two secondary occasions here as this matches the data used in this thesis.

Extension theorem

Parameter redundancy results can be generalised to any dimension of model using the extension theorem, first proposed by Catchpole and Morgan (1997) and extended to any exhaustive summary in Cole et al. (2010).

Suppose we have a full rank model for a fixed total number of sampling occasions, in which the exhaustive summary is denoted by $\kappa_1(\theta_1)$ with parameters θ_1 and the derivative matrix is $\mathbf{D}_{1,1} = [\partial \kappa_1(\theta_1)/\partial \theta_1]$. The model is then extended, by adding an extra sampling occasion, to give an exhaustive summary $\kappa = [\kappa_1, \kappa_2]$. We denote the extra parameters (if any) by θ_2 and let $\theta = [\theta_1, \theta_2]$. The extra exhaustive summary terms are $\kappa_2(\theta)$. Furthermore, the new derivative matrix of the extended model is formed by,

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{1,1}(\theta_1) & \mathbf{D}_{2,1}(\theta_1) \\ \mathbf{D}_{1,2}(\theta_2) & \mathbf{D}_{2,2}(\theta_2) \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{1,1}(\theta_1) & \mathbf{D}_{2,1}(\theta_1) \\ 0 & \mathbf{D}_{2,2}(\theta_2) \end{bmatrix},$$

where $\mathbf{D}_{1,2}(\theta_2) = [\partial \kappa_1 / \partial \theta_2] = 0$, $\mathbf{D}_{2,1}(\theta_1) = [\partial \kappa_2 / \partial \theta_1]$ and $\mathbf{D}_{2,2}(\theta_2) = [\partial \kappa_2 / \partial \theta_2]$.

Catchpole and Morgan (1997) proved that if both $\mathbf{D}_{1,1}(\theta_1)$ and $\mathbf{D}_{2,2}(\theta_2)$ are full rank, then the extended model is also full rank. Induction can be used to generalise this result to any number of sampling occasions.

If there is no extra parameter of the extended model, then the new derivative matrix can be further simplified to,

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{1,1}(\theta_1) & \mathbf{D}_{2,1}(\theta_1) \\ 0 & \mathbf{D}_{2,2}(\theta_2) \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{1,1}(\theta_1) & \mathbf{D}_{2,1}(\theta_1) \end{bmatrix},$$

which is always full rank if $\mathbf{D}_{1,1}(\theta_1)$ is full rank.

Furthermore, if there is only one extra parameter in the extended model, then $\mathbf{D}_{2,2}(\theta_2)$ is always full rank because it always has rank of one as a row vector. Therefore the extended model will be full rank if $\mathbf{D}_{1,1}(\theta_1)$ is full rank and there is only one new parameter in θ_2. We demonstrate the use of the extension theorem using the example of RMER.SR^{t}C model as below.

Example 4.2.1 continued: RMER.SR^{t}C model

Step 1: prove the model is full rank for a certain number of sampling occasions.

This step is done in Example 4.2.1. We have demonstrated the RMER.SR^{t}C model is not parameter redundant when there are $K = 8$ sampling occasions with 4 primary periods and 2 secondary samples. Let κ_1 denote the original exhaustive summary and $\theta_1 = [p \quad \phi_1^{12} \quad \phi_2^{12} \quad \phi_3^{12}]$ for the model with $K = 8$ sampling occasions. Furthermore, the derivative matrix $D_{1,1}(\theta_1) = [\partial \kappa_1(\theta_1) / \partial \theta_1]$ is found to be full rank in Section 4.2.1.

Step 2: extend the results in Step 1 to any number of occasions.

Because the derivative matrix $D_{1,1}(\theta_1)$ is full rank, we can then apply the extension theorem to prove this model is full rank for any number of sampling occasions $K \geq 4$.

The model can be extended by considering one additional primary period with 2 secondary occasions. Let $\boldsymbol{\theta_2} = [\phi_4^{12}]$ denote the extra parameter and $\boldsymbol{\kappa_2}(\boldsymbol{\theta_2})$ denote the extra terms in the exhaustive summary. The full set of parameters after the model extension is $\theta = [\boldsymbol{\theta_1} \quad \boldsymbol{\theta_2}] = [p \quad \phi_1^{12} \quad \phi_2^{12} \quad \phi_3^{12} \quad \phi_4^{12}]$. The exhaustive summary of this extended model is denoted by $\kappa(\theta) = [\boldsymbol{\kappa_1}(\boldsymbol{\theta_1}) \quad \boldsymbol{\kappa_2}(\boldsymbol{\theta_2})]$. The extension theorem states that if both $\mathrm{D}_{1,1}(\boldsymbol{\theta_1})$ and $\mathrm{D}_{2,2}(\boldsymbol{\theta_2})$ are full rank, we can explicitly conclude that the extended model is also full rank for any larger number of sampling occasions. Due to the complexity of $\boldsymbol{\kappa_2}(\boldsymbol{\theta_2})$ and $\mathrm{D}_{2,2}(\boldsymbol{\theta_2})$, we don't show their exact forms, the detailed expressions are available in the `Maple` file `4.2.1.cont.mw`. We find $\mathrm{D}_{2,2}(\boldsymbol{\theta_2})$ has rank 1 which is full rank as there is only 1 parameter in $\boldsymbol{\theta_2}$. Therefore, the extended model is proved to be explicitly full rank for any $K \geq 10$ sampling occasions.

Reparameterisation theorem

If the original model is not full rank, then the extension theorem cannot be used for that model. Instead, we need to find a reparameterization of the original model that is full rank to begin with, then we can apply the extension theorem to the reparameterized model to determine the appropriate general parameter redundancy result (Cole et al., 2010). We reparameterize our models in terms of the estimable parameter combinations. We demonstrate how to use the reparameterization theorem for the model RMER.NN^{t}C in the example 4.2.3 below.

Example 4.2.3: RMER.NN^{t}C model

The RMER.NN^{t}C model does not consider any constraint for time-dependent transition probabilities and initial state, as described in Chapter 3. In this example we demonstrate how to use the reparametersation theorem to show that the deficiency for the RMER.NNtC model is $K/2-1$ for $K \geq 4$. `Maple` code is in the `ex4.2.3.mw` file.

Step 1: prove the model is parameter redundant

Consider the RMER.NN^{t}C model with time-varying transition probabilities for $K = 4$ sampling occasions where there are 2 primary periods and 2 secondary samples.

The vector of parameters is $\boldsymbol{\theta_1} = [p \quad \pi \quad \phi_1^{12} \quad \phi_1^{21}]$. The exhaustive summary is

$$\kappa_{\mathbf{1}}(\boldsymbol{\theta_1}) = \left[L_{i,j}\right] = \begin{bmatrix} \pi p \\ \pi(1-p)p \\ \{\pi(1-p)^2(1-\phi_1^{12}) + (1-\pi)\phi_1^{21}\}p \\ \{\pi(1-p)^2(1-\phi_1^{12}) + (1-\pi)\phi_1^{21}\}(1-p)p \end{bmatrix}$$

where $L_{i,j}$ is the probability of individuals being removed at the jth secondary occasion within the ith primary period. The derivative matrix is given as

$$\mathbf{D}_{1,1}(\boldsymbol{\theta_1}) = \left[\frac{\partial \kappa_{\mathbf{1}}(\boldsymbol{\theta_1})}{\partial \boldsymbol{\theta_1}}\right] =$$

$$\begin{bmatrix} \pi & -\pi p + \pi(1-p) & -2\pi(1-p)(1-\phi_1^{12})p + \pi(1-p)^2(1-\phi_1^{12}) + (1-\pi)\phi_1^{21} & \cdots \\ p & (1-p)p & \{(1-p)^2(1-\phi_1^{12}) - \phi_1^{21}\}p & \cdots \\ 0 & 0 & -\pi(1-p)^2 p & \cdots \\ 0 & 0 & (1-\pi)p & \cdots \end{bmatrix},$$

which has rank 3. As there are 4 parameters in $\boldsymbol{\theta_1}$, the model is parameter redundant with deficiency of 1. To find if any of the parameters can be estimated individually and which combinations of parameters are estimable, we solve $\alpha^T \mathbf{D}_{1,1} = 0$, which gives

$$\alpha = \begin{bmatrix} 0 & 0 & -\dfrac{\pi - 1}{\pi(-1+p)^2} & 1 \end{bmatrix}.$$

We find there are two zeros in α which correspond to the first and second parameters in $\boldsymbol{\theta_1}$, i.e p and π. So in theory p and π are individually estimable. In order to find the other estimable combinations of parameters, we solve the following partial differential equation,

$$-\frac{\partial f}{\partial \phi_1^{12}} \frac{\pi - 1}{\pi(-1+p)^2} + \frac{\partial f}{\partial \phi_1^{21}} = 0.$$

Its solution shows that we can estimate $[\{(-1+p)^2\phi_1^{12} + \phi_1^{21}\}\pi - \phi_1^{21}]/(\pi - 1)$. Therefore, p, π and $[\{(-1+p)^2\phi_1^{12} + \phi_1^{21}\}\pi - \phi_1^{21}]/(\pi - 1)$ are the estimable parameter combinations when there are $K = 4$ sampling occasions.

Step 2: Reparameterise the model

We have proved that the model with 4 parameters is parameter redundant with deficiency of 1 for $K = 4$ occasions. In order to obtain its general results of parameter redundancy, we need to find a reparameterisation of the model which is full rank before we apply the extension theorem to determine the parameter redundancy for general values of K. Let $\mathbf{s}$ be a vector of three combinations of original parameters in the reparameterised model as shown below.

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 \end{bmatrix} = \begin{bmatrix} p \\ \pi \\ \{\pi(1-p)^2(1-\phi_1^{12}) + (1-\pi)\phi_1^{21}\} \end{bmatrix}.$$

We can then reparameterise the original model in terms of reparameterised parameters $\mathbf{s} = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \mathbf{s}_3]$ and the reparameterised $\kappa_1(\mathbf{s})$ is

$$\kappa_1(\mathbf{s}) = \begin{bmatrix} \mathbf{s}_1\mathbf{s}_2 \\ \mathbf{s}_2(1-\mathbf{s}_1)\mathbf{s}_1 \\ \mathbf{s}_1\mathbf{s}_3 \\ \mathbf{s}_1(1-\mathbf{s}_1)\mathbf{s}_3 \end{bmatrix}.$$

The derivative matrix $\mathbf{D}_{1,1}(\mathbf{s}) = \partial\kappa_1/\partial\mathbf{s}$ for the reparameterised model can be derived as matrix (4.2) with rank 3. Therefore the reparameterised model is full rank as there are three parameters in the model.

$$\mathbf{D}_{1,1}(\mathbf{s}) = \begin{bmatrix} \mathbf{s}_2 & -\mathbf{s}_1\mathbf{s}_2 + \mathbf{s}_2(1-\mathbf{s}_1) & \mathbf{s}_3 & (1-\mathbf{s}_1)\mathbf{s}_3 - \mathbf{s}_1\mathbf{s}_3 \\ \mathbf{s}_1 & (1-\mathbf{s}_1)\mathbf{s}_1 & 0 & 0 \\ 0 & 0 & \mathbf{s}_1 & (1-\mathbf{s}_1)\mathbf{s}_1 \end{bmatrix} \tag{4.2}$$

Step 3: Use extension theorem on the reparameterised model

As we have derived a full rank reparameterisation of the original model in Step 2, we then could apply the extension theorem for the reparameterised model. If we consider one additional primary period with 2 secondary occasions to the model, we will have two new parameters $\mathbf{s}' = \begin{bmatrix}\phi_2^{12} & \phi_2^{21}\end{bmatrix}$. There will be two terms in the extended part of exhaustive summary after adding one primary period with two secondary samples. The extended exhaustive summary $\kappa_2(\mathbf{s}')$ in terms of the new parameter $\mathbf{s}'$ is shown below,

$$\kappa_2(\mathbf{s}') = \begin{bmatrix} \mathbf{s}_1[\{(1-\phi_2^{12})\mathbf{s}_3 + \mathbf{s}_2\phi_2^{21}\}\mathbf{s}_1^2 + \{(2\phi_2^{12}-2)\mathbf{s}_3 - 2\mathbf{s}_2\phi_2^{21}\}\mathbf{s}_1 + (-\phi_2^{12} - \phi_2^{21} + 1)\mathbf{s}_3 + \phi_2^{21}] \\ \mathbf{s}_1(1-\mathbf{s}_1)[\{(1-\phi_2^{12})\mathbf{s}_3 + \mathbf{s}_2\phi_2^{21}\}\mathbf{s}_1^2 + \{(2\phi_2^{12}-2)\mathbf{s}_3 - 2\mathbf{s}_2\phi_2^{21}\}\mathbf{s}_1 + (-\phi_2^{12} - \phi_2^{21} + 1)\mathbf{s}_3 + \phi_2^{21}] \end{bmatrix}$$

We can then obtain $\mathbf{D}_{2,2}(\mathbf{s}^{'})$ by differentiating $\kappa_{\mathbf{2}}(\mathbf{s}^{'})$ with respect to $\mathbf{s}^{'} = [\phi_2^{12} \quad \phi_2^{21}]$ as below,

$$\mathbf{D}_{2,2}(\mathbf{s}^{'}) = \left[\begin{array}{cc} \mathbf{s_1}(-\mathbf{s_1^2}\mathbf{s_3} + 2\mathbf{s_1}\mathbf{s_3} - \mathbf{s_3}) & \mathbf{s_1}(1 - \mathbf{s_1})(-\mathbf{s_1^2}\mathbf{s_3} + 2\mathbf{s_1}\mathbf{s_3} - \mathbf{s_3}) \\ \mathbf{s_1}(\mathbf{s_1^2}\mathbf{s_2} - 2\mathbf{s_1}\mathbf{s_2} - \mathbf{s_3} + 1) & \mathbf{s} - \mathbf{s_1}(-1 + \mathbf{s_1})(\mathbf{s_1^2}\mathbf{s_2} - 2\mathbf{s_1}\mathbf{s_2} - \mathbf{s_3} + 1) \end{array} \right].$$

$\mathbf{D}_{2,2}(\mathbf{s}^{'})$ has a rank of 1. As there are two parameters in $\mathbf{s}^{'}$, the extended part of the exhaustive summary has a deficiency of 1 when we consider an additional primary period with two secondary samples.

For the RMER.NN^{t}C model with 2 secondary samples within each primary occasion, the deficiency of the extended terms in the exhaustive summary is always equal to the extra number of primary periods considered. This is because when we consider i additional primary periods, there will always be $2i$ new transition probabilities ϕ_i^{12} and ϕ_i^{21} in the extended model, the rank of $\mathbf{D}_{2,2}$ is always equal to half of the number of new parameters, i.e. the number of additional primary periods, so the deficiency of the extended part is always equal to the number of extra primary periods as well. See `Maple` file `ex4.2.3.mw` for proof.

We have proved that the model with $K = 4$ sampling occasions has rank 3 and deficiency of 1 in Step 1, then for a removal study of length K with two secondary samples within each primary period the extended part of the study has length of $K - 4$ with $(K-4)/2$ additional primary periods and $K - 4$ new transition probabilities. Therefore the rank of $\mathbf{D}_{2,2}$ is equal to $(K-4)/2 = K/2 - 2$ and the deficiency is equal to $(K-4) - (K-4)/2 = K/2 - 2$ as well. Bringing together the results in Step 1, the model with K sampling occasions will have rank of $(K/2 - 2) + 3 = K/2 + 1$ and deficiency of $(K/2 - 2) + 1 = K/2 - 1$.

Note that for a removal study with K sampling occasions and two secondary samples within each primary period, there will be $K/2 - 1$ primary periods, $K - 2$ transition probabilities for the RMER.NN^{t}C model, one constant capture probability p and one initial state parameter π, therefore there are a total of K parameters in the model. Hence, the deficiency can be calculated as $K - (K/2 + 1) = K/2 - 1$, where $K/2 + 1$ is the rank as derived before. Note that results only apply for $K \geq 4$, because the deficiency cannot be less than zero, i.e. $K/2 - 1 \geq 0$

Therefore, the deficiency of the model can be derived explicitly for any number of K sampling occasions where there are 2 secondary samples within each primary period.

4.2.4 The parameter n_0

In the exhaustive summaries considered so far we have not included any terms involving the parameter n_0, i.e. the number of individual never captured. This parameter appears in the multinomial constant of the full likelihood as shown in Equation (3.1) in Section 3.2.2. That term has been excluded from the exhaustive summary, because it only provides information on n_0 so n_0 can always be estimated in theory if other parameters can be estimated. We demonstrate this in the following example, where we consider n_0 as an extra parameter for the RMER.SR^{t}C discussed in the Example 4.2.1.

Note that the parameter n_0 also appears in the term $L_0^{n_0}$ in Equation (3.1). However, we never consider the probability that an individual is not captured in the exhaustive summary because it can be rewritten as a function of other exhaustive summary terms as mentioned in Example 4.2.1.

Example 4.2.4: consider n_0 in the RMER.SR^{t}C model

Similar to the Example 4.2.1, we still consider the model with $K = 8$ sampling occasions with 4 primary periods and 2 secondary samples. As we consider n_o as an additional parameter, the vector of parameters in this case is $\theta = [n_0 \quad p \quad \phi_1^{12} \quad \phi_2^{12} \quad \phi_3^{12}]$. We then include $n_0 + n_{1,1} + \cdots + n_{4,2}$ as an extra term in the exhaustive summary of a vector of probabilities of capturing individuals at the jth secondary period within the ith primary period. Including the term $n_0 + n_{1,1} + \cdots + n_{4,2}$ is equivalent to consider the coefficient $N! = (n_0 + \sum_{i=1,j=1}^{i=4,j=2} n_{i,j})!$ in the full multinomial likelihood function. Then, the exhaustive summary in this case is,

$$\kappa(\theta) = \left[L_{i,j}\right] =$$

$$\begin{bmatrix} n_0 + n_{1,1} + n_{1,2} + n_{2,1} + n_{2,2} + n_{3,1} + n_{3,2} + n_{4,1} + n_{4,2} \\ (1 - \frac{1}{3}\phi_1^{12} - \frac{1}{3}\phi_2^{12} - \frac{1}{3}\phi_3^{12})p \\ (1 - \frac{1}{3}\phi_1^{12} - \frac{1}{3}\phi_2^{12} - \frac{1}{3}\phi_3^{12})(1-p)p \\ \{(1 - \frac{1}{3}\phi_1^{12} - \frac{1}{3}\phi_2^{12} - \frac{1}{3}\phi_3^{12})(1-p)^2(1-\phi_1^{12}) + (\frac{1}{3}\phi_1^{12} - \frac{1}{3}\phi_2^{12} - \frac{1}{3}\phi_3^{12})(1-\phi_1^{12})\}p \\ \vdots \end{bmatrix}.$$

The derivative matrix of the above exhaustive summary is derived as

$$\mathbf{D} = \left[\frac{\partial \kappa(\theta)}{\partial \theta}\right] =$$

$$\begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & (1-\frac{1}{3}\phi_1^{12}-\frac{1}{3}\phi_2^{12}-\frac{1}{3}\phi_3^{12}) & (1-\frac{1}{3}\phi_1^{12}-\frac{1}{3}\phi_2^{12}-\frac{1}{3}\phi_3^{12})(1-2p) & \cdots \\ 0 & -\frac{1}{3}p & -\frac{1}{3}(1-p)p & \cdots \\ 0 & -\frac{1}{3}p & -\frac{1}{3}(1-p)p & \cdots \\ 0 & -\frac{1}{3}p & -\frac{1}{3}(1-p)p & \cdots \end{bmatrix}$$

The rank of this derivative matrix is $r = 5$, which is equal to the number of parameters $h = 5$. Therefore the model has deficiency of zero and is determined to be full rank. The full results are available in `Maple` file `ex4.2.4.mw`.

Note that the first row vector in **D** is (1,0,0,0,0,0,0,0,0), and it contributes to one rank for the additional n_0 parameter, so the deficiency of the model is the same with that in the Example 4.2.1. Therefore we consider a simpler exhaustive summary without the term $N! = (n_0 + \sum_{i=1,j=1}^{i=4,j=2} n_{i,j})!$ and without n_0 in the parameter vector throughout the thesis.

4.2.5 Near parameter redundancy

Catchpole et al. (2001) found that a full rank model can perform badly in practice, because the model is close to a nested parameter redundant model; this is known as near redundancy. It is useful to look at the values of the smallest eigenvalue of the expected information matrix for the parameter space as shown in Catchpole and Morgan (1997). Similar methods were used for examining practical identifiability of mathematical dynamic models (Raue et al., 2009; Chis et al., 2016).

In a parameter redundant model the expected information matrix will be exactly singular (Rothenberg, 1971), which occurs if and only if the rank of **D** is less than the number of original parameters h, which can be tested using symbolic algebra as described in Section 4.2.1. As a result, the expected information matrix will have at least one zero eigenvalue. In a near parameter-redundant model the smallest eigenvalue will be close to zero rather than exactly zero (Catchpole et al., 2001).

Chis et al. (2016) suggested to use a cut-off point of 0.001 in the parameter sloppiness, which can be calculated as the smallest eigenvalue divided by the largest eigenvalue of the Hessian of the log-likelihood function. The same threshold of 0.001 for the standardised smallest eigenvalue in the expected information matrix is also suggested in the concept of near parameter redundancy, i.e. a model is near parameter redundant if its standardised smallest eigenvalue $< 10^{-3}$.

We discussed the results of near parameter redundancy for removal models in Chapter 3 in Section 4.3. We are interested in looking at the standardised smallest eigenvalue under different scenarios. We observe that for models we discussed in the thesis the cut-off of 0.001 for the standardised smallest eigenvalue cannot be used, and we recommend to compare those eigenvalues obtained from models under the same scenario to avoid misleading conclusions.

4.3 Results

In this section we present general parameter redundancy results for removal models presented in Chapter 3. Tables 4.1 and 4.5 are produced to indicate whether the model is parameter redundant (PR), full rank (FR) or near parameter redundant (NR). An even number of total sampling occasions is used, as both simulated data and real data used in Chapter 3 have two secondary samples within each primary periods. In addition to numbering all the models with model codes, we also denote different models by their constituent parameters. The results of deficiency are shown in Tables 4.1 and 4.5, where the transition probabilities are assumed to be constant or time-dependent respectively. In addition, we demonstrate that some of the full rank models can give biased simulation results as shown in Chapter 3, which is caused by near parameter redundancy.

4.3.1 $\phi^{12}(c)$ Scenarios

Considering the general results in Table 4.1, it is clear that RMER models are full rank for all cases. As a result, the robust design improves the estimation of the models in general, as the secondary samples within each primary period provide an additional source of information about capture probability. However, MER models are problematic with only nine out of 12 models being full rank. Three models are parameter redundant with deficiency of one, where the capture probability p is always confounded with the rest of the parameters.

Table 4.1 Generalised parameter redundancy statistics (PRS) of RMER and MER models with constant transition probabilities in Chapter 3. h is the number of parameters, r is the rank of the derivative matrix, $d = h - r$ is the deficiency. Models are termed full rank (FR), parameter redundant (PR), and near parameter redundant (NR).

Model	h	RMER				MER			
		Model code	r	d	PRS	Model code	r	d	PRS
$\pi, \phi^{12}(c), \phi^{21}(c), p(c)$	4	RMER.NNC	4	0	FR, NR	MER.NNC	3	1	PR
$\pi, \phi^{12}(c), \phi^{21}(c), p(\text{cov})$	5	RMER.NNZ	5	0	FR, NR	MER.NNZ	5	0	FR, NR
$\phi^{12}(c), \phi^{21}(c), p(c)$	3	RMER.SNC	3	0	FR	MER.SNC	4	0	FR, NR
$\pi, \phi^{12}(c), p(c)$	3	RMER.NRC	3	0	FR	MER.NRC	2	1	PR
$\pi, \phi^{12}(c), p(c)$	3	RMER.NEC	3	0	FR	MER.NEC	3	0	FR, NR
$\phi^{12}(c), \phi^{21}(c), p(\text{cov})$	4	RMER.SNZ	4	0	FR	MER.SNZ	4	0	FR, NR
$\pi, \phi^{12}(c), p(\text{cov})$	4	RMER.NRZ	4	0	FR	MER.NRZ	4	0	FR, NR
$\pi, \phi^{12}(c), p(\text{cov})$	4	RMER.NEZ	4	0	FR	MER.NEZ	4	0	FR, NR
$\phi^{12}(c), p(c)$	2	RMER.SRC	2	0	FR	MER.SRC	1	1	PR
$\phi^{12}(c), p(c)$	2	RMER.SEC	2	0	FR	MER.SEC	2	0	FR, NR
$\phi^{12}(c), p(\text{cov})$	3	RMER.SRZ	3	0	FR	MER.SRZ	3	0	FR, NR
$\phi^{12}(c), p(\text{cov})$	3	RMER.SEZ	3	0	FR	MER.SEZ	3	0	FR, NR

Table 4.2 Estimable combinations of parameters (EP) for parameter redundant MER models in Table 4.1

Model	Model code	h	r	d	EP
$\pi, \phi^{12}(c), \phi^{21}(c), p(c)$	MER.NNC	4	3	1	πp $\phi^{12} p$ $(\phi^{12} - 1)p - \phi^{12} - \phi^{21}$
$\pi, \phi^{12}(c), p(c)$	MER.NRC	3	2	1	πp $(\phi^{12} - 1)p$
$\phi^{12}(c), p(c)$	MER.SRC	2	1	1	$(\phi^{12} - 1)p$

Note that we initially use $K = 6$ sampling occasions to calculate the rank of all of the models in Table 4.1. If the model is full rank when $K = 6$, then it remains full rank for any $K \geq 6$ because there is no new parameter in the extended model with constant parameters according to the extension theorem. If it is not full rank when $K = 6$, then we examine the parameter redundancy of the extended model. If the model is parameter redundant for a fixed value of K, we need to reparamerise the model so that it becomes full rank and then apply the extension theorem to the reparameterised model rather than the original model.

Table 4.3 Results of standardized smallest eigenvalues (S.Eig) for RMER.NNC, RMER.SNC, RMER.NRC and RMER.SRC models under Scenarios 1-4.

Scenario	Model	S.Eig
Scenario 1	RMER.NNC	1.34×10^{-12}
	RMER.SNC	1.55×10^{-2}
	RMER.NRC	8.61×10^{-2}
	RMER.SRC	7.85×10^{-2}
Scenario 2	RMER.NNC	1.34×10^{-12}
	RMER.SNC	9.96×10^{-2}
	RMER.NRC	1.29×10^{-1}
	RMER.SRC	2.19×10^{-1}
Scenario 3	RMER.NNC	1.03×10^{-11}
	RMER.SNC	3.06×10^{-3}
	RMER.NRC	1.67×10^{-2}
	RMER.SRC	1.58×10^{-2}
Scenario 4	RMER.NNC	6.25×10^{-12}
	RMER.SNC	9.86×10^{-3}
	RMER.NRC	1.39×10^{-2}
	RMER.SRC	2.24×10^{-2}

Near Parameter Redundancy

We have described near parameter redundancy for a full rank model in Section 4.2.5. Following the method of detecting near parameter redundancy in Catchpole et al. (2001), we calculated the expected information matrix in `Maple` and obtained the smallest standardized eigenvalue for the full rank RMER models, which can be computed as the absolute value of the ratio of the smallest eigenvalue and the largest eigenvalue. The results are shown in Tables 4.3 and 4.4, where we described Scenarios 1-4 and Scenarios 5-8 in Chapter 3. All models are diagnosed with $K = 8$ sampling occasions, where there are $T = 4$ primary periods and $k_1 = k_2 = k_3 = k_4 = 2$ secondary samples for each primary period. Covariates are simulated from the uniform(10,12) distribution.

RMER.NNC has extremely small standardized eigenvalues across all scenarios in Table 4.3, which results in biased parameter estimates in the simulation studies given in Chapter 3. In Table 4.4 we show that the standardised smallest eigenvalues for the RMER.SEC model are larger than those in the RMER.NEC model. Also, we observe similar results in the simulation study in Chapter 3, where the performance of RMER.SEC is better than RMER.NEC under each scenario.

Table 4.4 Results of standardized smallest eigenvalues (S.Eig) for RMER.NEC and RMER.SEC models under Scenarios 5-8.

Scenario	RMER Model	S.Eig
Scenario 5	RMER.NEC	9.10×10^{-2}
	RMER.SEC	1.72×10^{-2}
Scenario 6	RMER.NEC	1.35×10^{-1}
	RMER.SEC	4.84×10^{-1}
Scenario 7	RMER.NEC	7.47×10^{-3}
	RMER.SEC	2.23×10^{-2}
Scenario 8	RMER.NEC	9.48×10^{-3}
	RMER.SEC	6.27×10^{-2}

We also notice there is no clear cut-off point between large and small standardized eigenvalues across all scenarios in the cases we considered. So, we suggest that any interpretations of standardized eigenvalues have to be made cautiously.

4.3.2 $\phi^{12}(t)$ Scenarios

We have derived the parameter redundancy results for RMER models with constant transition probabilities in Section 4.3.1. In this section, we determine the results for RMER models with $\phi^{12}(t)$. The results are summarised in Table 4.5.

Table 4.5 shows that all models with both fully time-dependent transition probabilities ϕ_i^{12} and ϕ_i^{21} (i.e without any constraint for ϕ_i^{12} and ϕ_i^{21}) are parameter redundant with deficiency greater than one. RMER.NN^{t}C becomes full rank with constraint “R^t”. Although RMER.NR^{t}C is not parameter-redundant, we find it is near parameter redundant. Furthermore, the issue of near parameter redundancy for RMER.NR^{t}C is overcome if constraint “S” is additionally considered in place of “N” for the initial state parameter.

We also investigate the parameter redundancy of IRMER models using two populations (numbered 1 and 2) with results shown in the last four models in Table 4.5. IRMER.NR^{t}C and IRMER.SR^{t}C use the same constraints as RMER.NR^{t}C and RMER.SR^{t}C respectively and the two populations are assumed to share the same transition probabilities (i.e. $\phi_{i,1}^{12} = \phi_{i,2}^{12}$, $i = 1, \ldots, T-1$). IRMER.S$_{1,2}$R$_a^t$C accommodates both constraint “R$_a^t$” and constraint “S” for both of the populations. IRMER.S$_1$R$_a^t$C employs constraint “R$_a^t$”, but constraint “S” is only considered for population 1. All of these IRMER models are determined to be full rank. However, we find IRMER.NR^{t}C is near parameter redundant. Hence, we conclude that we need to apply at least constraints “S” and “R^t” in order to avoid parameter redundancy and near parameter redundancy.

Table 4.5 Generalised parameter redundancy status (PRS) of RMER and MER models with time-varying transition probabilities in Chapter 3. h is the number of parameters, r is the rank of the derivative matrix, $d = h - r$ is the deficiency. Models are termed full rank (FR), parameter redundant (PR), and near parameter redundant (NR).

Model code	Model	h	r	d	PRS
	RMER model				
RMER.NN^{t}C	$\pi, \phi^{12}(t), \phi^{21}(t), p(c)$	K	$K/2+1$	$K/2-1$	PR
RMER.NN^{t}Z	$\pi, \phi^{12}(t), \phi^{21}(t), p(\text{cov})$	$K+1$	$K/2+2$	$K/2-1$	PR
RMER.SN^{t}C	$\phi^{12}(t), \phi^{21}(t), p(c)$	$K-1$	$K/2+1$	$K/2-2$	PR
RMER.NR^{t}C	$\pi, \phi^{12}(t), p(c)$	$K/2+1$	$K/2+1$	0	FR, NR
RMER.NE^{t}C	$\pi, \phi^{12}(t), p(c)$	$K/2+1$	$K/2+1$	0	FR, NR
RMER.N2^{t}C	$\pi, \phi^{12}(t), \phi^{21}(t), p(c)$	$K-2$	$K/2+1$	$K/2-3$	PR
RMER.SN^{t}Z	$\phi^{12}(t), \phi^{21}(t), p(\text{cov})$	K	$K/2+2$	$K/2-2$	PR
RMER.NR^{t}Z	$\pi, \phi^{12}(t), p(\text{cov})$	$K/2+2$	$K/2+2$	0	FR, NR
RMER.NE^{t}Z	$\pi, \phi^{12}(t), p(\text{cov})$	$K/2+2$	$K/2+2$	0	FR, NR
RMER.N2^{t}Z	$\pi, \phi^{12}(t), \phi^{21}(t), p(\text{cov})$	$K-1$	$K/2+2$	$K/2-3$	PR
RMER.SR^{t}C	$\phi^{12}(t), p(c)$	$K/2-1$	$K/2-1$	0	FR
RMER.SE^{t}C	$\phi^{12}(t), p(c)$	$K/2-1$	$K/2-1$	0	FR, NR
RMER.S2^{t}C	$\phi^{12}(t), \phi^{21}(t), p(c)$	$K-3$	$K/2+1$	$K/2-4$	PR
RMER.NR$^t_2$C	$\pi, \phi^{12}(t), p(c)$	$K/2$	$K/2$	0	FR, NR
RMER.NE$^t_2$C	$\pi, \phi^{12}(t), p(c)$	$K/2$	$K/2$	0	FR, NR
RMER.SR^{t}Z	$\phi^{12}(t), p(\text{cov})$	$K/2+1$	$K/2+1$	0	FR
RMER.SE^{t}Z	$\phi^{12}(t), p(\text{cov})$	$K/2+1$	$K/2+1$	0	FR, NR
RMER.S2^{t}Z	$\phi^{12}(t), \phi^{21}(t), p(\text{cov})$	$K-2$	$K/2+2$	$K/2-4$	PR
RMER.NR$^t_2$Z	$\pi, \phi^{12}(t), p(\text{cov})$	$K/2+1$	$K/2+1$	0	FR, NR
RMER.NE$^t_2$Z	$\pi, \phi^{12}(t), p(\text{cov})$	$K/2+1$	$K/2+1$	0	FR, NR
RMER.SR$^t_2$C	$\phi^{12}(t), p(c)$	$K/2-1$	$K/2-1$	0	FR
RMER.SR$^t_2$Z	$\phi^{12}(t), p(\text{cov})$	$K/2$	$K/2$	0	FR
	IRMER model				
IRMER.NR^{t}C	$\pi_1, \pi_2, \phi^{12}(t), p(c)$	$K/2+2$	$K/2+2$	0	FR, NR
IRMER.S$_{1,2}$R^{t}C	$\phi^{12}(t), p(c)$	$K/2$	$K/2$	0	FR
IRMER.S$_{1,2}$R^t_aC	$\phi^{12}(t), \phi^{12}(t+\gamma_2), p(c)$	$K/2+1$	$K/2+1$	0	FR
IRMER.S$_1$R^t_aC	$\pi_2, \phi^{12}(t), \phi^{12}(t+\gamma_2), p(c)$	$K/2+2$	$K/2+2$	0	FR

Table 4.6 Estimable combinations of parameters (EP) for parameter redundant RMER models in Table 4.5. d is the deficiency of the derivative matrix.

Model code	Model	d	EP
RMER.NN^{t}C	$\pi, \phi^{12}(t), \phi^{21}(t), p(c)$	$K/2-1$	p, π $f(p, \pi, \phi_1^{12}, \phi_1^{21})$ $\ldots$ $f(p, \pi, \phi_1^{12}, \phi_1^{21}, \cdots, \phi_{K/2-1}^{12}, \phi_{K/2-1}^{21})$
RMER.NN^{t}Z	$\pi, \phi^{12}(t), \phi^{21}(t), p(\text{cov})$	$K/2-1$	α, β, π $f(\alpha, \beta, \pi, \phi_1^{12}, \phi_1^{21})$ $\ldots$ $f(\alpha, \beta, \pi, \phi_1^{12}, \phi_1^{21}, \cdots, \phi_{K/2-1}^{12}, \phi_{K/2-1}^{21})$
RMER.SN^{t}C	$\phi^{12}(t), \phi^{21}(t), p(c)$	$K/2-2$	p $f(\phi_1^{12}, \phi_1^{21}, \cdots, \phi_{K/2-1}^{12}, \phi_{K/2-1}^{21})$ $f(p, \phi_1^{12}, \phi_1^{21})$ $\ldots$ $f(p, \phi_1^{12}, \phi_1^{21}, \cdots, \phi_{K/2-1}^{12}, \phi_{K/2-1}^{21})$
RMER.N2^{t}C	$\pi, \phi^{12}(t), \phi^{21}(t), p(c)$	$K/2-3$	$p, \pi, \phi_{K/2-2}^{12}, \phi_{K/2-2}^{21})$ $f(p, \pi, \phi_1^{12}, \phi_1^{21})$ $\ldots$ $f(p, \pi, \phi_1^{12}, \phi_1^{21}, \cdots, \phi_{K/2-2}^{12}, \phi_{K/2-2}^{21})$
RMER.SN^{t}Z	$\phi^{12}(t), \phi^{21}(t), p(\text{cov})$	$K/2-2$	α, β $f(\phi_1^{12}, \phi_1^{21}, \cdots, \phi_{K/2-1}^{12}, \phi_{K/2-1}^{21})$ $f(\alpha, \beta, \phi_1^{12}, \phi_1^{21})$ $\ldots$ $f(\alpha, \beta, \phi_1^{12}, \phi_1^{21}, \cdots, \phi_{K/2-1}^{12}, \phi_{K/2-1}^{21})$
RMER.N2^{t}Z	$\pi, \phi^{12}(t), \phi^{21}(t), p(\text{cov})$	$K/2-3$	$\alpha, \beta, \pi, \phi_{K/2-2}^{12}, \phi_{K/2-2}^{21})$ $f(\alpha, \beta, \pi, \phi_1^{12}, \phi_1^{21})$ $\ldots$ $f(\alpha, \beta, \pi, \phi_1^{12}, \phi_1^{21}, \cdots, \phi_{K/2-2}^{12}, \phi_{K/2-2}^{21})$

Near parameter redundancy

In this section, we show results of standardized eigenvalues for full rank models in Table 4.5 with constant capture probability. Relatively smaller standardized eigenvalues for RMER.NR^{t}C and RMER.NR$^t_2$C models in Tables 4.7 explain their poor behaviors in Chapter 3. Although we observe relatively large standardized eigenvalues for models with constraint "E" in Table 4.8 especially under Scenarios 6 and 8, none of these models yield unbiased simulation results in the simulation study in Chapter 3.

