Improving Parameter Estimation for Integral Projection Models in Fluctuating Environments

by

Gopal Nath

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 8, 2020

Keywords: Matrix Projection Model, Integral Projection Model, Robust Integral Projection
Model, Permuted Integral Projection Model

Copyright 2020 by Gopal Nath

Approved by

Asheber Abebe, Chair, Professor of Mathematics and Statistics
Nedret Billor, Professor of Mathematics and Statistics
Mark Carpenter, Professor of Mathematics and Statistics
Philippe Gaillard, Associate Professor of Mathematics and Statistics

Abstract

The matrix projection model (MPM) is one of the mainstream approaches for population and environmental dynamics in biological and ecological sciences. MPM is very popular in modern biological sciences mainly due to its simplicity and the ease in which results can be explained. The main basis of MPM is a discrete staged transition matrix. MPM then projects the current population into the next census. However, MPM forces continuous trait, such as drought index, length, or mass into discrete staged classes. Easterling proposed the integral projection model (IPM) to avoid this artificial breakpoints of a continuous trait. The application of IPMs is rapidly growing in forest and wildlife ecology and it started attracting statisticians only recently. While IPM has advantages over the MPM, existing IPM estimation techniques are sensitive to outliers or mixing of population traits. We propose a robust fitting approach for IPMs and we analyze how the gain in robustness in the continuous size variable affects the estimation of population growth rate using a simulation study. We demonstrate the benefits of the proposed approach by analyzing the population dynamics of African elephants (*Loxodonta africa*) in Amboseli National Park, Kenya, where drought is thought to influence the population dynamics. Furthermore, existing IPM fails to generate a fecundity kernel in the situation of incomplete reproduction information. We propose a permutation based model to overcome this situation and demonstrate the applicability using a real life dataset.

Acknowledgments

This work would not have been possible without the guidance and help of several individuals. It is a pleasure to convey my gratitude to them all in my humble acknowledgment. At first, I would like to express my gratitude to Dr. Ash Abebe, Dr. Nedret Billor, Dr. Mark Carpenter and Dr. Philippe Gaillard , especially, Dr. Abebe, my advisor, for his supervision, advice and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work. He provided me unflinching encouragement and support in various ways. I am indebted to him more than he knows. The faculty of this department has provided me with tremendous graduate education. They have taught me how to think about a problem and to approach it as a statistician. Without their knowledge and assistance, this study would not have been successful. I am grateful to all of them. Finally, I would like to express my love and gratitude to my family and thanks to my friends who were important to the successful realization of the thesis.

Table of Contents

List of Figures

Chapter 1

Introduction

Population dynamics of plants and animals is determined by changes in various vital rates including fecundity, reproduction, survival, and growth. The integral Projection Model (IPM) is one of the modern population dynamics model that was developed relatively recently by Easterling and Elner in 2000. Interest in the application of IPMs is growing in the fields of forest and wildlife ecology. It is only recently that it is attracting the attention of statisticians. The Matrix Projection Model (MPM) is the model that is commonly used to quantify population dynamics. MPM is very popular in biological sciences because of its simplicity and because the results can be easily explained. However, MPM may underestimate population parameters since it forces continuous size variable (mass, length, drought index) into discrete stages. In this dissertation, we demonstrate a comprehensive set of tools for a smooth transition from MPM to IPM and then improve parameter estimation for IPM. This dissertation consists of two parts - robust integral projection models and permuted integral projection models. Rees et al. (2014) demonstrate population dynamics of Soay-sheep using standard IPM approach where the size variable, z (mass), follows a single Gaussian distribution. Heather et al. (2018) shows the effect of temperature on the growth of gilthead seabream using the IPM approach where we observe nearly perfect linear relationship between the Otolith radius at time $t - 1$ and the Otolith radius at time t. However, there is not enough explanation how to apply IPM in the situation of two combined similar species where one has slightly larger Otolith radius or one small group of gilthead seabream has larger variability in Otolith radius. In the recent literature, Merow, Dahlgren and Metcalf (2014) improve the IPMs in the presence of outliers or influential observations by using simple linear regression models and adjusting variances.

We are interested in constructing population dynamic models that take into account possible mixture of populations with respect to some of the vital rates. In this case, we are specifically concerned with the case where the size of the plants or animals, which is a determinant of the population growth rate, is not coming from a single Gaussian distribution. For example, we can think of situations where an animal population under study is a mixture of two where a small percentage of animals have a larger variability in their weight distribution. This calls for robust methods of estimation to be incorporated in the modeling of population dynamics. This is precisely the purpose of the first part of this study.

We will begin by reviewing robust regression, M estimation in particular. We will then compare integral projection models for population dynamics based on the usual approach of using the least squares estimation method to that of M estimation under various size distribution scenarios.

In the second part of this dissertation, we focus on how to minimize the cost of collecting the population data, especially for long-lived animals that may not be easily tracked continuously. Sometimes, it may take more than 60 years to achieve a complete life cycle of an elephant herd (Moss, 2001). To track this life cycle for long period of time would be extremely expensive. Again, collecting individual reproduction outcomes is time consuming, and in some scenarios it is almost impossible (fish, turtle).

Without the exact knowledge of reproduction (who gave birth to whom), it is not possible to calculate the population growth rate using standard IPM approaches. The motivation of this research is part of a big elephant research project which started initially by Moss and Croze in 1972 (Moss 2001). Dr. Cynthia Moss is an American wildlife researcher was born in Ossining, New York, in 1940 (Yount, 1999). She moved to Africa in 1968 and has been collecting the dataset ever since. The main focus the second part of this dissertation is to propose a permutation based IPM estimation where there is missing information (eg. exact mother is not known, exact size is missing, etc). In this dissertation, we explore a real data set of Amboseli elephants from greater Amboseli National Park, Kenya. The dataset consists of 91 unique individual samples (Moss, Croze, & Lee, 2011). This data set has complete information;

thus, it will be used as a benchmark to validate the fidelity of our proposed approaches in Chapter 4 and 5.

Reproduction Probability vs Drought Index
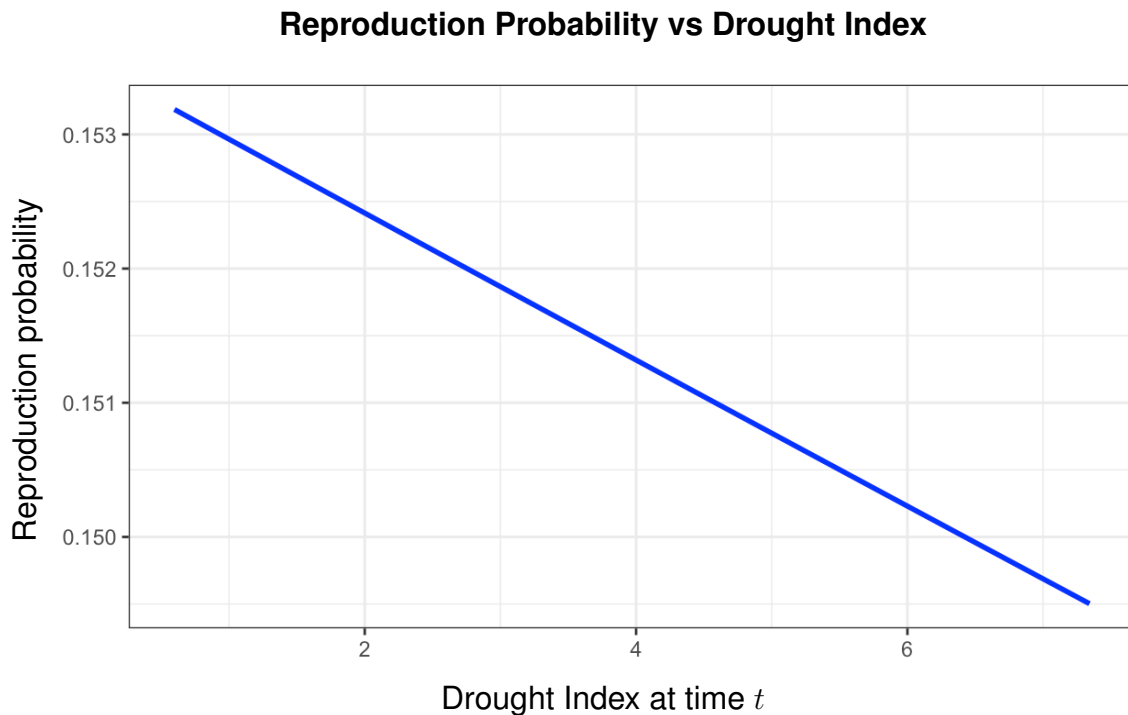
Figure 1.1: Reproduction probability of Amboseli elephants

Figures 1.1 and 1.2 give the survival and reproduction probability, respectively, of Amboseli elephants as a function of drought estimated using an IPM. A question of interest is if there is any significant effects of drought index on asymptotic population growth rate of the Amboseli elephant population. This will be explored in Chapter 4.

Another question of interest is to recover the asymptotic annual population growth rate, λ, assuming no prior complete information of reproduction. The complete reproduction information for one of the families (EB) is presented in Figure 1.3. In Chapter 5, we propose a model that will provide estimates of vital rates even when this information is not completely known.

Survival Probability vs Drought Index
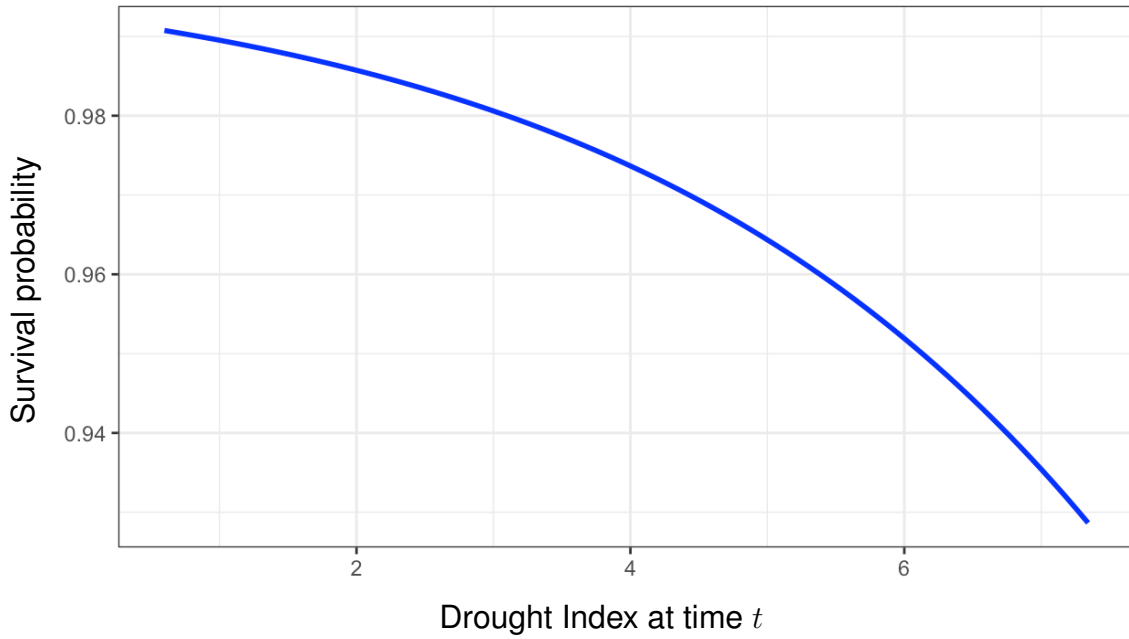
Figure 1.2: Survival probability of Amboseli elephants

Annual Reproduction Information of EB Family
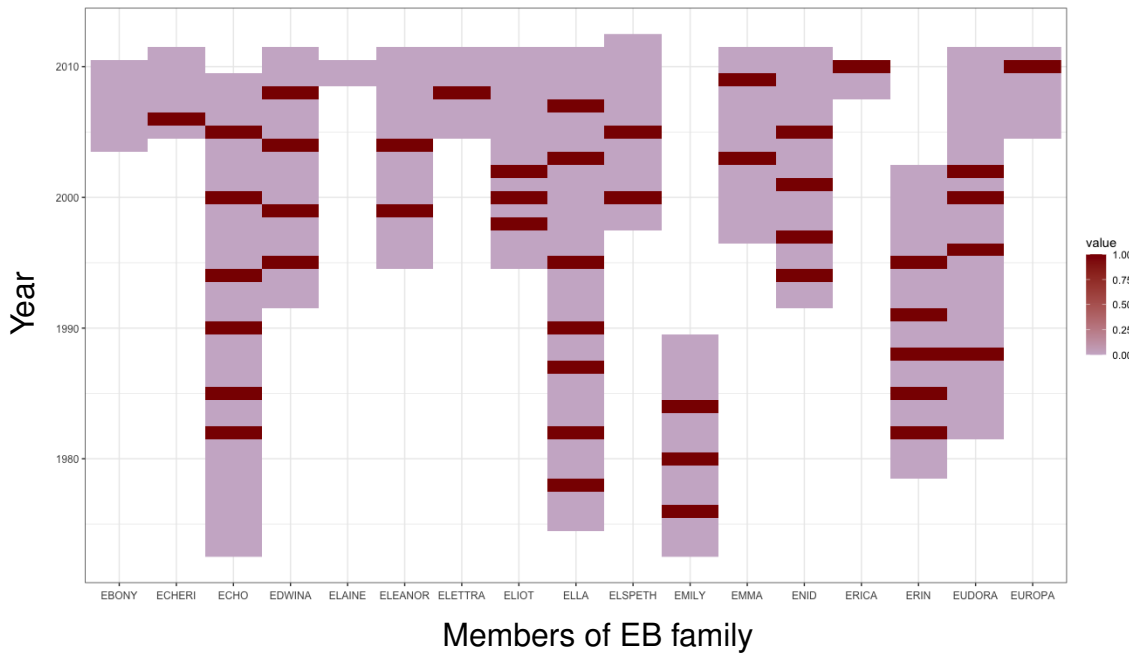
Figure 1.3: True reproductive information of Amboseli elephants (who is giving birth): Individuals who reproduce offspring are represented by the darkened (repr=1) area, Individuals who fail to reproduce offspring are represented by the thistled area (repr=0), and individuals who no longer survive (surv=0) are represented by the no color zone.

Chapter 2

Background

This background chapter covers the three entities we will be combining in the study of population dynamics of species. These are the matrix projection model (MPM), integral projection model (IPM), and robust regression.

2.1   Matrix Projection Model

A principal goal in ecology and biological sciences is to determine the factors that influence the population growth rate. Individual get born, grow or shrink from one stage to another stage, and reproduce or not reproduce conditional on survival. Matrix projection models contribute a connection between individuals and the population creating a descriptive life cycle. Age specific information is required to build matrix projection models. The best strategy to predict population dynamics is to explore the current condition of the population (Keyfitz, 1972). The population vital rates such as survival and reproduction probability depend on individual age-structured stages can be expressed as a transition matrix which project the current population dynamics to a singular step forward in time.
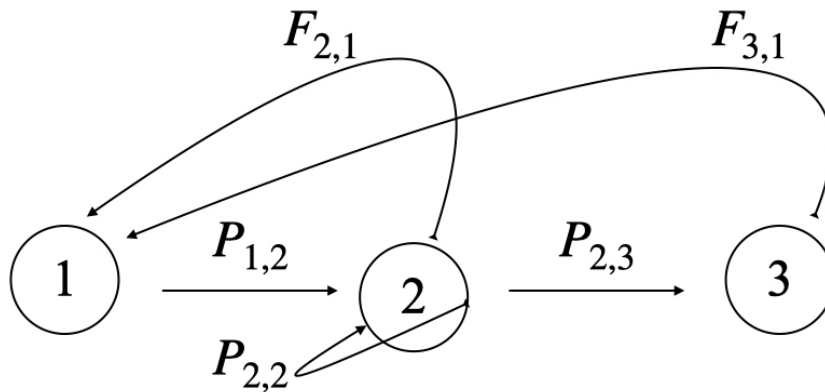
Figure 2.1: A life cycle for a species with a age-classified life history.

The equation of a matrix projection model can be expressed as follows (Caswell, 2001):

$$\mathbf{N}(t+1)_{n\times 1} = \mathbf{A}_{n\times n}\,\mathbf{N}(t)_{n\times 1} \tag{2.1}$$

where,

$\mathbf{A}$ =   transition or projection matrix

$\mathbf{N}(t+1)$ =   population distribution of different stages at time, $t+1$

$\mathbf{N}(t)$ =   population distribution of different stages at time, t

It is recommended to draw the life cycle for a species with stage or size structure before formulating the transition matrix. Consider the individuals who transit from one stage to the next stage with a probability, P. The individuals may reproduce or not reproduce while transitioning from one class to the next or remains same class, with a probability F. Figure 2.1 shows a life cycle of an age classified individuals as stated. The upper curved arrows at the top of the life cycle represent the fecundity probability. The bottom curved arrow represents the probability of individuals surviving and remaining in the same stage. The middle straight arrows represent the transition probability.

Adopting the life cycle in Figure 2.1, we can rewrite the equation (2.1) as follows:

$$
\begin{bmatrix} N_1 \\ N_2 \\ N_3 \end{bmatrix}_{t+1} = \underbrace{\begin{bmatrix} 0 & F_{2,1} & F_{3,1} \\ P_{1,2} & P_{2,2} & 0 \\ 0 & P_{2,3} & 0 \end{bmatrix}}_{\text{transition matrix}} \begin{bmatrix} N_1 \\ N_2 \\ N_3 \end{bmatrix}_t
\tag{2.2}
$$

where $F_{3,1}$ is the fecundity of stage 3 and $P_{2,1}$ is the probability of transition from stage 1 to 2. If we continue multiplication with a current population vector by the transition matrix, the population gradually reaches a stable stage distribution. At this stage the proportion of individual remains constant over time. Thereupon, we can compute population growth rate using the following equation:

$$
\lambda = \frac{\mathbf{N}_{t+1}}{\mathbf{N}_t}
\tag{2.3}
$$

The growth rate calculated at this point is called asymptotic population growth rate, λ. Another way, to compute the asymptotic population growth rate is to calculate the dominant eigenvalue of the transition matrix, $\mathbf{A}$.

Consider an elephant population that matches the description in Figure 2.1 with a transition matrix

$$
\mathbf{A} = \begin{bmatrix} 0.00 & 0.23 & 0.28 \\ 0.90 & 0.60 & 0.00 \\ 0.00 & 0.92 & 0.00 \end{bmatrix}
$$

Assume that elephant population consists of 5 members in group 1, 10 members in group 2, and 4 members in group 3. Using the equation 2.2, we can compute abundances to a singular step forward in time. Continuing this process of multiplication of the current abundances by the transition matrix, after approximately ten iterations the abundances reach a stable stage distribution as reported in Figure 2.3. At this stage, we can calculate the growth rate using equation (2.3).

Population Abundances for all Aged-Structure vs Time
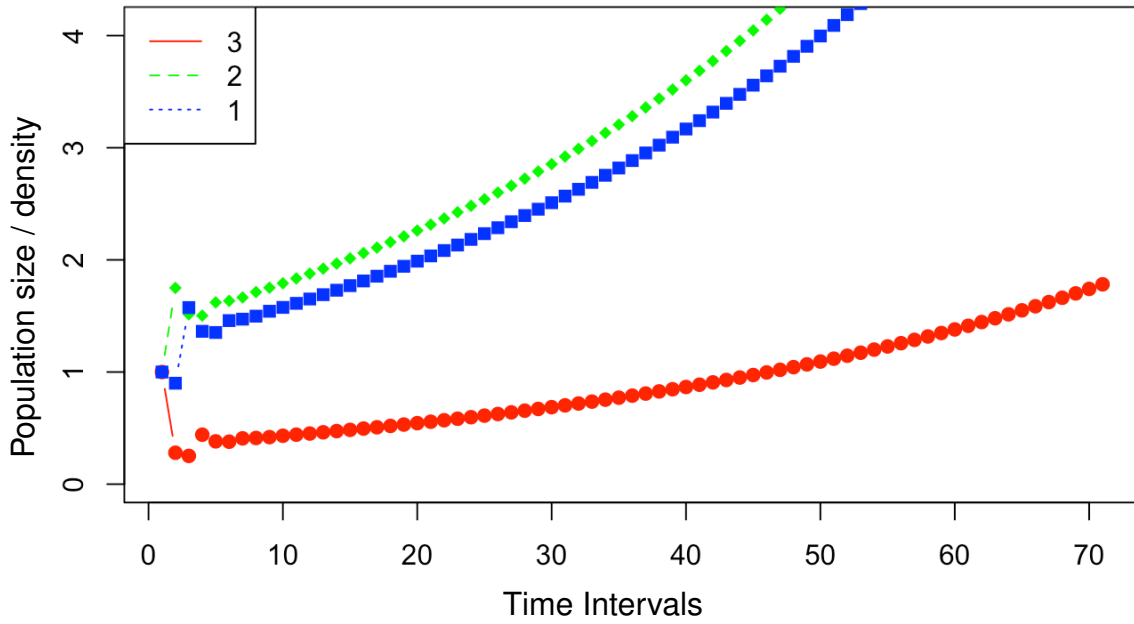
Figure 2.2: Population abundances for all aged-structure group over time

Log Scale of Population Abundances for all Aged-Structure vs Time
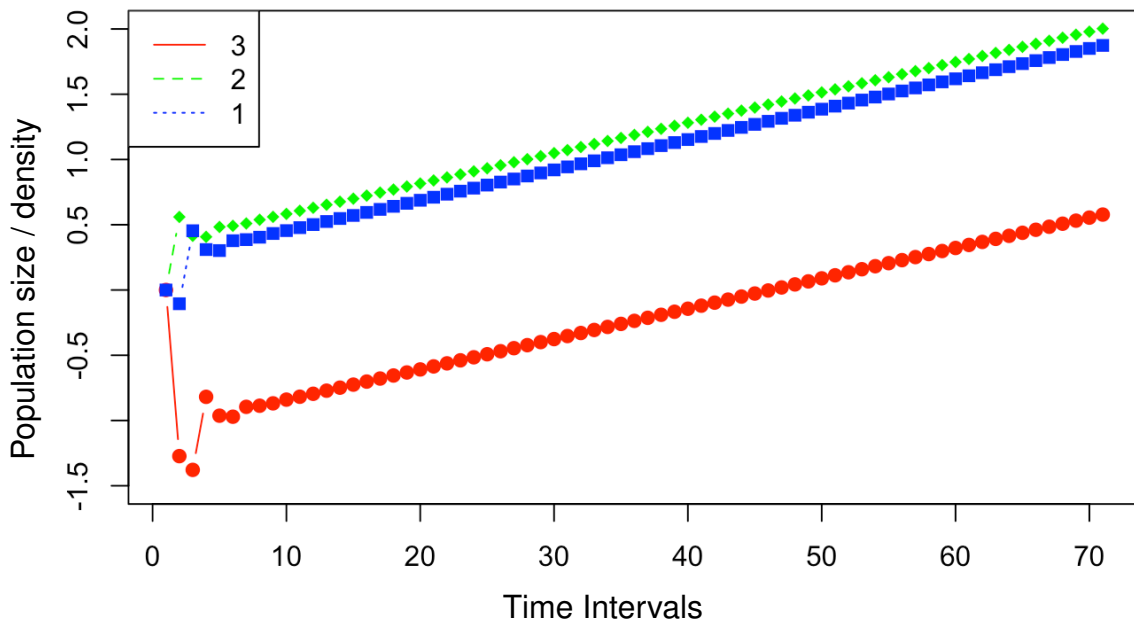
Figure 2.3: Log scale of population abundances for all aged-structure group over time.

Asymptotic population growth rate, $\lambda = \frac{1.576352 + 1.791492 + 0.4317041}{1.541800 + 1.751502 + 0.4195401} = 1.023353$. This asymptotic growth rate can be calculated directly by obtaining the dominant eigenvalue of the transition matrix, $\mathbf{A}$. In this case, we find $\lambda = 1.0235387$. Thus for this contrived example, the elephant population growth rate is about 2.3% each year.

Computational Algorithm

The outline of the algorithm pertaining to the matrix projection model (MPM) is as follows:

Step 1: formulate a transition / projection matrix

Step 2: using the members of each age-classified group define a initial vector abundance.

Step 3: multiply vector abundance by the transition matrix to get the vector abundance to a singular step forward in time.

Step 4: repeat **Step 3** until we get a stable stage distribution.

Step 5: at this stage calculate the asymptotic population growth rate.

While the matrix projection model (MPM) is the most used model to evaluate the population dynamics in biological sciences, it may underestimates the population parameters since it discretizes continuous variables as discussed in the introduction. Easterling proposed the integral projection model (IPM) to avoid this artificial breakpoints of a continuous trait. In the section, we will discuss the basic ideas behind the IPM and its wide applications.

2.2 Integral Projection Model

We are interested in estimating vital rates that contribute to population dynamics of species. Instead of discrete population groups, we will assume that the populations are continuously structured where some continuous measure of individual state may influence growth, survival or reproduction. For continuously structured populations, integral projection models (IPMs) are commonly applied. IPMs combine statistical estimation through regression models with population dynamic models from mathematical biology to study the contribution of different phenomena to the growth of populations.

Integral Projection Models (IPMs) have gained widespread popularity since they were developed by Easterling and Elner in 2000. IPMs are commonly used for the analysis of dynamics in wildlife ecology, conservation biology, and forestry and fisheries ecology. Although IPMs are based on regression models, they have not yet received a lot of attention from statisticians.

The base of an IPM is the kernel. Cory Merow defined the IPM kernel as - "a function that describes how the state of an individual at one time dictates its state and that of its offspring at some future time. Individuals can be characterized in terms of one or more state variables that explain variation in vital rates; often the state variable is some measure of size". The kernel explains how the size (z_t) at time t,, transit to the next state of size z_{t+1} at time $t + 1$. It is recommended to draw the life cycle carefully to represent the complete life census interval. The distribution of the size z_t is denoted by n_{zt} and the integration of n_{zt} over z_t expresses number of individuals conditional on survival probability.

The kernel is represented as a function $K(z_{t+1}, z_t)$ of size that maps the current size, z_t at time t into a size z_{t+1} at time $t + 1$. Kernel, K represents the individual survival, reproduction and transition in size variable/s. Size, z_t (mass, width, length, drought index etc) is a generic variable which is correlated to survival and reproduction probability. Heather and Childs, 2018 used otolith radius and temperature as a size variables to construct an IPM for the growth rate of gilthead seabream in Gulf of Lions (Mediterranean). They suggested that with increasing water temperature reduce the growth rate of gilthead seabream in their study. Rees *et al* (2014) used sheep mass as a size variable and suggested higher mass tends to increase population growth

rate. Merow *et al.* (2014) considered log of the sum of the products of leaf number and highest leaf length as the size variables. The kernel, which is at the the core of the IPM, performs a similar role that the transition matrix plays in the MPM. We present a graphical comparison between IPM and MPM in the diagram at end of this chapter.

IPM predicts population $n_{(t+1)}(z_{t+1})$ at time $t + 1$ by the following: (Easterling *at el.* (2000), Merow *at el.* (2014) and Rees *at el.* (2014))

$$n_{(t+1)}(z_{t+1}) = \int_{all\ sizes} [P(z_{t+1}, z_t) + F(z_{t+1}, z_t)]n_t(z_t)dz \qquad (2.4)$$

$$= \int_{\mathcal{L}}^{\mathcal{U}} K(z_{t+1}, z)n(z, t)dz \qquad (2.5)$$

where,

z_t	=	individual size at time t
z_{t+1}	=	individual size at $t + 1$
$\mathcal{L}$	=	the lower size limits
$\mathcal{U}$	=	the upper size limits
$\int n_t(z_t)dz$	=	represents the number of individuals
$n_t(z_t)$	=	size distribution at time t
$n_{(t+1)}(z_{t+1})$	=	size distribution at time $t + 1$
$P(z_{t+1}, z_t)$	=	growth/survival kernel
$F(z_{t+1}, z_t)$	=	fecundity /reproduction kernel
and $K(z_{t+1}, z_t)$	=	the full kernel $= P(z_{t+1}, z_t) + F(z_{t+1}, z_t)$.

