

DISSERTATION

BAYESIAN ANALYSIS OF AGE-AT-HARVEST DATA  
WITH FOCUS ON WILDLIFE MONITORING PROGRAMS

Submitted by

Paul Bayne Conn

Department of Fish, Wildlife, and Conservation Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2007

COLORADO STATE UNIVERSITY

March 29, 2007

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED  
UNDER OUR SUPERVISION BY PAUL BAYNE CONN ENTITLED  
BAYESIAN ANALYSIS OF AGE-AT-HARVEST DATA WITH FOCUS ON  
WILDLIFE MONITORING PROGRAMS BE ACCEPTED AS FULFILLING IN  
PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Committee on Graduate Work

---

Dr. Kenneth P. Burnham

---

Dr. Paul F. Doherty, Jr.

---

Dr. Kenneth A. Logan

---

Dr. Hari K. Iyer

---

**Adviser**      Dr. Gary C. White

---

**Department Head**      Dr. Kenneth R. Wilson

ABSTRACT OF DISSERTATION  
BAYESIAN ANALYSIS OF AGE-AT-HARVEST DATA WITH FOCUS ON  
WILDLIFE MONITORING PROGRAMS

State and federal agencies often collect hunter harvest data at check stations. When age- and sex-classes can be determined at the time of harvest, such data provide information about population structure. For instance, such summaries are used extensively in quantitative fisheries stock assessment. However, statistically defensible approaches for using age-at-harvest data to monitor terrestrial wildlife populations have not appeared until recently, and are deficient in several respects.

The primary focus of this dissertation is on developing better methods for analyzing wildlife age-at-harvest data, and on applying these methods to real and hypothetical populations. Chapter one starts by developing statistical methods necessary for fitting population dynamics models to age-at-harvest data. As an example, I analyze marking and harvest records from female black bears (*Ursus americanus*) in Pennsylvania.

In chapter two, I describe numerical implementation issues, as well as results from several extensive rounds of simulation testing. I show that Markov chains will typically need to be quite long to accurately summarize the posterior distribution of model parameters. Nonetheless, estimators are shown to display

little bias, to have satisfactory credible interval coverage, and to have a high degree of precision. I show that abundance estimators are quite robust to aging errors, although using data from marked animals twice may lead to overstated measures of precision.

In chapter three, I conduct a power analysis to determine if it would be feasible to monitor black bear in Colorado with age-at-harvest and radio telemetry data. My focus in this chapter is on detecting and estimating population trend for varying levels of effort. I show that five year studies are typically too short for all anticipated levels of marking effort, but that ten year studies can yield meaningful estimates of population trend.

In chapter four, I address methods that can be used to correct age-at-harvest data for misclassification errors. When the aging criterion is inexact, it is possible to correct for errors if additional information is available on error rates. I illustrate proposed methodology with a black bear dataset from Pennsylvania.

Paul B. Conn  
Department of Fish, Wildlife, and Conservation Biology  
Colorado State University  
Fort Collins, CO 80523  
Spring 2007

## ACKNOWLEDGEMENTS

First, I would like to thank my advisor Dr. Gary White for taking me on, for giving me a large degree of independence, and for being so resourceful when previously anticipated funding fell through. I would also like to thank my graduate committee members, Dr. Ken Burnham, Dr. Paul Doherty, Dr. Hari Iyer, and Dr. Ken Logan for their input on my research and for their contributions to my professional development. Dr. Jeff Laake provided much of the original impetus for this research, and Drs. Duane Diefenbach and Mark Terner provided useful input on issues associated with modeling Pennsylvania black bear data. Colorado Division of Wildlife researchers Dr. Mat Alldredge and Dr. Dave Freddy provided valuable input on power analyses for Colorado black bear. Dr. Paul Lukacs provided some helpful suggestions and example code to get me started in C++ and SAS/IML, and Dr. Evan Cooch let me use his multi-processor for some of my simulations and analyses. Funding for this research was provided by the Colorado Division of Wildlife, and by NSF-IGERT grant DGE-0221595003, administered by the PRIMES program at Colorado State University. Last but not least, I would like to thank my parents for instilling in me a hard work ethic and an appreciation for science, and my wife Kathy for kicking so much butt.

Paul B. Conn

March 2007

# TABLE OF CONTENTS

<b>1</b>	<b>Bayesian analysis of wildlife age-at-harvest data</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Model Development . . . . .	4
1.2.1	Data Requirements . . . . .	4
1.2.2	Population Process Models . . . . .	4
1.2.3	The Sampling Model . . . . .	7
1.2.4	Likelihood . . . . .	8
1.2.5	Auxiliary data . . . . .	8
1.3	Example . . . . .	9
1.4	Discussion . . . . .	24
<b>2</b>	<b>Estimating demographic parameters with mark-recovery and age-at-harvest data</b>	<b>26</b>
2.1	Introduction . . . . .	26
2.2	Model Development . . . . .	27
2.2.1	An alternative likelihood . . . . .	27

2.2.2	Estimation . . . . .	29
2.2.3	Prior distributions . . . . .	29
2.2.4	Model extensions: Random effects . . . . .	32
2.2.5	Posterior summaries . . . . .	34
2.2.6	Model selection . . . . .	37
2.3	Efficient MCMC Candidate Generation . . . . .	38
2.3.1	Adaptive algorithms . . . . .	42
2.3.2	Mixing properties of different proposal schemes . . . . .	43
2.3.3	Approximating the posterior: number of iterations . . . . .	51
2.4	Computing . . . . .	58
2.5	Parameter Identification . . . . .	59
2.6	Goodness-of-Fit . . . . .	61
2.7	Simulation Study . . . . .	63
2.7.1	Simulation Module I: Large Sample Performance . . . . .	64
2.7.2	Simulation Module II: Effects of Aging Error . . . . .	90
2.7.3	Simulation Module III: Detecting overdispersion . . . . .	101
2.7.4	Simulation Module IV: Using marked individuals twice . . . . .	103
2.8	Discussion . . . . .	109
<b>3</b>	<b>Potential for using age-at-harvest and radio telemetry data to monitor black bear populations in Colorado</b>	<b>113</b>
3.1	Introduction . . . . .	113
3.2	Model Development . . . . .	115

3.2.1	Incorporating Population Trend into $L_1$ . . . . .	115
3.2.2	Likelihood for Telemetry Data . . . . .	116
3.3	Analysis Methods . . . . .	117
3.3.1	Simulating Data . . . . .	117
3.3.2	Simulation Design . . . . .	119
3.3.3	Estimation . . . . .	124
3.4	Estimator Performance . . . . .	126
3.4.1	Measures of Estimator Performance . . . . .	126
3.4.2	Predictors of Markov Chain Convergence . . . . .	126
3.4.3	Statistical Power for Detecting Population Declines . . . . .	127
3.4.4	Credible Interval Coverage . . . . .	129
3.4.5	Bias . . . . .	132
3.4.6	Coefficient of Variation . . . . .	135
3.5	Discussion . . . . .	136
<b>4</b>	<b>Adjusting Age and Stage Distributions for Misclassification Errors</b>	<b>140</b>
4.1	Introduction . . . . .	140
4.2	Statistical Methods . . . . .	144
4.3	Black Bear Example . . . . .	146
4.4	Discussion . . . . .	153
<b>A</b>	<b>Bias of the Lincoln-Petersen Estimator Under Encounter Covari-</b>	
	<b>ance</b>	<b>158</b>



<b>B</b>	<b>A User's Guide to Program AGEHARV</b>	<b>161</b>
B.1	Introduction . . . . .	161
B.2	User Defined Files . . . . .	162
B.2.1	The Input File . . . . .	162
B.2.2	The Encounter History File . . . . .	168
B.2.3	The Age-at-harvest File . . . . .	169
B.2.4	The Design Matrix File . . . . .	169
B.3	Implementation . . . . .	173
B.3.1	Command Line Instructions . . . . .	173
B.3.2	Output . . . . .	174

## Chapter 1

# Bayesian analysis of wildlife age-at-harvest data

### 1.1 Introduction

The age- and sex-class of harvested animals are often recorded at hunter check stations as part of wildlife monitoring programs. Such data are relatively easy and cost effective to collect and thus are frequently used to inform management decisions. In most cases, harvest numbers are often taken at face value, with trends in harvest samples thought to mirror trends in the population. Unfortunately, this assumption may lead to flawed inferences about population status if trends in harvest data are related to trends in harvest or reporting rates rather than to abundance trends. For instance, even with a standardized reporting system, hunter reporting rates have changed over time for deer in Pennsylvania (Rosenberry et al., 2004).

These considerations have led several researchers to propose more rigorous

approaches to modeling age-at-harvest data for terrestrial species. For instance, assuming that immigration and emigration are negligible, several authors have conceptualized annual harvest counts of a cohort of animals (i.e., animals that entered the population at the same time) as realizations of a multinomial process, with initial cohort abundance as an index, and success probabilities determined by functions of annual survival and harvest rates (Dupont, 1983; Laake, 1992; Gove et al., 2002). Not all parameters could be identified when using the age-at-harvest likelihood alone; however, joint analyses using auxiliary data sources such as radio-telemetry enabled estimation of abundance, survival, and harvest parameters (Gove et al., 2002).

While the work of Laake (1992), Gove et al. (2002), and related authors (cf. Skalski et al., 2005) has proven useful for wildlife practitioners, I see several areas for improvement. First, the product multinomial formulation requires the assumption that the true age of a harvested animal can be determined definitively. For many species, only the first few stages of life can be distinguished. For others, a finer resolution is possible but the reliability of the aging criterion diminishes with the age of the animal (e.g., Harshyne et al., 1998; Hewison et al., 1999). In these cases, one possibility is to group individuals greater or equal to some threshold age into a ‘+’ age class, and use the E-M algorithm to perform maximum likelihood inference (Laake, 1992). Alternatively, one may consider a hierarchical model with additional abundance variables and conduct a Bayesian analysis (e.g., Link et al., 2003).

Another issue with the product multinomial formulation for age-at-harvest data is that the initial size of a cohort has been assumed to be independent of past levels of abundance. As such, the capacity to relate recruitment to previous

population size is missing. Clearly the capacity for relating recruitment to previous levels of abundance is desirable; not only could one incorporate biological knowledge in the form of prior distributions for recruitment process parameters, but one could also consider models with a reduced number of effective parameters.

In this chapter, I describe a general model for the analysis of age-at-harvest data. This model is composed of two parts: an observation model and a population dynamics model. My approach thus falls into a class of models that some authors have described as “hidden process” (Newman et al., 2006). After constructing an appropriate likelihood, I use Markov Chain Monte Carlo (MCMC) for parameter estimation. This framework allows one to properly model stochasticity in abundance classes, to easily summarize marginal posterior distributions for parameters commonly of interest to biologists, to consider a wide variety of random effects models, and to compare the parsimony of alternative models of population dynamics. I illustrate the application of my model with a demographic analysis of black bears (*Ursus americanus*) in Pennsylvania from 1986-1999; data from a mark-recovery study during the same time period were used to inform the estimation of survival and harvest parameters.

## 1.2 Model Development

### 1.2.1 Data Requirements

The fundamental data needed for age-at-harvest analysis is an age-at-harvest matrix,  $\mathbf{C}$ , which summarizes annual harvests by sex and age class. I assume that there is no error associated with aging techniques up to some threshold age,  $A$ . Upon reaching this threshold age, individuals are grouped into a ‘+’ category. For the purposes of this chapter, I further assume that  $\mathbf{C}$  only includes data from the female portion of the population, although extensions to model males are relatively straightforward. I also assume that the investigator has additional information with which to help model the processes of survival and harvest, either through expert knowledge or an auxiliary dataset.

### 1.2.2 Population Process Models

In order to characterize a general class of wildlife population dynamics models, I first condition on  $\mathbf{N}_1$ , the vector of age-specific population sizes in year one immediately prior to harvest. Letting  $[G|H]$  denote the conditional distribution of  $G$  given  $H$ , I write the joint probability mass function (pmf) of abundance in year two as

$$[\mathbf{N}_2 | \mathbf{N}_1, \mathbf{f}_1, \mathbf{S}_1] = [N_{22} | N_{11}, S_{11}] \cdots [N_{2,A+1} | N_{1A}, S_{1A}] \times \quad (1.1) \\ [N_{21} | \mathbf{N}_1, \mathbf{f}_1].$$

Here,  $\mathbf{S}_1$  denotes a vector of age-specific survival probabilities in year 1,  $\mathbf{f}_1$  denotes a vector of age-specific recruitment parameters in year one, and remaining notation is defined in Table 1.1. Note that this formulation assumes that there is no immigration to or emigration from the harvestable population. Note also that I explicitly include  $N_{2,A+1}$  to allow for the possibility that survival of age class  $A$  individuals differs from preceding age classes. Using the current notation, if  $A = 5$ , then one would treat  $N_{2,A+1}$  as  $N_{2,6+}$ , or the number of individuals in year two that are age 6 or older.

Joint pmfs for subsequent years are similar, but an additional allowance is made for  $N_{i,A+1}$  when  $i > 1$ :

$$\begin{aligned}
[\mathbf{N}_{i+1,\cdot} | \mathbf{N}_i, \mathbf{f}_i, \mathbf{S}_i] &= [N_{i+1,2} | N_{i1}, S_{i1}] \cdots [N_{i+1,A} | N_{i,A-1}, S_{i,A-1}] \times & (1.2) \\
& [N_{i+1,A+1} | N_{iA} + N_{i,A+1}, S_{1A}] \times \\
& [N_{i+1,1} | \mathbf{N}_i, \mathbf{f}_i].
\end{aligned}$$

As such, age-specific population structure is modeled as a first-order Markov process. Conditional on the vector of abundances in the first year and parameters for survival and recruitment processes, the probability of all future age- and time-specific abundances may be written as

$$[\mathbf{N} | \mathbf{N}_1, \mathbf{S}, \mathbf{f}] = \prod_{i=1}^{Y-1} [\mathbf{N}_{i+1,\cdot} | \mathbf{N}_i, \mathbf{f}_i, \mathbf{S}_i]$$

In the preceding formulation, population size in year  $i + 1$  consists of individuals who have survived from year  $i$  as well as new recruits to the population.

Table 1.1: Definitions of parameters, latent variables, and statistics used in the joint age-at-harvest, mark-recovery likelihood

---



---

Parameters and Variables	
$S_{ij}$	Probability that an age $j$ individual survives to time $i+1$ given it was alive at time $i$
$h_{ij}$	Probability that an age $j$ individual is harvested and reported to wildlife personnel in $[i, i+1]$ , given that it was alive at time $i$
$f_{ij}$	Per breeder recruitment rate over $[i, i+1]$ , with reference to the number of age $j$ breeders in the population at time $i$ and the number of new recruits at time $i+1$
$N_{ij}$	Number of age $j$ , unmarked individuals in the population in year $i$ immediately prior to harvest. The $N_{1j}$ are parameters while the remaining $N_{kj}$ ( $k > 1$ ) are treated as latent variables
Statistics	
$C_{ij}$	Number of age $j$ unmarked individuals that are harvested and reported to wildlife personnel in year $i$
$M$	Total number of individuals marked and released over the course of the experiment
$H_k$	Encounter history for individual $k$
$t_{k1}$	Year in which animal $k$ is first captured, marked, and released
$t_{k2}$	Year in which animal $k$ is harvested and reported, if encountered again
$I_k$	Indicator variable equal to 1 if animal harvested and reported at some time, 0 otherwise
$a_{ki}$	Age of animal $k$ at time $i$
$A$	Age at which an individual's age cannot be reliably distinguished from older age classes
$Y$	Duration of the study (e.g., years)

---

As written, the number of new recruits depends on abundance in the previous year; however, in some cases, individuals may not enter the harvestable population for several years after they are born. If this is the case, we may simply condition on the augmented vector  $[\mathbf{N}_{1\cdot}, N_{21}, \dots, N_{A_r+1,1}]$ , replacing  $[N_{i+1,1}|\mathbf{N}_{i\cdot}, \mathbf{f}_i]$  with  $[N_{i+A_r+1,1}|\mathbf{N}_{i\cdot}, \mathbf{f}_i]$ . Here,  $A_r$  gives the age at which animals are recruited to the population at risk of harvest. Choices of probability mass functions for survival and recruitment will depend on the population in question, but I suspect that binomial and Poisson models will commonly be appropriate for each process, with possible overdispersion incorporated via random effects.

### 1.2.3 The Sampling Model

I assume that age-at-harvest counts are realizations of stochastic processes, so that we may write

$$[\mathbf{C}|\mathbf{N}, \mathbf{h}] = [C_{1A}|N_{1A}, h_{1A}] \prod_{i=2}^Y [C_{iA}|N_{iA} + N_{i,A+1}, h_{iA}] \times \prod_{i=1}^Y \prod_{j=1}^{A-1} [C_{ij}|N_{ij}, h_{ij}].$$

This formulation assumes that individuals of common age class have the same probability of being harvested; however, extensions allowing for individual heterogeneity (as with a beta-binomial model) would be fairly straightforward. I suspect that binomial models for each component of the harvest process will suffice in most cases.



## 1.2.4 Likelihood

I suggest that inference be based on the likelihood

$$L_1 = [\mathbf{C}, \mathbf{N}|\mathbf{N}_{1.}, \mathbf{S}, \mathbf{h}, \mathbf{f}] = [\mathbf{C}|\mathbf{N}, \mathbf{h}][\mathbf{N}|\mathbf{N}_{1.}, \mathbf{S}, \mathbf{f}]. \quad (1.3)$$

Following Link et al. (2003), I retain the latent variables  $\mathbf{N}_2, \mathbf{N}_3, \dots, \mathbf{N}_Y$  in the likelihood, in part because of the computational difficulty in integrating them out, and in part because they may be of considerable interest to biologists. Indeed, predictions of total female population size in year  $i$  may be made with the quantity  $\sum_{j=1}^{A+1} N_{ij}$ . However, because of the complexity of the model, a Bayesian approach to estimation is required.

## 1.2.5 Auxiliary data

The likelihood in (3) is overparameterized. In order to generate sensible estimates of model parameters, additional information and/or simplifications to the model structure are needed. Gove et al. (2002) suggested basing inference on a joint likelihood similar to

$$L = L_1 L_2, \quad (1.4)$$

where  $L_2$  gives the likelihood for an auxiliary dataset pertaining to survival and harvest parameters, such as from a radio telemetry study. Informative prior distributions on survival and harvest parameters are another alternative for making

Table 1.2: Number of female black bears harvested in Pennsylvania by year and age. Also reported are the total number marked ( $n_1$ ) and number of marked recoveries ( $m_2$ ) each year.

Year	Age at previous winter						$n_1$	$m_2$
	0	1	2	3	4	5+		
1986	152	172	83	60	39	128	170	29
1987	169	197	95	90	38	125	113	17
1988	160	222	108	114	66	149	120	14
1989	274	313	149	136	68	206	96	20
1990	147	149	62	70	39	135	102	15
1991	176	245	85	93	55	147	35	3
1992	192	181	99	107	57	185	104	20
1993	161	312	104	111	51	154	105	18
1994	148	165	89	72	45	156	114	14
1995	208	316	100	124	44	200	134	33
1996	169	261	107	103	54	178	107	21
1997	252	234	115	130	60	245	172	35
1998	241	477	150	127	83	201	135	32
1999	171	256	92	107	39	201	112	15

abundance identifiable, the usual caveats about using informative priors notwithstanding.

### 1.3 Example

I collected statewide marking and harvest records for black bears in Pennsylvania for the period 1986–1999 from the Pennsylvania Game Commission (Table 1.2). During this time period, PGC personnel captured bears throughout their range, determined their sex and age, and released them with individually identifiable metal ear tags. Following a 3-day hunting season in November, hunters were required to present all harvested bears to PGC check stations, where age and sex information was collected and the identity of marked bears was recorded.

Several unique features of the bear data motivated me to elaborate on the general model for age-at-harvest data presented in section 1.2. First, I set recruitment parameters for cubs and yearlings to 0, as Pennsylvania black bear typically do not become pregnant before age 2 (Kordek and Lindzey, 1980; Alt, 1989). Second, I replaced the recruitment terms  $[N_{i+1,1}|\mathbf{N}_i, \mathbf{f}_i.]$  in (1) and (2) with  $[N_{i+1,1}|\mathbf{N}_i - \mathbf{C}_i, \mathbf{f}_i.]$ . In the case of black bears in Pennsylvania, the probability that a hunter reports a harvested bear to wildlife officials is thought to be near 1.0. As such, subtracting the observed harvest from population size the previous year represents the population size immediately after harvest. Conditioning on this quantity should remove one potentially time-varying component from the recruitment process; formulating prior distributions for  $\mathbf{f}_i.$  should also be simplified in this case, although this consideration is beyond the scope of the chapter.

I assumed binomial models for survival and harvest processes, and a Poisson model for the recruitment process. As such, the age-at-harvest likelihood may be

written as

$$\begin{aligned}
L_1 &= L(\mathbf{C}, \mathbf{N} | \mathbf{S}, \mathbf{h}, \mathbf{f}, \mathbf{N}_1.) \\
&= \prod_{i=1}^{13} \prod_{j=1}^5 \binom{N_{ij}}{N_{i+1,j+1}} S_{ij}^{N_{i+1,j+1}} (1 - S_{ij})^{D_{ij}} \\
&\quad \times \binom{N_{16}}{N_{27}} S_{16}^{N_{27}} (1 - S_{16})^{D_{16}} \prod_{i=2}^{13} \binom{N_{i6} + N_{i7}}{N_{i+1,7}} S_{i6}^{N_{i+1,7}} (1 - S_{i6})^{D_{i6}} \\
&\quad \times \prod_{i=1}^{y-1} \frac{\exp(-\lambda_i) \lambda_i^{N_{i+1,1}}}{N_{i+1,1}!} \\
&\quad \times \prod_{i=1}^{14} \prod_{j=1}^5 \binom{N_{ij}}{C_{ij}} h_{ij}^{C_{ij}} (1 - h_{ij})^{N_{ij} - C_{ij}} \\
&\quad \times \binom{N_{16}}{C_{16}} h_{16}^{C_{16}} (1 - h_{16})^{X_{16}} \prod_{i=2}^{14} \binom{N_{i6} + N_{i7}}{C_{i,6}} h_{i6}^{C_{i,6}} (1 - h_{i6})^{X_{i6}}.
\end{aligned}$$

Here,

$$D_{ij} = \begin{cases} N_{ij} - N_{i+1,j+1}, & j < 6 \text{ or } i = 1, \\ N_{i6} + N_{i7} - N_{i+1,7}, & \text{otherwise,} \end{cases}$$

$$X_{ij} = \begin{cases} N_{ij} - C_{ij}, & j < 6 \text{ or } i = 1, \\ N_{i6} + N_{i7} - C_{i,6}, & \text{otherwise} \end{cases}$$

and

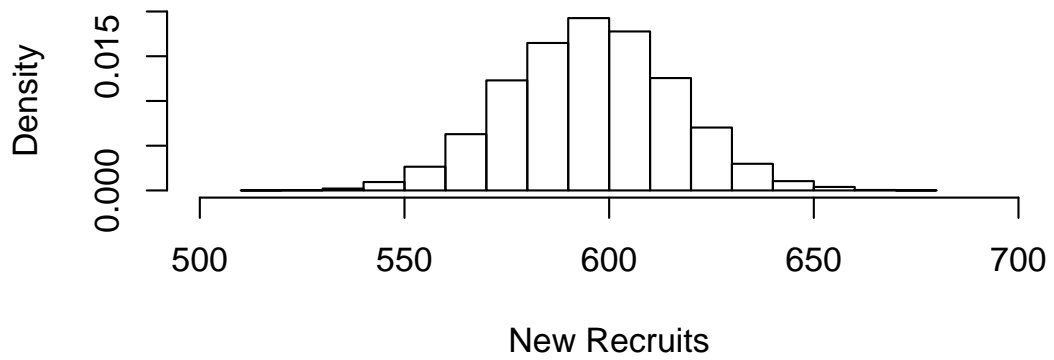
$$\lambda_i = \sum_{j=3}^7 f_{ij} X_{ij}.$$

In order to address concerns that the Poisson distribution may not be a good

choice for the recruitment process, I employed Monte Carlo simulation to compare expected distributions of recruitment under a) biologically realistic levels of variation, and b) a Poisson recruitment model. In the first case, I assumed that there were 1000 breeding age females, approximately half of which (500) would be breeding in any given year. Of those 500, I assumed that on average of 48% would give birth to 3 cubs, 23% would give birth to 4 cubs, 21% would give birth to 2 cubs, 5% would give birth to 1 cub, and 3% would give birth to 5 cubs (Alt, 1989). For each breeding female, I drew a random variate from this distribution to determine the total number of cubs for that female. Conditional on this number, I assumed that the number of female cubs was a binomial draw with the total number of cubs as an index and a success probability of 0.5 (i.e., a 50/50 sex ratio). Of these female cubs, I assumed the number surviving to the start of the harvest season could be drawn from a binomial distribution with success probability 0.8. This process was repeated for every breeding female in a given population for a total of 10,000 populations to yield a Monte Carlo distribution for the number of female cubs that could be expected the following year (Figure 1.1A). Taking the mean number of female cubs per breeding age female to be 0.596 (the mean of this Monte Carlo sample), I also generated a sample of 10,000 from a Poisson distribution with mean and variance equal to 596 (Figure 1.1B). The resulting distributions were almost identical, and I concluded that a Poisson model for recruitment would be satisfactory in this case.

In addition to age-at-harvest data, I also compiled mark-recovery histories for all females marked over the course of the study. A likelihood for these data was formulated along the lines of Brownie et al. (1985); however, I directly mod-

### A. Biological Model



### B. Poisson Model

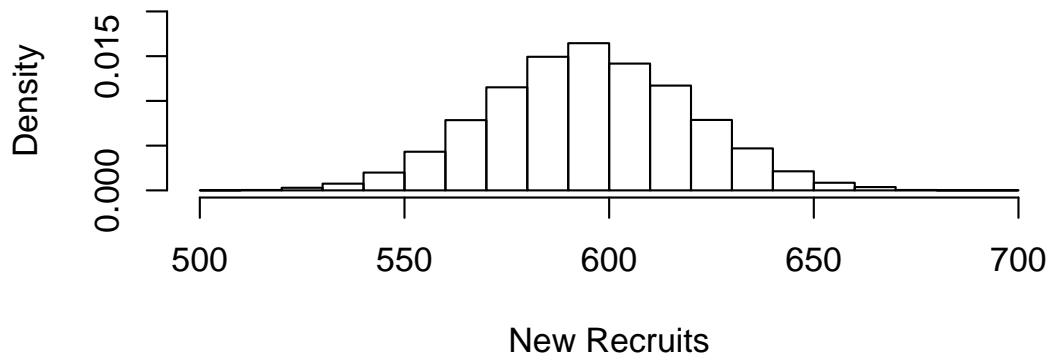


Figure 1.1: Distribution of female cubs that could be expected from a hypothetical population of 1,000 breeding age females, incorporating A) a biologically based model for the recruitment process, and B) a Poisson model for the recruitment process.

eled encounter histories instead of using minimum sufficient statistics. I write this likelihood as

$$L_2 \propto \prod_{k=1}^M Pr(H_k),$$

where

$$Pr(H_k) = \begin{cases} h_{t_{k2}, a_k, t_{k2}} \prod_{i=t_{k1}}^{t_{k2}-1} S_{i, a_k, i}, & I_k = 1, \\ 1 - h_{t_{k1}, a_k, t_{k1}} - \sum_{i=t_{k1}}^{y-1} \left( \prod_{j=t_{k1}}^i S_{j, a_{kj}} \right) h_{i+1, a_k, i+1}, & I_k = 0, \end{cases}$$

and definitions of statistics and parameters are given in Table 1.1. Formulating  $L_2$  in terms of encounter histories provides flexibility; in this form, parameters may easily be expressed as functions of individual covariates such as age (Pollock, 2002).