We notice there is no clear cut-off point between large and small standardized eigenvalues across all Scenarios for these models. So, we suggest that any interpretations of standardized eigenvalues have to be made cautiously.

Table 4.7 Results of standardized smallest eigenvalues (S.Eig) for RMER.NR^{t}C, RMER.SR^{t}C, RMER.S2^{t}C, RMER.NR$^t_2$C and RMER.SR$^t_2$C models under Scenarios 1-4.

Scenario	Model	S.Eig
Scenario 1	RMER.NR^{t}C	1.12×10^{-1}
	RMER.SR^{t}C	1.42×10^{-1}
	RMER.NR$^t_2$C	1.04×10^{-1}
	RMER.SR$^t_2$C	2.18×10^{-1}
Scenario 2	RMER.NR^{t}C	3.13×10^{-4}
	RMER.SR^{t}C	4.52×10^{-2}
	RMER.NR$^t_2$C	1.96×10^{-2}
	RMER.SR$^t_2$C	7.73×10^{-2}
Scenario 3	RMER.NR^{t}C	2.02×10^{-2}
	RMER.SR^{t}C	2.12×10^{-2}
	RMER.NR$^t_2$C	1.98×10^{-2}
	RMER.SR$^t_2$C	2.61×10^{-1}
Scenario 4	RMER.NR^{t}C	2.82×10^{-3}
	RMER.SR^{t}C	5.06×10^{-3}
	RMER.NR$^t_2$C	4.42×10^{-3}
	RMER.SR$^t_2$C	4.30×10^{-2}

Table 4.8 Results of standardised smallest eigenvalues (S.Eig) for RMER.NEC, RMER.SEC and RMER.NE$_2$C models under Scenarios 5-8.

Scenario	Model	S.Eig
Scenario 5	RMER.NE^{t}C	1.09×10^{-3}
	RMER.SE^{t}C	4.46×10^{-3}
	RMER.SE$^t_2$C	3.70×10^{-3}
Scenario 6	RMER.NE^{t}C	1.84×10^{-1}
	RMER.SE^{t}C	2.09×10^{-1}
	RMER.SE$^t_2$C	1.67×10^{-1}
Scenario 7	RMER.NE^{t}C	1.37×10^{-4}
	RMER.SEC	1.02×10^{-3}
	RMER.SE$^t_2$C	1.15×10^{-3}
Scenario 8	RMER.NE^{t}C	2.43×10^{-2}
	RMER.SE^{t}C	3.92×10^{-2}
	RMER.SE$^t_2$C	2.58×10^{-2}

Table 4.9 Results of standardized smallest eigenvalues (S.Eig) under different Scenarios, where true parameters are listed in p_{true} and ψ_{true}. True transition probabilities used are the same (0.8,0.7,0.8,0.3,0.6,0.7,0.8,0.6,0.6) for every scenario. All models are diagnosed with $K = 8$ sampling occasions where there are $T = 4$ primary periods and two secondary samples within each primary period. $*$ indicates all individuals are assume to survive during the study.

Scenario	Model	p_{true}	ψ_{true}	S.Eig
1	RMER.NR^{t}C	0.3	$*$	2.02×10^{-2}
1	RMER.SR^{t}C	0.3	$*$	2.12×10^{-2}
F1	RMER.SR^t_dC	0.3	0.95	1.39×10^{-2}
F2	RMER.SR^t_dC	0.3	0.8	1.30×10^{-2}
F3	RMER.SR^t_dC	0.3	0.4	1.07×10^{-2}
3	RMER.NR^{t}C	0.55	$*$	1.15×10^{-1}
3	RMER.SR^{t}C	0.55	$*$	1.42×10^{-1}
G1	RMER.SR^t_d C	0.55	0.95	1.14×10^{-1}
G2	RMER.SR^t_dC	0.55	0.8	1.10×10^{-1}
G3	RMER.SR^t_dC	0.55	0.4	6.05×10^{-2}

The RMER.SR^t_dC model

In Section 3.4, we have explored whether we can estimate a constant survival probability ψ in the model RMER.SR^t_dC, where the subscript “d" represents an additional death state in the transition matrix in the RMER.SR^{t}C model.

This model is determined to be full rank. We also investigate its eigenvalues under various Scenarios and compare those obtained from the RMER.NR^{t}C and RMER.SR^{t}C models. We list the true values of parameters we used in Table 4.9. When capture probability is low its standardized smallest eigenvalue under Scenario F1, F2 or F3 is smaller than those both RMER.NR^{t}C and RMER.SR^{t}C models under Scenario 1 as shown in Table 3.2. A similar conclusion is obtained when capture probability is high. Therefore it is not a reliable model to use as the standardized smallest eigenvalues are smaller than those from the RMER.NR^{t}C model which is a near parameter redundant model.

4.4 Discussion

It is vital to check for parameter redundancy of a model before fitting it to the data. In Chapter 4 we present symbolic methods for detecting parameter redundancy that provide efficient ways to assess the parameter identification for a model. The purpose of this chapter is to identify removal models developed in this thesis that are not parameter redundant using symbolic methods (Catchpole and Morgan, 1997; Cole et al., 2010).

We derived results for parameter redundancy for removal models in Chapters 3 and 4 presented in Section 4.3. We showed that the use of robust design with certain forms of parameter constraints (constraint “S” and “R”) could overcome parameter redundancy for removal models, which theoretically explains the behavior of the models in Chapter 3.

All of the parameter redundancy results here are based on the assumption that at least one animal is removed at every sampling occasion. This ensures that all of the exhaustive summary terms are present as we are interested in detecting parameter redundancy in the model itself instead of parameter redundancy due to the lack of data.

Future work could be done considering the missing data points for removal models. If there are missing data for a specific data set, then the corresponding terms in the exhaustive summary will be zero. In that case, the estimable parameters could become non-identifiable for the data set.

Chapter 5

The Use of Penalised Likelihood for Removal Data

5.1 Introduction

In previous chapters, we developed new removal models to enable the estimation of temporary emigration for data sampled under the robust design protocol. The main argument for classic removal modelling is that when the imperfect availability of the population is not correctly addressed, capture probability is underestimated, which leads to population size being overestimated. Although removal data sets with the robust design structure can be analysed using the models accounting for temporary emigration, as presented in Chapter 3, many data sets do not have such a structure. Matechou et al. (2016) recently developed a Bayesian framework for modelling new arrivals of individuals observed under the standard sampling protocol. However, their approach assumes that new individuals arrive at the study area in an unknown number of groups. In Chapter 5, we relax this assumption to allow new individuals to become available for capture according to time-dependent entry parameters as defined in Section 5.2.1. We develop new removal models in the classic framework that allow populations to be open to new individuals via birth/arrival for data sets without the robust design (e.g. a single observation within each primary period). The model proposed assumes that the population is depleted only via the removal process and individuals cannot depart from the survey site.

New individuals are an important source of arrivals when estimating abundance and ignoring them in removal modelling may have substantial consequences. For example, in removal projects of reptiles and amphibians, most juveniles are born between July and September and therefore can become available for capture during the translocation studies (Edgar et al., 2010). Additionally, hibernation usually occurs from late October to early March, so adult individuals can potentially disperse from hibernation sites and migrate to the study area during the course of the spring season. Failing to account for new arrivals in the study leads to underestimated capture probability and overestimated population sizes (see Section 5.4.3).

Although entry parameters can be considered for populations in an open population removal model accounting for new birth/arrivals of individuals, problems occur with maximum likelihood estimation of such removal models without further development. The issues that arise with maximum likelihood estimation are discussed in Section 5.2.2, where we demonstrate the capture probability is always estimated on the boundary.

In order to address the problem of boundary estimates, we employ penalisation methods considering three different forms of penalty functions. Penalised maximum likelihood methods, or regularisation methods, are widely used in machine learning and statistical modelling in order to control the complexity of a model and reduce the variance of estimators of parameters (Hastie et al., 2009). The idea of penalised maximum likelihood estimation is to maximise the likelihood function with a penalty term which penalises the method for complexity and can discard undesirable estimates by shrinking them towards zero. The estimates of parameters are obtained by maximising a new objective function, which brings together both the likelihood function and the penalty term. This approach has been intensively discussed in the context of linear regression, and recently it also has become an emerging topic in statistical ecology such as occupancy models (Moreno and Lele, 2010; Hutchinson et al., 2015) and capture-recapture models (Viallefont, 2011).

In Chapter 5, we consider three penalty functions. The first penalty is similar to ridge regression (Hoerl and Kennard, 1970; Hastie et al., 2009), and penalises the sum of squares of parameters. The second penalty term is an analogue to Lasso (Tibshirani, 1996), and penalises the sum of the absolute values of parameters. The third penalty is akin to the fused lasso, and penalises the sum of absolute differences between two consecutive parameters (Tibshirani et al., 2005). We perform a simulation study that shows that the proposed penalised likelihood estimation considerably improves the maximum likelihood estimation, which fails completely and gives boundary estimates of capture probability for all scenarios we considered.

The chapter is structured as follows: Section 5.2 specifies the open population removal models using maximum likelihood estimation and penalised likelihood methods. In Section 5.3 we use symbolic approaches discussed in Chapter 4 for determining parameter redundancy status of the new models. Section 5.4 investigates the model using simulation. A real example is shown in Section 5.5, and Section 5.6 presents a discussion and conclusion.

5.2 Open population removal models

In this section, we first introduce the notation in Section 5.2.1 and the formulation for the open population removal model (denoted by OPR) using maximum likelihood estimation in Section 5.2.2. We then consider penalised likelihood methods in Section 5.2.3, where the use of three different penalty terms is explored. The first method we consider is similar to ridge regression or l_2 regularsation (denoted by OPR-l_2), the second model to the lasso or l_1 regularsation (denoted by OPR-l_1), adapted specifically for removal data, and the third penalty term corresponds to the fused lasso (denoted by OPR-fl_1), where we aim to make the two consecutive entry parameters smooth.

5.2.1 Notation

Suppose the total number of sampling occasions is K. The population size is denoted by N. n_k is the number of individuals being removed at the kth sampling occasion, where $k = 1, ..., K$. n_0 is the number of individuals we failed to capture by the end of the study. We assume a constant capture probability p over time. The entry parameters β_k represent the proportion of individuals that become available for removal for the first time at the kth sampling occasion, where $\sum_{k=1}^{K} \beta_k = 1$.

5.2.2 Maximum likelihood estimation

The probability an individual is removed at the kth sampling occasion is

$$L_k = \sum_{j=1}^{k} \beta_j (1-p)^{k-j} p.$$

The probability an individual is not removed by the end of the study is

$$L_0 = \sum_{j=1}^{k} \beta_j (1-p)^{k-j+1}.$$

The full multinomial likelihood is

$$L(n_0, \beta_k, p | n_k) = \frac{N!}{n_0! \prod_{k=1}^{K} n_k!} L_0^{n_0} \prod_{k=1}^{K} L_k^{n_k}. \tag{5.1}$$

We observe that this model is near parameter redundant (see Section 5.3) using the approaches for detecting parameter redundancy discussed in Chapter 4. In Figure 5.1 we illustrate the fitted number of individuals removed obtained from maximising the Equation (5.1) for a simulated data set with $K = 20$ sampling occasions, $p = 0.3$ and $N = 1000$ individuals in the population. The true values of β are shown in Figure 5.1(B). It is clear that the model is overfitted to the simulated data set, where the predicted counts always equal to the observed counts across the study. Furthermore, the estimates of p and n_0 are 0.99 and 1.20×10^{-5} respectively, which means the model produces boundary estimates of p and n_0. The results agree with what we found in a simulation study in Section 5.4.3 for the OPR model, where the estimates of p are always on the boundary of one for all scenarios we considered.

Therefore the OPR model will not produce reliable estimates in practice. In order to overcome the issue of near parameter redundancy, we additionally consider a penalty term for the log of likelihood in the form of Equation (5.1) in Section 5.2.3.

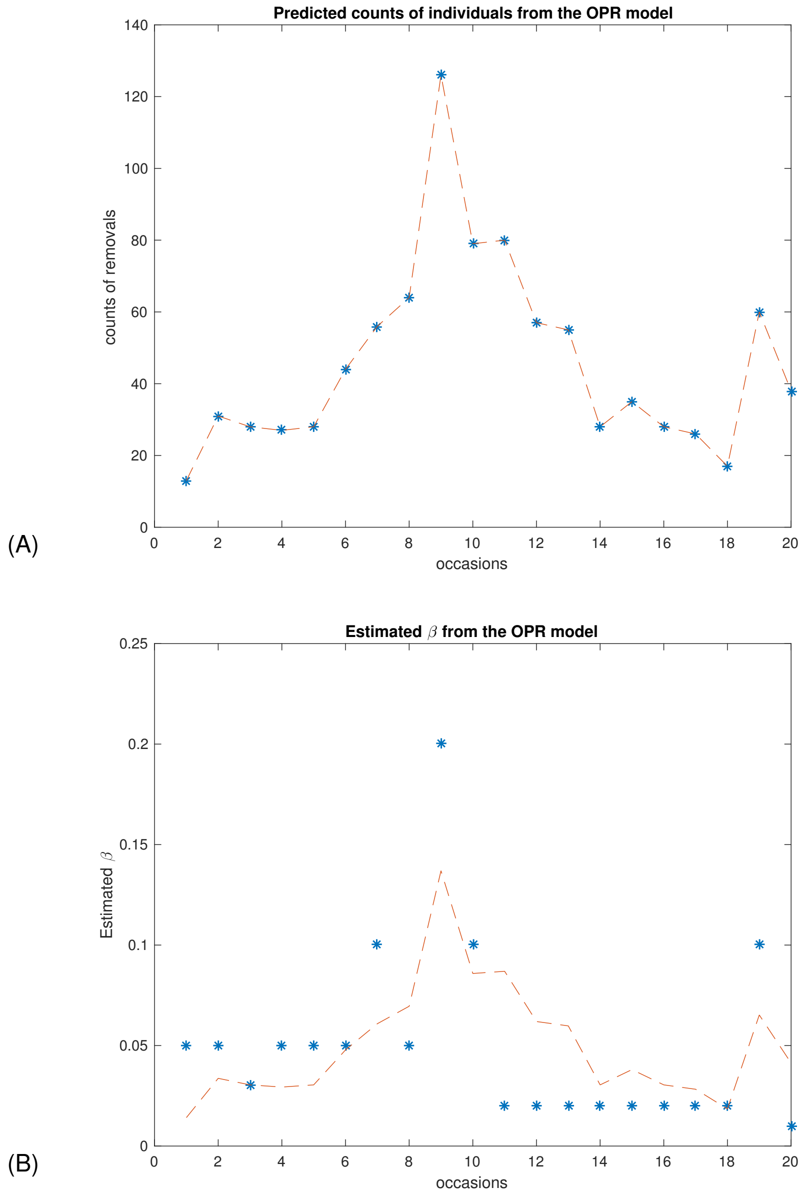


Fig. 5.1 (A) Fitted counts of individuals (dashed line) removed obtained from the OPR model for one simulated data set (*) with $N = 1000$, $p = 0.3$ and $K = 20$ sampling occasions. (B) Estimates of β (dash line) and true values of β (stars).

5.2.3 Penalty functions for removal data

The use of penalty terms is originally developed in the context of linear regression with the aims of reducing the prediction errors by penalising the magnitudes of their coefficients (Hastie et al., 2015). One of the key aims of fitting linear regression models is to interpret the effect of the explanatory variables on the response variable, so it is critical to select which set of variables are included in the model. The most widely used approaches for variable selection include stepwise procedures and criterion-based selection, such as the AIC and predicted residual sum of squares. However, stepwise selection only improves prediction accuracy when there are only a few covariates with strong relationship with the response variable (Hurvich and Tsai, 1990; Harrell, 2001).

In Section 5.2.2 we have discussed the issues of the OPR model using maximum likelihood estimation. In this Section, we adapt penalised likelihood methods for removal data and introduce a penalty function in terms of entry parameters β for the log-likelihood function of the OPR model. The idea is that instead of maximising $\log\{L(n_0,\beta_k,p|n_k)\}$, we use a new objective function in the form of Equation (5.2) for optimisation (Tibshirani, 1996),

$$O(n_0,\beta,p|n_k) = \log\{L(n_0,\beta_k,p|n_k)\} - \lambda\Omega(\beta), \tag{5.2}$$

where $\Omega(\beta)$ is a function that penalises some form of the estimates of entry parameters β. $\lambda \geq 0$ is a tuning (or smoothing) parameter that controls the trade-off between the fit and the amount of smoothness and plays a crucial role in the penalised likelihood estimation.

Note that when $\lambda = 0$, the penalty term has no effect, and the estimation will produce the maximum likelihood estimates, where the model is near parameter redundant. Furthermore, we do not estimate the tuning parameter λ but instead we fix it to a positive value because β and λ are not individually identifiable and the highest value of the objective function would be obtained when the smoothing parameter is equal to zero. This would result in overfitting to the data which is not desirable. Therefore, we need to select the λ using an alternative criterion, namely cross-validation, where we use the mean squared error (MSE) as a criterion, which is the average squared distance between the estimate and the true value. We will give more detail about cross-validation in the subsequent section.

The inferential theory for models involving the use of penalised likelihood is not standard. This is because of the presence of the penalty, which undermines the use of classic asymptotic likelihood results for practical modelling. Instead, bootstrapped replicates are used for computing standard errors and confidence intervals (Hastie et al., 2009).

Ridge for the OPR$-l_2$ model

The first penalty term for β that we considered for the removal data is a ridge penalty as shown in Equation (5.3). The use of ridge penalty was first proposed by Hoerl and Kennard (1970) and it is a popular technique for improving prediction accuracy of regression coefficients. It can improve prediction errors by shrinking large regression coefficients towards zero.

$$\Omega^{\text{ridge}}(\beta) = \sum_{j=1}^{K} \beta_j^2. \tag{5.3}$$

Therefore, the new objective function for the OPR$-l_2$ model is penalised by subtracting a penalty function $\lambda\Omega = \lambda \sum_{j=1}^{K} \beta_j^2$ as shown in Equation (5.4),

$$O^{\text{ridge}}(n_0, \beta, p|n_k) = \log\{L(n_0, \beta_k, p|n_k)\} - \lambda \sum_{j=1}^{K} \beta_j^2, \tag{5.4}$$

where $\lambda \geq 0$ is a tuning parameter that controls the amount of shrinkage. The larger the value of λ, the greater the amount of shrinkage. The entry parameters β are shrunk towards zero which is similar to shrinking coefficients towards zero in ridge regression.

As we mentioned above, we expect the OPR-l_2 to shrink the estimates of β towards zero as λ increases. In Figure 5.2 the estimates of β are shown from modeling a simulated data set using three different values of $\lambda = 20$, 1000 and 30000. It is clear that the OPR$-l_2$ model is able to shrink the peak estimate of β_1 towards zero, therefore the rest of β gains more weight. In an extreme case of $\lambda = 30000$, bigger estimates of β_k in Figure 5.2 (B) move towards zero, and similarly smaller $\hat{\beta}_k$ becomes larger (i.e. $\hat{\beta}_1$), which is because the constraint of $\sum_k \beta_k = 1$.

In Figure 5.3 we also demonstrate how the estimate of capture probability changes with different values of λ for the same simulated data set used in Figure 5.2. When λ is less than about 90, the estimates of p are still on the boundary of one as the amount of shrinkage in the model is small. As λ becomes larger, the model is able to recover reasonable estimates of p which are quite close to the true value of p. When λ gets sufficiently large, the estimated p first converges to the true value used for simulation and then remains relatively flat with several $\hat{p}$s on the boundary of zero. The reason for obtaining some boundary estimates of $\hat{p}$ in Figure 5.3(B) is because the maximum number of iterations allowed is set to be 10000 when the “quasi-newton” algorithm is used for optimisation in `Matlab`. Therefore, it is possible that the numerical algorithm stops when the maximum number of iterations is reached. We have used different starting values to initialise the optimisation, however, when λ becomes extremely large obtaining the global maximum becomes more difficult and optimisations may still produce boundary estimates of $\hat{p}$ after 10000 iterations.

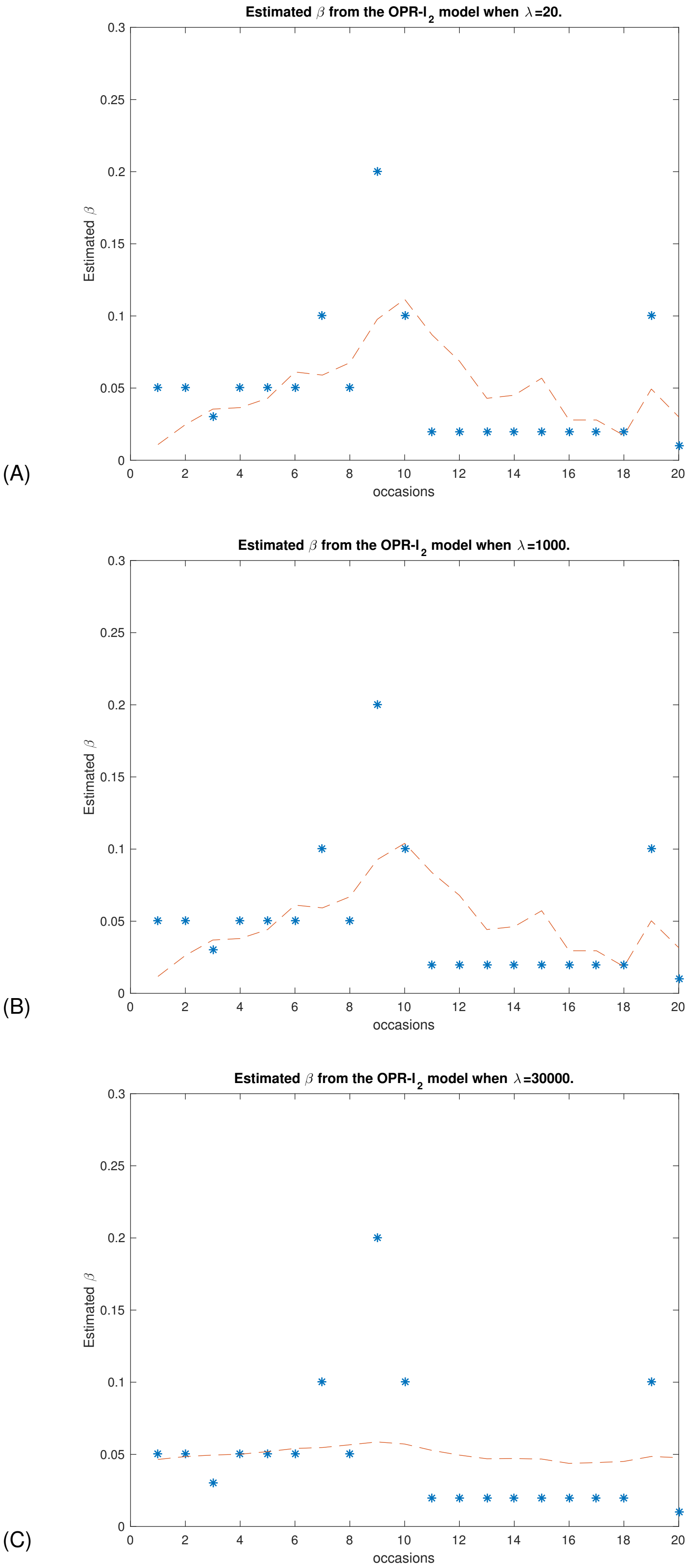


Fig. 5.2 Fitted values of β (dashed lines) obtained from the OPR-l_2 model with different λ values, i.e. (A) $\lambda = 20$, (B) $\lambda = 1000$ and (C) $\lambda = 30000$. Data is simulated using $p = 0.3$, $N = 1000$ for $K = 20$ sampling occasions. True values of β are indicated by stars.

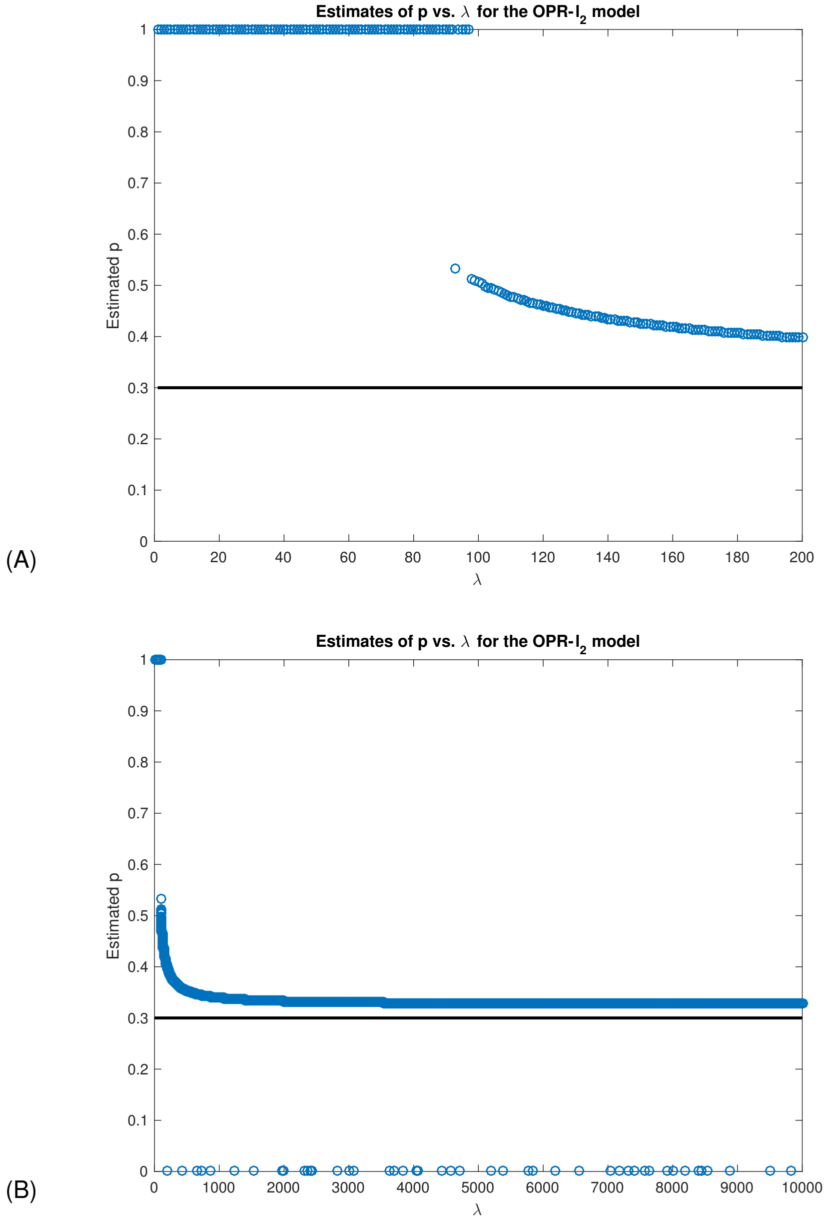


Fig. 5.3 Estimates of p vs. the tuning parameter λ used for the OPR$-l_1$ model. The black horizontal line is the true value of p used to simulate the data. (A) $\lambda = (0,\ldots,200)$. (B) $\lambda = (0,\ldots,10000)$

Lasso for the OPR$-l_1$ model

The Lasso is another shrinkage method that has been widely used to enhance not only the prediction accuracy but also the interpretability of the regression model. It was originally formulated by Tibshirani (1996) based on the Leo Breiman's nonnegative garrote (Geisser, 1993; Kohavi, 1995) for linear least squares models.

The second penalty we considered for removal data is analogous to the Lasso as shown in Equation (5.5),

$$\Omega^{\text{lasso}}(\beta) = \sum_{j=2}^{K} |\beta_j| = \sum_{j=2}^{K} \beta_j. \tag{5.5}$$

The Lasso penalises the sum of the absolute values of the coefficients in linear regression, therefore the absolute values of entry parameters in the first part of Equation (5.5). As β is a vector of probabilities which are non-negative, we can drop the modulus. Additionally, the β add up to 1, so we cannot penalise the sum of all of them. So we choose to leave the first one as a free parameter as the expectation is that a part of the population will be available at the start of the study by study design. When λ is sufficiently large, the lasso could penalise some of the β to be exactly zero and assign more weight to the first entry parameter β_1, so the model assumes there are more individuals available for capture for the first time at the beginning of the study.

The new objective function for the OPR$-l_1$ model is shown in Equation (5.6):

$$O^{\text{lasso}}(n_0, \beta, p|n_k) = \log\{L(n_0, \beta_k, p|n_k)\} - \lambda \sum_{j=2}^{K} \beta_j, \tag{5.6}$$

where the $\lambda \geq 0$ is again a fixed value, which controls the shrinkage of the coefficients towards zero.

Although ridge regression has been popular in reducing prediction error, it does not perform variable selection as the ridge penalty is less likely to make the coefficients exactly zero, therefore it does not make the model more interpretable. Tibshirani (1996) details the nature of shrinkage for Lasso and why in contrast to methods such as ridge regression, the use of Lasso may improve the interpretation of models by shrinking coefficients of variables towards exactly zero, in order to select only a subset of covariates rather than keeping all of them in the model. So Lasso essentially can be regarded as an alternative to variable selection.

In the case of modelling removal data using the objective function in Equation (5.6), it can be interpreted as an entry parameter selection, as it potentially can shrink some of the β_j to be exactly zero if λ is large enough. In Figure 5.4 we obtain estimates of β from modeling a simulated data set using three different values of $\lambda =$, 0 (Figure 5.4.A), 66 (Figure 5.4.B) and 90 (Figure 5.4.C). The estimate of the first entry parameter β_1 becomes larger as we increase the value of λ. This is because the OPR$-l_1$ model shrinks some of the β with very small values to be exactly zero, therefore β_1 gains more weight. In an extreme case of $\lambda = 90$, all of β_k, $k = 2, \ldots, K$ become zero, so β_1 is estimated to be one as shown in Figure 5.4 (C), which means all individuals are available for capture at the first sampling occasion, which is equivalent to the closure assumption in the classic removal model.

As the motivation of the development of the OPR model is near parameter redundancy and the resulting boundary estimate of capture probability, in Figure 5.5 we illustrate $\hat{p}$ obtained from modelling the same simulated data set used in Figure 5.4 for a vector of fixed values of λ. When λ is small, the estimates of p are still on the boundary of one. As λ increases, we observe some estimates of p which are closer to the true value used for simulating the data set. When λ is too large, the estimated p are on the boundary of zero, which matches the simulation results for the classic removal model in Section 5.4.3. This is because when λ is sufficiently large, $\beta_2, \ldots, \beta_K \rightarrow 0$ and $\beta_1 \rightarrow 1$ in the OPR$-l_1$ model, so the model equivalently becomes the classic removal model where all individuals in the population are exposed to removal sampling effort across the study as we mentioned before. Although it seems that we never hit the true value of $p = 0.3$ for this particular simulated data set, in Section 5.4.1 we show that simulation results suggest we could obtain unbiased estimate of p when population size is large.

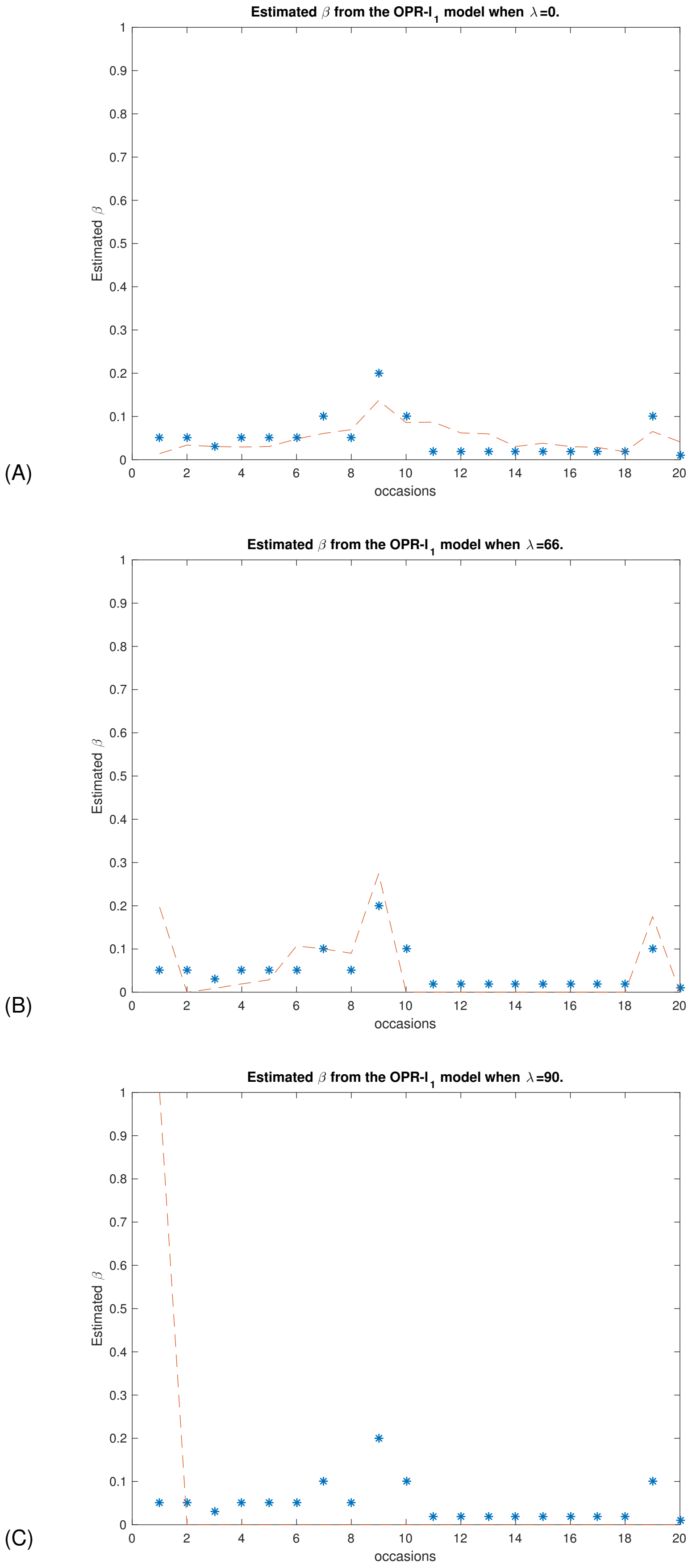


Fig. 5.4 Fitted values of β (dashed lines) obtained from the OPR-l_1 model with different λ values, i.e. (A) $\lambda = 0$, (B) $\lambda = 66$ and (C) $\lambda = 90$. Data is simulated using $p = 0.3$, $N = 1000$ for $K = 20$ sampling occasions. The true values of β are indicated by stars.

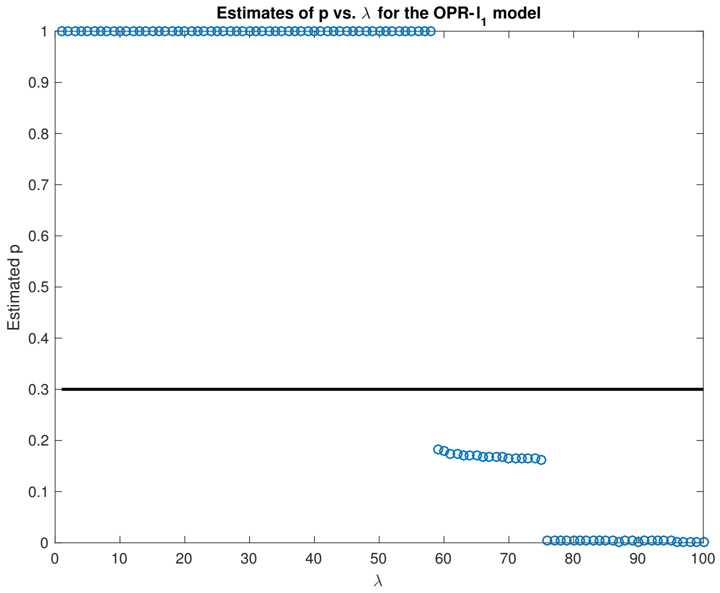


Fig. 5.5 Estimates of p vs. the tuning parameter λ used for the OPR$-l_1$ model. The black horizontal line is the true value of p for simulating the data.

Fused lasso for the OPR$-fl_1$ model

The fussed lasso is one of the commonly used Lasso variants, which penalises large changes with respect to temporal or spatial features (Tibshirani et al., 2005). It is especially useful for modelling spatial or temporal structures, such as time series or image based data. One simple example of the fused lasso is to model ordered data, so the fused lasso is essentially solving the classic change point problem, which has a wide range of applications.

The last penalty term we considered is motivated by the fussed Lasso for removal data and is given by:

$$\Omega^{\text{fussed-lasso}}(\beta) = \sum_{j=1}^{K-1} |\beta_j - \beta_{j+1}|. \tag{5.7}$$

The objective function for the OPR$-fl_1$ model is given in Equation (5.8):

$$O^{\text{fussed-lasso}}(n_0, \beta, p|n_k) = \log\{L(n_0, \beta_k, p|n_k)\} - \lambda \sum_{j=1}^{K-1} |\beta_j - \beta_{j+1}|. \tag{5.8}$$

The idea of the OPR-fl_1 model is to penalise the differences between two consecutive entry parameters and to force the coefficients to vary as a smooth function. When λ becomes larger, we expect that the differences between the two adjacent βs tend to be smaller. Therefore, if λ is extremely large, the estimates of β will become equal in the OPR$-fl_1$ model (i.e. $\hat{\beta_1} = \cdots = \hat{\beta_K}$).

There are also other variants of Lasso that have been developed for different types of dependencies in the covariates in regression models. For example, the Elastic net lasso additionally considers a ridge penalty function (Zou and Hastie, 2005) and it improves the performance of the Lasso when the number of predictors is greater than the sample size. The Group Lasso allows the selection of strongly correlated covariates together when variables can be categorised into different groups (Yuan and Lin, 2006).