The integration completes the projection from size z_t at time t to the next size z_{t+1} transition at time $t + 1$. This integration performs on main kernel K which consists of the survival/growth kernel P and the fecundity/reproduction kernel F. The kernel P explains population that survive from one state to the next state and while the kernel F explains all possible births of new recruits. Now let us explore the components of IPM discussed in equation 2.5.

The **growth kernel** consists of survival and size transition functions. It represents the size transitions for those that survive. The growth kernel can be defined as follows:

$$P(z_{t+1}, z_t) = g(z_{t+1}, z_t) * s(z_t) \tag{2.6}$$

where,

$$g(z_{t+1}, z_t) \quad = \quad \text{size transition function}$$

$$s(z_t) \quad = \quad \text{survival probability}$$

Here the functions $g(\cdot, \cdot)$ and $s(\cdot)$ are taken as

$$g(z_{t+1}, z_t) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left(\frac{(z_t - \mu_z)^2}{2\sigma_z^2}\right) \tag{2.7}$$

$$s(z_t) = \frac{\exp(\beta_0 + \beta_1 z_t)}{1 + \exp(\beta_0 + \beta_1 z_t)} \tag{2.8}$$

here $\mu_z = az_t + b$ is calculated using a regression model that predicts the size, z_{t+1} at time $t+1$ given the size z_t at time t. This regression line is calculated from the scatter plot of size at time t versus the size at time $t+1$ and the standard deviation, σ_z in the Equation (2.7) is calculated from this regression model. The higher the σ_z the more scatter of the data from the regression line. Survival probability in Equation (2.8) is estimated using a logistic regression model.

On the other hand, the **fecundity kernel** consists of reproduction, offspring size, survival, and recruitment function. The fecundity kernel can be defined as follows:

$$F(z_{t+1}, z_t) = O_f(z_{t+1}, z_t) * s(z_t) * (1/2) * r_p(z_t) \tag{2.9}$$

where,

$$O_f(z_{t+1}, z_t) \quad = \quad \text{offspring size function}$$

$$r_p(z_t) \quad = \quad \text{reproduction probability}$$

$$s(z_t) \quad = \quad \text{survival probability}$$

Similarly the fecundity kernel components can be expressed as follows:

$$O_f(z_{t+1}, z_t) = \frac{1}{\sqrt{2\pi\sigma_{rz}^2}} \exp\left(\frac{(z_t - \mu_{rz})^2}{2\sigma_{rz}^2}\right) \tag{2.10}$$

$$r_p(z_t) = \frac{\exp(\beta_2 + \beta_3 z_t)}{1 + \exp(\beta_2 + \beta_3 z_t)} \tag{2.11}$$

$$s(z_t) = \frac{\exp(\beta_0 + \beta_1 z_t)}{1 + \exp(\beta_0 + \beta_1 z_t)} \tag{2.12}$$

where $\mu_{rz} = a_o z_t + b_o$ is calculated using a regression model that predicts the size of offspring z_{t+1} at time $t + 1$ given the maternal size z_t at time t. This regression line is calculated from the scatter plot of maternal size at time t versus the offspring size at time $t + 1$. The standard deviation σ_{rz} in the Equation (2.10), which measures the vertical scatter in this scatterplot, is also calculated from this regression model. The higher the σ_{rz}, the higher the variability of the data around the regression line.

After estimating the growth and fecundity kernels, we step forward for the final step of IPM. At this stage, we calculate the full kernel using Equation (2.5). A step by step computational algorithm to calculate population growth rate using IPM is presented at the end of this section. The graphic at the end of this section that compares MPM and IPM gives a visual representation of the algorithm.

The full Integral Projection Model components can be summarized as shown below in Table 2.1:

Comparison Diagrams

MPM

Size at t

Size at t+1

$$\begin{bmatrix} 122 \\ 100 \\ 44 \\ 10 \\ 55 \\ 45 \\ 11 \\ 18 \end{bmatrix} = \begin{bmatrix} 89 \\ 131 \\ 44 \\ 10 \\ 55 \\ 23 \\ 3 \\ 5 \end{bmatrix}$$

Transition Matrix

Repeat this process

Population size/density

Time Intervals

biasS1
biasS3
biasS5
biasS4
biasS6
biasS7
biasS8

Population Distribution will be stable at this point

Calculate population growth rate at this stage which is called asymptotic population growth rate.

IPM

Size at t+1

Size at t

Survival probability

Size at t

σ_z μ_z

a_s b_s

$$\frac{1}{\sqrt{2\pi\sigma_z^2}}\exp\left(\frac{(z-\mu_z)^2}{2\sigma_z^2}\right)$$

$$\mathbf{logit}(a_s z + b_s)$$

Size at t

Size at t

Size at t+1

Size at t+1

Growth Kernel, P

Fecundity Kernel, F

$$\frac{1}{\sqrt{2\pi\sigma_{rz}^2}}\exp\left(\frac{(z-\mu_{rz})^2}{2\sigma_{rz}^2}\right)$$

$$\mathbf{logit}(a_{rp} + b_{rp})$$

σ_{rz} μ_{rz}

a_{rp} b_{rp}

Offspring size at t+1

Size at t

Reproduction probability

Size at t

Size at t

Size at t+1

Full Kernel, K

λ

Table 2.1: Parameterize the IPM using vital rate parameters.

Function	Details
Growth function, $g(z_{t+1}, z_t)$	g follows normal distribution with mean, $\mu_z = a_z + b_z z_t$, sd $= \sigma_z$ a_z, b_z, σ_z are estimated from regression model $[lm(z_{t+1} \sim z_t)]$
Survival probability $s(z_t)$	logit$[s(z_t)]= a_s + b_s z_t$ a_s and b_s are estimated from logistic regression $[glm(survival \sim z_t)]$
Reproduction probability $r_p(z_t)$	$\log[r_p(z_t)] = a_{r_p} + b_{r_p} z_t$ parameters are estimated using logistic regression $[glm(reproduction \sim z_t)]$
Offspring growth function, $O_f(z_{t+1}, z_t)$	O_f follows normal distribution with $\mu_{rz} = a_{rz} + b_{rz} z_t$,sd $= \sigma_{rz}$ $a_{rz}, b_{rz},$ and σ_{rz} are estimated using regression model $[lm(recruitement \sim z_t)]$
Recruitment probability, r_{cr}	logit$[r_{cr}] = a_{rcr}$ parameter is estimated using logistic regression model $[glm(Recr \sim 1)]$

Computational Algorithm

The outline of the algorithm pertaining to the integral projection model (IPM) is as follows:

Step 1: Define a function $\text{Par}(a_z, b_z, \sigma_z, a_s, b_s, a_{r_p}, b_{r_p}, a_{rz}, b_{rz})$ contains all parameters of vital rate.

Step 2: Using generalized linear regression calculate all vital rates parameters and update the function in step 1.

Step 3: Define growth function, $g(z_{t+1}, z_t) = \frac{1}{\sqrt{2\pi\sigma_z^2}}\exp\left(\frac{(z_t - \mu_z)^2}{2\sigma_z^2}\right)$ that returns output as size, z_{t+1} at time $t+1$ given the size, z_t at time t as follows:

- plot individual size, z_t at time t versus the individual size, z_{t+1} at time $t+1$

- using m-estimation calculate regression parameters (a_z and b_z) and standard deviation σ_z

- define $\mu_z = a_z + b_z * z_t$, where z_t is the size at time t

- generate a normal distribution with mean μ_z and standard deviation σ_z

- input the size, z_t at time t using **steps 10 through 12**

- the growth function $g(z_{t+1}, z_t)$ returns size z_{t+1} at time $t+1$ provided the size, z_t at time t.

Step 4: Define survival function, $s(z_t) = \frac{1}{1+\exp(-\mu_{lgs})}$ which returns the survival probability of size z_t at time t as follows:

- define $\mu_{lgs} = a_s + b_s * z_t$, where z_t is the size at time t

- input the size, z_t at time t using **steps 10 through 12**

- return the survival probability of size z_t at time t

Step 5: Define reproduction function, $r_p(z_t) = \frac{1}{1+\exp(-\mu_{lgp})}$ which returns the reproduction probability of size z_t at time t as follows:

- where, $\mu_{lgp} = a_{r_p} + b_{r_p} z_t$, and z_t is the size at time t

16

- input the size, z_t at time t using **steps 10 through 12**

- return the reproduction probability of size z_t at time t

Step 6: Define recruitment function, $O_f(z_{t+1}, z_t$ which returns the size distribution of offspring at time $t + 1$ as follows:

- plot maternal size, z_t at time t versus the offspring size, z_{t+1} at time $t + 1$

- using m-estimation calculate regression parameters a_{rz} and b_{rz} and standard deviation σ_{rz}

- calculate mean $\mu_{rz} = a_{rz} + b_{rz} * z_t$ and the standard deviation σ_{rz} of the size,z_t

- generate a normal distribution with mean μ_{rz} and standard deviation σ_{rz}

- input the size, z_t at time t using **steps 10 through 12**

- return the offspring size transition of z_{t+1} at time $t + 1$

Consider all individuals are recruited to the next census conditional on survival. So, in this case $r_{cr} = 1$

Step 7: Define survival or growth kernel function, $P(z_{t+1}, z_t) = g(z_{t+1}, z_t) * s(z_t)$.

Step 8: Define reproduction kernel function, $F(z_{t+1}, z_t) = O_f(z_{t+1}, z_t) * s(z_t) * (1/2) * r_p(z_t)$.

Step 9: Define full kernel $K(z_{t+1}, z_t) = P(z_{t+1}, z_t) + F(z_{t+1}, z_t)$.

Step 10:Calculate the size range of the data set as follows: $L = l - 0.5$ and $U = u + 0.5$ where l and u are min and max of the size respectively.

Step 11: Define number of elements in the matrix $n = 200$.

Step 12: Calculate mesh points, $z_t = L + k * (U - L)/n - (U - L)/2n$, where, $k = 0, 1, 2, \ldots n$.

Step 13: Define a function which evaluates all pairwise components of the vector z_t (mesh points) and returns all elements of $n \times n$ matrix as an output of the function $K(z_t, z_t)$.

Step 14: Calculate dominant eigenvalue and eigenvector of the $n \times n$ matrix.

2.3 Robust Regression

Robust regression approaches provide an alternative to least squares or maximum likelihood estimation for regression models and they are used when there is suspicion that data are contaminated with outliers or influential observations. It can also be used for the purpose of detecting influential observations. It is well-known that ecological data will have outliers since they are collected in uncontrolled conditions. Thus it is reasonable to explore whether using robust regression to estimate components of the IPM provides an improvement over the traditional maximum likelihood and least squares approaches. There are several techniques of robust regression including M-estimation (Huber,1964, Almewally & Almong, 2018 and Hayes et. al. 2007) and R-estimation (Wilcox, 1998, 2010). However, we chose to use M-estimation in this dissertation for two reasons:

- Two of the regression components in IPM have binary responses (reproduction and survival) and M-estimates have a stable algorithm for fitting generalized linear models, including those with binary and count responses.

- M-estimates can be computed efficiently. There will be no appreciable difference between M and R estimates (and perhaps other robust regression methods) from a practical point of view to justify the additional computational time. We did not see a large difference in estimates when continuous responses (size) were fit using R or M estimation.

M -Estimation

Let us consider the current COVID-19 pandemic, where total confirmed cases are known to depend on population density, the level of social distance, and mobility (indicates individual movement). A multiple regression model for this relationship can be represented as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \tag{2.13}$$

where,

$$
\begin{aligned}
y &= \text{total confirmed cases of COVID-19} \\
x_1 &= \text{population density / sqmi} \\
x_2 &= \text{social distancing} \\
x_3 &= \text{mobility} \\
\varepsilon &= \text{a random error}
\end{aligned}
$$

The term linear indicates that the mean is a linear function of the unknown parameters β_0, β_1, β_2 and β_3.

In general for p regressors, the multiple linear regression can be defined as

$$
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{2.14}
$$

where y is the response variable, $x_j, j = 0, 1, \cdots, p$, are the regressor variables, the parameters $\beta_j, j = 0, 1, \cdots, p$, are the regression coefficients, $E(\varepsilon) = 0$, and $\text{Var}(\varepsilon) = \sigma^2$.

For a given predictor vector $\mathbf{x}' = [1, x_1, x_2, \cdots, x_p]$ and an arbitrary vector $\mathbf{b} = [b_0, b_1, \cdots, b_p]'$, the model residuals are defined as

$$
e = y - \mathbf{x}'\mathbf{b} \ .
$$

Given data $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, M-estimation (Huber,1964) determines the estimate of $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_p]'$ by minimizing a particular ***objective function*** over all $\mathbf{b}$

$$
S = \sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i'\mathbf{b})
$$

For ordinary least squares estimator $\rho(e) = e^2$ and for least absolute deviations estimator $\rho(e) = |e|$.

A reasonable ρ should have the following properties:

1. $\rho(e) \geq 0$

2. $\rho(0) = 0$

3. $\rho(e) = \rho(-e)$

4. $\rho(e_i) \geq \rho(e_j)$ for $|e_i| > |e_j|$

For least squares regression we take the derivative of $S = \sum e_i^2$ with respect to the parameter to obtain the score function which we equate to 0 (estimating equation)

$$\frac{\partial S}{\partial \beta_k} = 0 \rightarrow \sum_{i=1}^{n} e_i x_{ki} = 0$$

and for M-estimator, the estimating equation becomes

$$\frac{\partial S}{\partial \beta_k} = 0 \rightarrow \sum_{i=1}^{n} \frac{\partial \rho}{\partial e_i} x_{ki} = 0$$

If we define a **weight function** $w(e) = \frac{1}{e_i} \frac{\partial \rho}{\partial e_i}$, the estimating equation can be written as

$$\sum_{i=1}^{n} w_i e_i \mathbf{x}_i' = 0$$

Thus M-estimators can be found using the iteratively reweighted least squares approach. We first assume the weights, fit, then calculate the residuals. Use those residuals to calculate new weights and repeat until convergence.

The Huber M-estimator attempts to get the best of both the least-squares estimator (easy to find the minimum) and the absolute deviation estimator (more robust)

$$\rho(e) = \begin{cases} \frac{e^2}{2}, & \text{if } |e| \leq k \\ k|e| - k^2/2, & |e| > k \end{cases} \tag{2.15}$$

Huber picked $k = 1.345\sigma$, which gives 95% efficiency (almost the same as OLS). There are other weight functions that can be used. Some are presented in Table 2.2 and Figure 2.4.
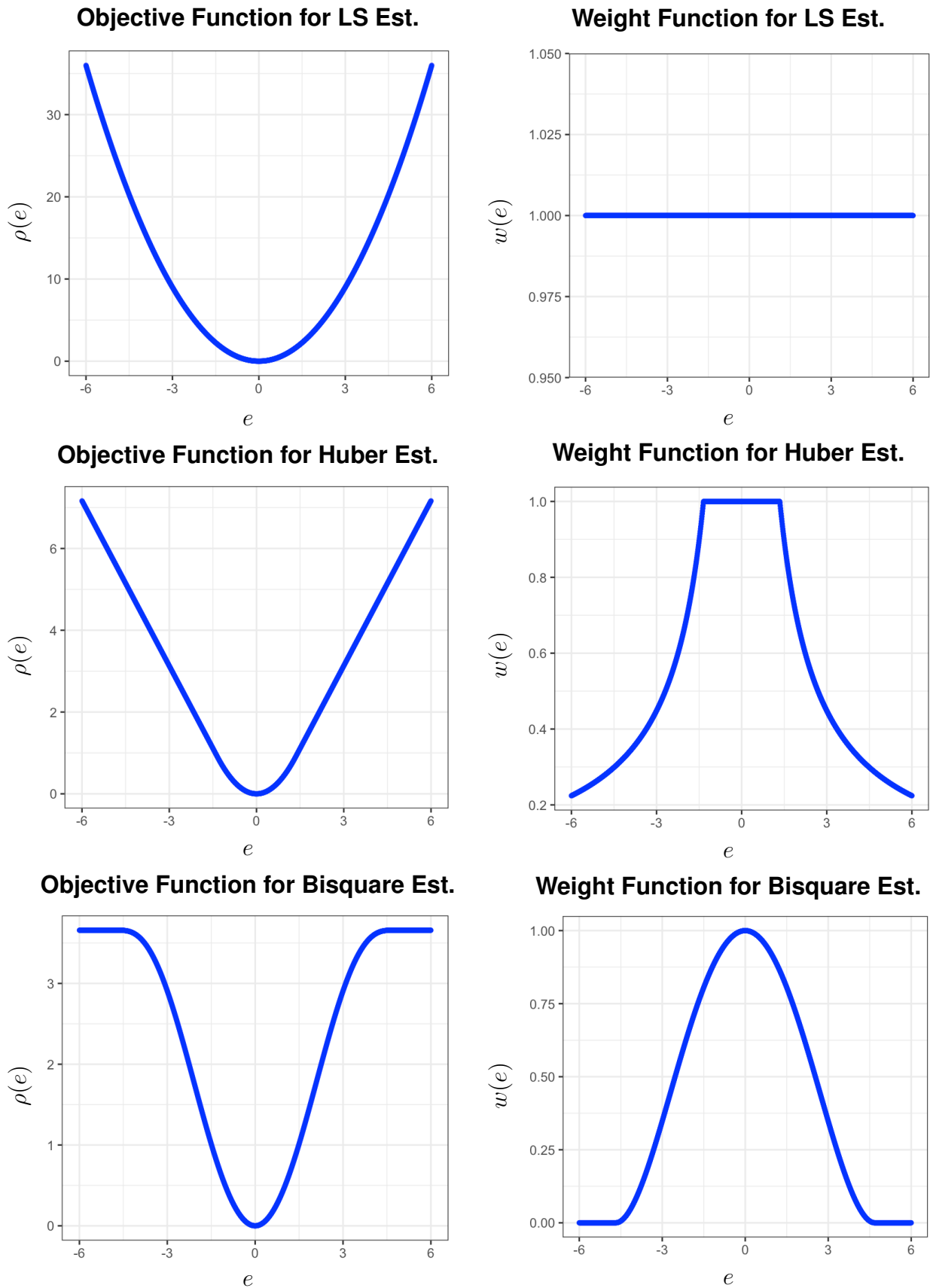
Figure 2.4: **Top panel:** Objective and weight functions for the least-squares, **Middle panel:** Objective and weight functions for the Huber estimators, and **Bottom panel:** Objective and weight functions for the Bi-square estimators. The tuning constants are $k = 1.345$ for the Huber estimator and $k = 4.685$ for the bi-square estimator.

Table 2.2: Least-Squares, Huber and Bisquare Estimation

Method	Objective Function	Weight Function
Least-Square	$\rho_{LS}(e) = e^2$	$w_{LS} = 1$
Huber	$\rho_H(e) = \begin{cases} \frac{e^2}{2}, & |e| \leq k \\ k|e| - k^2/2, & |e| > k \end{cases}$	$w_H(e) = \begin{cases} 1, & |e| \leq k \\ k/|e|, & |e| > k \end{cases}$
Bisquare	$\rho_B(e) = \begin{cases} \frac{k^2}{6}\left\{1 - \left[1 - (\frac{e^2}{k^2})\right]^3\right\}, & |e| \leq k \\ \frac{k^2}{6}, & |e| > k \end{cases}$	$w_B(e) = \begin{cases} \left[1 - (\frac{e^2}{k^2})\right]^2, & |e| \leq k \\ 0, & |e| > k \end{cases}$

We explore current data to visualize the benefits of M-estimation. We used a COVID-19 data set (Worldometer, 2020)-collection period between January 20, 2020 and May 4, 2020. The scatter plot of population density/ sqmi against total COVID-19 confirmed cases is presented in Figure 2.5. We presented Huber weights and residuals (first and last five) of the observations in Table 2.2 and observe the same pattern in Figure 2.4: means the absolute value of residual, e increases , the weight, $w(e)$ decreases.

Table 2.3: Huber weights and residuals

Obs	State	resid	weight
32	New York	295748.539	0.02547065
39	Rhode Island	-65146.791	0.11560598
13	Illinois	45295.824	0.16630999
30	New Jersey	40807.658	0.18465999
5	California	35534.851	0.21199818
45	Vermont	-6306.092	1.00000000
46	Virginia	1627.353	1.00000000
47	Washington	5440.525	1.00000000
48	West Virginia	-6397.277	1.00000000
49	Wisconsin	-1801.677	1.00000000
50	Wyoming	-2194.753	1.00000000

Again, we presented the Huber and Bi-square weights in Figure 2.6 and observe the similar pattern defined in equation (2.15) as the absolute value of residuals increase, the Huber weights and Bisquare weights both decrease. Again, in Figure 2.5, we observe a difference between
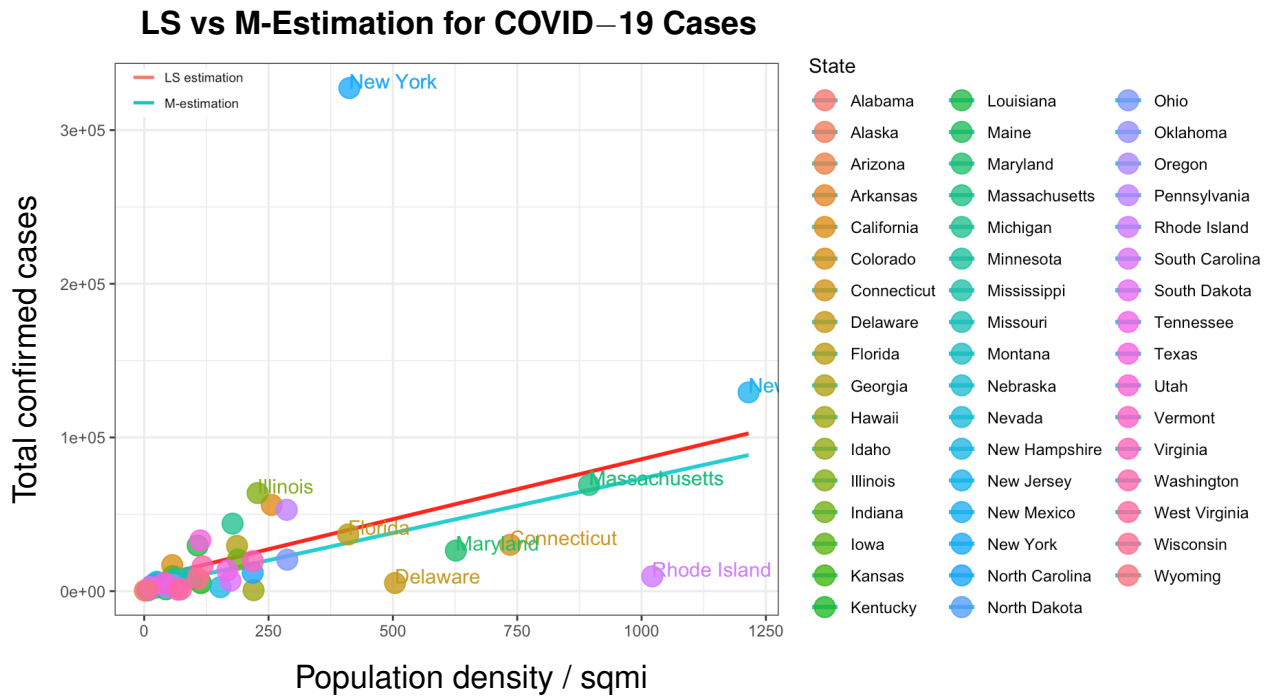
Figure 2.5: COVID-19 Cases

M-regression (darked cyan line) and LS regression (red line) that provide the information of existing outliers which are highly influencing the model parameters.
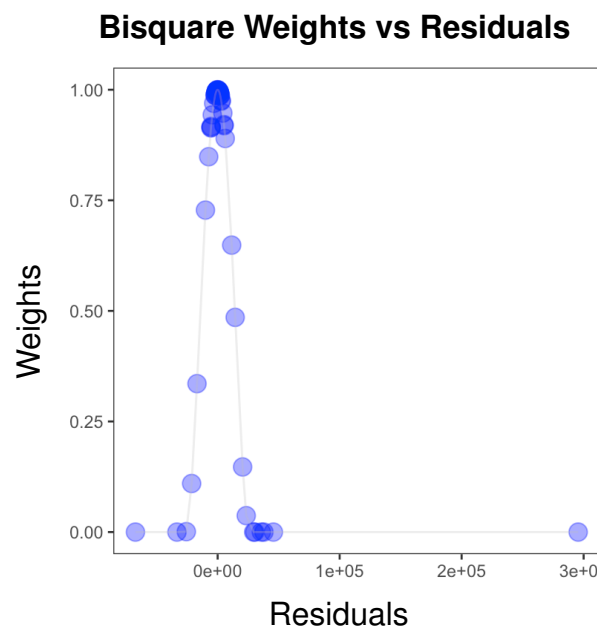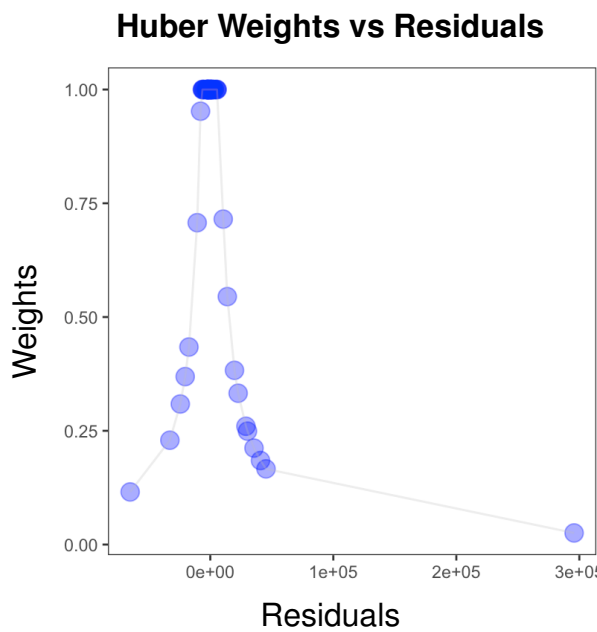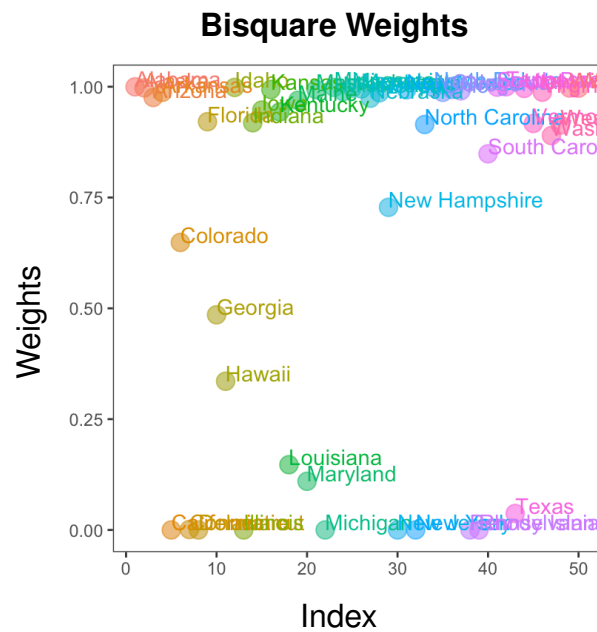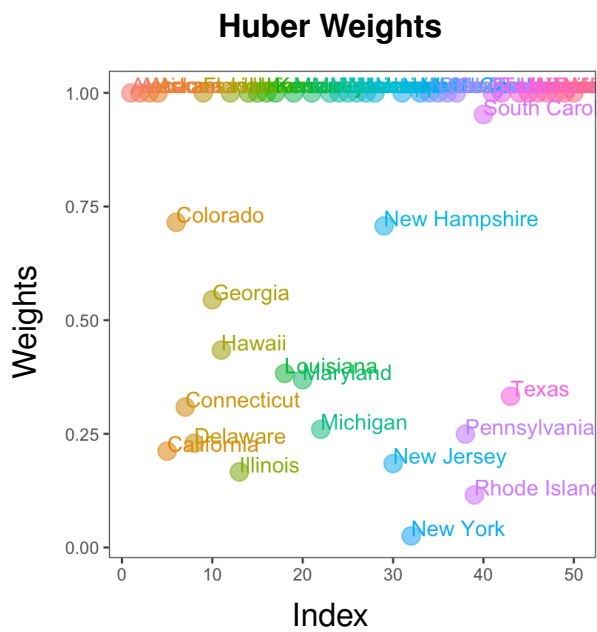
Figure 2.6: Comparison between Huber and Bi-square weight function

Chapter 3

Simulation Study on Robust IPM

In this study we used the data of Soay sheep population of the island of Hirta in Scotland. Rees first simulated this data in 2014 to mimic the true population. To simulate the notion of mixing populations, we will generate some outliers and observe how IPM deals with those extreme sizes. The resulting work-flow can be applied to almost any life history that can be approximated as a sequence of transitions in discrete time. We have decided to put together our model with respect to this population since it has been a significant objective of research into the elements and advancement of wild populations. The data set contains 6 variables as follows:

$\mathbf{z}_t$ =   Mass of a Soay Sheep at time t (continuous)

$\mathbf{z}_{t+1}$ =   Mass of a Soay Sheep at time $t+1$ (continuous)

Surv   =   Survive or not survive at time t (binary)

rep   =   Reproduce or not reproduce (binary)

rcr   =   Recruited or not recruited the offspring for the next census at time, $t+1$ (binary)

rcz   =   Recruited size (continuous)

First, we have explored the data set in Figures 3.1 and, 3.2. Thereupon, population growth rate was estimated using standard IPM before mixing the population. The growth distribution is $g(z_{t+1}, z_t) \sim N(\mu_z, \sigma_z)$, where μ_z and σ_z were estimated using a regression model from the scatter plot of size z_{t+1} at time $t+1$ versus size z_t at time time t as presented in Figure 3.1 (b). Thus, we projected the size, z_{t+1} at $t+1$ by randomly generating a normal distribution with

mean μ_z and standard deviation σ_z given its current size z_t at time t. For example, in R, we can use dnorm(z_t, μ_z, σ_z).