I made all the common assumptions typical for mark-recovery studies (Williams et al., 2002): animals behave independently, marks are not lost or overlooked, and there is no individual heterogeneity (at least that cannot be explained by age class). Further, I assumed that there was no mortality between the time a bear is marked and the harvest season. Several of these assumptions may be violated to some degree; for instance, sampling effort to mark black bears was distributed between March and November, although mortality was thought to be low during this period for individuals greater than one year of age. Nevertheless, the recovery rates of cubs were likely underestimated, ostensibly causing a positive bias in the number of new recruits each year. Cubs' fates were also certainly not independent from those of their litter mates or those of their mothers, which would tend to cause negative bias

in variance estimates. Tag loss occurred, but was of small magnitude for females (Diefenbach and Alt, 1998). I further assumed that the population was closed to immigration and emigration, which is a reasonable assumption given the sheer size of the geographic area being considered. As suggested in section 2, I based inference on (1.4), which assumes that mark-recovery and age-at-harvest data are independent. Nevertheless, I included data from marked individuals in the age-at-harvest matrix so that total population abundance could be estimated. In some settings, this assumption may result in measures of uncertainty that are too precise (see Chapter 2).

My goal was to estimate abundance, survival, recovery rate, and recruitment for female black bears. However, the model as formulated was overparameterized. In the field of fisheries stock assessment, this problem is often remedied by assuming an additive model for sampling parameters (the “separability assumption”; Megrey, 1989). Similarly, additive and/or random effects models have been proposed for survival and recruitment parameters in age-structured models for whooping cranes (Link et al., 2003), and for the analysis of data from marked individuals in general (Lebreton et al., 1992; Barry et al., 2003).

I adopted aspects of each of these approaches, fitting a total of 4 models to the data which varied by the number of fixed and random effects on the logits of survival and recovery rate, and on the log of recruitment rate (Table 1.3). Preliminary analysis indicated substantial differences in observed and expected harvest data when a model such as

$$\text{Logit}(h_{ij}) = v_i + \phi_j$$



Table 1.3: Models fit to age-at-harvest and mark-recovery data sorted by estimated deviance information criterion (DIC), where  $\Delta\text{DIC}$  gives the difference in DIC from the highest ranked model. Age is modeled as a fixed effect, while time is modeled as a random effect (when included). Also presented are Bayesian p-values ( $p_B$ ) for each model.

Model Name	Logit( $S_{ij}$ )	Logit( $h_{ij}$ )	Log( $f_{ij}$ )	$\Delta\text{DIC}$	$p_B$
$S(a+t)h(a+t)f(t)$	$\gamma_j + \alpha_i$	$\phi_j + \nu_i + \epsilon_{ij}$	$\beta + \kappa_i$	0.0	0.67
$S(a+t)h(a+t)f(\cdot)$	$\gamma_j + \alpha_i$	$\phi_j + \nu_i + \epsilon_{ij}$	$\beta$	6.0	0.58
$S(a)h(a)f(t)$	$\gamma_j$	$\phi_j + \epsilon_{ij}$	$\beta + \kappa_i$	28.9	0.5
$S(a)h(a)f(\cdot)$	$\gamma_j$	$\phi_j + \epsilon_{ij}$	$\beta$	34.1	0.49

was fit to the data. These differences likely resulted from annual changes in the temporal distribution of denning dates as well as the percentage of a given cohort that was pregnant. Pregnant females typically den prior to the rest of the population, and thus are more likely to be unavailable for harvest during the hunting season. It was difficult to model these biological processes directly, so I followed the approach of Barry et al. (2003) and included overdispersion terms,  $\epsilon_{ij}$ , in the formulation for harvest rates. These were modeled as random effects on the logit scale and were assumed to have a normal distribution with mean zero and precision parameter  $\tau_\epsilon$ . Similarly, when included in the model structure, year effects (i.e.,  $\alpha_i$ ,  $\nu_i$ , and  $\kappa_i$ ) were modeled as normally distributed random effects on the logit scale with mean zero and precision parameters  $\tau_\alpha^2$ ,  $\tau_\nu^2$ , and  $\tau_\kappa^2$ , respectively. Remaining parameters were all modeled as fixed effects.

Conducting a Bayesian analysis required that I specify prior distributions for model parameters. In particular, I assigned the following diffuse priors:

$$[N_{1j}] \propto c,$$

$$[\gamma_j], [\phi_j] \sim \text{Normal}(0, 3),$$

$$[\beta] \sim \text{Normal}(0.25, 1), \text{ and}$$

$$[\tau_\alpha], [\tau_v], [\tau_\kappa], [\tau_\epsilon] \sim \text{Gamma}(0.1, 0.1).$$

I used Gibbs sampling to sequentially update each parameter and latent variable. The full conditionals for precision parameters were available in closed form, and were simulated directly as

$$[\tau_\alpha | \boldsymbol{\alpha}] \sim \text{Gamma} \left( \frac{Y-1}{2} + 0.1, \frac{\sum \alpha_i^2}{2} + 0.1 \right),$$

$$[\tau_v | \mathbf{v}] \sim \text{Gamma} \left( \frac{Y}{2} + 0.1, \frac{\sum v_i^2}{2} + 0.1 \right),$$

$$[\tau_\kappa | \boldsymbol{\kappa}] \sim \text{Gamma} \left( \frac{Y-1}{2} + 0.1, \frac{\sum \kappa_i^2}{2} + 0.1 \right), \text{ and}$$

$$[\tau_\epsilon | \boldsymbol{\epsilon}] \sim \text{Gamma} \left( \frac{YA}{2} + 0.1, \frac{\sum \epsilon_{ij}^2}{2} + 0.1 \right).$$

All remaining parameters and variables were updated with Metropolis-Hastings steps. Proposals in a given iteration were normally distributed with a mean at the previous iteration's parameter value, and a standard deviation was chosen so as to achieve a 35-40% acceptance rate.

For each model, I ran two independent Markov Chains of length 1 million with overdispersed starting values. If after 500,000 iterations Gelman-Rubin statistics (Gelman et al., 2004) indicated convergence, I combined the second halves of each chain to generate a sample of one million from the posterior distribution. In order

to save disk space, this sample was thinned to 200,000 by recording every fifth iteration. Marginal posterior distributions were then summarized by calculating moments and 90% HPD Bayesian credible intervals. Deviance information criterion (DIC; Spiegelhalter et al., 2002) was also calculated for purposes of model selection.

I implemented a goodness-of-fit test based on a Bayesian p-value (Gelman et al., 2004) for the age-at-harvest portion of the likelihood. For a given sample  $i$  from the posterior distribution, I simulated harvest data,  $\mathbf{C}_i^{\text{rep}}$ , given  $\boldsymbol{\theta}_i$ , the set of parameter values at iteration  $i$ . Next, deviance for the age-at-harvest portion of the likelihood was calculated for observed,  $D_i(\mathbf{C}, \boldsymbol{\theta}_i)$ , and for simulated,  $D_i(\mathbf{C}_i^{\text{rep}}, \boldsymbol{\theta}_i)$ , data. The Bayesian p-value was then obtained as

$$p_B = \frac{1}{K} \sum_{i=1}^K I_{[0, \infty)} [D_i(\mathbf{C}_i^{\text{rep}}, \boldsymbol{\theta}_i) - D_i(\mathbf{C}, \boldsymbol{\theta}_i),]$$

where  $K$  denotes the total number of samples in which age-at-harvest data is simulated, and  $I_{\Omega}(x)$  denotes an indicator function for the set  $\Omega$ . For purposes of this chapter, I set  $K = 200$ , spacing samples evenly across Markov chain iterations.

When fit to the data, models including time- and age- varying survival and recovery rate were heavily favored by DIC (Table 1.3). There also appeared to be some evidence for time-varying recruitment. I chose to base inference on  $S(a+t)h(a+t)f(t)$  for two reasons. First, it was the highest-ranked model according to DIC; second, it included the most complexity. For instance, there is a tendency for overstating precision when data from marked animals are used in both portions of the likelihood, particularly when model complexity is low (see Chapter 2).

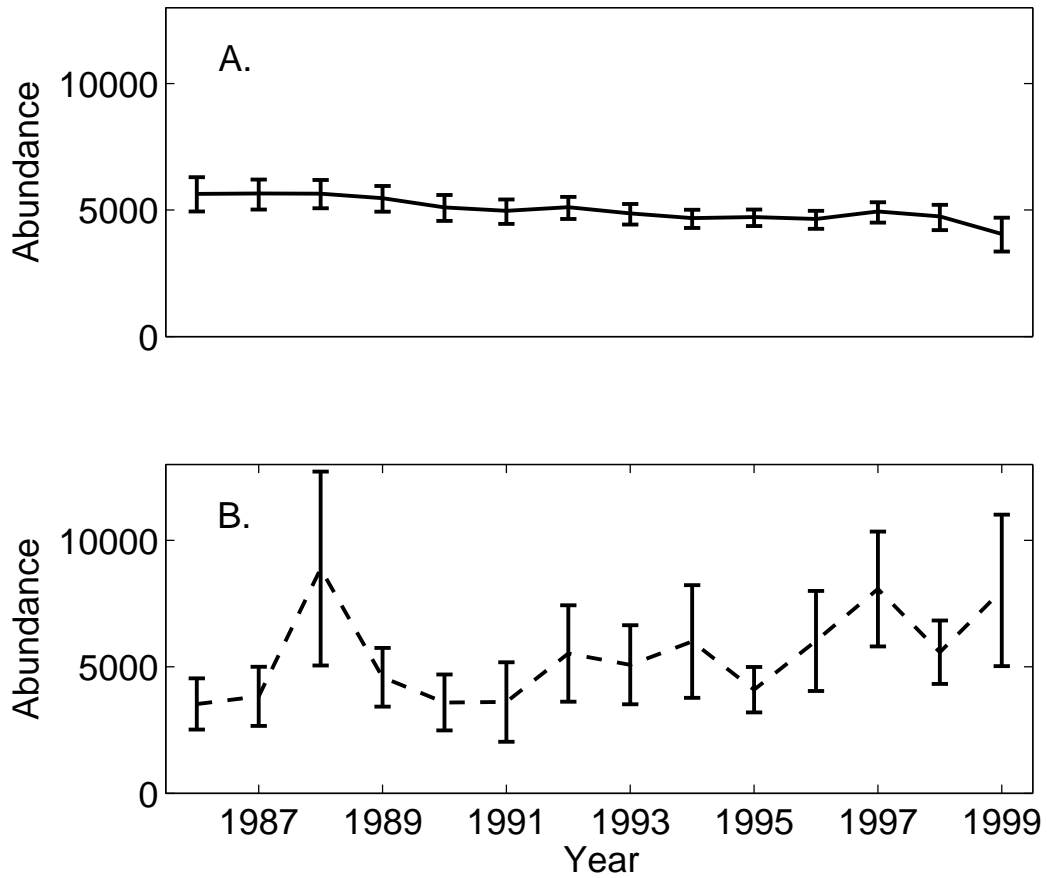


Figure 1.2: Estimated abundance for female black bears in Pennsylvania from 1986 to 1998 as estimated from A) the joint age-at-harvest and mark-recovery estimator, and B) the bias corrected Lincoln-Petersen estimator. Posterior means and 90% Bayesian credible intervals are presented from the highest ranked DIC model, while point estimates and 90% asymptotic confidence intervals are presented for the Lincoln-Petersen estimator.

Posterior summaries indicated that female black bear abundance declined slightly in Pennsylvania from 1986 to 1999 (Figure 1.2). In order to contrast our estimator with another commonly used abundance estimator, I computed year-specific bias-corrected Lincoln-Petersen (LP) estimates of abundance, together with accompanying variances (see Seber, 1982). Point estimates from the LP estimator were considerably more variable, and had substantially higher standard errors (Figure 1.2). Further, consecutive point estimates using the LP approach were often out of the realm of biological possibility. Diefenbach et al. (2004) noted this tendency, suggesting that annual changes in the availability of pregnant females for harvest could lead to a high degree of variability in single season estimators of abundance. In addition to abundance, I was also able to summarize posterior distributions for survival probability (Figure 1.3), recovery probability (Figure 1.4), and recruitment rate (Figure 1.5), three quantities of fundamental interest to population biologists and managers. These estimates reconfirm that the black bear population in Pennsylvania is one of the most productive in the United States (Alt, 1989).

Despite positive advances in methodology, I am hesitant to make any definite conclusions about population trend toward the end of the study. For instance, the trend indicated by age-at-harvest abundance estimates was different than that from LP estimates, and also from previously published estimates generated from a Horvitz-Thompson (HT) type estimator (Diefenbach et al., 2004). Both of these estimators indicated an increase in abundance over the same time period. One possible explanation for differences in estimated abundance is temporary emigration from the harvestable population, which can cause negative biases in survival

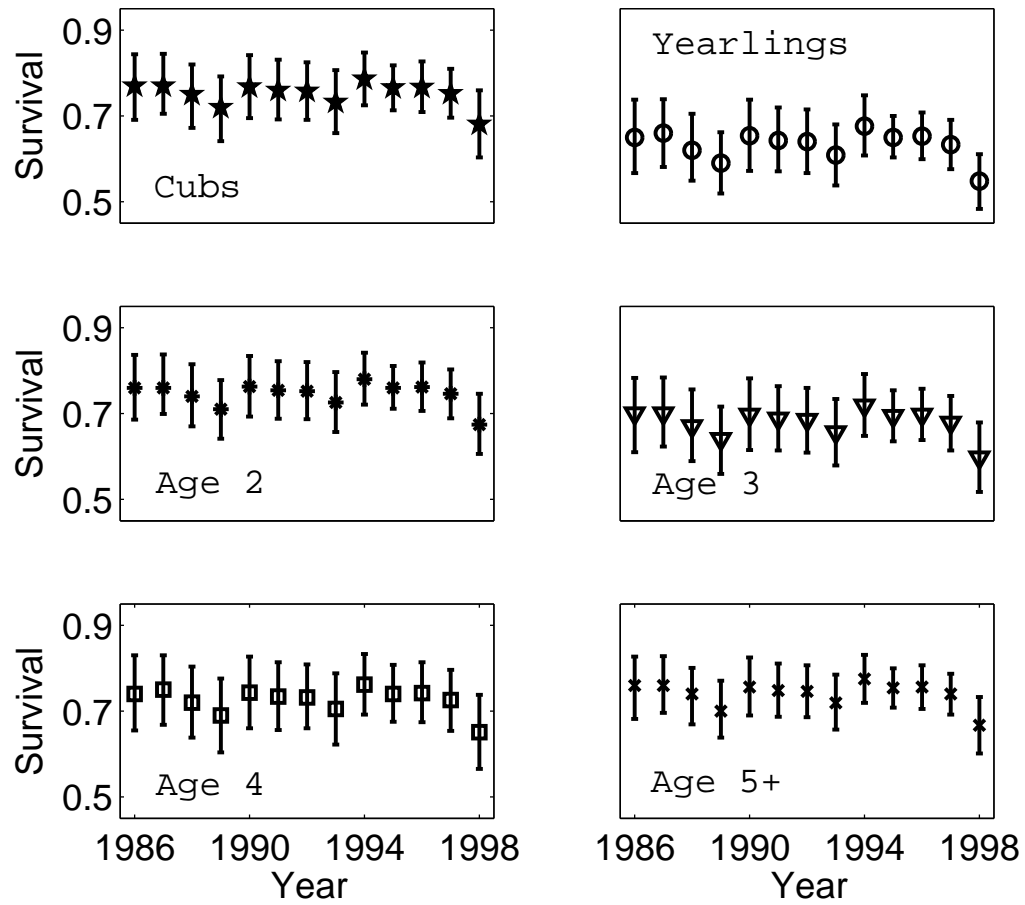


Figure 1.3: Female black bear survival in Pennsylvania from 1986 to 1998 as estimated from age-at-harvest and mark-recovery data. Posterior means and 90% Bayesian credible intervals are presented from the highest ranked DIC model.

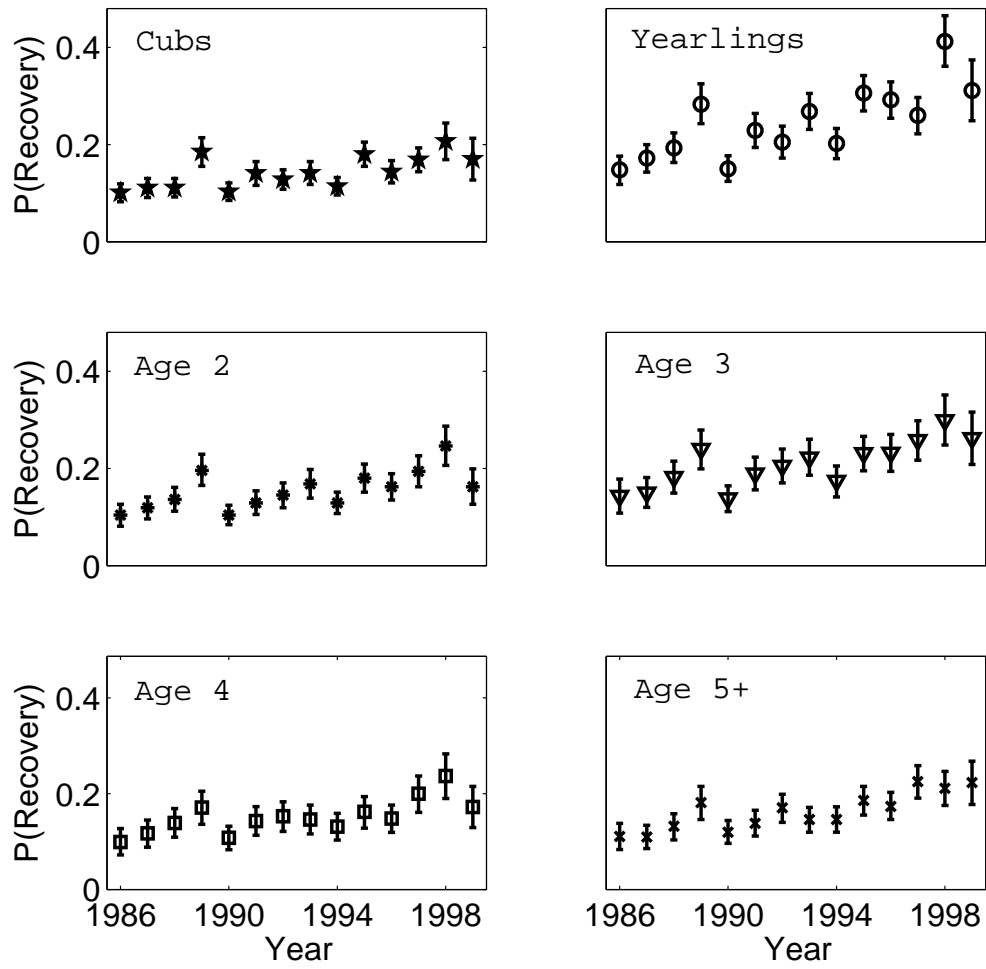


Figure 1.4: Recovery rates for female black bears in Pennsylvania from 1986 to 1998 as estimated from a joint analysis of age-at-harvest and mark-recovery data. Posterior means and 90% Bayesian credible intervals are presented from the highest ranked DIC model.

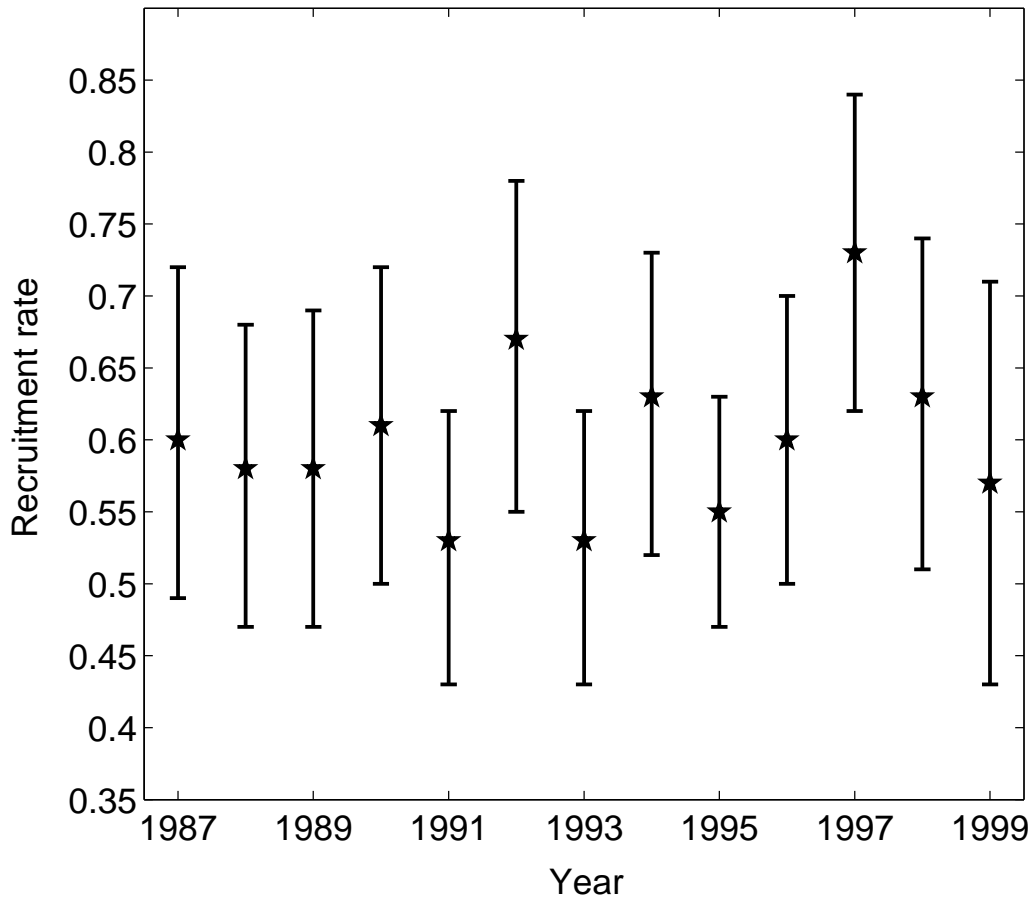


Figure 1.5: Estimated recruitment rate (number of female cubs immediately prior to hunting season the following year per breeding age female that survived the current hunting season) for Pennsylvania black bear from 1987 to 1999. Presented are posterior means and 90% Bayesian credible intervals from the highest ranked DIC model.



probability toward the end of mark-recovery time series (W. L. Kendall, *personal communication*). Pregnant females typically enter dens at earlier dates than other bears, and thus in some sense have temporarily emigrated. Assumption violations such as preharvest mortality, tag loss, and nonrandom selection of bears for marking may also have contributed to a biased picture of temporal trend in abundance. These considerations are a focus of current research.

## 1.4 Discussion

In this chapter, I was able to extract a wealth of information on population demography by jointly modeling the age structure of harvests and mark-recovery data. Parameter estimates were required to be internally consistent with other population parameters, leading to higher precision and less temporal variation in point estimates than the Lincoln-Petersen estimator. Posterior summaries were also available for other demographic parameters such as survival and recruitment. By adopting a Bayesian approach to the problem, I was able to conduct analysis using a likelihood that would be intractable for traditional maximum likelihood techniques. This allowed for improvements upon previously proposed models for analyzing age-at-harvest data. In particular, I have permitted population models with a self-loop for older age classes as well as an explicit recruitment process. Expert opinion concerning biologically plausible values for demographic parameters can also be incorporated in the form of informative prior distributions.

With regard to Pennsylvania black bear, future work should explore the ten-

ability of several key model assumptions, as with the assumption of no preharvest mortality following marking. This assumption will be violated to some degree, particularly for cubs that are marked in dens during March, and may induce positive bias in abundance estimates. Tag loss could also induce bias when multiple years are analyzed within the mark-recovery framework. A number of other assumptions are also likely to be violated. For instance cubs fates are dependent on those of their mothers, which may serve to introduce unmodeled overdispersion.

Finally, I note that the modeling framework developed here can easily be extended to incorporate additional data sources. For instance, if recaptures are available in addition to recoveries, one may simply replace  $L_2$  with a likelihood similar to that attributed to Burnham (1993). Likewise, information from radio-telemetry studies could be incorporated (e.g., Gove et al., 2002, also see Chapter 3). I believe this flexibility is essential for biologists, who are often confronted with a diverse array of sampling challenges.

## Chapter 2

# Estimating demographic parameters with mark-recovery and age-at-harvest data

### 2.1 Introduction

Chapter one outlined one possible strategy for estimating population parameters from age-at-harvest and mark-recovery data. However, other parameterizations and posterior simulation strategies may be more efficient in terms of the number of iterations required to generate an accurate summary of the joint posterior distribution. Further, even if a good sampler can be developed, there is no guarantee that posterior moment estimators will perform well, especially if model assumptions are violated.

In this chapter, I explore the problem of estimating demographic parame-

ters from age-at-harvest and mark-recovery data in further depth. In particular, I describe alternate parameterizations for model structure, summarize Bayesian approaches for calculating posterior summaries and measures of model fit, and use simulation to explore the performance of different posterior samplers, as well as the performance of population parameter estimators under a number of biologically meaningful scenarios. In addition to employment of a response surface design to explore general estimator performance, I also use simulation to assess the ability of a posterior predictive check to diagnose lack of fit from overdispersion, and to examine the robustness of several key model assumptions, such as when aging errors occur or when mark-recovery and age-at-harvest likelihoods are dependent. The latter case can arise, for instance, when data from marked animals are included as part of the age-at-harvest matrix.

## **2.2 Model Development**

### **2.2.1 An alternative likelihood**

Chapter 1 described a general modeling framework for age-at-harvest data. However, the observation model was intentionally selected to fit in nicely with the mark-recovery parameterization used by Brownie et al. (1985). An alternative formulation of the mark-recovery likelihood was given by Seber (1982), which can be written in

terms of individual encounter histories as follows:

$$L_2 \propto \prod_{k=1}^M \Pr(H_k),$$

where

$$\Pr(H_k) = \begin{cases} (1 - S_{t_k', a_{t_k'}}) r_{t_k', a_{t_k'}} \prod_{i=t_k}^{t_k'-1} S_{i, a_i}, & I_k = 1, \\ 1 - \sum_{i=t_k-1}^{y-1} \left( \prod_{j=t_k}^i S_{j, a_j} \right) (1 - S_{i+1, a_{i+1}}) r_{i+1, a_{i+1}}, & I_k = 0. \end{cases}$$

Definitions of statistics and parameters are given in Table 2.1. Use of Seber's parameterization in conjunction with age-at-harvest data requires reformulation of the age-at-harvest model so that parameters can be shared between  $L_1$  and  $L_2$ . This can be accomplished by letting

$$\begin{aligned} L_1 &= L(\mathbf{S}, \mathbf{r}, \mathbf{f}, \mathbf{N}, \mathbf{D} | \mathbf{C}) \\ &= \prod_{i=1}^Y \prod_{j=1}^A \text{Bin}(D_{ij}; N_{ij}, 1 - S_{ij}) \\ &\quad \times \prod_{i=1}^{y-1} \text{Pois}(N_{i+1,1}; f_{i1}N_{i1} + f_{i2}N_{i2} + \cdots + f_{iA}N_{iA} + f_{iA}N_{i,A+1}) \\ &\quad \times \prod_{i=1}^y \prod_{j=1}^A \text{Bin}(C_{ij}; D_{ij}, r_{ij}) \end{aligned}$$

Here, the notation  $\text{Bin}(x; N, p)$  indicates that  $x$  is binomially distributed with index  $N$  and success probability  $p$ . Similarly,  $\text{Pois}(x; \theta)$  denotes that the variable  $x$  follows a Poisson distribution with parameter  $\theta$ . Bold face characters represent year- and

age-specific parameter vectors. The relationship between latent variables  $N_{ij}$  and  $D_{ij}$  can be summarized as follows:

$$N_{i+1,j} = \begin{cases} N_{i,j-1} - D_{i,j-1}, & j < A, \\ N_{i,A-1} + N_{iA} - D_{i,A-1} - D_{iA}, & j = A, \end{cases}$$

### 2.2.2 Estimation

Estimation via traditional maximum likelihood techniques is prohibitively difficult in this case, due to the large number of (latent) abundance parameters that need to be integrated out of the likelihood to make this feasible. As Link et al. (2003) pointed out, there may be considerable interest in these parameters. For instance, calculating annual abundance,  $N_{i\cdot} = \sum_j N_{ij}$  requires knowledge of all of the  $N_{ij}$  parameters. For these reasons, as well as for the ease with which hierarchical extensions to model structure can be implemented, Bayesian estimation using Markov Chain Monte Carlo (MCMC) is preferable for estimating model parameters.