In Figure 5.6 we obtain estimates of β from modeling a simulated data set using three different values of λ, 50, 400 and 1896. As expected the estimates of β becomes flat as we increase the value of λ. When λ is extremely large, all of β become equal as shown in Figure 5.6 (C). In Figure 5.7 we demonstrate a plot of $\hat{p}$ against λ obtained from fitting the same simulated data set used in Figure 5.6. When λ is small, $\hat{p}$ is on the boundary of one. As λ becomes larger, the estimates of p steadily move away from the boundary of one and start getting closer to the true value of p after reaching a minimum point that the model can produce.

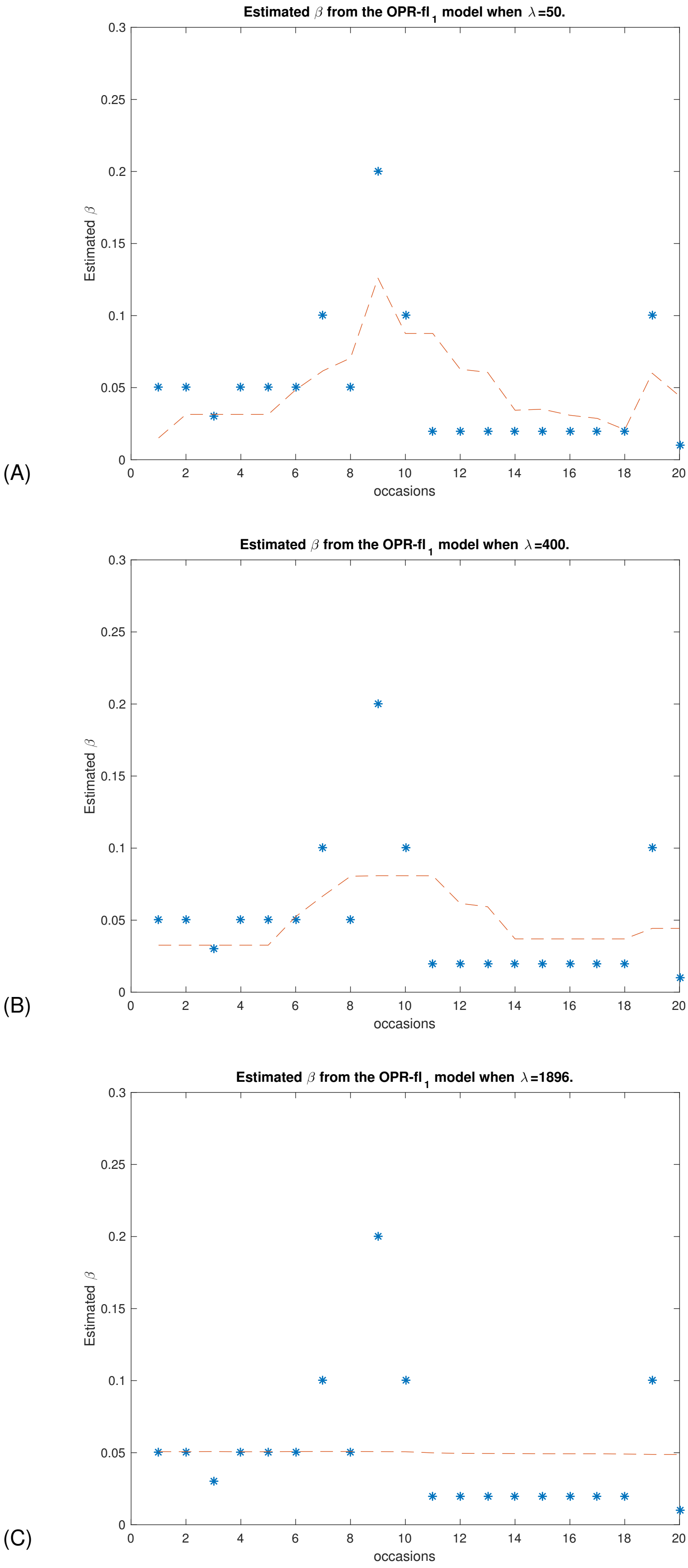


Fig. 5.6 Fitted values of β (dash lines) obtained from the OPR-fl_1 model with different λ values, i.e. (A) $\lambda = 50$, (B) $\lambda = 400$ and (C) $\lambda = 1896$. Data is simulated using $p = 0.3$, $N = 1000$ for $K = 20$ sampling occasions. True values of β is indicated by stars.

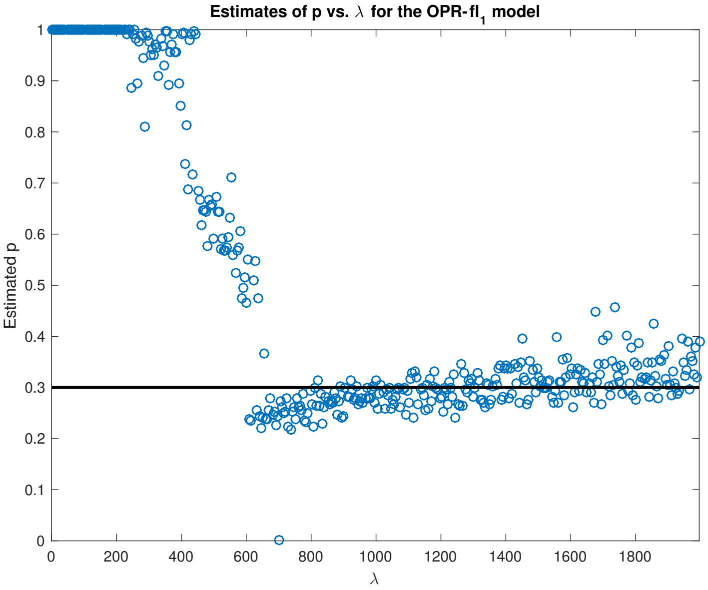


Fig. 5.7 Estimates of p vs. the tuning parameter λ used for the OPR$-fl_1$ model. The black horizontal line is the true value of p for simulating the data.

Cross-validation

In previous sections, we discussed the use of penalty terms for removal data and we demonstrated that the boundary estimate of p can be potentially overcome with some values of λ in the penalised likelihood estimation, where $\lambda \geq 0$ controls the trade-off between the penalty and the fit. Therefore, we require a method to determine which of the models under consideration is the best, thus we need a method determining a fixed value for the λ. In this section we discuss the cross-validation method for choosing λ.

Cross-validation provides a way to select a value for λ in order to implement regularisation methods. In short, we first need to choose a grid of possible λ values, and compute a cross-validation error for each value of λ. We then select the value for which the cross-validation error is the smallest. Finally, the model is fitted using all of the observations with the selected value of λ.

In order to obtain the cross-validation error, we need to separate the original data into the two parts, a training set and a test set. The approach of separating training and test sets is commonly used for classification and regression tree models (Breiman et al., 1984). A similar approach is carried out for neural network models (Reed et al., 1993). Both training and test sets are used for model selection. The training set is used to maximise the performance of a pre-defined criteria and the test set is used to validate its performance for the same value of λ used in modelling the training set. The values of the parameters are chosen not as the ones with optimal criterion on the training test, but those where the performance of the test set is the best. Therefore, as the model fitting for the training set proceeds, we need to monitor the performance on the test set by evaluating the criterion. We could terminate the training when the performance of the test set begins to deteriorate. It is expected that the performance on the test set will first improve, then begin to deteriorate beyond a certain model complexity.

Typically mean squared error (MSE) is used for quantitative data sets, where it is calculated as the average squared difference between the estimated data and the observed data. Furthermore, misclasssification rate is assessed for the qualitative responses. Estimates with smaller MSE are preferred.

There are many existing methods of cross-validation with k-fold cross validation are the most widely used. The idea is to randomly divide the original data into k equally-sized samples, leave out one of the samples and fit the model to the other parts. The squared error is obtained from the left-out sample and the procedure is repeated until all of k samples have been left out once. The MSE is calculated at the end. As we can see, the cost of k-fold cross validation increases as k gets larger. Most common choice is k=5 or 10 as it provides a good compromise for the bias-variance trade-off

(Zhang, 1993). Kohavi (1995) suggested that the 10-fold cross validation is the best for model selection, even if more folds can be considered for computers with large computational power. Another usual choice for cross validation is the leave-one-out method, where the test set consists of a single data point. The cross validation estimate of MSE is then the average over all possible training sets. Monte Carlo cross validation (Burman, 1989) is also an option. In this case, the original data set is randomly split into the training and test sets with replacement. The model is fitted to each training data set and the predictive accuracy (e.g. mean squared error) is assessed using the test data sets. Once we fit the model to each resample, the results are averaged over all the repeated samples. This approach may exhibit Monte Carlo variation, so the results may vary with a different number of random resamples. As the number of random resamples gets large, the results of validation tend to be similar. Hastie et al. (2011) suggested the use of at least 50 Monte Carlo resamples for cross validation is needed .

All the cross-validation approaches mentioned before are not valid for time series models without adaption due to the inherent serial correlation of the data. Similarly, the order of removal data we used is critical because of potential peaks of new arrivals of individuals, standard cross-validations might be problematic. In this chapter, we adapt the Monte Carlo cross-validation for removal data with 200 resamples. The algorithm is listed below.

- Step 1: rewrite the removal data in terms of removal records for each observed individual. For example, if $n_3 = 5$ at $k = 3$, we rewrite it as five threes, i. e. 3, 3, 3, 3, 3.

- Step 2: randomly split the removal records into a training (70%) and a test (30%) dataset.

- Step 3: for a given value of λ, fit the model to the training data, save the estimates of β_k.

- Step 4: fit the model to the test data given the estimates of β_k obtained from the training data and regard p and N as free parameters. Calculate its squared error.

- Step 5: repeat Step 2 - Step 4 a large number of times, and calculate the MSE for the given value of λ.

- Step 6: repeat Step 5 for a vector of values of λ, and choose the λ for which the resulting MSE is the smallest.

Since the test and training data sets will give rise to different estimates for N and N and p are highly correlated. Therefore we need to obtained new estimates for these two parameters using the test data set.

In Figure 5.8 we show the cross-validation results obtained from fitting the OPR$-l_1$ model to the same simulated data we used in the previous sections. Because we potentially can observe non-boundary estimates of p for λ within an approximate range of (55,75), we calculate MSE for 200 resamples from $\lambda = 50$ to $\lambda = 90$ to reduce computational costs. We observe that the MSE gradually becomes smaller at the start, then rapidly gets bigger after the minimum MSE value has been reached.

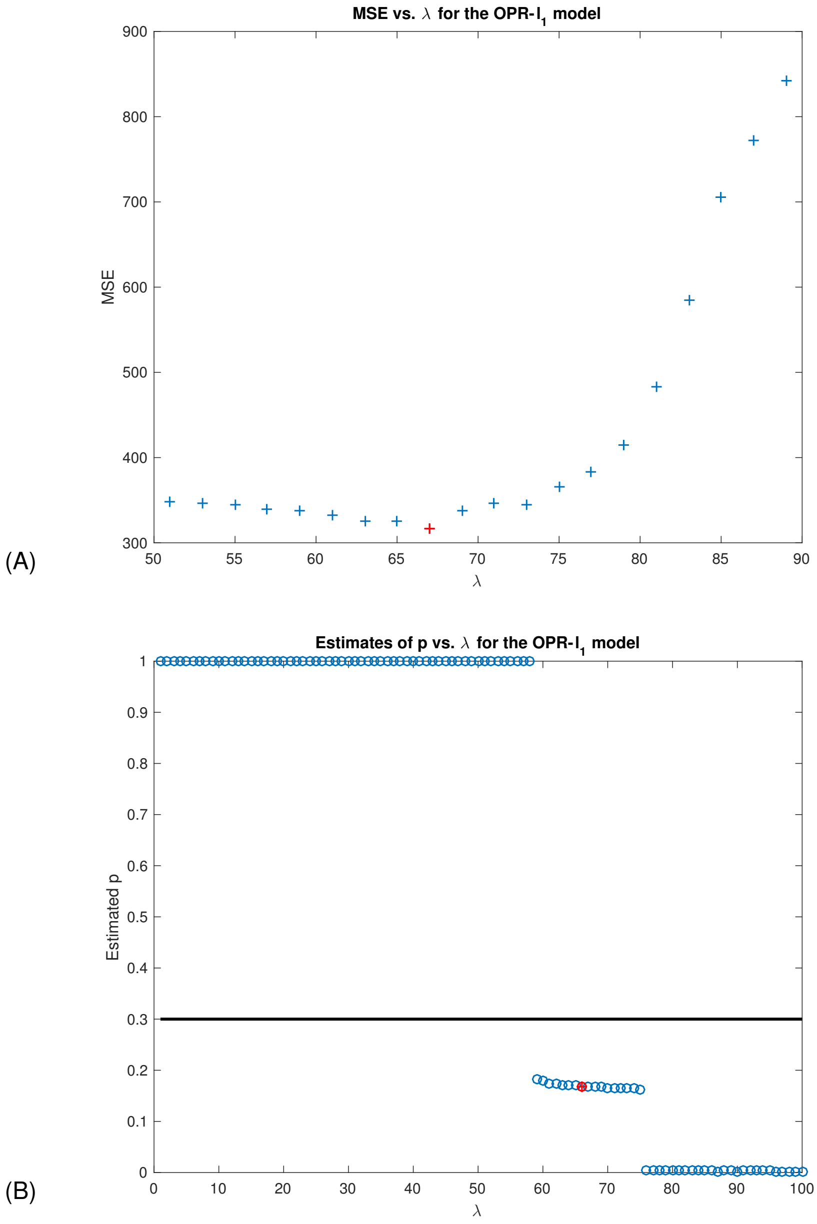


Fig. 5.8 Results of mean squared errors (A) and estimates of p (B) obtained from the OPR-l_1 model. The selected estimate of p (red circle cross) is indicated corresponding to the minimal MSE (red cross).

5.3 Parameter redundancy

In Chapter 4, we discussed methods for determining whether a model is parameter redundancy. Recall that we employ the approaches of Cole et al. (2010) by forming a derivative matrix $\mathbf{D} = \partial \boldsymbol{\kappa}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$, where $\boldsymbol{\kappa}(\boldsymbol{\theta})$ is a vector of parameter combinations that represents the structure of the model for a set of parameters $\boldsymbol{\theta}$ of length h. Once $\mathbf{D}$ is formed, we then calculate the rank of $\mathbf{D}$, r, and the deficiency of the model, $d = h - r$. If $d > 0$, the model is parameter redundant, otherwise if $d = 0$ the model is not parameter redundant, and termed full rank. Once we obtain the results for a fixed number of sampling occasions, we obtain the generalised results for any larger number of samples using the extension theorem or reparameterisation theorem. We investigate the parameter redundancy status for the removal models we developed in this chapter. We demonstrate an example for the OPR model in Section 5.3.1. For the penalised OPR models, we consider the corresponding penalty term as an extra element in the exhaustive summary. We observe that the OPR, OPR-l_1, OPR-l_2 and OPR-fl_1 models are full rank. However, the OPR model is near parameter redundant and we discuss its results in Section 5.3.2. `Maple` code for deriving the results in this section is available in the file `ex5.3.1.mw` in the electronic appendix provided.

5.3.1 Example 5.3.1: the OPR model

We consider the model with $K = 10$ sampling occasions and we define the exhaustive summary as a vector of probabilities of an individual being removed at the the kth occasion. The $h = 10$ parameters we used for detecting parameter redundancy are the entry parameters and the constant capture probability, i.e. $\theta = [p \quad \beta_1 \quad \cdots \quad \beta_9]$, where we constrain the last $\beta_{10} = 1 - \sum_{j=1}^{9} \beta_j$.

We only show the first four terms in the exhaustive summary $\kappa(\theta)$ below. Full results are available in the `Maple` code in the file `ex5.3.1.mw`.

$$\kappa(\theta) = [L_k] = \begin{bmatrix} \beta_1 p \\ \beta_1(1-p)p + \beta_2 p \\ \beta_1(1-p)^2 p + \beta_2(1-p)p + \beta_3 p \\ \beta_1(1-p)^3 p + \beta_2(1-p)^2 p + \beta_3(1-p)p + \beta_4 p \\ \vdots \end{bmatrix}, \tag{5.9}$$

where L_k is the probability of capture an individual on the kth sampling occasion. We remove the more complicated probability that an animal is never captured to make the exhaustive summary simpler, as we assume we observe at least one individual for all possible other capture histories and all the probabilities sum up to one and (Catchpole and Morgan, 1997). We also exclude the term for individuals never captured in the exhaustive summary because it will not affect the result of the deficiency of derivative matrix for dataset following a multinomial distribution as discussed in Section 4.2.4 in Chapter 4.

The derivative matrix is given by

$$\mathbf{D}(\theta) = \left[\frac{\partial \kappa(\theta)}{\partial \theta}\right] = \begin{bmatrix} \beta_1 & \beta_1(1-p)-\beta_1 p+\beta_1 & \beta_1(3p^2-4p+1)+\beta_2(1-2p)+\beta_3 & \cdots \\ p & (1-p)p & (1-p)^2 p & \cdots \\ 0 & p & (1-p)p & \cdots \\ 0 & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix},$$

and is found to have symbolic rank $r = 10$. Therefore the model is not parameter redundant as the deficiency is $d = h - r = 10 - 10 = 0$.

Because the OPR is determined as full rank for a fixed number of sampling occasions ($K = 10$), we can employ the extension theorem to generalise the parameter redundancy results to any length of the study as discussed in Chapter 4. The model can be extended by additionally considering one more sampling occasion. Let $\theta^{'} = [\beta_{11}]$ denote the extra parameter and $\kappa^{'}(\theta^{'}) = \beta_1(1-p)^{10}p + \beta_2(1-p)^9 p + \cdots + \beta_{11}p$ denote the additional extra term in the exhaustive summary. We use the theory of Catchpole and Morgan (1997) and Cole et al. (2010) that if both $\mathbf{D}(\theta)$ and $\mathbf{D}^{'}(\theta^{'})$ are full rank, then we can explicitly conclude that the extended model is also full rank for any larger number of samples.

The extended part of the derivative matrix is $\mathbf{D}^{'}(\theta^{'}) = [p^2]$ and has full rank 1 as there is only 1 parameter in $\theta^{'}$. As a result, the extended model is full rank for any $K \geq 10$ sampling occasions.

5.3.2 Near parameter redundancy

We observe that the OPR model is full rank in theory but it performs badly in practice, therefore is a near parameter redundant model. Recall that we have discussed near parameter redundancy in Section 4.2.4. In a near parameter redundant model the smallest eigenvalue of the expected information matrix will be very close to zero (Catchpole et al., 2001). The smallest standardised eigenvalues (s.eigen) of the OPR

Table 5.1 Results of standardized smallest eigenvalues for the OPR model under different simulation settings in Section 5.4.1.

Simulation setting A			
Scenario A1	Scenario A2	Scenario A3	Scenario A4
1.81×10^{-2}	1.39×10^{-4}	9.67×10^{-3}	2.15×10^{-4}
Simulation setting B			
Scenario B1	Scenario B2	Scenario B3	Scenario B4
1.94×10^{-2}	2.26×10^{-4}	1.81×10^{-2}	1.39×10^{-4}

model are very small as shown in Table 5.1 for simulations using parameter values in Table 5.2 when there are $K = 10$ sampling occasions.

In order to further examine the near parameter redundancy results of the OPR model, we investigate the flatness of profile likelihood using simulated data, where optimisation is conducted for a given value of parameter of interest with respect to the rest of parameters (Venzon and Moolgavkar, 1988; Murphy and van der Vaart, 2000). Similar approaches were used for assessing the practical identifiability of mathematical dynamic models (Raue et al., 2009; Chis et al., 2016). In Figure 5.9 we show profile likelihood plots for p and n_0 obtained from the OPR model under various simulation scenarios. It is clear that the profile likelihood surface of p is very flat under each scenario, and it reaches a peak on the boundary of one which agrees with what we discussed before. Similarly, the number of individuals that we failed to to capture tends to zero, which means that the model infers that all the individuals are captured during the study.

To investigate whether adding a penalty term to the model can improve near redundancy, we investigate the smallest standardised eigenvalues obtained from the penalised models for a vector of λ. Results of the s.eigen against λ for the OPR$-l_1$ and OPR$-l_2$ models under Scenarios A1-A4 are shown in Figures 5.10. They suggest a potential of improvement over the OPR model with some s.eigen that are larger than those obtained from the OPR model if we additionally consider the penalty term in the model. We also show results of s.eigen for the OPR$-fl_1$ model under Scenarios B1-B4 in Figure 5.11. We observe that the smallest standardised eigenvalues are smaller than those obtained from the OPR model under Scenarios B1, B3 and B4 for the values of λ considered. Under Scenario B2 there are some s.eigen that are larger than s.eigen in the OPR model, however simulation results in Section 5.4.2 demonstrate the poor performance of the OPR$-fl_1$ model across all scenarios.

Table 5.2 True values of parameters used for the different simulation settings.

Model	Scenario	p	β
OPR-l_1/OPR-l_2	A1	0.3	(0.1, 0.1, 0.1, 0.2, 0.05, 0.1, 0.15, 0.1, 0.05, 0.05)
	A2	0.7	(0.1, 0.1, 0.1, 0.2, 0.05, 0.1, 0.15, 0.1, 0.05, 0.05)
	A3	0.3	(0.1, 0.25, 0.1, 0.02, 0, 0 , 0.25, 0, 0, 0.1)
	A4	0.7	(0.1, 0.25, 0.1, 0.02, 0, 0 , 0.25, 0, 0, 0.1)
OPR-fl_1	B1	0.3	(0.1, 0.15, 0.1, 0.2, 0.05, 0, 0.1, 0.015, 0.01, 0.05)
	B2	0.7	(0.1, 0.15, 0.1, 0.2, 0.05, 0, 0.1, 0.015, 0.01, 0.05)
	B3	0.3	(0.1, 0.1, 0.1, 0.2, 0.05, 0.1, 0.15, 0.1, 0.05, 0.05)
	B4	0.7	(0.1, 0.1, 0.1, 0.2, 0.05, 0.1, 0.15, 0.1, 0.05, 0.05)

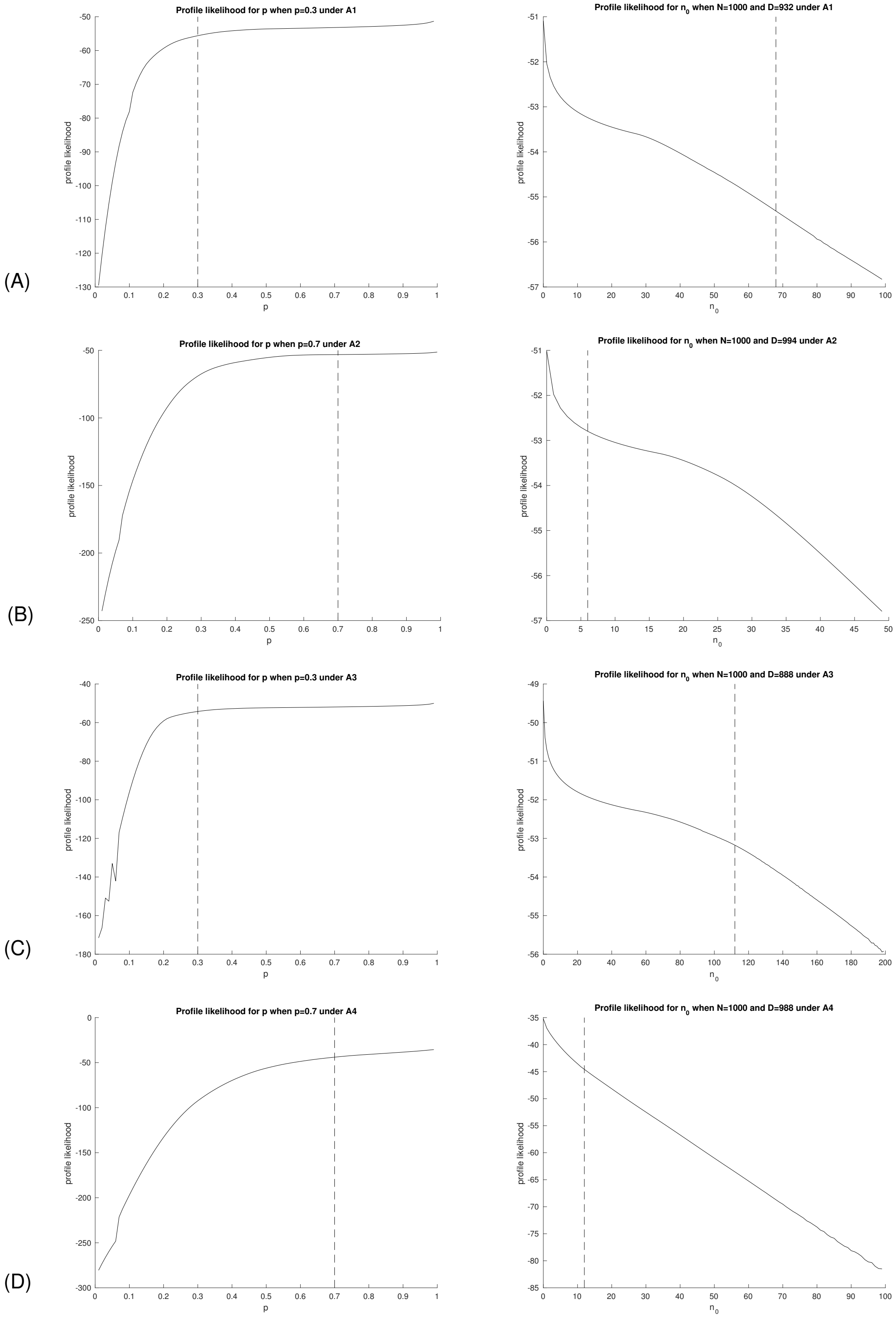


Fig. 5.9 Profile likelihood plots for p (left) and n_0 (right) obtained from the OPR model under various simulation scenarios. The black vertical dash lines are the true values of parameters.

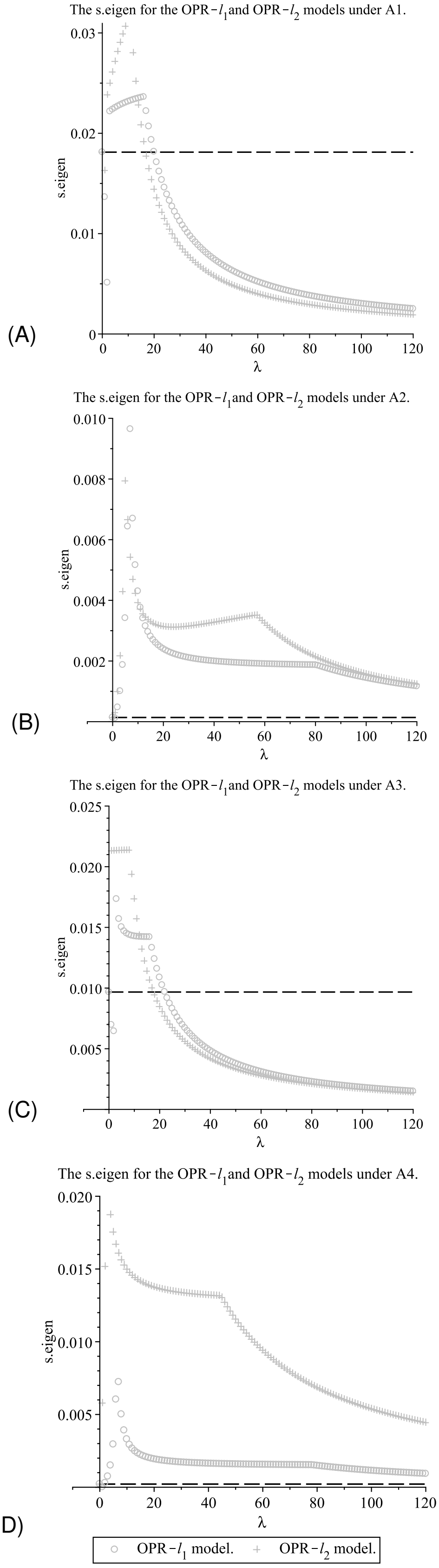


Fig. 5.10 The smallest standardised eigenvalues (s.eigen) vs. λ for the OPR$-l_1$ and OPR$-l_2$ models under simulation scenarios A1(A), A2(B), A3(C) and A4(D). The black horizontal dashed lines are the s.eigen values obtained from the OPR model.

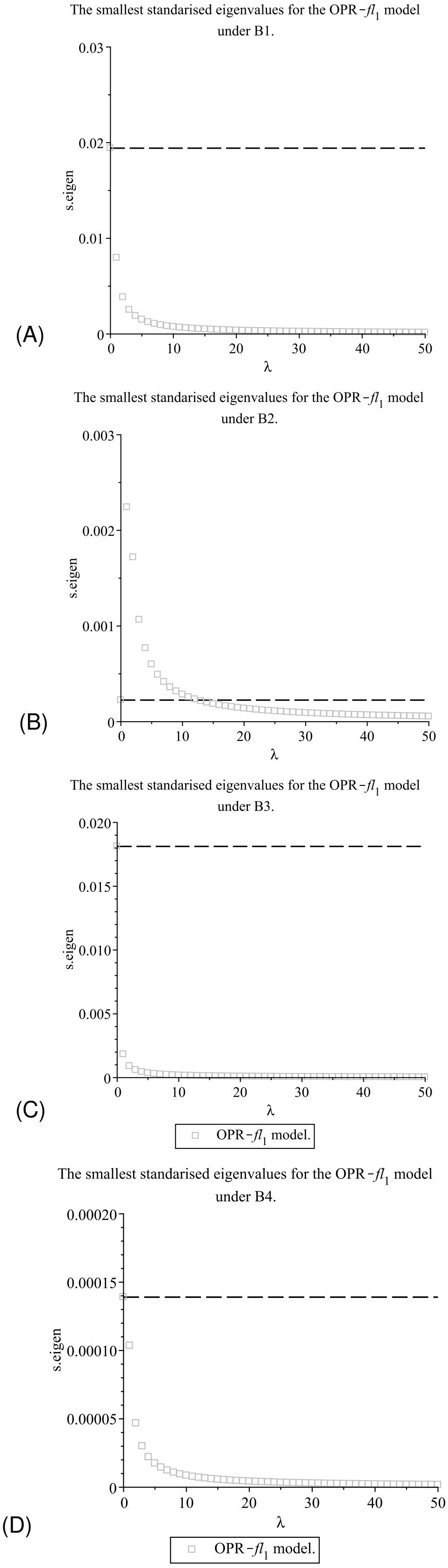


Fig. 5.11 The smallest standardised eigenvalues (s.eigen) vs. λ for the OPR$-fl_1$ models under simulation scenarios B1(A), B2(B), B3(C) and B4(D). The black horizontal dash lines are the s.eigen values obtained from the OPR model.

5.4 Simulation study

In this section we use simulation to assess the performance of the proposed models. We simulate data sets using two simulation settings, *Setting A* and *Setting B*. For each simulation setting, we simulate 100 data sets with $K = 20$ sampling occasions to examine the precision of the estimates obtained from maximising the three penalised likelihoods. Population size is set equal to 1000 or 300 for each simulation scenario. The true values of parameters used for the different simulation settings are given in Table 5.3. For each simulated data, we generate 200 training and validation sets and calculate the MSE for cross-validation. The estimates with minimum MSE are chosen for the corresponding simulated data sets. Lists of simulation scenarios and results are shown in the subsequent subsections.

5.4.1 Setting A

For OPR-l_1 and OPR-l_2 models, we consider the following simulation scenarios A1-A4.

- Scenario A1: No zero in β and low capture probability.
- Scenario A2: No zero in β and high capture probability.
- Scenario A3: Six zeros in β and low capture probability.
- Scenario A4: Six zeros in β and high capture probability.

Results of estimated capture probability p and population size N are shown in Figures 5.12 and 5.13 for $N = 1000$ and $N = 300$ individuals respectively. As shown in Figure 5.12, unbiased estimates of both capture probability and population size are obtained from the OPR-l_1 model when $N = 1000$. However, when $N = 300$ the OPR-l_1 model struggles to obtain the true values of parameters used, and instead obtains biased estimates that are often on the boundary under Scenarios A1 and A2 (Figure 5.13). In contrast, when there are more zeros in β under Scenarios A3 and A4 when $N = 300$, the frequency of boundary estimates reduces, as suggested in Figure 5.13. However, those estimates are negatively biased for capture probability and hence positively biased for population size.

For the OPR-l_2 model, capture probability is slightly overestimated and therefore population size is underestimated across all scenarios we considered. Figures 5.12 and 5.13 suggest that the variability of the estimates is larger when $N = 300$ compared to the case when $N = 1000$. However, increasing the population size cannot improve the accuracy of parameters in the OPR-l_2 model.

5.4.2 Setting B

This simulation setting is performed to assess the performance of the OPR-fl_1 model.

- Scenario B1: No equal in β and low capture probability.
- Scenario B2: No equal in β and high capture probability.
- Scenario B3: Ten equals in β and low capture probability.
- Scenario B4: Ten equals in β and high capture probability.

Figures 5.14 and 5.15 illustrate that the variability in the estimated capture probabilities is lower when there are $N = 1000$ individuals in the population compared to the cases when $N = 300$. However, the median values of the estimates of capture probabilities are slightly greater than the true values across all scenarios. Positive tails in the distribution of estimated p are clearly noticeable when capture probability is low under Scenarios B1 and B3, where longer tails are observed when population size is smaller. The estimates of population size are all underestimated and have negative tails for scenarios with low capture probabilities.

We have explored the performance of the three proposed penalized likelihood methods using simulations. The new approaches show improvement over the maximum likelihood estimation for all scenarios we considered. However, we only observe unbiased estimates for the OPR-l_1 model when population size is big enough, and none of the proposed penalties performs well for small population sizes. We conclude that the OPR-l_1 model can be used when N is large.

Table 5.3 True values of parameters used for different simulation setting.

Setting	Model	Scenario	p	N	β
A	OPR-l_1/OPR-l_2	A1	0.3	1000/300	(0.05, 0.05, 0.03, 0.05, 0.05, 0.05, 0.1, 0.05, 0.2, 0.1, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.1, 0.01)
		A2	0.7	1000/300	(0.05, 0.05, 0.03, 0.05, 0.05, 0.05, 0.1, 0.05, 0.2, 0.1, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.1, 0.01)
		A3	0.3	1000/300	(0.05, 0.05, 0.02, 0, 0, 0 , 0.15, 0, 0.3, 0.1, 0.02, 0.02, 0.02, 0.02, 0.02 , 0.02, 0, 0, 0.2, 0.01)
		A4	0.7	1000/300	(0.05, 0.05, 0.02, 0, 0, 0 , 0.15, 0, 0.3, 0.1, 0.02, 0.02, 0.02, 0.02, 0.02 , 0.02, 0, 0, 0.2, 0.01)
B	OPR-fl_1	B1	0.3	1000/300	(0.01, 0.055, 0.06, 0.07, 0.1, 0.15, 0.021, 0.04, 0.08, 0.03, 0.01, 0.0120, 0.043, 0.015, 0.003, 0.09, 0.08, 0.02, 0.08, 0.031)
		B2	0.7	1000/300	(0.01, 0.055, 0.06, 0.07, 0.1, 0.15, 0.021, 0.04, 0.08, 0.03, 0.01, 0.0120, 0.043, 0.015, 0.003, 0.09, 0.08, 0.02, 0.08, 0.031)
		B3	0.3	1000/300	(0.05, 0.05, 0.03, 0.05, 0.05, 0.05, 0.1, 0.05, 0.2, 0.1, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.1, 0.01)
		B4	0.7	1000/300	(0.05, 0.05, 0.03, 0.05, 0.05, 0.05, 0.1, 0.05, 0.2, 0.1, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.02, 0.1, 0.01)

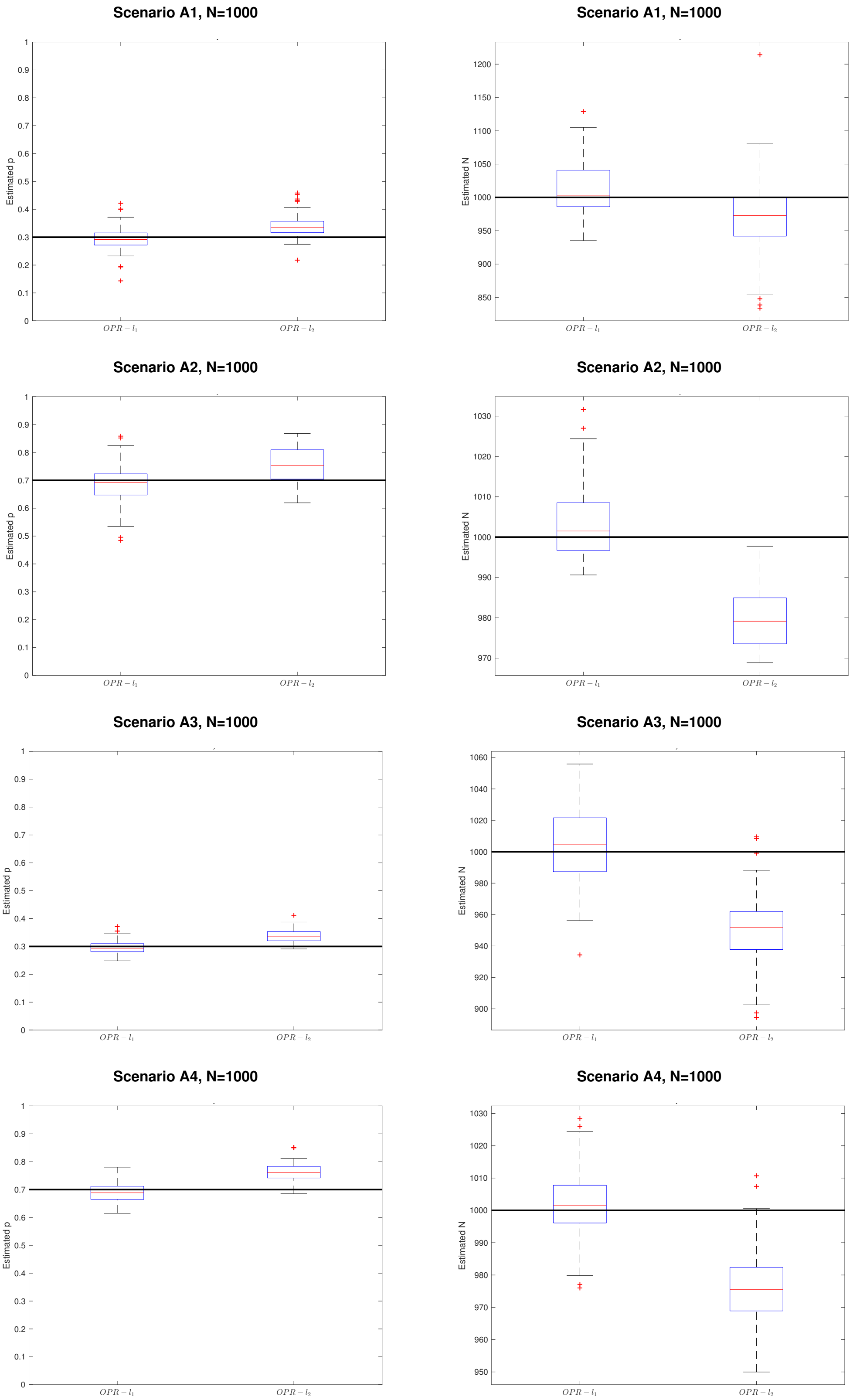


Fig. 5.12 Estimates of capture probability p (left) and population size N (right) obtained from the OPR-l_1 and OPR-l_2 models for simulations with $N = 1000$ individuals and $K = 20$ sampling occasions under Scenarios A1 to A4. The black lines indicate the true values of parameters.