From the size dependent survival plot in Figure 3.1 (a), we observed that the individual with higher in size, z_t tends to have a higher survival probability. We also noted a similar pattern in size transition to the next census in Figure 3.1 (b). Thus, we calculated growth/survival kernel $P(z_{t+1}, z_t)$ as the product of size dependent survival probability $s(z_t)$ and the growth transition $g(z_{t+1}, z_t)$ using the IPM algorithm discussed in Chapter 2. The growth kernel projected the population at time t to the next census at time $t + 1$. We visually observed that the growth kernel in Figure 3.1 (c) is a good representation of size transition plotted in Figure 3.1 (b).

Again, Figure 3.2 (a) shows the size dependent reproduction probability. We observed that individuals larger in size at time t (large z_t) tended to have a higher survival probability. We also observed a similar pattern in offspring size transition to the next census in Figure 3.2 (b).

Furthermore, the fecundity kernel is calculated as $F(z_{t+1}, z_t)$ as half of the product of size dependent reproduction probability $r_p(z_t)$, offspring size transition $O_f(z_{t+1}, z_t)$, and size dependent survival probability $s(z_t)$. We only take half since the focus is only on the population of females. The fecundity kernel projected the population (offspring) at time t to the next census at time $t + 1$. We observed that the fecundity kernel in Figure 3.2 (c) is a good representation of offspring size transition plotted in Figure 3.2 (b).

The full kernel, $K(z_{t+1}, z_t) = P(z_{t+1}, z_t) + F(z_{t+1}, z_t)$ for the Soay sheep data is shown in in Figure 3.3 which is a 3D surface plot of the big matrix (100×100 as we chose our bins $=100$). The higher ridge represents the growth kernel which transits near the diagonal and increases as a function of size. Equivalently, this represents individuals who survive to the next census without much growing or shrinking in size. The second lower ridge is for fecundity kernel, which represents the size variation in offspring. From the contour plot in Figure 3.3 (c) a higher value of the kernel occurs at the top of the diagonal (≈ 0.1) where size of the Soay sheep is higher. Therefore, sheep with size in this area contribute the most to the asymptotic population growth rate λ.
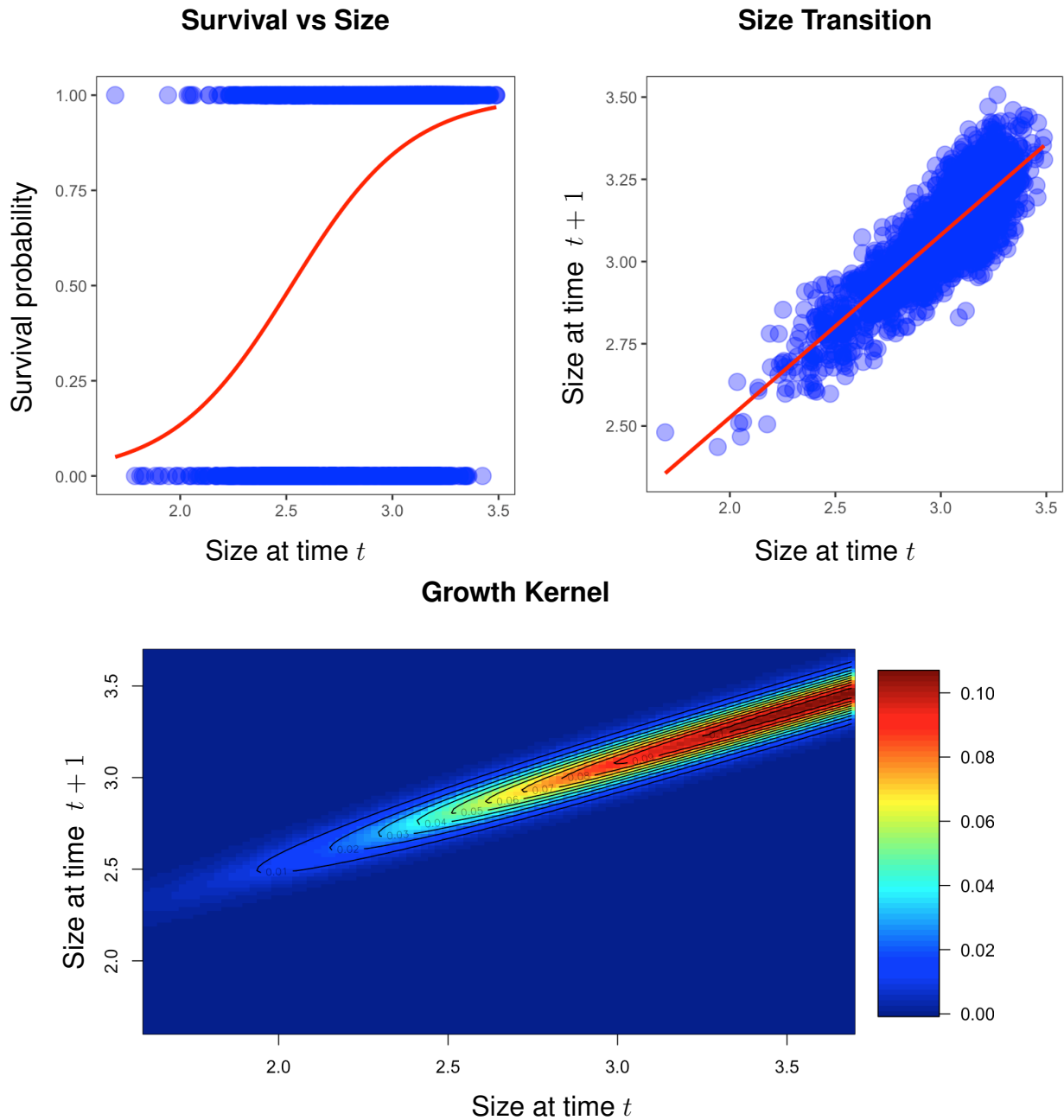
Figure 3.1: (a) Vital rates manifestation. (a) Size, z_t dependent survival probability. Red curve is estimated using logistic regression. (b) The scatter plot of maternal size , z_{t+1} at time $t + 1$ versus the maternal size, z_t at time, t. The red line is estimated using ls-estimation. (c) Growth kernel, $P(z_{t+1}, z_t) =$ size dependent survival probability, $s(z_t)$ * growth transition, $g(z_{t+1}, z_t)$.

Reproduction vs Size

Offspring Size Transition

Fecundity Kernel
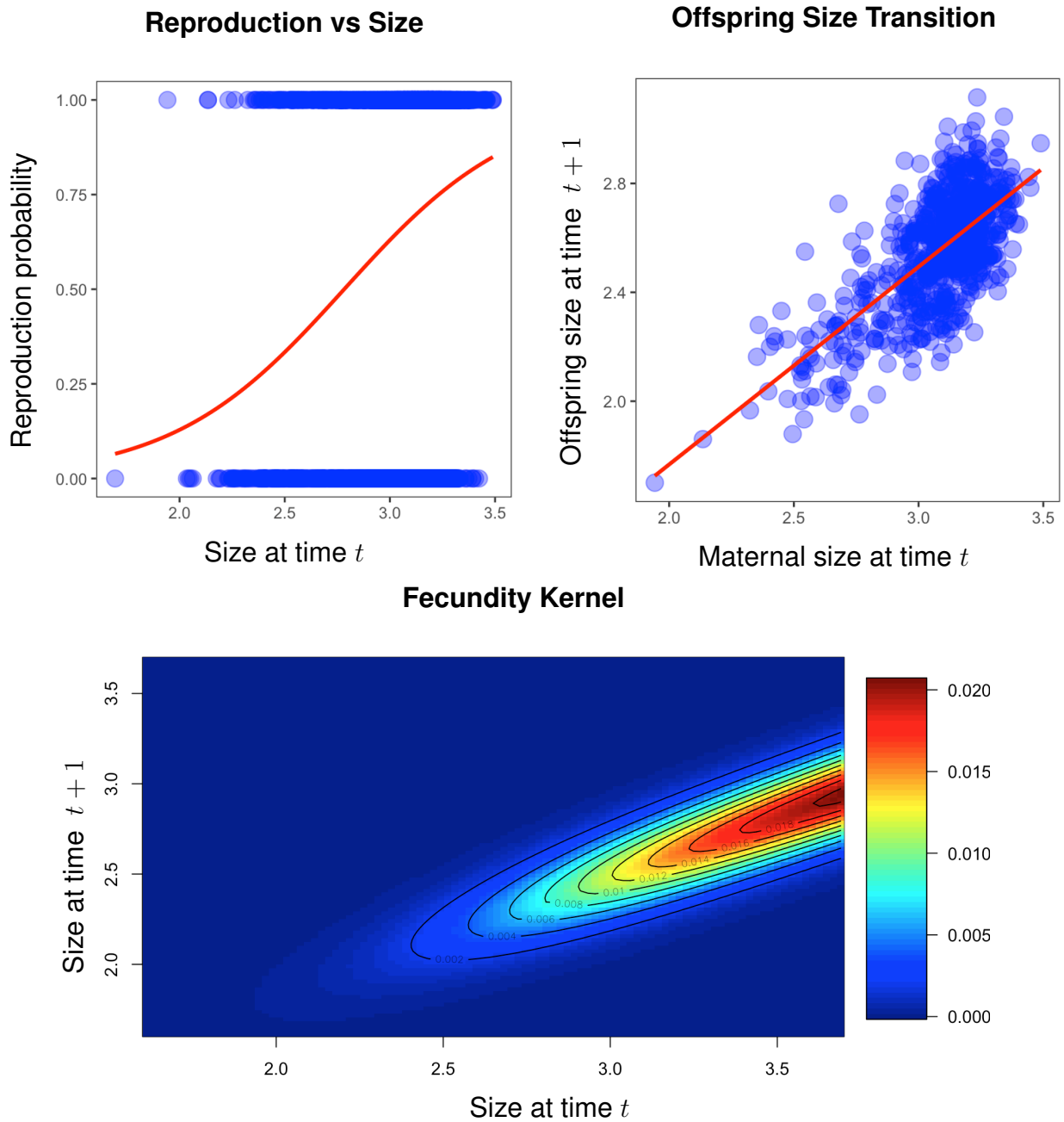
Figure 3.2: (a) Analysis of vital rates. (a) Size, z_t dependent reproduction probability. Red curve is estimated using logistic regression. (b) The scatter plot of offspring size , z_{t+1} at time $t + 1$ versus the maternal size, z_t at time, t. The red line is estimated using ls-estimation. (c) Fecundity kernel, $F(z_{t+1}, z_t) =$ size dependent reproduction probability, $r_p(z_t)$ $*$ offspring size transition, $O_f(z_{t+1}, z_t) * 0.5 *$ size dependent survival probability, $s(z_t)$

225° View of Full Kernel

45° View of Full Kernel
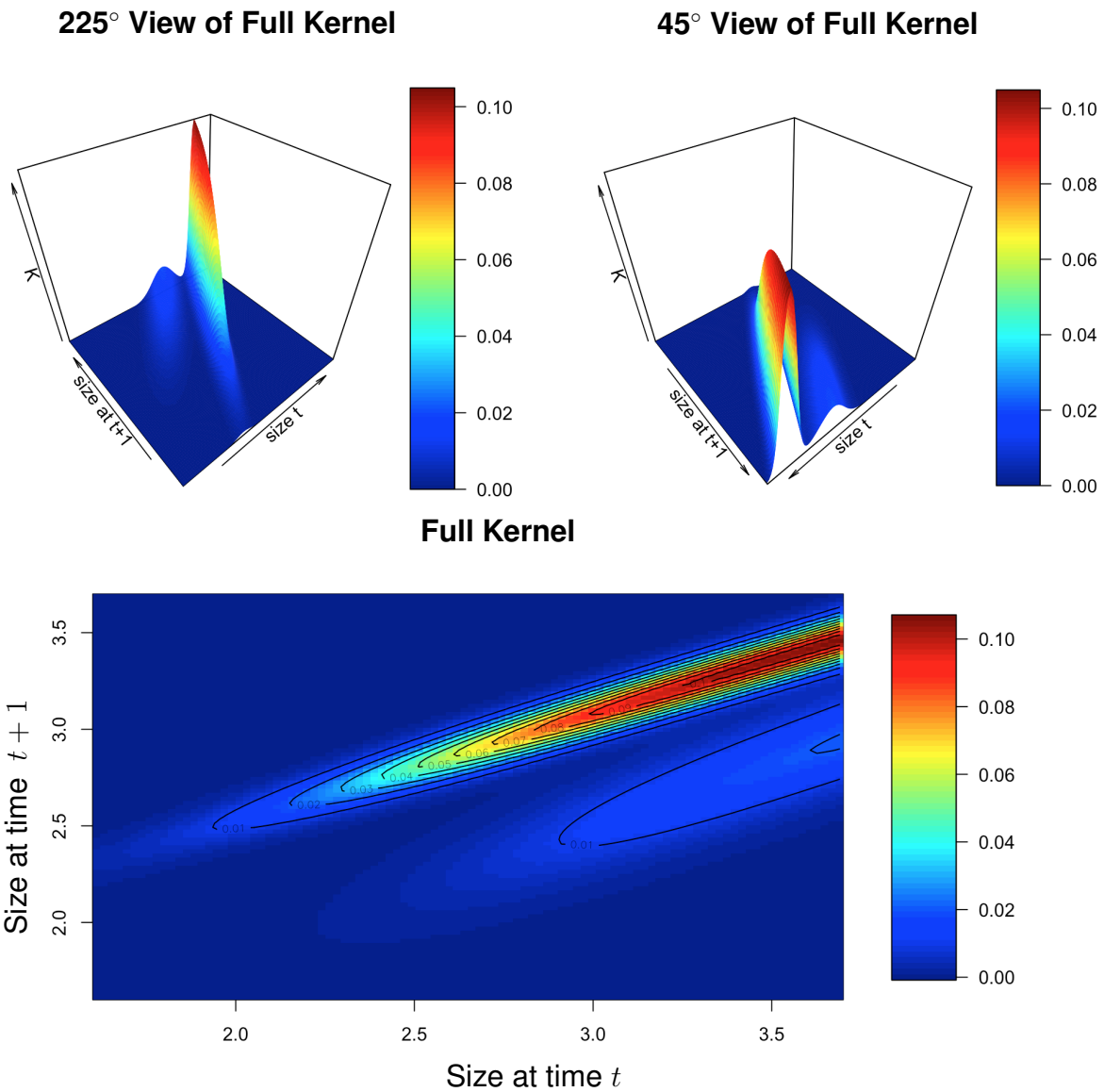
Full Kernel

Figure 3.3: Full kernel, $K(z_{t+1}, z_t) = P(z_{t+1}, z_t) + F(z_{t+1}, z_t)$. The higher ridge dominates the growth kernel which varies near diagonal and increases as size increases. The another interpretation we made from this diagonal ridge as - individuals who survive to the next census, without much growing or shrinking in size. The second lower ridge is for fecundity kernel, which represents the size variation of offspring.

Stable Size Distribution
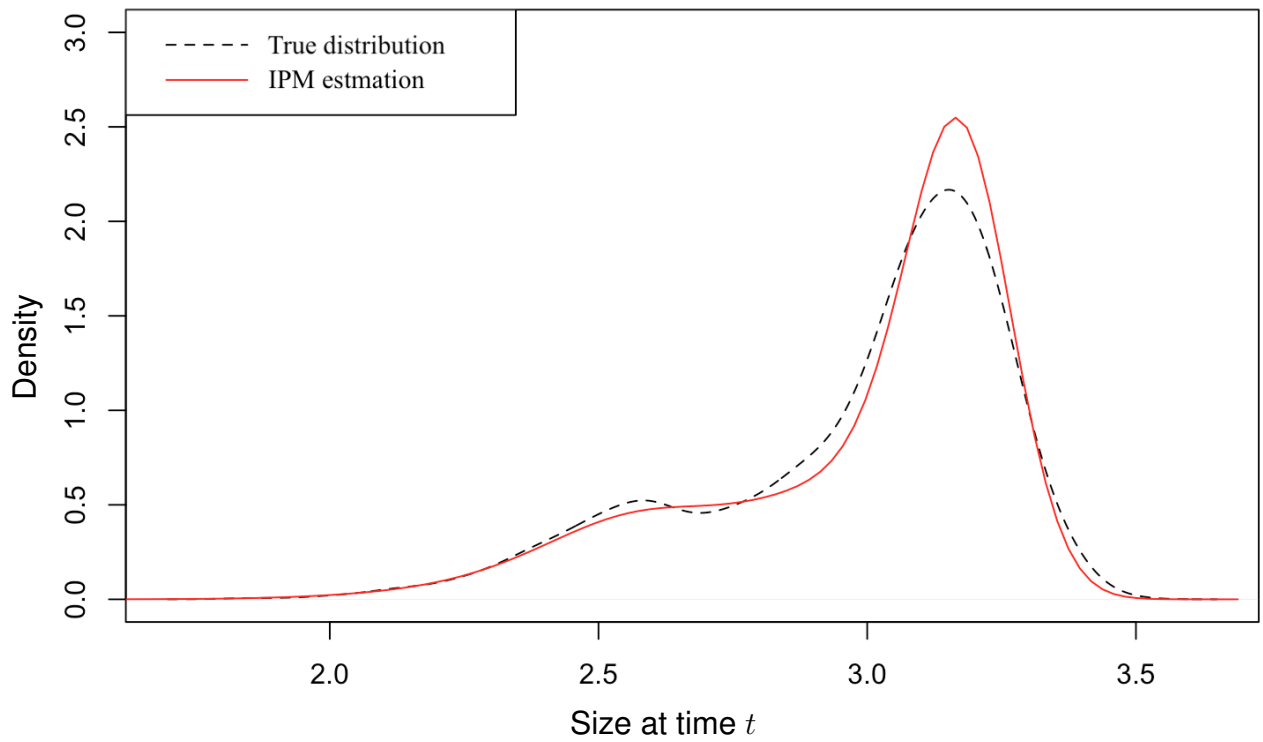
Figure 3.4: (a) The distribution of Soay Sheep masses used for fitting the IPM and the stable size distribution estimated using IPM

The dominant eigenvalue (real part) of the big matrix represents the asymptotic population growth rate, λ. This gives us $\lambda = 1.025$. Equivalently, the Soay sheep population growth rate is 2.5% yearly.

The right eigenvector of the big matrix corresponding to the dominant eigenvalue, λ, is the stable size distribution for the integral projection model as shown in Figure 3.4(red line) (Easterling *et. al*, 2000). The predicted stable distribution is skewed to the left which is similar to the observed size distribution in the dataset used to fit the model as shown in Figure 3.4 (dotted line).

In the situation where individuals have higher variance in size induced by external factors (eg. drought) or mixing caused by migration from a different population, standard IPM approaches may underestimate or overestimate the asymptotic population growth rate since the traditional estimation of IPM is highly sensitive to the presence of outliers. In this situation, robust fitting of regression components of the IPM will likely provide estimators of the population growth rate with better performance in terms of bias and variance. In a simulated scenario, we created a virtual environment where size distribution follows a mixed normal distribution where 5% are sampled from the t distribution with 10 degrees of freedom. This is then scaled to have mean 2.978575 and scale 0.2896221. The remaining 95% are sampled from the normal distribution with mean 2.978575 and standard deviation 0.2896221. The mixed normal distribution can be written as:

$$z \sim \begin{cases} \delta(0 \leq X < 0.95) & N(\mu = 2.978575, \sigma = 0.2896221) \\ \delta(0.95 \leq X \leq 1) & t(\mathrm{df} = 10) * 0.2896221 + 2.978575 \end{cases} \tag{3.1}$$

where X is a uniform random variable on the interval $(0, 1)$ and $\delta(p)$ is 1 if p is true and 0 if p is false.

Survival probability curves are estimated using logistic regression estimated via maximum likelihood as well as M estimation with 200 replications from mixed normal (gray curves). The red curve shows the true survival curve in Figure 3.5. Both logistic regression and robust logistic regression give similar estimated survival curves (Figure 3.5(b)). Regression lines between the size z of female sheep at t and size z' at time $t + 1$ are estimated using the least squares

31
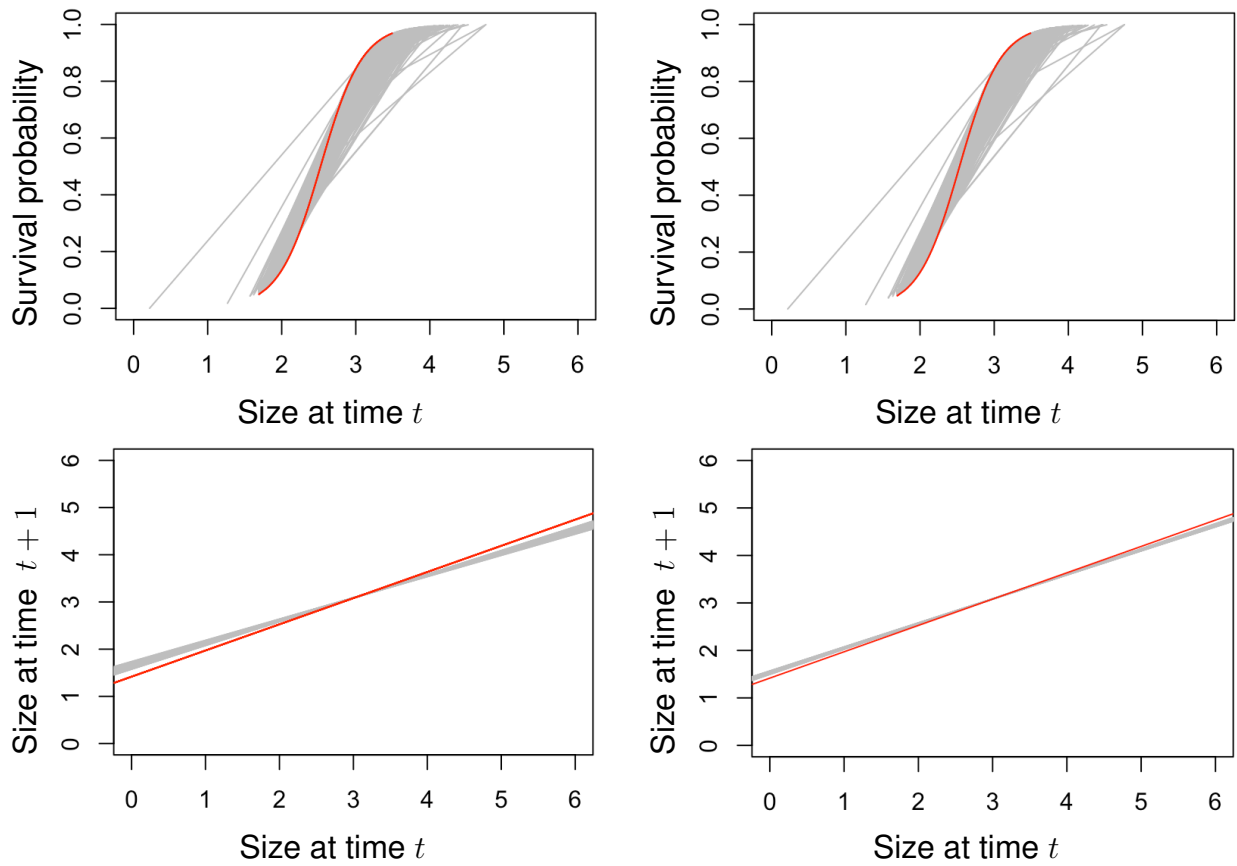
Figure 3.5: (a) Estimated survival curve using logistic regression(gray curve), true survival curve (red line) (b) survival curve using logistic M estimation(gray curve), true survival curve (red line). (c) Regression line using least square regression(gray line), true regression line(red line) and (d) Regression line using M- estimation (gray line), true regression line (red line) for 200 replication of mixed normal distribution
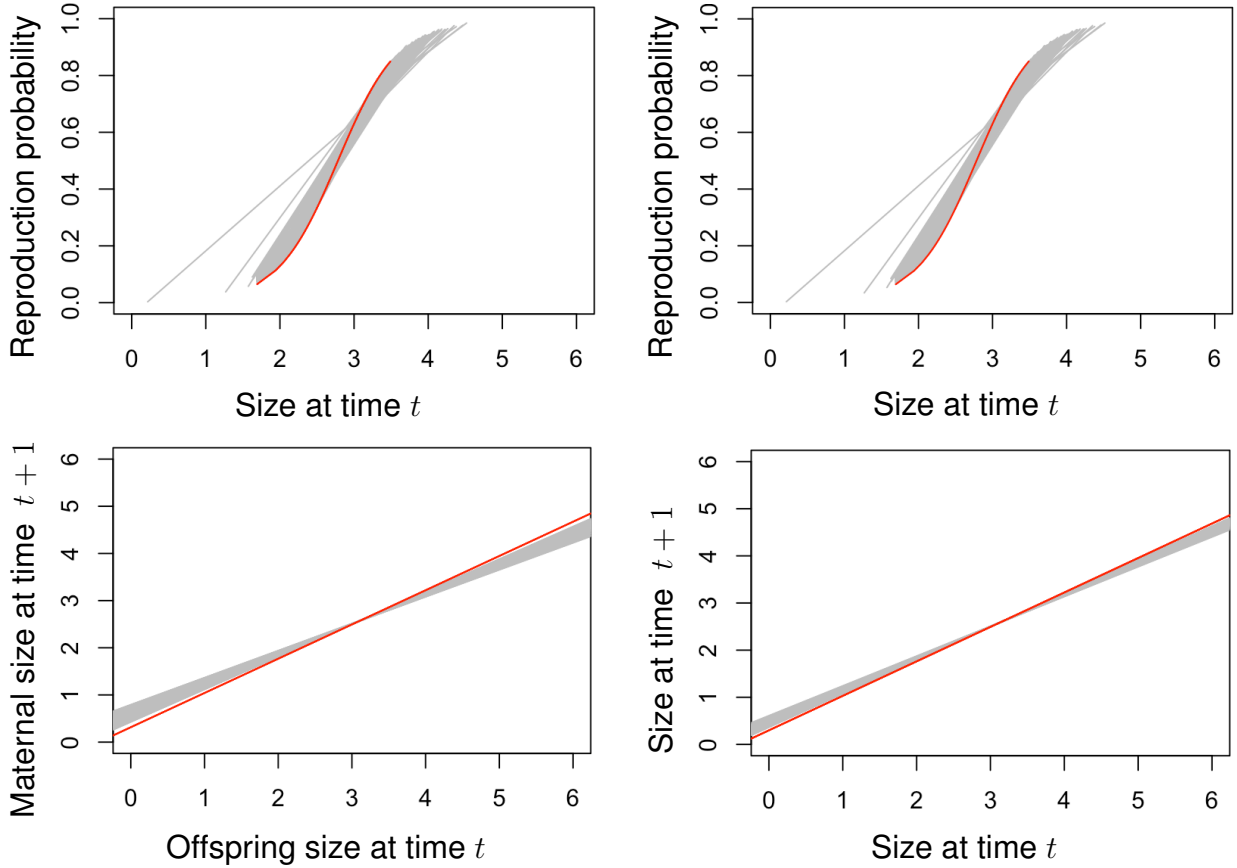
Figure 3.6: (a)Estimated reproduction curve(The probability of reproduction) using logistic regression (b) Reproduction curve using logistic M estimation. (c) Regression line between offspring and parent size using least square regression and (d) Regression line using M- estimation for 200 replication of mixed normal distribution

approach of 200 replicates from mixed normal (gray curves in Figure 3.5 (c)) and the red line shows true growth line. In Figure 3.5(d) regression lines are estimated using M-estimation (gray curves) and the red line shows true growth regression line.