### 2.2.3 Prior distributions

Bayesian implementation requires the formulation of prior distributions for a number of model parameters. For instance, for the Seber parameterization we require prior distributions for the following sets of parameters:  $\mathbf{S}$ ,  $\mathbf{r}$ ,  $\mathbf{N}_{1\cdot}$ , and  $\mathbf{f}$ . Here,  $\mathbf{N}_{1\cdot}$  denotes the set of abundance parameters in year 1. Typically, Bayesian analysts select priors that lead to conjugate posterior distributions (where the posterior has the same form

Table 2.1: Definitions of parameters, latent variables, and statistics used in the joint age-at-harvest, mark-recovery likelihood

---



---

Parameters and Variables	
$S_{ij}$	Probability that an age $j$ individual survives to time $i+1$ given it was alive at time $i$
$r_{ij}$	Probability that an age $j$ individual is reported, given that it died in $[i, i+1]$
$h_{ij}$	Probability that an age $j$ individual is harvested and reported in $[i, i+1]$ , given that it was alive at time $i$
$f_{ij}$	Per breeder recruitment rate over $[i, i+1]$ , with reference to the number of age $j$ breeders in the population at time $i$ and the number of new recruits at time $i+1$
$N_{ij}$	Number of age $j$ , individuals in the population at time $i$ . The $N_{1j}$ are parameters while the remaining $N_{kj}$ ( $k > 1$ ) are treated as latent variables
$D_{ij}$	A latent variable giving the number of age $j$ individuals at time $i$ that die in the interval $[i, i+1]$
Statistics	
$C_{ij}$	Number of age $j$ individuals that are harvested and reported in year $i$
$M$	Total number of individuals marked and released over the course of the experiment
$H_k$	Encounter history for individual $k$
$t_k$	Year in which animal $k$ is first captured, marked, and released
$t_k'$	Year in which animal $k$ is harvested and reported, if encountered again
$I_k$	Indicator variable equal to 1 if animal $k$ harvested and reported at some time, 0 otherwise
$a_{t_k}$	Age of animal $k$ at time $t$
$A$	Age at which an individuals age cannot be reliably distinguished from older age classes
$Y$	Duration of the study (e.g., years)

---

as the prior). Owing to the complicated form of the likelihood, this is not possible with either parameterization of the combined mark-recovery age-at-harvest model. Nevertheless, the following prior specification strategies may be useful in practice.

Several prior distributions have been used for abundance in a mark-recapture context. For instance, Link et al. (2003) used a constant, improper prior on abundance; that is,  $\Pr(N_{1j}) \propto c$ . This is equivalent to specifying a  $\text{Uniform}(0, x)$  prior on abundance, where  $x$  is an arbitrarily large constant selected so that the probability of reaching this value in the course of MCMC simulation is effectively 0. Another standard improper but uninformative prior on  $N_{1j}$  can be specified as  $\Pr(N_{1j}) \propto 1/N_{1j}$  (Madigan and York, 1997). In univariate models for abundance estimation, this selection corresponds to a transformation invariant (i.e., Jeffreys') prior (Jeffreys, 1961).

If information other than harvest data and mark-recovery data are available on initial population sizes, one may incorporate this information by specifying an informative prior. A typical choice is the Poisson distribution, which in this case would amount to specifying  $N_{1j} \sim \text{Pois}(x; \lambda_{1j})$ , where  $\lambda_{1j}$  could be subject to hierarchical extension if prior information is imprecise (see, e.g., Madigan and York, 1997; Smith, 1991).

It is convenient to consider models for  $S_{ij}$ ,  $r_{ij}$ , and  $h_{ij}$  that occur on the logit scale, so that one can entertain model extensions incorporating covariates and random effects ( $h_{ij}$  being recovery rate if the parameterization of the previous chapter is adopted). As such, priors for these parameters should be selected such that they are approximately uninformative when transformed to (0,1) space (unless, of course,



there are auxiliary data available). As an approximately uninformative prior, the  $\text{Normal}(0, 3)$  distribution on the logit scale is a convenient selection, as it leads to back transformed random deviates that are approximately uniformly distributed on  $(0,1)$ . This is not the case when additive models are used, however, as the inverse logit transformation of the sum of two  $\text{Normal}(0, 3)$  random variables has a slightly parabolic shape.

The recruitment parameters  $\lambda_{ij}$  must be nonnegative, so it is natural to consider a log link for these parameters. Appropriate priors will likely be somewhat informative, and should ultimately rest on the biology of the species being analyzed. For example, if the taxa in question averages five or less young per year, an informative but diffuse prior could be specified as

$$\log(\lambda_{ij}) \sim \text{Normal}(0.25, 1),$$

which has substantial support on all likely parameter values. In cases where informative prior distributions such as this one are utilized, prior sensitivity analyses are a natural way of checking to see how strongly posterior inference relies on specification of the prior (Gelman et al., 2004).

#### **2.2.4 Model extensions: Random effects**

Random effects models are useful for describing variation in parameters that are modeled as randomly distributed across time, space, or other factors, while keeping the overall number of parameters low. For instance, if survival varies over time in a

“random” manner, the model

$$\text{logit}(S_i) = \beta_0 + \epsilon_i,$$

where  $\epsilon_i \sim \text{Normal}(0, \sigma^2)$  can represent this variation with fewer parameters than if a separate  $S_i$  were estimated at each time period (Royle and Link, 2002). Of course, this assumes that the number of observations are distributed sufficiently over time or space; for example, there would be little use in implementing a random effects model for a 3-occasion mark-recovery study. Random effects have also been used to accommodate lack of fit in the Bayesian implementation of mark-recovery models (Barry et al., 2003).

A possible extension for including random effects on survival and reporting rate is to model

$$\text{logit}(S_{ij}) = \mu_S + \epsilon_{ij}^S, \text{ and } \text{logit}(r_{ij}) = \mu_r + \epsilon_{ij}^r, \text{ where}$$

$$\epsilon_{ij}^S \sim \text{Normal}(0, 1/\tau_S), \text{ and } \epsilon_{ij}^r \sim \text{Normal}(0, 1/\tau_r).$$

Similarly to fixed effects models, priors of  $\mu_S \sim \text{Normal}(0, 3)$ , and  $\mu_r \sim \text{Normal}(0, 3)$ , are approximately uninformative on the real scale. Informative priors of  $\tau_S \sim \text{Gamma}(\alpha_1, \beta_1)$  and  $\tau_r \sim \text{Gamma}(\alpha_2, \beta_2)$  for the precision of random effects terms are natural to consider, where  $\beta$  denotes a rate (rather than a scale) parameter. The priors on  $\tau_S$  and  $\tau_r$  are conjugate; that is, the marginal conditional posterior is available in closed form and thus it is possible to simulate from these directly

without a Metropolis-Hastings step. In particular,

$$\tau_S \sim \text{Gamma} \left( \frac{n}{2} + \alpha, \frac{\sum \epsilon_i^2}{2} + \beta \right)$$

Royle and Link (2002) suggest specifying a diffuse prior on precisions by letting  $\alpha = 0.1$  and  $\beta=0.1$ .

## 2.2.5 Posterior summaries

One advantage of Bayesian inference is the ability to obtain marginal posterior distributions that completely summarize the current state of knowledge about a given model parameter (Gelman et al., 2004). Researchers can use these distributions directly to predict future system states, as with a posterior predictive distribution, or as a basis for formulating management decisions according to a pre-specified utility function. However, in certain cases, one may also wish to provide a point estimate together with an analog to a frequentist confidence interval, as with a Bayesian credible interval. Since most researchers are familiar with these types of summaries, their use may better communicate results. Further, typical estimator performance statistics, such as mean squared error, percent relative bias, and interval coverage rely on point and quantile estimates.

Selection of a Bayesian point estimator is often made based on quantitative consideration of the ramifications of choosing a given estimator  $\hat{\theta}$  as articulated through a loss function. Acknowledging uncertainty about truth (that is, about the “true” value of  $\theta$ ), one minimizes the loss function by choosing an estimator that

minimizes

$$\int_{\boldsymbol{\theta}} l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta},$$

where  $l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$  gives the specified loss function and  $p(\boldsymbol{\theta}|\mathbf{x})$  gives the posterior distribution of  $\boldsymbol{\theta}$  given data  $\mathbf{x}$  (Bernardo and Juárez, 2003). In general, the posterior mean satisfies *squared error loss*, the posterior median satisfies *absolute error loss*, and the posterior mode satisfies *zero-one loss* (Bernardo and Juárez, 2003). Zero-one loss results when

$$l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \begin{cases} 0 & \hat{\boldsymbol{\theta}} = \boldsymbol{\theta} \\ 1 & \text{otherwise} \end{cases}.$$

Describing alternative point estimates in this manner may help to explain why many authors in the Bayesian literature have chosen to use the posterior mean as a point estimator. However, when a simulation study is undertaken with respect to a “true” parameter vector  $\boldsymbol{\theta}$ , there is no uncertainty about its value and thus we would desire a point estimator that minimizes

$$l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x}) = c \times l(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}),$$

where  $c$  is a constant. For each of the loss functions described above, this quantity is minimized when  $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ . In the case where noninformative priors are chosen, choosing  $\hat{\boldsymbol{\theta}}$  to be the posterior mode is approximately equivalent to choosing the maximum likelihood estimator for  $\boldsymbol{\theta}$ . Since maximum likelihood estimators are consistent (Casella and Berger, 1990), it follows (heuristically) that the posterior

mode is the appropriate estimator for calculating simulation performance statistics.

Estimating a posterior mode using a sample from the posterior distribution is not as straightforward a task as estimation of a posterior mean or median. The approach that I implemented was to use a kernel density (KD) estimator for the posterior distribution (Silverman, 1986), and then to find the mode of this estimator. For situations where parameters are estimated away from boundaries (e.g., 0 or 1 for survival), a KD approach may be preferable over spline-based estimators (K. P. Burnham, *personal communication*). Implementation of KD involves selecting a kernel,  $\kappa$  as well as a bandwidth  $h$  that controls the degree of smoothing. I chose to use a normal distribution for the kernel, and selected a bandwidth that would be asymptotically optimal if the posterior was normally distributed, namely  $h = 1.06 \times \min(I\hat{Q}R/1.34, \hat{\sigma}) \times n^{-1/5}$ , where  $I\hat{Q}R$  and  $\hat{\sigma}$  give the estimated interquartile range and standard deviation of the posterior, respectively, and  $n$  gives the number of independent samples from the posterior distribution (Silverman, 1986). Adaptive bandwidth selections may result in a better approximation of the posterior than this default choice, but these approaches typically require computer-intensive cross-validation procedures (Silverman, 1986).

Bayesian credible intervals can be used to give the probability that a parameter falls into a given interval given the current state of knowledge for that parameter (as summarized by the marginal posterior distribution). Despite a slightly different interpretation, credible intervals are roughly analogous to frequentist confidence intervals. One approach to obtaining a credible interval for parameter  $\theta$  is simply to sort marginal samples from lowest to highest, and then to use order statistics

$(\theta^{(n\alpha)}, \theta^{(n(1-\alpha))})$  as the lower and upper bounds for a  $100(1-\alpha)$  interval (Chen et al., 2000). However, as Chen et al. (2000) note, highest probability density (HPD) intervals are preferable because they are typically shorter (at least if the target distribution is non-symmetric). In the case of a unimodal distribution, finding an HPD interval amounts to finding all  $\theta^{(L)}$  and  $\theta^{(U)}$  such that  $(U - L) = n(1 - \alpha)$  and then taking the values for  $L$  and  $U$  that minimize  $\theta^{(U)} - \theta^{(L)}$  (Chen et al., 2000).

### 2.2.6 Model selection

Several options exist for comparing the relative parsimony of different model parameterizations. In particular, a Bayesian analog of AIC called the deviance information criterion (DIC; Spiegelhalter et al., 2002) has been developed, as has a criterion based on asymptotic approximation of Bayes factors called the Bayesian information criterion (BIC; Schwarz, 1978). While model averaging approaches based on reversible jump MCMC (Green, 1995) also appear promising, requisite computing time is dauntingly large even for simpler model structures than considered here (e.g., Barry et al., 2003), and thus were omitted from consideration.

For purposes of this dissertation, I opted to use DIC for model selection. Current scientific literature employing Bayesian inference often uses DIC, which makes no assumption about the “true” model being in the model set (as is the case with BIC). One nuance with DIC, however, is that one must specify a point estimate with which to calculate deviance. In their original derivation of DIC, Spiegelhalter et al. (2002) used the posterior mean as the point estimate to calculate DIC; however,

one could also contemplate use of the posterior mode (see Section 3.3).

## 2.3 Efficient MCMC Candidate Generation

With the exception of random effects precision terms, the full conditional posterior distributions for model parameters do not belong to recognizable families of probability density functions (e.g., normal, gamma, etc.). Therefore, it is not possible to simulate parameter values directly as in traditional Gibbs sampling. One possibility for simulating these parameters is to use the so called Metropolis-within-Gibbs hybrid update (e.g., Brooks, 1999; Gelman et al., 2004). A typical approach is to cyclically update model parameters one at a time, with a single iteration’s proposed parameter value determined by generating a random variate from a symmetric probability density function centered at the previous iteration’s value, and a scale parameter chosen so as to achieve a desired acceptance rate. This is often referred to as a “random walk” chain or simply a “Metropolis” proposal scheme (Gelman et al., 2004). A proposal for parameter  $i$ ,  $\theta_i^*$ , at iteration  $t$  is accepted with probability dependent on the Metropolis ratio, i.e.,

$$\min\left(1, \frac{P(\theta_i^*|\mathbf{Y}, \boldsymbol{\theta}_{-i}^t)}{P(\theta_i^{t-1}|\mathbf{Y}, \boldsymbol{\theta}_{-i}^t)}\right)$$

Here,  $P(\theta_i^*|\mathbf{Y}, \boldsymbol{\theta}_{-i}^t)$  denotes the conditional posterior distribution of  $\theta_i$  given data ( $\mathbf{Y}$ ) and the remaining parameters of the model ( $\boldsymbol{\theta}_{-i}^t$ ) evaluated at the proposed parameter value  $\theta_i^*$ . Typically, the vector  $\boldsymbol{\theta}_{-i}^t$  will consist of  $[\theta_1^t, \theta_2^t, \dots, \theta_{i-1}^t, \theta_{i+1}^{t-1}, \dots, \theta_P^{t-1}]$ ,

where  $P$  denotes the total number of parameters in the model. One may also consider randomly changing the order that parameters are updated at each iteration, but I did not implement that here.

While the Metropolis proposal has been shown to work well in Bayesian analysis of mark-recovery and mark-recapture data alone (Brooks, 1999), the addition of age-at-harvest data into the likelihood presents additional challenges. Abundance is an integer value, subject to a number of constraints; thus, proposals must also be integers and the Metropolis ratio must be corrected for asymmetries resulting from these constraints. One possibility for generating univariate proposals for abundance, which I adopt throughout, involves the following algorithm:

1. Generate

$$\theta_i^* \sim \frac{\text{Normal}(\theta_i^{(t)}, \sigma_i^2)}{\int_{\theta_i^{min}-0.5}^{\theta_i^{max}+0.5} \text{Normal}(\theta_i^{(t)}, \sigma_i^2)}$$

for  $(\theta_i^{min} - 0.5) \leq \theta_i^* < (\theta_i^{max} + 0.5)$  by rejection sampling.

2. Round  $\theta_i^*$  to the nearest integer
3. Calculate Metropolis-Hastings ratio as

$$\frac{P(\theta_i^* | \mathbf{Y}, \boldsymbol{\theta}_{-i}^t) \int_{\theta_i^{min}-0.5}^{\theta_i^{max}+0.5} \text{Normal}(\theta_i^t, \sigma_i^2)}{P(\theta_i^{t-1} | \mathbf{Y}, \boldsymbol{\theta}_{-i}^t) \int_{\theta_i^{min}-0.5}^{\theta_i^{max}+0.5} \text{Normal}(\theta_i^t, \sigma_i^2)}$$

Here,  $\theta_i^{min}$  and  $\theta_i^{max}$  give the minimum and maximum values that are permissible given the constraints (with  $\theta_i^{max} = \infty$  if there is no upper constraint).

Addition of abundance into the likelihood has large ramifications for the convergence of Markov chains, as well as the amount of iterations required to reasonably



represent features of the posterior distribution. In particular, when the quantity of mark-recovery data is limited, or when one considers models for survival or recovery rate that are sufficiently complex, the Metropolis-within-Gibbs updating procedure requires a large number of iterations to accurately represent the posterior distribution (e.g., 1 to 2 million), making large scale simulation experiments difficult.

In this section, I consider the efficacy of several candidate proposal schemes which may help with the “mixing” ability of Markov chains, and thus for faster exploration of the posterior distribution. Because there is a high degree of sampling covariance between parameters, autocorrelation between successive values of the Markov chain is quite high. I hoped to reduce the degree of autocorrelation by devising proposal schemes that accounted for sampling covariance. In addition to comparing the Seber and Brownie model parameterizations, I thus explored the effects of generating multivariate Metropolis-Hastings proposals for vectors of parameters.

### *Joint proposals*

The abundance of a cohort at discrete points in time are considered individual parameters in the joint age-at-harvest mark-recovery model. However, they are inextricably linked by the process of survival. Thus, in addition to the numerous constraints that rule out certain combinations of parameter values (such as abundance of 2-year-olds being higher in year 2 than the number of 1-year-olds in year 1, for instance), we also expect an underlying symmetry; for example, if abundance in year 1 is “high,” abundance is also likely to be “high” in later years.

One possibility for accelerating convergence in this case is to use a nondiagonal

proposal covariance matrix for abundance parameters that are in the same cohort to better “mimic the posterior surface” (Carlin and Louis, 2000, pg. 157). In order to get a proposal covariance matrix that fulfills this purpose, one can first obtain an estimate of the posterior correlation structure by estimating posterior covariances from a traditional Metropolis-within-Gibbs updating scheme. A multivariate normal distribution can then be used to generate proposals, where

$$\begin{bmatrix} \theta_1^* \\ \theta_2^* \\ \vdots \\ \theta_k^* \end{bmatrix} \sim MVNormal \left( \begin{bmatrix} \theta_1^{t-1} \\ \theta_2^{t-1} \\ \vdots \\ \theta_k^{t-1} \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_k^2 \end{bmatrix} \right)$$

Gelman et al. (2004) suggest setting proposal variances such that they yield acceptance rates of 0.4 when parameters are altered one at a time, and 0.2 when updating vectors of parameters. We thus may perturb the proposal variances (the  $\sigma_i^2$ ) to achieve this goal, while setting

$$\sigma_{ij} = \hat{\rho}_{ij} \sigma_i \sigma_j$$

to preserve the correlation structure (here,  $\hat{\rho}_{ij}$  gives the estimated posterior correlation between parameters  $i$  and  $j$  from the standard Metropolis-within-Gibbs update).

### 2.3.1 Adaptive algorithms

As mentioned, Gelman et al. (2004) recommend that proposal acceptance rates be near 0.4 for univariate updates, and 0.2 for multivariate updates. It is possible to obtain these rates via trial and error (that is, sequential iteration of proposal standard deviations), but ultimately this strategy is not time efficient. Instead, Gelman et al. (2004) recommend implementing some sort of tuning algorithm to adjust proposal standard deviations to yield appropriate acceptance rates before the final phase of MCMC estimation.

A natural way to construct such an algorithm is to monitor the number or proposal acceptances that are made over a certain window, and then to adjust the proposal standard deviation upward if too many are accepted, and downward if too many are rejected. In practice, I found that a large window was necessary to properly tune model parameters, as too short of a window could lead to some proposal standard deviations tending towards 0, and thus a “stuck” Markov Chain. This was likely due to the nature of the dynamic, highly correlated system that was under consideration. Using a downward multiplier of 0.95, an upward multiplier of 1.0526 ( $1/0.95$ ), and a window of 1000 iterations seemed sufficient, with perhaps 30 such windows being required to reach targeted acceptance rates for all parameters using the standard Metropolis within Gibbs update. For multivariate normal proposals, I simply multiplied the entire proposal variance covariance matrix  $\Sigma$  by a downward or upward multiplier; this procedure preserved correlation structure while allowing acceptance rates to be adjusted to somewhere near 0.2.

### 2.3.2 Mixing properties of different proposal schemes

The surface of the posterior distribution may differ markedly depending on model parameterization. For instance, a model with survival and reporting rate varying by the same factor (e.g., age or time) will generally result in higher posterior correlation between these parameters than a model in which  $S$  and  $r$  vary by different factors (e.g., survival varies by age but reporting rate varies by time). We thus might expect that the efficiency of different proposal schemes will vary by estimation model. Rather than conduct an exhaustive study of a large number of estimation models, I selected two estimation models that might be “useful” in practice. In particular,  $S(\cdot)r(\cdot)f(\cdot)$  denotes the model where survival, reporting rate, and recruitment rate are constant over time. While of limited biological plausibility (survival will almost certainly vary by age, for instance), this model may be all that some data sets can support, at least given uninformative priors. Similarly,  $S(a+t)r(a)f(\cdot)$  specifies a model with additive time and age effects on survival (i.e.,  $\text{logit}(S_{ij}) = \alpha_i + \beta_j$ ), age effects on reporting rate, and constant recruitment rate. In contrast to the “dot” model, this model may represent a level of complexity realistic for rich data sets.

In order to best compare different proposal schemes, it was instructive to conduct analyses on the same data set, varying only the proposal scheme. However, I considered different data sets for each estimation model in order to represent sparse and rich data sets for models  $S(\cdot)r(\cdot)f(\cdot)$  and  $S(a+t)r(a)f(\cdot)$ , respectively. In both cases, I assumed 5 distinguishable age classes, 5 years of data, a survival rate of 0.6, and a reporting rate of 0.2. I used expected (rather than simulated) data, rounding

harvest numbers to the nearest integer while keeping the frequencies of observed mark-recovery histories as rational numbers. Each population was assumed to be stationary ( $\lambda=1.0$ ) and to have a stable stage distribution (Caswell, 2001).

For the simple model  $S(\cdot)r(\cdot)f(\cdot)$  I used an initial population size of 1000 individuals, and assumed that 20 individuals were captured, marked, and released per year, partitioned into the 5 age classes depending on the proportion of individuals that were of a given age. For model  $S(a+t)r(a)f(\cdot)$ , I used an initial population size of 5000 and assumed that 300 individuals were released per year. For both models, a constant, improper prior was specified for abundance parameters (i.e.,  $P(N_{1j}) \propto c$ ), and priors for  $\mathbf{S}$ ,  $\mathbf{r}$ , and  $\mathbf{f}$  were specified as in section 2.2.3.

Construction of an efficient proposal scheme may ultimately rely on knowledge about the shape of the posterior distribution. Examination of the posterior surface often revealed large correlations between certain model parameters, suggesting a particular reparameterization or multivariate proposal scheme. For instance, one sample from the posterior (summarized from 200,000 iterations) for model  $S(\cdot)r(\cdot)f(\cdot)$  exhibited correlations of 0.94 - 0.97 for consecutive abundance parameters (e.g.,  $N_{11}$  and  $N_{22}$ ), with correlations between  $\text{logit}(r)$  and abundance reaching -0.9 for age 0 cohorts, decreasing to the -0.4 to -0.7 range as age class increased. This was likely due to there being more mark-recovery data for earlier ages since the number released per age class in a given year was proportional to the abundance of that age class. Correlation between  $\text{logit}(S)$  and other parameters was comparably low (0.0 to 0.6), including a correlation between  $\text{logit}(S)$  and  $\text{logit}(r)$  of 0.26. This last relationship was surprising given that high posterior correlation between survival

and reporting rate have been reported in the context of mark-recovery models (e.g., Barry et al., 2003). Ostensibly, the correlation between these parameters has been reduced because of the even higher dependence between abundance and reporting rate.

In contrast, the posterior surface associated with model  $S(a+t)r(a)f(\cdot)$  revealed only small correlations between consecutive cohort abundances ( $0 < \rho < 0.3$ ). However, correlations were high between same-age cohorts across time. For example, the posterior correlation between  $N_{11}$  and  $N_{21}$  was substantial (0.88). This relationship likely results from the specified model for reporting rate, which varies by age in this case. In addition to correlations between abundances, there were also substantial correlations between age-specific beta parameters for reporting rate and survival. The linear model for survival was formulated as

$$\begin{bmatrix} \text{logit}(S_{11}) \\ \text{logit}(S_{12}) \\ \text{logit}(S_{13}) \\ \text{logit}(S_{14}) \\ \text{logit}(S_{15}) \\ \text{logit}(S_{21}) \\ \text{logit}(S_{22}) \\ \text{logit}(S_{23}) \\ \text{logit}(S_{24}) \\ \text{logit}(S_{25}) \\ \text{logit}(S_{31}) \\ \text{logit}(S_{32}) \\ \text{logit}(S_{33}) \\ \text{logit}(S_{34}) \\ \text{logit}(S_{35}) \\ \text{logit}(S_{41}) \\ \text{logit}(S_{42}) \\ \text{logit}(S_{43}) \\ \text{logit}(S_{44}) \\ \text{logit}(S_{45}) \\ \text{logit}(S_{51}) \\ \text{logit}(S_{52}) \\ \text{logit}(S_{53}) \\ \text{logit}(S_{54}) \\ \text{logit}(S_{55}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \beta_0^S \\ \beta_{t_1}^S \\ \beta_{t_2}^S \\ \beta_{t_3}^S \\ \beta_{t_4}^S \\ \beta_{a_1}^S \\ \beta_{a_2}^S \\ \beta_{a_3}^S \\ \beta_{a_4}^S \end{bmatrix}$$

while the linear model for reporting rate was formulated as

$$\begin{bmatrix} \text{logit}(r_{11}) \\ \text{logit}(r_{12}) \\ \text{logit}(r_{13}) \\ \text{logit}(r_{14}) \\ \text{logit}(r_{15}) \\ \text{logit}(r_{21}) \\ \text{logit}(r_{22}) \\ \text{logit}(r_{23}) \\ \text{logit}(r_{24}) \\ \text{logit}(r_{25}) \\ \text{logit}(r_{31}) \\ \text{logit}(r_{32}) \\ \text{logit}(r_{33}) \\ \text{logit}(r_{34}) \\ \text{logit}(r_{35}) \\ \text{logit}(r_{41}) \\ \text{logit}(r_{42}) \\ \text{logit}(r_{43}) \\ \text{logit}(r_{44}) \\ \text{logit}(r_{45}) \\ \text{logit}(r_{51}) \\ \text{logit}(r_{52}) \\ \text{logit}(r_{53}) \\ \text{logit}(r_{54}) \\ \text{logit}(r_{55}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \beta_0^r \\ \beta_{a_1}^r \\ \beta_{a_2}^r \\ \beta_{a_3}^r \\ \beta_{a_4}^r \end{bmatrix}.$$

Large correlations existed between parameters  $\beta_{a_1}^S$  and  $\beta_{a_1}^r$  (0.80),  $\beta_{a_2}^S$  and  $\beta_{a_2}^r$  (0.86),  $\beta_{a_3}^S$  and  $\beta_{a_3}^r$  (0.87), and  $\beta_{a_4}^S$  and  $\beta_{a_4}^r$  (0.90), ostensibly because survival and reporting rate both varied by a shared factor (age) in these models. There were also reasonably large correlations between time effects for survival, all of which were between 0.82 and 0.94.