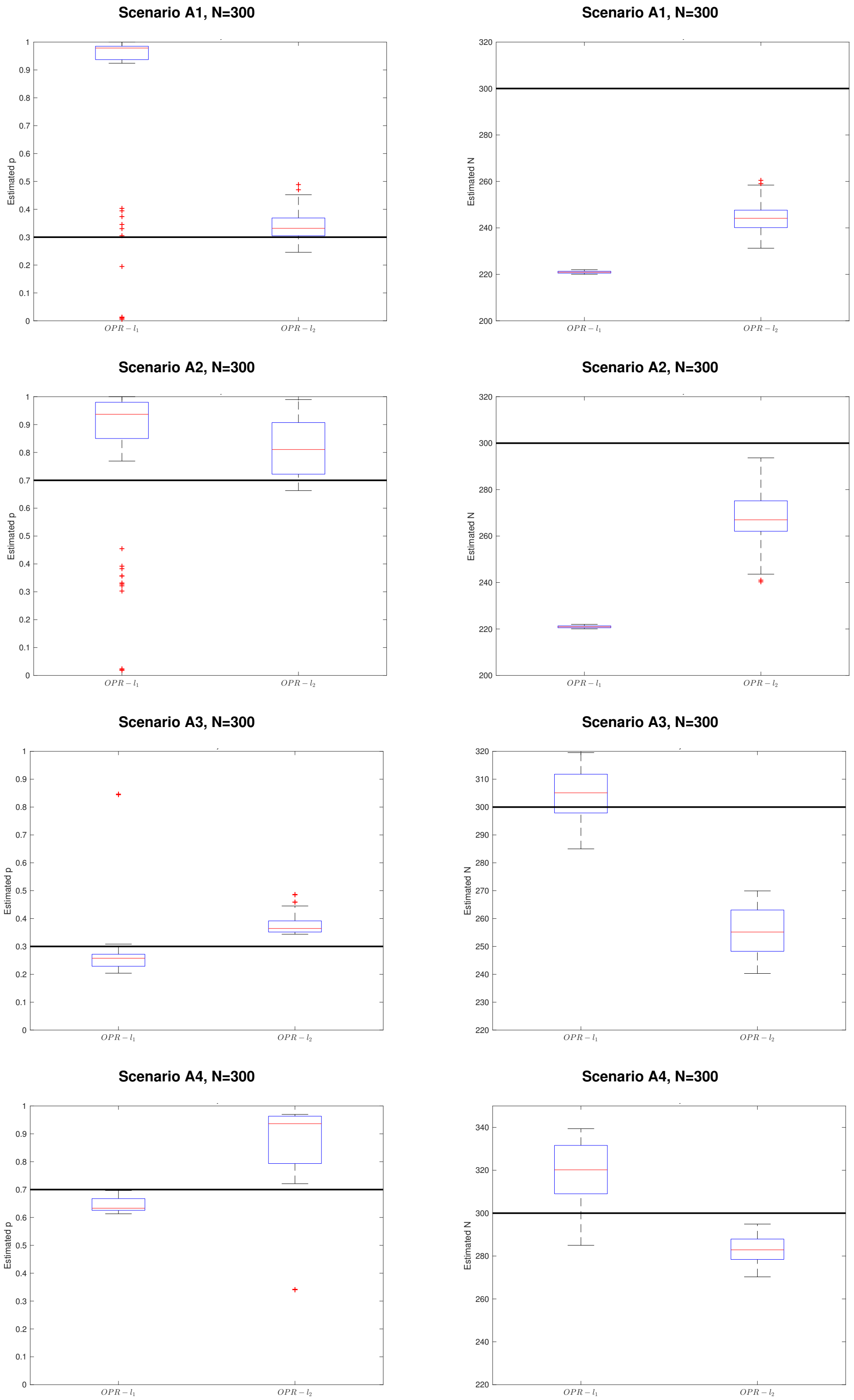


Fig. 5.13 Estimates of capture probability p (left) and population size N (right) obtained from the OPR-l_1 and OPR-l_2 models for simulations with $N = 300$ individuals and $K = 20$ sampling occasions under Scenarios A1 to A4. The black lines indicate the true values of parameters.

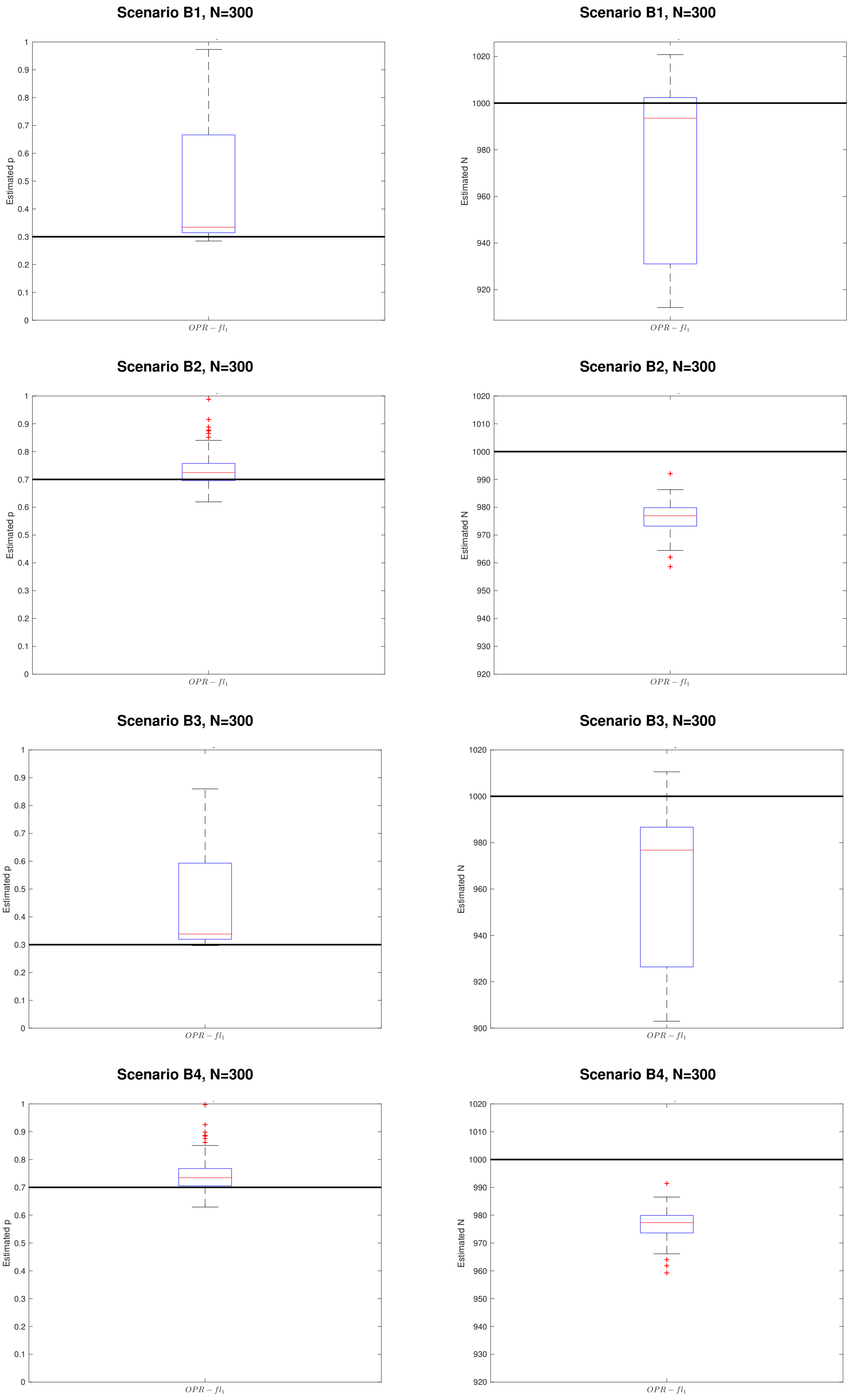


Fig. 5.14 Estimates of capture probability p (left) and population size N (right) obtained from the OPR-fl_1 model for simulations with $N = 1000$ individuals and $K = 20$ sampling occasions under Scenarios A1 to A4. The black lines indicate the true values of parameters.

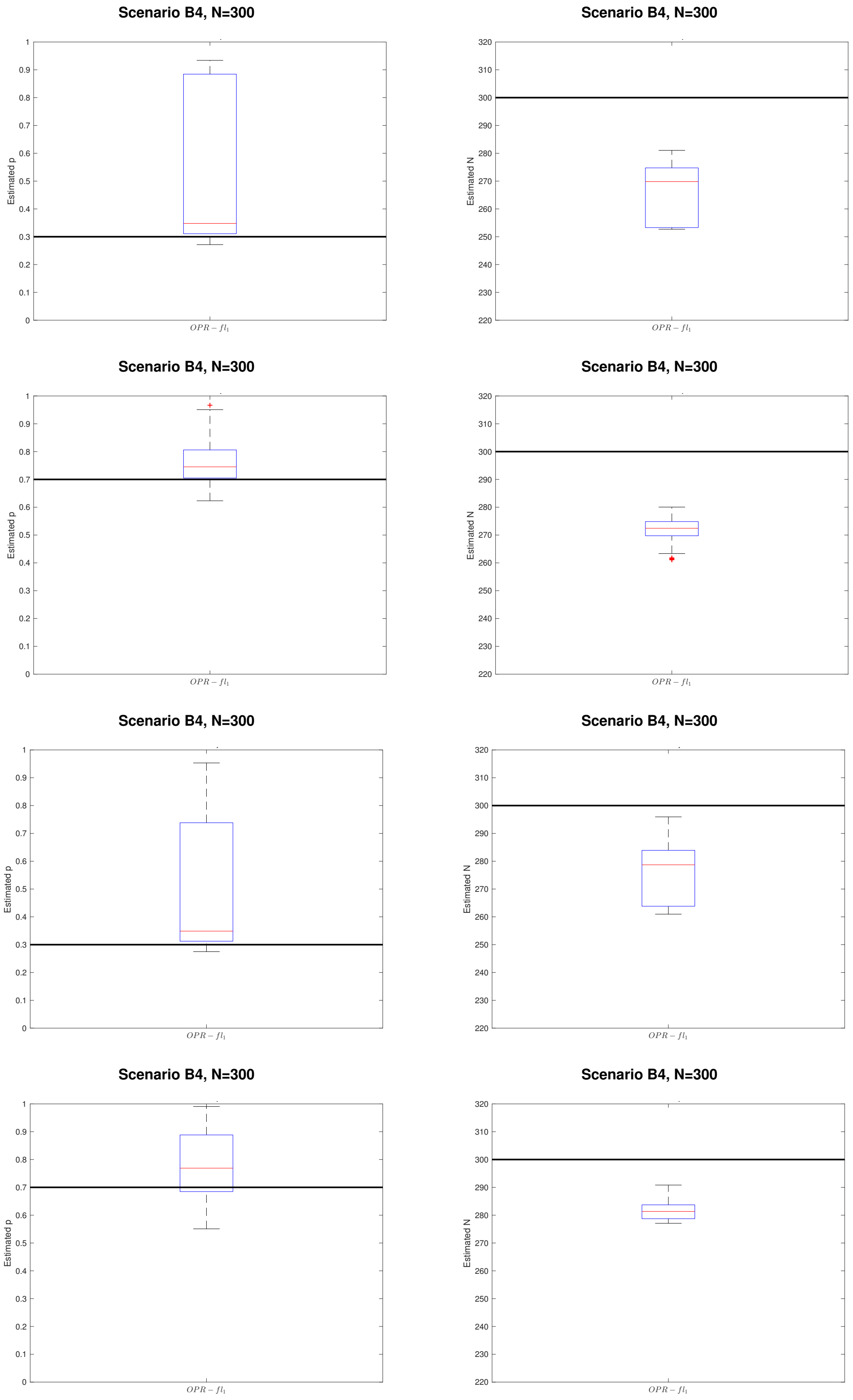


Fig. 5.15 Estimates of capture probability p (left) and population size N (right) obtained from the OPR-fl_1 model for simulations with $N = 300$ individuals and $K = 20$ sampling occasions under Scenarios A1 to A4. The black lines indicate the true values of parameters.

Table 5.4 Median relative bias (MBias) in estimates $\hat{p}$, $\hat{\psi}$ and $\hat{N}$ from the geometric removal model under different Scenarios. The true values of parameters p_{true} and N_{true} are listed below. The true values for the entry parameters used are shown in Table 5.3.

Scenario	p_{true}	N_{true}	$\hat{p}_{\text{MBias}}$	$\hat{N}_{\text{MBias}}$
A1	0.3	1000	-0.99	2.06×10^7
A1	0.3	300	-0.99	2.79×10^6
A2	0.7	1000	-0.97	2.31×10^3
A2	0.7	300	-0.97	6.58×10^2
A3	0.3	1000	-0.99	2.39×10^9
A3	0.3	300	-0.99	3.24×10^0
A4	0.7	1000	-0.99	2.04×10^8
A4	0.7	300	-0.99	7.01×10^7
B1	0.3	1000	-0.99	1.34×10^8
B1	0.3	300	-0.99	2.93×10^7
B2	0.7	1000	-0.97	2.36×10^3
B2	0.7	300	-0.98	1.25×10^3
B3	0.3	1000	-0.99	2.06×10^7
B3	0.3	300	-0.99	2.79×10^6
B4	0.7	1000	-0.97	2.31×10^3
B4	0.7	300	-0.97	6.58×10^2

5.4.3 Results from the geometric and OPR models

In addition to investigating the performance of the OPR$-l_1$, OPR$-l_2$ and OPR$-fl_1$ models using simulation, we also conduct simulations for the geometric removal model and the OPR model under each simulation setting.

We show median relative bias obtained from the geometric removal models using data simulated under simulation Settings A and B in Table 5.4. We observe that population sizes are overestimated and the estimates of constant capture probabilities are extremely small compare to the true values of parameters across all scenarios. As a result, we demonstrate that the geometric removal model underestimates the constant capture probability and yields extremely large estimates of population sizes if new arrivals of individuals are ignored in the study.

In Table 5.5 we show simulation results of the OPR model for data simulated using the true parameters in Table 5.3. It is clear that the capture probabilities p are overestimated across all scenarios because the estimates of p are always on the boundary of one due to near parameter redundancy. Similarly, the OPR model underestimates the

Table 5.5 Median relative bias (MBias) in estimates $\hat{p}$, $\hat{\psi}$ and $\hat{N}$ from the OPR model under different Scenarios. The true values of parameters p_{true} and N_{true} are listed below. True entry parameters used are shown in Table 5.3.

Scenario	p_{true}	N_{true}	$\hat{p}_{\text{MBias}}$	$\hat{N}_{\text{MBias}}$
A1	0.3	1000	2.33	-8.35×10^{-2}
A1	0.3	300	2.34	-8.33×10^{-2}
A2	0.7	1000	0.43	-1.29×10^{-2}
A2	0.7	300	0.42	-1.33×10^{-2}
A3	0.3	1000	2.33	-1.23×10^{-1}
A3	0.3	300	2.33	-1.19×10^{-1}
A4	0.7	1000	0.42	-2.09×10^{-2}
A4	0.7	300	0.43	-2.00×10^{-2}
B1	0.3	1000	2.33	-1.08×10^{-1}
B1	0.3	300	2.33	-1.10×10^{-1}
B2	0.7	1000	4.29	-1.20×10^{-2}
B2	0.7	300	4.28	-1.20×10^{-2}
B3	0.3	1000	2.33	-8.35×10^{-2}
B3	0.3	300	2.34	-8.33×10^{-2}
B4	0.7	1000	0.43	-1.29×10^{-2}
B4	0.7	300	0.42	-1.33×10^{-2}

population sizes as shown in Table 5.5. Therefore, the OPR model cannot be used in practice.

5.5 Data analysis

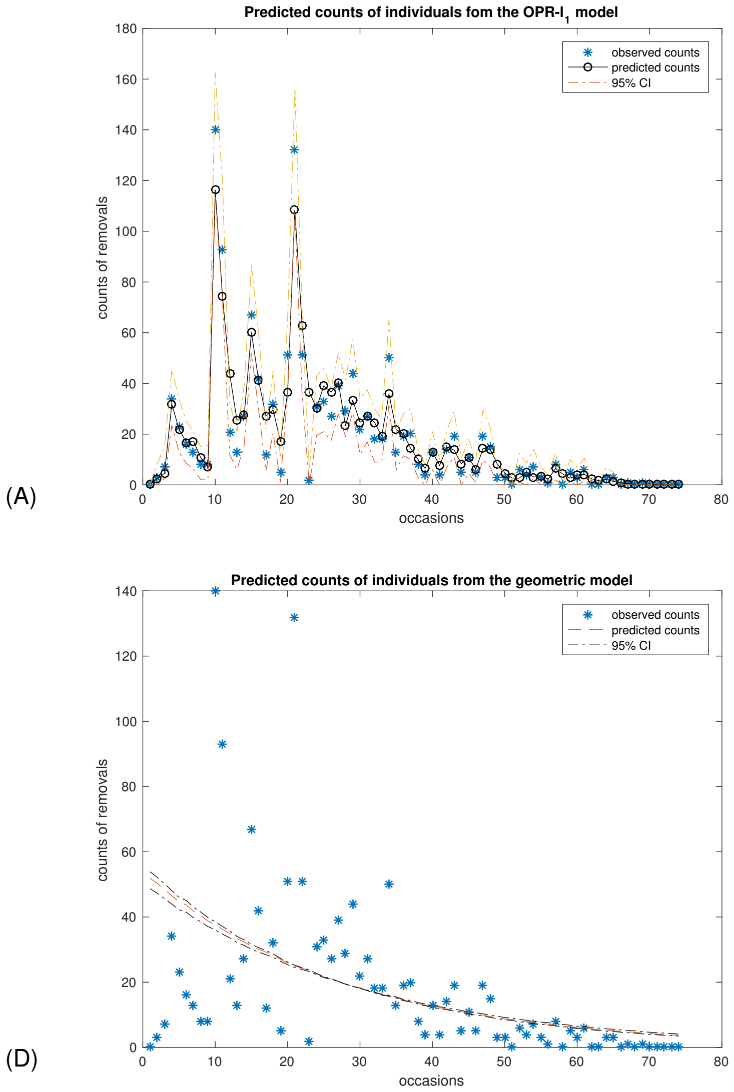
In this section we present results for the real data analysed using the OPR-l_1 and geometric removal models, as simulation results indicate only the OPR-l_1 model can produce unbiased estimates when population size is large enough. Translocation of common lizards, *Zootoca vivipara*, was conducted from the 22th of July 2016 to the 17th of October 2016 before a land development project commenced at Discovery park in Sandwich. 1364 common lizards were captured and removed from the pathway of the development site over $K = 74$ removal sampling occasions. As we observed more than a thousand individuals which is greater than the population size used for simulation (N=1000), we fit the OPR-l_1 model to the real data. Furthermore, there is no covariate (e.g. temperature or humidity) information collected during the study. Standard errors and confidence intervals are computed using bootstrap (100 resamples).

The fitted counts of common lizards removed at each sampling occasion and corresponding 95% confidence intervals are shown in Figure 5.16. The improvement in model fitting of the OPR-l_1 model for the real data we consider is clearly visible compared with the fitted geometric removal model. In Table 5.6 we present the estimated capture probabilities and number of individuals that failed to be captured with standard errors and confidence intervals for both the OPR-l_1 and geometric removal models. The OPR$-l_1$ model suggests that we nearly captured all of the individuals in the population. However, the geometric removal model produces a larger $\hat{n}_0$ suggesting there are still many individuals left behind. This is because the geometric removal model tends to underestimate p and overestimate population size as we mentioned in Section 5.4.3.

Additionally, the estimates of entry parameters β and the corresponding 95% confidence intervals are displayed in Figure 5.17. Adult common lizards emerge from their hibernation sites in early spring. Additionally, birth of common lizard typically occurs around July and August and most young individuals are born from late June to early September (Edgar et al., 2010). Therefore, the estimated peaks in the entry parameter are more likely due to new birth of common lizard rather than emergence from hibernation.

Table 5.6 Estimates and their standard errors (SE) and 95% confidence intervals (95% CI) for the OPR-l_1 and geometric removal models for the common lizards data. $\hat{n}_0$ is the estimate of the number of individuals we failed to capture by the end of the study. $\hat{p}$ is the estimated capture probabilities.

Model	OPR-l_1	Geometric
$\hat{p}$	0.793	0.035
SE	0.182	9.703×10^{-4}
95% CI	(0.483, 0.980)	(0.033, 0.037)
$\hat{n}_0$	2.450	106.923
SE	1.326	8.230
95% CI	(2.023, 5.321)	(88.696, 119.160)


Predicted counts of individuals fom the OPR-l_1 model

Predicted counts of individuals from the geometric model

Fig. 5.16 Results of predicted counts of individuals removed at each occasion, shown by the circle line. Stars are the real data. 95% confidence intervals obtained from 100 bootstrap samples are indicated within dotted dashed lines. (A) Fitted counts from the OPR-l_1 model. (D) Fitted counts from the geometric removal model.

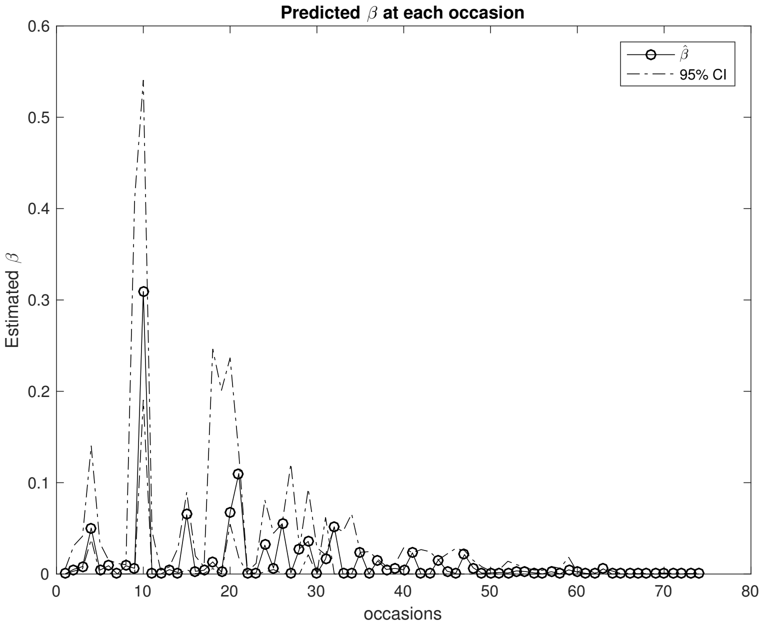


Fig. 5.17 Results of estimated entry parameters β, indicated by the circle lines connected by the solid line. 95% confidence intervals obtained from 200 bootstrap samples are indicated within the dotted dashed lines.

5.6 Conclusion

Populations are rarely closed when we conduct sampling in the study area. Previous studies suggest the estimation of temporary emigration (Zhou et al., 2018) and recruitment (Gould and Pollock, 1997) in the context of removal sampling relies on the use of robust design. Chapter 5 develops new approaches to model new births/arrivals in removal sampling without robust design, which consists a possible violation of the closure assumption in the geometric removal model (Moran, 1951). We show that penalised maximum likelihood estimation reduces the frequency of boundary estimates obtained from optimising the likelihood without a penalty term. Additionally, penalised likelihood estimation provides an alternative to removal models analysed in a Bayesian framework (Matechou et al., 2016) when modelling new arrivals of individuals in the classic framework without prior information

We investigate the ways to improve parameter estimates and reduce the incidence of boundary estimates for both big and small data sets using simulation. When there is a large number of individuals in the population, the penalised likelihood of the OPR-l_1 model produces unbiased estimates. However, both the OPR-l_2 and OPR-fl_1 models exhibit biased estimates. The performance of all of the proposed models becomes worse for populations with small sizes.

Small data sets are not uncommon in ecology due to practical constraints and limited budgets for fieldwork. Our results show that for small data sets, all three penalised likelihood methods exhibit positive bias in capture probabilities, which results in negative bias in population sizes. However, all of the new approaches are considerably less biased than the non-penalised maximum likelihood estimation. We conclude that, although the penalised likelihood methods are potentially useful for removal data, they have to be used with caution in cases of small population sizes. When population size is expected to be small (e.g. few hundreds), we recommend to sample individuals using the robust design protocol so that we can use the new robust design removal models discussed in Chapter 3.

There are also other concerns regarding the use of penalised likelihood methods. Firstly, they require the use of cross-validation to choose a value for the tuning parameter λ. Therefore, estimation can be time-consuming if there are a large number of candidate values for λ. We suggest to start with several λ, then narrow down the possible intervals of λ based on the MSE criteria and repeat the process until a minimum MSE is found for the defined accuracy of λ (e.g. single digit). Secondly, we cannot directly compute standard errors and confidence intervals from penalised likelihood estimation using standard asymptotic theory, as in maximum likelihood estimation. Bootstrapping can be used instead, however, it can be computationally intensive to perform the cross-validation process for all of the boostrapped resamples, which may be not feasible without access to high performance computing facilities.

We have discussed the difficulties of fitting the proposed models to data sets without robust design. However, if the robust design is additionally incorporated in removal sampling, the observed data can be analysed using the model in Zhou et al. (2018) where maximum likelihood estimation is much quicker in practice. Therefore, a study should be designed carefully before data collection begins, as solving issues with a good study design is always preferable than modelling the way out of it. In Chapter 6, we will investigate the sampling design aspects of robust design.

Chapter 6

Optimal Design for Robust Design Removal Sampling

6.1 Introduction

Statistical design has been recognised as a crucial component of animal population studies (Morrison et al., 2008; Williams et al., 2002). Although many ecologists have experience with the design of natural resource surveys from a classical perspective (Thompson, 2002), a lot of them have been in a situation where some problems with their collected data could be traced back to a flaw in study design. To ensure that studies provide meaningful results and therefore valuable resources are not wasted, it is crucial to pay attention to survey design (Legg and Nagy, 2006). It is not only important to design the study so that biologically significant results can be observed, but also to ensure that this is achieved in an efficient way.

Bohrmann and Christman (2013) recently explored how to optimally allocate total sampling effort for multiple depletion sites by maxisiming the Fisher information of the constant capture probability in the classic removal model. Their optimal design for removal sampling assumes that all individuals are available for capture throughout the study. However, this assumption will be violated if individuals temporarily emigrate to an area outside the survey site during the study. We found that no work has been done for designing removal surveys taking into account temporary emigration of individuals.

In Chapter 3 we developed new removal models under the robust design for modelling temporary emigration and we showed that the best performing models consider at least the combination of constraints “R” (random movement constraint) and “S” (initial state constraint). In Chapter 6 we will use the structure of the RMER.SRC model, a robust design removal model with constraints “R”, “S” and a constant capture probability, to explore how to plan removal sampling. We assume an equal number of secondary samples for each primary period and a fixed total sampling effort. The aim of the research is to consider how to optimally allocate the total sampling effort for the removal sampling under the robust design sampling protocol, i.e. to determine how many primary periods and how many repeated samples within each primary period to include in the sampling for a fixed level of total survey effort.

Optimal design theory provides useful tools to determine the accuracy of statistical models without additional costs before we conduct the study. Optimal designs have been developed explicitly for many models (Pukelsheim, 2006; Atkinson et al., 2007). Several algorithms have been developed for their numerical construction if the optimal designs in particular cases are not available in explicit form (Yu, 2010; Yang et al., 2013). We compare several approaches in this chapter where we compare the results obtained using four criteria, A-optimality, D-optimality, E-optimality and T-optimality. The results will help ecologists to plan future removal studies with the use of the robust design, improving the precision of parameters for a fixed level of total survey effort.

We describe the formulation of the likelihood function of the RMER.SRC model using the machinery of the forward algorithm in hidden Markov models in Section 6.2. In Section 6.3 we explore four study design criteria analytically and derive theoretical results for each of those criteria. We evaluate the theoretical results obtained in section 6.3 using simulations and present the results of a simulation study in Section 6.4. Conclusions and a discussion are given in Section 6.5.

6.2 Formulation

6.2.1 Description and notation

Suppose the total survey effort is $K = Tk$, where T is the number of primary periods and k is the number of secondary samples within each primary period. The population size at a depletion site is denoted by N, which represents the number of individuals that are exposed to sampling effort at least once during the removal study. The removal data set is a matrix with entry $n_{i,j}$ representing the number of individuals removed at the jth secondary sample within the ith primary period, where $i = 1, \ldots, T$ and $j = 1, \ldots, k$. Once the individuals are caught, they are permanently removed from the study area. The total number of individuals removed is denoted by $n^{\dagger}$, where $n^{\dagger} = \sum_{i=1}^{T} \sum_{j=1}^{k} n_{i,j}$. Furthermore, n_0 denotes the number of individuals that are missed and never captured during the study, where $n_0 = N - n^{\dagger}$.

We suppose there are two states in the model, denoted by s: individuals in state 1 ($s = 1$) are present and available for removal, while individuals in state 2 ($s = 2$) are absent from the study area and therefore unavailable for removal. Transitions between the two states are only allowed between primary sampling periods and the population is assumed to be closed within each primary occasion.

We only consider constant parameters in the removal model accounting for temporary emigration in this derivation, because we want to minimise the number of parameters considered in the model and simplify the complexity of the chosen design criterion. Although in practice this simplification may not always be reasonable, it is convenient to provide general study design guidelines. We define π to be the proportion of individuals in state 1 at the beginning of the study and its complement, $1-\pi$, as the proportion of the population in state 2. The probability of individuals transitioning from state 1 to state 2 is denoted by ϕ^{12} in Chapter 3. In this chapter we use ϕ for simplicity. Similarly, the transition probability from state 2 to state 1 is denoted by $\phi^{'}$. Let p denote a constant capture probability over time. Note that individuals can only be captured and removed if they are in state 1, so $p = 0$ when individuals stay in state 2.

6.2.2 Constraints

In addition to assuming parameters are constant over time, we consider the following parameter constraints in order to reduce the number of parameters in the model and ease the derivation of the Fisher information in Section 6.3:

- $\pi = \phi^{'}/(\phi + \phi^{'})$, the initial state parameter π is constrained using this expression if we assume that the initial state distribution $\pi = (\pi, 1-\pi)$ is the stationary distribution of the chain.
- $\phi + \phi^{'} = 1$. This constraint is equivalent to the random emigration/movement model for capture-recapture data sampled with robust design as described in Kendall et al. (1995, 1997). It suggests that the probability of being in the unobservable state between the ith and $(i+1)$th primary session is the same for individuals in and individuals outside the study area. In this case, we treat the transition probability ϕ as a free parameter to be estimated in the model, and the transition probability $\phi^{'}$ is reparameterized using the constraint, i.e. $\phi^{'} = 1-\phi$. We note that with this constraint, the state that individuals stayed in at the $i+1$th primary period is independent of the state individuals resided

in at the ith primary period. Therefore, the model becomes a special case of an independent mixture model.

The use of these constraints for removal models with robust design is discussed in Chapter 3 and Zhou et al. (2018), where it has been demonstrated that we need to consider at least the combination of these two constraints to obtain unbiased estimates of parameters.

6.2.3 The likelihood function

We adopt the forward algorithm for hidden Markov models for removal data (Lystig and Hughes, 2002; Turner, 2008) and calculate the likelihood function recursively as shown below. When deriving sampling design criteria in Section 6.3, we use symbolic calculation in `Matlab`. As symbolic computation is slow and the likelihood is complicated, obtaining the likelihood function recursively can reduce the execution times and memory requirement for the limited computational memory.

The probability of individuals being removed at the jth secondary sample within the ith primary period is denoted by $\boldsymbol{\alpha}_{i,j}$. The forward algorithm for hidden Markov models is implemented for the removal data to calculate the likelihood $\boldsymbol{\alpha}_{i,j}$ recursively in terms of p and ϕ.

We can define the probability of an individual being removed at the first secondary occasion in the first primary period to be,

$$\boldsymbol{\alpha}_{1,1} = \pi p.$$

Define $\boldsymbol{\beta}_{i,j}(s)$ to be the probability of an individual in state s but not removed at the jth secondary sample within the ith primary period. Then, $\boldsymbol{\beta}_{1,1}(s)$ are formed as

$$\boldsymbol{\beta}_{1,1}(s=1) = \boldsymbol{\alpha}_{1,1}\frac{1-p}{p} = \pi(1-p)$$

$$\boldsymbol{\beta}_{1,1}(s=2) = 1 - \frac{\boldsymbol{\alpha}_{1,1}}{p} = 1-\pi.$$

Similarly, $\boldsymbol{\alpha}_{1,j}$ and $\boldsymbol{\beta}_{1,j}$ for the second and subsequent secondary samples within the first primary period, can be computed by

$$\boldsymbol{\alpha}_{1,j} = \boldsymbol{\alpha}_{1,j-1}\frac{1-p}{p}p = \boldsymbol{\alpha}_{1,j-1}(1-p)$$

$$\boldsymbol{\beta}_{1,j}(s=1) = \boldsymbol{\beta}_{1,j-1}(s=1)(1-p)$$

$$\boldsymbol{\beta}_{1,j}(s=2) = \boldsymbol{\beta}_{1,j-1}(s=1) = 1-\pi$$

for $j = 2,\ldots,k$.

Then $\boldsymbol{\alpha}_{i,1}$ and $\boldsymbol{\beta}_{i,1}$ for the first secondary sample within the second and subsequent primary periods, can be calculated recursively using,

$$\boldsymbol{\alpha}_{i,1} = \Big\{\sum_{r=1}^{2} \boldsymbol{\beta}_{i-1,k}(s=r)\phi^{r1}\Big\}p = \Big\{\sum_{r=1}^{2} \boldsymbol{\beta}_{i-1,k}(s=r)(1-\phi)\Big\}p$$

$$\boldsymbol{\beta}_{i,1}(s=1) = \boldsymbol{\alpha}_{i,1}\frac{1-p}{p}$$

$$\boldsymbol{\beta}_{i,1}(s=2) = \sum_{u=1}^{2} \boldsymbol{\beta}_{i-1,k}(s=u)\phi^{u2} = \sum_{u=1}^{2} \boldsymbol{\beta}_{i-1,k}(s=u)\phi$$

as $\phi^{r1} = 1-\phi$ for $r = 1,2$ and $\phi^{u2} = 1-\phi$ for $u = 1,2$ for $i = 2,\ldots,T$ and $j = 1$.

Therefore $\boldsymbol{\alpha}_{i,j}$ and $\boldsymbol{\beta}_{i,j}$ for the second and subsequent secondary samples within the second and subsequent primary periods are,

$$\boldsymbol{\alpha}_{i,j} = \boldsymbol{\alpha}_{i,j-1}\frac{1-p}{p}p = \boldsymbol{\alpha}_{i,j-1}(1-p)$$

$$\boldsymbol{\beta}_{i,j}(s=1) = \boldsymbol{\beta}_{i,j-1}(s=1)(1-p)$$

$$\boldsymbol{\beta}_{i,j}(s=2) = \boldsymbol{\beta}_{i,j-1}(s=2)$$

for $i = 2,\ldots,T$ and $j = 2,\ldots,k$.

The full multinomial likelihood function in terms of p and ϕ is defined as,

$$L(p,\phi;n_{i,j}) = \frac{(n_0+n^{\dagger})!}{n_0!\cdot\prod_{i=1}^{i=T}\prod_{j=1}^{j=k} n_{i,j}!}\cdot\boldsymbol{\alpha}_0^{n_0}\cdot\prod_{i=1}^{i=T}\prod_{j=1}^{j=k}\boldsymbol{\alpha}_{i,j}^{n_{i,j}}.$$

In order to account for the uncertainty in the number of individuals never captured n_0 in the survey design, we derive the conditional probabilities $\boldsymbol{\alpha}^{*}_{i,j}$ which is conditional on the observed individuals (Bohrmann and Christman, 2013), i.e $\boldsymbol{\alpha}^{*}_{i,j} = \frac{\boldsymbol{\alpha}_{i,j}}{\boldsymbol{\alpha}^{\dagger}_{i,j}}$, where $\boldsymbol{\alpha}^{\dagger}_{i,j} = \sum_i^T\sum_j^k \boldsymbol{\alpha}_{i,j}$ is the sum of $\boldsymbol{\alpha}_{i,j}$, so that we can use the law of iterated expectation (Catchpole et al., 2002) when calculating its Fisher information as shown in Section 6.3.1.

6.3 Criteria

In the robust design sampling protocol under consideration, there are different ways to allocate the total sampling effort available: one may survey the study site more intensively (more secondary samples) with less replicated primary periods or vice versa. In this section we derive several criteria for removal sampling within the robust design framework. When there is one parameter in a model, minimising its variance is equivalent to maximising the Fisher information. When there are several parameters, we have to assess the variance matrix or information matrix (the inverse of the variance matrix). There are different criteria that can be used for optimal sampling design when there is more than one parameter of interest. For a discussion of the merits of the different methods see Atkinson and Donev (1992). We investigate four optimality criteria, A-optimality, D-optimality, E-optimality and T-optimality and we aim to optimise these information criteria using real-valued functions.