Reproduction probability curves are estimated using logistic regression estimated via maximum likelihood as well as M estimation with 200 replications from mixed normal (gray curves). The black line shows true reproduction curve (Figure 3.6). Again, the reproduction probability profile obtained using maximum likelihood and robust logistic regression (Figure 3.6(b)) are similar. Regression lines between the size z of female soay sheep at t and the offspring size z_0 at time $t + 1$ are estimated for 200 replications from the mixed normal distribution using the least squares approach (gray curves in Figure 3.6 (c)). The red line in the figure shows true growth line of offspring size. This process is repeated using M estimation.

Comparison among true λ, IPM, and RIPM Est.
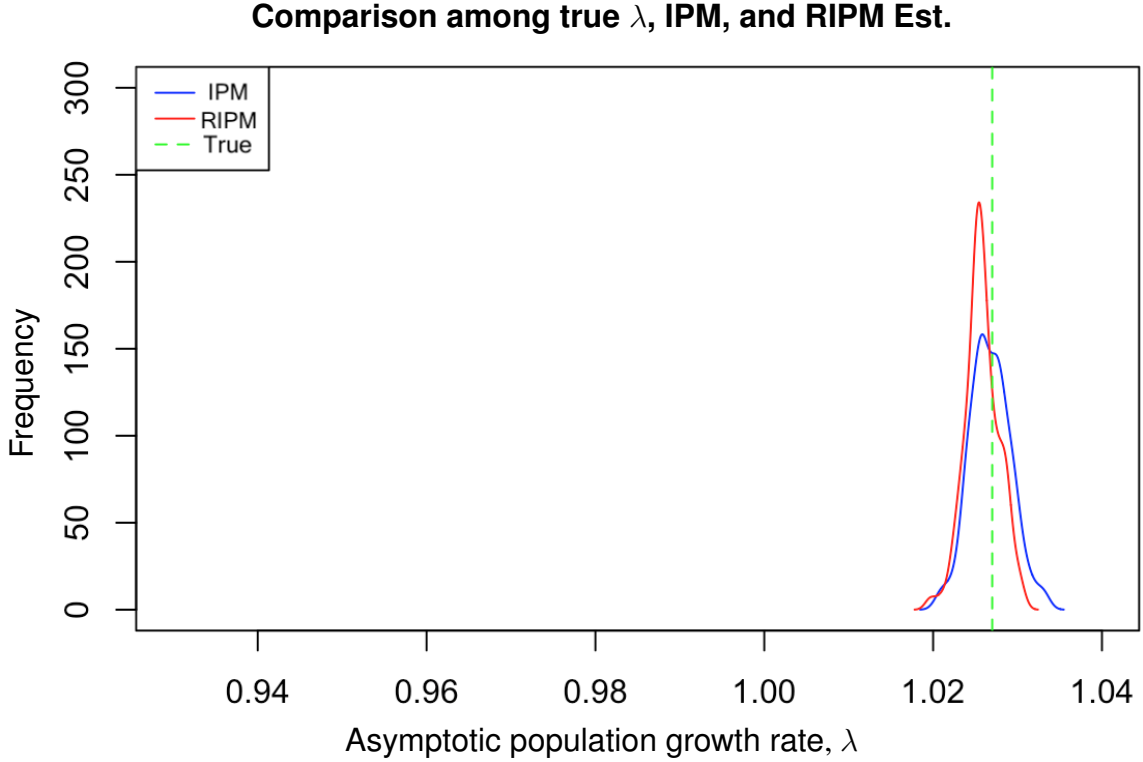
Figure 3.7: Density plot of the population growth rate for 200 replications of mixed normal distribution using IPM (blue line) and RIPM (red line). Green line is the true asymptotic population growth rate.

The results are given as the gray curves in Figure 3.6(d) and the red line shows true regression line of offspring size.

In Figure 3.7 the density curve of 200 estimated population growth rates λ using M-estimation (red line) and LS estimation (blue line). We observe a similar pattern in both scenarios.

For our second set of simulation experiments, we took scenarios in which we have a smaller proportion of more extreme outliers.

For the first of these, we simulated data where 2% of the data are sampled from the normal distribution with mean 5.978575 and standard deviation 0.2896221 and the remaining 98% of the data are sampled from the normal distribution with mean 2.978575 standard deviation 0.2896221. The mixed normal distribution can be written as:

$$z \sim \begin{cases} \delta(0 \leq X < 0.98) & N(\mu = 2.978575, \sigma = 0.2896221) \\ \delta(0.98 \leq X \leq 1) & N(\mu = 5.978575, \sigma = 0.2896221) \end{cases} \tag{3.2}$$
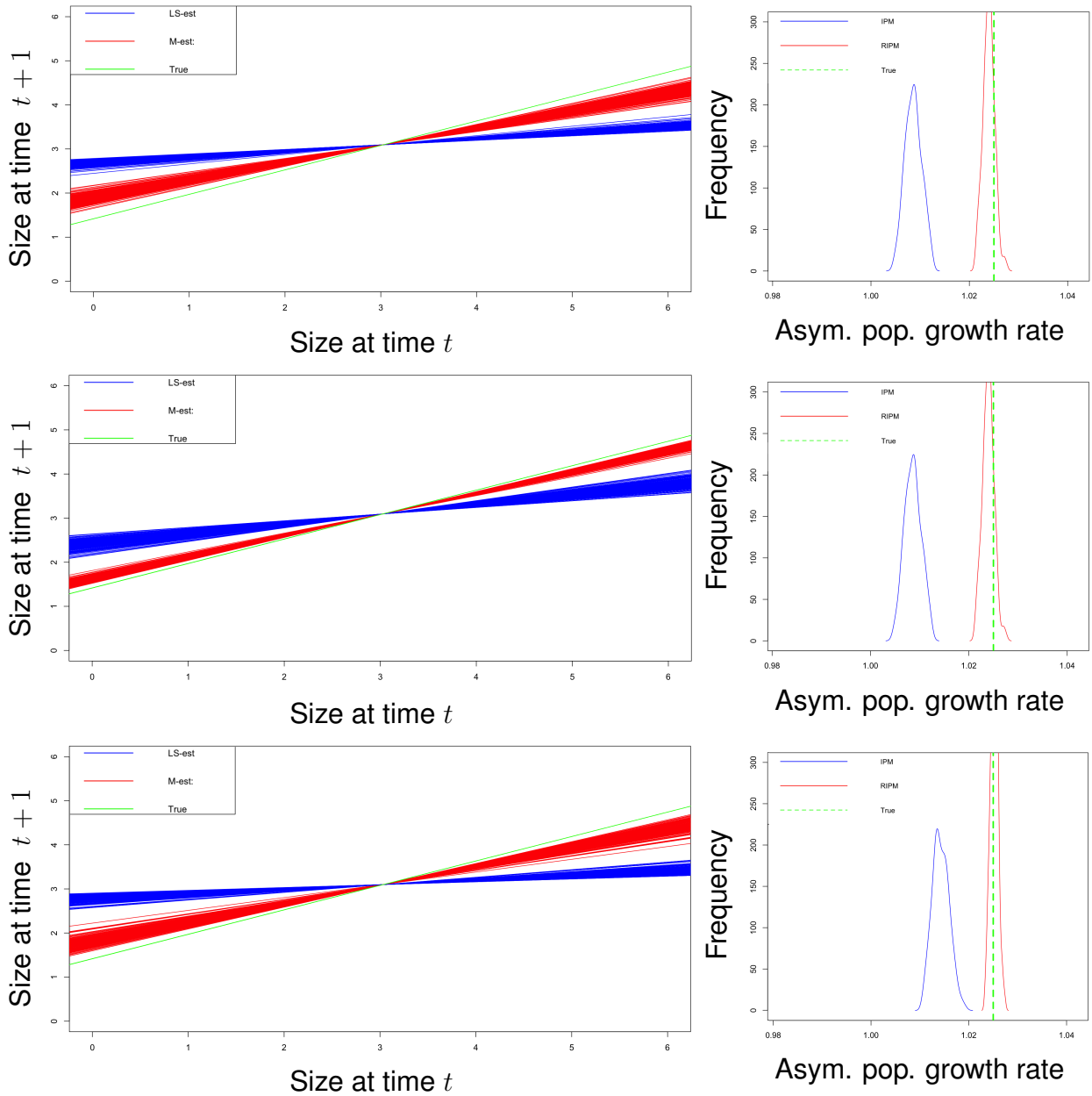
Figure 3.8: **Right panel:** Size regression lines using least squares regression (blue lines), true line(green line) and regression lines using M- estimation (red lines) for 200 replication of mixed normal distribution. **Left panel:** Density plot of the asymptotic population growth rate using classical IPM (blue curve) and robust IMPM (red curve). Green curve is the true asymptotic population growth rate. **Top panel:** 2% contamination, mean shift of 5, **Middle panel:** 1% contamination, mean shift of 5, **Bottom panel:** 1% contamination, mean shift of 7

where, as above, X is a uniform random variable on the interval $(0, 1)$ and $\delta(p)$ is 1 if p is true and 0 if p is false.

For the second of these, we simulated data with 1% of the data are sampled from the normal distribution with mean 5.978575 and standard deviation 0.2896221 and the remaining 99% are sampled from the normal distribution with mean 2.978575, standard deviation 0.2896221.

For the third of these, we simulated data with 1% of the data are sampled from the normal distribution with mean 7.978575 and standard deviation 0.2896221 and the remaining 99% are sampled from the normal distribution with mean 2.978575, standard deviation 0.2896221.

In each of these scenarios, the contaminant distribution is the original distribution shifted up by either 5 or 7. For the sake of brevity, we only report the linear fit of size transition and the estimated λ values. These are given in Figure 3.8.

In all the cases, we observe that the classical IPM woefully misses the true value while the robust IPM remains consistent.

Chapter 4

Robust IPM: A Case Study on Elephant Population Dynamics

We explored the population dynamics of an elephant population that resides in Amboseli National Park in Southern Kenya. The data are collected for the period 1973 to 2013. The Park, which was established in 1974, covers an area of 392 km^2 and the Amboseli ecosystem extends over a much wider area of approximately 3500 km^2 (Western, 1975).

The predominant habitat is semi-arid savannah (Pratt, Greenway & Gwynne, 1966) interspersed with woodlands and permanent swamps. Rain falls mainly during 2 seasons: the 'long rains' of March, April and May and the 'short rains' of November and December. Interspersed are 2 dry seasons: January, February, and most of March and then June to October. As is common in semi-arid savannahs the distribution of rainfall is unpredictable and highly variable from year to year. Rainfall averages 341 mm per year (140, n=26). The minimum annual rainfall was 136 mm and the maximum 837 mm (Moss, 2001).

AERP, the Amboseli Elephant Research Project is the Trust's research arm. Since 1972 AERP has studied the Amboseli elephants, making it today one of the longest studied populations of free living large mammals in the world. AERP was first initiated in 1972 by Cynthia Moss and Harvey Croze.

Relatively few poachers have been active in Amboseli Park. This is especially due to the Maasai people, and the constant presence of tourists and researchers. Thus, Amboseli is one of the few regions in Africa where the age structure of elephants has remained undistorted. The area is monitored by game wardens and scientists throughout the year.
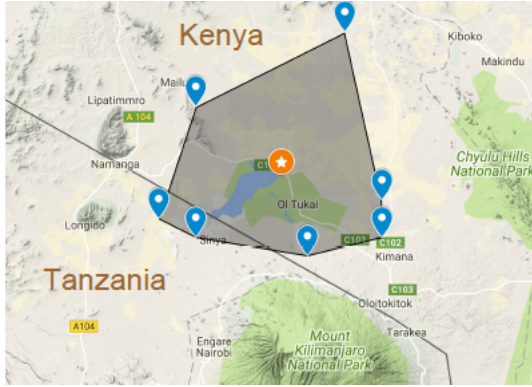
(a) Amboseli National Park

(b) Elephants in Amboseli National Park. *Image credit: Amoghavarsha JS, 2012*

The subjects of the Amboseli Elephant Research Project, mostly notably the elephant matriarch Echo, have been described at length in documentaries on PBS and Animal Planet (Amboseli Trust, 2014).

In this study we use the longitudinal data of Amboseli elephants from 1972 to 2012 to determine lifetime annual growth rate. In our IPMs, the Amboseli elephant is individual unit.

4.1   Analysis

Drought conditions are known to influence elephant population dynamics. Thus in the analysis of the Amboseli elephant data, we take z_t to be the level of drought in a particular year and z_{t+1} the level of drought the following year. The variable 'repr' represents if an individual reproduced (repr =1) or not (repr = 0) and 'surv' represents if an individual survived (surv = 1) or not (surv = 0).

It is strongly recommended to draw a life cycle diagram before implementing IPM. We presented the life cycle diagram of an elephant in Figure 4.1. This is a pre-reproductive census as reproduction occurs right after the census at time t. The upper and lower part of this life cycle make growth kernel, $P(z_{t+1}, z_t)$, and fecundity kernel, $F(z_{t+1}, z_t)$, respectively. The Amboseli elephant life cycle begins at time t and the individuals who survive ($s(z_t) = 1$) transit to the next stage. There is a size transition $g(z_{t+1}, z_t)$ to the next time census at $t + 1$. On the other hand, the elephants reproduce ($r_p(z_t) = 1$) and each offspring has to survive ($s(z_O) = 1$) to move forward to the next stage. At this stage the offspring may or may not grow in size according
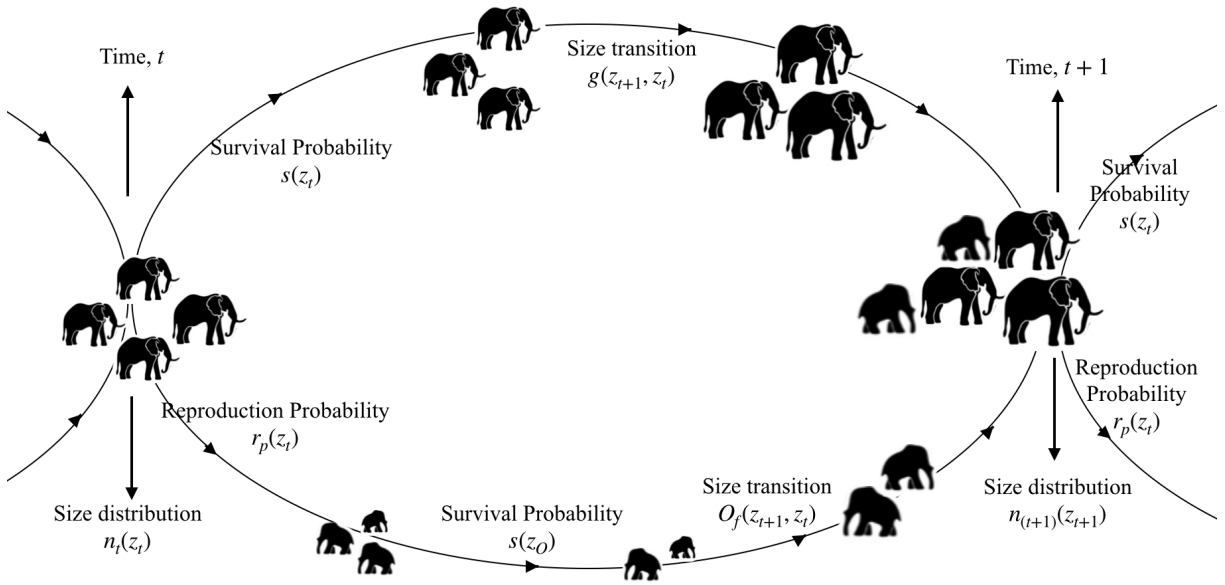
Figure 4.1: Life cycle of elephant

to size transition $O_f(z_{t+1}, z_t)$ as presented in Figure 4.1. The life cycle of the elephant is used as a guideline to formulate the main kernel $K = s(z_t)g(z_{t+1}, z_t) + s(z_O) * r_p(z_t) * [f(z) = 1] * 0.5 * O_f(z_{t+1}, z_t)$ for modeling female elephant population dynamics.

4.2 IPM execution

To build the IPM of the Amboseli elephant data set, first we find the parameters of the vital rates functions as presented in the Table 2.1.

At this stage we construct an IPM kernel which consists of survival and reproduction kernel. For Amboseli elephant IPM kernel can be written as:

$$K(z_{t+1}, z_t) = s(z_t) * g(z_{t+1}, z_t) + s(z_t)r_p(z_t)f(z_t) * 0.5 * O_f(z_{t+1}, z_t)$$

where, $K(z_{t+1}, z_t)$ is the full kernel. $s(z_t)$ is the survival probability from time t to time $t + 1$. $g(z_{t+1}, z_t)$ is the growth distribution. p_r is the reproduction probability. We assume there are equal male and female elephant births in Amboseli park and we only track the dynamics of female elephants; hence the multiplication of the kernel by 0.5. $f(z_t)$ is the size distribution of recruits. $N_t(z_t)$ is the size distribution at time t and the integration of $N_t(z_t)$ over size is
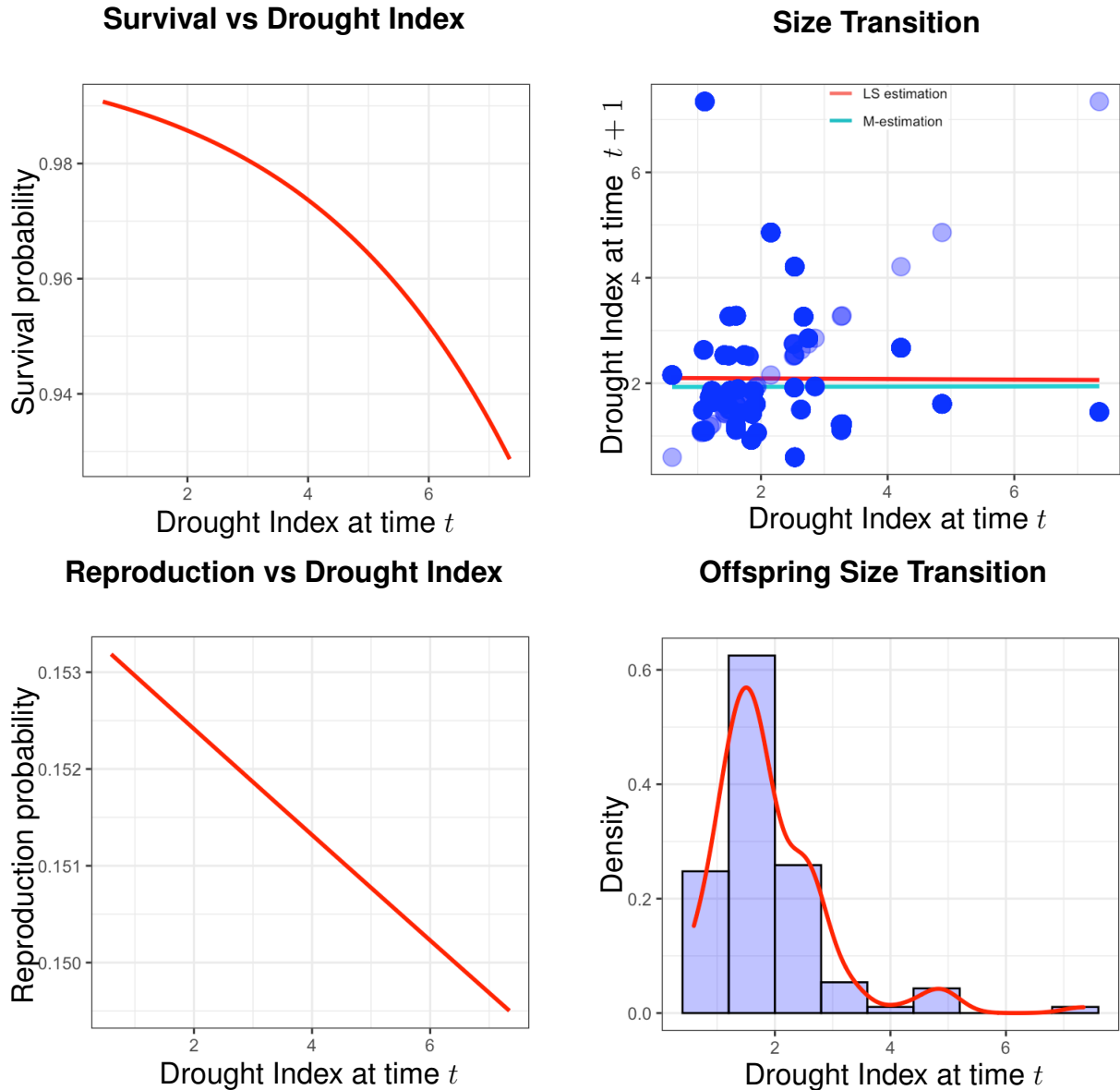
Figure 4.2: Estimating the model parameters of Amboseli elephant data. (a) Scatter plot of survival against drought,: the red line is logistic regression estimation of the survival probability. (b) Scatter plot of the drought,z_t in current year, t against to the drought, z_{t+1} in the next year $t + 1$.: The darked turquoise line and the red line are the M-estimation and LS-estimation of drought index in time, t respectively. (c) Scatter plot of reproduction against drought index, the red line is the logistic regression estimation of the reproduction probability. (d) Size distribution of recruits.

the number of individuals in that time interval. $K(z_{t+1}, z_t)$ is a prediction function of z_t that projects the current size z_t to the future size z_{t+1} for survivors and recruits. In our case, size is a proxy for drought. So,

$$N_{t+1}(z_{t+1}) = \int_{z^*} K(z_{t+1}, z_t) N_t(z_t) \, dz$$

where z^* consists of the lower (0.597) and upper (7.342) limits of the drought index.

Robust integral projection model (RIPM) performs better when there is a suspicion that the data are contaminated with outliers or influential observations. We demonstrated the benefits of RIPM over standard IPM in Chapter 3. It is well-known that ecological data will have outliers since they are collected in uncontrolled conditions. Thus, it is practicable to explore whether using RIPM to estimate vital rates provides an improvement over the standard IPM. Furthermore, in this case study, drought index was used as a generic size variable. Outliers had been observed in the plot of drought index as reported in Figure 4.2 as a suspicion of the output of drought in 1976, 1984, 2008, 2009, and 2010. In this situation, we will compute the vital rates using RIPM and IPM, and thereupon compare the results with the true demographic information of Amboseli elephants.

Using standard IPM we calculated the growth, $P(z_{t+1}, z_t)$, and fecundity, $F(z_{t+1}, z_t)$ kernels which are the main two components for IPM as reported in Figure 4.4. We fit the classical IPM and robust IPM and we observed that the ridge of the growth kernel using standard IPM is lower than the ridge of growth kernel using robust IPM. Again, we observed a similar pattern in the full kernels given in Figures 4.5 and 4.6. The effect of this discrepancy on asymptotic growth rate is discussed in Section 4.4. However, the ridge and shape of the graph of the fecundity kernel in both models remains unchanged as reported in Figures 4.3 and 4.4.

For the IPM matrices, we used 100 bins, which gave rise to a big matrix of 100×100 of all pairwise combinations of all sizes (drought indices). The dominant eigenvalue of this big matrix is 1.025 which is the asymptotic population growth rate λ of the Amboseli elephant population. Equivalently, the growth rate of the Amboseli elephant population over the years 1972 - 2012 is 2.5% yearly.