I considered the following proposal schemes for model  $S(\cdot)r(\cdot)f(\cdot)$ :

- Independent updates using the Seber parameterization for recoveries (IUS)
- Independent updates using the Brownie parameterization for recoveries (IUB)
- Multivariate normal updates on cohort-specific abundance (MVN1)



This selection was made primarily to compare the traditional independent update (IUS) with the multivariate update that seeks to exploit information about the shape of the posterior distribution (MVN1). This proposal scheme involved jointly updating 12 parameter groups:

$$[N_{11}, N_{22}, N_{33}, N_{44}, N_{55}]$$

$$[N_{12}, N_{23}, N_{34}, N_{45}, N_{56}]$$

$$[N_{13}, N_{24}, N_{35}, N_{46}]$$

$$[N_{14}, N_{25}, N_{36}]$$

$$[N_{15}, N_{26}]$$

$$[N_{21}, N_{32}, N_{43}, N_{54}]$$

$$[N_{31}, N_{42}, N_{53}]$$

$$[N_{41}, N_{52}]$$

$$[N_{51}]$$

$$[\text{logit}(S)]$$

$$[\text{logit}(r)]$$

$$[\log(f)].$$

I did not expect much difference in performance between the (IUS) and (IUB) updates because of the low correlation between  $\text{logit}(S)$  and  $\text{logit}(r)$ , but included (IUB) to afford contrasts between the efficiency of proposal schemes for different models (e.g., to  $S(a+t)r(a)f(\cdot)$ ).

The following proposal schemes were considered for model  $S(a+t)r(a)f(\cdot)$ :

- Independent updates using the Seber parameterization for recoveries (IUS)

- Independent updates using the Brownie parameterization for recoveries (IUB)
- Multivariate normal updates on age-specific abundance using the Brownie parameterization for recoveries (MVN2)
- Multivariate normal updates on age-specific abundance using the Brownie parameterization for recoveries with a multivariate normal update on beta parameters for year effects on survival (MVN3)

In this case, I guessed the Brownie parameterization would result in better mixing because of the large posterior correlations between age effects of the logits of survival and reporting rate. However, it is important to note that the models  $S(a+t)r(a)f(\cdot)$  and  $S(a+t)h(a)f(\cdot)$  do not provide the same inference since recovery rate  $h$  should actually be a function of both age and time if the substitution  $h = (1 - S)r$  were directly made. Nevertheless, choices like these may sometimes be necessary.

The MVN2 and MVN3 proposal schemes required multivariate updates with respect to the following blocks of parameters:

$$[N_{11}, N_{21}, N_{31}, N_{41}, N_{51}]$$

$$[N_{12}, N_{22}, N_{32}, N_{42}, N_{52}]$$

$$[N_{13}, N_{23}, N_{33}, N_{43}, N_{53}]$$

$$[N_{14}, N_{24}, N_{34}, N_{44}, N_{54}]$$

$$[N_{15}, N_{25}, N_{35}, N_{45}, N_{55}]$$

$$[N_{16}, N_{26}, N_{36}, N_{46}, N_{56}].$$

The remaining parameters associated with survival, reporting rate, and recruitment were updated independently in the case of MVN2, while for MVN3, parameters

Table 2.2: Mean lag-200 autocorrelations for total abundance in year  $i$  ( $N_i$ ) and accompanying standard errors for different proposal schemes and likelihood parameterizations (“Est Type”). Table 2.2A gives results for model  $S(\cdot)r(\cdot)f(\cdot)$ , while Table 2.2B. gives results for model  $S(a+t)r(a)f(\cdot)$ . Data sets were different for the two models; details are provided in the text.

Est Type	Lag-200 Autocorrelation				
	$N_1$ .	$N_2$ .	$N_3$ .	$N_4$ .	$N_5$ .
<b>A.</b>					
IUS	0.91 (0.01)	0.94 (0.01)	0.95 (0.01)	0.94 (0.01)	0.90 (0.01)
IUB	0.90 (0.01)	0.93 (0.01)	0.95 (0.01)	0.94 (0.01)	0.91 (0.01)
MVN1	0.91 (0.01)	0.93 (0.01)	0.93 (0.01)	0.93 (0.01)	0.90 (0.01)
<b>B.</b>					
IUS	0.79 (0.01)	0.81 (0.01)	0.82 (0.01)	0.86 (0.01)	0.87 (0.01)
IUB	0.64 (0.01)	0.76 (0.01)	0.78 (0.01)	0.78 (0.01)	0.73 (0.01)
MVN2	0.71 (0.01)	0.80 (0.01)	0.81 (0.01)	0.80 (0.01)	0.74 (0.01)
MVN3	0.73 (0.01)	0.81 (0.01)	0.83 (0.01)	0.81 (0.01)	0.74 (0.01)

$$[\beta_{t_1}^S, \beta_{t_2}^S, \beta_{t_3}^S, \beta_{t_4}^S]$$

were updated as a group.

I compared the performance of each proposal scheme by monitoring autocorrelation of total annual abundance in year  $i$  ( $N_i$ ), and by examining typical sample paths. For each proposal scheme, I conducted 10 separate MCMC data analyses, using 200,000 iterations for each, starting each chain off at true initial values, and using a burn-in of 20,000 iterations to perturb Markov Chains from their starting values. I then computed the mean lag-200 autocorrelation across analyses for each year’s abundance and accompanying standard errors (Table 2.2).

For model  $S(\cdot)r(\cdot)f(\cdot)$ , all samplers produced roughly the same autocorrelation in annual abundances. This suggested that an investigator should base their selection of samplers on which parameterization is most pragmatic for their purposes, or perhaps on which sampler takes the least computing time. Computing

time is most directly related to the number of log likelihood evaluations that are made; when multiple parameters are updated at once (as with MVN1), the number of such evaluations are reduced. However, this procedure also requires extra time to summarize posterior correlations and adjust acceptance rates prior to initialization of the final estimation run. Thus a choice based on computing time ultimately relies on the number of iterations it takes to satisfactorily summarize the posterior distribution, a topic covered in the next section.

In contrast to the dot model, better sampler performance was obtained when the Brownie parameterization was employed for model  $S(a+t)r(a)f(\cdot)$ . This was likely because age effects were included on both survival and reporting rate. Inclusion of multivariate normal proposals appeared to negatively affect chain mixing, however, so that the IUB sampler seems the most efficient in this case.

Contrary to my prior expectations, at no time did multivariate normal proposals substantially help the mixing of Markov chains. Further, there appeared to be a tendency for some samplers to work better than others depending upon the estimation model considered. These considerations highlight the sometimes counterintuitive nature of the challenges that a Bayesian analyst faces when seeking to analyze highly dynamic, correlated systems such as this one.

### **2.3.3 Approximating the posterior: number of iterations**

The small experiment in the previous section provides guidelines on the relative efficiency of various sampling strategies, but does not provide any direct guide-

lines concerning the number of iterations required to accurately summarize features of the posterior distribution. In this section, I summarize the variation between multiple Markov Chains at different landmarks of simulation time for a number of design points, considering the standard Metropolis-within-Gibbs update (IUS) for model  $S(\cdot)r(\cdot)f(\cdot)$ , and the standard Metropolis-within-Gibbs update (IUB) for model  $S(a+t)h(a)f(\cdot)$ .

To start, I generated expected value data for a number of design points, rounding harvest numbers to the nearest integer. Data were simulated for 18 different design points for model  $S(\cdot)r(\cdot)f(\cdot)$  and for 12 design points for model  $S(a+t)h(a)f(\cdot)$ . Possible configurations depended on the model; for model  $S(\cdot)r(\cdot)f(\cdot)$  configurations included

- True abundance, number of releases/year = (1000,100), (1000,20), or (200,20)
- Reporting rate ( $r$ ) = 0.2 or 0.5
- Number of years ( $Y$ ), cohorts ( $A$ )= (3,3), (5,3), or (5,5)

while for model  $S(a+t)h(a)f(\cdot)$  they included

- True abundance, number of releases/year = (5000,150), (5000,300)
- Recovery rate ( $h$ ) = 0.08 or 0.2
- Number of years ( $Y$ ), cohorts ( $A$ )= (3,3), (5,3), or (5,5)

All expected value data sets assumed a time- and cohort-constant survival probability of 0.6, a stable stage distribution (Caswell, 2001) and a stable population

( $\lambda=1.0$ ). For each design point and candidate proposal scheme, I employed 10 sets of Markov chains to approximate features of the posterior distribution. I was interested in the similarity of posterior distributions approximated by independent Markov chains of a specified length; that is, how repeatable results were as a function of chain length. I quantified “similarity” by calculating a) the average standard deviation of the posterior mean of total abundance (where the standard deviation is calculated across Markov chains for each year of the study and then averaged across years), b) the average standard deviation of the posterior mode of total abundance, c) the standard deviation of the deviance information criterion (DIC) when the posterior mean is used to calculate DIC, and d) the standard deviation of DIC as calculated with the posterior mode. Each chain was run for 2.1 million iterations, with the first 100,000 iterations discarded as a burn-in and the chain thinned by recording only 1 in 10 observations to reduce memory requirements. Markov chains were then summarized at  $1.0 \times 10^5$ ,  $5.0 \times 10^5$ ,  $1.0 \times 10^6$ ,  $1.5 \times 10^6$  and  $2.0 \times 10^6$  iterations.

Several observations may be made based on results of this experiment (Table 2.3). First, the number of iterations required to reliably summarize the posterior distribution decreases as the quantity of mark-recovery data increases. This could occur either through inclusion of additional years of data, increased reporting rates, or increased numbers of initial releases. Number of age classes did not appear to have much of an effect on convergence rates, at least with the range of experimental inputs considered here.

A second issue that arose had to do with sparse mark-recovery data. In particular, when numbers of releases and reporting rates were low (e.g.,  $R=20$  and  $r=0.2$ ),

the variance in posterior estimates of abundance from different Markov chains was extremely high even after 2.1 million iterations. Apparently, more data are required to get sensible estimates in these cases, or perhaps stronger priors.

In general, posterior estimates of abundance tended to vary less when using the posterior mean instead of the posterior mode, and this was also the case with computation of DIC. This computational feature is likely to be a general one given the extra error involved in approximating posterior modes from KD estimates. For example, even after 2.1 million iterations with relatively high sample sizes and sampling rates it would not be surprising to have an estimated posterior mode about 1% different from its true value, even if expected value data were analyzed.

Table 2.3: Performance of Markov Chains in summarizing marginal posterior distributions for animal abundance at different landmarks of simulation time for models  $S(\cdot)r(\cdot)f(\cdot)$  and  $S(a+t)h(a)f(\cdot)$ . Tables 2.3A and 2.3B give the average between chain standard deviation of annual abundance as determined by the posterior mean and mode, respectively. Tables 2.3C and 2.3D give DIC values for cases where the DIC is computed relative to the posterior mean and mode, respectively.  $N$  gives initial population abundance,  $R$  gives the number of releases of marked animals per year, and “Model” specifies whether results pertained to estimation model  $S(\cdot)r(\cdot)f(\cdot)$  (“Model”=1), or  $S(a+t)h(a)f(\cdot)$  (“Model”=2). Remaining notation is defined Table 2.1.

$A$	$Y$	$N$	$R$	$r$	Model	MCMC Iterations ( $\times 10^6$ )				
						1	5	10	15	20
<b>A.</b>										
3	3	1000	100	0.2	1	29.2	9.4	6.8	5.2	4.4
3	3	1000	100	0.5	1	11.2	4.7	3.0	2.8	2.2
3	3	1000	20	0.2	1	870.0	272.0	111.3	86.6	73.2
3	3	1000	20	0.5	1	51.1	18.5	16.5	15.9	14.9
3	3	200	20	0.2	1	51.4	27.0	21.1	31.3	23.4
3	3	200	20	0.5	1	7.1	3.0	1.7	1.4	1.0
3	5	1000	100	0.2	1	13.7	8.0	6.9	4.6	3.6
3	5	1000	100	0.5	1	5.5	2.5	1.6	1.1	1.2
3	5	1000	20	0.2	1	87.0	38.1	22.8	24.1	18.8
3	5	1000	20	0.5	1	24.4	10.2	7.0	7.5	5.3
3	5	200	20	0.2	1	7.9	5.1	2.7	2.2	2.0
3	5	200	20	0.5	1	2.6	1.2	0.76	0.61	0.4
5	5	1000	100	0.2	1	14.6	7.8	4.6	3.9	3.5
5	5	1000	100	0.5	1	5.2	2.4	1.1	1.2	1.3
5	5	1000	20	0.2	1	141.8	72.8	52.9	49.5	44.1
5	5	1000	20	0.5	1	27.1	11.2	6.5	8.1	6.4
5	5	200	20	0.2	1	21.0	8.7	3.0	3.1	3.0
5	5	200	20	0.5	1	3.3	1.2	0.9	0.8	0.5
3	3	5000	150	0.2	2	210.9	71.1	34.2	32.7	44.8
3	3	5000	150	0.5	2	25.2	12.8	10.7	10.1	9.4
3	3	5000	300	0.2	2	81.7	28.1	29.3	25.8	18.9
3	3	5000	300	0.5	2	27.7	11.8	8.8	5.7	4.9
3	5	5000	150	0.2	2	115.0	69.9	45.2	34.5	27.0
3	5	5000	150	0.5	2	19.5	12.9	13.6	9.0	7.6
3	5	5000	300	0.2	2	39.8	13.2	15.9	10.3	8.0
3	5	5000	300	0.5	2	11.5	7.9	5.1	3.8	3.5



$A$	$Y$	$N$	$R$	$r$	Model	MCMC Iterations ( $\times 10^6$ )				
						1	5	10	15	20
5	5	5000	150	0.2	2	111.7	39.7	45.3	22.1	17.9
5	5	5000	150	0.5	2	22.4	14.1	10.9	5.5	2.2
5	5	5000	300	0.2	2	67.7	20.4	12.9	14.1	9.5
5	5	5000	300	0.5	2	10.0	5.5	2.9	3.1	2.1
<b>B.</b>										
3	3	1000	100	0.2	1	40.1	18.8	16.8	13.2	13.4
3	3	1000	100	0.5	1	14.0	9.9	9.0	6.4	5.8
3	3	1000	20	0.2	1	187.1	90.0	66.8	61.1	49.4
3	3	1000	20	0.5	1	51.2	28.8	22.9	13.8	13.7
3	3	200	20	0.2	1	39.5	21.2	14.7	12.2	9.9
3	3	200	20	0.5	1	8.1	5.0	4.6	4.8	4.1
3	5	1000	100	0.2	1	25.2	11.6	9.2	7.6	7.2
3	5	1000	100	0.5	1	8.6	4.7	4.4	4.5	3.9
3	5	1000	20	0.2	1	57.5	44.3	42.0	42.3	37.9
3	5	1000	20	0.5	1	31.7	13.5	10.4	10.1	10.0
3	5	200	20	0.2	1	16.6	9.0	7.2	4.7	4.7
3	5	200	20	0.5	1	4.6	3.3	2.9	2.8	3.0
5	5	1000	100	0.2	1	22.0	13.8	10.3	9.2	7.8
5	5	1000	100	0.5	1	10.2	5.1	4.0	4.0	4.1
5	5	1000	20	0.2	1	96.0	47.0	51.1	59.4	59.5
5	5	1000	20	0.5	1	33.3	20.3	15.0	13.7	14.1
5	5	200	20	0.2	1	18.2	10.2	6.8	6.7	5.3
5	5	200	20	0.5	1	5.3	3.0	2.4	2.6	1.7
3	3	5000	150	0.2	2	312.8	113.4	75.9	56.1	54.2
3	3	5000	150	0.5	2	48.3	30.4	19.0	23.7	16.1
3	3	5000	300	0.2	2	122.5	58.9	45.9	39.9	32.0
3	3	5000	300	0.5	2	53.9	23.5	21.5	15.0	16.0
3	5	5000	150	0.2	2	191.6	73.2	43.3	39.8	34.7
3	5	5000	150	0.5	2	40.1	25.8	23.6	22.1	17.6
3	5	5000	300	0.2	2	84.5	37.8	30.4	26.2	24.4
3	5	5000	300	0.5	2	27.4	17.1	18.3	15.7	17.9
5	5	5000	150	0.2	2	132.8	97.7	80.7	36.7	42.7
5	5	5000	150	0.5	2	50.4	27.9	28.7	22.8	19.5
5	5	5000	300	0.2	2	85.4	41.2	33.7	30.4	25.3
5	5	5000	300	0.5	2	30.4	19.9	16.3	15.9	16.0

<i>A</i>	<i>Y</i>	<i>N</i>	<i>R</i>	<i>r</i>	Model	MCMC Iterations ( $\times 10^6$ )				
						1	5	10	15	20
<b>C.</b>										
3	3	1000	100	0.2	1	0.3	0.1	0.1	0.1	0.0
3	3	1000	100	0.5	1	0.3	0.1	0.1	0.1	0.1
3	3	1000	20	0.2	1	19.5	8.0	4.4	3.0	2.3
3	3	1000	20	0.5	1	0.6	0.4	0.3	0.2	0.2
3	3	200	20	0.2	1	2.6	1.3	1.4	1.5	0.7
3	3	200	20	0.5	1	0.4	0.2	0.3	0.3	0.2
3	5	1000	100	0.2	1	0.3	0.1	0.1	0.1	0.1
3	5	1000	100	0.5	1	0.3	0.2	0.1	0.0	0.1
3	5	1000	20	0.2	1	0.6	0.2	0.2	0.2	0.2
3	5	1000	20	0.5	1	0.4	0.2	0.2	0.1	0.1
3	5	200	20	0.2	1	0.3	0.2	0.1	0.1	0.1
3	5	200	20	0.5	1	0.3	0.2	0.2	0.1	0.1
5	5	1000	100	0.2	1	0.4	0.3	0.2	0.2	0.1
5	5	1000	100	0.5	1	0.3	0.2	0.1	0.0	0.1
5	5	1000	20	0.2	1	1.5	0.6	0.4	0.4	0.3
5	5	1000	20	0.5	1	0.7	0.4	0.2	0.2	0.2
5	5	200	20	0.2	1	0.9	0.4	0.2	0.3	0.2
5	5	200	20	0.5	1	0.3	0.3	0.2	0.1	0.2
3	3	5000	150	0.2	2	6.5	2.6	1.7	1.8	1.8
3	3	5000	150	0.5	2	0.4	0.2	0.1	0.1	0.1
3	3	5000	300	0.2	2	3.3	1.8	0.9	0.6	0.6
3	3	5000	300	0.5	2	0.5	0.2	0.1	0.1	0.1
3	5	5000	150	0.2	2	16.4	4.6	3.8	2.7	1.6
3	5	5000	150	0.5	2	0.4	0.3	0.2	0.2	0.2
3	5	5000	300	0.2	2	4.3	1.0	0.7	0.4	0.5
3	5	5000	300	0.5	2	0.2	0.2	0.1	0.1	0.1
5	5	5000	150	0.2	2	8.1	5.5	4.4	3.5	3.2
5	5	5000	150	0.5	2	0.6	0.2	0.2	0.2	0.1
5	5	5000	300	0.2	2	4.5	1.5	1.3	1.1	1.4
5	5	5000	300	0.5	2	0.4	0.2	0.2	0.1	0.1

A	Y	N	R	r	Model	MCMC Iterations ( $\times 10^6$ )					
						1	5	10	15	20	
<b>D.</b>											
3	3	1000	100	0.2	1	2.8	3.2	3.5	2.4	2.5	
3	3	1000	100	0.5	1	1.6	1.3	0.5	0.5	0.5	
3	3	1000	20	0.2	1	195.8	28.8	8.3	8.5	9.3	
3	3	1000	20	0.5	1	16.3	5.5	3.9	3.7	2.5	
3	3	200	20	0.2	1	13.1	13.2	12.1	9.1	2.8	
3	3	200	20	0.5	1	3.5	4.0	2.1	1.8	2.3	
3	5	1000	100	0.2	1	3.1	2.0	1.1	1.3	1.2	
3	5	1000	100	0.5	1	1.9	0.9	0.4	0.5	0.5	
3	5	1000	20	0.2	1	13.5	6.5	5.4	7.8	5.2	
3	5	1000	20	0.5	1	16.6	2.0	2.2	2.5	1.5	
3	5	200	20	0.2	1	10.5	8.7	3.0	4.2	3.4	
3	5	200	20	0.5	1	4.4	3.8	1.9	3.4	1.7	
5	5	1000	100	0.2	1	5.1	3.3	2.1	2.0	2.3	
5	5	1000	100	0.5	1	2.9	1.8	1.8	2.1	1.9	
5	5	1000	20	0.2	1	21.3	9.3	11.8	10.7	9.5	
5	5	1000	20	0.5	1	3.2	3.8	2.1	3.7	2.5	
5	5	200	20	0.2	1	7.4	7.4	4.9	5.0	4.0	
5	5	200	20	0.5	1	2.7	3.1	2.2	2.0	1.8	
3	3	5000	150	0.2	2	541.0	88.2	61.9	42.5	30.4	
3	3	5000	150	0.5	2	86.9	16.3	8.0	5.5	3.8	
3	3	5000	300	0.2	2	150.1	19.0	17.2	18.7	21.3	
3	3	5000	300	0.5	2	30.3	10.3	3.7	4.7	2.7	
3	5	5000	150	0.2	2	322.3	36.4	24.1	28.6	23.6	
3	5	5000	150	0.5	2	29.3	13.0	10.0	6.3	4.1	
3	5	5000	300	0.2	2	80.3	56.0	20.1	21.9	16.3	
3	5	5000	300	0.5	2	11.4	5.2	3.7	3.4	3.4	
5	5	5000	150	0.2	2	239.1	132.2	49.9	36.6	39.1	
5	5	5000	150	0.5	2	60.3	10.8	22.1	18.1	6.8	
5	5	5000	300	0.2	2	134.2	39.6	18.1	17.1	17.2	
5	5	5000	300	0.5	2	24.3	5.1	6.7	7.3	3.9	

## 2.4 Computing

I developed a C++ program, AGEHARV, to perform analyses. This program requires the use of three files: 1 file for encounter history data, 1 file for age-at-harvest data, and an input file which provides user specifications regarding MCMC, as well

as the particular formulation of model structure. Upon completion, AGEHARV outputs specific posterior summaries according to specifications in the input file. A simple user’s manual is provided in Appendix B; the executable and source code are available from the author.

## 2.5 Parameter Identification

The likelihood presented in section 2 is over-parameterized. In practice, several suggestions may help to solve this problem. For instance, in the field of fisheries stock assessment, an additive model has often been assumed for sampling parameters (the “separability assumption”; Megrey, 1989). Similarly, additive and/or random effects models have been proposed for survival and recruitment parameters in age-structured models for whooping cranes (Link et al., 2003). Nonetheless, even if model simplifications are made, it is still desirable to identify parameter redundancies or situations in which there are not enough data to estimate a given parameter. In the context of analyzing data from marked individuals, the issue of parameter identification has been previously explored using a combination of analytical tools and numerical methods.

For product multinomial models, Catchpole and Morgan (1997) identified a procedure for determining which parameters are identifiable. This approach involves calculating a partial derivative matrix of the log of non-zero multinomial cell probabilities. The symbolic rank of this matrix (as calculated in MAPLE for example) gives the number of identifiable parameters. If the matrix is rank deficient,

this indicates the level of parameter redundancy (see also Giminez et al., 2004). Using this approach, it is possible to determine whether survival and reporting parameters are theoretically identifiable. The addition of the “+” age group makes determining parameter identification from the age-at-harvest portion of the likelihood analytically intractable, although it is possible if this class is omitted (Gove, 1997). However, if survival and reporting rate are identifiable, it should be possible to estimate abundance and recruitment parameters, since the distribution of abundances are unimodal given survival and reporting rate in the age-at-harvest portion of the likelihood.

Just because a parameter is theoretically identifiable does not guarantee that one can practically estimate it, even if there are a plethora of data. Thus, in addition to aforementioned analytic tools, numerical methods have been developed to identify singularities in the estimated variance-covariance matrix via a singular value decomposition (e.g., Cooch and White, 2005). In the case of multi-state mark-recapture models, this approach has been successfully applied to detect cases where parameters associated with unobservable states may be estimated (Kendall and Nichols, 2002).

At first glance, adopting a Bayesian perspective might seem to alleviate some problems related to parameter identification; after all, specification of proper prior distributions on model parameters leads to proper posterior distributions and thus to some form of an “answer.” However, in the end arguments of this type fail because the only information about certain parameters may come from the prior distribution. Indeed, in a Bayesian context, a parameter can be said to be non-identifiable

if the marginal posterior distribution for a given parameter is equal to the prior distribution (Gustafson, 2005). In practice, we might expect a slightly less rigid definition to suffice, as when there are some data to inform the estimation of a given parameter, but not enough to overcome the prior, or when functions of 2 or more parameters are identifiable.

K. P. Burnham (*Personal Communication*) has suggested an interesting approach, which would involve comparing posterior distributions resulting from successively larger and larger datasets. For cases in which functions of 2 or more non-identifiable parameters are estimable, posterior distributions may differ markedly from prior distributions. However, the posterior for these parameters should be relatively insensitive to sample size. Doubling or even quadrupling a dataset and comparing marginal posterior distributions may provide a way to diagnose nonidentifiable parameters in this case.

## 2.6 Goodness-of-Fit

In any statistical analysis, care should be taken that the model(s) under consideration actually fit the data. Under the Bayesian paradigm, such checks often take the form of a Bayesian p-value ( $p_B$ ; Gelman et al., 2004), in which a test statistic is defined and compared between actual data and samples from the posterior predictive distribution of replicated data. It differs from the traditional p-value in that test statistics depend on the posterior distribution of parameters rather than a fixed vector of estimated parameters (as with MLE's). In practice, the following steps

may be used to approximate a Bayesian p-value for a given model (Gelman et al., 2004):

1. Obtain  $i = 1$  to  $L$  samples of  $\boldsymbol{\theta}$  from their posterior distribution
2. For each such sample,
  - (a) Simulate a new data set,  $y_i^{rep}$  using the current values of  $\boldsymbol{\theta}_i$
  - (b) Compute  $T(y_i^{rep}, \boldsymbol{\theta}_i)$ , the test statistic for the  $i$ th simulated dataset
  - (c) Compute  $T(y, \boldsymbol{\theta}_i)$ , the test statistic for the real dataset
3. Calculate  $p_B$  as the proportion of samples for which the the simulated data produce a test statistic value more extreme than the observed data.

Values of  $p_B$  that approach either zero or one indicate some lack of fit.

Defining a reasonable, omnibus test statistic to use for goodness-of-fit has proven to be one of the more vexing challenges in the analysis of mark-recapture and mark-recovery data, and is not necessarily alleviated by adopting a Bayesian perspective. For instance, expected multinomial cell frequencies associated with many mark-recapture histories are often quite low, which tends to inflate deviance and chi-squared based test statistics whenever an unlikely observation is obtained. Pooling data may be one solution to this issue, but how exactly this is done is quite subjective and likely to affect interpretation. Alternatively, some authors have employed the Freeman-Tukey statistic (cf. Brooks et al., 2000),

$$D(\mathbf{y}|\boldsymbol{\theta}) = \sum_j (\sqrt{y_j} - \sqrt{e_j})^2,$$

where  $y_j$  and  $e_j$  denote observed and expected cell frequencies. However, while this discrepancy measure does not include small terms in its denominator, it has not as of yet (to my knowledge) been evaluated rigorously as a measure for examining goodness-of-fit in mark-recapture or recovery models.

In contrast to mark-recovery data, age-at-harvest data may yield higher expected (and observed) cell frequencies, making goodness-of-fit testing more fool-proof. For the purposes of this dissertation, then, I only consider goodness-of-fit with regard to posterior predictive simulation of age-at-harvest data. To this end I use deviance as a test statistic whenever goodness-of-fit is assessed with a Bayesian p-value. The utility of this approach is examined more in Simulation Module 3, described in section 7 below.