6.3.1 Fisher information

Let θ be a vector of length two of parameters in the model, i.e. $\theta = [p \quad \phi]$. The element in the ith row and jth column in the Fisher information matrix is defined as the negative expectation of the second derivative of the log-likelihood function with respect to the parameters,

$$[\mathbf{I}(\theta)_{i,j}] = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\{L(p,\phi)\}\right] \tag{6.1}$$

where $i = 1,2$ and $j = 1,2$.

As we only consider two parameters in the model, we can write the Fisher information matrix as

$$\mathbf{I}(\theta) = -\mathrm{E}\left[\begin{pmatrix} \frac{\partial^2}{\partial p^2} & \frac{\partial^2}{\partial p\partial\phi} \\ \\ \frac{\partial^2}{\partial\phi\partial p} & \frac{\partial^2}{\partial\phi^2} \end{pmatrix}\log\{L(p,\phi;n_{i,j})\}\right] = \begin{pmatrix} \mathbf{I}_{p,p} & \mathbf{I}_{p,\phi} \\ \mathbf{I}_{\phi,p} & \mathbf{I}_{\phi,\phi} \end{pmatrix}.$$

We drive $\mathbf{I}_{p,\phi}$, $\mathbf{I}_{p,p}$ and $\mathbf{I}_{\phi,\phi}$ and the results are shown in Equations (6.3) and (6.5) for $\mathbf{I}_{p,\phi}$ and $\mathbf{I}_{p,p}$ respectively. As the result for $\mathbf{I}_{\phi,\phi}$ in Equation (6.6) is similar to that obtained for $\mathbf{I}_{p,p}$, we show the details in Appendix C.

The second step for deriving $\mathbf{I}_{p,\phi}$ and $\mathbf{I}_{p,p}$, as shown in Equations (6.2) and (6.4), is suggested by the law of iterated expectation (Casella and Berger, 2002) to account for the randomness in the total number of individuals observed, $n^{\dagger}$. The law of iterated expectation states that for two random variables A and B, $\mathrm{E(A)} = \mathrm{E\{E(A|B)\}}$, which allows us to derive the expected Fisher information by first computing the expectation of the observations $n_{i,j}$ with respect to the conditional distribution of $n^{\dagger}$, and then use that to derive the Fisher information with respect to the distribution of $n^{\dagger}$.

$$
\begin{aligned}
&\mathbf{I}_{p,\phi}\\
&= -\mathrm{E}\left[\frac{\partial^2}{\partial p \partial \phi}\log\{L(p,\phi)\}\right]\\
&= -\mathrm{E}_{n^{\dagger}}\left[\mathrm{E}_n\left\{\frac{\partial^2}{\partial p \partial \phi}\log\left(\frac{n^{\dagger}!}{\prod_{i=1}^{T}\prod_{j=1}^{k} n_{i,j}!}\cdot\prod_{i=1}^{T}\prod_{j=1}^{k}(\alpha^*_{i,j})^{n_{i,j}}\right)|n^{\dagger}\right\}\right] \qquad (6.2)\\
&= -\mathrm{E}_{n^{\dagger}}\left[\mathrm{E}_n\left\{\frac{\partial^2}{\partial p \partial \phi}\left(\log(n^{\dagger}!) - \sum_{i=1}^{T}\sum_{j=1}^{k}\log(n_{i,j}!) + \sum_{i=1}^{T}\sum_{j=1}^{k} n_{i,j}\cdot\log(\alpha^*_{i,j})\right)|n^{\dagger}\right\}\right]\\
&= -\mathrm{E}_{n^{\dagger}}\left[\mathrm{E}_n\left\{\frac{\partial^2}{\partial p \partial \phi}\sum_{i=1}^{T}\sum_{j=1}^{k} n_{i,j}\cdot\log(\alpha^*_{i,j})|n^{\dagger}\right\}\right]\\
&= -\mathrm{E}_{n^{\dagger}}\left[\mathrm{E}_n\left\{\sum_{i=1}^{T}\sum_{j=1}^{k} n_{i,j}\cdot(\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p\partial\phi} - \frac{1}{(\alpha^*_{i,j})^2}\cdot\frac{\partial\alpha^*_{i,j}}{\partial p}\cdot\frac{\partial\alpha^*_{i,j}}{\partial\phi})|n^{\dagger}\right\}\right]\\
&= -\mathrm{E}_{n^{\dagger}}\left\{\sum_{i=1}^{T}\sum_{j=1}^{k}\mathrm{E}(n_{i,j}|n^{\dagger})\cdot(\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p\partial\phi} - \frac{1}{(\alpha^*_{i,j})^2}\cdot\frac{\partial\alpha^*_{i,j}}{\partial p}\cdot\frac{\partial\alpha^*_{i,j}}{\partial\phi})\right\}\\
&= -\mathrm{E}_{n^{\dagger}}\left\{\sum_{i=1}^{T}\sum_{j=1}^{k} n^{\dagger}\cdot\alpha^*_{i,j}\cdot(\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p\partial\phi} - \frac{1}{(\alpha^*_{i,j})^2}\cdot\frac{\partial\alpha^*_{i,j}}{\partial p}\cdot\frac{\partial\alpha^*_{i,j}}{\partial\phi})\right\}\\
&= -\sum_{i=1}^{T}\sum_{j=1}^{k}\mathrm{E}(n^{\dagger})\cdot\alpha^*_{i,j}\cdot(\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p\partial\phi} - \frac{1}{(\alpha^*_{i,j})^2}\cdot\frac{\partial\alpha^*_{i,j}}{\partial p}\cdot\frac{\partial\alpha^*_{i,j}}{\partial\phi})\\
&= -\sum_{i=1}^{T}\sum_{j=1}^{k} N\cdot\alpha^{\dagger}\cdot\frac{\alpha_{i,j}}{\alpha^{\dagger}}\cdot(\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p\partial\phi} - \frac{1}{(\alpha^*_{i,j})^2}\cdot\frac{\partial\alpha^*_{i,j}}{\partial p}\cdot\frac{\partial\alpha^*_{i,j}}{\partial\phi})\\
&= -\sum_{i=1}^{T}\sum_{j=1}^{k} N\cdot\alpha_{i,j}\cdot(\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p\partial\phi} - \frac{1}{(\alpha^*_{i,j})^2}\cdot\frac{\partial\alpha^*_{i,j}}{\partial p}\cdot\frac{\partial\alpha^*_{i,j}}{\partial\phi})\\
&= N\cdot\left[-\sum_{i=1}^{T}\sum_{j=1}^{k}\alpha_{i,j}\cdot(\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p\partial\phi} - \frac{1}{(\alpha^*_{i,j})^2}\cdot\frac{\partial\alpha^*_{i,j}}{\partial p}\cdot\frac{\partial\alpha^*_{i,j}}{\partial\phi})\right]\\
&= N\cdot\left[-\sum_{i=1}^{T}\sum_{j=1}^{k}\alpha^{\dagger}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p\partial\phi} - \frac{(\alpha^{\dagger})^2}{\alpha_{i,j}}\cdot\frac{\partial\alpha^*_{i,j}}{\partial p}\cdot\frac{\partial\alpha^*_{i,j}}{\partial\phi}\right]. \qquad (6.3)
\end{aligned}
$$

To evaluate the bracketed expressions in Equations (6.3), (6.5) and (6.6), we need to calculate the following derivatives,

$$\frac{\partial^2 \alpha^*_{i,j}}{\partial p \partial \phi}, \frac{\partial \alpha^*_{i,j}}{\partial p}, \frac{\partial^2 \alpha^*_{i,j}}{\partial p^2}, \frac{\partial \alpha^*_{i,j}}{\partial \phi} \text{ and } \frac{\partial^2 \alpha^*_{i,j}}{\partial \phi^2}.$$

They can be computed recursively by elementary calculus (Lystig and Hughes, 2002; Turner, 2008). The bracketed equation of $\mathbf{I}_{p,\phi}$, $\mathbf{I}_{p,p}$ and $\mathbf{I}_{\phi,\phi}$ can be computed over a grid of p and ϕ values and possible integer values of k and T.

$$
\begin{aligned}
&\mathbf{I}_{p,p}\\
&= -\mathrm{E}\left[\frac{\partial^2}{\partial p^2}\log\left\{L(p,\phi)\right\}\right]\\
&= -\mathrm{E}_{n^\dagger}\left[\mathrm{E}_n\left\{\frac{\partial^2}{\partial p^2}\log\left(\frac{n^\dagger!}{\prod_{i=1}^{T}\prod_{j=1}^{k}n_{i,j}!}\cdot\prod_{i=1}^{T}\prod_{j=1}^{k}(\alpha^*_{i,j})^{n_{i,j}})|n^\dagger\right)\right\}\right] \qquad (6.4)\\
&= -\mathrm{E}_{n^\dagger}\left[\mathrm{E}_n\left\{\frac{\partial^2}{\partial p^2}\log(n^\dagger!)-\sum_{i=1}^{T}\sum_{j=1}^{k}\log(n_{i,j}!)+\sum_{i=1}^{T}\sum_{j=1}^{k}n_{i,j}\cdot\log(\alpha^*_{i,j})\middle|n^\dagger\right\}\right]\\
&= -\mathrm{E}_{n^\dagger}\left[\mathrm{E}_n\left\{\frac{\partial^2}{\partial p^2}\sum_{i=1}^{T}\sum_{j=1}^{k}n_{i,j}\cdot\log(\alpha^*_{i,j})|n^\dagger\right\}\right]\\
&= -\mathrm{E}_{n^\dagger}\left[\mathrm{E}_n\left\{\sum_{i=1}^{T}\sum_{j=1}^{k}n_{i,j}\cdot\frac{\partial^2\log(\alpha^*_{i,j})}{\partial p^2}|n^\dagger\right\}\right]\\
&= -\mathrm{E}_{n^\dagger}\left[\mathrm{E}_n\left\{\sum_{i=1}^{T}\sum_{j=1}^{k}n_{i,j}\cdot\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{1}{(\alpha^*_{i,j})^2}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2|n^\dagger\right\}\right]\\
&= -\mathrm{E}_{n^\dagger}\left\{\sum_{i=1}^{T}\sum_{j=1}^{k}\mathrm{E}(n_{i,j}|n^\dagger)\cdot\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{1}{(\alpha^*_{i,j})^2}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2\right\}\\
&= -\mathrm{E}_{n^\dagger}\left\{\sum_{i=1}^{T}\sum_{j=1}^{k}n^\dagger\cdot\alpha^*_{i,j}\cdot\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{1}{(\alpha^*_{i,j})^2}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2\right\}\\
&= -\sum_{i=1}^{T}\sum_{j=1}^{k}\mathrm{E}(n^\dagger)\cdot\alpha^*_{i,j}\cdot\left\{\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{1}{(\alpha^*_{i,j})^2}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2\right\}\\
&= -\sum_{i=1}^{T}\sum_{j=1}^{k}N\cdot\alpha^\dagger\cdot\alpha^*_{i,j}\cdot\left\{\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{1}{(\alpha^*_{i,j})^2}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2\right\}\\
&= -\sum_{i=1}^{T}\sum_{j=1}^{k}N\cdot\alpha^\dagger\cdot\frac{\alpha_{i,j}}{\alpha^\dagger}\cdot\left\{\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{1}{(\alpha^*_{i,j})^2}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2\right\}\\
&= -\sum_{i=1}^{T}\sum_{j=1}^{k}N\cdot\alpha_{i,j}\cdot\left\{\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{1}{(\alpha^*_{i,j})^2}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2\right\}\\
&= N\cdot\left[-\sum_{i=1}^{T}\sum_{j=1}^{k}\alpha_{i,j}\cdot\left\{\frac{1}{\alpha^*_{i,j}}\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{1}{(\alpha^*_{i,j})^2}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2\right\}\right]\\
&= N\cdot\left[-\sum_{i=1}^{T}\sum_{j=1}^{k}\alpha^\dagger\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial p^2}-\frac{(\alpha^\dagger)^2}{\alpha_{i,j}}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial p})^2\right\}\Bigg]. \qquad (6.5)
\end{aligned}
$$

$$
\mathbf{I}_{\phi,\phi} = N\cdot\left[-\sum_{i=1}^{T}\sum_{j=1}^{k}\alpha^\dagger\cdot\frac{\partial^2\alpha^*_{i,j}}{\partial\phi^2}-\frac{(\alpha^\dagger)^2}{\alpha_{i,j}}\cdot(\frac{\partial\alpha^*_{i,j}}{\partial\phi})^2\right\}\Bigg]. \qquad (6.6)
$$

6.3.2 A-optimality

One common approach of study design is to minimise the trace of the inverse of the information matrix, proposed by Chernoff (1953), which results in minimising the average variance of the estimates of parameters (i.e. min{[var(p)+var(ϕ)]/2} = min{var(p)+var(ϕ)}).

The determinant of $\mathbf{I}(\theta)$ is calculated as

$$\begin{aligned}\det[\mathrm{I}(\theta)] &= \mathbf{I}_{p,p}\mathbf{I}_{\phi,\phi} - \mathbf{I}^2_{p,\phi} \\ &= \left[-\mathrm{E}\left\{\frac{\partial^2}{\partial p^2}l(p,\phi)\right\}\right]\left[-\mathrm{E}\left\{\frac{\partial^2}{\partial \phi^2}l(p,\phi)\right\}\right] - \left[-\mathrm{E}\left\{\frac{\partial^2}{\partial p \partial \phi}l(p,\phi)\right\}\right]^2\end{aligned} \tag{6.7}$$

where $l(p,\phi) = \log\{L(p,\phi)\}$.

The inverse of the information matrix is:

$$\begin{aligned}\mathbf{I}(\theta)^{-1} &= \begin{pmatrix} \mathbf{I}_{p,p} & -\mathbf{I}_{p,\phi} \\ -\mathbf{I}_{\phi,p} & \mathbf{I}_{\phi,\phi} \end{pmatrix}^{-1} \\ &= \frac{1}{\det[\mathbf{I}(\theta)]}\begin{pmatrix} \mathbf{I}_{\phi,\phi} & -\mathbf{I}_{p,\phi} \\ -\mathbf{I}_{\phi,p} & \mathbf{I}_{p,p} \end{pmatrix} \\ &= \frac{1}{\det[\mathbf{I}(\theta)]} \quad \mathrm{E}\left[\begin{pmatrix} -\frac{\partial^2}{\partial \phi^2} & \frac{\partial^2}{\partial p \partial \phi} \\ \\ \frac{\partial^2}{\partial \phi \partial p} & -\frac{\partial^2}{\partial p^2} \end{pmatrix} l(p,\phi)\right].\end{aligned} \tag{6.8}$$

The trace of the inverse of the information matrix is:

$$\begin{aligned}\mathrm{tr}[\mathbf{I}(\theta)^{-1}] &= \frac{1}{\det[\mathbf{I}(\theta)]}\left(\mathbf{I}_{\phi,\phi} + \mathbf{I}_{p,p}\right) \\ &= \frac{1}{\det[\mathbf{I}(\theta)]}\left[-\mathrm{E}\left\{\frac{\partial^2}{\partial \phi^2}l(p,\phi)\right\} - \mathrm{E}\left\{\frac{\partial^2}{\partial p^2}l(p,\phi)\right\}\right].\end{aligned} \tag{6.9}$$

This criterion has been used in designing occupancy models (Guillera-Arroita et al., 2010) in terms of a fixed level of sampling effort. It also has been widely discussed in experimental design for clinical trials where subjects receive sequences of treatments (Jones and Donev, 1996; Kerr and Churchill, 2001; Jacroux, 2012).

6.3.3 D-optimality

D-optimality was introduced by Kiefer (1959) and it seeks to maximise the determinant of the information matrix (i.e. det[I(θ)]) of the design, which is equivalent to minimising the determinant of the dispersion matrix, because of the following formula,

$$\det[\mathrm{I}(\theta)]^{-1} = \det[\mathrm{I}(\theta)^{-1}], \tag{6.10}$$

as discussed in (Pukelsheim, 2006).

The inverse of the information matrix, $\mathrm{I}(\theta)^{-1}$, is identified to be the standardised dispersion matrix of the optimal estimator in the parameter space. The determinant of $\mathrm{I}(\theta)^{-1}$, $\det[\mathrm{I}(\theta)^{-1}]$, is the generalised variance, which is the origin of the popularity of this criterion (Pukelsheim, 2006). Therefore D-optimality results in minimizing the generalized variance of the parameter estimates for a model, which is equivalent to minimising the volume of the confidence ellipsoid.

Similar to A-optimality, the use of this criterion has been discussed in designing occupancy models (Guillera-Arroita et al., 2010). Furthermore, Covey-Crump and Silvey (1970) and Yu (2010); Yang et al. (2013) discussed the use of D-optimal designs for multivariable polynomial models.

6.3.4 E-optimality

This design was proposed by Ehrenfeld (1955), and maximises the minimum eigenvalue of the information matrix. The eigenvalues of the information matrix are the roots of the characteristic equation,

$$\begin{aligned}
\det[\mathbf{I}(\theta) - \omega \boldsymbol{E}_2] &= 0 \\
\det\left[\begin{pmatrix} \mathbf{I}_{p,p} & \mathbf{I}_{p,\phi} \\ \mathbf{I}_{\phi,p} & \mathbf{I}_{\phi,\phi} \end{pmatrix} - \omega \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] &= 0 \\
\det\left[\begin{pmatrix} \mathbf{I}_{p,p} - \omega & \mathbf{I}_{p,\phi} \\ \mathbf{I}_{\phi,p} & \mathbf{I}_{\phi,\phi} - \omega \end{pmatrix}\right] &= 0 \\
(\mathbf{I}_{p,p} - \omega)(\mathbf{I}_{\phi,\phi} - \omega) - \mathbf{I}_{p,\phi}^2 &= 0 \\
\omega^2 - (\mathbf{I}_{p,p} + \mathbf{I}_{\phi,\phi})\omega + \mathbf{I}_{p,p}\mathbf{I}_{\phi,\phi} - \mathbf{I}_{p,\phi}^2 &= 0 \qquad (6.11)
\end{aligned}$$

where $\boldsymbol{E}_2$ is the 2×2 identity matrix.

We solve the second order Equation (6.11) to obtain the two eigenvalues of the information matrix. These eigenvalues are then standardised by dividing by the maximum eigenvalues and we aim to maximise the smallest standardised eigenvalue, which guards against the worst possible variance.

This criterion is also discussed in Sections 4.2.4 and 5.3 in evaluating standardised eigenvalues in the derivative matrix for detecting parameter redundancy. It has been used in Markov decision processes (Eugene and Shwartz, 2002; Bonet, 2002) as well as regression models (Pukelsheim and Studden, 1993) but, there is no previous sampling design studies using this criterion in statistical ecology.

6.3.5 T-optimality

This criterion was introduced by Atkinson and Fedorov (1975) in regression models, and maximises the trace of the information matrix. The trace of the information matrix is:

$$\begin{aligned}\mathrm{tr}[\mathbf{I}(\theta)] &= \mathbf{I}_{\phi,\phi} + \mathbf{I}_{p,p} \\ &= -\mathrm{E}\left\{\frac{\partial^2}{\partial \phi^2} l(p,\phi)\right\} - \mathrm{E}\left\{\frac{\partial^2}{\partial p^2} l(p,\phi)\right\}. \end{aligned} \tag{6.12}$$

It is obvious that this criterion is linear (Pukelsheim, 2006; Atkinson and Donev, 1992), so it can be easily affected by either $\mathbf{I}_{\phi,\phi}$ or $\mathbf{I}_{p,p}$ than other three criterions considered. Although this criterion is week, it can be useful if accompanied by further conditions which prevent it from going astray (Pukelsheim, 2006). From the prospective of computational complexity, this criterion is particularly simple to evaluate since it only requires the computation of the diagonal entries of the information matrix.

We find no sampling design studies in statistical ecology using this criterion. However, there are numerous applications such as chemistry of pharmaceutics (Asprey and Macchietto, 2000; Foo and Duffull, 2011). In addition, Dette et al. (2012) observe clear benefits with respect to the power of the F-test for the T- optimal design in cubic regression models.

6.3.6 Results

We calculate the criteria values using all four optimality criteria for a grid values of p and ϕ $(= 0.1, 0.2, \cdots, 0.9)$ for a study with $K = 20$ sampling occasions, where the number of secondary periods per primary period is $k = 2, 4, 5, 10$. The optimal number of secondary samples for different values of p and ϕ is summarised in Table 6.1, which describes analytically how estimator precision is affected by changes in the true values of the design parameters.

In the robust design sampling protocol under consideration there are different ways to allocate the available sample effort. We could survey more primary periods (e.g. days) with less secondary replication within each day or vice versa. Increasing the number of secondary samples has two opposing effects. On the one hand, it provides more information for capture probability p and therefore capture probability is estimated more precisely. On the other hand, it reduces the number of primary periods, so there is less chance to study the transitions between study area and the unobservable state. Additionally, individuals are assumed to remain at their current state within each primary period for a longer period of time (more secondary samples).

Some general observations can be drawn from the results in Table 6.1. On the one hand, the higher the capture probability the lower the optimal number of secondary samples, which is expected as fewer secondary replications are necessary to observe the transition patterns between two states. When p is high, the optimal design suggests $k = 2$ more frequently, which means any effort invested in extra secondary samples ($k > 2$) is wasted. On the other hand, the lower ϕ, i.e. fewer individuals emigrate to the unobservable state, the larger the optimal k is. The extra variance due to capture probability has more impact for low ϕ, so it is more beneficial to have more secondary samples in these cases. Therefore, we generally recommend sampling more days less intensively for rare species, which tend to stay outside the study area. Similarly, surveying fewer days more intensively is optimal for common species, which tend to be present in the study area.

For each fixed value of ϕ, A-optim, D-optim and E-optim all suggest that we need to survey the study area more intensively (i.e. with more secondary samples) within each primary period when capture probability p is expected to be lower. This fact agrees with results of the classic removal model as investigated in Bohrmann and Christman (2013), where they observed that the it is more efficient to increase the number of samples collected when p tends to be low as that way we obtain more information about p and therefore minimize the uncertainty of population size N. Similar results are obtained in Guillera-Arroita et al. (2010) for study design of occupancy models, where they demonstrated that the optimal number of replicates increases with lower probability of detection for each possible value of occupancy probability.

For a fixed value of p, A-optimality, D-optimality and E-optimality designs suggest that the optimal number of secondary samples increases when there are fewer individuals that temporarily emigrate to an unobservable area, i.e. when ϕ is lower. Therefore when more individuals tend to be present in the sampling area and become available for capture across the study, it is optimal to conduct more secondary surveys within each primary period. On the other hand, when most individuals in the population tend to temporarily stay in a hidden area, we need to reduce the number of secondary samples, and therefore increase the number of primary periods to give those individuals more chances of moving back to the study area between primary periods and becoming exposed to sampling effort. Guillera-Arroita et al. (2010) observed similar conclusions for occupancy models, where they suggest to increase the number of samples at each sampling site when occupancy probability is expected to be high.

However, the T-optimality criterion does not show similar patterns with the other three criteria except from the parameter space of high capture probability ϕ and low p. Therefore, it is critical to consider the determinant of $\mathbf{I}(\theta)$ for the RMER.SRC model for sampling design.

	A-optimality								
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	10	10	10	10	10	10	5	5	5
0.2	10	10	10	10	5	5	5	5	4
0.3	10	10	10	5	5	5	5	4	2
0.4	10	10	10	5	5	5	4	4	2
0.5	10	10	5	5	5	4	4	4	2
0.6	10	5	5	5	4	4	4	2	2
0.7	5	5	5	4	4	4	2	2	2
0.8	5	5	5	4	4	2	2	2	2
0.9	5	4	4	4	4	2	2	2	2
	D-optimality								
0.1	10	10	10	10	10	10	10	10	10
0.2	10	10	10	10	10	10	5	5	5
0.3	10	10	10	10	5	5	5	5	4
0.4	10	10	5	5	5	5	4	4	4
0.5	10	5	5	5	5	4	4	4	2
0.6	5	5	5	5	4	4	4	2	2
0.7	5	5	5	4	4	4	2	2	2
0.8	5	5	4	4	4	2	2	2	2
0.9	5	4	4	4	2	2	2	2	2
	E-optimality								
0.1	10	10	10	10	10	5	5	4	2
0.2	10	10	10	10	5	5	5	4	4
0.3	10	10	10	5	5	5	4	4	4
0.4	10	10	5	5	5	5	4	4	4
0.5	10	5	5	5	5	4	4	4	4
0.6	10	5	5	5	4	4	4	4	4
0.7	10	5	5	4	4	4	4	4	4
0.8	10	5	4	4	4	4	4	4	2
0.9	5	2	2	2	2	2	2	2	2
	T-optimality								
0.1	2	2	2	2	10	10	10	10	10
0.2	2	2	2	2	10	10	10	10	10
0.3	10	2	2	2	10	10	10	10	10
0.4	10	10	2	2	2	2	2	2	2
0.5	10	5	5	5	2	2	2	2	2
0.6	5	5	5	4	2	2	2	2	2
0.7	5	5	5	4	4	2	2	2	2
0.8	5	5	4	4	4	2	2	2	2
0.9	5	4	4	4	4	2	2	2	2

Table 6.1 Optimal number of secondary samples based on A-optimality, D-optimality, E-optimality and T-optimality.

$k=2$									
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	9.59^{*}	4.66^{*}	2.50^{*}	1.43^{*}	$7.58^{\dagger}$	$4.46^{\dagger}$	$2.78^{\dagger}$	$1.87^{\dagger}$	$1.78^{\dagger}$
0.2	1.10	3.89^{*}	1.74^{*}	$8.88^{\dagger}$	$4.74^{\dagger}$	$2.95^{\dagger}$	$1.93^{\dagger}$	$1.50^{\dagger}$	$1.63^{\dagger}$
0.3	1.13	3.34^{*}	1.39^{*}	$6.63^{\dagger}$	$3.82^{\dagger}$	$2.36^{\dagger}$	$1.68^{\dagger}$	$1.38^{\dagger}$	$\mathbf{1.59}^{\dagger}$
0.4	1.28	3.42^{*}	1.15^{*}	$5.86^{\dagger}$	$3.30^{\dagger}$	$2.13^{\dagger}$	$1.56^{\dagger}$	$1.32^{\dagger}$	$\mathbf{1.56}^{\dagger}$
0.5	1.49	2.70^{*}	1.07^{*}	$5.29^{\dagger}$	$3.05^{\dagger}$	$2.01^{\dagger}$	$1.47^{\dagger}$	$1.28^{\dagger}$	$\mathbf{1.55}^{\dagger}$
0.6	1.31	2.74^{*}	$9.84^{\dagger}$	$4.94^{\dagger}$	$2.89^{\dagger}$	$1.90^{\dagger}$	$1.43^{\dagger}$	$1.26^{\dagger}$	$\mathbf{1.55}^{\dagger}$
0.7	1.64	2.94^{*}	1.07^{*}	$4.89^{\dagger}$	$2.83^{\dagger}$	$1.91^{\dagger}$	$1.44^{\dagger}$	$\mathbf{1.28}^{\dagger}$	$\mathbf{1.58}^{\dagger}$
0.8	2.12	3.64^{*}	1.15^{*}	$5.65^{\dagger}$	$3.23^{\dagger}$	$2.09^{\dagger}$	$\mathbf{1.56}^{\dagger}$	$\mathbf{1.39}^{\dagger}$	$\mathbf{1.70}^{\dagger}$
0.9	3.71	5.71^{*}	1.86^{*}	$8.30^{\dagger}$	$4.52^{\dagger}$	$\mathbf{2.91}^{\dagger}$	$\mathbf{2.16}^{\dagger}$	$\mathbf{1.91}^{\dagger}$	$\mathbf{2.37}^{\dagger}$
$k=4$									
0.1	4.38^{*}	1.35^{*}	$4.57^{\dagger}$	$2.10^{\dagger}$	$2.10^{\dagger}$	$6.70^{\dagger\dagger}$	$5.21^{\dagger\dagger}$	$5.75^{\dagger\dagger}$	$1.01^{\dagger}$
0.2	4.09^{*}	$8.30^{\dagger}$	$2.98^{\dagger}$	$1.37^{\dagger}$	$1.37^{\dagger}$	$5.38^{\dagger\dagger}$	$4.75^{\dagger\dagger}$	$5.61^{\dagger\dagger}$	$\mathbf{1.02}^{\dagger}$
0.3	3.45^{*}	$6.31^{\dagger}$	$2.32^{\dagger}$	$1.11^{\dagger}$	$1.11^{\dagger}$	$4.97^{\dagger\dagger}$	$4.59^{\dagger\dagger}$	$\mathbf{5.56}^{\dagger\dagger}$	$1.02^{\dagger}$
0.4	3.31^{*}	$5.79^{\dagger}$	$2.00^{\dagger}$	$1.01^{\dagger}$	$1.01^{\dagger}$	$4.79^{\dagger\dagger}$	$\mathbf{4.54}^{\dagger\dagger}$	$\mathbf{5.58}^{\dagger\dagger}$	$1.03^{\dagger}$
0.5	3.30^{*}	$5.38^{\dagger}$	$1.90^{\dagger}$	$9.63^{\dagger\dagger}$	$9.63^{\dagger\dagger}$	$\mathbf{4.77}^{\dagger\dagger}$	$\mathbf{4.59}^{\dagger\dagger}$	$\mathbf{5.68}^{\dagger\dagger}$	$1.04^{\dagger}$
0.6	3.35^{*}	$5.38^{\dagger}$	$1.90^{\dagger}$	$9.78^{\dagger\dagger}$	$\mathbf{9.78}^{\dagger\dagger}$	$\mathbf{4.94}^{\dagger\dagger}$	$\mathbf{4.79}^{\dagger\dagger}$	$\mathbf{5.98}^{\dagger\dagger}$	$1.10^{\dagger}$
0.7	4.24^{*}	$5.77^{\dagger}$	$2.11^{\dagger}$	$\mathbf{1.06}^{\dagger}$	$\mathbf{1.06}^{\dagger}$	$\mathbf{5.41}^{\dagger\dagger}$	$\mathbf{5.30}^{\dagger\dagger}$	$6.60^{\dagger\dagger}$	$1.22^{\dagger}$
0.8	5.25^{*}	$7.52^{\dagger}$	$\mathbf{2.60}^{\dagger}$	$\mathbf{1.30}^{\dagger}$	$\mathbf{1.30}^{\dagger}$	$6.68^{\dagger\dagger}$	$6.57^{\dagger\dagger}$	$8.19^{\dagger\dagger}$	$1.51^{\dagger}$
0.9	9.71^{*}	$\mathbf{1.31}^{*}$	$\mathbf{4.28}^{\dagger}$	$\mathbf{2.15}^{\dagger}$	$\mathbf{2.15}^{\dagger}$	$1.09^{\dagger}$	$1.08^{\dagger}$	$1.33^{\dagger}$	$2.48^{\dagger}$
$k=5$									
0.1	3.11^{*}	$7.92^{\dagger}$	$2.65^{\dagger}$	$1.17^{\dagger}$	$6.20^{\dagger\dagger}$	$4.30^{\dagger\dagger}$	$\mathbf{4.03}^{\dagger\dagger}$	$\mathbf{5.19}^{\dagger\dagger}$	$\mathbf{1.01}^{\dagger}$
0.2	2.85^{*}	$4.74^{\dagger}$	$1.68^{\dagger}$	$7.96^{\dagger\dagger}$	$\mathbf{4.82}^{\dagger\dagger}$	$\mathbf{3.79}^{\dagger\dagger}$	$\mathbf{3.87}^{\dagger\dagger}$	$\mathbf{5.18}^{\dagger\dagger}$	$1.00^{\dagger}$
0.3	2.25^{*}	$3.82^{\dagger}$	$1.34^{\dagger}$	$\mathbf{6.83}^{\dagger\dagger}$	$\mathbf{4.40}^{\dagger\dagger}$	$\mathbf{3.64}^{\dagger\dagger}$	$\mathbf{3.85}^{\dagger\dagger}$	$5.21^{\dagger\dagger}$	$1.01^{\dagger}$
0.4	2.12^{*}	$3.40^{\dagger}$	$1.21^{\dagger}$	$\mathbf{6.37}^{\dagger\dagger}$	$\mathbf{4.25}^{\dagger\dagger}$	$\mathbf{3.62}^{\dagger\dagger}$	$3.90^{\dagger\dagger}$	$5.26^{\dagger\dagger}$	$1.03^{\dagger}$
0.5	2.03^{*}	$3.32^{\dagger}$	$\mathbf{1.19}^{\dagger}$	$\mathbf{6.24}^{\dagger\dagger}$	$\mathbf{4.29}^{\dagger\dagger}$	$3.72^{\dagger\dagger}$	$4.02^{\dagger\dagger}$	$5.49^{\dagger\dagger}$	$1.07^{\dagger}$
0.6	2.31^{*}	$\mathbf{3.32}^{\dagger}$	$\mathbf{1.23}^{\dagger}$	$\mathbf{6.52}^{\dagger\dagger}$	$4.52^{\dagger\dagger}$	$3.98^{\dagger\dagger}$	$4.32^{\dagger\dagger}$	$5.93^{\dagger\dagger}$	$1.15^{\dagger}$
0.7	$\mathbf{2.54}^{*}$	$\mathbf{3.80}^{\dagger}$	$\mathbf{1.35}^{\dagger}$	$7.34^{\dagger\dagger}$	$5.12^{\dagger\dagger}$	$4.54^{\dagger\dagger}$	$4.94^{\dagger\dagger}$	$6.76^{\dagger\dagger}$	$1.32^{\dagger}$
0.8	$\mathbf{3.50}^{*}$	$\mathbf{4.82}^{\dagger}$	$1.71^{\dagger}$	$9.27^{\dagger\dagger}$	$6.54^{\dagger\dagger}$	$5.81^{\dagger\dagger}$	$6.36^{\dagger\dagger}$	$8.72^{\dagger\dagger}$	$1.69^{\dagger}$
0.9	$\mathbf{6.51}^{*}$	$8.36^{\dagger}$	$2.95^{\dagger}$	$1.58^{\dagger}$	$1.11^{\dagger}$	$9.95^{\dagger\dagger}$	$1.09^{\dagger}$	$1.51^{\dagger}$	$2.93^{\dagger}$
$k=10$									
0.1	$\mathbf{9.43}^{\dagger}$	$\mathbf{1.41}^{\dagger}$	$\mathbf{4.49}^{\dagger\dagger}$	$\mathbf{2.52}^{\dagger\dagger}$	$\mathbf{2.24}^{\dagger\dagger}$	$\mathbf{2.56}^{\dagger\dagger}$	$3.37^{\dagger\dagger}$	$5.05^{\dagger\dagger}$	$1.00^{\dagger}$
0.2	$\mathbf{6.39}^{\dagger}$	$\mathbf{9.19}^{\dagger\dagger}$	$\mathbf{3.51}^{\dagger\dagger}$	$\mathbf{2.34}^{\dagger\dagger}$	$2.25^{\dagger\dagger}$	$2.64^{\dagger\dagger}$	$3.49^{\dagger\dagger}$	$5.22^{\dagger\dagger}$	$1.05^{\dagger}$
0.3	$\mathbf{5.23}^{\dagger}$	$\mathbf{8.06}^{\dagger\dagger}$	$\mathbf{3.33}^{\dagger\dagger}$	$2.37^{\dagger\dagger}$	$2.35^{\dagger\dagger}$	$2.78^{\dagger\dagger}$	$3.66^{\dagger\dagger}$	$5.49^{\dagger\dagger}$	$1.10^{\dagger}$
0.4	$\mathbf{4.74}^{\dagger}$	$\mathbf{7.93}^{\dagger\dagger}$	$\mathbf{3.45}^{\dagger\dagger}$	$2.53^{\dagger\dagger}$	$2.53^{\dagger\dagger}$	$3.02^{\dagger\dagger}$	$3.97^{\dagger\dagger}$	$5.93^{\dagger\dagger}$	$1.19^{\dagger}$
0.5	$\mathbf{4.90}^{\dagger}$	$\mathbf{8.23}^{\dagger\dagger}$	$3.72^{\dagger\dagger}$	$2.80^{\dagger\dagger}$	$2.84^{\dagger\dagger}$	$3.38^{\dagger\dagger}$	$4.45^{\dagger\dagger}$	$6.69^{\dagger\dagger}$	$1.34^{\dagger}$
0.6	$\mathbf{5.38}^{\dagger}$	$9.16^{\dagger\dagger}$	$4.24^{\dagger\dagger}$	$3.26^{\dagger\dagger}$	$3.32^{\dagger\dagger}$	$3.96^{\dagger\dagger}$	$5.24^{\dagger\dagger}$	$7.82^{\dagger\dagger}$	$1.56^{\dagger}$
0.7	$6.31^{\dagger}$	$1.10^{\dagger}$	$5.21^{\dagger\dagger}$	$4.08^{\dagger\dagger}$	$4.17^{\dagger\dagger}$	$4.99^{\dagger\dagger}$	$6.57^{\dagger\dagger}$	$9.80^{\dagger\dagger}$	$1.98^{\dagger}$
0.8	$9.01^{\dagger}$	$1.53^{\dagger}$	$7.34^{\dagger\dagger}$	$5.75^{\dagger\dagger}$	$5.92^{\dagger\dagger}$	$7.01^{\dagger\dagger}$	$9.26^{\dagger\dagger}$	$1.41^{\dagger}$	$2.77^{\dagger}$
0.9	1.75^{*}	$2.92^{\dagger}$	$1.39^{\dagger}$	$1.09^{\dagger}$	$1.12^{\dagger}$	$1.35^{\dagger}$	$1.76^{\dagger}$	$2.64^{\dagger}$	$5.28^{\dagger}$

Table 6.2 A-optimality values when $k = 2, 4, 5, 10$. $*$ indicates $\times 10^{-1}$, $\dagger$ indicates $\times 10^{-2}$, $\dagger\dagger$ indicates $\times 10^{-3}$ and $\ddagger$ indicates $\times 10^{-4}$. Optimal values are highlighted in bold.