The comparison among estimated population using robust IPM, standard IPM, and the true population from 1972 to 2012 are reported in Figure 4.13. The right eigenvector provides stable size distribution as reported in Figure 4.15. Furthermore, in Figure 4.14 revealed the reproductive distribution of the population, which is the left eigenvector of the big matrix.
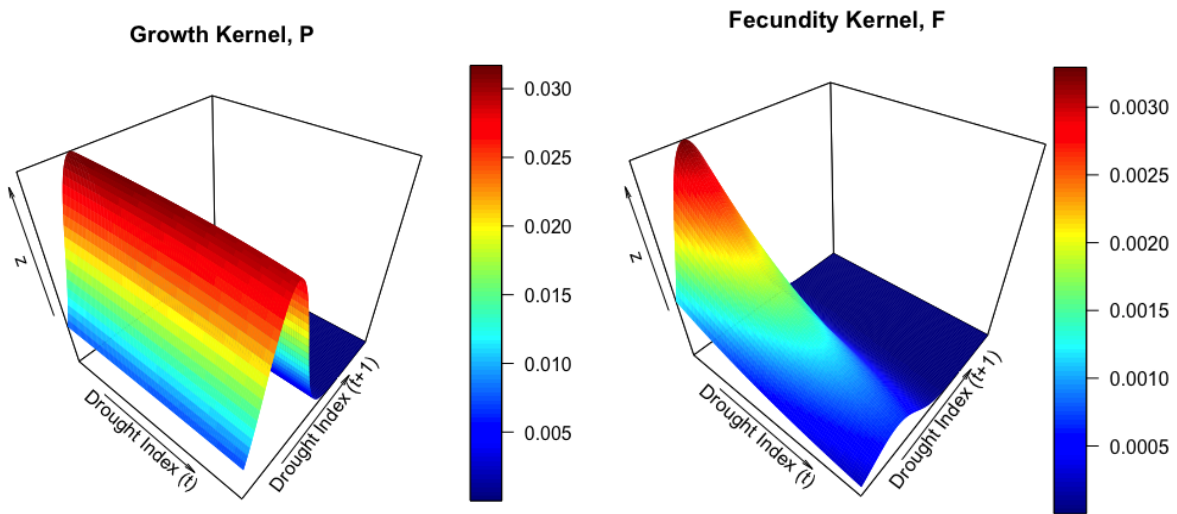
Figure 4.3: Growth and fecundity kernel using robust IPM. In the fecundity kernel, $F(z_{t+1}, z_t)$ we notice the higher elevation on the lower drought index side indicates the elephants produce more offspring during the weather of lower drought index. Again, the steep slope represents the individuals who transit from time t to the next next census at time $t+1$ conditional on survival $(s(z_o) = 1)$ decrease in size (drought index) gradually. On the other hand, in the growth kernel the ridge located along the vertical of drought index at time t indicates the individuals who survive and transit to the next census without changing much in drought index. We also observed that growth kernel is higher than the fecundity kernel, indicating that growth kernel contributes more on asymptotic population growth rate. The distribution of this contribution will be discussed in section 4.4.
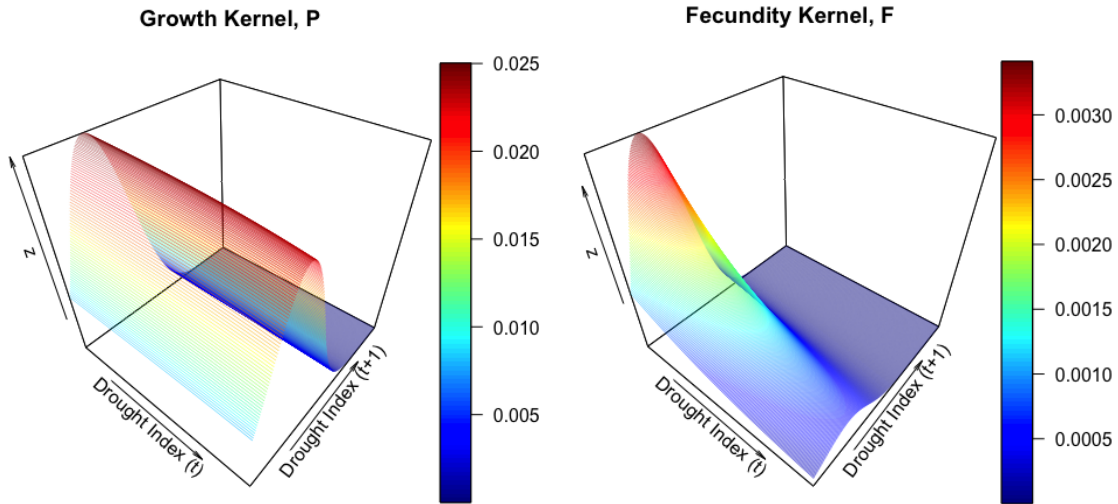
Figure 4.4: Growth and Fecundity Kernel using standard IPM
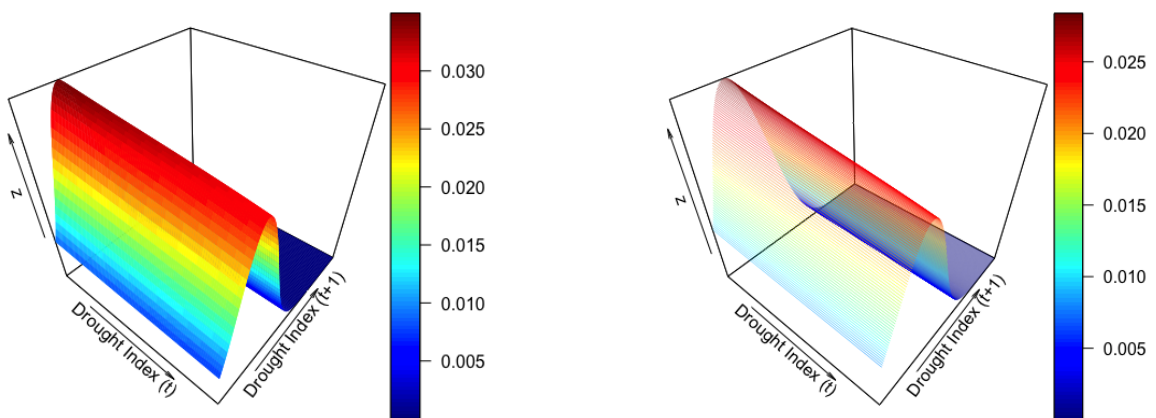
Figure 4.5: Full kernel, $K(z_{t+1}, z_t) = P(z_{t+1}, z_t) + F(z_{t+1}, z_t)$ using robust IPM.

Figure 4.6: Full kernel, $K(z_{t+1}, z_t) = P(z_{t+1}, z_t) + F(z_{t+1}, z_t)$ using standard IPM.

In Figure 4.4 we observed a similar pattern in the growth and fecundity kernel using standard IPM as we observed in Figure 4.3 using robust IPM. The only difference we noticed was that the height of the growth kernel is lower using standard IPM than the robust IPM. This means important areas of the growth kernel are better highlighted in the robust IPM that the standard IPM.

4.3   Model Diagnostics

Dynamics of population varies population to population, proposing a good model is an iterative process of finding errors and then find its solutions [1]. The diagnostics plots for the elephant growth kernel in Figure 4.7 (a) shows no particular pattern in the residuals versus corresponding fitted values. The red line should be approximately horizontal at zero. The presence of a pattern may indicate a problem with some aspect of the linear model. So we can assume linear relationship between the regressors and the response variables. The Q-Q plot of residuals can be used to check the normality assumption. The Q-Q plot of a normal distribution should approximately follow a straight line along the diagonal. In Figure 4.7 (b) all the points do not fall approximately along this reference line, so we can assume non-normality. Again, using Shapiro-Wilk normality test data we get $W = 0.81409$ and p-value $< 2.2e - 16$. Hence, we reject null hypothesis saying that the residual data is normally distributed. Scale-location plot in Figure 4.7 (c) to check the assumption of homoscedasticity. Here, we check to see if there is a pattern in the residuals . If the red line in the plot is flat and horizontal with equally and randomly spread data points, then we can assume that the homoscedasticity assumption is satisfied. In our case, this is not satisfied. In the Figure 4.7 (d) the data does not present any influential points. We do not observe red dashed line (Cook's Distance) in the Residuals vs Leverage plot, so we can assume that all points are well inside of the Cook's distance bounds.

To diagnose the survival probability we plot survival probability vs drought index in Figure 4.8 (a). The black line the estimated line using logistic regression and the red line is the predicted line using generalized additive model via the restricted maximum likelihood (REML). Figure 4.8 (b) shows fitted spline regression (blue line, not straight), so we can assume the
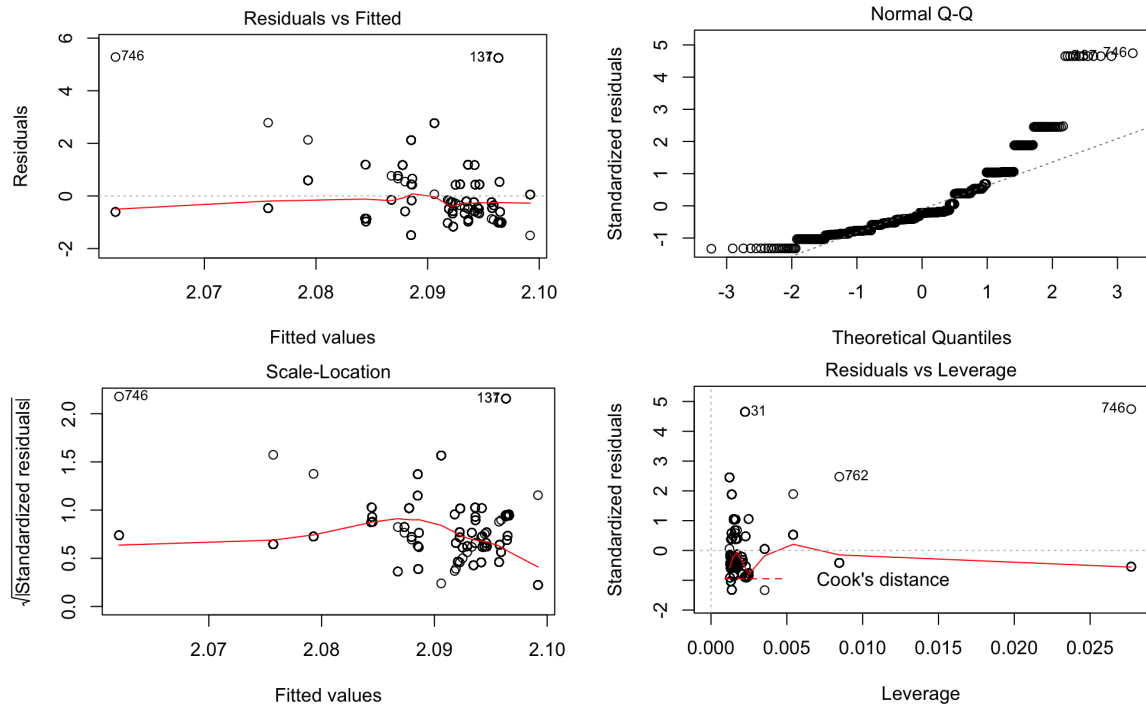
Figure 4.7: (a) Residuals vs fitted plot (b) Normality of the residuals plot
(c) Scale-Location plot to check homoscedasticity (d) Residuals vs Leverage plot

generalized additive model and logistic regression are not equivalent for the estimation of sur-
vival probability. To diagnose the reproduction probability we plot reproduction probability
vs drought index in Figure 4.8 (c). The black line the estimated line using logistic regression
and the red dot is the predicted line using generalized additive model (REML). Figure 4.8 (d)
shows fitted spline regression (straight blue line), so in this case generalized additive model and
logistic regression give more or less equivalent results for reproduction probability .
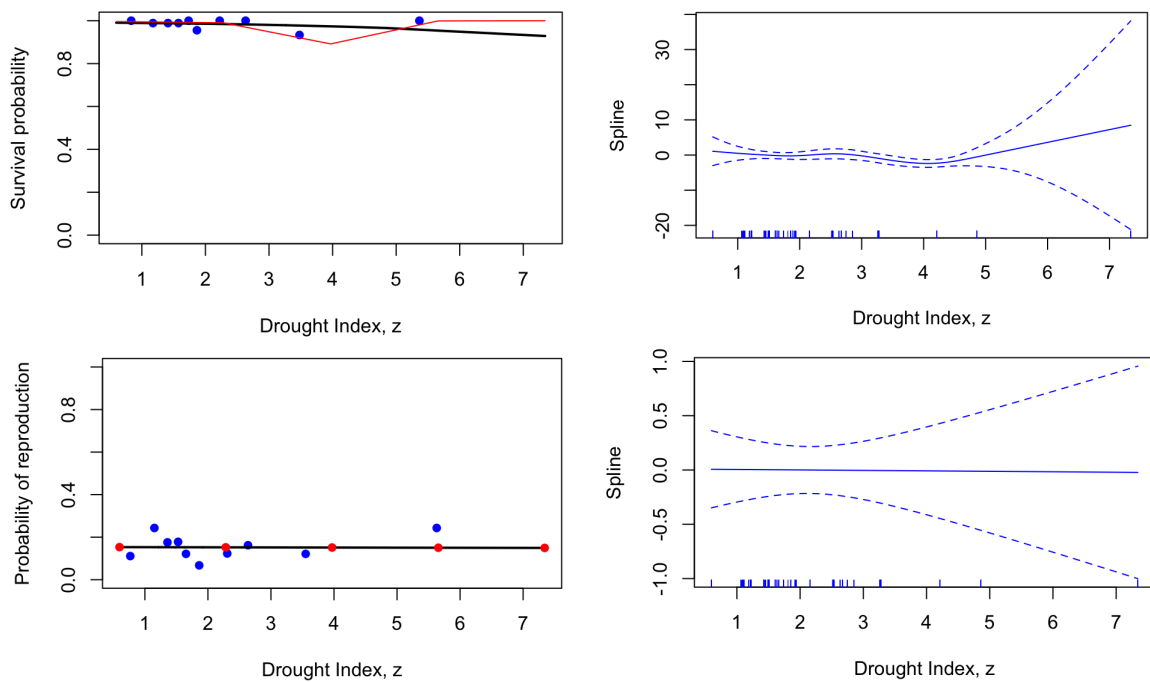
Figure 4.8: (a) Survival probability versus drought index, z (b) Spline vs Drought index, z (for survival) (c) Reproduction probability versus drought index, z (d) Spline vs Drought index, z (for reproduction)

4.4 Sensitivity and Elasticities

The main contemplate in this study is to suggest the best ways to the authority of Amboseli National Park to administer the elephant as an endangered species. By exploring sensitivity and elasticity, we can provide the information to the authority that which part of the generic size variable (Drought Index) are contributing most to the asymptotic population growth rate. Having the knowledge, authority can significantly increase the population growth rate by monitoring the relevant information.

The first step for the sensitivity analysis is to observe the degree to which λ changes when additional impediment is applied to the full kernel. We do this analysis by impeding full kernel $K(z_{t+1}, z_t)$ to $K(z_{t+1}, z_t) + \epsilon I(z_{t+1}, z_t)$ where ϵ is a small constant and $I(z_{t+1}, z_t)$ is impediment kernel.

The sensitivity of an element of the kernel matrix can be defined as

47
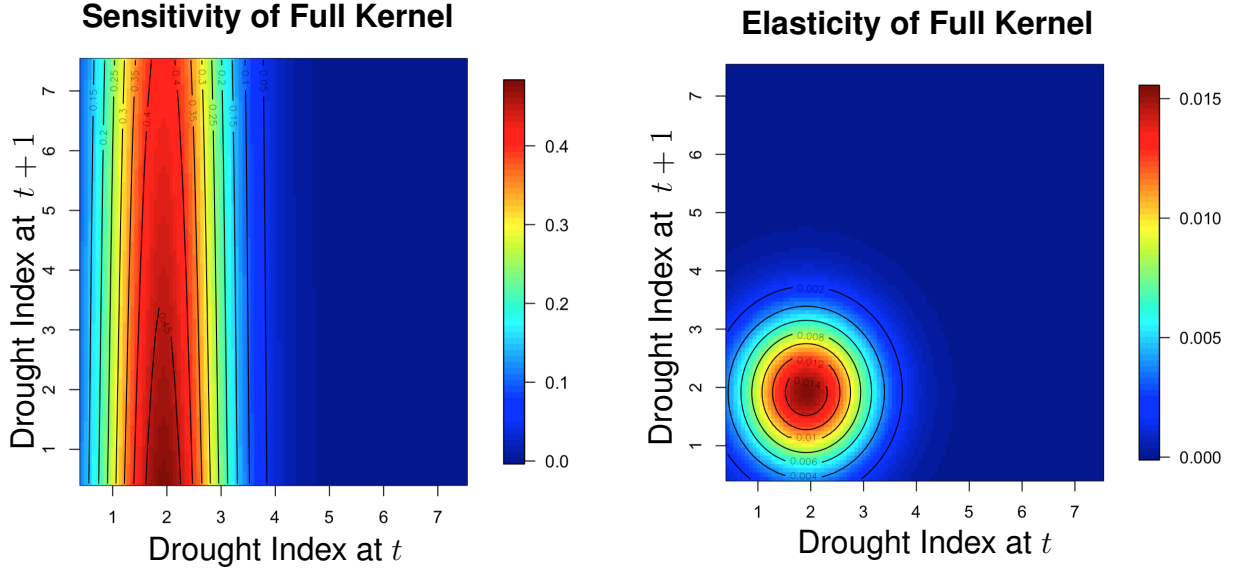
Figure 4.9: Sensitivity and elasticity function of the kernel, K using robust IPM.

$$s_{ij} = \frac{\partial \lambda}{\partial K_{ij}} = \frac{v_i w_j}{< w, v >} \tag{4.1}$$

where, v_i is the ith element of the reproductive value vector and w_j is the jth element of the stable size vector (Caswell 1978). So sensitivity analysis reveals how very small changes in K affect the asymptotic population growth when other elements in K matrix are held constant.

One problem in interpreting the sensitivity is that demographic variables are measured in different units. Elasticity analysis estimates the effect of a proportional change of the vital rates on λ. The elasticity of a matrix element, e_{ij} is defined as :

$$e_{ij} = \frac{\partial \log \lambda}{\partial \log K_{ij}} = \frac{\frac{\partial \lambda}{\lambda}}{\frac{\partial K_{ij}}{K_{ij}}} = \frac{K_{ij} s_{ij}}{\lambda} \tag{4.2}$$

where K_{ij} is the kernel elements (Zuidema and Franco, 2001). Elasticity , e_{ij} of a matrix element, K_{ij} can be acquired by multiplying $\frac{K_{ij}}{\lambda}$ by sensitivity, s_{ij} in equation 4.1.

The matrix elements with higher sensitivity contribute more to the changes in asymptotic growth rate. We observed the sensitivity surfaces of K with the transitions from the lowest drought index areas have the greatest effect on asymptotic growth rate as shown in the Figure 4.9. Thus the sensitivity analysis provides an important insight on the impact of drought z_t on Amboseli elephants population growth rate. The Amboseli National Park's authority may

48
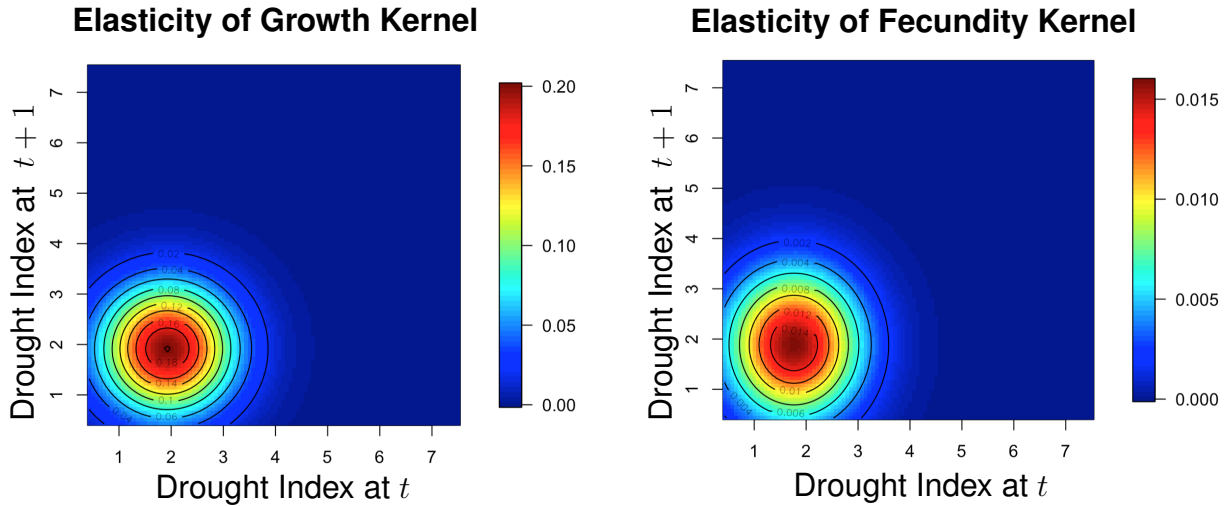
Figure 4.10: Elasticity Analysis of kernel $P(z_{t+1}, z_t)$ and $F(z_{t+1}, z_t)$ using robust IPM.

use this information to provide drought mitigating measures to decrease the impact of seasonal droughts when the drought index shows high values. This will have the impact of increase the asymptotic population growth rate, λ. Again, the elasticity function shows the elements of the kernel have greatest impact on population growth rate in the area of smaller drought transitions. To compare which kernel (survival or reproduction) has the greatest impact on λ we need to find elasticity of survival and reproduction kernel individually.

Figure 4.10 gives the elasticity function of $P(z_{t+1}, z_t)$ and $F(z_{t+1}, z_t)$ for the robust IPM. We see survival/growth kernel has the greatest impact on population growth rate. To compare this two elasticity function we can double integrate this surface individually to get the total volume. For growth/survival and reproduction kernel elasticity function we get 0.9206276 and 0.0793724, respectively. So survival and growth contribute 92% to the population growth rate while reproduction only contributes about 8% to the total population growth rate. These results are not very surprising in light of the fact that elephants are long-lived animals with relatively low reproduction rates.
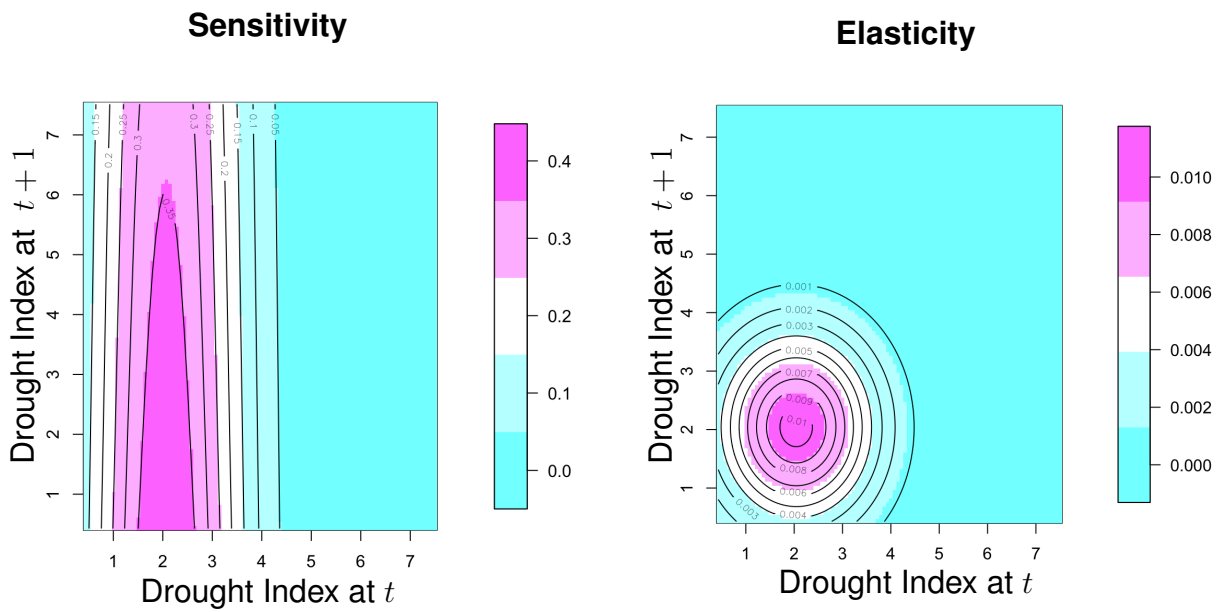
Figure 4.11: Sensitivity and elasticity analysis using standard IPM.
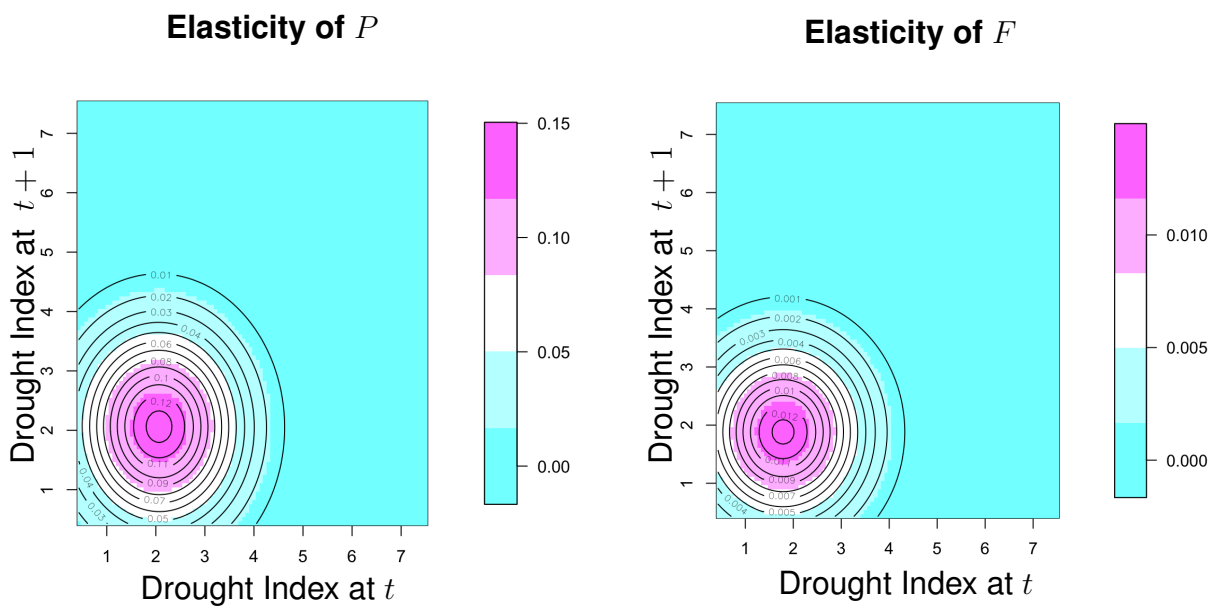
Figure 4.12: Elasticity analysis of $P(z_{t+1}, z_t)$ and $F(z_{t+1}, z_t)$ using standard IPM.

For the standard IPM, we observed similar size transitions in Figures 4.11 and, 4.12 as we did in Figures 4.9 and 4.10. However, we noticed a significant difference in the height of elasticities in fecundity kernels. We concluded from this comparison that the standard IPM underestimated around 33% of the fecundity kernel than the robust IPM, once again highlighting the better performance of robust IPM. The effect of this is further confirmed when we compare the population estimated via the robust IPM versus the standard IPM (Figure 4.13). Again, the difference between the two estimates is stark. While the standard IPM is affected by outlying values of the drought index and an aberrant population decline in 2011-12, the robust IPM remained true to the overall trend.

Finally, the reproductive value computes the "worth" of individuals of different stage classes in terms of future offspring they are destined to contribute to the next generation, adjusted for the growth rate of the population (Fisher 1930). We noticed in Figure 4.14 higher drought indices contribute less to the future generations. That is reproduction and/or survival are both negatively affected by drought conditions.

Comparision among RIPM Estimation, IPM Estimation, and True Population
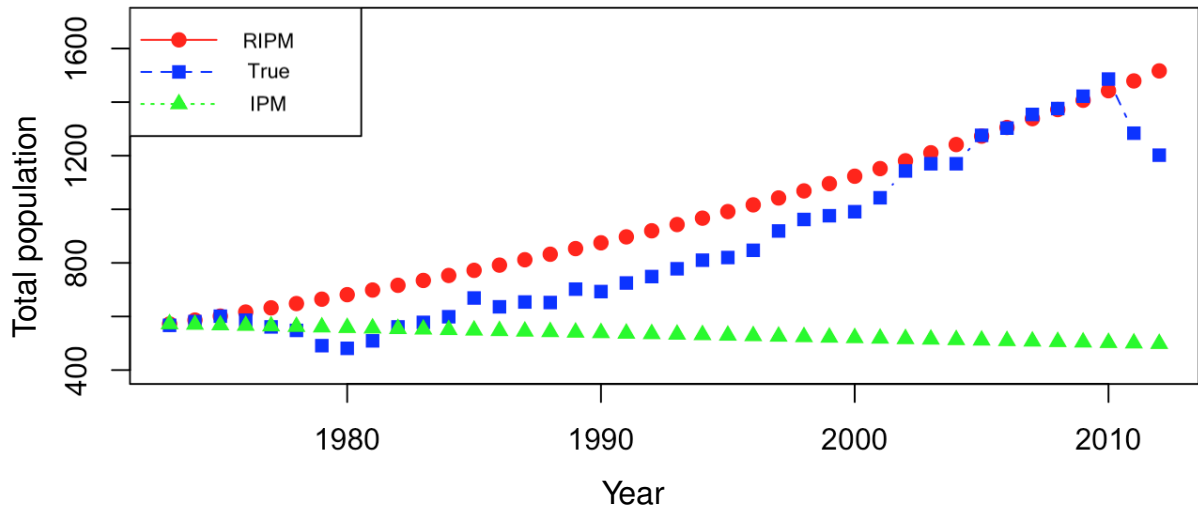
Figure 4.13: Total population of Amboseli Elephant(blue square), red circle and green triangle are the estimated population using robust IPM and standard IPM estimation respectively.