## 2.7 Simulation Study

While estimators of abundance appear to converge if Markov chains are run for long enough, this does not preclude bias or guarantee estimators with good properties. Thus, I used simulation to investigate estimator performance under a variety of hypothetical scenarios. In total, I considered 4 simulation modules to summarize estimator performance and diagnostics. In the first module, I evaluated bias, coefficient of variation (CV) and 90% Bayesian credible interval coverage (BCOV) for a variety of models, parameter combinations, number of age classes, years of data, and number of individuals marked per year. In the second module, I examined the performance of several estimators when there were errors in age determination.



In the third module, I examined the efficacy of Bayesian p-values for diagnosing overdispersion. Finally, in the fourth module, I examine consequences of using data from marked animals in both portions (age-at-harvest and mark-recovery) of the likelihood. I discuss each of these modules in further detail below.

### **2.7.1 Simulation Module I: Large Sample Performance**

In the first simulation module, my goal was to quantify estimator performance when model assumptions were perfectly satisfied and when enough data were available to avoid issues with parameter estimability. I thus assumed combinations of initial population sizes, number of marked individuals, and estimation models such that estimation models included a reasonable level of complexity for the given values of population size and number of marked animals. I further assumed that marked individuals were not part of the population being estimated (that is, they were effectively from a similar but different population). An additional consideration was to minimize computing time, as with input configurations that would not require a great number of iterations to run. For instance, if only 20 individuals are marked per year, it is unlikely that parameters will be well estimated no matter how simple the model structure is (Table 2.3). A complete representation for each combination of initial population size, number of individuals marked per year, and estimation model are presented in Table 2.4. For simplicity, and also for potential reduction in computing time, the IUB update was used for all estimation models. I generated data with values for  $\lambda$  of 0.9 and 1.0, values for  $S$  of 0.6 or 0.8, and values for  $r$  of

Table 2.4: Combinations of initial population size, number of individuals marked and released per year, and estimation model that are considered in simulation module I

Initial N	Number of Releases	Estimation Model
1000	50/year	$S(\cdot)h(\cdot)f(\cdot)$
1000	100/year	$S(\cdot)h(\cdot)f(\cdot)$
1000	100/year	$S(a)h(t)f(\cdot)$
2000	100/year	$S(a)h(t)f(\cdot)$
2000	300/year	$S(a)h(t)f(\cdot)$
2000	300/year	$S(a+t)h(a)f(t)$
5000	300/year	$S(a+t)h(a)f(t)$

0.2 or 0.5. Thus, I considered 8 possible combinations of initial parameter values, with a possibility of 3, 5, or 7 years of data, and 2, 3, or 6 age classes.

I used different population models to simulate data depending on the number of age classes. In each case, I used specified values of  $S$ ,  $\lambda$ , and population size to solve for recruitment and to specify a stable stage distribution (all simulations were started with a stable stage distribution). In particular, recruitment was determined by setting  $\det |\mathbf{A} - \lambda \mathbf{I}| = \mathbf{0}$  and solving for  $f$  in terms of  $\lambda$  and  $S$ . The stable stage distribution was determined by solving the system of equations  $[\mathbf{A} - \lambda \mathbf{I}] \mathbf{N}^T = \mathbf{0}$ , subject to the constraint  $N = N_{11} + N_{12} + \dots + N_{1A}$ , where  $\mathbf{N} = [N_{11}, N_{12}, \dots, N_{1A}]$  (Caswell, 2001). Possible population models, presented here as matrices (cf. Caswell, 2001), included

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & f & f & f & f \\ S & 0 & 0 & 0 & 0 & 0 \\ 0 & S & 0 & 0 & 0 & 0 \\ 0 & 0 & S & 0 & 0 & 0 \\ 0 & 0 & 0 & S & 0 & 0 \\ 0 & 0 & 0 & 0 & S & S \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & f \\ S & 0 & 0 \\ 0 & S & S \end{bmatrix}, \text{ and}$$

$$\mathbf{A}_3 = \begin{bmatrix} 0 & f \\ S & S \end{bmatrix}.$$

Here, the matrix model format is used only to represent structural features of the population models; actual simulations assumed stochasticity in all processes. Model  $\mathbf{A}_1$  describes a model with 6 age classes and a pre-breeding census; the last 4 age classes can produce young. Individuals older than five years old cannot be differentiated, a feature incorporated with a self-loop. This scenario corresponds roughly to the population biology and age identification criterion of black bear in Pennsylvania (see Chapter 1). Model  $\mathbf{A}_2$  describes a model for a population with 3 age classes, as when juveniles, yearlings, and subadults/adults may be differentiated, but finer scale resolution is not possible. In this case, only the subadult/adult category contributes to recruitment in the following year. Note that this model is structurally equivalent to  $\mathbf{A}_1$  if one were to pool individuals in age classes 3-6 into a single age class. Finally, model  $\mathbf{A}_3$  describes a population where only 2 age classes are recognizable, as with many waterfowl species.

Values for recruitment and stable stage distribution for the 3 models were as follows:

MODEL  $\mathbf{A}_1$

$$f = \frac{\lambda^2 (\lambda - S)}{S^2} \quad N_{11} = \frac{1}{a}N \quad N_{12} = \frac{S}{a\lambda}N \quad N_{13} = \frac{S^2}{a\lambda^2}N$$

$$N_{14} = \frac{S^3}{a\lambda^3}N \quad N_{15} = \frac{S^4}{a\lambda^4}N \quad N_{1,6+} = \frac{S^5}{a\lambda^4 (\lambda - S)}N \quad a = \frac{\lambda}{\lambda - S}$$

MODEL  $\mathbf{A}_2$

$$f = \frac{\lambda^2(\lambda - S)}{S^2} \quad N_{11} = \frac{1}{a}N \quad N_{12} = \frac{S}{a\lambda}N \quad N_{1,3+} = \frac{S^2}{a\lambda(\lambda - S)}N$$

MODEL **A<sub>3</sub>**

$$f = \frac{\lambda(\lambda - S)}{S} \quad N_{11} = N - S\lambda N \quad N_{1,2+} = S\lambda N$$

Limiting the number of simulation inputs was not enough to decrease computing time to the level that would allow a “normal” number of Monte Carlo simulations to be run per input configuration (i.e., 1,000-10,000). The number of possible simulation input combinations was 504, and computing time for each combination ranged from around 30 minutes to several days, depending upon the problem’s dimensionality. I thus conceptualized the problem as one of estimating a response surface (e.g., Box and Draper, 1987), where the number of simulation runs per input configuration was low ( $n = 3$ ), but where strength could be borrowed from the entire ensemble of simulations to summarize estimator performance in different regions of the input parameter space.

Although certain performance measures such as bias, CV, and BCOV are fairly standard, how to summarize estimator performance was perhaps the most difficult decision that needed to be made in this section. There were a large number of initial parameter combinations and estimation models, and for each estimation model there were a large number of estimated parameters. To condense the number of

estimator performance variables to a manageable number, I thus calculated a mean performance value for each parameter type ( $S$ ,  $h$ ,  $f$ , and  $N$ ) in a given simulation. If there were 10 estimates of survival (time- or age- specific, for instance) at a given level of simulation inputs, response variables would be calculated for a given simulation as the average value across all 10 estimates. An exception was for BCOV, in which the the number of estimates in a given simulation run were taken to be binomial trials, and the number of successes (i.e., the true value of a parameter was within the 90% credible interval) was recorded. Further, mean performance of abundance estimators is taken with respect to annual estimates of total population size (that is, the sum of cohort-specific abundance in a given year) instead of age-specific abundance. For each parameter type, I calculated estimator performance statistics with respect to the posterior mode, mean, and median. Response variables were only recorded for a given simulation if the estimation model converged. If a simulation did not converge, simulations were repeated until a response variable could be obtained. The percent of non-converging simulations per input configuration (%Conv) was also recorded as a potentially confounding factor.

I determined convergence by running 2 Markov chains per simulation and monitoring whether Gelman-Rubin statistics were less than 1.2 immediately after the burn-in period (Gelman et al., 2004). Starting abundance values for the first Markov chain were set equal to the lowest possible values with positive support given the data, and overestimates of survival, recovery rate, and recruitment rate were provided. For the second chain, initial values were automated to produce overestimates of abundance and underestimates of other parameters. Ostensibly, if the

ratio of between-chain to within-chain variance (as measured by the Gelman-Rubin statistic) declined sufficiently, this would be evidence that the effects of opposite types of overdispersed starting values had been overcome.

My approach to summarizing posterior distributions involved running each Markov chain for 1,000,000 iterations. If convergence was determined to have occurred after 500,000 iterations, I combined the second halves of each Markov chain together to produce a sample from the posterior distribution. However, in order to conserve memory, I thinned each chain by only recording every fifth iteration throughout the estimation process, thus yielding a sample of 200,000 from the posterior. Over the course of the study, only 2 out of approximately 1500 simulations did not converge according to the Gelman-Rubin convergence diagnostic.

### **Bayesian credible interval coverage (BCOV)**

Because each simulation produced multiple estimates for each type of parameter (except in case of a  $(\cdot)$  model), I treated the problem of estimating the effects of dependent variables on estimator coverage as one of estimating the regression coefficients of a generalized linear model with binomial error (cf. Fox, 2002). Under this approach, success probability for the  $i$ th simulation,  $p_i$ , is determined according to relationship

$$\text{logit}(p_i) = \beta_0 + \beta_1 y_{i1} + \cdots + \beta_k y_{ik},$$

for  $k$  dependent variables and regression coefficients. The number of times that a specific type of parameter overlaps its posterior interval in simulation  $i$ ,  $X_i$ , is then

modeled as

$$X_i \sim \text{Binomial}(x|Y_i, p_i),$$

where  $Y_i$  is the number of real parameters of a given type estimated in simulation  $i$ . For instance, if there are 7 years of data, there are 7 Bayesian credible intervals for abundance (one for each year). Using this approach, I assumed that particular simulations were more or less prone to failure of coverage, depending upon the predictors.

One advantage of using this approach is the ease to which one can incorporate extrabinomial variation. For instance, in the preceding scenario there will typically be posterior correlations among model parameters. It might also be the case that abundance estimators systematically differ in coverage based on the year of the study, or that undetected problems with Markov chain convergence lead some simulations to produce excessive coverage failures. In these cases, modeling extrabinomial variation becomes important, and can be accomplished by estimating an overdispersion parameter (cf. McCullagh and Nelder, 1989).

Potential dependent variables affecting BCOV included  $A$ ,  $Y$ ,  $R$ ,  $N$ ,  $\lambda$ ,  $S$ ,  $r$ , and the estimation model (*EstMod*). Since I did not have any a priori beliefs about which factors were important determinants of BCOV, I considered all possible combinations of main effects. I fit all models using the function “glm” in the statistical programming language R, as described in Fox (2002). However, only AIC is available in this package, which is not suited for overdispersed data (Burnham and Anderson, 2002). In order to perform model selection, I thus used “glm” to estimate

Table 2.5: Summary of model selection results for factors affecting Bayesian credible interval coverage on annual abundance parameters. Reported are  $K$ , the number of parameters in each generalized linear model, and  $\Delta\text{QAIC}_c$ , the difference in  $\text{QAIC}_c$  score from the top model. Overdispersion was estimated as  $\hat{c} = 2.57$  from the most general model and was applied to each  $\text{QAIC}_c$  calculation. The effective sample size was 1512, the total number of simulations. Reported are models within 2.0  $\Delta\text{QAIC}_c$  units of the top-ranked model.

Model	$K$	$\Delta\text{QAIC}_c$
$A + R + S$	3	0.0
$A + R + S + r$	4	0.6
$A + R + S + Y$	4	1.8
$A + R + S + \lambda$	4	2.0

an overdispersion parameter,  $c$ , from the most general model, and calculated  $\text{QAIC}_c$  for each model as suggested by Burnham and Anderson (2002).

Using this approach, it appeared that survival probability, the number of age classes, and the number of newly released marked animals were the most consistent predictors of BCOV for annual abundance (Table 2.5). All other predictors appearing in the top ranked  $\text{QAIC}_c$  models had confidence intervals overlapping zero. Parameter estimates from the  $\text{QAIC}_c$  favored model indicated that credible interval coverage increased with  $R$ , but decreased with  $S$  and  $A$  (Figure 2.1). Only simulations which included 300 newly marked animals per year, a survival probability of 0.6 and 2 age classes were predicted to approach a “nominal” 90% Bayesian interval coverage on annual abundance; predicted coverage in this case was 0.94. In contrast, simulations with  $R = 50$ ,  $S = 0.8$ , and  $A = 6$  were predicted to have a coverage of 0.82; thus, it would appear that the uncertainty about annual abundance estimators is routinely underestimated, particularly when sample sizes are small. However, for larger studies with more marked animals, coverage was satisfactory.



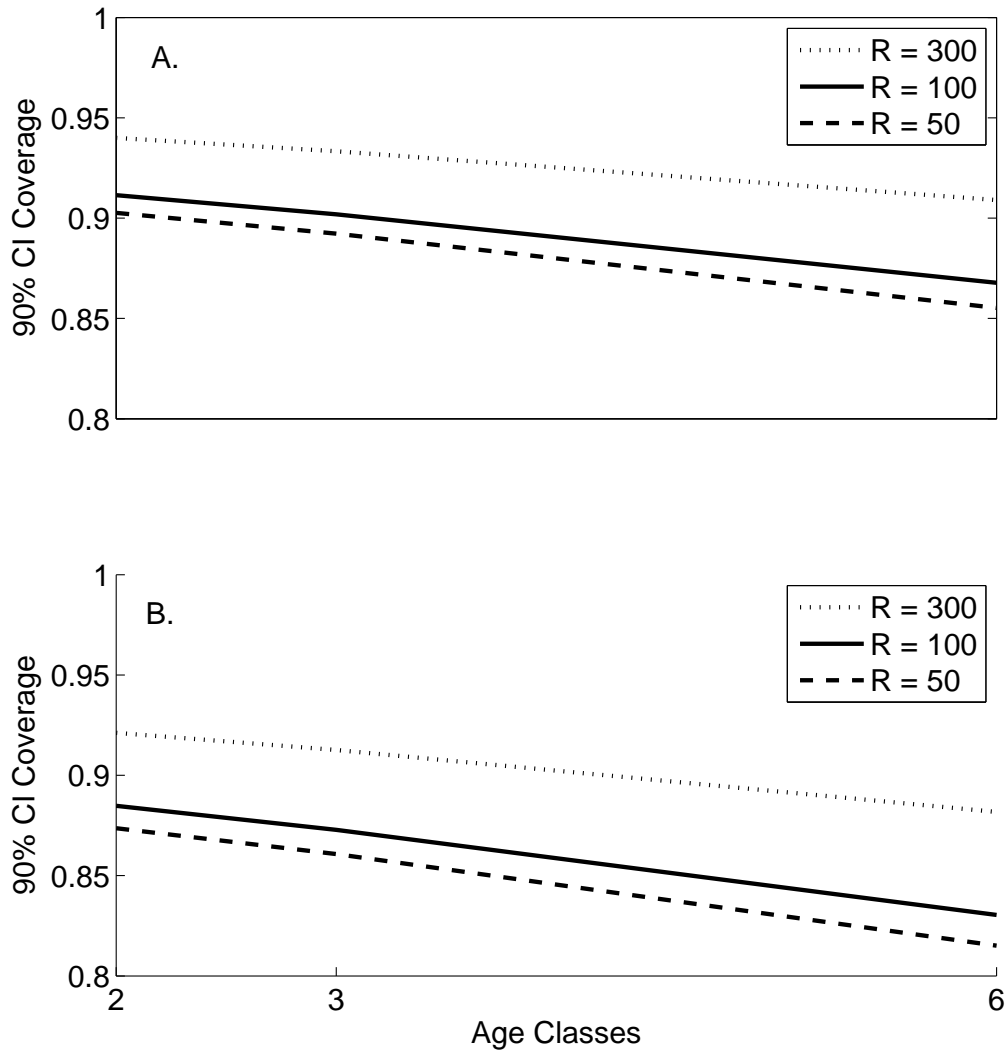


Figure 2.1: 90% Bayesian credible interval coverage for abundance at different levels of number of released animals per year ( $R$ ) and number of age classes. Panel A gives performance for the case where  $S=0.6$ , while panel B is for  $S=0.8$ .

Table 2.6: Summary of model selection results for factors affecting Bayesian credible interval coverage on year- and age- specific survival parameters. Reported are  $K$ , the number of parameters in each generalized linear model, and  $\Delta\text{QAIC}_c$ , the difference in  $\text{QAIC}_c$  score from the top model. Overdispersion was estimated as  $\hat{c} = 2.23$  from the most general model and was applied to each  $\text{QAIC}_c$  calculation. The effective sample size was 1512, the total number of simulations. Reported are models within 2.0  $\Delta\text{QAIC}_c$  units of the top-ranked model.

Model	$K$	$\Delta\text{QAIC}_c$
$A + r$	2	0.0
$A + r + N$	3	0.0
$A + r + Y$	3	0.8
$A + r + Y + N$	4	1.2
$A + r + S$	3	1.2
$A + r + S + N$	4	1.2
$A + r + \lambda$	3	1.6
$Y + r + N + \lambda$	4	1.7
$A + r + R$	3	2.0

For survival parameters,  $\text{QAIC}_c$  favored models that included number of age classes and reporting probability (Table 2.6). Other effects appearing in top models, such as population size or number of years of the study had 90% confidence intervals strongly overlapping 0. In general, confidence interval coverage increased with  $r$  and  $A$  (Figure 2.2). Estimated coverage was between 0.86 and 0.92 depending on input configuration.

For recovery rate parameters ( $h_{ij}$ ),  $\text{QAIC}_c$  clearly favored models that included survival probability,  $\lambda$ , and estimation model (Table 2.7). Other effects appearing in top models included reporting rate, number of age classes, and population size; however, these had 90% confidence intervals overlapping 0. For simplicity, I thus used the 6th ranked model,  $EstMod + \lambda + S$  to portray differences in coverage when there are variations in these parameters (Figure 2.3). Differences in coverage

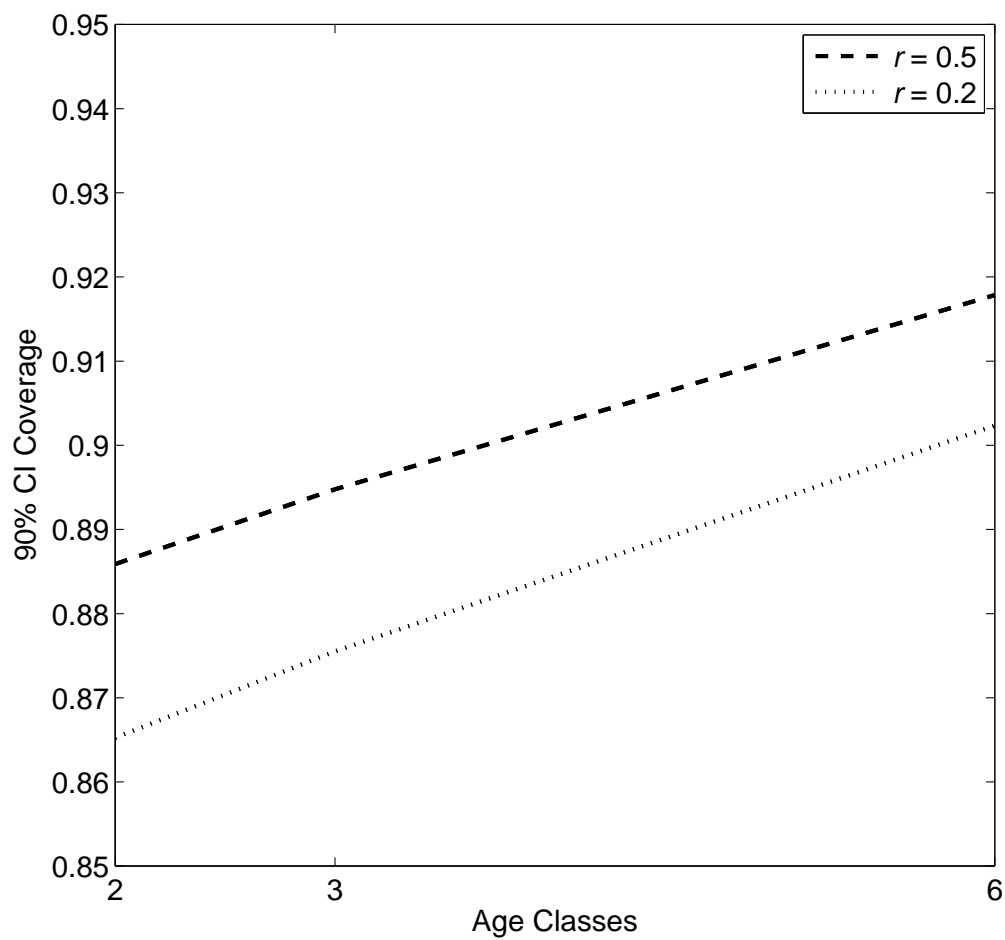


Figure 2.2: 90% Bayesian credible interval coverage for survival at different age classes and reporting rates as estimated from the highest ranked  $QAIC_c$  model.

Table 2.7: Summary of model selection results for factors affecting Bayesian credible interval coverage on year- and age- specific recovery rate parameters. Reported are  $K$ , the number of parameters in each generalized linear model, and  $\Delta QAIC_c$ , the difference in  $QAIC_c$  score from the top model. Overdispersion was estimated as  $\hat{c} = 2.57$  from the most general model and was applied to each  $QAIC_c$  calculation. The effective sample size was 1512, the total number of simulations. Reported are models within 2.0  $\Delta QAIC_c$  units of the top-ranked model.

Model	$K$	$\Delta QAIC_c$
$A + EstMod + \lambda + S + r$	6	0.0
$A + EstMod + \lambda + S$	5	0.2
$A + EstMod + \lambda + S + r + N$	7	0.4
$A + EstMod + \lambda + S + N$	5	0.4
$EstMod + \lambda + S + r$	5	0.6
$EstMod + \lambda + S$	4	0.7
$A + EstMod + \lambda + S + r + R$	7	0.8
$EstMod + \lambda + S + r + N$	6	0.9
$EstMod + \lambda + S + N$	5	0.9
$A + EstMod + \lambda + S + R$	6	1.0
$\lambda + S + r$	3	1.1
$\lambda + S$	2	1.2
$EstMod + \lambda + S + r + R$	6	1.4
$A + \lambda + S + r$	4	1.5
$EstMod + \lambda + S + R$	6	1.5
$A + \lambda + S$	3	1.7
$A + EstMod + \lambda + S + r + Y$	7	2.0

were only reported for estimation model  $S(a)h(t)f(\cdot)$  in relation to the other two models; models  $S(\cdot)h(\cdot)f(\cdot)$  and  $S(a+t)h(a)f(t)$  had only a minuscule difference in coverage. Since more complex models were employed when there were more data, these differences were likely a function of the quantity of data in addition to the complexity of the model. In general, credible interval coverage increased with  $\lambda$  and decreased as survival increased, and was between 0.77 and 0.93 depending upon input configuration (Figure 2.3).

For recruitment rates, there were a total of 32 predictive models for coverage

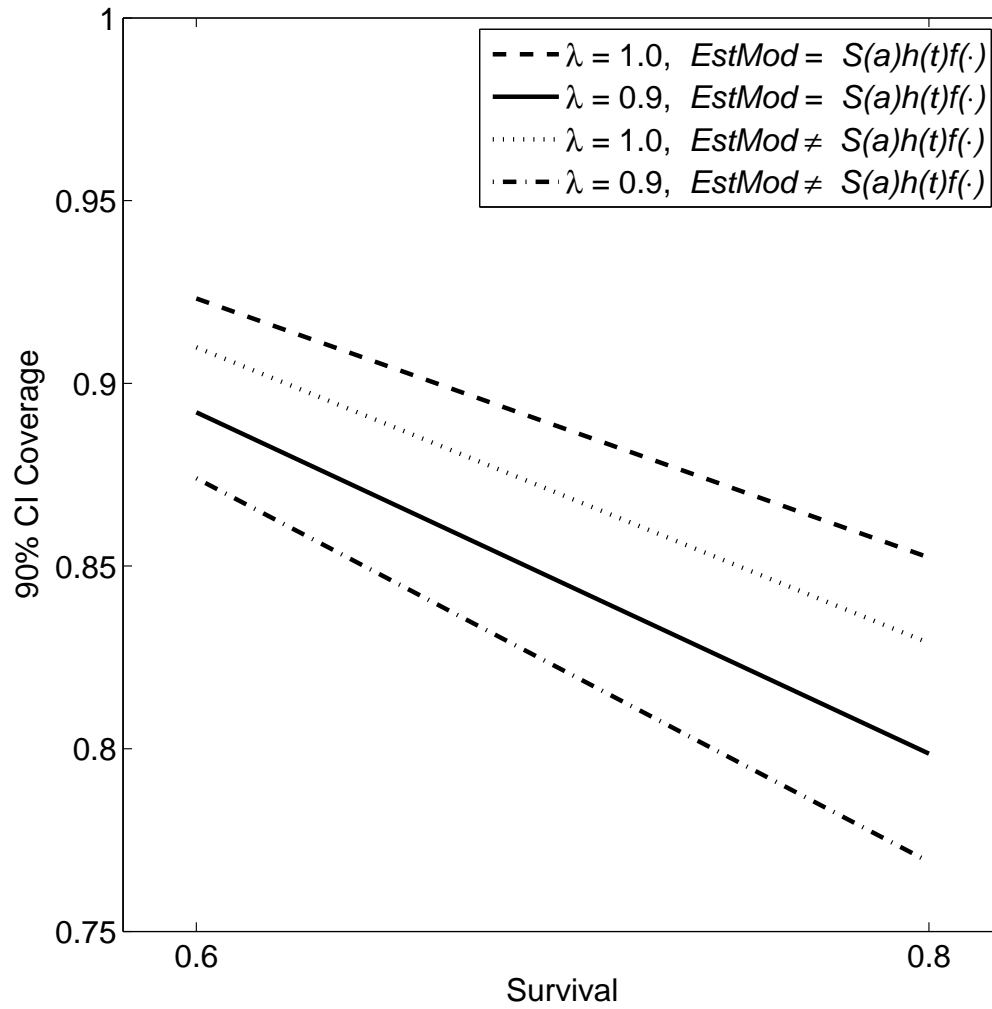


Figure 2.3: 90% Bayesian credible interval coverage for recovery rate, differing by estimation model, level of survival, and finite rate of population change.

within 2.0 QAIC<sub>c</sub> units of the top-ranked model; model selection results were thus suppressed. The most consistent predictors of BCOV were  $\lambda$ ,  $S$ , and *EstMod*, with other variables such as  $A$ ,  $r$ , and  $N$  receiving less support. I thus summarized performance by using a generalized linear model with the first 3 predictors only (Figure 2.4). In general, coverage was positively correlated with  $\lambda$ , but negatively correlated with survival and model complexity. Estimated coverage was between 0.82 and 0.96 (Figure 2.4).

### Percent Relative Bias

Disregarding input simulation parameter specifications, average percent relative bias for estimated abundance was 7.6% (SE=0.9%), 1.8% (SE=0.4%), and 5.8% (SE=0.7%) for posterior mean, mode, and median moment estimators, respectively (Figure 2.5). As a moment estimator for abundance, the mode thus appeared to have the least bias, which is expected given the manner in which data were simulated.