$k=2$									
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.08	0.46	1.03	1.87	3.37	6.45	13.45	31.52	93.61
0.2	0.07	0.46	1.10	2.00	3.44	6.06	11.33	23.38	62.08
0.3	0.06	0.45	1.16	2.16	3.69	6.24	11.04	21.56	54.87
0.4	0.05	0.42	1.19	2.34	4.05	6.80	11.76	22.41	55.93
0.5	0.04	0.38	1.16	2.45	4.42	7.57	13.14	24.92	**61.89**
0.6	0.03	0.31	1.06	2.41	4.61	8.21	14.59	**28.04**	**70.07**
0.7	0.02	0.23	0.85	2.11	4.32	8.11	**14.96**	**29.48**	**74.71**
0.8	0.01	0.13	0.54	1.47	3.24	**6.45**	**12.45**	**25.32**	**65.38**
0.9	0.00	0.04	0.19	0.58	**1.37**	**2.91**	**5.89**	**12.40**	**32.69**
$k=4$									
0.1	0.45	3.03	7.86	15.48	26.96	42.44	61.27	87.08	150.00
0.2	0.39	2.70	6.98	12.91	20.33	28.99	39.22	54.17	92.66
0.3	0.32	2.34	6.17	11.25	17.20	23.85	31.71	43.48	**74.24**
0.4	0.25	1.95	5.29	9.74	14.86	20.49	**27.12**	**37.13**	**63.38**
0.5	0.19	1.54	4.32	8.10	12.45	**17.23**	**22.84**	**31.28**	53.39
0.6	0.13	1.12	3.25	6.24	**9.72**	**13.54**	**18.00**	24.68	42.14
0.7	0.08	0.71	2.15	**4.23**	**6.69**	**9.39**	12.52	17.19	29.37
0.8	0.04	0.36	**1.12**	**2.26**	**3.63**	5.13	6.87	9.45	16.15
0.9	0.01	**0.10**	**0.33**	**0.68**	1.10	1.57	2.11	2.91	4.97
$k=5$									
0.1	0.76	5.42	14.54	27.89	43.82	59.03	72.96	92.47	150.09
0.2	0.63	4.50	11.55	20.20	28.72	36.10	**43.25**	**54.32**	**88.06**
0.3	0.50	3.66	9.27	15.73	**21.71**	**26.81**	**31.87**	**39.94**	64.72
0.4	0.38	2.86	**7.29**	**12.28**	**16.81**	**20.64**	24.49	30.67	49.69
0.5	0.28	**2.13**	**5.47**	**9.24**	**12.64**	15.52	18.40	23.05	37.35
0.6	**0.19**	**1.46**	**3.80**	**6.47**	8.88	10.91	12.94	16.21	26.26
0.7	**0.11**	**0.88**	**2.33**	4.00	5.51	6.78	8.05	10.08	16.33
0.8	**0.05**	**0.42**	1.13	1.95	2.70	3.33	3.96	4.96	8.04
0.9	**0.01**	0.11	0.31	0.54	0.75	0.92	1.10	1.37	2.23
$k=10$									
0.1	**3.40**	**26.66**	**56.25**	**65.36**	**60.85**	**55.64**	**55.04**	**63.28**	**100.00**
0.2	**2.32**	**14.49**	**25.04**	**26.58**	**24.16**	**21.99**	21.75	25.00	39.51
0.3	**1.56**	**8.45**	**13.35**	**13.70**	12.35	11.23	11.10	12.76	20.16
0.4	**1.03**	**5.02**	7.52	7.59	6.81	6.19	6.12	7.03	11.11
0.5	**0.64**	2.92	4.24	4.23	3.79	3.44	3.40	3.91	6.17
0.6	0.38	1.61	2.28	2.26	2.02	1.83	1.81	2.08	3.29
0.7	0.19	0.80	1.11	1.09	0.97	0.88	0.87	1.00	1.59
0.8	0.08	0.32	0.43	0.42	0.38	0.34	0.34	0.39	0.62
0.9	0.02	0.07	0.10	0.09	0.08	0.08	0.08	0.09	0.14

Table 6.3 D-optimality values when $k = 2, 4, 5, 10$. Optimal values are highlighted in bold.

$k=2$									
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2.00‡	1.00††	4.40††	1.41†	3.51†	7.36†	1.39*	2.63*	**6.05***
0.2	2.00‡	1.60††	6.10††	1.83†	4.47†	9.39†	1.83*	3.61*	7.19*
0.3	4.00‡	2.50††	8.80††	2.40†	5.48†	1.11*	2.11*	4.13*	7.08*
0.4	7.00‡	4.20††	1.31†	3.18†	6.55†	1.23*	2.24*	4.29*	7.37*
0.5	1.50††	7.70††	2.08†	4.31†	7.77†	1.32*	2.24*	4.15*	8.03*
0.6	3.50††	1.54†	3.49†	6.00†	9.23†	1.39*	2.19*	3.86*	8.46*
0.7	9.50††	3.45†	6.10†	8.47†	1.11*	1.49*	2.17*	3.64*	8.23*
0.8	3.34†	8.44†	1.08*	1.19*	1.34*	1.64*	2.24*	3.62*	8.13*
0.9	1.60*	**2.10***	**1.76***	**1.59***	**1.62***	**1.87***	**2.47***	**3.90***	**8.70***
$k=4$									
0.1	9.00‡	8.00††	3.88†	1.20*	2.45*	3.57*	4.49*	**5.88***	9.88*
0.2	1.50††	1.25†	5.65†	1.65*	3.32*	5.13*	6.90*	**9.29***	**6.24***
0.3	2.50††	2.01†	8.20†	2.16*	4.08*	6.15*	**8.25***	**8.45***	**5.02***
0.4	4.50††	3.32†	1.20*	2.80*	4.85*	6.97*	**8.76***	**7.36***	**4.36***
0.5	8.50††	5.60†	1.74*	3.58*	5.68*	**7.67***	**8.29***	**6.48***	**3.83***
0.6	1.68†	9.57†	2.52*	4.53*	**6.53***	**7.93***	**7.38***	**5.65***	**3.34***
0.7	3.48†	1.62*	3.55*	**5.55***	**7.10***	**7.40***	**6.37***	**4.84***	**2.86***
0.8	7.15†	2.59*	**4.68***	**6.28***	**6.90***	**6.43***	**5.38***	**4.07***	**2.41***
0.9	1.28*	3.57*	5.37*	6.18*	6.07*	5.40*	4.48*	3.38*	2.00*
$k=5$									
0.1	1.60††	1.55†	7.79†	2.23*	3.77*	**4.54***	**5.11***	6.26*	9.89*
0.2	2.60††	2.42†	1.13*	3.06*	**5.39***	**7.17***	**8.52***	9.39*	5.82*
0.3	4.30††	3.78†	1.55*	**3.79***	**6.36***	**8.33***	8.55*	7.00*	4.33*
0.4	7.40††	5.83†	**2.06***	**4.40***	**6.59***	**7.44***	6.86*	5.57*	3.45*
0.5	1.29†	**8.79**†	**2.60***	**4.73***	**6.11***	6.21*	5.58*	4.53*	2.80*
0.6	2.27†	**1.27***	**3.05***	**4.66***	5.30*	5.12*	4.56*	3.69*	2.28*
0.7	3.90†	**1.69***	**3.27***	4.26*	4.46*	4.20*	3.73*	3.01*	1.86*
0.8	6.17†	**2.01***	3.19*	3.71*	3.70*	3.44*	3.04*	2.46*	1.52*
0.9	**8.23**†	2.10*	2.89*	3.13*	3.05*	2.81*	2.48*	2.01*	1.24*
$k=10$									
0.1	**8.10**††	**1.07***	**4.40***	**8.04***	**9.24***	8.33*	8.22*	9.45*	6.69*
0.2	**1.08**†	**1.02***	**2.66***	**3.78***	4.53*	5.04*	5.11*	4.45*	2.81*
0.3	**1.32**†	**8.55**†	**1.68***	2.18*	2.58*	2.87*	2.90*	2.53*	1.60*
0.4	**1.50**†	**6.87**†	1.14*	1.42*	1.67*	1.85*	1.88*	1.63*	1.03*
0.5	**1.59**†	5.46†	8.14†	9.97†	1.16*	1.29*	1.31*	1.14*	7.20†
0.6	**1.59**†	4.35†	6.07†	7.33†	8.53†	9.46†	9.58†	8.33†	5.27†
0.7	**1.51**†	3.50†	4.67†	5.58†	6.48†	7.18†	7.27†	6.33†	4.00†
0.8	**1.39**†	2.85†	3.68†	4.36†	5.06†	5.61†	5.68†	4.94†	3.13†
0.9	1.25†	2.35†	2.96†	3.48†	4.04†	4.47†	4.53†	3.94†	2.49†

Table 6.4 E-optimality values when $k = 2,4,5,10$. $*$ indicates $\times 10^{-1}$, † indicates $\times 10^{-2}$, †† indicates $\times 10^{-3}$ and ‡ indicates $\times 10^{-4}$. Optimal values are highlighted in bold.

				$k=2$					
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**23.10**	**21.00**	**15.31**	**11.69**	10.14	10.05	11.19	13.83	19.97
0.2	**17.38**	**17.23**	**13.48**	**10.62**	9.16	8.79	9.31	10.95	15.97
0.3	12.49	**13.55**	**11.59**	**9.73**	8.65	8.34	8.76	10.21	15.04
0.4	8.47	10.12	**9.62**	**8.84**	**8.37**	**8.35**	**8.87**	**10.33**	**15.13**
0.5	5.32	7.08	7.62	7.86	**8.13**	**8.58**	**9.37**	**10.97**	**15.83**
0.6	3.01	4.56	5.70	6.72	**7.72**	**8.75**	**9.95**	**11.81**	**16.80**
0.7	1.46	2.65	3.96	5.42	6.94	**8.48**	**10.11**	**12.28**	**17.37**
0.8	0.57	1.34	2.48	3.93	5.57	**7.30**	**9.13**	**11.39**	**16.26**
0.9	0.16	0.54	1.23	2.21	3.38	**4.68**	**6.09**	**7.84**	**11.46**
				$k=4$					
0.1	22.18	19.64	14.80	12.71	13.07	14.80	16.93	19.32	24.50
0.2	16.09	14.89	11.75	10.31	10.42	11.38	12.74	14.73	19.79
0.3	11.24	11.02	9.38	8.77	9.14	10.06	11.31	13.23	18.26
0.4	7.51	7.93	7.45	7.55	8.22	9.20	10.44	12.33	17.32
0.5	4.75	5.54	5.84	6.46	7.34	8.38	9.60	11.45	16.33
0.6	2.82	3.74	4.50	**5.39**	6.38	7.41	8.58	10.34	14.99
0.7	1.55	2.44	3.34	**4.29**	**5.25**	6.20	7.26	8.85	13.03
0.8	0.77	1.48	**2.27**	**3.09**	**3.87**	4.64	5.50	6.78	10.16
0.9	0.32	**2.72**	**1.20**	**1.69**	**2.16**	2.63	3.15	3.92	5.99
				$k=5$					
0.1	21.77	18.99	14.73	13.67	**14.85**	**16.58**	**18.05**	**19.76**	**24.50**
0.2	15.57	13.96	11.28	10.62	**11.23**	**12.18**	**13.20**	**14.75**	**19.46**
0.3	10.81	10.21	8.92	8.88	**9.56**	**10.40**	**11.32**	**12.84**	**17.52**
0.4	7.27	7.42	7.17	7.61	8.38	9.19	10.07	11.55	16.14
0.5	4.70	**5.35**	**5.78**	**6.51**	7.33	8.10	8.95	10.37	14.78
0.6	**2.92**	**3.82**	**4.61**	5.46	6.26	6.98	7.76	9.07	13.17
0.7	**1.74**	**2.67**	**3.54**	4.37	5.08	5.71	6.38	7.53	11.11
0.8	**0.96**	**1.73**	2.48	3.15	3.70	4.19	4.71	5.60	8.38
0.9	**0.43**	0.88	1.33	1.72	2.04	2.32	2.62	3.14	4.76
				$k=10$					
0.1	20.59	17.50	16.28	16.26	15.61	14.98	14.91	15.92	20.40
0.2	14.81	13.15	12.28	11.55	10.61	9.93	9.86	10.83	15.19
0.3	**12.01**	10.79	10.42	9.65	8.71	8.05	7.98	8.90	13.03
0.4	**8.38**	**9.13**	9.06	8.34	7.45	6.85	6.78	7.63	11.44
0.5	**6.46**	7.72	7.80	7.16	6.37	5.83	5.76	6.53	9.93
0.6	4.94	6.35	6.50	5.96	5.28	4.82	4.77	5.42	8.32
0.7	3.64	4.93	5.10	4.67	4.13	3.76	3.72	4.24	6.55
0.8	2.43	3.42	3.56	3.26	2.88	2.62	2.59	2.95	4.58
0.9	1.24	1.78	1.86	1.70	1.50	1.37	1.35	1.54	2.40

Table 6.5 Simulated T-optimality values when $k = 2, 4, 5, 10$. Optimal values are highlighted in bold.

6.4 Simulation study

In Section 6.3 we have discussed several design criteria, but the choice of an appropriate optimality criterion requires further investigation. In this section we use simulation to assess the performance of the design with respect to those criteria in the RMER.SRC model. The simulation setting considered is described in Section 6.4.1.

Small sample sizes are not uncommon in population ecology. In particular they are frequently encountered in surveys linked to conservation projects, as these often have limited resources and tend to focus on rare and protected species. Pilot studies also tend to deal with relatively small amount of data. Designing of a study based on large sample approximations may be not appropriate if the expected population size is small. Under these circumstances, the actual properties of the estimators may be different from those predicted by the analytical expressions. In these cases, the use of simulations may be more appropriate for designing a study. As the analytic results are based on the large sample properties of the estimators in the model, we also conduct simulations when the population size is small. We illustrate the need to use simulations for study design when the population size is small.

6.4.1 Setting

We conducted a simulation study to assess the performance of these analytic results in Section 6.3.6. We are interested in which criterion performs the best based on the results obtained in previous sections.

We simulate removal data sets with $K = 20$ sampling occasions and we vary the number of secondary periods within each primary period $k = (2,4,5,10)$. The Population size is set to be $N = 500$ or $N = 100$ individuals. The true values of capture probabilities p and transition probabilities ϕ are chosen from $(0.1, 0.2, \cdots, 0.9)$. For each possible combination of values of k, p and ϕ, 1000 data sets are simulated. For each simulation setting we obtain the Hessian matrix and calculate the median criteria values as shown in Tables 6.8 to 6.15.

6.4.2 Results

The optimal number of secondary samples to be conducted for each primary period is presented in Tables 6.6 and 6.7 for simulations with $N = 500$ and $N = 100$ respectively. As shown in Table 6.6, simulation results of A-optimality generally support our analytic results where greater optimal number of secondary samples is recommended when capture probability and ϕ are low. When $N = 100$, the patterns in A-optimality are similar to those of Table 6.6 and 6.1 except that when the true value of p is extremely small (e.g. $p = 0.1$). Therefore, we recommend to use simulations as a tool for sampling design if population size and capture probability are both expected to be small.

There are noticeable disagreements for other criteria especially in the left lower and right upper corners of Tables 6.6 and 6.7. Therefore the D-optimality, E-optimality and T-optimality should not be used in practice for removal data collected under the robust design.

A-optimality									
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	10	10	10	10	10	10	5	5	5
0.2	10	10	10	10	5	5	5	4	4
0.3	10	10	10	5	5	5	5	4	4
0.4	10	10	5	5	5	5	4	4	2
0.5	10	5	5	5	5	4	4	4	2
0.6	10	5	5	5	4	4	4	4	2
0.7	5	5	5	4	4	4	4	2	2
0.8	5	5	4	4	4	4	2	2	2
0.9	2	4	4	4	4	2	2	2	2
D-optimality									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	10	10	10	10	10	10	10	10	10
0.2	10	10	10	10	10	10	10	10	10
0.3	10	10	10	10	10	10	10	10	10
0.4	10	10	10	10	10	10	10	10	10
0.5	10	10	10	10	5	5	5	5	5
0.6	10	10	5	5	5	5	5	5	5
0.7	10	5	5	5	4	4	4	4	4
0.8	10	5	5	4	4	4	4	2	2
0.9	10	10	5	4	2	2	2	2	2
E-optimality									
0.1	10	10	10	10	10	10	10	10	10
0.2	10	10	10	10	10	10	10	10	10
0.3	10	10	10	10	10	10	10	10	10
0.4	10	10	5	10	10	10	10	10	10
0.5	10	5	5	5	5	5	5	5	5
0.6	10	5	4	4	4	4	5	5	5
0.7	10	5	4	4	4	4	4	4	4
0.8	5	4	4	2	2	2	2	2	2
0.9	10	5	2	2	2	2	2	2	2
T-optimality									
0.1	2	2	2	10	10	10	10	5	5
0.2	2	2	2	10	10	10	10	10	5
0.3	2	2	10	10	10	10	10	10	10
0.4	2	2	10	10	10	10	10	10	10
0.5	2	2	10	10	10	10	5	5	10
0.6	2	2	2	10	10	5	5	5	5
0.7	2	2	2	2	5	5	4	4	4
0.8	2	2	2	2	2	4	2	2	2
0.9	2	2	2	2	2	2	2	2	2

Table 6.6 Optimal number of secondary samples based on A-optimality, D-optimality, E-optimality and T-optimality when $N = 500$.

A-optimality									
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	10	10	10	10	10	10	5	5	5
0.2	4	10	10	10	5	5	5	5	5
0.3	4	10	10	5	5	5	4	4	4
0.4	5	10	5	5	5	5	4	4	2
0.5	4	5	5	5	4	4	4	4	2
0.6	4	5	5	5	4	4	4	4	2
0.7	5	5	5	4	4	4	4	2	2
0.8	4	5	4	4	4	4	2	2	2
0.9	2	4	4	4	4	4	2	2	2
D-optimality									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	10	10	10	10	10	10	10	10	10
0.2	10	10	10	10	10	10	10	10	10
0.3	10	10	10	10	10	10	10	10	10
0.4	10	10	10	10	10	10	10	10	10
0.5	10	10	10	10	5	5	5	5	5
0.6	10	10	5	5	5	5	5	5	4
0.7	10	5	5	5	5	4	4	4	4
0.8	10	5	5	4	4	4	4	2	2
0.9	10	5	4	4	2	2	2	2	2
E-optimality									
0.1	10	10	10	10	10	10	10	10	10
0.2	10	10	10	10	10	10	10	10	10
0.3	10	10	10	10	10	10	10	10	10
0.4	10	10	10	10	10	10	10	10	10
0.5	10	10	5	5	5	5	5	5	5
0.6	10	5	4	5	5	5	5	5	5
0.7	10	5	4	4	4	4	4	4	4
0.8	10	5	4	4	4	4	4	4	4
0.9	10	10	5	4	4	4	4	4	4
T-optimality									
0.1	2	2	2	10	10	10	10	10	10
0.2	2	2	10	10	10	10	10	10	10
0.3	2	2	10	10	10	10	10	10	10
0.4	2	2	10	10	10	10	10	10	10
0.5	2	2	10	10	10	10	10	5	10
0.6	2	2	2	10	10	5	5	5	5
0.7	2	2	2	2	5	5	5	4	4
0.8	2	2	2	2	2	4	4	2	2
0.9	2	2	2	2	2	2	2	2	2

Table 6.7 Optimal number of secondary samples based on A-optimality, D-optimality, E-optimality and T-optimality when $N = 100$.

					$k=2$				
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	$2.87^{\dagger}$	$1.84^{\dagger}$	$1.08^{\dagger}$	$5.56^{\dagger\dagger}$	$2.91^{\dagger\dagger}$	$1.56^{\dagger\dagger}$	$8.34^{\ddagger}$	$4.33^{\ddagger}$	$2.13^{\ddagger}$
0.2	$3.10^{\dagger}$	$2.16^{\dagger}$	$1.12^{\dagger}$	$5.36^{\dagger\dagger}$	$2.62^{\dagger\dagger}$	$1.46^{\dagger\dagger}$	$8.16^{\ddagger}$	$4.70^{\ddagger}$	$2.57^{\ddagger}$
0.3	$3.48^{\dagger}$	$2.23^{\dagger}$	$1.00^{\dagger}$	$4.44^{\dagger\dagger}$	$2.36^{\dagger\dagger}$	$1.31^{\dagger\dagger}$	$7.92^{\ddagger}$	$4.71^{\ddagger}$	$\mathbf{2.73}^{\ddagger}$
0.4	$4.37^{\dagger}$	$2.37^{\dagger}$	$8.10^{\dagger\dagger}$	$3.90^{\dagger\dagger}$	$2.06^{\dagger\dagger}$	$1.22^{\dagger\dagger}$	$7.56^{\ddagger}$	$4.60^{\ddagger}$	$\mathbf{2.69}^{\ddagger}$
0.5	$3.96^{\dagger}$	$1.62^{\dagger}$	$6.73^{\dagger\dagger}$	$3.26^{\dagger\dagger}$	$1.85^{\dagger\dagger}$	$1.15^{\dagger\dagger}$	$7.12^{\ddagger}$	$4.40^{\ddagger}$	$\mathbf{2.56}^{\ddagger}$
0.6	$2.58^{\dagger}$	$1.38^{\dagger}$	$5.20^{\dagger\dagger}$	$2.81^{\dagger\dagger}$	$1.68^{\dagger\dagger}$	$1.06^{\dagger\dagger}$	$6.77^{\ddagger}$	$4.23^{\ddagger}$	$\mathbf{2.39}^{\ddagger}$
0.7	$2.25^{\dagger}$	$1.08^{\dagger}$	$4.87^{\dagger\dagger}$	$2.51^{\dagger\dagger}$	$1.58^{\dagger\dagger}$	$1.04^{\dagger\dagger}$	$6.76^{\ddagger}$	$\mathbf{4.17}^{\ddagger}$	$\mathbf{2.33}^{\ddagger}$
0.8	$2.14^{\dagger}$	$1.02^{\dagger}$	$4.43^{\dagger\dagger}$	$2.71^{\dagger\dagger}$	$1.74^{\dagger\dagger}$	$1.14^{\dagger\dagger}$	$\mathbf{7.35}^{\ddagger}$	$\mathbf{4.46}^{\ddagger}$	$\mathbf{2.51}^{\ddagger}$
0.9	$\mathbf{2.43}^{\dagger}$	$1.25^{\dagger}$	$6.37^{\dagger\dagger}$	$3.83^{\dagger\dagger}$	$2.45^{\dagger\dagger}$	$\mathbf{1.60}^{\dagger\dagger}$	$\mathbf{1.03}^{\dagger\dagger}$	$\mathbf{6.30}^{\ddagger}$	$\mathbf{3.49}^{\ddagger}$
					$k=4$				
0.1	$1.89^{\dagger}$	$5.59^{\dagger\dagger}$	$1.81^{\dagger\dagger}$	$8.08^{\ddagger}$	$4.83^{\ddagger}$	$3.48^{\ddagger}$	$2.76^{\ddagger}$	$2.21^{\ddagger}$	$1.63^{\ddagger}$
0.2	$2.20^{\dagger}$	$5.41^{\dagger\dagger}$	$1.70^{\dagger\dagger}$	$8.02^{\ddagger}$	$5.11^{\ddagger}$	$3.88^{\ddagger}$	$3.21^{\ddagger}$	$2.68^{\ddagger}$	$\mathbf{2.09}^{\ddagger}$
0.3	$2.37^{\dagger}$	$4.62^{\dagger\dagger}$	$1.54^{\dagger\dagger}$	$7.76^{\ddagger}$	$5.32^{\ddagger}$	$4.22^{\ddagger}$	$3.57^{\ddagger}$	$\mathbf{3.04}^{\ddagger}$	$2.46^{\ddagger}$
0.4	$2.21^{\dagger}$	$4.20^{\dagger\dagger}$	$1.39^{\dagger\dagger}$	$7.72^{\ddagger}$	$5.54^{\ddagger}$	$4.49^{\ddagger}$	$\mathbf{3.84}^{\ddagger}$	$\mathbf{3.32}^{\ddagger}$	$2.73^{\ddagger}$
0.5	$2.05^{\dagger}$	$3.61^{\dagger\dagger}$	$1.35^{\dagger\dagger}$	$7.95^{\ddagger}$	$5.90^{\ddagger}$	$\mathbf{4.86}^{\ddagger}$	$\mathbf{4.20}^{\ddagger}$	$\mathbf{3.66}^{\ddagger}$	$3.06^{\ddagger}$
0.6	$1.88^{\dagger}$	$3.33^{\dagger\dagger}$	$1.37^{\dagger\dagger}$	$8.66^{\ddagger}$	$\mathbf{6.56}^{\ddagger}$	$\mathbf{5.46}^{\ddagger}$	$\mathbf{4.77}^{\ddagger}$	$\mathbf{4.18}^{\ddagger}$	$3.55^{\ddagger}$
0.7	$1.90^{\dagger}$	$3.29^{\dagger\dagger}$	$1.55^{\dagger\dagger}$	$\mathbf{1.02}^{\dagger\dagger}$	$\mathbf{7.84}^{\ddagger}$	$\mathbf{6.58}^{\ddagger}$	$\mathbf{5.79}^{\ddagger}$	$5.14^{\ddagger}$	$4.43^{\ddagger}$
0.8	$1.99^{\dagger}$	$4.16^{\dagger\dagger}$	$\mathbf{2.02}^{\dagger\dagger}$	$\mathbf{1.37}^{\dagger\dagger}$	$\mathbf{1.07}^{\dagger\dagger}$	$\mathbf{9.04}^{\ddagger}$	$7.98^{\ddagger}$	$7.15^{\ddagger}$	$6.27^{\ddagger}$
0.9	$2.65^{\dagger}$	$\mathbf{7.07}^{\dagger\dagger}$	$\mathbf{3.62}^{\dagger\dagger}$	$\mathbf{2.52}^{\dagger\dagger}$	$\mathbf{1.96}^{\dagger\dagger}$	$1.67^{\dagger\dagger}$	$1.48^{\dagger\dagger}$	$1.34^{\dagger\dagger}$	$1.20^{\dagger\dagger}$
					$k=5$				
0.1	$1.42^{\dagger}$	$3.32^{\dagger\dagger}$	$1.02^{\dagger\dagger}$	$4.88^{\ddagger}$	$3.35^{\ddagger}$	$2.79^{\ddagger}$	$\mathbf{2.46}^{\ddagger}$	$\mathbf{2.12}^{\ddagger}$	$\mathbf{1.61}^{\ddagger}$
0.2	$1.70^{\dagger}$	$3.07^{\dagger\dagger}$	$9.83^{\ddagger}$	$5.24^{\ddagger}$	$\mathbf{3.91}^{\ddagger}$	$\mathbf{3.37}^{\ddagger}$	$\mathbf{3.04}^{\ddagger}$	$\mathbf{2.70}^{\ddagger}$	$2.20^{\ddagger}$
0.3	$1.69^{\dagger}$	$2.81^{\dagger\dagger}$	$9.53^{\ddagger}$	$\mathbf{5.67}^{\ddagger}$	$\mathbf{4.39}^{\ddagger}$	$\mathbf{3.87}^{\ddagger}$	$\mathbf{3.55}^{\ddagger}$	$3.21^{\ddagger}$	$2.70^{\ddagger}$
0.4	$1.75^{\dagger}$	$2.60^{\dagger\dagger}$	$9.75^{\ddagger}$	$\mathbf{6.19}^{\ddagger}$	$\mathbf{4.98}^{\ddagger}$	$\mathbf{4.44}^{\ddagger}$	$4.10^{\ddagger}$	$3.77^{\ddagger}$	$3.23^{\ddagger}$
0.5	$1.55^{\dagger}$	$2.55^{\dagger\dagger}$	$\mathbf{1.06}^{\dagger\dagger}$	$\mathbf{7.04}^{\ddagger}$	$\mathbf{5.80}^{\ddagger}$	$5.21^{\ddagger}$	$4.85^{\ddagger}$	$4.49^{\ddagger}$	$3.95^{\ddagger}$
0.6	$1.54^{\dagger}$	$\mathbf{2.58}^{\dagger\dagger}$	$\mathbf{1.22}^{\dagger\dagger}$	$\mathbf{8.46}^{\ddagger}$	$7.07^{\ddagger}$	$6.39^{\ddagger}$	$5.97^{\ddagger}$	$5.59^{\ddagger}$	$4.99^{\ddagger}$
0.7	$\mathbf{1.51}^{\dagger}$	$\mathbf{3.02}^{\dagger\dagger}$	$\mathbf{1.51}^{\dagger\dagger}$	$1.09^{\dagger\dagger}$	$9.20^{\ddagger}$	$8.40^{\ddagger}$	$7.89^{\ddagger}$	$7.44^{\ddagger}$	$6.75^{\ddagger}$
0.8	$\mathbf{1.92}^{\dagger}$	$\mathbf{4.10}^{\dagger\dagger}$	$2.20^{\dagger\dagger}$	$1.61^{\dagger\dagger}$	$1.37^{\dagger\dagger}$	$1.25^{\dagger\dagger}$	$1.18^{\dagger\dagger}$	$1.12^{\dagger\dagger}$	$1.03^{\dagger\dagger}$
0.9	$\mathbf{2.61}^{\dagger}$	$7.68^{\dagger\dagger}$	$4.23^{\dagger\dagger}$	$3.16^{\dagger\dagger}$	$2.69^{\dagger\dagger}$	$2.50^{\dagger\dagger}$	$2.36^{\dagger\dagger}$	$2.26^{\dagger\dagger}$	$2.11^{\dagger\dagger}$
					$k=10$				
0.1	$\mathbf{5.41}^{\dagger\dagger}$	$\mathbf{6.62}^{\ddagger}$	$\mathbf{2.87}^{\ddagger}$	$\mathbf{2.48}^{\ddagger}$	$\mathbf{2.54}^{\ddagger}$	$\mathbf{2.66}^{\ddagger}$	$2.67^{\ddagger}$	$2.52^{\ddagger}$	$2.01^{\ddagger}$
0.2	$\mathbf{6.24}^{\dagger\dagger}$	$\mathbf{9.15}^{\ddagger}$	$\mathbf{4.89}^{\ddagger}$	$\mathbf{4.28}^{\ddagger}$	$4.36^{\ddagger}$	$4.49^{\ddagger}$	$4.49^{\ddagger}$	$4.31^{\ddagger}$	$3.81^{\ddagger}$
0.3	$\mathbf{7.02}^{\dagger\dagger}$	$\mathbf{1.28}^{\dagger\dagger}$	$\mathbf{7.74}^{\ddagger}$	$6.99^{\ddagger}$	$7.04^{\ddagger}$	$7.10^{\ddagger}$	$7.08^{\ddagger}$	$6.91^{\ddagger}$	$6.39^{\ddagger}$
0.4	$\mathbf{8.20}^{\dagger\dagger}$	$\mathbf{1.83}^{\dagger\dagger}$	$\mathbf{1.20}^{\dagger\dagger}$	$1.09^{\dagger\dagger}$	$1.09^{\dagger\dagger}$	$1.11^{\dagger\dagger}$	$1.10^{\dagger\dagger}$	$1.08^{\dagger\dagger}$	$1.02^{\dagger\dagger}$
0.5	$\mathbf{1.01}^{\dagger}$	$\mathbf{2.65}^{\dagger\dagger}$	$1.81^{\dagger\dagger}$	$1.68^{\dagger\dagger}$	$1.66^{\dagger\dagger}$	$1.67^{\dagger\dagger}$	$1.68^{\dagger\dagger}$	$1.67^{\dagger\dagger}$	$1.61^{\dagger\dagger}$
0.6	$\mathbf{1.28}^{\dagger}$	$3.86^{\dagger\dagger}$	$2.80^{\dagger\dagger}$	$2.56^{\dagger\dagger}$	$2.58^{\dagger\dagger}$	$2.59^{\dagger\dagger}$	$2.60^{\dagger\dagger}$	$2.54^{\dagger\dagger}$	$2.50^{\dagger\dagger}$
0.7	$1.83^{\dagger}$	$6.12^{\dagger\dagger}$	$4.53^{\dagger\dagger}$	$4.20^{\dagger\dagger}$	$4.15^{\dagger\dagger}$	$4.24^{\dagger\dagger}$	$4.26^{\dagger\dagger}$	$4.13^{\dagger\dagger}$	$4.04^{\dagger\dagger}$
0.8	$2.84^{\dagger}$	$1.04^{\dagger}$	$8.03^{\dagger\dagger}$	$7.36^{\dagger\dagger}$	$7.51^{\dagger\dagger}$	$7.43^{\dagger\dagger}$	$7.59^{\dagger\dagger}$	$7.40^{\dagger\dagger}$	$7.21^{\dagger\dagger}$
0.9	$5.06^{\dagger}$	$2.06^{\dagger}$	$1.62^{\dagger}$	$1.54^{\dagger}$	$1.54^{\dagger}$	$1.55^{\dagger}$	$1.51^{\dagger}$	$1.45^{\dagger}$	$1.46^{\dagger}$

Table 6.8 A-optimality values when $k = 2, 4, 5, 10$ and $N = 500$. † indicates $\times 10^{-2}$, †† indicates $\times 10^{-3}$ and ‡ indicates $\times 10^{-4}$ Optimal values are highlighted in bold.