Reproductive Values
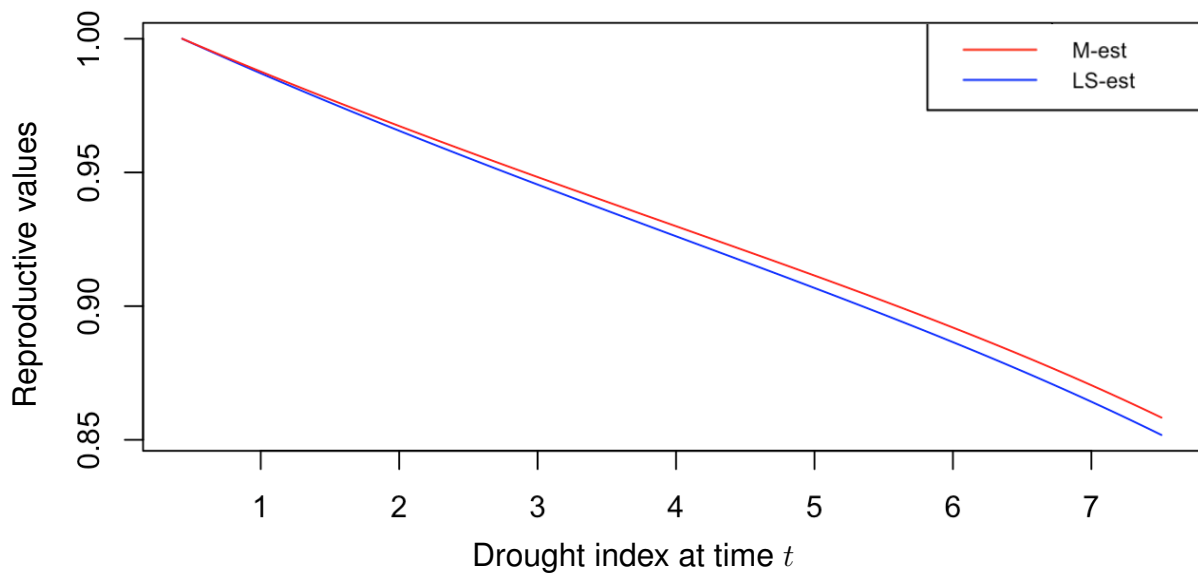
Figure 4.14: Reproduction values

Stable Size Distribution
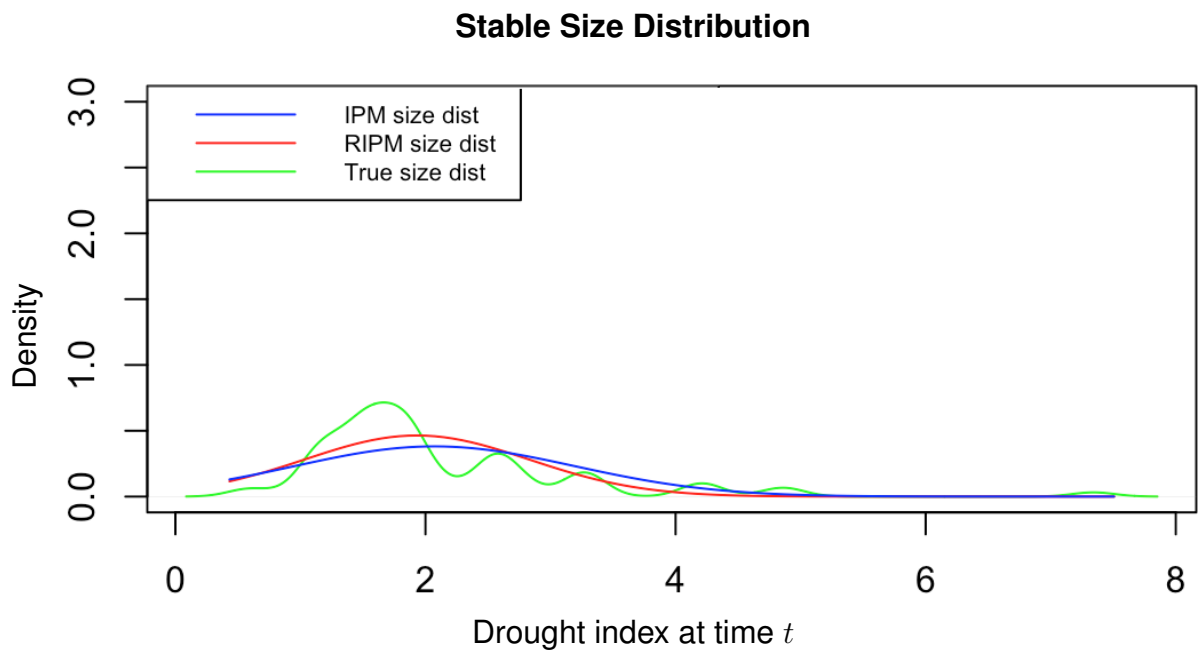
Figure 4.15: The density plot of initial size distribution(green line), estimated stable size distribution using Robust IPM(red line)

Chapter 5

Estimating the Population Growth Rate Using Permuted Integral Projection Model

5.1 Introduction

Integral projection model (IPM) consists of two parts, growth/survival kernel and fecundity/reproduction kernel. Fecundity kernel is the multiplication of fecundity probability, reproduction probability, and the recruitment size distribution.

Without the information of reproduction probability, it is almost impossible to calculate population growth rate using the standard integral projection model. To calculate reproduction probability, it is essential to know who is giving birth; however, collecting this information is significantly expensive and time consuming, in some scenarios it is almost impossible (fish, turtle). The existing integral projection model fails to calculate the fecundity kernel in the situation of incomplete reproduction probability. In this research study, we propose a permuted integral projection model (PIPM) in the situation of missing information of reproduction (who is giving birth). We demonstrate the benefits of the proposed approach by analyzing the population dynamics of African elephants in Amboseli National Park, Kenya.

5.2 Permuted Integral Projection Model

Reproduction is the process of producing new offspring. Individuals who transit from time, t to time, $t + 1$ given that individuals survive, may or may not reproduce (Repr = 1 or 0). We usually assign the random variable of reproduction 1, if individuals reproduce and 0, if not. Matrix projection model forces continuous trait into discrete stages while the integral projection model restrain from these artificial subdivisions. The first step is to estimate the reproduction probability given particular size (drought index) using the logistic regression

model. The estimated logistic curve is used as the function of reproduction probability. if any size (drought index) is given, then the function will provide the reproduction probability as an output. The reproduction probability is estimated as :

$$p(z_{t+1}) = \frac{\exp(\beta_0 + \beta_1 z_t)}{1 + \exp(\beta_0 + \beta_1 z_t)} \tag{5.1}$$

where z_{t+1} and z_t are the size at time $t+1$ and t respectively. β_0, and β_1 are estimated from true data (reproduction vs. size at time t) using the logistic regression model.

In the situation where data has incomplete reproduction (1 or 0) information (who is giving birth), we are unable to estimate β_0 and β_1. Also, it may be very expensive and time consuming to track the reproduction information (who is giving birth). If we know how many babies are born yearly, reproduction probability can be recovered by a restricted permutation as discussed in section 5.4.

The asymptotic population growth rate λ can be calculated using the algorithm given in the following section. It involves a complex permutation scheme using binary matrices with various restrictions on the margins and minimum gaps among the 1 values.

5.3   Computational Algorithm

The outline of the algorithm pertaining to the permuted integral projection model (PIPM) is as follows:

Step 1: Generate a null vector of size m to save all asymptotic population growth rate, λ

Step 2: Run a for-loop for m iterations.

Step 3: Define a function $\text{Par}(a_z, b_z, \sigma_z, a_s, b_s, a_{r_p}, b_{r_p}, \mu_{rz}, \sigma_{rz})$ contains all parameters of vital rate.

Step 4: Using generalized linear regression calculate all vital rates parameters and update the function in step 1.

Step 5: Define growth function, $g(z_{t+1}, z_t) = \frac{1}{\sqrt{2\pi\sigma_z^2}}\exp\left(\frac{(z_t - \mu_z)^2}{2\sigma_z^2}\right)$ that returns output as size, z_{t+1} at time $t+1$ given the size, z_t at time t as follows:

- plot individual size, z_t at time t versus the individual size, z_{t+1} at time $t+1$

55

- using m-estimation calculate regression parameters (a_z and b_z) and standard deviation σ_z

- define $\mu_{z_t} = a_z + b_z * z_t$, where z_t is the size at time t

- generate a normal distribution with mean μ_z and standard deviation σ_z

- input the size, z_t at time t using **steps 17 through 19**

- the growth function $g(z_{t+1}, z_t)$ returns size z_{t+1} at time $t+1$ provided the size, z_t at time t.

Step 6: Define survival function, $s(z_t) = \frac{1}{1+\exp(-\mu_{lgs})}$ which returns the survival probability of size, z_t at time, t as follows:

- define $\mu_{lgs} = a_s + b_s * z_t$, where z_t is the size at time t

- input the size, z_t at time t using **steps 17 through 19**

- return the survival probability of size z_t at time t

Step 7: Define reproduction function, $r_p(z_t) = \frac{1}{1+\exp(-\mu_{lgp})}$ which returns the reproduction probability of size, z_t at time, t as follows:

Step 8: Since we have incomplete reproduction information, by using steps 9 through 12, reproduction information can be permuted.

Step 9: Calculate the total offspring in a year

Step 10 : Find the total possible candidates who may give birth by restricting their age and survival information. If an individual gives birth within the current year, then this individual would be considered as possible the candidate for the next two or three years depending on their gestation period.

Step 11: Permute all offspring using the strategy discussed in section 5.4 to the possible candidates (mother).

Step 12: Save the reproduction probability corresponding to the individual as $r_p(z_t)$.

- where, $\mu_{lgp} = a_{r_p} + b_{r_p} z_t$, and z_t is the size at time t

- input the size, z_t at time t using **steps 17 through 19**

- return the reproduction probability of size z_t at time t

Step 13: Define recruitment function, $O_f(z_{t+1}, z_t$ which returns the size distribution of offspring at time $t + 1$ as follows:

- plot maternal size, z_t at time t versus the offspring size, z_{t+1} at time $t + 1$

- using m-estimation calculate regression parameters a_{rz} and b_{rz} and standard deviation σ_{rz}

- calculate mean $\mu_{rz} = a_{rz} + b_{rz}z_t$ and the standard deviation σ_{rz} of the size,z_t

- generate a normal distribution with mean μ_{rz} and standard deviation σ_{rz}

- input the size, z_t at time t using **steps 17 through 19**

- return the offspring size transition of z_{t+1} at time $t + 1$

Step 14: Define survival or growth kernel function, $P(z_{t+1}, z_t) = g(z_{t+1}, z_t) * s(z_t)$.

Step 15: Define reproduction kernel function, $F(z_{t+1}, z_t) = O_f(z_{t+1}, z_t) * s(z_t) * (1/2) * r_p(z_t)$.

Step 16: Define full kernel $K(z_{t+1}, z_t) = P(z_{t+1}, z_t) + F(z_{t+1}, z_t)$.

Step 17: Calculate the size range of the data set as follows: $L = l - 0.5$ and $U = u + 0.5$ where l and u are min and max of the size respectively.

Step 18: Define number of elements in the matrix $n = 200$.

Step 19: Calculate mesh points, $z_t = L + k * (U - L)/n - (U - L)/2n$, where, $k = 0, 1, 2, \ldots n$.

Step 20: Define a function which evaluates all pairwise components of the vector z_t (mesh points) and returns all elements of $n \times n$ matrix as an output of the function $K(z_t, z_t)$.

Step 21: Calculate dominant eigenvalue of the $n \times n$ matrix.

Step 22: Repeat **Step 3** through **Step 21** .

Step 23: Draw the histogram and box-plot of λ as reported in Figure 5.7.

5.4 Restricted Permutation

Let us consider a 5×3 matrix.

A	B	C
1	0	0
0	1	0
0	0	0
1	0	0
0	1	0

We need to determine how many unique ways this matrix can be permuted by restricting row sum and spaces (at least two zeros) between two ones in a column. For small matrices, answers should be straightforward, but it would be a tedious process to calculate all possible combinations satisfying those restrictions for high dimensional matrices. Thus, we developed a function that will solve this problem numerically.

For the example above, the matrix has five rows and the row sums are 1,1,0,1 and 1. Let $\oplus$ be the row sums. For the above matrix $\oplus = [1, 1, 0, 1, 1]^t$. Again, let ϕ denote the minimum allowable space(s) between the non zero elements in each column and π denote the total number of distinct permuted matrices. For $\phi = 3$, we get total 6 distinct matrices as follows:

$$
\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}
\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}
\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}
$$

In the above matrices, it is observed that the row sum is always equal to $\oplus = [1, 1, 0, 1, 1]^t$. There are at least three zeros ($\phi = 3$) between ones in each column of the matrices.

Table 5.1: Restricted Permutation Output.

Iteration	ϕ	π
30	1	20
500	1	36
5000	1	36
30	2	18
500	2	24
5000	2	24
30	3	6
500	3	6
5000	3	6

We explored some more scenarios in Table 5.1 and observed that the value of ϕ increased as the value of π decreased. From these observations, it was found that increasing the value for ϕ decreased the room for non-zero elements to permute. For $n = 30$ (number of iterations) and $\phi = 1$, π did not converge. However, for $n = 500$ and 5000, π converged in all scenarios ($\phi =1$, 2 and 3). We observed the total number of unique permuted matrices under the restriction of $\oplus$ and ϕ were always equal to the multiple of matrices column numbers (p).

To observe the effect on π in increasing the column (p) of a matrix, we explored some scenarios in Table 5.2. We continued to increase a column $[1\,0\,0\,0\,0]^t$ to the above 5×3 matrix. We observed as p increased, the value for π also increased, and in all scenarios π is multiple of p.

These observations give us much needed information on the number of permutations that need to be considered for actual populations (eg. the elephant population data) as these computations are time consuming.

Table 5.2: Scenarios of Restricted Permutation in increasing the column of a matrix .

Iteration	p	ϕ	π
500	4	1	137
5000	4	1	144
10000	4	1	144
10000	4	2	108
10000	4	3	24
5000	5	1	400
15000	5	1	400
15000	5	2	320
15000	5	3	60
5000	6	1	898
15000	6	1	900
15000	6	2	750
15000	6	3	120
10000	7	1	1755
25000	7	1	1764
50000	7	1	1764
25000	7	2	1512
25000	7	3	210

5.5   Real Data Analysis

We explore the EB family population (elephants) in Amboseli National Park in Southern Kenya from 1973 to 2013. The park, which was established in 1974, covers an area of 392 km^2 and the Amboseli ecosystem extends over a much wider area of approximately 3500 km^2 (Western, 1975). Using robust IPM discussed in chapter 3, we calculated $\lambda =1.031226$; therefore, the asymptotic population growth rate of the EB family is 3.1226% yearly. If we had incomplete information concerning reproduction, there are steps we would take. If this occurred, we would set up all possible combinations who may give birth in way where total number of offspring and possible candidates (mother) are preserved. We performed 200 permutations (possible reproduction paths), 46 out of these 200 permutations the asymptotic population growth rate (λ) converged to 1.031226, where we set up the restrictions within a 2

years gap to take into account the elephant gestation period which means if an individual gives birth during the current year she will not be considered as a possible candidate (mother) for the next two years.

If we use variables like a drought index as a proxy for size, the elephant family will have a common 'size' as in the above example and the asymptotic population growth rate converges to 1.031226. Now consider an elephant family such that size varies individual to individual. Assuming all variables remain same, we can estimate the distribution of a female's foot length (size) from the age of the female using the asymptotic equation by Lee and Moss (1995) as Females Foot length $= 42.05 - 24.32 * \exp(-0.1090 * \text{ age })$. We generated this artificial data which represents the actual data as presented in Figure 5.1. Using the robust IPM discussed in Chapter 3, we get the vital rate in Figure 5.6 and asymptotic population growth rate $\lambda = 1.061468$. In the situation where incomplete information of reproduction probability using permuted integral projection model (PIPM), we can recover asymptotic population growth rate (λ).

Out of 100 iterations 81 provide a complete path and follows the distributions in Figure 5.7 with mean 1.059546 which is very close to the true asymptotic population growth rate 1.061468. When the number of iterations is increased to 10000, we find 7973 complete paths and the resulting growth rate follows the distributions in Figure 5.7 (c) and (d) with mean 1.060573, which is very close to the true asymptotic population growth rate 1.061468.
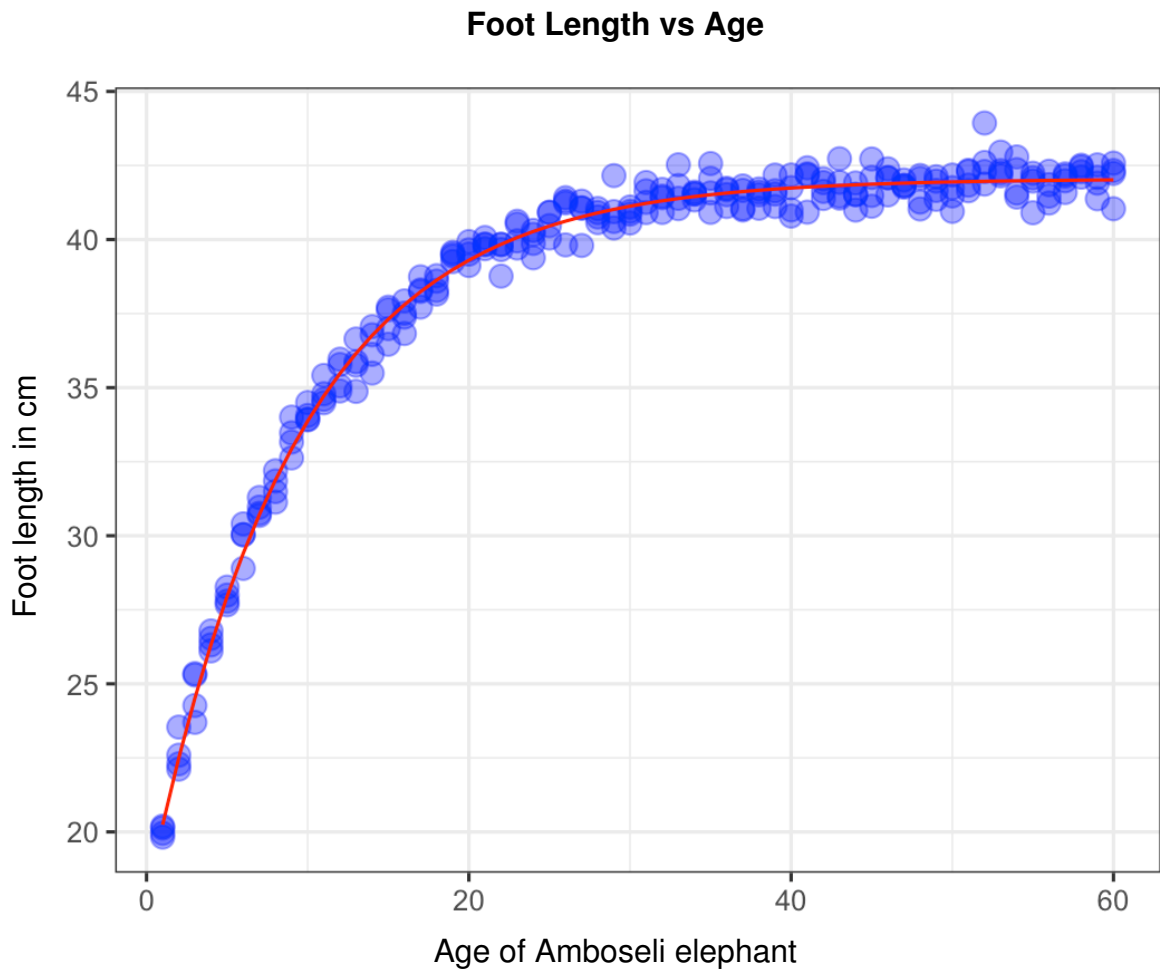
Figure 5.1: Foot length of Amboseli elephant. The red curve is the mean (μ_{length}) of females foot length. Moss *et al.* (2001) reported the relationship between age and foot length as, $\mu_{length} = 42.05 - 24.32 * \exp(-0.1090 * \text{ age })$.
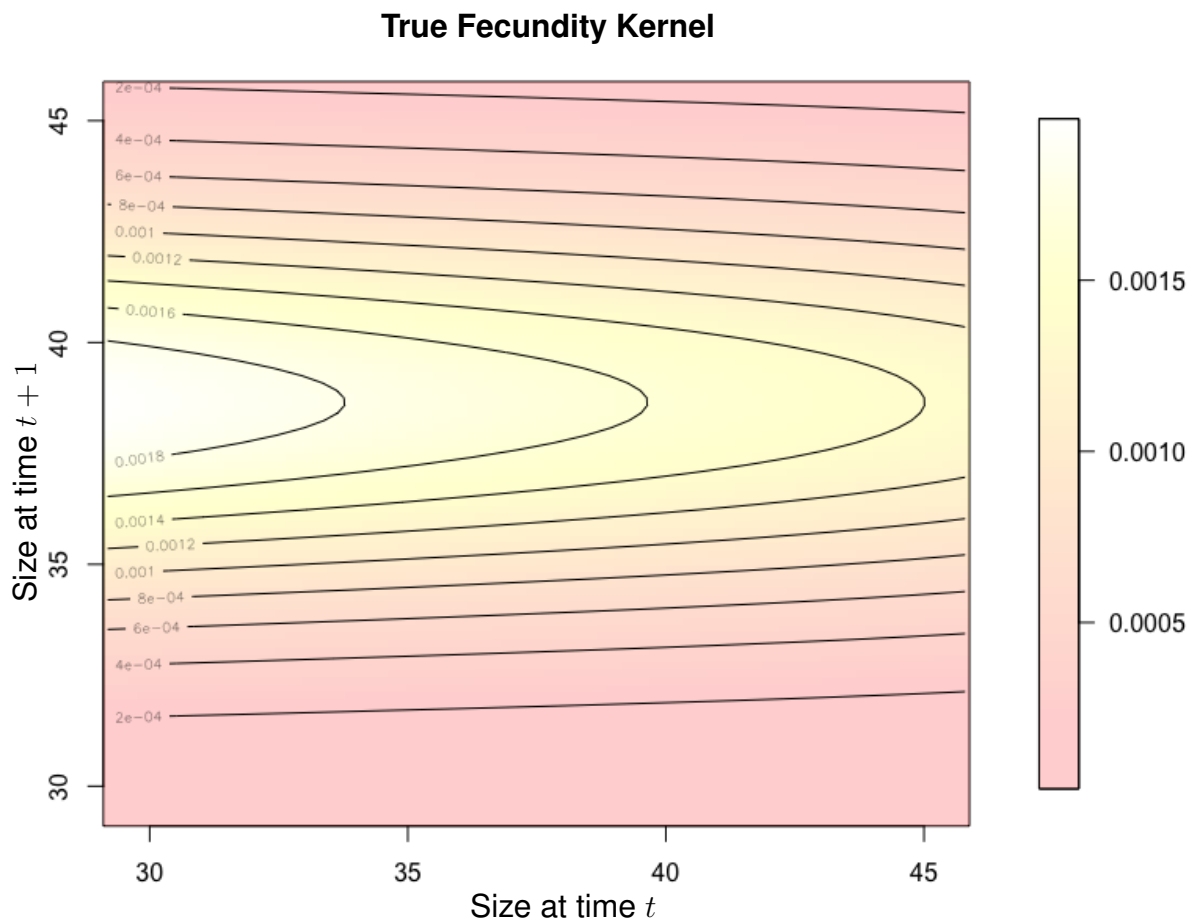
Figure 5.2: True fecundity kernel, $F(z_{t+1}, z_t)$ using robust IPM. $F(z_{t+1}, z_t)$ revealed the higher elevation on the medium foot size side indicates the elephant produce more offspring during the foot size transition approximately from 29 to 42 cm.

Table 5.3 shows the true scenarios of permuted distribution of $50{\times}100$ iterations. First rows include all individuals of EB family where $\hat{\lambda}$ is the estimated asymptotic population growth rate assuming incomplete information of reproduction probability where we used restricted permutation discussed in section 5.4 to recover the reproduction probability. On the other hand λ is the true asymptotic population growth calculated using robust IPM method using the true reproduction probability of EB family.

The rest of the rows show the result of the cumulative member/s removal to compare the individual contribution and span (how long an individual is considered a possible candidate) effect on λ. For example second row shows the results excluding the mother Europa and third row excluding mothers Europa and Eudora. Removing the mother Ebony from the group increases the asymptotic population growth rate by 0.45% because Ebony did not have any offsprings in reality. Adding more columns with different span decreases the success rate of complete path and predicts λ with higher accuracy.
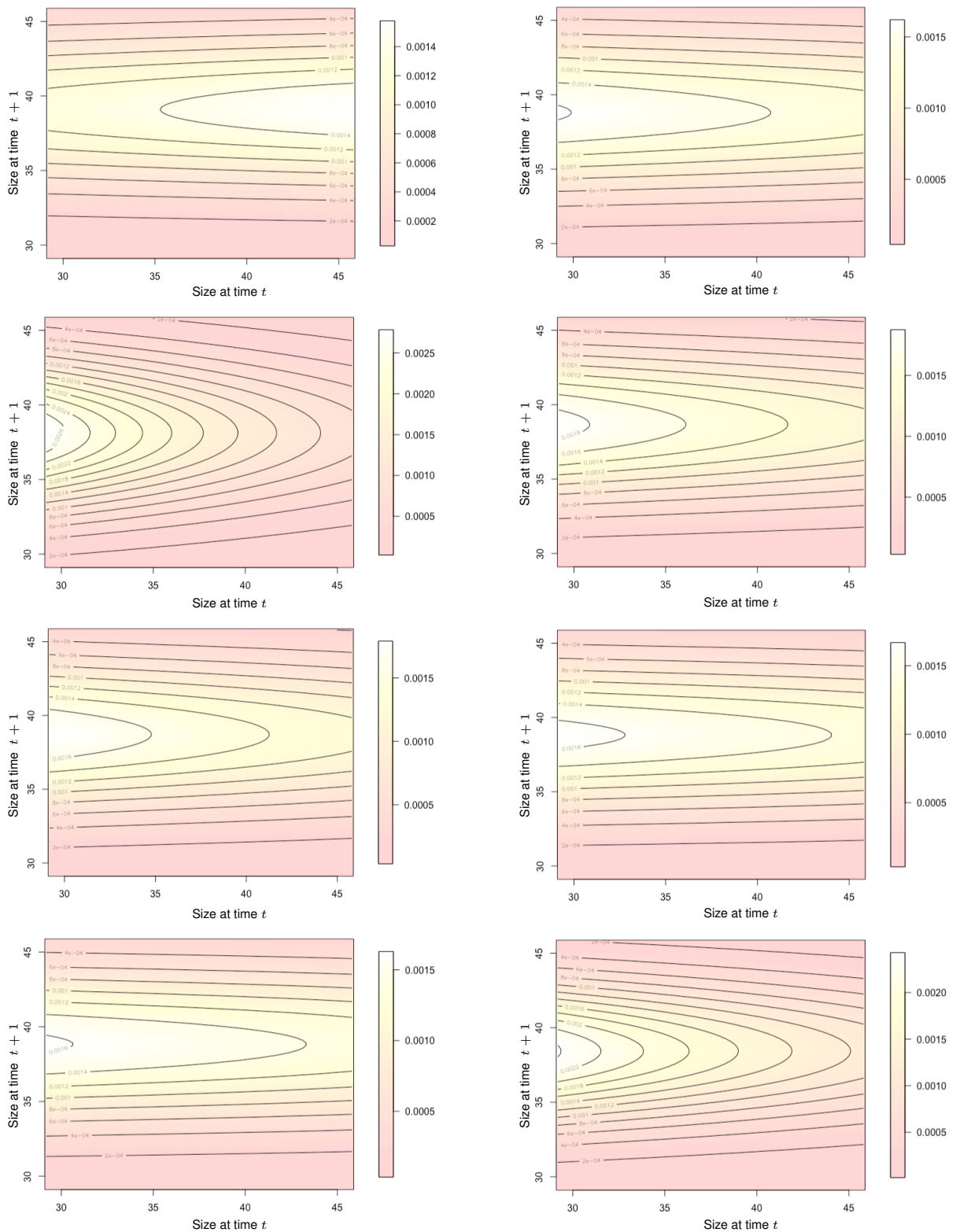
Figure 5.3: Permuted fecundity kernels, $F(z_{t+1}, z_t)$ using PIPM. 168 complete paths were received that fulfilled all restrictions out of 200 iterations. Using these permuted reproduction information, asymptotic population growth rates were calculated. First 8 corresponding fecundity kernels reported here. Similar size transition patterns of true fecundity kernel were observed as well.