In order to investigate which factors affected the bias of the posterior mode estimator for abundance, I compared the parsimony of alternative linear models that expressed the response variable,  $|\%BIAS|$ , which here is defined as average percent absolute relative bias for a given design point ( $n = 3$ ), as a function all possible combinations of predictor variables ( $A$ ,  $Y$ ,  $R$ ,  $N$ ,  $\lambda$ ,  $S$ ,  $r$ , and the *EstMod*). Inspection of quantile plots in the statistical programming language *R* indicated that residuals from the most general model were not normally distributed, and that there were 3 major outliers. All outliers were associated with the case when  $R = 50$ ,  $S = 0.8$ , and  $r = 0.2$ . I removed these outliers and systematically considered different power

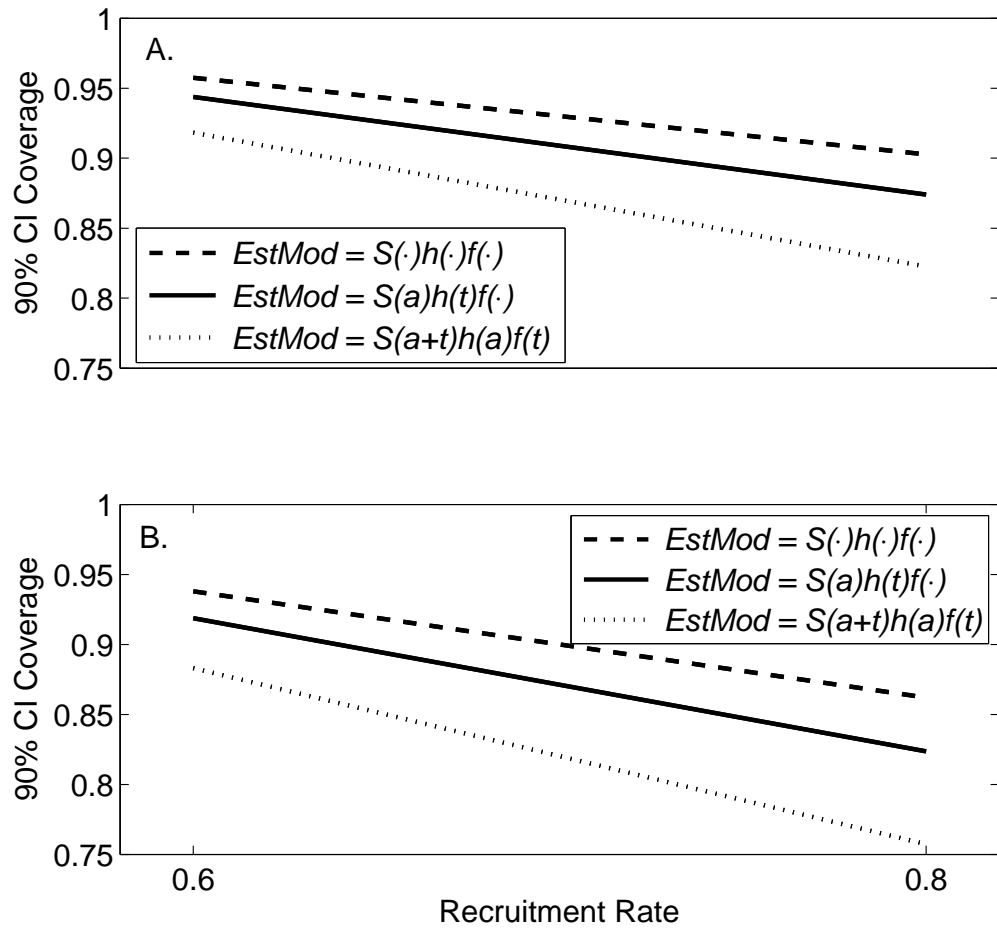


Figure 2.4: 90% Bayesian credible interval coverage for recruitment rate for different estimation models, levels of survival, and finite rate of population change. Panel A. gives coverage estimates for  $\lambda = 1.0$ , whereas Panel B. gives results from  $\lambda = 0.9$ .

Table 2.8: Summary of model selection results for factors affecting average % absolute relative bias for annual abundance parameters. Reported are  $K$ , the number of parameters in each linear model, and  $\Delta\text{AIC}_c$ , the difference in  $\text{AIC}_c$  score from the top model. The effective sample size was 501, the total number of design points minus three outliers. Reported are models within 2.0  $\Delta\text{AIC}_c$  units of the top-ranked model.

Model	$K$	$\Delta\text{AIC}_c$
$A + Y + r + \text{EstMod}$	5	0.0
$A + Y + r + \text{EstMod} + S$	6	0.5
$A + Y + r + \text{EstMod} + N$	6	0.8
$A + Y + r + \text{EstMod} + \lambda$	6	0.8
$A + Y + r + \text{EstMod} + S + N$	7	1.3
$A + Y + r + \text{EstMod} + S + \lambda$	7	1.4
$A + Y + r + \text{EstMod} + N + \lambda$	7	1.6

transformations for  $|\% \text{BIAS}|$  until quantile plots indicated that residuals were normally distributed. A power transformation of 0.3 seemed sufficient for this purpose, and a plot of studentized residuals against fitted values further indicated that the residual variance was largely constant under this approach.

Inspection of parameter estimates and standard errors from top-ranked  $\text{AIC}_c$  models (Table 2.8) indicated that  $A$ ,  $Y$ ,  $r$ , and  $\text{EstMod}$  were the most important variables for predicting  $|\% \text{BIAS}|$  for abundance. I thus based inferences on the highest ranked  $\text{AIC}_c$  model. In general,  $|\% \text{BIAS}|$  increased with  $A$ , and decreased with  $Y$  and  $r$ , although an asymptotic 95% confidence interval for the regression coefficient of  $A$  overlapped zero (-0.0004,0.0116). Absolute bias was also negatively associated with estimation model complexity (Figure 2.6).

As a whole, simulations revealed a negative bias in moment estimators for survival (Figure 2.7). Average percent relative bias for the posterior mean, median, and mode was -3.6% (SE=0.2%), -3.3% (SE=0.2%), and -2.4% (SE=0.2%), respectively.



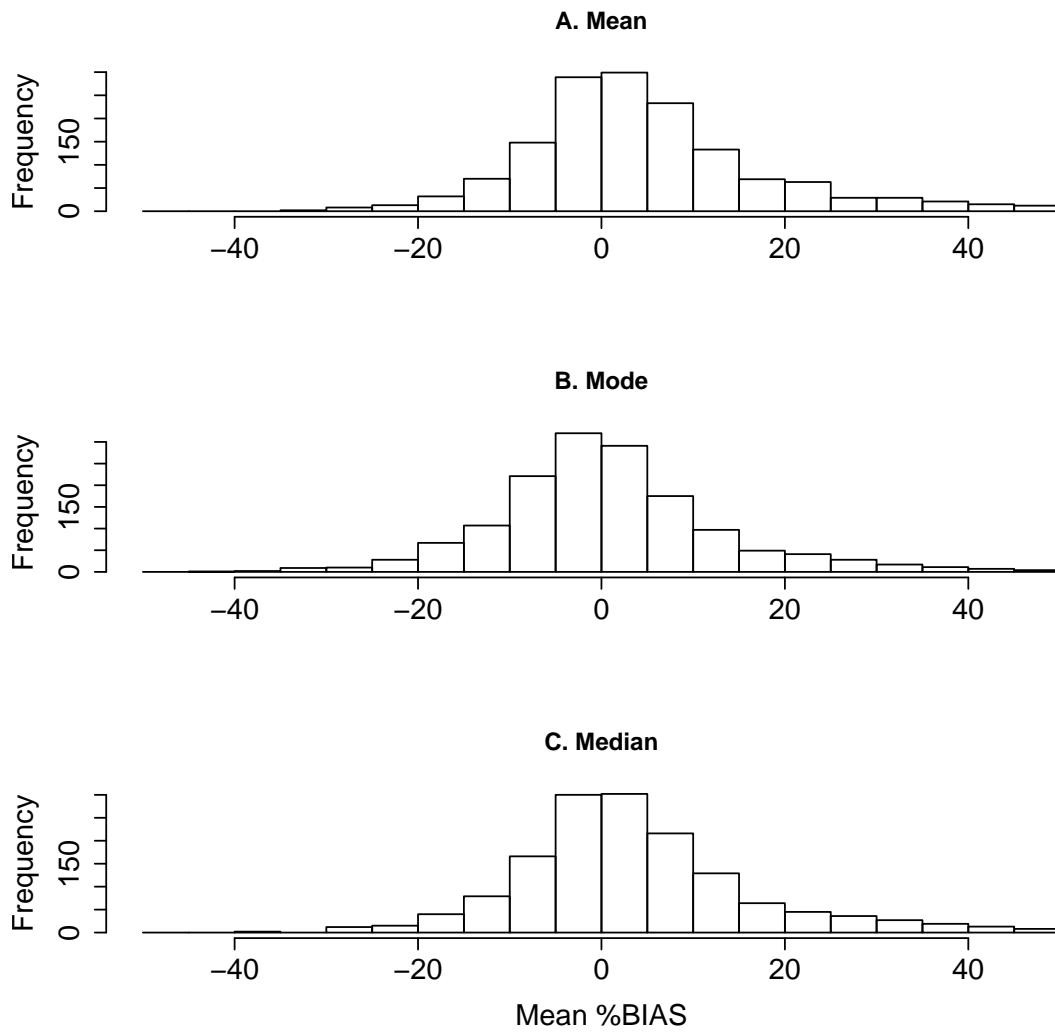


Figure 2.5: The distribution of average percent relative bias for abundance over all simulations, as determined by (a) the posterior mean, (b) the posterior mode, and (c) the posterior median. One value is included for every simulation; thus both sampling variation and bias are represented.

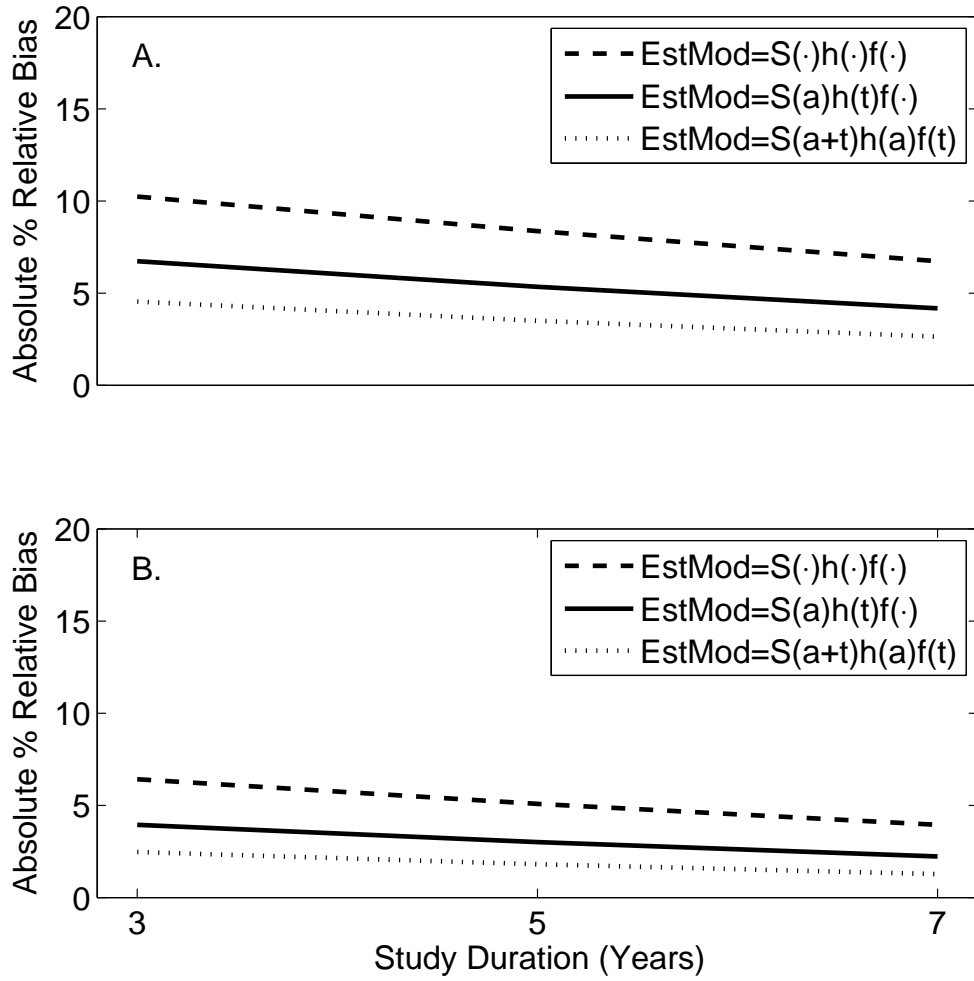


Figure 2.6: Estimated relationship between absolute average % relative bias for abundance and  $Y$ ,  $r$ , and  $EstMod$ . All predictions are given for the case where  $A = 3$ . Panel (A) gives results for  $r = 0.2$  and panel (B) depicts the case when  $r = 0.5$ .

Table 2.9: Summary of model selection results for factors affecting average % absolute relative bias for survival. Reported are  $K$ , the number of parameters in each linear model, and  $\Delta\text{AIC}_c$ , the difference in  $\text{AIC}_c$  score from the top model. The effective sample size was 504, the total number of simulation design points. Reported are models within 2.0  $\Delta\text{AIC}_c$  units of the top-ranked model.

Model	$K$	$\Delta\text{AIC}_c$
$A + R + N + r$	4	0.0
$A + R + N + r + Y$	5	0.7
$A + R + N + r + \lambda$	5	1.3
$A + R + N$	3	1.8
$A + r$	2	1.8

The same suite of linear models was fit to simulation data, with  $|\% \text{BIAS}|^{0.25}$  for survival as the response variable. The 0.25 power transformation resulted in better conformation of residuals to linear model assumptions. Inspection of parameter estimates and standard errors from highly ranked  $\text{AIC}_c$  models (Table 2.9) indicated that the predictors  $A$ ,  $r$ ,  $N$ , and  $R$  were important determinants of  $|\% \text{BIAS}|^{0.25}$ , but that an asymptotic 95% confidence interval for the  $r$  effect overlapped zero (-0.235,0.004) when all 4 variables were included in model structure. I thus used the fourth ranked  $\text{AIC}_c$  model to describe factors affecting absolute mean percent relative bias  $|\% \text{BIAS}|$  (Figure 2.8). In general,  $|\% \text{BIAS}|$  was predicted to increase with the number of age classes and abundance, and to decrease with the number of releases.

As with survival, simulations pointed to a negative bias in moment estimators for recovery probability (Figure 2.9). Average percent relative bias for the posterior mean, mode, and median were -1.5% (SE=0.3%), -3.9% (SE=0.3%), and -2.3% (SE=0.3%), respectively. The same suite of linear models were fit to simulated

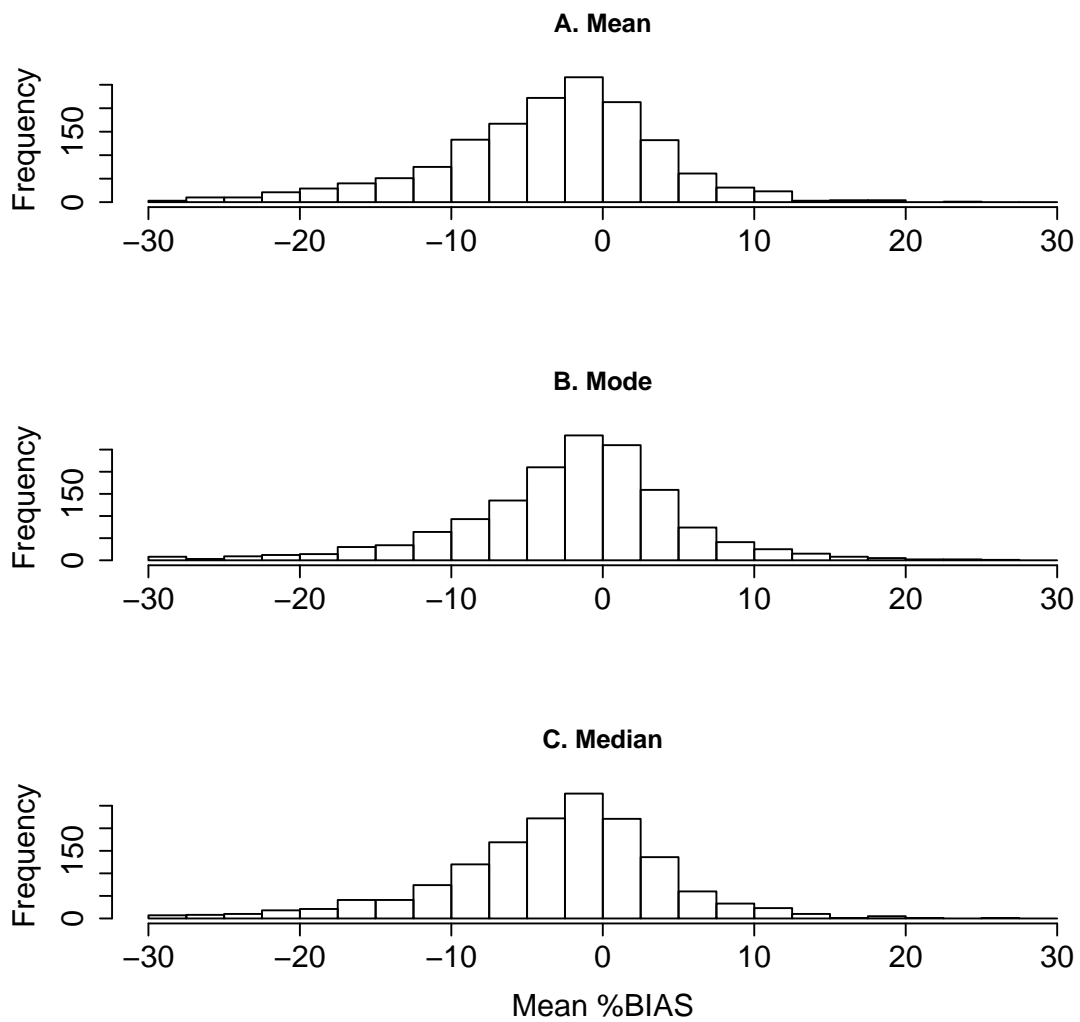


Figure 2.7: The distribution of average percent relative bias for survival over all simulations, as determined by (a) the posterior mean, (b) the posterior mode, and (c) the posterior median. One value is included for every simulation; thus both sampling variation and bias are represented.

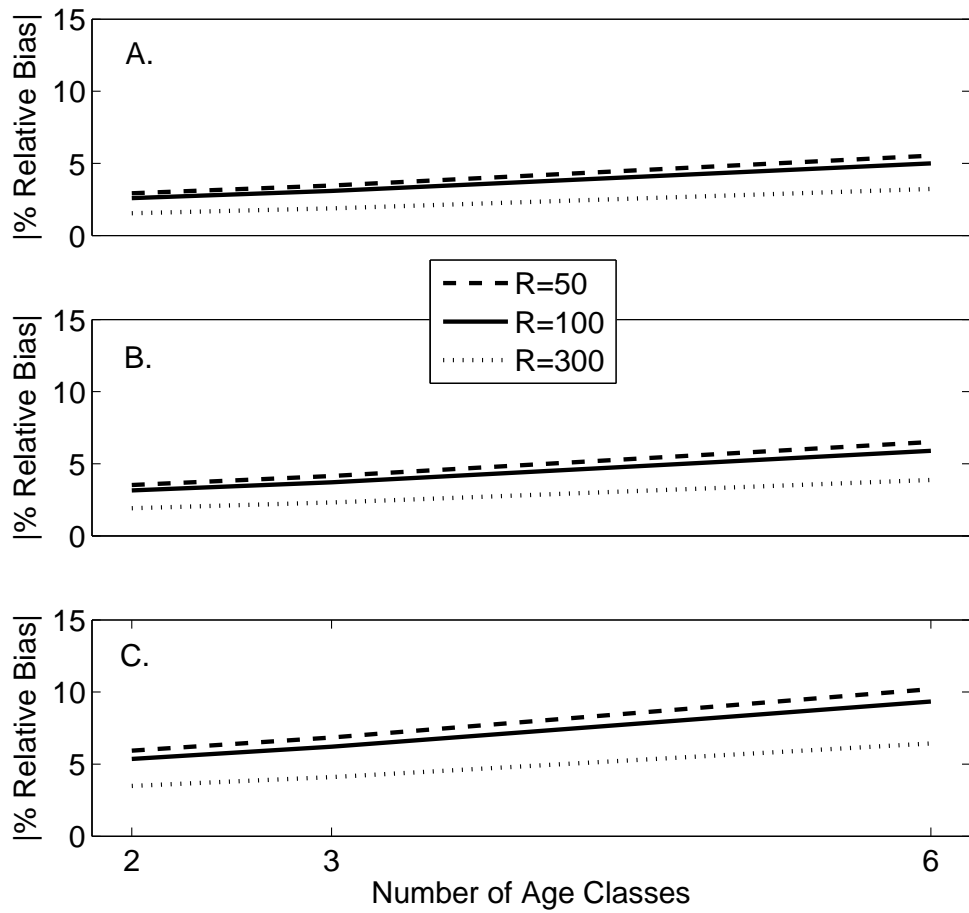


Figure 2.8: Estimated relationship between absolute average % relative bias for survival and  $R$ ,  $N$ , and  $A$ . Panel (A) gives results for  $N = 1000$ , panel (B) depicts the case when  $N = 2000$  and panel (C) gives predictions when  $N = 5000$ .

Table 2.10: Summary of model selection results for factors affecting average % absolute relative bias for recovery rate ( $h$ ). Reported are  $K$ , the number of parameters in each linear model, and  $\Delta AIC_c$ , the difference in  $AIC_c$  score from the top model. The effective sample size was 1512, the total number of simulations. Reported are models within 2.0  $\Delta AIC_c$  units of the top-ranked model.

Model	$K$	$\Delta AIC_c$
$A + Y + R + S + EstMod + r + \lambda$	8	0.0
$A + Y + R + S + EstMod + r + \lambda + N$	9	1.9

data, instead using  $|\%BIAS|^{0.4}$  of the posterior mode of recovery probability as the response variable. In this case, a 0.4 power transformation satisfied linear model assumptions. All highly ranked  $AIC_c$  models included every predictor except  $N$  and interaction terms (Table 2.10). Unfortunately, there were too many highly influential predictors to produce an economical set of displays. However, predictions could be made from the equation

$$|\%BIAS|^{0.4} = 0.594 - .00038R - .018Y - .213\lambda + .241S - .277r \\ - .019MOD1 - .065MOD2.$$

Here,  $MOD1 = 1$  if the estimation model is  $S(\cdot)h(\cdot)r(\cdot)$ , and 0 otherwise. Similarly,  $MOD2 = 1$  if the estimation model is  $S(a)h(t)r(\cdot)$ , and 0 otherwise.

Simulations revealed a positive bias in moment estimators for recruitment (Figure 2.10); average percent relative bias for the posterior mean, median, and mode were 9.7% (SE=0.8%), 7.8% (SE=0.7%), and 4.2% (SE=0.6%), respectively. Again, I employed linear models with all combinations of explanatory variables to examine the relationship between  $|\%BIAS|$  and simulation inputs. This time, a 0.35

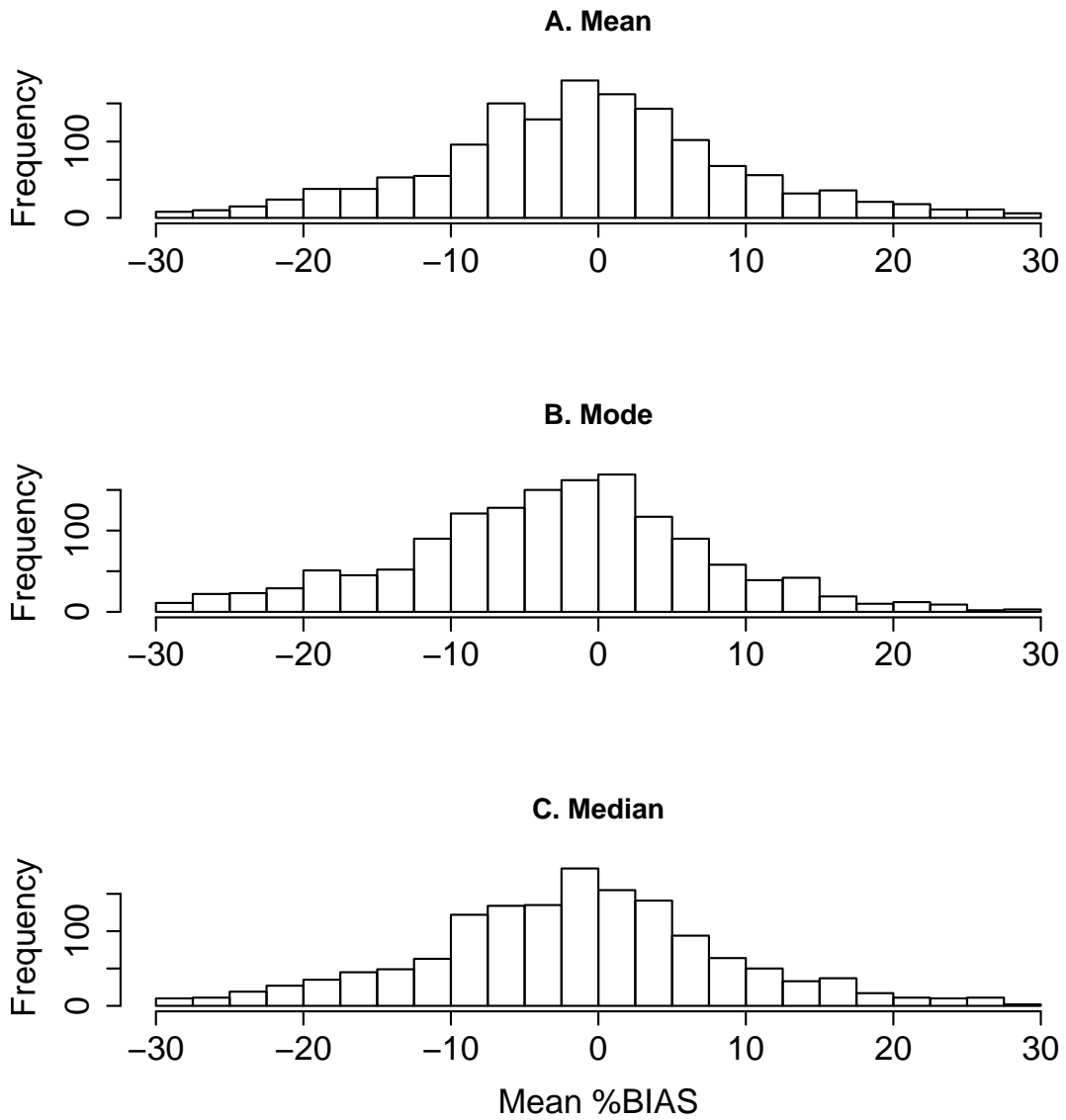


Figure 2.9: The distribution of average percent relative bias for recovery rate over all simulations, as determined by (a) the posterior mean, (b) the posterior mode, and (c) the posterior median. One value is included for every simulation; thus both sampling variation and bias are represented.

Table 2.11: Summary of model selection results for factors affecting average % absolute relative bias for recruitment. Reported are  $K$ , the number of parameters in each linear model, and  $\Delta AIC_c$ , the difference in  $AIC_c$  score from the top model. The effective sample size was 1512, the total number of simulations. Reported are models within 2.0  $\Delta AIC_c$  units of the top-ranked model.