$k=2$									
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	$2.92^{\otimes}$	$1.41^{\star}$	$3.96^{\otimes}$	$3.44^{\dagger}$	$5.32^{\dagger}$	$2.33^{\dagger}$	$3.53^{\dagger\dagger}$	$1.29^{\dagger\dagger}$	$2.90^{\dagger\dagger}$
0.2	$1.43^{\star}$	$1.00^{\star\star}$	$5.36^{\star}$	$1.07^{\otimes}$	$3.61^{\otimes}$	$4.90^{\otimes}$	$2.21^{\otimes}$	$3.15^{\dagger}$	$3.73^{\dagger\dagger}$
0.3	$4.49^{\star}$	$2.83^{\star\star}$	$2.30^{\star\star}$	$8.53^{\star}$	2.44	1.72	1.42	$3.58^{\otimes}$	$2.25^{\otimes}$
0.4	$5.60^{\star}$	$4.09^{\star\star}$	$7.40^{\star\star}$	$4.90^{\star\star}$	$2.23^{\star\star}$	$2.61^{\star}$	3.50	$7.49^{\otimes}$	$1.14^{\dagger}$
0.5	$4.49^{\star}$	$7.59^{\star\star}$	$1.47^{\diamond}$	$1.68^{\diamond}$	$1.31^{\diamond}$	$8.57^{\star\star}$	$4.56^{\star\star}$	$2.04^{\star\star}$	$7.42^{\star}$
0.6	$4.22^{\star}$	$7.57^{\star\star}$	$2.51^{\diamond}$	$3.77^{\diamond}$	$4.34^{\diamond}$	$4.17^{\diamond}$	$3.53^{\diamond}$	$2.76^{\diamond}$	$1.75^{\diamond}$
0.7	$1.82^{\star}$	$5.32^{\star\star}$	$2.40^{\diamond}$	$5.61^{\diamond}$	$8.50^{\diamond}$	$1.04^{\diamond\diamond}$	$1.12^{\diamond\diamond}$	$1.03^{\diamond\diamond}$	$7.30^{\diamond}$
0.8	$3.01^{\otimes}$	$1.99^{\star\star}$	$1.38^{\diamond}$	$3.89^{\diamond}$	$7.96^{\diamond}$	$1.24^{\diamond\diamond}$	$1.60^{\diamond\diamond}$	$\mathbf{1.70}^{\diamond\diamond}$	$\mathbf{1.32}^{\diamond\diamond}$
0.9	2.41	$1.56^{\star}$	$1.51^{\star\star}$	$6.55^{\star\star}$	$\mathbf{1.70}^{\diamond}$	$\mathbf{3.28}^{\diamond}$	$\mathbf{4.98}^{\diamond}$	$\mathbf{5.88}^{\diamond}$	$\mathbf{4.86}^{\diamond}$
$k=4$									
0.1	$3.53^{\star}$	$1.17^{\star\star}$	$6.14^{\star}$	$8.75^{\dagger\dagger}$	$8.17^{\dagger}$	$1.90^{\otimes}$	$1.39^{\dagger}$	$7.24^{\dagger\dagger}$	1.35
0.2	$1.15^{\star\star}$	$6.41^{\star\star}$	$6.16^{\star\star}$	$2.99^{\star\star}$	$5.32^{\star}$	3.42	$4.39^{\otimes}$	$1.50^{\otimes}$	3.09
0.3	$2.12^{\star\star}$	$1.68^{\diamond}$	$2.43^{\diamond}$	$2.32^{\diamond}$	$1.73^{\diamond}$	$1.32^{\diamond}$	$1.03^{\diamond}$	$7.37^{\star\star}$	$3.84^{\star\star}$
0.4	$2.42^{\star\star}$	$2.74^{\diamond}$	$5.87^{\diamond}$	$7.49^{\diamond}$	$7.77^{\diamond}$	$7.45^{\diamond}$	$6.58^{\diamond}$	$4.99^{\diamond}$	$2.65^{\diamond}$
0.5	$2.23^{\star\star}$	$3.47^{\diamond}$	$9.75^{\diamond}$	$1.53^{\diamond\diamond}$	$1.93^{\diamond\diamond}$	$2.05^{\diamond\diamond}$	$1.89^{\diamond\diamond}$	$1.47^{\diamond\diamond}$	$7.77^{\diamond}$
0.6	$1.53^{\star\star}$	$3.14^{\diamond}$	$1.13^{\diamond\diamond}$	$2.11^{\diamond\diamond}$	$2.99^{\diamond\diamond}$	$3.38^{\diamond\diamond}$	$3.29^{\diamond\diamond}$	$2.55^{\diamond\diamond}$	$1.38^{\diamond\diamond}$
0.7	$5.65^{\star}$	$1.87^{\diamond}$	$8.20^{\diamond}$	$1.83^{\diamond\diamond}$	$\mathbf{2.81}^{\diamond\diamond}$	$\mathbf{3.44}^{\diamond\diamond}$	$\mathbf{3.40}^{\diamond\diamond}$	$\mathbf{2.70}^{\diamond\diamond}$	$\mathbf{1.45}^{\diamond\diamond}$
0.8	$1.12^{\star}$	$5.65^{\star\star}$	$3.09^{\diamond}$	$\mathbf{7.70}^{\diamond}$	$\mathbf{1.31}^{\diamond\diamond}$	$\mathbf{1.67}^{\diamond\diamond}$	$\mathbf{1.69}^{\diamond\diamond}$	$1.35^{\diamond\diamond}$	$7.26^{\diamond}$
0.9	44.0	$3.77^{\star}$	$2.60^{\star\star}$	$\mathbf{7.41}^{\star\star}$	$1.34^{\diamond}$	$1.77^{\diamond}$	$1.86^{\diamond}$	$1.46^{\diamond}$	$7.91^{\star\star}$
$k=5$									
0.1	$7.26^{\star}$	$2.39^{\star\star}$	$1.56^{\star\star}$	$8.68^{\otimes}$	$4.23^{\otimes}$	$3.39^{\dagger}$	$1.06^{\dagger}$	$2.64^{\dagger\dagger}$	1.44
0.2	$1.97^{\star\star}$	$1.26^{\diamond}$	$1.41^{\diamond}$	$1.05^{\diamond}$	$7.13^{\star\star}$	$4.87^{\star\star}$	$3.48^{\star\star}$	$2.31^{\star\star}$	$1.13^{\star\star}$
0.3	$3.25^{\star\star}$	$2.93^{\diamond}$	$5.12^{\diamond}$	$5.74^{\diamond}$	$5.65^{\diamond}$	$5.25^{\diamond}$	$4.44^{\diamond}$	$3.11^{\diamond}$	$1.59^{\diamond}$
0.4	$3.68^{\star\star}$	$4.53^{\diamond}$	$1.07^{\diamond\diamond}$	$1.47^{\diamond\diamond}$	$1.72^{\diamond\diamond}$	$1.72^{\diamond\diamond}$	$1.47^{\diamond\diamond}$	$1.09^{\diamond\diamond}$	$5.40^{\diamond}$
0.5	$3.34^{\star\star}$	$5.15^{\diamond}$	$1.50^{\diamond\diamond}$	$2.52^{\diamond\diamond}$	$\mathbf{3.18}^{\diamond\diamond}$	$\mathbf{3.32}^{\diamond\diamond}$	$\mathbf{2.92}^{\diamond\diamond}$	$\mathbf{2.10}^{\diamond\diamond}$	$\mathbf{1.07}^{\diamond\diamond}$
0.6	$1.85^{\star\star}$	$4.34^{\diamond}$	$\mathbf{1.49}^{\diamond\diamond}$	$\mathbf{2.82}^{\diamond\diamond}$	$\mathbf{3.81}^{\diamond\diamond}$	$\mathbf{4.10}^{\diamond\diamond}$	$\mathbf{3.65}^{\diamond\diamond}$	$\mathbf{2.65}^{\diamond\diamond}$	$\mathbf{1.35}^{\diamond\diamond}$
0.7	$7.46^{\star}$	$\mathbf{2.35}^{\diamond}$	$\mathbf{9.70}^{\diamond}$	$\mathbf{1.98}^{\diamond\diamond}$	$2.82^{\diamond\diamond}$	$3.11^{\diamond\diamond}$	$2.82^{\diamond\diamond}$	$2.05^{\diamond\diamond}$	$1.05^{\diamond\diamond}$
0.8	$1.66^{\star}$	$\mathbf{6.59}^{\star\star}$	$\mathbf{3.14}^{\diamond}$	$7.03^{\diamond}$	$1.04^{\diamond\diamond}$	$1.18^{\diamond\diamond}$	$1.08^{\diamond\diamond}$	$7.84^{\diamond}$	$4.02^{\diamond}$
0.9	47.2	$4.63^{\star}$	$\mathbf{2.52}^{\star\star}$	$6.08^{\star\star}$	$9.28^{\star\star}$	$9.44^{\star\star}$	$9.43^{\star\star}$	$6.57^{\star\star}$	$3.34^{\star\star}$
$k=10$									
0.1	$\mathbf{3.05}^{\star\star}$	$\mathbf{1.82}^{\diamond}$	$\mathbf{2.72}^{\diamond}$	$\mathbf{2.98}^{\diamond}$	$\mathbf{2.64}^{\diamond}$	$\mathbf{2.21}^{\diamond}$	$\mathbf{1.63}^{\diamond}$	$\mathbf{1.18}^{\diamond}$	$\mathbf{5.54}^{\star\star}$
0.2	$\mathbf{7.76}^{\star\star}$	$\mathbf{7.04}^{\diamond}$	$\mathbf{1.30}^{\diamond\diamond}$	$\mathbf{1.47}^{\diamond\diamond}$	$\mathbf{1.44}^{\diamond\diamond}$	$\mathbf{1.20}^{\diamond\diamond}$	$\mathbf{8.90}^{\diamond}$	$\mathbf{6.02}^{\diamond}$	$\mathbf{2.97}^{\diamond}$
0.3	$\mathbf{1.07}^{\diamond}$	$\mathbf{1.12}^{\diamond\diamond}$	$\mathbf{2.33}^{\diamond\diamond}$	$\mathbf{2.87}^{\diamond\diamond}$	$\mathbf{2.76}^{\diamond\diamond}$	$\mathbf{2.27}^{\diamond\diamond}$	$\mathbf{1.72}^{\diamond\diamond}$	$\mathbf{1.15}^{\diamond\diamond}$	$\mathbf{5.77}^{\diamond}$
0.4	$\mathbf{1.09}^{\diamond}$	$\mathbf{1.22}^{\diamond\diamond}$	$\mathbf{2.68}^{\diamond\diamond}$	$\mathbf{3.36}^{\diamond\diamond}$	$\mathbf{3.25}^{\diamond\diamond}$	$\mathbf{2.74}^{\diamond\diamond}$	$\mathbf{2.06}^{\diamond\diamond}$	$\mathbf{1.37}^{\diamond\diamond}$	$\mathbf{6.84}^{\diamond}$
0.5	$\mathbf{7.50}^{\star\star}$	$\mathbf{9.58}^{\diamond}$	$\mathbf{2.21}^{\diamond\diamond}$	$\mathbf{2.81}^{\diamond\diamond}$	$2.73^{\diamond\diamond}$	$2.27^{\diamond\diamond}$	$1.71^{\diamond\diamond}$	$1.15^{\diamond\diamond}$	$5.72^{\diamond}$
0.6	$\mathbf{4.14}^{\star\star}$	$\mathbf{5.64}^{\diamond}$	$1.33^{\diamond\diamond}$	$1.71^{\diamond\diamond}$	$1.65^{\diamond\diamond}$	$1.38^{\diamond\diamond}$	$1.04^{\diamond\diamond}$	$6.99^{\diamond}$	$3.50^{\diamond}$
0.7	$\mathbf{1.51}^{\star\star}$	$2.23^{\diamond}$	$5.40^{\diamond}$	$7.06^{\diamond}$	$6.81^{\diamond}$	$5.54^{\diamond}$	$4.16^{\diamond}$	$2.83^{\diamond}$	$1.40^{\diamond}$
0.8	$\mathbf{2.55}^{\star}$	$4.86^{\star\star}$	$1.18^{\diamond}$	$1.58^{\diamond}$	$1.40^{\diamond}$	$1.19^{\diamond}$	$8.65^{\star\star}$	$5.87^{\star\star}$	$3.03^{\star\star}$
0.9	$\mathbf{1.08}^{\otimes}$	$\mathbf{2.85}^{\star}$	$7.15^{\star}$	$7.49^{\star}$	$8.79^{\star}$	$6.72^{\star}$	$5.79^{\star}$	$4.47^{\star}$	$1.87^{\star}$

Table 6.9 Simulated D-optimality values when $k = 2, 4, 5, 10$ and $N = 500$. $\otimes$ indicates $\times 10^2$, $\star$ indicates $\times 10^3$, $\star\star$ indicates $\times 10^4$, $\diamond$ indicates $\times 10^5$ and $\diamond\diamond$ indicates $\times 10^6$. Optimal values are highlighted in bold.

ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$k=2$				
0.1	1.22^{*}	$1.91^{\ddagger}$	$8.49^{\ddagger}$	1.42^{*}	9.48^{*}	2.74^{*}	3.45^{*}	1.39^{*}	4.80^{*}
0.2	7.26^{*}	$9.68^{\ddagger}$	$3.49^{\dagger\dagger}$	8.40^{*}	2.12^{*}	1.93^{*}	7.24^{*}	1.04^{*}	1.61^{*}
0.3	$2.47^{\ddagger}$	$1.90^{\dagger\dagger}$	$6.18^{\dagger\dagger}$	$2.57^{\dagger\dagger}$	5.73^{*}	2.88^{*}	2.10^{*}	4.70^{*}	4.00^{*}
0.4	$3.55^{\ddagger}$	$2.18^{\dagger\dagger}$	$9.19^{\dagger\dagger}$	$8.69^{\dagger\dagger}$	$2.75^{\dagger\dagger}$	$2.52^{\ddagger}$	2.91^{*}	5.63^{*}	1.10^{*}
0.5	$3.81^{\ddagger}$	$3.26^{\dagger\dagger}$	$1.06^{\dagger}$	$1.83^{\dagger}$	$1.05^{\dagger}$	$5.18^{\dagger\dagger}$	$2.33^{\dagger\dagger}$	$9.91^{\ddagger}$	$4.61^{\ddagger}$
0.6	$5.84^{\ddagger}$	$3.34^{\dagger\dagger}$	$1.16^{\dagger}$	$2.57^{\dagger}$	$2.34^{\dagger}$	$1.72^{\dagger}$	$1.24^{\dagger}$	$9.20^{\dagger\dagger}$	$7.56^{\dagger\dagger}$
0.7	$5.54^{\ddagger}$	$3.17^{\dagger\dagger}$	$1.03^{\dagger}$	$2.50^{\dagger}$	$3.34^{\dagger}$	$3.21^{\dagger}$	$2.93^{\dagger}$	$2.60^{\dagger}$	$2.39^{\dagger}$
0.8	$3.44^{\ddagger}$	$2.77^{\dagger\dagger}$	$8.86^{\dagger\dagger}$	$\mathbf{1.80}^{\dagger}$	$\mathbf{2.75}^{\dagger}$	$\mathbf{3.31}^{\dagger}$	$\mathbf{3.57}^{\dagger}$	$\mathbf{3.64}^{\dagger}$	$\mathbf{3.63}^{\dagger}$
0.9	1.88^{*}	$1.20^{\dagger\dagger}$	$\mathbf{3.54}^{\dagger\dagger}$	$\mathbf{6.92}^{\dagger\dagger}$	$\mathbf{1.03}^{\dagger}$	$\mathbf{1.28}^{\dagger}$	$\mathbf{1.55}^{\dagger}$	$\mathbf{1.71}^{\dagger}$	$\mathbf{1.82}^{\dagger}$
					$k=4$				
0.1	$1.75^{\ddagger}$	$1.72^{\dagger\dagger}$	$3.13^{\dagger\dagger}$	3.05^{*}	2.07^{*}	3.42^{*}	2.83^{*}	2.10^{*}	1.40^{*}
0.2	$6.19^{\ddagger}$	$6.59^{\dagger\dagger}$	$1.10^{\dagger}$	$3.35^{\dagger\dagger}$	$4.50^{\ddagger}$	2.86^{*}	4.01^{*}	2.26^{*}	8.31^{*}
0.3	$1.30^{\dagger\dagger}$	$1.24^{\dagger}$	$2.24^{\dagger}$	$1.37^{\dagger}$	$8.26^{\dagger\dagger}$	$5.88^{\dagger\dagger}$	$4.96^{\dagger\dagger}$	$4.53^{\dagger\dagger}$	$4.37^{\dagger\dagger}$
0.4	$1.87^{\dagger\dagger}$	$1.54^{\dagger}$	$3.25^{\dagger}$	$2.80^{\dagger}$	$2.30^{\dagger}$	$1.99^{\dagger}$	$1.83^{\dagger}$	$1.76^{\dagger}$	$1.74^{\dagger}$
0.5	$2.20^{\dagger\dagger}$	$1.68^{\dagger}$	$3.63^{\dagger}$	$3.91^{\dagger}$	$3.79^{\dagger}$	$3.60^{\dagger}$	$3.43^{\dagger}$	$3.37^{\dagger}$	$3.35^{\dagger}$
0.6	$2.37^{\dagger\dagger}$	$1.49^{\dagger}$	$\mathbf{3.17}^{\dagger}$	$\mathbf{3.98}^{\dagger}$	$\mathbf{4.27}^{\dagger}$	$\mathbf{4.32}^{\dagger}$	$4.30^{\dagger}$	$4.27^{\dagger}$	$4.26^{\dagger}$
0.7	$1.83^{\dagger\dagger}$	$1.06^{\dagger}$	$\mathbf{2.18}^{\dagger}$	$\mathbf{2.94}^{\dagger}$	$\mathbf{3.39}^{\dagger}$	$\mathbf{3.60}^{\dagger}$	$\mathbf{3.67}^{\dagger}$	$\mathbf{3.67}^{\dagger}$	$\mathbf{3.68}^{\dagger}$
0.8	$9.09^{\ddagger}$	$\mathbf{5.56}^{\dagger\dagger}$	$\mathbf{1.12}^{\dagger}$	$1.47^{\dagger}$	$1.81^{\dagger}$	$1.92^{\dagger}$	$2.01^{\dagger}$	$2.01^{\dagger}$	$2.02^{\dagger}$
0.9	$2.85^{\ddagger}$	$1.57^{\dagger\dagger}$	$2.89^{\dagger\dagger}$	$3.96^{\dagger\dagger}$	$4.94^{\dagger\dagger}$	$4.95^{\dagger\dagger}$	$5.42^{\dagger\dagger}$	$5.44^{\dagger\dagger}$	$5.75^{\dagger\dagger}$
					$k=5$				
0.1	$3.67^{\ddagger}$	$3.41^{\dagger\dagger}$	$4.65^{\dagger\dagger}$	$1.65^{\ddagger}$	5.62^{*}	3.74^{*}	1.46^{*}	7.12^{*}	1.38^{*}
0.2	$1.13^{\dagger\dagger}$	$1.24^{\dagger}$	$1.55^{\dagger}$	$7.49^{\dagger\dagger}$	$3.96^{\dagger\dagger}$	$2.64^{\dagger\dagger}$	$2.41^{\dagger\dagger}$	$3.00^{\dagger\dagger}$	$2.94^{\dagger\dagger}$
0.3	$2.17^{\dagger\dagger}$	$2.01^{\dagger}$	$2.95^{\dagger}$	$2.30^{\dagger}$	$1.91^{\dagger}$	$1.81^{\dagger}$	$1.88^{\dagger}$	$1.83^{\dagger}$	$1.82^{\dagger}$
0.4	$2.95^{\dagger\dagger}$	$2.35^{\dagger}$	$\mathbf{3.87}^{\dagger}$	$3.86^{\dagger}$	$3.74^{\dagger}$	$3.68^{\dagger}$	$3.60^{\dagger}$	$3.58^{\dagger}$	$3.54^{\dagger}$
0.5	$3.56^{\dagger\dagger}$	$\mathbf{2.18}^{\dagger}$	$\mathbf{3.86}^{\dagger}$	$\mathbf{4.45}^{\dagger}$	$\mathbf{4.64}^{\dagger}$	$\mathbf{4.65}^{\dagger}$	$\mathbf{4.62}^{\dagger}$	$\mathbf{4.59}^{\dagger}$	$\mathbf{4.60}^{\dagger}$
0.6	$2.88^{\dagger\dagger}$	$\mathbf{1.69}^{\dagger}$	$3.06^{\dagger}$	$3.82^{\dagger}$	$4.15^{\dagger}$	$4.28^{\dagger}$	$\mathbf{4.30}^{\dagger}$	$\mathbf{4.28}^{\dagger}$	$\mathbf{4.29}^{\dagger}$
0.7	$2.07^{\dagger\dagger}$	$\mathbf{1.08}^{\dagger}$	$1.95^{\dagger}$	$2.48^{\dagger}$	$2.77^{\dagger}$	$2.87^{\dagger}$	$2.93^{\dagger}$	$2.94^{\dagger}$	$2.94^{\dagger}$
0.8	$\mathbf{1.21}^{\dagger\dagger}$	$5.28^{\dagger\dagger}$	$9.08^{\dagger\dagger}$	$1.14^{\dagger}$	$1.28^{\dagger}$	$1.37^{\dagger}$	$1.42^{\dagger}$	$1.38^{\dagger}$	$1.39^{\dagger}$
0.9	$2.72^{\ddagger}$	$\mathbf{1.63}^{\dagger\dagger}$	$2.62^{\dagger\dagger}$	$3.24^{\dagger\dagger}$	$3.75^{\dagger\dagger}$	$3.30^{\dagger\dagger}$	$3.61^{\dagger\dagger}$	$3.54^{\dagger\dagger}$	$3.56^{\dagger\dagger}$
					$k=10$				
0.1	$\mathbf{1.64}^{\dagger\dagger}$	$\mathbf{1.59}^{\dagger}$	$\mathbf{1.59}^{\dagger}$	$\mathbf{1.24}^{\dagger}$	$\mathbf{1.21}^{\dagger}$	$\mathbf{1.42}^{\dagger}$	$\mathbf{1.87}^{\dagger}$	$\mathbf{2.93}^{\dagger}$	$\mathbf{5.49}^{\dagger}$
0.2	$\mathbf{4.54}^{\dagger\dagger}$	$\mathbf{2.87}^{\dagger}$	$\mathbf{3.38}^{\dagger}$	$\mathbf{3.07}^{\dagger}$	$\mathbf{3.25}^{\dagger}$	$\mathbf{3.89}^{\dagger}$	$\mathbf{5.09}^{\dagger}$	$\mathbf{6.90}^{\dagger}$	$\mathbf{6.94}^{\dagger}$
0.3	$\mathbf{6.66}^{\dagger\dagger}$	$\mathbf{2.99}^{\dagger}$	$\mathbf{4.06}^{\dagger}$	$\mathbf{4.13}^{\dagger}$	$\mathbf{4.42}^{\dagger}$	$\mathbf{5.21}^{\dagger}$	$\mathbf{5.79}^{\dagger}$	$\mathbf{5.76}^{\dagger}$	$\mathbf{5.76}^{\dagger}$
0.4	$\mathbf{6.59}^{\dagger\dagger}$	$\mathbf{2.46}^{\dagger}$	$3.50^{\dagger}$	$\mathbf{3.95}^{\dagger}$	$\mathbf{4.03}^{\dagger}$	$\mathbf{4.05}^{\dagger}$	$\mathbf{4.09}^{\dagger}$	$\mathbf{4.08}^{\dagger}$	$\mathbf{4.08}^{\dagger}$
0.5	$\mathbf{5.42}^{\dagger\dagger}$	$1.70^{\dagger}$	$2.38^{\dagger}$	$2.58^{\dagger}$	$2.63^{\dagger}$	$2.67^{\dagger}$	$2.62^{\dagger}$	$2.62^{\dagger}$	$2.58^{\dagger}$
0.6	$\mathbf{3.90}^{\dagger\dagger}$	$1.08^{\dagger}$	$1.43^{\dagger}$	$1.55^{\dagger}$	$1.54^{\dagger}$	$1.55^{\dagger}$	$1.54^{\dagger}$	$1.57^{\dagger}$	$1.53^{\dagger}$
0.7	$\mathbf{2.34}^{\dagger\dagger}$	$5.57^{\dagger\dagger}$	$7.17^{\dagger\dagger}$	$7.96^{\dagger\dagger}$	$8.06^{\dagger\dagger}$	$7.65^{\dagger\dagger}$	$7.71^{\dagger\dagger}$	$8.12^{\dagger\dagger}$	$8.21^{\dagger\dagger}$
0.8	$1.07^{\dagger\dagger}$	$2.46^{\dagger\dagger}$	$3.24^{\dagger\dagger}$	$3.43^{\dagger\dagger}$	$3.22^{\dagger\dagger}$	$3.35^{\dagger\dagger}$	$3.07^{\dagger\dagger}$	$3.27^{\dagger\dagger}$	$3.36^{\dagger\dagger}$
0.9	$\mathbf{3.61}^{\ddagger}$	$7.73^{\ddagger}$	$9.81^{\ddagger}$	$8.43^{\ddagger}$	$9.64^{\ddagger}$	$8.55^{\ddagger}$	$1.11^{\dagger\dagger}$	$1.22^{\dagger\dagger}$	$1.08^{\dagger\dagger}$

Table 6.10 Simulated E-optimality values when $k = 2, 4, 5, 10$ and $N = 500$. † indicates $\times 10^{-2}$, †† indicates $\times 10^{-3}$, ‡ indicates $\times 10^{-4}$, and $*$ indicates $\leq \times 10^{-4}$. Optimal values are highlighted in bold.

				$k=2$				
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**896.26**	**778.97**	**654.24**	537.67	426.37	329.07	249.59	195.71	160.91
0.2	**906.96**	**786.25**	**651.98**	534.19	434.74	356.58	303.85	280.14	260.10
0.3	**907.62**	**821.81**	684.90	569.10	479.19	418.84	387.05	375.15	359.62
0.4	**913.56**	**862.43**	737.58	628.67	549.34	499.63	479.09	474.73	460.63
0.5	**901.10**	**913.12**	808.37	708.96	633.80	593.31	578.79	575.47	560.41
0.6	**855.69**	**955.30**	**887.86**	799.01	735.36	697.76	684.19	680.17	663.64
0.7	**782.45**	**959.38**	**953.36**	**892.17**	836.86	804.77	793.81	786.90	767.22
0.8	**628.91**	**885.31**	**946.94**	**937.70**	**911.51**	893.55	**886.52**	**883.17**	**865.72**
0.9	**369.45**	**614.22**	**738.43**	**789.33**	**813.95**	**835.18**	**850.05**	**856.85**	**849.07**
				$k=4$					
0.1	873.36	733.73	584.39	470.36	394.26	357.90	333.74	281.76	198.74
0.2	866.87	718.11	575.23	482.49	445.61	442.24	430.35	379.28	295.22
0.3	869.81	741.48	610.42	541.47	529.31	538.74	530.42	482.38	397.91
0.4	858.53	781.55	672.26	623.75	625.03	641.25	632.53	583.60	501.38
0.5	845.46	827.41	751.25	717.65	730.07	746.03	736.18	690.27	607.87
0.6	805.19	869.02	824.18	806.81	827.03	842.05	837.49	790.10	714.52
0.7	729.02	858.92	863.86	872.95	898.17	919.22	**911.38**	**874.01**	**802.95**
0.8	577.62	766.27	812.90	854.32	886.90	**909.00**	907.38	878.47	818.60
0.9	339.04	509.91	586.85	636.27	672.65	701.02	696.09	683.54	642.77
				$k=5$					
0.1	857.20	710.83	560.54	460.68	420.37	405.70	367.86	**294.57**	**199.13**
0.2	849.99	689.01	554.60	496.18	498.03	501.15	466.67	392.51	**296.90**
0.3	845.72	715.08	599.95	573.70	593.36	602.89	569.17	495.80	402.67
0.4	838.45	758.41	672.24	664.37	694.54	706.63	672.43	604.01	507.47
0.5	824.13	805.72	754.57	762.08	797.61	808.20	**776.31**	**706.98**	617.18
0.6	775.39	841.71	819.71	846.30	883.54	**892.16**	**862.69**	**800.63**	**715.84**
0.7	706.63	820.29	848.28	883.28	**923.67**	**940.21**	913.65	857.70	787.16
0.8	557.75	715.11	776.32	831.88	876.33	884.25	866.50	830.17	772.20
0.9	324.01	466.10	534.01	583.98	612.51	637.13	619.27	599.90	563.73
				$k=10$					
0.1	805.58	623.87	575.22	**604.79**	**574.41**	**499.22**	**402.09**	305.45	205.85
0.2	776.10	630.15	658.61	**704.58**	**683.93**	**607.71**	**514.34**	**419.01**	322.32
0.3	751.01	683.95	**748.21**	**806.70**	**785.85**	**710.04**	**622.21**	**533.33**	**443.02**
0.4	741.44	738.58	**826.54**	**886.69**	**867.50**	**803.56**	**717.86**	**636.58**	**548.48**
0.5	721.29	775.58	**872.40**	**927.91**	**912.03**	**852.11**	784.20	705.44	**638.58**
0.6	675.55	766.22	872.00	**916.49**	**907.74**	856.92	801.77	732.27	676.81
0.7	588.69	713.52	813.69	855.26	835.26	807.97	755.97	703.13	645.97
0.8	451.13	573.51	649.88	686.55	679.40	652.64	626.89	585.59	545.77
0.9	260.92	338.90	387.53	410.67	402.66	391.06	367.79	350.42	333.05

Table 6.11 Simulated T-optimality values when $k=2,4,5,10$ and $N=500$. Optimal values are highlighted in bold.

					$k=2$				
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	3.15†	3.37†	2.78†	2.35†	1.80†	1.14†	7.45††	4.32††	2.11††
0.2	4.81†	3.98†	3.89†	2.62†	1.91†	1.32†	8.16††	4.46††	2.56††
0.3	3.93†	5.01†	3.77†	3.08†	2.00†	1.27†	7.84††	4.84††	2.71††
0.4	5.44†	5.12†	3.96†	2.80†	1.93†	1.17†	7.42††	4.70††	**2.68**††
0.5	4.46†	4.29†	4.06†	2.64†	1.85†	1.07†	7.03††	4.44††	**2.49**††
0.6	7.11†	4.87†	3.94†	2.60†	1.70†	1.07†	6.93††	4.26††	**2.33**††
0.7	6.22†	5.47†	4.00†	2.47†	1.56†	1.04†	6.73††	**4.21**††	**2.32**††
0.8	6.45†	5.66†	3.93†	2.50†	1.68†	1.12†	**7.09**††	**4.44**††	**2.49**††
0.9	**5.86**†	6.53†	5.09†	3.59†	2.37†	1.64†	**9.96**††	**6.47**††	**3.37**††
					$k=4$				
0.1	2.39†	2.02†	1.14†	7.11††	4.69††	3.49††	2.77††	2.22††	1.62††
0.2	**2.76**†	2.51†	1.35†	7.44††	5.00††	3.86††	3.19††	2.67††	2.08††
0.3	**3.29**†	2.58†	1.38†	7.49††	5.10††	3.99††	**3.39**††	**2.86**††	**2.29**††
0.4	4.33†	2.95†	1.40†	7.68††	5.50††	4.48††	**3.84**††	**3.31**††	2.70††
0.5	**3.97**†	2.76†	1.36†	7.89††	**5.79**††	**4.87**††	**4.16**††	**3.63**††	3.02††
0.6	**4.70**†	2.82†	1.39†	8.58††	**6.57**††	**5.45**††	**4.70**††	**4.11**††	3.48††
0.7	5.42†	3.17†	1.59†	**1.03**†	**7.98**††	**6.53**††	**5.71**††	5.08††	4.33††
0.8	**5.41**†	3.78†	**2.03**†	**1.34**†	**1.06**†	**8.96**††	7.81††	7.01††	6.15††
0.9	6.24†	**5.02**†	**3.30**†	**2.29**†	**1.86**†	**1.57**†	1.41†	1.25†	1.08†
					$k=5$				
0.5	5.15†	**2.48**†	**1.07**†	**7.01**††	5.80††	5.12††	4.81††	4.40††	3.83††
0.6	5.34†	**2.63**†	**1.20**†	**8.41**††	7.00††	6.28††	5.94††	5.46††	4.85††
0.7	**5.30**†	**2.99**†	**1.52**†	1.10†	9.05††	8.30††	7.82††	7.40††	6.51††
0.8	5.57†	**3.66**†	2.15†	1.57†	1.32†	1.19†	1.15†	1.06†	9.99††
0.9	6.65†	5.91†	3.54†	2.78†	2.34†	2.18†	2.03†	1.90†	1.81†
					$k=10$				
0.1	**2.24**†	**5.72**††	**2.72**††	**2.26**††	**2.37**††	**2.52**††	2.51††	2.33††	1.86††
0.2	3.35†	**8.89**††	**4.72**††	**4.06**††	4.09††	4.09††	4.17††	3.94††	3.48††
0.3	4.68†	**1.29**†	**7.08**††	6.57††	6.52††	6.68††	6.53††	6.48††	6.10††
0.4	6.40†	**1.79**†	1.11†	1.00†	1.01†	1.02†	1.02†	9.85††	9.62††
0.5	7.54†	2.48†	1.63†	1.52†	1.59†	1.54†	1.58†	1.53†	1.49†
0.6	8.74†	3.58†	2.47†	2.35†	2.35†	2.36†	2.25†	2.33†	2.25†
0.7	1.04*	5.27†	3.83†	3.26†	3.61†	3.24†	3.49†	3.61†	3.36†
0.8	1.14*	7.37†	5.19†	5.07†	5.51†	5.08†	5.09†	5.09†	4.84†
0.9	9.60†	1.09*	8.36†	8.14†	8.39†	7.15†	8.26†	8.19†	6.82†

Table 6.12 Simulated A-optimality values when $k = 2,4,5,10$ and $N = 100$. † indicates $\times 10^{-2}$, †† indicates $\times 10^{-3}$ and ‡ indicates $\times 10^{-4}$. Optimal values are highlighted in bold.

$k=2$									
ϕ \ p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.21	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.2	9.19	5.63	0.00	0.00	0.00	0.00	0.01	0.01	0.01
0.3	0.92	30.86	0.90	0.01	0.01	0.01	0.01	0.03	0.02
0.4	4.22	41.78	34.59	0.07	0.02	0.02	0.03	0.03	0.03
0.5	3.06	44.85	102.28	63.82	5.02	0.06	0.07	0.08	0.04
0.6	2.51	44.88	204.62	253.67	218.61	134.52	57.15	8.79	0.13
0.7	0.03	41.53	190.05	469.65	670.67	756.70	751.75	624.69	402.84
0.8	0.02	9.03	101.41	371.61	703.66	1074.81	1357.51	**1415.64**	**1094.98**
0.9	0.00	0.14	12.14	49.67	**148.30**	**276.43**	**450.44**	**525.78**	**430.96**
$k=4$									
0.1	3.04	2.40	0.00	0.00	0.00	0.01	0.01	0.01	0.01
0.2	2.33	23.67	2.86	0.02	0.03	0.04	0.06	0.07	0.04
0.3	8.73	112.81	100.60	15.05	0.11	0.09	0.08	0.07	0.03
0.4	10.64	194.42	374.96	355.73	270.58	175.49	106.48	65.04	27.05
0.5	21.34	312.96	732.27	1066.45	1261.23	1335.79	1238.57	857.62	440.97
0.6	5.37	269.87	910.25	1726.24	2430.28	2819.55	2575.60	2008.17	**977.38**
0.7	1.21	139.17	681.19	1496.81	2454.63	**2994.04**	**2982.37**	**2307.07**	**1202.16**
0.8	0.09	39.85	253.11	**731.48**	**1162.94**	**1483.81**	**1511.72**	1218.42	641.23
0.9	0.00	1.24	**22.78**	**83.07**	146.20	203.46	199.93	166.66	73.61
$k=5$									
0.1	2.01	3.64	0.00	0.00	0.01	0.02	0.02	0.02	0.01
0.2	9.29	54.37	18.26	0.03	0.04	0.07	0.07	0.09	0.04
0.3	32.33	197.93	252.98	150.97	77.98	38.51	9.01	1.33	3.68
0.4	17.84	368.58	764.55	918.27	1036.02	916.47	799.79	564.44	279.63
0.5	14.79	396.32	1337.65	1896.80	**2626.85**	**2445.05**	**2227.56**	**1566.27**	**748.86**
0.6	8.60	309.94	**1377.96**	**2468.14**	**3097.01**	**3573.75**	**3136.41**	**2214.24**	1095.70
0.7	2.06	**177.98**	**855.21**	**1681.76**	**2434.62**	2769.28	2517.34	1721.41	880.28
0.8	0.10	**44.33**	**274.69**	634.43	957.83	1088.63	963.00	724.04	338.78
0.9	0.01	**2.87**	25.42	63.82	99.44	120.30	116.92	69.34	29.84
0.1	**11.43**	**69.51**	**58.21**	**13.72**	**13.04**	**11.44**	**7.48**	**4.52**	**3.40**
0.2	**37.55**	**426.25**	**871.89**	**911.76**	**879.38**	**635.97**	**534.40**	**348.98**	**171.41**
0.3	**65.76**	**857.30**	**1776.85**	**2245.56**	**2108.36**	**1806.24**	**1334.57**	**875.67**	**422.47**
0.4	**58.95**	**971.72**	**2186.03**	**2830.12**	**2762.42**	**2285.53**	**1775.66**	**1149.48**	**544.16**
0.5	**34.59**	**782.58**	**1849.88**	**2326.80**	2310.17	1924.18	1459.60	942.86	465.62
0.6	**13.83**	**459.23**	1114.24	1479.76	1373.55	1135.31	902.11	557.92	287.40
0.7	**4.56**	152.70	460.87	630.27	593.90	521.79	351.02	240.15	110.08
0.8	**0.46**	33.67	117.69	163.16	156.86	139.03	88.84	59.96	28.42
0.9	**0.01** 2.88	8.82	13.96	17.08	10.92	6.86	4.01	1.36	

Table 6.13 Simulated D-optimality values when $k = 2, 4, 5, 10$ and $N = 100$. Optimal values are highlighted in bold.