A Successful Permuted Reproduction Matrix
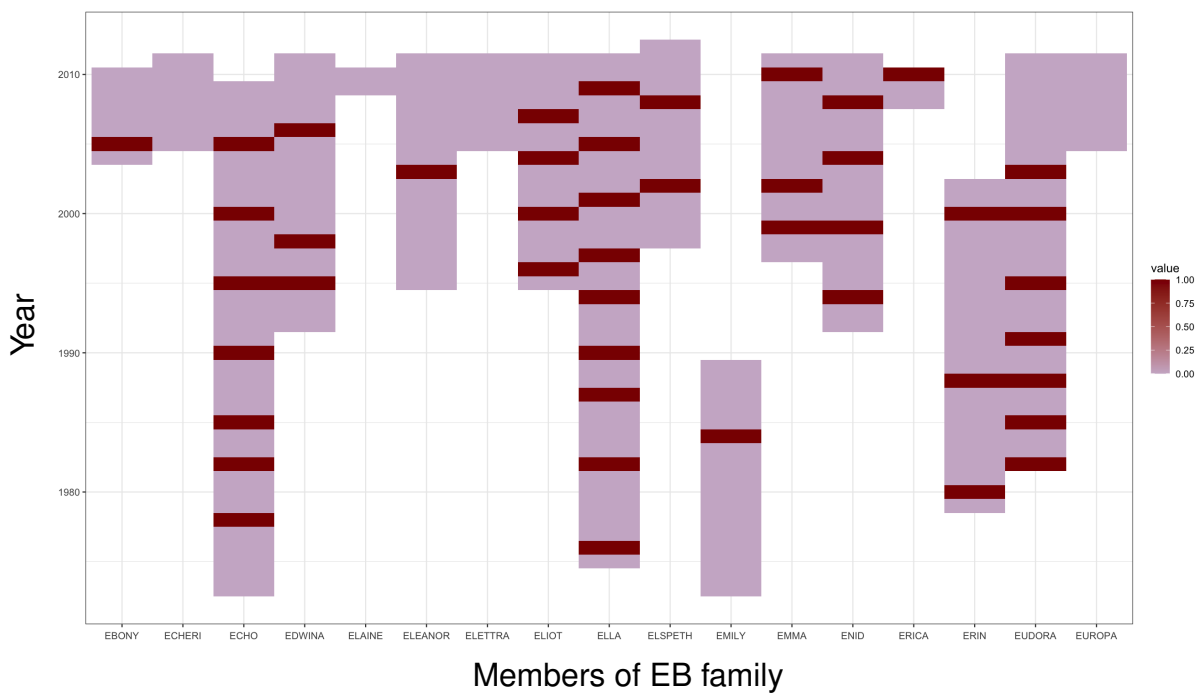
Figure 5.4: A successful iteration. If an individual gives birth within the current year, then this individual would be considered as the possible candidate for the next two or three years depending on her gestation period. The thistled column chunk is the time span of possible candidate who may give birth and there should not be any gap (white color). Darkened area within the thistled column chunk represent the year of reproduction. Row sum (total number of babies reproduced in the current year) of the successful iteration should be constant.

A Failed Permuted Reproduction Matrix
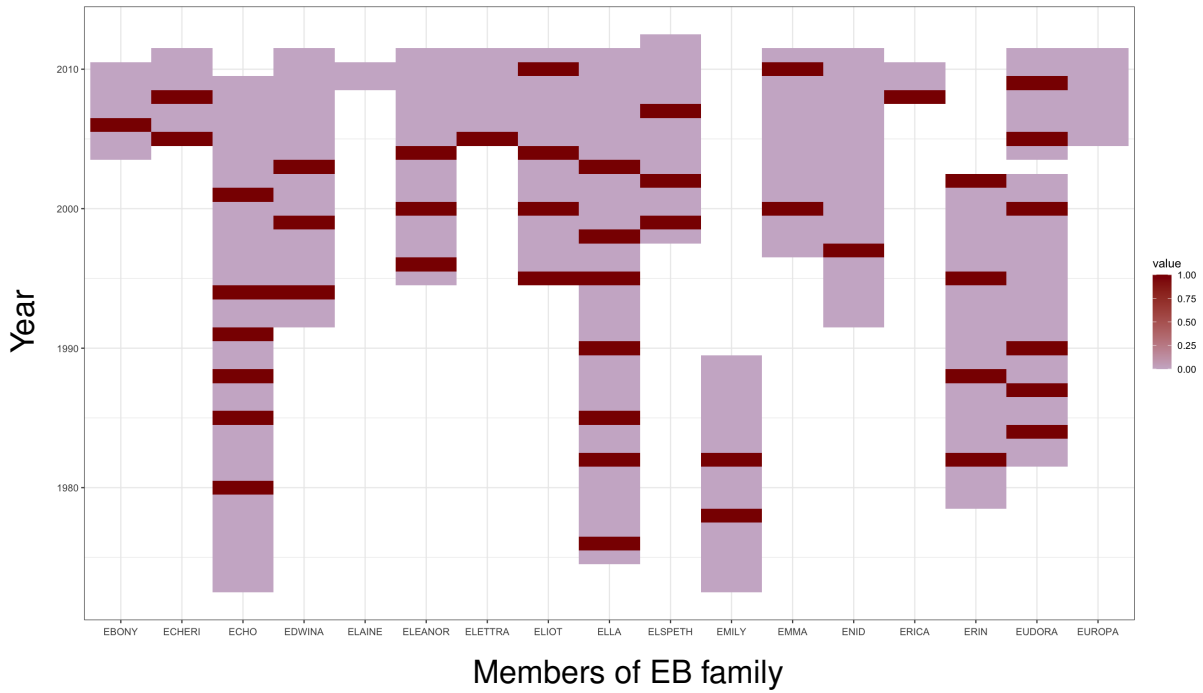
Figure 5.5: A iteration that fails to satisfy the constraints. In the current year (row), we observe a gap in the column of "EUDORA" means there is not enough possible candidates to distribute fixed number of babies randomly.

Table 5.3: True Scenarios in IPM Permuted Distribution.

Iteration	Cumulative Removal ↓	Success (mean)	Sd	Possible Candidate	Offspring	Span	$\hat{\lambda}$	Sd	True λ
50× 100	——	27.88	3.2428	18	46	282	1.0612	0.0042	1.0615
50× 100	Europa	26.8	3.9383	17	45	275	1.0611	0.0040	1.0620
50× 100	Eudora	34.4	4.9693	16	41	245	1.062	0.0045	1.0607
50× 100	Erin	30.14	4.1748	15	36	221	1.0627	0.0046	1.0673
50× 100	Erica	29.34	4.4245	14	35	218	1.061	0.0040	1.0648
50× 100	Enid	72.44	4.4454	13	31	198	1.0589	0.0043	1.0575
50× 100	Emma	65.74	5.6489	12	29	183	1.0599	0.0041	1.0593
50× 100	Emily	66	4.6247	11	26	166	1.0669	0.0051	1.0666
50× 100	Elspeth	100	0	10	25	151	1.0709	0.0048	1.0661
50× 100	Elliot	100	0	9	22	135	1.0669	0.0049	1.0653
50× 100	Elettra	100	0	8	21	128	1.0669	0.0047	1.0660
50× 100	Eleanor	100	0	7	19	111	1.0702	0.0048	1.0677
50× 100	Elaine	100	0	6	19	109	1.0753	0.0053	1.0823
50× 100	Edwina	100	0	5	15	89	1.0649	0.0032	1.0690
50× 100	Echeri	100	0	4	14	82	1.0718	0.0060	1.0688
50× 100	Ebony	100	0	3	14	75	1.0776	0.0046	1.0733

Survival vs Foot Length

Foot Length Transition

Reproduction vs Foot Length

Offspring Size Distribution
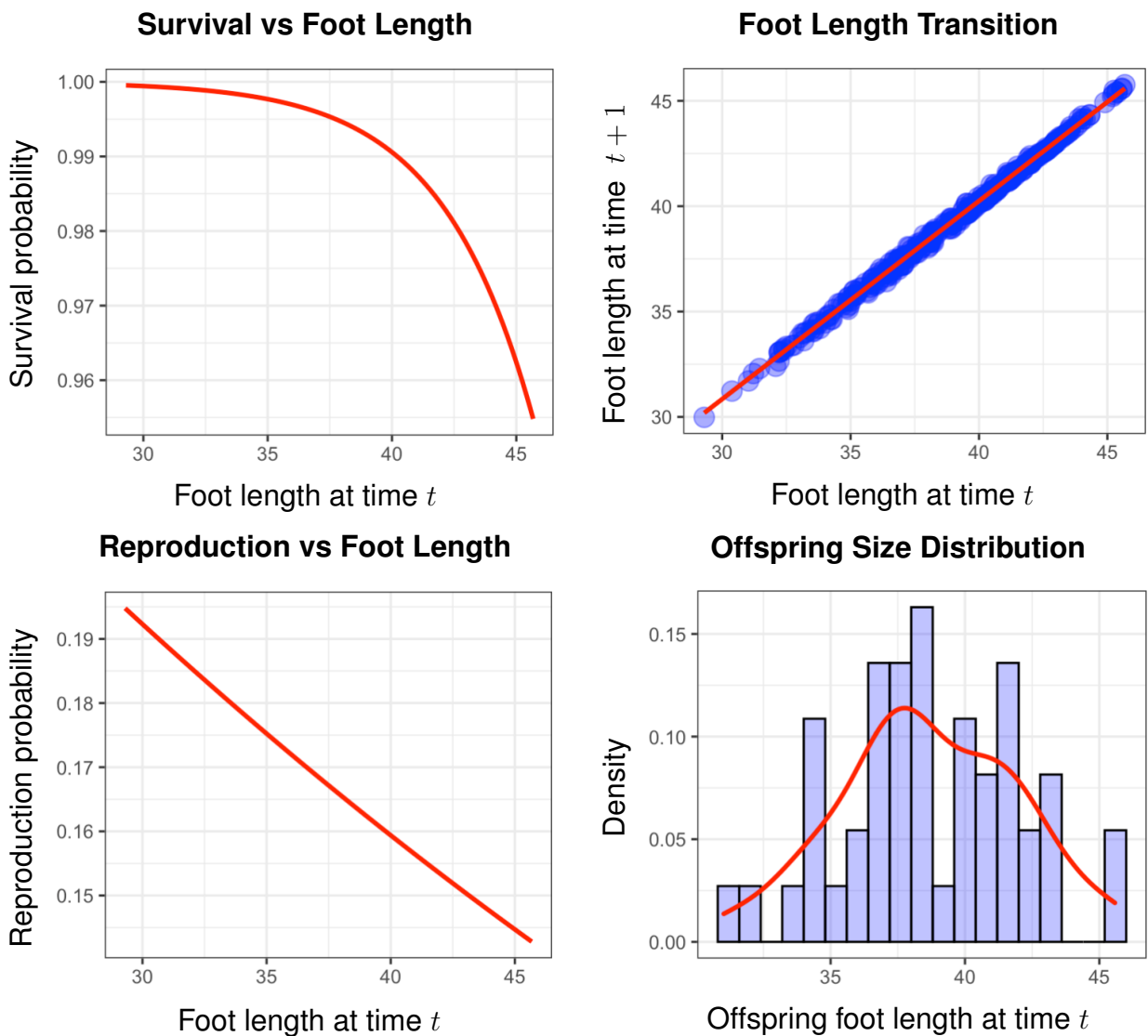
Figure 5.6: Estimating the model parameters of EB family. (a) Scatter plot of survival against foot length, red line is logistic regression estimation of the survival probability. (b) Scatter plot of the foot length ,z in current year, t against to the foot length, $z1$ in the next year $t+1$. Redline is M-estimation of foot length in time, t. (c) Scatter plot of reproduction against foot length, red line is the logistic regression estimation of the reproduction probability. (d) Size distribution of recruits.
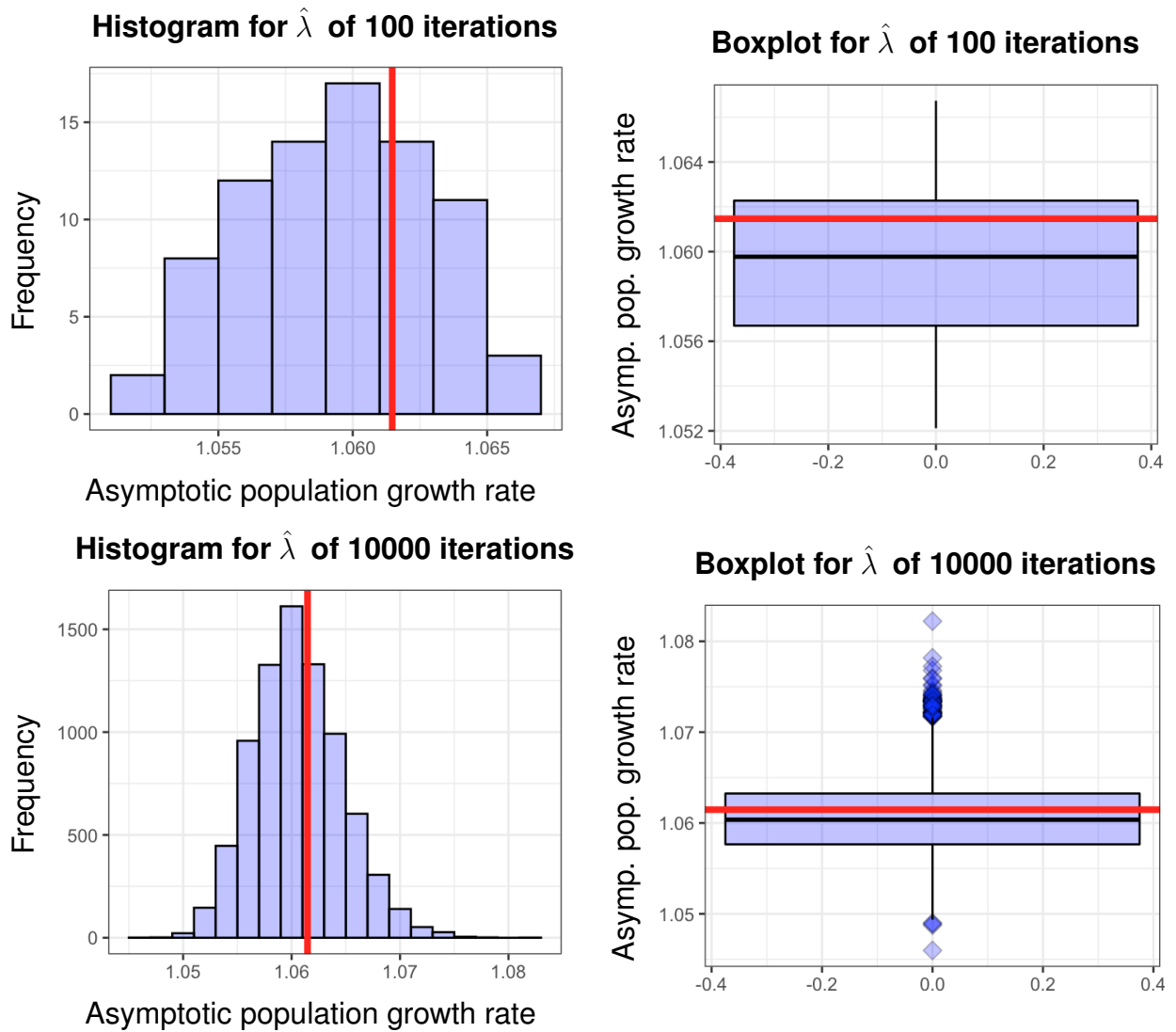
Figure 5.7: **Top panel:** Distribution of asymptotic population growth rate, λ for 100 iterations. **Bottom panel:** Distribution of asymptotic population growth rate, λ for 10000 iterations. Solid red lines on the graphs shows true λ.

Table 5.4: Hypothetical Scenarios in IPM Permuted Distribution.

Name	Iteration	Success (mean)	Sd	Possible Candidate	Offspring	Span	$\hat{\lambda}$	Sd	True λ
EUROPA	50× 100	80.12	3.474	16	46	275	1.064	0.004	–
EUDORA	50× 100	39.26	5.756	15	46	245	1.073	0.004	–
ERICA	50× 100	39.56	3.737	14	46	242	1.076	0.004	–
ENID	50× 100	38.68	4.757	13	46	222	1.081	0.002	–
EMMA	50× 100	37.6	4.953	12	46	207	1.086	0.003	–
EMILY	50× 100	0	0	11	46	190	–	–	–

Table 5.5: Hypothetical Scenarios in IPM Permuted Distribution.

Name	Iteration	Success (mean)	Sd	Possible Candidate	Offspring	Span	$\hat{\lambda}$	Sd	True λ
EUROPA	50× 100	80.12	3.474	16	46	275	1.064	0.004	–
EUDORA	50× 100	39.26	5.756	15	46	245	1.073	0.004	–
ERICA	50× 100	39.56	3.737	14	46	242	1.076	0.004	–
ENID	50× 100	38.68	4.757	13	46	222	1.081	0.002	–
EMMA	50× 100	37.6	4.953	12	46	207	1.086	0.003	–
ELSPETH	50× 100	0	0	11	46	192	–	–	–

Table 5.6: Hypothetical Scenarios in IPM Permuted Distribution.

Name	Iteration	Success (mean)	Sd	Possible Candidate	Offspring	Span	$\hat{\lambda}$	Sd	True λ
EUROPA	50× 100	80.12	3.474	16	46	275	1.064	0.004	—
EUDORA	50× 100	39.26	5.756	15	46	245	1.073	0.004	—
ERICA	50× 100	39.56	3.737	14	46	242	1.076	0.004	—
ENID	50× 100	38.68	4.757	13	46	222	1.081	0.002	—
EMMA	50× 100	37.6	4.953	12	46	207	1.086	0.003	—
ECHERI	50× 100	37.7	4.683	11	46	200	1.092	0.002	—
EBONY	50× 100	38.56	5.61	10	46	193	1.095	0.003	—
EBONY	100 × 100	38.64	4.994	10	46	193	1.095	0.003	—
ELAINE	50× 100	38.48	3.845	9	46	191	1.094	0.003	—
ELEANOR	50× 100	0	—	8	46	174	—	—	—

Table 5.7: Hypothetical Scenarios in IPM Permuted Distribution (Decreasing Offspring while PC constant).

Name	Iteration	Success (mean)	Sd	Possible Candidate	Offspring	Span	$\hat{\lambda}$	Sd	True λ
EUROPA	50× 100	80.3	4.047	17	45	282	1.061	0.004	—
EUDORA	50× 100	89.7	2.501	17	41	282	1.053	0.005	—
ERIN	50× 100	90	3.232	17	36	282	1.043	0.005	—
ERICA	50× 100	90.14	3.003	17	35	282	1.041	0.005	—
ENID	50× 100	90	3.123	17	31	282	1.035	0.005	—
EMMA	50× 100	89.78	3.334	17	29	282	1.031	0.005	—
EMILY	50× 100	90.28	2.718	17	26	282	1.025	0.005	—
ELLA	50× 100	100	0	17	19	282	1.013	0.006	—

Table 5.8: Hypothetical Scenarios in IPM Permuted Distribution (Decreasing Offspring while PC constant).

Name	Iteration	Success (mean)	Sd	Possible Candidate	Offspring	Span	$\hat{\lambda}$	Sd	True λ
EUROPA	50× 100	80.3	4.047	17	45	282	1.061	0.004	—
EUDORA	50× 100	89.7	2.501	17	41	282	1.053	0.005	—
ERIN	50× 100	90	3.232	17	36	282	1.043	0.005	—
ERICA	50× 100	90.14	3.003	17	35	282	1.041	0.005	—
ENID	50× 100	90	3.123	17	31	282	1.035	0.005	—
EMMA	50× 100	89.78	3.334	17	29	282	1.031	0.005	—
EMILY	50× 100	90.28	2.718	17	26	282	1.025	0.005	—
ELETTRA	50× 100	89.9	2.998	17	25	282	1.024	0.005	—
ELEANOR	50× 100	90.1	2.823	17	23	282	1.021	0.006	—
EDWINA	50× 100	90.2	2.347	17	19	282	1.015	0.006	—
ECHERI	50× 100	90.32	3.107	17	18	282	1.013	0.006	—
ELSPETH	50× 100	89.84	3.31	17	16	282	1.01	0.006	—
ELSPETH	100× 100	89.81	2.987	17	16	282	1.01	0.006	—

Tables 5.4 through 5.8 show the results of hypothetical scenarios of permuted distribution of 50×100 iterations. We explore a hypothetical group of elephants that represents the EB family. $\hat{\lambda}$ is the estimated asymptotic population growth rate assuming incomplete information of reproduction probability where we used the restricted permutation scheme, as discussed in section 5.4, to estimate the reproduction probability. In this situation, the true asymptotic population growth rate λ remains unknown.

The tables show the cumulative member(s) removal in order to observe the individual contribution, number of offspring, and the span (how long an individual is considered a possible candidate) all have an effect on the estimated λ. For example, in Table 5.7 the first row shows the results excluding Europa and the second row excluding Europa and Eudora. Tables 5.4 through 5.6 show that adding more column with different spans increases the success rate of complete path and decreases the asymptotic population growth rate $\hat{\lambda}$. This is a reasonable observation as increasing the number of candidates without changing the total number of offspring will decrease the population growth rate. On the other hand, as expected, decreasing the number of offspring decreases the asymptotic population growth rate, assuming the total possible candidates remain constant as showed in Tables 5.7 and 5.8. Again, decreasing the number of offspring increases the success rate of complete path in Tables 5.7 and 5.8 due to more room being available for offspring to permute.

On account of the presence of long duration (more than 30 years) as a possible candidate, removing Eudora from the group increases the asymptotic population growth rate by 0.9 %, in the hypothetical scenarios where the number of offspring remains constant as presented in Table 5.6.

All of these patterns follow our expectations. It is important to note that using the permutation scheme we are now able to quantify the uncertainties introduced due to incomplete knowledge of the identity of birth givers or even when we only have proxy knowledge of size (eg. age-footsize relationship). This is extremely useful for understanding population dynamics of species that are not living in captivity and we only have periodic census information.

5.6 Bayesian MCMC Simulation on elephant foot length and PIPM

In this section, we will demonstrate that the measurement of uncertainty in such cases can be accomplished using Bayesian Markov Chain Monte Carlo sampling schemes. For instance, for a given age, the distribution of size can be sampled to provide a size distribution without knowing individual sizes. This uncertainty is propagated to the estimate of the population growth rate. So, we expect wider confidence intervals in comparison to cases where we have complete information. The Bayesian version of the linear model as discussed in Figure 5.6 (top right) can be represented as:

$$\text{Likelihood: } z_{t+1} \sim N(\mu_z, \sigma_z^2)$$

$$\text{Linear model: } \mu_z = \beta_1 + \beta_2 * z_t$$

$$\text{Prior: } \beta_1 \sim N(0, 1/100000)$$

$$\text{Prior: } \beta_2 \sim N(0, 1/100000)$$

$$\text{Precision term: } \frac{1}{\sigma_z^2} = \Gamma(0.001, 0.001)$$

We ignored first 1000 iterations (burn in period) to update the model, thereupon, stored 150,000 iterations for posterior estimation of coefficients. Using these coefficients asymptotic population growth rates were estimated as reported in Figure 5.8.
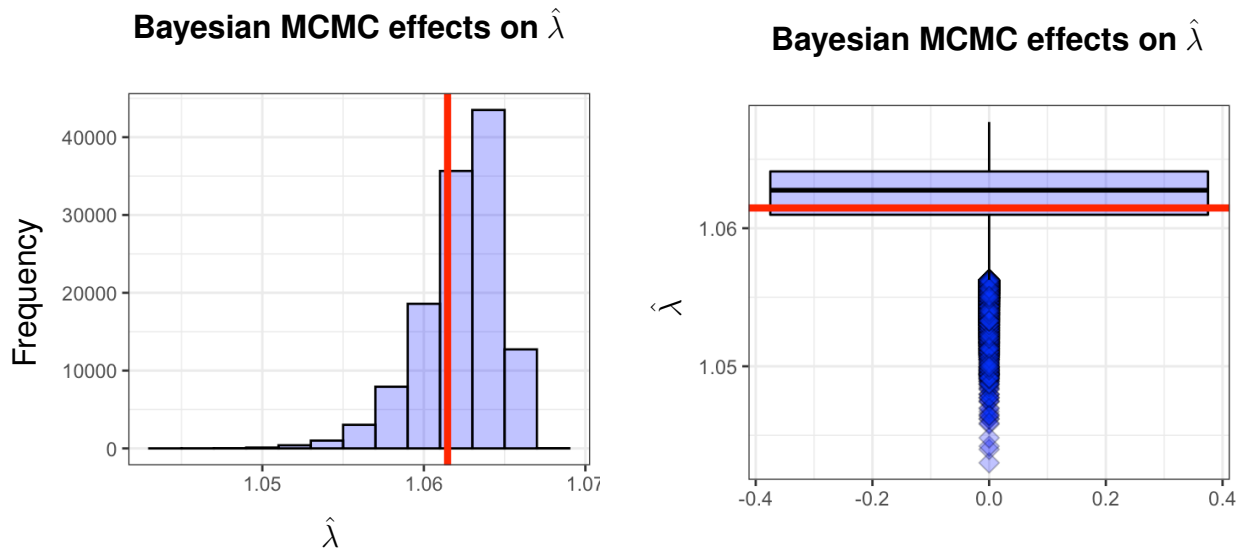
Figure 5.8: The effect of Monte Carlo simulation on true asymptotic population growth rate, λ in the situation of a complete reproduction information. Solid red lines on the graphs represent true, λ.