Model	$K$	$\Delta AIC_c$
$Y + R + S + EstMod + r + N + \lambda + EstMod * R + N * R$	10	0.0
$Y + R + S + EstMod + r + N + \lambda + A + EstMod * R + N * R$	11	0.9
$Y + S + EstMod + r + N + \lambda$	7	1.8

power transformation seemed appropriate to make residuals normally distributed with a relatively constant variance. The exercise in model selection indicated that most explanatory variables were important predictors (Table 2.10). As with recovery rate, it was thus difficult to make economical graphical displays; nevertheless, predictions could be made with coefficients from the highest ranked  $AIC_c$  model from the following equation:

$$|\%BIAS|^{0.35} = 0.668 - .00015R - .015Y - .435\lambda + .692S - .308r \\ - .190MOD1 - .158MOD2.$$

### Coefficient of Variation

Conducting preliminary explorations of the data, it appeared that there were a number of outliers associated with CV for abundance, all of which were associated with the case of 50 releases per year, 3 years of data, and survival and reporting rates of 0.8 and 0.2, respectively. In this scenario, the number of expected recoveries of marked animals is quite small (2.0, 3.6, and 4.9 for years 1,2, and 3, respectively)



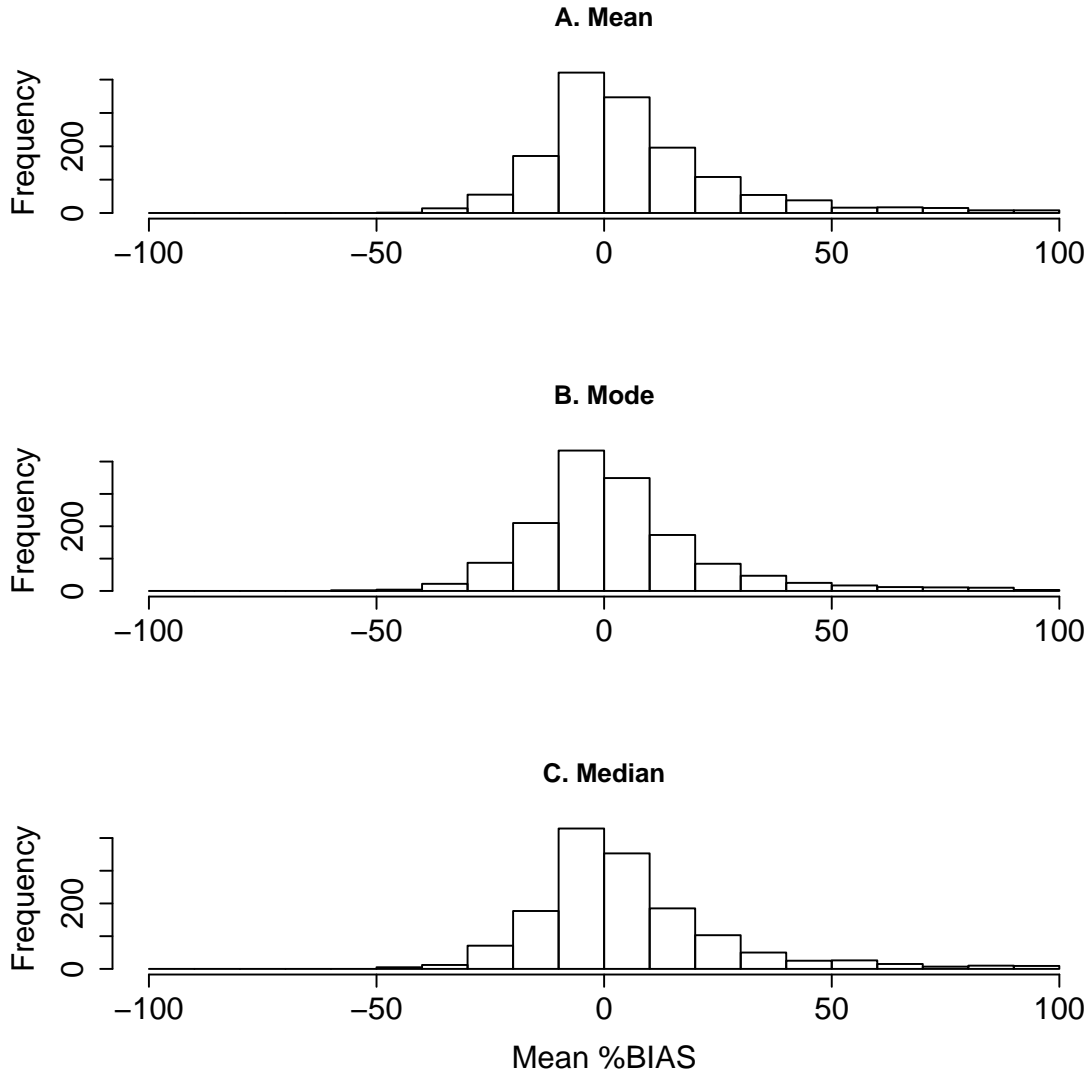


Figure 2.10: The distribution of average percent relative bias for recruitment over all simulations, as determined by (a) the posterior mean, (b) the posterior mode, and (c) the posterior median. One value is included for every simulation; thus both sampling variation and bias are represented.

and can lead to unstable estimates. Interestingly, other than one outlier removed from the analysis of bias in abundance, this combination of simulation inputs did not seem to produce unduly extreme outliers for other measures of model performance. Nevertheless, removing these points was essential for analysis of CV.

Analysis of CV was performed for abundance only, as a coding error resulted in incorrect calculations of this performance measure for other parameters. A power transformation of -0.4 on the response variable proved adequate for meeting linear model assumptions, and the same suite of models were fit to the data. The top  $AIC_c$  model in this case was the most general model, with the second closest model  $> 3 \Delta AIC_c$  units behind. As with the analysis of recruitment bias, there were too many important predictors in this case to visually portray cumulative effects of all simulation inputs; nevertheless, approximate predictions for mean CV can be made using estimated regression coefficients as

$$\begin{aligned} \left[ CV(\hat{N})^{-0.4} \right] &= 1.31 + 0.000402N + 0.00202R + 0.00324A + 0.111Y \\ &\quad + 0.427\lambda - 1.517S + 1.844r - 0.150MOD1 - 0.0234MOD2 \\ &\quad + 0.00285(R \times MOD1) + 4.93 \times 10^{-8}(N \times R). \end{aligned}$$

Because the response variable is taken to a negative power, the interpretation of the sign of regression coefficients is reversed; for instance, a positive coefficient leads to a decrease in CV. As with any regression analysis, it is important not to extrapolate predictions from the linear model past the particular input configurations of the response surface analysis, especially given that the design was not a complete

factorial. Over the support of simulation inputs, mean  $CV(\hat{N})$  was predicted to be lowest ( $= 0.016$ ) when  $N = 5000$ ,  $R = 300$ ,  $A = 6$ ,  $Y = 7$ ,  $\lambda = 1$ ,  $S = 0.6$ , and  $r = 0.5$ . Note that the only estimation model employed for this configuration was  $S(a+t)h(a)f(t)$ ; a lower coefficient of variation could be expected if a simpler estimation model were used here. The highest prediction for  $CV(\hat{N})$  was 0.24, which corresponded to the situation where  $N = 1000$ ,  $R = 100$ ,  $A = 2$ ,  $Y = 3$ ,  $\lambda = 0.9$ ,  $S = 0.8$ ,  $r = 0.2$ , and  $EstMod = S(a)h(t)f(\cdot)$ .

### 2.7.2 Simulation Module II: Effects of Aging Error

Misclassification of an individual's age might be expected to produce bias in parameter estimates, particularly if error rates vary systematically by age. For instance, small magnitude positive biases are typically observed in age estimates of young black bears, while larger negative biases are observed for older bears (e.g., Beck, 1991; Costello et al., 2004; Harshyne et al., 1998). Even if there are no systematic age-mediated changes in the direction and magnitude of bias, age misclassification will still tend to result in misrepresentation of "strong" age classes (Fournier and Archibald, 1982). In effect, the power to discriminate cohorts with large abundances will be somewhat obscured, and process error in recruitment will be underestimated. In the context of fisheries stock assessment models, it has also been shown that aging error can lead to negative bias in estimates of harvest mortality and measures of abundance, possibly leading to management advice which would favor over-harvesting (Reeves, 2003).

Published levels of aging error rates differ dramatically between taxa and even among populations of the same taxa. For instance, aging estimates for black bears in west-central Colorado were shown to be both inaccurate and imprecise (Beck, 1991), while estimates from known-age bears were markedly better for studies occurring in California (Keay, 1995), Pennsylvania (Harshyne et al., 1998), and New Mexico (Costello et al., 2004). Discrepancies may be due to differences in diet and variability in food availability in different areas affecting cementum deposition (Costello et al., 2004), but also may be due to standardization of techniques in the latter three studies to conform with subsequently published work (cf. Coy and Garshelis, 1992). The latter three studies also used the same commercial laboratory (Matson's Laboratory) to perform aging assignments.

In this module, I examined the effect of several magnitudes and types of aging error on estimates of abundance and other parameters under a number of hypothetical scenarios. In each case, I consider a model for aging errors whereby assigned age  $A'$  is related to true age  $A$  by the relationship

$$A' = \text{round}(A + \epsilon_j)$$

where  $\epsilon_j$  denotes a random effect associated with age  $j$ . Variation in error type and magnitude are produced by considering different models for  $\epsilon_j$ . In total, five a priori models for aging error were considered:

- Model 1: No aging error (i.e.,  $\epsilon_j = 0$ )

- Model 2:

$$\epsilon_j \sim \frac{\text{Laplace}(0, \sigma_j) I_{[0.5-j, \infty)}(\epsilon_j)}{1 - \int_{-\infty}^{0.5-j} \text{Laplace}(0, \sigma_j)} \quad \sigma_j = 0.1j$$

- Model 3:

$$\epsilon_j \sim \frac{\text{Laplace}(0, \sigma_j) I_{[0.5-j, \infty)}(\epsilon_j)}{1 - \int_{-\infty}^{0.5-j} \text{Laplace}(0, \sigma_j)} \quad \sigma_j = 0.2j$$

- Model 4:

$$\epsilon_j \sim \frac{\text{Laplace}(\mu_j, \sigma_j) I_{[0.5-j, \infty)}(\epsilon_j)}{1 - \int_{-\infty}^{0.5-j} \text{Laplace}(\mu_j, \sigma_j)} \quad \sigma_j = 0.1j, \quad \mu_j = .4 - .2j$$

- Model 5:

$$\epsilon_j \sim \frac{\text{Laplace}(\mu_j, \sigma_j) I_{[0.5-j, \infty)}(\epsilon_j)}{1 - \int_{-\infty}^{0.5-j} \text{Laplace}(\mu_j, \sigma_j)} \quad \sigma_j = 0.2j, \quad \mu_j = .4 - .2j$$

Here,  $I_{[\Omega]}()$  gives an indicator function for the set  $\Omega$ , and the Laplace (also known as the double exponential) distribution is the probability density function

$$f_X(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma}\right).$$

The Laplace distribution has fatter tails than a normal distribution, and may be more useful for describing aging error in natural populations (see, for instance, Chapter 4). Models 2 and 4 specify a relatively high precision on age estimates, while models 3 and 5 are relatively imprecise, particular at older ages. Models 2 and 3 assume no bias in age estimation (other than induced from a nonnegative support for

the true age), while models 4 and 5 assume a positive aging bias for yearlings and a negative one for individuals greater than 2 years of age. In particular, the magnitude of bias increases with age. For all models, I assumed that age 0 individuals were aged definitively, as is typically the case with black bear cubs.

For simplicity, I only considered two biological scenarios for which to quantify possible effects of aging error on estimator performance, both of which corresponded roughly to the demography, harvest numbers, and sampling effort associated with female black bears in Pennsylvania. For each scenario, I set initial population size at 5000, and assumed that 400 new individuals were marked and released per year, the ages of which were in proportion to their relative abundance in the population. I treated the population as if it consisted of 3 demographically relevant age classes: cubs (0-1 year old), yearlings (1-2 years old), and adults (ages 2+), with associated survival rates ( $S$ ) of 0.4, 0.55, and 0.7, and harvest rates ( $h$ ) of 0.1, 0.3, and 0.2, respectively. The two different biological scenarios were described by different parameterizations for recruitment ( $f$ ). Scenario *A* assumed that recruitment was a Poisson process with a mean of 1.364 female cubs per adult female, a number approximately necessary for a stationary population. Scenario *B* assumed the same mean, but with a hyperprior on  $\lambda$ , such that  $f_i \sim \text{Poisson}(\lambda)$ , where  $\lambda \sim \text{Gamma}(10, 7.33)$ . Thus the 2 scenarios embodied quite different assumptions about the nature of process error in recruitment (Figure 2.11). I considered this potentially relevant since aging error will typically serve to obscure the detection of high abundance cohorts, and thus may lead to underestimates of recruitment process error.

Instead of pooling virtual animals into a ‘+’ category to start with, I allowed

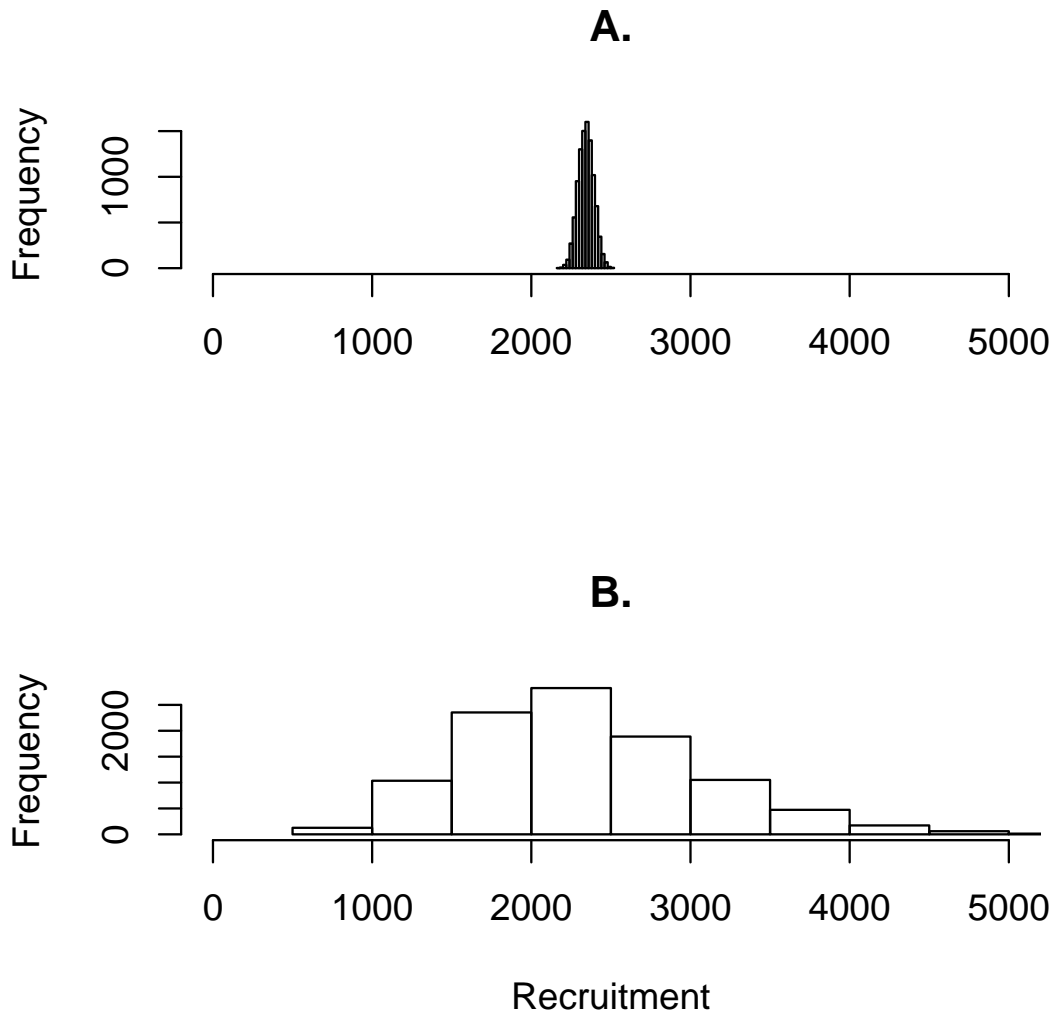


Figure 2.11: Histograms representing the relative likelihood of obtaining different values of recruitment under Scenarios A and B in simulation module II when populations are started at a stable stage distribution and an initial population size of 5000. In particular, Scenario A dictates a relatively fixed number of recruits, while Scenario B allows recruitment to differ drastically from year to year.

them to advance to a possible age of 20, at which time they were automatically removed from the population. In this manner, aging error could be appropriately applied to an animal's real age, and data could be pooled to a pre-specified level for analysis. If aging error increases with the age of an animal, there may be some 'optimal' way in which to pool older age classes in terms of estimator performance, at least for a given hypothetical scenario. For each simulated set of "complete" data, I thus considered two possible levels of pooling: 3 age classes and 7 age classes. In either case, only adults (2-year-olds and older) were assumed to reproduce, and populations were started at a stable stage distribution, as in *Simulation Module I*.

I ran a total of 50 simulations for each combination of the 5 aging error models and 2 biological scenarios. For each simulation and pooling option, I ran 2 chains of length 1,000,000 starting at overdispersed values. If after 500,000 iterations Gelman-Rubin statistics confirmed that the chains had approximately the same within- and between- chain variance, I combined the final 500,000 samples of each chain to arrive at a sample of 1,000,000 from the posterior distribution, which was thinned to 200,000 to save memory. I calculated the same statistics as in *Simulation Module I* to quantify estimator performance; model  $S(a)h(a)f(\cdot)$  was used to estimate parameters for all Scenario A simulations, while model  $S(a)h(a)f(t)$  was used for all Scenario B simulations. Here, an  $a$  denotes the case where 3 parameters are estimated, corresponding to cubs, yearlings, and adults.

To quantify effects of aging error on estimator performance, I once again compared the relative parsimony of different models where estimator performance was related to all possible subsets of predictor variables. Here, I considered average ab-



solute percent relative bias and 90% credible interval coverage for each parameter type (i.e.,  $\mathbf{N}$ ,  $\mathbf{S}$ ,  $\mathbf{h}$ ,  $\mathbf{f}$ ) as possible response variables; predictor variables were  $\mu$  (2 levels),  $\sigma$  (3 levels), and recruitment variance scenario (2 levels). To start, I only analyzed 7-age class data because more errors could be expected at older age classes, and thus it was more likely to see effects on estimator performance. Also, 3-age class data were obtained by pooling data from 7-age-classes, and so these data were not independent (a mixed model approach may have been used here to limit the number of degrees of freedom for pairwise blocking, but this seemed excessive). I entertained the possibility of including two way interaction terms whenever the variables comprising the interaction were included in the list of predictor variables. However, one of the interactions between  $\sigma$  and  $\mu$  was not estimable because simulations did not include the design point where  $\mu_j = 0.4 - 0.2j$  and  $\sigma_j = 0$ .

A total of 36 models were fit for each response variable. When bias was modeled, my strategy was to find a transformation of the response variable which approximately satisfied linear model assumptions for the most general model, as indicated by quantile plots and plots of studentized residuals versus fitted values. In this case, I used  $AIC_c$  for model selection. When coverage was of interest, I once again considered a binomial model for the response variable within a generalized linear model framework where success probability was related to predictors. In this case, I used a logit link, estimated an overdispersion parameter,  $c$ , and used  $QAIC_c$  for model selection.

For bias on estimators of annual abundance, a power transformation of 0.1 seemed adequate for meeting linear model assumptions. An intercept model, in

which bias was constant across simulation inputs, was ranked second with  $\Delta\text{AIC}_c = 0.3$ . Parameter estimates from other top ranked models actually predicted that bias would decrease with the amount of aging error (for both  $\sigma$  and  $\mu$ ), although confidence intervals overlapped zero. I thus concluded that there was no simulation-based evidence that aging error increased bias in annual abundance estimators (see also Figure 2.12).

Models selected for 90% Bayesian credible interval coverage on annual abundance tended to include  $\mu$  and  $\sigma$  effects on coverage, as well as interactions with these terms and the level of recruitment variance. Investigation of parameter estimates and confidence intervals from top ranked models indicated that coverage decreased slightly when  $\sigma > 0$  and when recruitment variance was high. Strangely, coverage was predicted to increase when  $\mu$  was different than the true age class in this case (Figure 2.13).

For survival, I used a power transformation of 0.4 to relate bias to predictor variables. Top models tended to include  $\mu$ , predicting that bias would increase when the expected mode of aging error was different from zero. However, confidence intervals for this effect overlapped zero, and the intercept-only model was only 1.8  $\Delta\text{AIC}_c$  units behind the top model, indicating that the levels of aging error used here did not greatly affect bias in survival estimators (see also Figure 2.12).

The number of age-specific survival parameters that were covered in a given simulation suffered from substantial overdispersion ( $\hat{c} = 4.3$ ). Even so, recruitment variance appeared in all top models, with point estimates predicting that credible interval coverage would decrease slightly under high recruitment variance

( $\hat{\beta} = -0.41, SE(\hat{\beta}) = 0.22$ ). However, given the high value of  $\hat{c}$ , it is perhaps not surprising that the intercept-only model was ranked close to the top, with a  $\Delta\text{QAIC}_c = 2.1$ .

For recovery rate, no transformation of the response variable was needed when modeling bias. Recruitment variance,  $\sigma$ , and  $\mu$  all appeared in top models; bias was predicted to increase with  $\sigma$  and recruitment variance, but was predicted to decline when  $\mu \neq 0$ . The latter effect was once again counterintuitive, but somewhat borne out by observed data (Figure 2.12). Observed coverage again appeared overinflated given the simulation run; overdispersion was estimated as  $\hat{c} = 3.7$ . This lack of fit obscured my ability to discern any effect of predictors on recovery rate coverage, the intercept only model was ranked second with  $\Delta\text{QAIC}_c = 0.1$ ; no explanatory variables consistently appeared in top-ranked models.

As with recovery rate, transformation was not required when modeling recruitment bias as a response variable. Top models all included  $\sigma$  and process variance in recruitment as predictor variables, with higher biases predicted when standard deviations of aging errors increased and when recruitment variance increased. For recruitment credible interval coverage, overdispersion was estimated as  $\hat{c} = 2.9$ . Again, there were no clearly important predictors, as the intercept model was highly ranked ( $\Delta\text{QAIC}_c = 1.2$ ). The model of highest rank, for instance, included only an effect of high variance in the aging process ( $\sigma_j = 0.2j$ ), indicating a slight decrease in coverage under this scenario ( $\hat{\beta} = -0.49, SE(\hat{\beta}) = 0.28$ ).

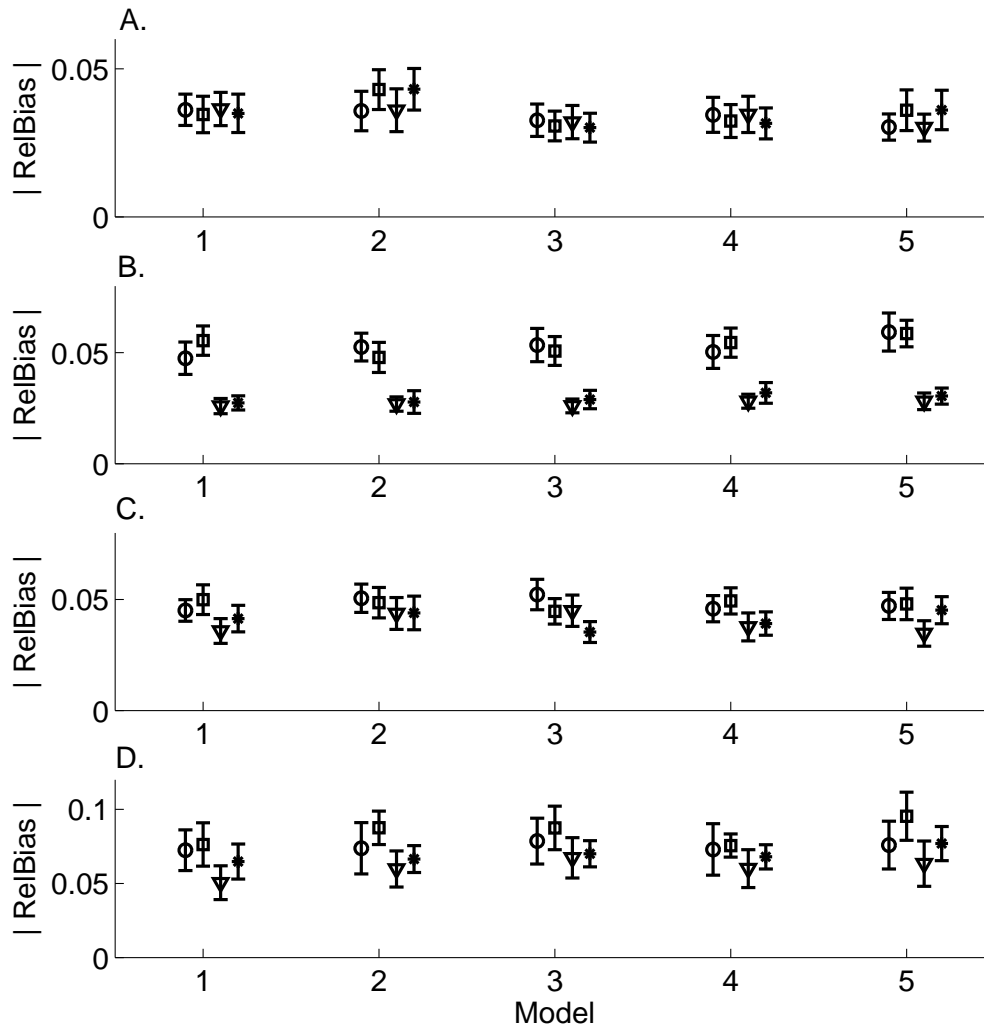


Figure 2.12: Average absolute relative bias and approximate 90% confidence intervals ( $\pm 2$  SE) for different parameters, aging error models, and number of age classes. Panels (A), (B), (C), and (D) give results for abundance, survival, recovery rate, and recruitment, respectively. Circles give results for 3 age classes and low process variance in recruitment; squares give results for 3 age classes and high process variance in recruitment; triangles denote 7 age classes and low recruitment variance; stars denote 7 age classes and high recruitment variance. The five models for aging error are defined in the text.

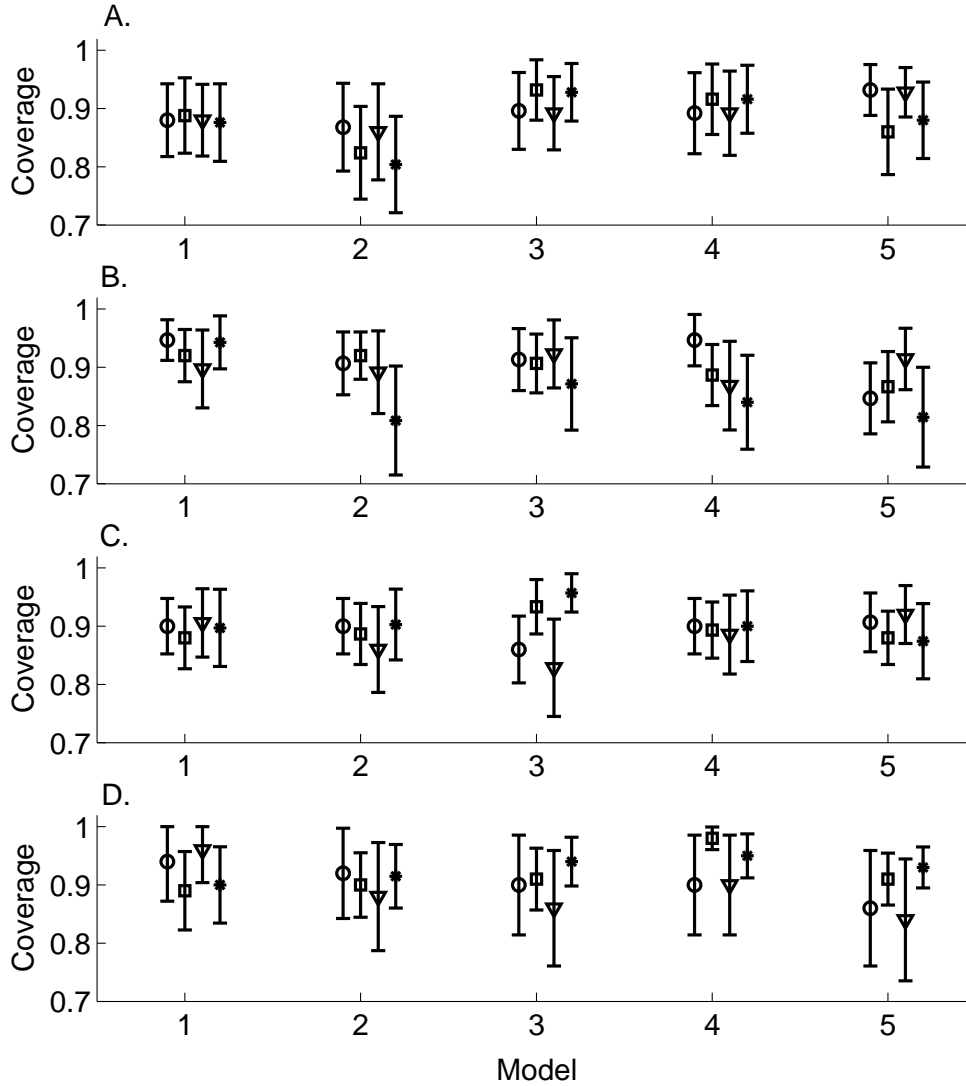


Figure 2.13: Estimates of 90% Bayesian credible interval coverage, together with approximate 95% confidence intervals ( $\pm 2$  SE) for different parameters, aging error models, and number of age classes. Panels (A), (B), (C), and (D) give results for abundance, survival, recovery rate, and recruitment, respectively. Circles give results for 3 age classes and low process variance in recruitment; squares give results for 3 age classes and high process variance in recruitment; triangles denote 7 age classes and low recruitment variance; stars denote 7 age classes and high recruitment variance. The five models for aging error are defined in the text.