$k=2$									
ϕ \ p	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
0.1	1.29^{*}	8.06^{*}	1.95^{*}	1.43^{*}	1.63^{*}	2.83^{*}	7.82^{*}	2.03^{*}	6.29^{*}
0.2	6.78^{*}	1.07^{*}	4.25^{*}	1.97^{*}	2.15^{*}	5.33^{*}	1.05^{*}	2.02^{*}	2.81^{*}
0.3	5.74^{*}	4.01^{*}	8.17^{*}	4.06^{*}	3.69^{*}	5.30^{*}	7.14^{*}	1.16^{*}	1.29^{*}
0.4	3.54^{*}	3.35^{*}	9.63^{*}	2.40^{*}	5.31^{*}	5.61^{*}	5.82^{*}	7.70^{*}	7.63^{*}
0.5	3.06^{*}	3.27^{*}	$1.20^{\ddagger}$	$1.16^{\ddagger}$	9.75^{*}	7.66^{*}	7.97^{*}	8.98^{*}	6.12^{*}
0.6	4.67^{*}	2.22^{*}	$1.36^{\ddagger}$	$2.54^{\ddagger}$	$1.99^{\ddagger}$	$1.05^{\ddagger}$	3.73^{*}	5.44^{*}	1.04^{*}
0.7	2.40^{*}	2.63^{*}	$1.06^{\ddagger}$	$2.85^{\ddagger}$	$3.70^{\ddagger}$	$3.63^{\ddagger}$	$3.17^{\ddagger}$	$2.56^{\ddagger}$	$2.27^{\ddagger}$
0.8	4.28^{*}	1.45^{*}	6.88^{*}	$1.84^{\ddagger}$	$2.97^{\ddagger}$	$3.96^{\ddagger}$	$4.26^{\ddagger}$	$4.32^{\ddagger}$	$4.24^{\ddagger}$
0.9	2.82^{*}	2.96^{*}	4.11^{*}	7.61^{*}	$1.50^{\ddagger}$	$1.90^{\ddagger}$	$2.38^{\ddagger}$	$2.51^{\ddagger}$	$2.54^{\ddagger}$
$k=4$									
0.1	$1.77^{\ddagger}$	$8.31^{\ddagger}$	1.74^{*}	1.67^{*}	1.59^{*}	1.50^{*}	1.98^{*}	4.19^{*}	1.03^{*}
0.2	$2.16^{\ddagger}$	$4.29^{\dagger\dagger}$	$8.37^{\ddagger}$	2.69^{*}	2.78^{*}	3.48^{*}	4.97^{*}	9.81^{*}	9.33^{*}
0.3	$5.34^{\ddagger}$	$1.16^{\dagger}$	$1.02^{\dagger}$	$1.24^{\dagger\dagger}$	5.64^{*}	4.33^{*}	4.38^{*}	4.73^{*}	4.20^{*}
0.4	$9.07^{\ddagger}$	$1.27^{\dagger}$	$2.36^{\dagger}$	$1.56^{\dagger}$	$8.99^{\dagger\dagger}$	$5.05^{\dagger\dagger}$	$3.12^{\dagger\dagger}$	$2.41^{\dagger\dagger}$	$2.11^{\dagger\dagger}$
0.5	$1.79^{\dagger\dagger}$	$1.64^{\dagger}$	$3.07^{\dagger}$	$3.13^{\dagger}$	$2.73^{\dagger}$	$2.54^{\dagger}$	$2.33^{\dagger}$	$2.17^{\dagger}$	$2.21^{\dagger}$
0.6	$8.82^{\ddagger}$	$1.28^{\dagger}$	$\mathbf{2.94}^{\dagger}$	$3.59^{\dagger}$	$3.78^{\dagger}$	$3.82^{\dagger}$	$3.75^{\dagger}$	$3.71^{\dagger}$	$3.70^{\dagger}$
0.7	$3.79^{\ddagger}$	$8.33^{\dagger\dagger}$	$\mathbf{2.18}^{\dagger}$	$\mathbf{3.02}^{\dagger}$	$\mathbf{3.31}^{\dagger}$	$\mathbf{3.57}^{\dagger}$	$\mathbf{3.61}^{\dagger}$	$\mathbf{3.64}^{\dagger}$	$\mathbf{3.62}^{\dagger}$
0.8	7.25^{*}	$4.70^{\dagger\dagger}$	$\mathbf{1.12}^{\dagger}$	$\mathbf{1.74}^{\dagger}$	$\mathbf{2.08}^{\dagger}$	$\mathbf{2.14}^{\dagger}$	$\mathbf{2.34}^{\dagger}$	$\mathbf{2.28}^{\dagger}$	$\mathbf{2.30}^{\dagger}$
0.9	6.83^{*}	$9.62^{\ddagger}$	$5.11^{\dagger\dagger}$	$\mathbf{7.16}^{\dagger\dagger}$	$\mathbf{8.25}^{\dagger\dagger}$	$\mathbf{9.55}^{\dagger\dagger}$	$\mathbf{9.25}^{\dagger\dagger}$	$\mathbf{8.97}^{\dagger\dagger}$	$\mathbf{7.36}^{\dagger\dagger}$
$k=5$									
0.1	$1.52^{\ddagger}$	$8.35^{\ddagger}$	1.98^{*}	1.27^{*}	1.20^{*}	2.01^{*}	2.91^{*}	7.13^{*}	1.12^{*}
0.2	$6.27^{\ddagger}$	$7.57^{\dagger\dagger}$	$2.90^{\dagger\dagger}$	3.22^{*}	3.00^{*}	3.97^{*}	5.31^{*}	1.20^{*}	1.03^{*}
0.3	$2.08^{\dagger\dagger}$	$1.69^{\dagger}$	$1.72^{\dagger}$	$6.87^{\dagger\dagger}$	$3.04^{\dagger\dagger}$	$1.38^{\dagger\dagger}$	$4.00^{\ddagger}$	$1.23^{\ddagger}$	$5.31^{\ddagger}$
0.4	$1.43^{\dagger\dagger}$	$2.17^{\dagger}$	$3.14^{\dagger}$	$2.72^{\dagger}$	$2.39^{\dagger}$	$2.19^{\dagger}$	$2.12^{\dagger}$	$2.10^{\dagger}$	$2.11^{\dagger}$
0.5	$1.83^{\dagger\dagger}$	$1.81^{\dagger}$	$\mathbf{3.42}^{\dagger}$	$\mathbf{3.85}^{\dagger}$	$\mathbf{4.03}^{\dagger}$	$\mathbf{3.92}^{\dagger}$	$\mathbf{3.90}^{\dagger}$	$\mathbf{3.82}^{\dagger}$	$\mathbf{3.83}^{\dagger}$
0.6	$1.15^{\dagger\dagger}$	$\mathbf{1.36}^{\dagger}$	$2.91^{\dagger}$	$\mathbf{3.72}^{\dagger}$	$\mathbf{4.00}^{\dagger}$	$\mathbf{4.06}^{\dagger}$	$\mathbf{4.04}^{\dagger}$	$\mathbf{4.05}^{\dagger}$	$\mathbf{4.04}^{\dagger}$
0.7	$6.78^{\ddagger}$	$\mathbf{9.26}^{\dagger\dagger}$	$1.97^{\dagger}$	$2.58^{\dagger}$	$2.97^{\dagger}$	$2.95^{\dagger}$	$3.05^{\dagger}$	$3.02^{\dagger}$	$3.21^{\dagger}$
0.8	$1.24^{\ddagger}$	$\mathbf{4.81}^{\dagger\dagger}$	$9.67^{\dagger\dagger}$	$1.44^{\dagger}$	$1.58^{\dagger}$	$1.60^{\dagger}$	$1.70^{\dagger}$	$1.76^{\dagger}$	$1.64^{\dagger}$
0.9	8.47^{*}	$1.95^{\dagger\dagger}$	$\mathbf{5.27}^{\dagger\dagger}$	$6.91^{\dagger\dagger}$	$6.09^{\dagger\dagger}$	$6.71^{\dagger\dagger}$	$7.78^{\dagger\dagger}$	$6.92^{\dagger\dagger}$	$5.98^{\dagger\dagger}$
$k=10$									
0.1	$\mathbf{6.85}^{\ddagger}$	$\mathbf{8.99}^{\dagger\dagger}$	$\mathbf{4.80}^{\dagger\dagger}$	$\mathbf{7.56}^{\ddagger}$	$\mathbf{7.10}^{\ddagger}$	$\mathbf{8.57}^{\ddagger}$	$\mathbf{9.82}^{\ddagger}$	$\mathbf{1.44}^{\dagger\dagger}$	$\mathbf{3.75}^{\dagger\dagger}$
0.2	$\mathbf{2.73}^{\dagger\dagger}$	$\mathbf{2.27}^{\dagger}$	$\mathbf{2.66}^{\dagger}$	$\mathbf{2.24}^{\dagger}$	$\mathbf{2.31}^{\dagger}$	$\mathbf{2.59}^{\dagger}$	$\mathbf{3.53}^{\dagger}$	$\mathbf{5.00}^{\dagger}$	$\mathbf{5.37}^{\dagger}$
0.3	$\mathbf{4.11}^{\dagger\dagger}$	$\mathbf{2.58}^{\dagger}$	$\mathbf{3.46}^{\dagger}$	$\mathbf{3.58}^{\dagger}$	$\mathbf{3.76}^{\dagger}$	$\mathbf{4.33}^{\dagger}$	$\mathbf{4.99}^{\dagger}$	$\mathbf{5.35}^{\dagger}$	$\mathbf{5.35}^{\dagger}$
0.4	$\mathbf{3.55}^{\dagger\dagger}$	$\mathbf{2.29}^{\dagger}$	$\mathbf{3.37}^{\dagger}$	$\mathbf{3.64}^{\dagger}$	$\mathbf{3.84}^{\dagger}$	$\mathbf{3.99}^{\dagger}$	$\mathbf{4.20}^{\dagger}$	$\mathbf{4.31}^{\dagger}$	$\mathbf{4.22}^{\dagger}$
0.5	$\mathbf{3.08}^{\dagger\dagger}$	$\mathbf{1.83}^{\dagger}$	$2.68^{\dagger}$	$2.95^{\dagger}$	$2.80^{\dagger}$	$2.93^{\dagger}$	$2.92^{\dagger}$	$2.92^{\dagger}$	$2.82^{\dagger}$
0.6	$\mathbf{1.66}^{\dagger\dagger}$	$1.27^{\dagger}$	$1.85^{\dagger}$	$1.82^{\dagger}$	$1.77^{\dagger}$	$1.80^{\dagger}$	$1.98^{\dagger}$	$1.90^{\dagger}$	$1.91^{\dagger}$
0.7	$\mathbf{1.10}^{\dagger\dagger}$	$7.30^{\dagger\dagger}$	$1.04^{\dagger}$	$1.24^{\dagger}$	$9.75^{\dagger\dagger}$	$1.30^{\dagger}$	$1.11^{\dagger}$	$9.75^{\dagger\dagger}$	$1.13^{\dagger}$
0.8	$\mathbf{3.82}^{\ddagger}$	$4.04^{\dagger\dagger}$	$7.57^{\dagger\dagger}$	$5.65^{\dagger\dagger}$	$6.82^{\dagger\dagger}$	$7.88^{\dagger\dagger}$	$7.62^{\dagger\dagger}$	$7.45^{\dagger\dagger}$	$6.13^{\dagger\dagger}$
0.9	$\mathbf{1.00}^{*}$	$\mathbf{3.18}^{\dagger\dagger}$	$5.05^{\dagger\dagger}$	$6.11^{\dagger\dagger}$	$6.80^{\dagger\dagger}$	$6.81^{\dagger\dagger}$	$6.12^{\dagger\dagger}$	$3.07^{\dagger\dagger}$	$2.27^{\dagger\dagger}$

Table 6.14 Simulated E-optimality values when $k = 2, 4, 5, 10$ and $N = 100$. † indicates $\times 10^{-2}$, †† indicates $\times 10^{-3}$, ‡ indicates $\times 10^{-4}$, and $*$ indicates $\leq \times 10^{-4}$. Optimal values are highlighted in bold.

				$k=2$					
ϕ \ p	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
0.10	**88.74**	**76.96**	**65.47**	55.34	42.35	33.61	25.86	20.27	15.71
0.20	**89.45**	**79.10**	65.28	55.75	44.54	37.39	31.71	27.68	25.54
0.30	**88.67**	**82.02**	69.34	58.15	49.48	42.87	39.10	37.15	35.98
0.40	**87.68**	**85.48**	73.36	62.85	55.92	50.93	47.98	46.92	45.37
0.50	**85.35**	**87.84**	80.47	70.19	62.45	59.10	57.20	56.81	55.31
0.60	**79.90**	**90.31**	**86.21**	78.29	71.85	67.95	66.45	66.16	64.73
0.70	**68.08**	**89.65**	**90.51**	**86.38**	81.94	79.00	77.07	76.24	74.67
0.80	**52.76**	**79.10**	**88.04**	**89.83**	**90.46**	87.14	86.25	**85.47**	**84.52**
0.90	**28.20**	**52.04**	**65.17**	**73.75**	**77.67**	**81.61**	**81.43**	**82.85**	**81.60**
				$k=4$					
0.10	85.33	74.32	59.98	47.07	40.41	36.31	33.03	28.14	19.26
0.20	84.88	74.08	58.01	49.09	44.97	44.09	42.51	37.75	29.15
0.30	85.68	74.81	60.63	53.63	51.62	52.25	51.87	47.04	38.76
0.40	83.80	79.00	66.61	60.87	60.89	61.64	61.14	56.34	48.35
0.50	81.54	82.11	73.56	69.28	70.35	72.12	71.72	66.31	58.43
0.60	74.91	84.00	80.37	77.55	79.97	82.12	80.24	76.19	68.03
0.70	65.97	82.76	82.68	83.29	87.69	88.76	88.63	**83.46**	**77.58**
0.80	50.29	68.72	77.86	80.85	83.76	**88.22**	**85.90**	83.83	78.79
0.90	26.58	44.28	52.82	59.65	62.91	62.38	64.72	62.77	61.80
				$k=5$					
0.10	84.75	73.44	56.95	46.90	42.70	40.41	36.25	28.86	19.80
0.20	84.00	70.27	56.18	49.53	48.70	49.27	45.49	38.34	29.55
0.30	83.82	72.89	59.52	55.31	56.99	58.13	54.57	47.08	38.17
0.40	82.11	75.21	66.49	63.98	67.47	67.91	65.11	57.84	49.11
0.50	80.74	79.04	73.78	72.65	78.19	77.35	74.91	**67.63**	58.49
0.60	72.92	81.47	80.68	82.77	83.71	**86.27**	83.58	**76.96**	**68.27**
0.70	65.15	78.05	81.92	85.87	**88.12**	**91.90**	88.09	82.60	74.09
0.80	47.29	65.46	75.81	78.42	80.97	86.12	81.88	78.36	72.52
0.90	24.52	39.86	46.11	51.83	56.34	57.23	55.47	53.43	51.49
				$k=10$					
0.10	82.72	64.36	57.29	**57.94**	**55.21**	**48.20**	**38.35**	28.36	18.83
0.20	78.43	64.42	**63.90**	**67.28**	**65.38**	**56.95**	**48.30**	**39.27**	29.37
0.30	76.99	66.34	**71.12**	**76.88**	**73.94**	**67.91**	**59.15**	**49.74**	**41.35**
0.40	73.87	71.18	**78.18**	**83.94**	**82.07**	**75.68**	**68.41**	**59.50**	**51.78**
0.50	66.82	73.97	**80.97**	**87.07**	**87.29**	**80.37**	**74.40**	66.60	**58.80**
0.60	63.01	71.46	79.02	**86.88**	**86.01**	79.73	74.66	66.67	61.11
0.70	53.17 64.49	74.89	76.04	79.67	72.04	68.94	64.73	58.20	
0.80	38.67	50.21	55.95	62.56	58.88	57.26	51.53	48.98	46.22
0.90	17.54	26.33	28.20	31.77	29.91	29.47	28.35	27.95	25.45

Table 6.15 Simulated T-optimality values when $k=2,4,5,10$ and $N=100$. Optimal values are highlighted in bold.

6.5 Discussion

Many ecologists have extensive experience with the design of natural resource surveys from an empirical perspective, however it is also important to plan studies based on the properties of estimators, such as their variance, to sample adequate data that are usable for modelling and improve precision of parameter estimates. Incorporation of the robust design is highly preferable when sampling removal data because it overcomes the issues of parameter redundancy (Zhou et al., 2018). However, no work has been done on how to optimally design the removal sampling under the robust design.

Bohrmann and Christman (2013) present the first detailed investigation of study design in the context of classic removal modelling based on analytical consideration of the large sample properties of the capture probability and simulations, in particular the allocation of sampling effort into number of samples sites and amount of replication within each site.

We investigate the ways for choosing the optimal design for removal sampling accounting for temporary emigration analytically. This assessment can help ecologists to decide how many secondary samples and primary periods should be conducted to minimise the variance of parameters for a fixed level of total survey effort. We explored four criteria for optimal design for removal data with robust design and we concluded that the best performing criterion is A-optimality as theoretical results are supported by simulation results, but only when population size is large enough.

We recommend sampling more primary periods with less number of secondary samples for rare species, as such a population tends to stay outside the study area (i.e. $\phi > 0.5$ on average). Similarly, surveying fewer primary period with more secondary samples is better for common species when $\phi < 0.5$ is assumed on average.

It is always useful to perform simulation-based studies before conducting the actual field survey not only to evaluate the design in terms of its ability to generate useful estimates, but also so that you have an expectation of what the data will look like as they are being sampled. This would give you the ability to recognise some pathologies and possibly intervene to resolve issues before they conduct a whole study.

For robust design removal sampling we assume the population is closed between secondary samples within each primary period, with periods of closure assumed to be short relative to the life history of the species under study. If the closure assumption in the secondary periods is violated we would expect the capture probability to be underestimated, and could potentially lead to positive bias in the estimate of population size. In practice it can be difficult to guarantee a closed population, for example a dispersing individual may arrive just as observers leave the study site. For some species we may be able to avoid seasons where violation of demographic closure is highly likely happening such as during migration seasons in migratory birds or during the start and end of breeding seasons. However, for many species such ecological seasons might be less clear cut. For instance, rats can easily breed throughout the year if conditions are suitable. As a consequence, guidelines of what time frame adequately gives rise to a closed population can be vague. To choose a time that is appropriate, a prior study of biological behavior of the species subject to removal before the survey is highly recommended.

The frequency or length of temporal samples is one of the most important aspects of sampling design for population sampling. However, there are other aspects that should be considered in sampling. The importance of adequate trap placing has been investigated in capture-recapture studies. Dice (1938, 1941) recognized the importance of population home ranges and highlighted the need to distribute the traps in a way so that the array of traps exposes as many individuals as possible. This lead to the development of spatial models.

In order to initiate a study, we have to choose a study area which contains target populations that you may be able to capture and remove. For a translocation study, the area is likely to comprise of an area which is going to built on and its surrounding area which contains individuals that are subject to removal. Given a well-defined study area, we prefer a design with the arrangement of collecting traps that are capable of sampling more data. The design we present in this chapter assumes that the survey area is carefully chosen and the traps are able to cover home ranges of individuals in the population.

Even if we have carefully designed how to place the traps and to choose the study area, as a practical matter, some species might still have a small probability of being captured. Our models can only apply to individuals that have capture probabilities that are consistent with the model being considered. If there are individuals with $p = 0$, they cannot be estimated as the existence of a hole is not addressed in the model. Therefore the use of an optimal study design may not lead to precise estimation of the population size, as the RMER.SRC model assumes a homogeneous constant capture probability.

There are also other practical difficulties for sampling. Depending on environmental conditions, potentially there are parts of the study area which may be not accessible to field workers, due to dense vegetation cover or deep sea. Even when accessible, setting up traps in difficult habitat conditions can cost large amounts of time and resources.

In most cases, study design faces a limit of available resources or being logistically manageable. In the future, we could generate a framework incorporated with other considerations, e.g. logistics, costs and other practical restrictions. It is also useful to perform preliminary field studies before conducting the actual sampling, which would not only evaluate the design in terms of its ability to generate meaningful estimates, but also give you an opportunity to highlight potential issues.

Chapter 7

Conclusion

7.1 Discussion

Estimating the population size is of fundamental interest when studying animals for conservation science. The use of removal sampling is increasingly being used as it recently has been adopted as a conservation management tool for translocation projects and invasive species removal.

In Chapter 2 we discussed the limitation of the classic removal model, which assumes constant capture probability and that all individuals are within the sampling area throughout the study (Moran, 1951; Zippin, 1956, 1958), but these assumptions are often violated. If constant capture probability and perfect availability are incorrectly assumed, the estimate of population size will be positively biased. It is crucial to avoid bias of estimates in the model as biased estimators may lead to misleading conclusions for conservation purposes.

In Chapter 3, we developed new removal models accounting for temporary emigration for removal data sets with a robust design structure. The underlying movement pattern of individuals between the study area and an area outside of the study was modelled in a multievent framework.

In Chapter 4, we investigated which parameters of the models proposed in Chapter 3 can be individually estimated, i.e. whether or not a model is parameter redundant. Symbolic algebra is used to investigate parameter redundancy and to find the estimable combinations of parameters if the proposed model is parameter redundant. The methods presented in Chapter 4 provide almost identical findings to Chapter 3, without the need for intensive simulations.

Although we showed the benefits of the use of robust design in Chapters 3 and 4, a lot of removal data sets are sampled from a standard sampling protocol with only a single sampling occasion within each primary period. In Chapter 5, we proposed penalised likelihood approaches to model new arrivals of individuals, which allowed individuals to enter the study area during the removal study. We investigated three types of penalty terms, ridge, Lasso and fused Lasso for the proposed model. We demonstrated that the new approaches reduce the frequency of boundary estimates obtained from optimising the likelihood without a penalty term. However they have to be used with caution in cases of small sample sizes.

As the robust design is highly preferred for removal sampling as we discussed in Chapters 3 and 4, we investigated the survey design aspects of removal sampling at a single site while accounting for the availability of individuals under the robust design in Chapter 6. We used both simulations and analytic criteria to investigate how to allocate the total sampling effort for the removal sampling under the robust design sampling protocol where constant parameters are assumed.

7.2 Future work

We hope that the work conducted in this thesis will make a valuable contribution to modelling removal data and help guide removal studies in the future. There is still work to be done in the future on removal models. We list potential work which can be investigated below:

- This thesis presents new models for temporary emigration and new arrivals of individuals in the population as populations are rarely closed during sampling. These models can be used to deal with lack of closure and a better fit to the data than that obtained from the classic removal model may provide evidence that the collected data does not satisfy the closure assumption. However, it requires us to fit models to the data in order to conduct model comparison. Currently there is no statistical test for population closure for removal data. A number of such tests have been developed in traditional capture recapture studies. Otis et al. (1978) developed a test for population closure that can handle heterogeneity in capture probability, however the performance in the presence of time or behavioral variation is not promising. Stanley and Burnham (1999) developed a closure test for the closed population model M_t with time-varying p_t, which performs well when emigration is permanent and the majority of the population migrate. Both tests can be implemented in the program `CloseTest` (Stanley and Richards, 2005). However, it is not fully understood whether these tests can be adapted for removal data. In the future we could develop a closure test specifically for removal data, which would provide evidence of whether the closure assumption has been violated or not.

- For the robust design removal models explored in Chapter 3, simulation studies suggest that when capture probability is small and ϕ_i^{12} is large, the variances of fully time-dependent transition probabilities tend to be large. Similar results are observed when analysing real data. The poor precision is likely due to low capture probability (0.18 on average) and low availability of individuals ($\hat{\phi}_i^{12}$ is 0.70 and 0.57 on average for juveniles and adults respectively). In the future, we could propose an alternative method of estimation to improve the precision of estimates for the robust design removal models based on penalised likelihood.
- The penalised likelihood approaches investigated for removal data sets without the robust design structure are computationally intensive due to the required cross-validation. One possible way to resolve the problem is when other auxiliary data are available, we can integrate that information with removal data sets. For example, if the study area has been monitored by capture recapture sampling for a number of years, those data can potentially be considered in the model fitting. Alternatively, we could conduct concurrent capture-recapture sampling as suggested in Gould and Pollock (1997) or a few capture-recapture sampling occasions prior to removal sampling. We could evaluate modelling approaches and survey design in the presence of such data in the future.
- Prior information or guesses of parameter values is a typical requirement in the development of sampling design for statistical models. In Chapter 6, we have explored study design aspects of removal sampling under robust design and the results heavily depend on the true values of the parameter we used. These parameter values can be obtained from a systematic review of the literature or a pilot study. Potentially, meta-analysis can be performed for a family of species with similar biological behaviors that are analysed in previous studies.
- Although an idea of true values of parameters can be obtained from careful literature review, in most situations precise prior knowledge can be rarely available, as previous results may depend on combinations of environmental factors which are potentially different from those of the proposed study. Several studies have explored robust versions of the classic optimality criteria such as minimax D-optimality criteria, which is less sensitive with respect to the choice of parameters (Pronzato and Walter, 1985; Dette, 1997). Therefore, in the future we could explore other study design criteria for removal models with robust design.

- In the future, we could develop a user-friendly software program that is reproducible and efficient while being accessible and intuitive for applied practitioners. Similar ideas have been executed for fitting capture recapture models, e.g. `Mark` (White and Burnham, 1999) and `E-SURGE` (Choquet, 2009), and occupancy models, e.g. `Presence` (MacKenzie, 2018). We are currently developing web-based software for the robust design removal model. This program will be be easier to use for practitioners without prior programming knowledge in `R`.

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. in b. n. petrov and b. f. csaki (eds). *Second International Symposium on Information Theory*.

Asprey, S. P. and Macchietto, S. (2000). Statistical tools for optimal dynamic model building. *Computers and Chemical Engineering*, 24:1261–1267.

Atkinson, A., Donev, A., and Tobias, R. (2007). *Optimum Experimental Designs, with SAS (Oxford Statistical Science Series) 2nd edition*. Oxford University Press, USA.

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*. Clarendon Press, Oxford.

Atkinson, A. C. and Fedorov, V. V. (1975). The design of experiments for discriminating between two rival models. *Biometrika*, 62:57–70.

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58:540–547.

Bohrmann, T. F. and Christman, M. C. (2013). Optimal allocation of sampling effort in depletion surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(12):218–233.

Bonet, B. (2002). An e-optimal grid-based algorithm for partially obserable markov decision processes. *ICML*.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. Taylor & Francis.

Buckland, S. T. (1980). A modified analysis of the jolly-seber capture-recapture model. *Biometrics*, 36:419–435.

Buckland, S. T. and Garthwaite, P. H. (1991). Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics*, 47:255–268.

Burgiel, S. W. and Muir, A. A. (2010). *Invasive species, climate change and ecosystem-based adaptation: addressing multiple drivers of global change*. Global Invasive Species Programme, Washington DC.

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross validation and the repeated learing testing-model methods. *Bometrika*, 76:503–514.

Carle, F. and Strub, M. (1978). A new method for estimating population size from removal data. *Biometrics*, 34(4):621–630.

Carothers, A. D. (1973). The effect of unequal catchability on jolly-seber estimates. *Biometrics*, 29(1):79–100.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. N. Scituate: Duxbury.

Catchpole, E., Freeman, S. N., and Morgan, B. J. T. (1998). Estimation in parameter redundant models. *Biometrika*, 85:462–468.

Catchpole, E. and Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrics*, 84:187–196.

Catchpole, E., Morgan, B. J. T., and Viallefont, A. (2002). Solving problems in parameter redundancy using computer algebra. *Journal of Applied Statistics*, 29:625–636.

Catchpole, E. A., Kgosi, P. M., and Morgan, B. J. T. (2001). On the near-singularity of models for animal recovery data. *Biometrics*, 57:720–726.

Chandler, R. B., Royle, J. A., and King, D. J. (2011). Inference about density and temporary emigration in unmarked populations. *Ecology*, 92:1429–1435.

Chappell, M. J. and Gunn, R. N. (1998). A procedure for generating locally identifiable reparameterisations of unidentifiable non-linear systems by the similarity transformation approach. *Mathematical Biosciences*, 148:21–41.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 24:586–602.

Chis, O.-T., Villaverde, A. F., Banga, J. R., and Balsa-Canto, E. (2016). On the relationship between sloppiness and identifiability. *Mathematical Biosciences*, 282:147–161.

Choquet, R., R. L. . P. R. (2009). *Program E-SURGE: a software application for fitting multievent models*.

Clair, K., Dunton, E., and Giudice, J. (2013). A comparison of models using removal effort to estimate animal abundance. *Journal of Applied Statistics*, 40:527–545.

Cole, D. J. (2012). Determining parameter redundancy of multi-state mark-recapture models for sea birds. *Journal of Ornithology*, 152:305–315.

Cole, D. J., Morgan, B. J. T., Catchpole, E. A., and Hubbard, B. A. (2012). Parameter redundancy in mark-recovery models. *Biometrical Journal*, 54:507–523.

Cole, D. J., Morgan, B. J. T., McCrea, R. S., Pradel, R., Gimenez, O., and Choquet, R. (2014). Does your species have memory? analysing capture-recapture data with memory models. ecology and evolution. *Ecology and Evolution*, 4:2124–2133.

Cole, D. J., Morgan, B. J. T., and Titterington, D. M. (2010). The parametric structure of models. *Mathematical Biosciences*, 228:16–30.

Cormack, R. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, 51:429–438.

Covey-Crump, P. A. K. and Silvey, S. D. (1970). Optimal regression designs with previous observations. *Biometrika*, 57:551–566.

Cowx, I. G. (1983). Review of the methods for estimating fish population size from survey removal data. *Aquaculture Research*, 14:67–82.

Cox, D. R. and Miller, H. D. (1965). *The Theory of Stochastic Process*. Chapman & Hall, London.

Dette, H. (1997). Designing experiments with respect to "standardized" optimality criteria. *Journal of the Royal Statistical Society, Ser. B.*, 59:97–110.

Dette, H., Melas, V. B., and Shpilev, P. (2012). T -optimal designs for discrimination between two polynomial models. *The Annals of Statistics*, 10:188–205.

Dice, L. R. (1938). Some census methods for mammals. *The Journal of Wildlife Management*, 2:119–130.

Dice, L. R. (1941). Methods for estimating populations of mammals. *The Journal of Wildlife Management*, 5:398–407.

Dorazio, R. M. and Howard, L. J. (2005). Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics*, 61:1093–1101.

Edgar, P., Foster, J., and Baker, J. (2010). *Reptile Habitat Management Handbook*. Amphibian and Reptile Conservation, Bournemouth.

Ehrenfeld, S. (1955). On the efficiency of experimental designs. *The Annals of Mathematical Statistics*, 26:247–255.

Eugene, A. F. and Shwartz, A. (2002). *Handbook of Markov Decision Processes: Methods and Applications*. Springer Science, Business Media, New York.

Evans, N. D. and Chappell, M. J. (2000). Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems. *Mathematical Biosciences*, 168:137–159.

Foo, L. K. and Duffull, S. (2011). Optimal design of pharmacokinetic–pharmacodynamic studies. *Pharmacokinetics in Drug Development, Advances and Applications*, 3:175–194.

Geisser, S. (1993). *Predictive Inference: An Introduction*. Chapman & Hall, New York.

Germano, J. M. and Bishop, P. J. (2009). Suitability of amphibians and reptiles for translocation. *Conservation Biology*, 23:7–15.

Germano, J. M., Field, K. J., Griffiths, R. A., Clulow, S., Foster, J., Harding, G., and Swaisgood, R. R. (2015). Mitigation-driven translocations: Are we moving wildlife in the right direction? *Frontiers in Ecology and the Environment*, 13:100–105.

Gimenez, O., Viallefont, A., Catchpole, E. A., Choquet, R., and Morgan, B. J. T. (2004). Methods for investigating parameter redundancy. *Animal Biodiversity and Conservation*, 27:561–572.

Gould, W. R. and Pollock, K. H. (1997). Catch-effort estimation of population parameters under the robust design. *Biometrics*, 53(1):207–216.

Griffiths, B., Scott, J. M., Carpenter, J. W., and Reed, C. (1989). Translocation as a species conservation tool: status and strategy. *Science*, 245:477–480.

Griffiths, R., Foster, J., Wilkinson, J., , and Sewell, D. (2015). Science, statistics and surveys: a herpetological perspective. *Journal of Applied Ecology*, 52:1413–1417.

Guillera-Arroita, G., Hauser, C. E., and Michael A. McCarthy, M. A. (2014). Optimal surveillance strategy for invasive species management when surveys stop after detection. *Methods in Ecology and Evolution*, 4:1751–1760.

Guillera-Arroita, G., Ridout, M. S., and Morgan, B. J. T. (2010). Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, 1:131–139.

Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer-Verlag, New York.

Hastie, T., Tibshirani, R., and Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second ed.* Springer, New York.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Springer Series, New York, USA.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.

Hilborn, R. and Walters, C. J. (1992). *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Chapman and Hall, New York.

Hoerl, A. and Kennard, R. (1970). Ridge regression: biased estimation of nonorthogonal problems. *Technometrics*, 12:55–67.

Hubbard, B. A., Cole, D. J., and Morgan, B. J. T. (2014). Parameter redundancy in capture-recapture-recovery models. *Statistical Methodology*, 17:17–29.

Hurvich, C. M. and Tsai, C. L. (1990). The impact of model selection on inference in linear regression. *American Statistician*, 44:214–217.

Hutchinson, R. A., Valente, J. J., Emerson, S. C., Betts, M. G., and Dietterich, T. G. (2015). Penalized likelihood methods improve parameter estimates in occupancy models. *Methods in Ecology and Evolution*, 6:949–959.

IUCN/SSC (2013). *International Union for Conservation of Nature/Species Survival Commission. Version 1.0. Guidelines for reintroductions and other conservation translocations*. Gland, Switzerland: IUCN/SSC.

Jacroux, M. (2012). The a-optimality of block designs for comparing test treatments with a control. *Journal of the American Statistical Association*, 84:310–317.

Jolly, G. (1965). Explicit estimates from capture-recapture data with both death and immigration stochastic model. *Biometrika*, 52.

Jones, B. and Donev, A. N. (1996). Modelling and design of cross-over trials. *Statistics in Medicine*, 15:1435–1446.

Kendall, L. W. and Bjorkland, R. (2001). Using open robust design models to estimate temporary emigration from capture—recapture data. *Biometrics*, 57:1113–1122.

Kendall, L. W., Nichols, J. D., and Hines, J. E. (1997). Estimating temporary emigration using capture-recapture data with pollock's robust design. *Ecology*, 78:563–578.

Kendall, L. W., Pollock, K. H., and Brownie, C. (1995). A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*, 51:293–308.

Kendall, W. L. and Nichols, J. D. (2002). Estimating state-transition probabilities for unobservable states using capture-recapture/resighting data. *Ecology*, 83:3276–3284.

Kerr, M. K. and Churchill, G. A. (2001). Experimental design for gene expression microarrays. *Biostatistics*, 84:183–201.

Kiefer, J.and Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 30:271–294.

King, R. (2014). Statistical ecology. *Annual Review of Statistics and Its Application*, 1:401–426.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 4th International Joint Conference on Artificial Intelligence*, pages 1137–1143.

Laplanche, C. (2010). A hierarchical model to estimate fisher abundance in alpine streams by sing removal sampling data from multiple locations. *Biometrical Journal*, 52:209–221.

Lebreton, J. D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, 82:67–118.

Legg, C. J. and Nagy, L. (2006). Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management*, 78:194–199.

Lincoln, F. C. (1930). The waterflow flyways of north america. *US Department of Agriculture Circular*, 118:1–4.

Linklater, W. L., Adcock, K., du Preez, P., Swaisgood, R. R., Law, P. R., Knight, M. H., Gedir, J. V., and Kerley, G. I. H. (2011). Guidelines for large herbivore translocation simplified: black rhinoceros case study. *Journal of Applied Ecology*, 48:493–502.

Linnell, J. D. C., Aanes, R., Swenson, J. E., Odden, J., and Smith, M. E. (1997). Translocation of carnivores as a method for managing problem animals: a review. *Conservation Biology*, 6:1245–1257.

Lystig, T. C. and Hughes, J. P. (2002). Exact computation of the observed information matrix for hidden markov models. *Journal of Computational and Graphical Statistics*, 11(3):678–689.

MacKenzie, D. (2018). *PRESENCE 12.12 – Software to estimate patch occupancy and related parameters*. Patuxent Wildlife Research Center, U.S. Geological Survey.

Matechou, E., McCrea, R. S., Morgan, B. J. T., Nash, D. J., and Griffiths, R. A. (2016). Open models for removal data. *The Annals of Applied Statistics*, 10:1572–1589.

McCrea, R. and Morgan, B. (2014). *Analysis of Capture-Recapture Data*. Chapman and Hall.

McCrea, R. S., Morgan, B. J. T., Gimenez, O., Besbeas, P., Lebreton, J.-D., and Bregnballe, T. (2010). Multi-site integrated population modelling. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:539–561.

Mills, M. D., Rader, R. B., and Belk, M. C. (2004). Complex interactions between native and invasive fish: the simultaneous effects of multiple negative interactions. *Oecologia*, 141:713–721.

Moran, P. A. P. (1951). A mathematical theory of animal trapping. *Biometrika*, 38:307–311.

Moreno, M. and Lele, S. R. (2010). Improved estimation of site occupancy using penalized likelihood. *Ecology*, 91:341–346.

Morgan, B. (2008). *Applied Stochastic Modelling*. Chapman and Hall.

Morris, R. (2010). An unfortunate experiment. *Forest and Bird*, 337:14–18.

Morrison, M. L., Block, W. M., Strickland, M. D., Collier, B. A., and Peterson, M. J. (2008). *Wildlife study design*. Springer, New York, USA.

Murphy, S. and van der Vaart, A. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95:449–485.

Nees, J. G., Helm, W. T., and Threinen, C. W. (1957). Some vital statistics in a heavily exploited population. *Journal of Wildlife Management*, 21:279–292.

North, P. M. and Morgan, B. (1979). Modelling heron survival using weather data. *Biometrics*, 35:667–682.

Otis, D. L., Burnham, K. R., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monograph*, 62:163–169. Washington, D.C. The Wildlife Society.

Petersen, C. G. J. (1894). On the biology of our flat-fishes and on the decrease of our flat-fish fisheries. *Report of the Danish Biological Station IV*, pages 1893–1894.

Peterson, J. T., Thurow, R. F., and Guzevich, J. W. (2004). An evaluation of multipass electrofishing for estimating the abundance of stream-dwelling salmonids. *Transactions of the American Fisheries Society*, 133(2):462–475.

Pollock, K. (1991). Modelling 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*, 86:225–238.

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

Pradel, R. (2005). Multievent : An extension of multistate capture-recapture models to uncertain states. *Biometrics*, 61(2):442–447.

Pronzato, L. and Walter, E. (1985). Robust experimental design via stochastic approximation. *Mathematical Biosciences*, 75:103–120.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

Pukelsheim, F. and Studden, W. J. (1993). E-optimal designs for polynomial regression. *The Annals of Statistics*, 21:402–415.

Rafferty, J. P. (2011). *Reptiles and Amphibians*. Britannica Educational Publishing, New York.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25:1923–1929.

Rosenberger, A. E. and Dunham, J. B. (2005). Validation of abundance estimates from mark-recapture and removal techniques for rainbow trout captured by electrofishing in small streams. *North American Journal of Fisheries Management*, 25:1395–1410.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, 39(3):577–591.

Royle, A. J., Chandler, R. B., Sollmann, R., and Gardner, B. (2013). *Spatial Capture-Recapture*. Academic Press, Waltham, MA.

Sandland, R. L. and Kirkwood, G. P. (1981). Estimation of survival in marked populations with possibly dependent sighting probabilities. *Biometrika*, 68:531–541.

Sarno, R. J., Nixon, P. A., Mealey, B. K., Concoby, R. E., Mrykalo, R. J., and Grigione, M. M. (2012). Suitability of translocation sites for florida burrowing owls: Prey availability and diet. *Southeastern Naturalist*, 11(4):755–764.

Schwarz, C. J. and Arnason, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. *Mathematical Biosciences*, 52:860–873.

Seber, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika*, 52:249–259.

Seber, G. A. F. and Le Cren, E. D. (1967). Estimating population parameters from catches large relative to the population. *Journal of Animal Ecology*, 36:631–643.

Seber, G. A. F. and Whale, J. F. (1970). The removal method for two and three samples. *Biometrics*, 26:393–400.

Seddon, P. J.and Armstrong, D. P. and Maloney, R. F. (2007). Developing the science of reintroduction biology. *Conservation Biology*, 21:303–312.

Stanley, T. R. and Burnham, K. P. (1999). A closure test for time-specific capture-recapture data. *Environmental and Ecological Statistics*, 6:197–209.

Stanley, T. R. and Richards, J. D. (2005). Software review: A program for testing capture–recapture data for closure. *Wildlife Society Bulletin*, 33:782–785.

Sullivan, T. P. and Sullivan, D. S. (2013). Influence of removal sampling of small mammals on abundance and diversity attributes: scientific implications. *Human–Wildlife Interactions*, 7:85–98.

Tennant, E. N. and Germano, D. J. (2017). Survival of translocated heermann's kangaroo rats (dipodomys heermanni) in the san joaquin desert of california using hard and soft release methods. *Western Wildlife*, 4:1–11.

Thompson, S. K. (2002). On sampling and experiments. *Environmetrics*, 13:429–436.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58:267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B*, 67:91–108.

Turner, R. (2008). Direct maximization of the likelihood of a hidden markov model. *Computational statistics and data analysis*, 52:4147–4160.

Venzon, D. and Moolgavkar, S. (1988). A method for computing profile-likelihood-based confidence intervals. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 37:87–94.

Viallefont, A. (2011). Capture-recapture smooth estimation of age-specific survival probabilities in animal populations. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(1):131–141.

Wang, Y. G. (1999). Estimating equations for removal data analysis. *Biometrics*, 55(4):1263–1268.

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, Los Alamos, New Mexico, USA.

White, G. C. and Burnham, K. P. (1999). Program mark: survival estimation for populations of marked animals. *Bird Study*, 46:120–138.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and management of animal populations*. Academic Press, New York, USA.

Wyatt, R. J. (2002). Estimating riverine fish population size from single- and multiple-pass removal sampling using a hierarchical model. *Canadian Journal of Fisheries and Aquatic Sciences*, 59:695–706.

Yang, M., Biedermann, S., and Tang, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *Journal of the American Statistical Association*, 108:1411–1420.

Yu, Y. (2010). Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics*, 38:1593–1606.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B*, 68:49–67.

Zhang, P. (1993). Model selection via multifold cross validation. *The Annals of Statistics*, 21(1):299–313.

Zhou, M., McCrea, R. S., Matechou, E., Cole, D. J., and Griffiths, R. (2018). Removal models accounting for temporary emigration. *Biometrics*. Accepted.

Zippin, C. (1956). An evaluation of the removal method of estimating animal populations. *Biometrics*, 12:163–169.

Zippin, C. (1958). The removal method of population estimation. *The Journal of Wildlife Management*, 22(1):82–90.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 67:301–320.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An introduction using R, second edition*. CRC Press, Boca Raton.