To observe the effect of Monte Carlo simulation on permuted integral projection model, asymptotic population growth rates were calculated using the algorithm as discussed in Appendix. Out of 50 iterations in PIPM, 42 successful complete paths were received. Thereupon, in each successful iteration, uncertainty of the asymptotic population growth rate, λ was calculated using 3000 Monte Carlo simulations. The first 12 outputs out of 42 successful iterations were reported in Figure 5.9 and total combined asymptotic population growth rates were presented in Figure 5.10. By comparing the Figures 5.8 and 5.10, we ascertained that the estimation of asymptotic population growth rates, λ using permuted IPM have higher uncertainty than the Bayesian approach.
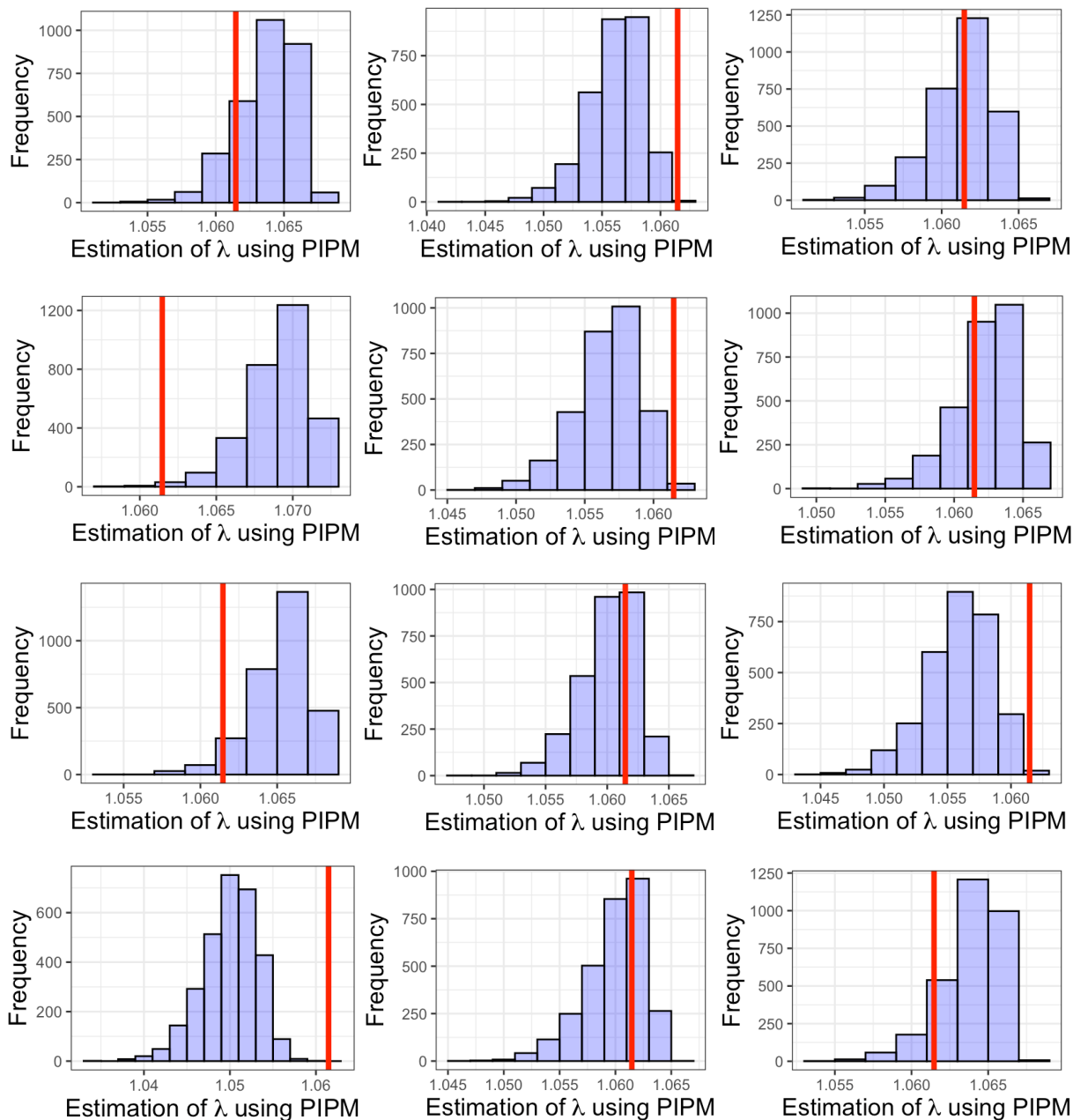
Figure 5.9: The effect of Monte Carlo simulation on PIPM. . Out of 50 iterations in PIPM, 42 successful complete paths were received. Thereupon, in each successful iteration, uncertainty of the asymptotic population growth rate, λ was calculated using Monte Carlo simulation. We presented the first 12 outputs out of 42 successful iterations.

Permuted Bayesian MCMC effects on $\hat{\lambda}$

Permuted Bayesian MCMC effects on $\hat{\lambda}$
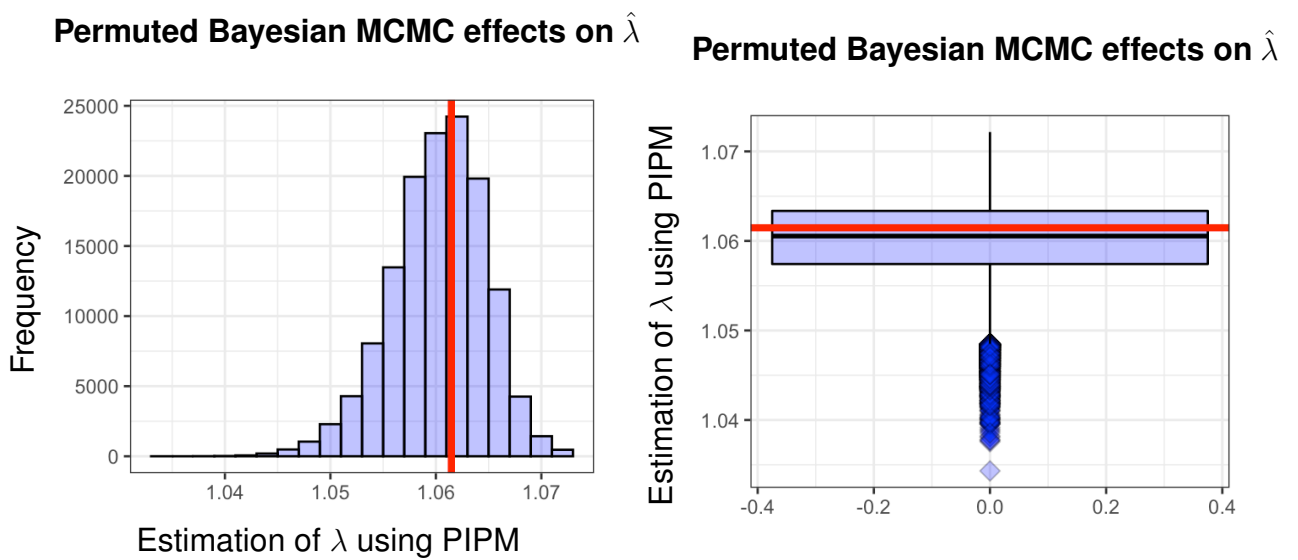
Figure 5.10: Total combined population growth rate, λ. 150,000 were used permuted integral projection model The solid red lines typify the true λ.

Chapter 6

Conclusion and Discussion

In this dissertation we presented a comprehensive set of tools to improve upon existing approaches for constructing an integral projection model of population dynamics.

Utilizing IPMs to bring continuous traits to population dynamics models has the potential to provide a powerful new approach to tackling an array of important questions in animal ecology and evolution. IPMs can be used to answer important questions regarding population dynamics in wildlife and plant ecology. In IPMs, regression is an essential component in which we project the current population to a singular step forward in time. We have shown empirically that standard implementations of integral projection models will generally severely underestimate the population growth rate in presence of outliers in size variables or mixed-size populations. In the presence of highly influential observations, LS based IPM estimation poorly estimated population growth rates as reported in Figure 3.8, while we get very satisfactory results using M-estimation-based IPM. Thus, we recommend using M-estimation in the cases where the population being modeled has high size variability. This has potential to be useful in modeling animal and plant populations when the environment is changing drastically, such as when a sub-population has larger or smaller than expected sizes.

In Chapter 5, an extension of the IPM was proposed in the circumstances of incomplete knowledge of reproduction probability. The benefits of the proposed model were demonstrated by exploring a real data set. The true reproductive information is presented in Figure 6.1 (left) and a successful iteration of reproductive information (right) using restricted permutation (discussed in Chapter 5). The question is raised: Is permuted reproductive matrix a good representation of true reproductive matrix? To answer this question, we estimated asymptotic
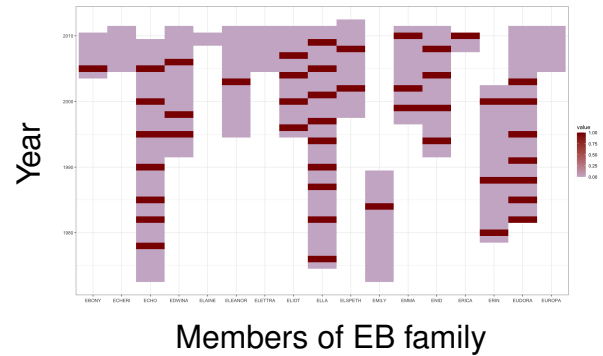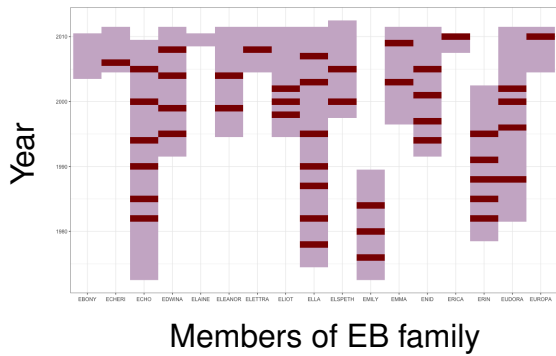
| A true reproductive information of Amboseli's elephants (who is giving birth). | A successful iteration of reproductive information using restricted permutation. |

Figure 6.1: Individuals who reproduce offspring are represented by the darkened (repr=1) area, individuals who fail to reproduce offspring are represented by the thistled area (repr=0), and individuals who no longer survive (surv=0) are represented by the no color zone.

population growth rates using true and permuted reproductive information. It was observed that the asymptotic population growth rate λ from the permuted IPMs converge to the true λ in the situation of common in size variables.

Furthermore, artificial individual size variable were generated to simulate the true size variable of the Amboseli elephant. The asymptotic population growth rate was estimated using a robust IPM, which is assumed to be the true population growth rate. Thereupon, the asymptotic population growth rate was calculated using the proposed permuted integral projection model assuming incomplete information of reproduction probability. It was observed that the asymptotic population growth rate converged to the true λ for $n > 500$.

Again, we presented the true fecundity kernel in Figure 5.1 (using true reproduction probability of Amboseli elephant). Additionally, fecundity kernels were estimated using proposed permuted IPM. Out of 200 iterations, we received 168 complete paths and the corresponding fecundity kernels, as reported in Figure 5.3 (presented only first 8 fecundity kernels due to the space limitation). A similar size transition pattern of true fecundity kernel was observed as well.

Therefore, with the estimated parameters, the permuted integral projection model can be used for prediction of future asymptotic population growth rate λ in the situation of incomplete information of reproduction probability. Collecting a long-term data set is very expensive and

in some scenarios almost impossible. Our proposed permuted IPM significantly reduces the cost time for collecting the long - term data on species that are difficult to monitor.

According to Moss, the African elephant faces two serious threats to its future:

(i) over-abuse for its ivory, meat, and skins and

(ii) loss of appropriate natural surroundings.

IPMs carrying continuous attributes to population dynamic models have the potential to implement all inquiries related to animal and wildlife ecology. Again, wildlife ecology is integral to the ecosystem. However, human prosperity and progress towards unceasing development are fundamentally gambled upon the planet's ecosystem. In order to settle on these sorts of choices for elephants, both all through secured zones, it is fundamental to comprehend their essential reproductive and survival parameters. It is anticipated that this dissertation will be of significance for future elephant data collection, modeling the effect of drought, elephant size dependent IPM, and manifesting IPM in the situation of incomplete information of reproduction probability. All of these factors will help with the point of successful, profitable, and sustainable future development.

Future work

There are a multitude of questions regarding the statistical estimation approaches for IPMs. Here are a few items we would like to explore in future work:

- The Integral Projection Model can be developed to the elephant population dynamics and the interaction with the environmental situations. So in future we will develop the robust IPMs for continuously structured population with finite size, including both demographic and environmental stochasticity in the dynamics.

- We will compare our results with those from State-Space Models. Confusingly these are sometimes known as integrated population models (IPMs) in the ecological modeling community although they are basically stochastic versions of MPMs with input variables.

- We will also include foot size/ shoulder size (PC Lee, CJ Moss, 1995) as a predictor in our model. We used a Bayesian MCMC approach to incorporate such information. It is of interest to study how the subgroup characteristics affect on population growth rate.

- A theoretical understanding of the permuted IPM (PIPM) is combinatorial in nature. Although from our simulations we can determine the proportion of complete paths based on the number of candidate mothers, we would like to know if there is a theoretical solution to this.

- A theoretical study of the Bayesian-PIPM is of interest as well. Our simulation shows that this can reduce the uncertainty in the estimation of vital rates than just using Bayesian MCMC alone ot PIPM alone.

References

[1] Rees M, Childs D and Ellner S. 2014. Building integral projection models: a user guide *Journal of Animal Ecologoy*, **83**, 528-545.

[2] Merow C, Dahlgren J , Metcalf J, Childs D and McMahon S. 2014. Advancing population ecology with integral projection models: a practical guide *Methods in Ecology and Evolution*, **5**, 99-110.

[3] Archie, E. A., Hollister-Smith, J. A., Poole, J. H., Lee, P. C., Moss, C. J., Maldonado, J. E., Fleischer, R. C. & Alberts, S. C. 2007. Behavioural inbreeding avoidance in wild African elephants. *Molecular Ecology*, **16**, 4138-4148.

[4] Vindenes, Y., Engen, S. & Saether, B.-E. (2011). Integral projection models for finite populations in a stochastic environment. *Ecology*, 92,1146–1156.

[5] Moss, Cynthia J. 2001. The demography of an African elephant (Loxodonta africana) population in Amboseli, Kenya *J. Zool., Lond* , **255** ,145-156.

[6] Easterling M, Ellner S and Dixon P. 2000. Size-Specific Sensitivity: Applying a New Structured Population Model *Ecology*, **3**, 694-708.

[7] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining Introduction to Linear Regression Analysis.

[8] Wallace, K., Leslie, A. & Coulson, T. (2013). Re-evaluating the effect of harvesting regimes on Nile crocodiles using an integral projection model. *Journal of Animal Ecology*, 82, 155–165.

[9] John Fox and Sanford Weisberg Applied regression analysis, linear models, and related models. Sage publications, Inc, 1997.

[10] Childs, D.Z, Rees, M., Rose, K.E., Grubb, P.J. & Ellner, S.P. 2003. Evolution of size-dependent flowering in a variable environment: construction and analysis of a stochastic integral projection model. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* **270**, 1829-1838.

[11] Rose, K.E., Louda, S.M. & Rees, M. (2005). Demographic and evolutionary impacts of native and invasive insect herbivores on Cirsium canescens. *Ecology*, 86, p453-465.

[12] Archie, E. A., Moss, C. J. & Alberts, S. C. 2003. Characterization of tetranucleotide microsatellite loci in the African savannah elephant (Loxodonta africana africana). *Molecular Ecology* Notes **3**, 244-246.

[13] G. Li, Robust regression In Exploring Data Tables, Trends, and Shapes, Wiley, 1985.

[14] Hollister-Smith, J. A., Poole, J. H., Archie, E. A., Vance, E. A., Georgiadis, N. J., Moss, C. J. & Alberts, S. C. 2007. Age, musth and paternity success in wild male African elephants, Loxodonta africana. *Animal Behaviour* **74**, 287-296.

[15] Lee, P. C. & Moss, C. J. (ed.) 1999. The social context for learning and behavioral development among wild African elephants. Mammalian social learning: comparative and ecological perspectives. London, UK: Cambridge University Press.

[16] F.J. Heather, D.Z. Childs, A.M Darnaude, J.L. Blanchard Using an integral projection model to assess the effect of temperature on the growth of gilthead seabream Sparus aurata. *Plos one*.

[17] Nicole, F., Dahlgren, J.P., Vivat, A., Till-Bottraud, I. & Ehrl"en, J. (2011). Interdependent effects of habitat quality and climate on population growth of an endangered plant. *Journal of Ecology*, 99, 1211–1218.

[18] Lee, P. C. & Moss, C. J. 1999. The social context for learning and behavioural development among wild African elephants. *In Symposia of the Zoological Society of London; Mammalian social learning: Comparative and ecological perspectives* pp. 102-125.

[19] Lee, P. C. & Moss, C. J. 1986. Early maternal investment in male and female African elephant calves. *Behavioral Ecology and Sociobiology* **18**, 353-361.

[20] McComb, K., Moss, C., Durant, S. M., Baker, L. & Sayialel, S. 2001. Matriarchs as repositories of social knowledge in African elephants. *Science* **292**, 491-494.

[21] Ramula, S., Rees, M. & Buckley, Y.M. Integral projection models perform better for small demographic data sets than matrix population models: a case study of two perennial herbs *Journal of Applied Ecology,* **46**, 1048-1053

[22] Bali, J. L., Boente, G., Tyler, D. E., Wang, J.-L., et al. 2011. Robust functional principal components: A projection-pursuit approach. *Annals of Statistics,* **39**(6):2852-2882.

[23] Archie, E. A., Maldonado, J. E., Hollister-Smith, J. A., Poole, J. H., Moss, C. J., Fleischer, R. C. & Alberts, S. C. 2008. Fine-scale population genetic structure in a fission-fusion society. *Molecular Ecology* **17**, 2666-2679.

[24] Moss, C., Croze, H. & Lee, P. C. (2011). The Amboseli Elephants: a long-term perspective on a long-lived species. : *University of Chicago Press*.

[25] Western, D. 1982. Amboseli. *Swara* 5(4):8-14.

[26] Chevin, L.M. (2015). Evolution of adult size depends on genetic variance in growth trajectories: a comment on analyses of evolutionary dynamics using integral projection models. *Ecology Evolution*. 6: 981-986.

[27] Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: 804 body size allometry. *Evolution* 33: 401-416.

[28] Yount, Lisa.1999. A Biographical Dictionary A to Z of Women in Science and Math. *New York: Facts on File, Inc.* pp. 156–158. ISBN 0816037973.

[29] Ozgul, A., Coulson, T., Reynolds, A., Cameron, T.C. & Benton, T.G. (2012). Population responses to perturbations: the importance of trait based analysis illustrated through a microcosm experiment. *The American Naturalist*, 179, 582–594.

[30] Western, D. 1973. *The structure, dynamics and changes of the Amboseli ecosystem.* Ph.D. thesis. Nairobi, Kenya: University of Nairobi.

[31] Western, D. 1982. Amboseli National Park: Enlisting landowners to conserve migratory wildlife. *Ambo* 11(5):302-308.

[32] Moss, C. 1977. The Amboseli elephants. *Wildlife News* 12(2):9-12.

[33] Xu K, Wikle CK, & Fox N.(2005). A kernel based spatio temporal dynamical model for nowcasting radar precipitation. *Journal of the American Statistical Association.* 100:1133–1144.

[34] Thouless, C., J. K., P. Omondi, P. K., and Douglas-Hamilton I. 2008. Report to Kenya Wildlife Service and Save the Elephants. Nairobi, Kenya. *The status of Kenya's elephants 1990-2002.*

[35] Zuidema PA, Jongejans E, Chien PD, During HJ, & Schieving F.(2010). Integral projection models for trees: A new parameterization method and a validation of model output. *Journal of Ecology.* 98:345–355.

[36] Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35 (1), 73-101.

[37] Almongy, H., & Almetwaly, E. (2018). Comparison between methods of robust estimation for reducing the effect of outliers. *The Egyptian Journal for Commercial Studies, Faculty of Commerce, Mansoura University, Egypt*, 4(41), 1-23.

[38] Dennis B, & Taper ML. (1994).Density dependence in time series observations of natural populations: estimation and testing. *Ecological Monographs.* 64:205–224.

[39] Almewally, E. M., & Almongy, H. M. (2018). Comparison Between M-estimation, S-estimation, And MM Estimation Methods of Robust Estimation with Application and Simulation. *International Journal of Mathematical Archive*, 9(11), 55-63.

[40] Meloun, M., Militký, J. (2011). Linear Regression Models *Statistical Data Analysis*.

[41] Mercier L, Mouillot D, Bruguier O, Vigliola L, & Darnaude A.M. (2012). Multi-element otolith fingerprints unravel sea-lagoon lifetime migrations of gilthead sea bream Sparus aurata. *Marine Ecology Progress Series*. 444:175-94.

[42] Dennis B, Desharnais R. A, Cushing J.M, & Costantino R.F. (1997) . Transitions in population dynamics: equilibria to periodic cycles to aperiodic cycles. *Journal of Animal Ecology*. 66:704–729.

[43] H. Wang, G. Li, & G. Jiang. (2007) Robust regression shrinkage and consistent variable selection through the lad-lasso. *Journal of Business and Economic Statistics,* 25(3):347–355.

[44] J. Jureckova. (1971). Nonparametric estimate of regression coefficients. *The Annals of Mathematical Statistics*, 1328–1338.

[45] West M. (1994). Outlier Models and Prior Distributions in Bayesian Linear Regression. *Journal of the Royal Statistica Society*. Series B. 46-3.

[46] Harding MC, & Hausman J. (2007). Using a Laplace approximation to estimate the random coefficients logit model by non-linear least squares. I*nternational Economic Review*. 48:1311–1328.

[47] Rebarber, R., Tenhumberg, B. & Townley, S. (2012). Global asymptotic stability of density-dependent integral population projection models. *Theoretical Population Biology*, 81, 81–87.

[48] Crowder, L.B., D.T. Crouse, S.S. Heppell & T.H. Martin. (1994). Predicting the impact of turtle excluder devices on loggerhead sea turtle populations. *Ecological Applications* 4:437–445.

[49] Klinkhamer, Meelis, de Jong & Weiner. (1992). On the analysis of size-dependent reproductive output in plants. *Functional Ecology*. 6, p308-316.

[50] Sletvold, N. (2005). Density-dependent growth and survival in a natural population of the facultative biennial. *Digitalis purpurea Journal of Ecology* 93, p727-736.

[51] Williams, J.L., Miller, T.E.X. & Ellner, S.P. (2012). Avoiding unintentional eviction from integral projection models. *Ecology*, 93,2008–2014.

Appendices

Appendix A

Monte Carlo Simulation on PIPM

A.1 Computational Algorithm

The outline of the algorithm pertaining to the Monte Carlo simulation (using JAGS) on permuted integral projection model is as follows:

Step 1: Generate a null vector of size m to save all asymptotic population growth rate, λ

Step 2: Run a for-loop for m iterations.

Step 3: Define a function $\text{Par}(a_z, b_z, \sigma_z, a_s, b_s, a_{r_p}, b_{r_p}, \mu_{rz}, \sigma_{rz})$ contains all parameters of vital rate.

Step 4: Using generalized linear regression calculate all vital rates parameters and update the function in step 1.

Step 5: Define growth function, $g(z_{t+1}, z_t) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left(\frac{(z_t - \mu_z)^2}{2\sigma_z^2}\right)$ that returns output as size, z_{t+1} at time $t+1$ given the size, z_t at time t as follows:

Step 6: we will determine the hierarchical form of the model.

Step 7: **likelihood of the data**:

- define a for-loop as index, i from 1 to number of observation.

- define the coefficients of the linear model as $b[1]$ and $b[2]$,

  foot size , at time, $t = z_t$, and foot size , at time, $t + 1 = z_{t+1}$.

- generate a normal distribution, $N[i]$ in each iteration with mean, $\mu[i] = b[1] + b[2] * z_t$, and standard deviation = precision.

- set i^{th} observation, $y[i] = N[i]$

Step 8: **Non-informative prior for the coefficients**:

- define a for-loop as index, j from 1 to 2 (total number of coefficients).

- generate normal distribution, with mean 0 and variance 1/1000000 [N.B. variance 1 million to make it non-informative prior. For precision of a normal distribution variance $= 1/1000000$] as : $b[j] = n(0, 0.000001)$

Step 9: **precision term** ($\frac{1}{\sigma^2} = \Gamma(\alpha, \beta)$):

- define prior sample size $= 5.0$ and prior guess for the variance $= 10$

- generate a gamma distribution, G with shape parameter $\underbrace{= 5.0/2.0}_{\text{divide by 2.0 to turn into gamma parameter}}$

  and rate parameter $= 5.0 * 10.0/2.0$

- set precision $= \underbrace{G}_{\text{Equivalent to an inverse gamma prior on the variance with same shape and scale parameters.}}$

- set variance $= \frac{1}{\text{precision}}$ (deterministic relation between precision and variance).

- $\sigma = \sqrt{\text{variance}}$

Step 10: put steps 2 through 4 in a string and save as mod1 (JAGS required)

Step 11: Input the data as a list.

Step 12: define a vector of parameters ($b's$ and σ).

Step 13: define a function as a list to assign initial values of the parameters as:

- $b[1]$ and $b[2]$ follow normal distributions with mean 0 and variance 100.

- precision follows a $\underbrace{\text{gamma distribution with shape 1 and rate 1}}_{\text{exponential distribution}}$.

Step 14 specify the JAGS model

Step 15 set burning period 1000 iterations.

Step 16 get the posterior simulation.

- plot individual size, z_t at time t versus the individual size, z_{t+1} at time $t+1$

- using m-estimation calculate regression parameters (a_z and b_z) and standard deviation σ_z

- define $\mu_{z_t} = a_z + b_z * z_t$, where z_t is the size at time t

- generate a normal distribution with mean μ_z and standard deviation σ_z

- input the size, z_t at time t using **steps 17 through 19**

- the growth function $g(z_{t+1}, z_t)$ returns size z_{t+1} at time $t+1$ provided the size, z_t at time t.

Step 17: Define survival function, $s(z_t) = \frac{1}{1+\exp(-\mu_{lgs})}$ which returns the survival probability of size, z_t at time, t as follows:

- define $\mu_{lgs} = a_s + b_s * z_t$, where z_t is the size at time t

- input the size, z_t at time t using **steps 17 through 19**

- return the survival probability of size z_t at time t

Step 18: Define reproduction function, $r_p(z_t) = \frac{1}{1+\exp(-\mu_{lgp})}$ which returns the reproduction probability of size, z_t at time, t as follows:

Step 19: Since we have incomplete reproduction information, by using steps 9 through 12, reproduction information can be permuted.

Step 20: Calculate the total offspring in a year

Step 21 : Find the total possible candidates who may give birth by restricting their age and survival information. If an individual gives birth within the current year, then this individual would be considered as possible the candidate for the next two or three years depending on their gestation period.

Step 22: Permute all offspring using the strategy discussed in section 5.4 to the possible candidates (mother).

Step 23: Save the reproduction probability corresponding to the individual as $r_p(z_t)$.

- where, $\mu_{lgp} = a_{r_p} + b_{r_p} z_t$, and z_t is the size at time t

- input the size, z_t at time t using **steps 17 through 19**

- return the reproduction probability of size z_t at time t

Step 24: Define recruitment function, $O_f(z_{t+1}, z_t$ which returns the size distribution of off-spring at time $t + 1$ as follows:

- plot maternal size, z_t at time t versus the offspring size, z_{t+1} at time $t + 1$

- using m-estimation calculate regression parameters a_{rz} and b_{rz} and standard deviation σ_{rz}

- calculate mean $\mu_{rz} = a_{rz} + b_{rz} z_t$ and the standard deviation σ_{rz} of the size,z_t

- generate a normal distribution with mean μ_{rz} and standard deviation σ_{rz}

- input the size, z_t at time t using **steps 17 through 19**

- return the offspring size transition of z_{t+1} at time $t + 1$

Step 25: Define survival or growth kernel function, $P(z_{t+1}, z_t) = g(z_{t+1}, z_t) * s(z_t)$.

Step 26: Define reproduction kernel function, $F(z_{t+1}, z_t) = O_f(z_{t+1}, z_t) * s(z_t) * (1/2) * r_p(z_t)$.

Step 27: Define full kernel $K(z_{t+1}, z_t) = P(z_{t+1}, z_t) + F(z_{t+1}, z_t)$.

Step 28: Calculate the size range of the data set as follows: $L = l - 0.5$ and $U = u + 0.5$ where l and u are min and max of the size respectively.

Step 29: Define number of elements in the matrix $n = 200$.

Step 30: Calculate mesh points, $z_t = L + k * (U - L)/n - (U - L)/2n$, where, $k = 0, 1, 2, \ldots n$.

Step 31: Define a function which evaluates all pairwise components of the vector z_t (mesh points) and returns all elements of $n \times n$ matrix as an output of the function $K(z_t, z_t)$.

Step 32: Calculate dominant eigenvalue of the $n \times n$ matrix.

Step 33: Repeat **Step 3** through **Step 21** .

Step 34: Draw the histogram and box-plot of λ as reported in Figures 5.9 and 5.10.