### 2.7.3 Simulation Module III: Detecting overdispersion

In this module, I assessed the efficacy of Bayesian p-values for diagnosing lack of fit. While there are many causes for lack-of-fit, I focused on scenarios in which the fates of individuals were not independent. The independence assumption is one of the most difficult to relax, and violation of this assumption will generally lead to credible intervals for model parameters that are too short.

Following an approach developed by G. C. White (*personal communication*), I considered several levels of overdispersion corresponding to different levels of dependence among individuals. These levels are summarized by a single overdispersion parameter,  $c$ , which relates the effective sample size to perceived sample size. For instance, when all animals' fates are independent  $c = 1$  and perceived sample size equals effective sample size. In contrast, when the fate of every individual in the population is the same as exactly one other individual in the population (as with a pair bond),  $c = 2$ , and perceived sample size is twice the effective sample size. Simulated data for different levels of  $c$  may be produced by duplicating fates of individuals. For instance, if there are 120 individuals released, and  $c = 2$ , the fates of 60 individuals would be simulated and each recorded twice.

Detection of lack-of-fit should be influenced by the value of  $c$  as well as sample sizes, the complexity of model structure, and number of age classes that can be reliably determined. For this reason, I examined scenarios that had different levels of overdispersion ( $c = 1$  or  $2$ ), and varied by the number of individuals newly captured and released per year (200 or 400), population size of unmarked animals (1000 or

5000), and the number of age classes (3 or 5). For simplicity, I considered age- and time-constant survival, recovery ( $h$ ), and recruitment rates of 0.6, 0.2, and 0.4, respectively, assuming that individuals in all age classes could reproduce. I employed the fixed effect model  $S(a + t)h(a)f(\cdot)$  to estimate parameters in each case; thus the Bayesian p-value applies to this model only. Detection of overdispersion may also be influenced by the number of years of data. However, for simplicity, I only considered the case where there were 5 years of data. I ran a total of 20 simulations for each combination of input factors, using the same MCMC configurations as in simulation module II.

In order to explore the relationship between simulation inputs and Bayesian p-values, I used the logit of the Bayesian p-value as a response variable and fit a series of linear models to the data. This series included all possible subsets of predictor variables and first order interactions;  $AIC_c$  was used for model selection. No other transformation of the response variable was needed to meet linear model assumptions, although I mapped the interval  $[0,1]$  to  $[0.025, 0.975]$  before applying the logit transformation in order to avoid numerical problems.

Top models fit to the data all included number of age classes, the overdispersion parameter used to generate the data, and their interaction as predictors. In general, Bayesian p-values were predicted to be near 0.5 when an overdispersion value of  $c=1.0$  was used to generate the data, and to be in the 0.1 to 0.2 range when  $c=2.0$  (Figure 2.14). Power to detect lack-of-fit increased markedly when using a higher number of age classes. However, even for the case where  $A = 5$  and  $c = 2$ , observed power to detect lack of fit was 0.36 when using a critical p-value of 0.05 ( $\hat{\alpha} = 0.0$ ).

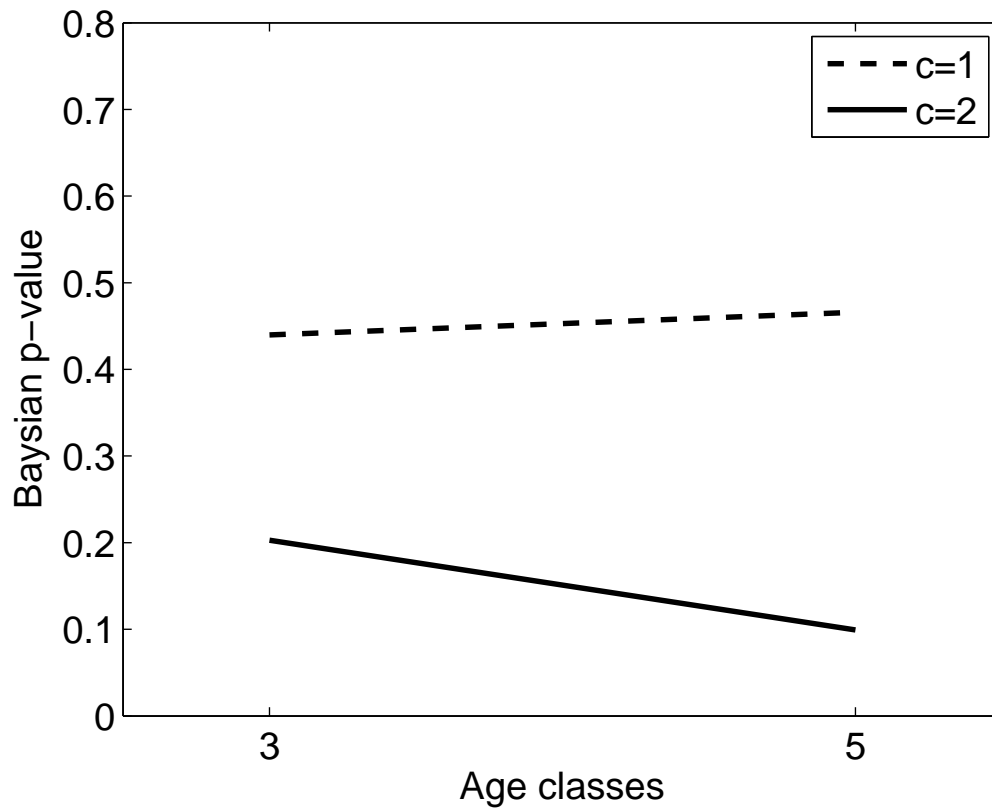


Figure 2.14: Expected Bayesian p-value as predicted by the number of age classes ( $A$ ) and the underlying overdispersion parameter ( $c$ ) used to generate the data.

When using a critical p-value of 0.20, observed power was 0.74, with  $\hat{\alpha} = 0.08$ . Here,  $\hat{\alpha}$  gives the estimated probability of a type 1 error. A visual depiction of the null distribution when  $A = 5$  and  $c = 1$  is also provided (Figure 2.15).

#### 2.7.4 Simulation Module IV: Using marked individuals twice

Up to this point, all evaluations of model performance have assumed that mark-recovery data come from a similar, but different population from the one being



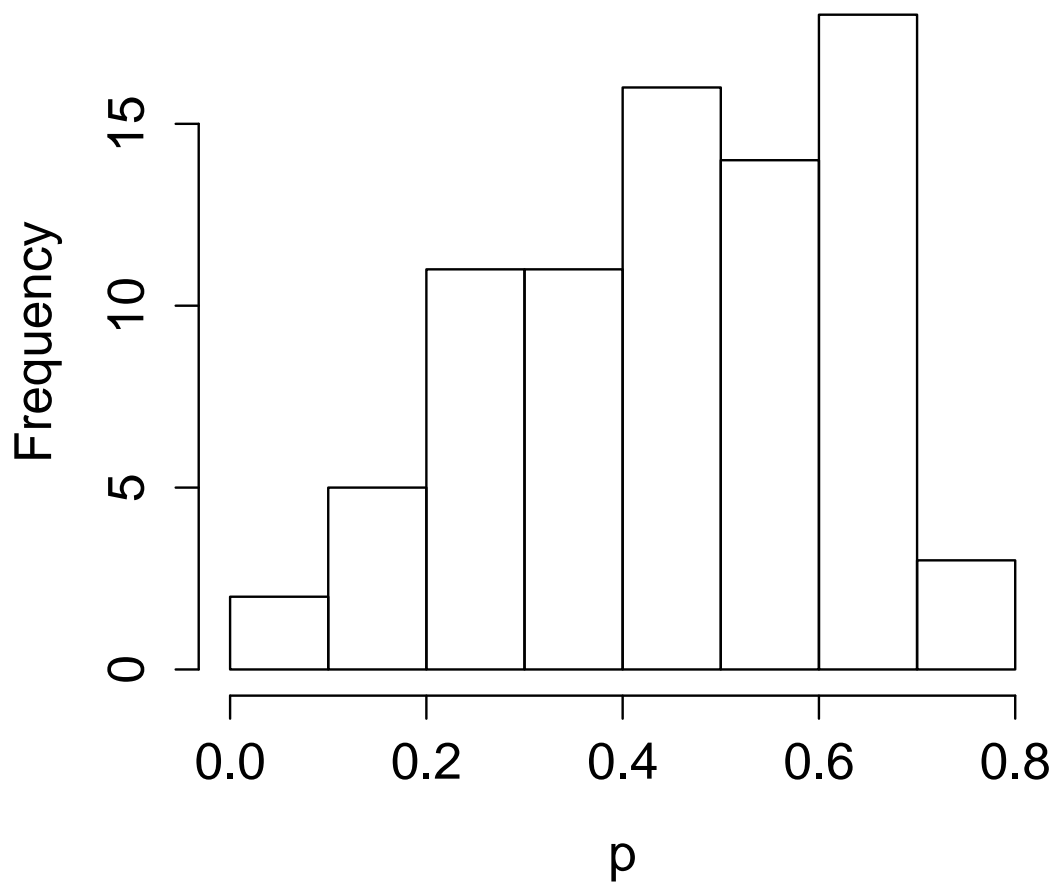


Figure 2.15: A histogram summarizing the approximate null distribution of the Bayesian p-value when  $c = 1$  and  $A = 5$ . The histogram summarizes results from 80 simulations.

estimated. In practical applications, however, auxiliary data from marked individuals will often come from the same population for which abundance is of interest. One option for getting an estimate of the total number alive is to include data from marked individuals in both portions of the likelihood (i.e., in both  $L_1$  and  $L_2$ ). Technically, this invalidates the independence assumption needed for a coherent joint likelihood. For instance, sample sizes will be inflated under this approach and thus we might expect an artificially high level of precision. However, the importance of this assumption for obtaining estimators with good properties has not been previously explored, and it is not clear how much overlap there is between the two data sets as far as the amount of information contained about each parameter.

In this module, I used simulation to compare the performance of estimators with and without data from marked individuals included in the age-at-harvest portion of the likelihood. I surmised that performance would likely be influenced by *i*) the number of animals marked each year, and *ii*) the complexity of the model fit to the data. For instance, if the number of marked animals is low in comparison to unmarked animals, then most of the data informing inference about population size comes from unmarked animals, which is independent of the mark-recovery portion of the likelihood. Similarly, if fewer parameters are used to describe the survival, harvest, and recruitment processes, there is more information in the age-at-harvest likelihood about them. I thus expected to see a greater degree of bias in coverage and CV when employing simpler models.

I considered a total of 6 scenarios to evaluate estimator performance, which differed by the number of marked animals released each year and by the estimation

model considered. The number of marked animals newly released each year was set to either 200 or 400, with the number released in each age category proportional to the number of animals in that age class. Three possible estimation models were considered in order to compare results across different levels of model complexity:  $S(\cdot)h(\cdot)f(\cdot)$ ,  $S(a)h(t)f(\cdot)$ , and  $S(a+t)h(a)f(t)$ .

For each scenario, I generated data for cases where marked animals were either included or ignored in the age-at-harvest matrix. Fifty replicate data sets were simulated in each case. I used age- and time-constant population parameters, with survival probability set to 0.6, a recovery rate of 0.2, and a recruitment probability of 0.4, with 5 age classes, an initial population size of 5000, and 5 years of data. When data from marked individuals were included in the age-at-harvest likelihood, population size included both marked and unmarked individuals; when only data from unmarked individuals were utilized, it included unmarked individuals only.

To quantify effects of non-independence on estimator performance, I once again compared the relative parsimony of different models where estimator performance was related to all possible subsets of predictor variables. In this case, I did not expect any changes with respect to bias, but I did expect that precision would be overestimated. Thus, I treated 90% BCOV for each parameter type (i.e., **N**, **S**, **h**, **f**) as possible response variables; predictor variables were the number of releases each year (200 or 400), the generating model *EstMod* (3 levels), and an indicator for whether or not data in each part of the likelihood were completely independent (*Ind*), which equaled 1 if all data were independent and 0 otherwise. In addition, whenever main effects terms were included in the model, I considered additional

models with all possible combinations of two way interactions. If all two way interactions were present, I also considered the possibility of a three way interaction.

When 90% BCOV for abundance was the response variable, overdispersion was estimated as  $\hat{c} = 2.5$ . Nevertheless, model selection using QAIC<sub>c</sub> favored highly parameterized models. I chose to base inference on the most general model, which was ranked second  $\Delta\text{QAIC}_c = 0.6$ , because it was the only model that included a highly influential interaction effect. In particular, when data from marked animals were also included in the age-at-harvest matrix, coverage was predicted to be much worse for the simple estimation model  $S(\cdot)h(\cdot)f(\cdot)$  (0.42) than for the other, more complex estimation models ( $\approx 0.7$ ) (Figure 2.16).

The story was similar for coverage of survival and recruitment parameters. For survival, a value of  $c$  was estimated as 1.6. The top QAIC<sub>c</sub> model included all effects except for the three way interaction and the two way interaction  $Mod \times Rel$ , and was used for inference because the second ranked model was 1.7  $\Delta\text{QAIC}_c$  units higher. Interval coverage was substantially lower than “nominal” when data from marked animals was included in the age-at-harvest matrix, especially when the simple estimation model  $S(\cdot)h(\cdot)f(\cdot)$  was used, and when the number of releases per year was high (Figure 2.17). For recruitment,  $\hat{c} = 1.1$ , so AIC<sub>c</sub> was used for model selection. In this case, the highest ranked model included all explanatory variables except for the three way interaction. Once again, coverage was particularly poor when  $L_1$  and  $L_2$  were dependent and  $S(\cdot)h(\cdot)f(\cdot)$  was used as the estimation model (Figure 2.18).

Surprisingly, coverage for recovery rates was not affected to the same degree

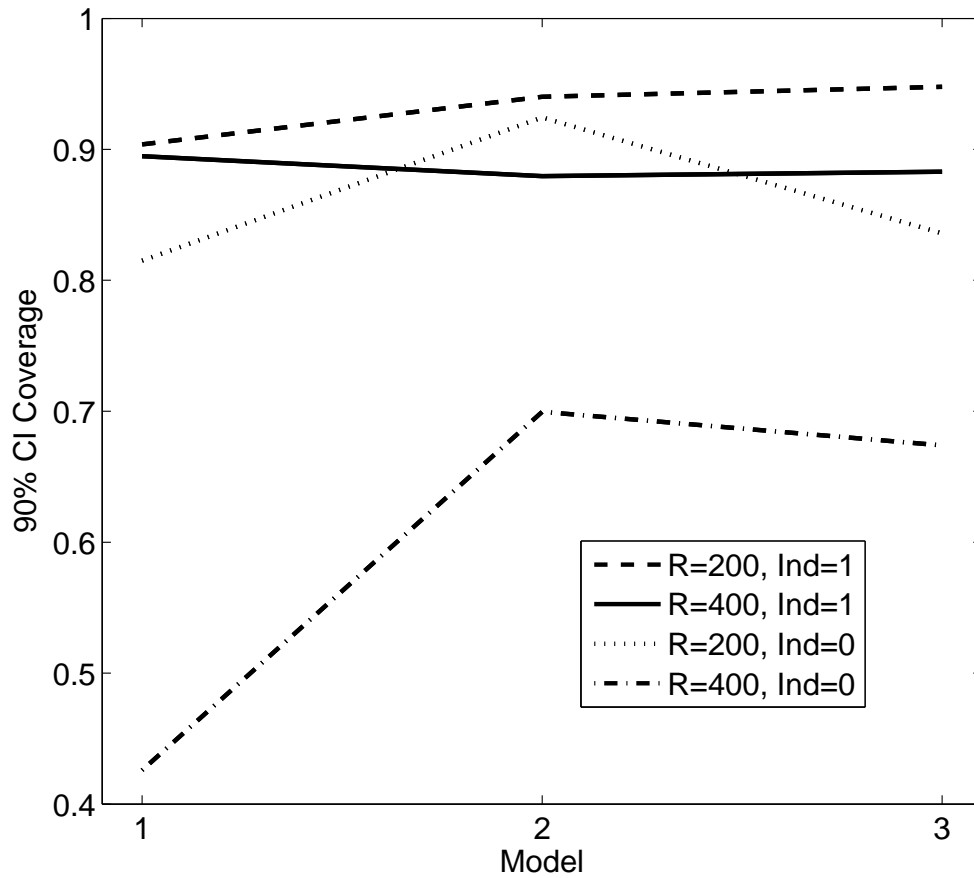


Figure 2.16: 90% Bayesian credible interval coverage on abundance ( $N$ ) for different model configurations, number of releases ( $R$ ), and depending on whether marked individuals were included in the age-at-harvest matrix ( $Ind = 0$ ) or not ( $Ind = 1$ ). Models “1,” “2,” and “3” were  $S(\cdot)h(\cdot)f(\cdot)$ ,  $S(a)h(t)f(\cdot)$ , and  $S(a + t)h(a)f(t)$ , respectively. Other notation is defined in the text.

by assumption violations. The top  $AIC_c$  model in this case included only  $R$  and  $Ind$  as explanatory variables. When  $Ind = 0$  and  $R = 400$ , coverage was predicted to be 0.84; when  $Ind = 0$  and  $R = 200$ , coverage was predicted to be 0.90; when  $Ind = 1$  and  $R = 400$ , coverage was predicted to be 0.89; when  $Ind = 1$  and  $R = 200$ , coverage was predicted to be 0.93.

## 2.8 Discussion

### *Sampling the Posterior Distribution*

None of the samplers I considered were able to explore from the posterior distribution very efficiently, but the Brownie parameterization for recoveries seemed to result in better mixing for at least some of the estimation models. In general, long Markov chains (along the order of  $1.0 \times 10^6$  iterations) were needed to generate consistent estimators of model parameters. Using marginal posterior means for calculation of DIC and point estimators usually resulted in a greater degree of repeatability.

### *Estimator Performance*

For the scenarios considered, Bayesian analysis of age-at-harvest and mark-recovery data generally resulted in estimators of population parameters with a low degree of bias and high degree of precision. Overall, bias was positive for abundance and recruitment, and negative for survival and recovery rate. As expected, bias decreased and precision increased with sample size, either through longer studies, a higher number of individuals marked each year, or increased harvest rates. Credible

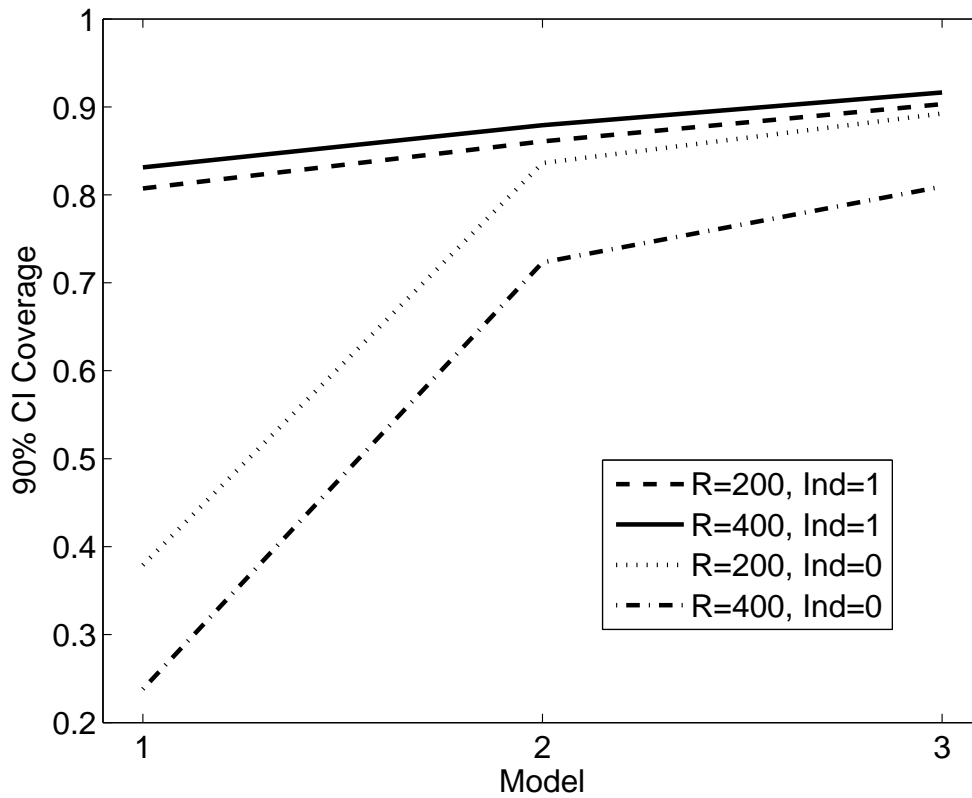


Figure 2.17: 90% Bayesian credible interval coverage on survival ( $S$ ) for different model configurations, number of releases ( $R$ ), and depending on whether marked individuals were included in the age-at-harvest matrix ( $Ind = 0$ ) or not ( $Ind = 1$ ). Models “1,” “2,” and “3” were  $S(\cdot)h(\cdot)f(\cdot)$ ,  $S(a)h(t)f(\cdot)$ , and  $S(a + t)h(a)f(t)$ , respectively. Other notation is defined in the text.

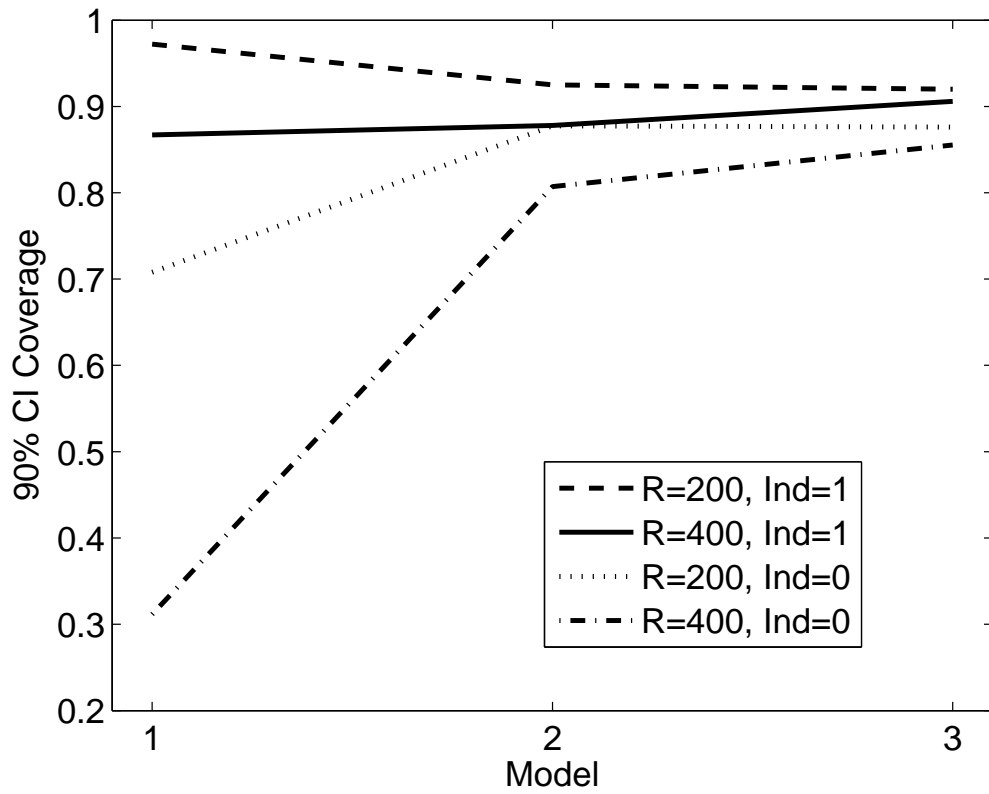


Figure 2.18: 90% Bayesian credible interval coverage on recruitment ( $f$ ) for different model configurations, number of releases ( $R$ ), and depending on whether marked individuals were included in the age-at-harvest matrix ( $Ind = 0$ ) or not ( $Ind = 1$ ). Models “1,” “2,” and “3” were  $S(\cdot)h(\cdot)f(\cdot)$ ,  $S(a)h(t)f(\cdot)$ , and  $S(a + t)h(a)f(t)$ , respectively. Other notation is defined in the text.



interval coverage was predicted to be close to “nominal” when all assumptions were met and when sample sizes were high.

### *Estimator Robustness*

Bias and coverage were quite robust to aging errors, at least with the range of models considered here. However, when data from marked animals were included in the age-at-harvest matrix, the potential for understatement of uncertainty was substantial. While use of estimation models with a sufficient degree of complexity helped, coverage was still poor when a large proportion of the population was marked each year. Ostensibly, including marked animals in both  $L_1$  and  $L_2$  artificially inflated sample sizes, resulting in inaccurate inferences about precision. When overdispersion was present, the goodness-of-fit test developed here showed good potential for diagnosing lack of fit in age-at-harvest data, especially when sample sizes were high.

Given that most applications of this methodology will involve estimating parameters for which marked animals are a part, an important consideration is how many animals to mark each year. In particular, it is important that enough animals be marked to reduce bias and to generate CV's that are acceptable for management purposes, but not enough to result in poor coverage (as in *Simulation Module 4*). This may prove to be quite a challenge for small populations. One ad hoc approach that may be useful in practice is to conduct simulations with input values relevant to the study population, artificially increasing the length of credible intervals until coverage is close to nominal.

## Chapter 3

# Potential for using age-at-harvest and radio telemetry data to monitor black bear populations in Colorado

### 3.1 Introduction

Results from Chapters 1 and 2 suggested that joint modeling with age-at-harvest data and data from marked individuals could be a viable monitoring strategy for wildlife. However, the sample sizes of marked individuals that were considered were larger than could be reasonably expected for many populations, including those of black bear in Colorado. There may be a multitude of factors that could contribute to limited sample sizes. For instance, habitat heterogeneity or nonuniform harvest

pressure could lead to substantial heterogeneity in survival and recovery rates, and perhaps to misleading inferences about abundance using this methodology. In this case, one solution is to perform stratification, but this tends to increase the number of parameters that need to be estimated, and thus increases requisite sample sizes. As another example, the effort required to mark individuals can vary substantially between taxa and/or the population in question. Pennsylvania black bears are highly productive in terms of recruitment and exhibit high population densities (Alt, 1989), while black bears in Colorado are considerably less fecund and sparsely distributed (Beck, 1991). Pennsylvania researchers also typically pool tagging data from research and management activities because they repeatedly fail to detect differences in demographic parameters between the two groups. In contrast, management bears in Colorado are subject to a “three strikes” rule and thus may be expected to have lower survival than the population as a whole. Thus, considerably more effort will be required to obtain a “representative” sample of marked individuals.

The purpose of this chapter is to determine whether joint analysis of age-at-harvest and radio telemetry data is a feasible method for monitoring black bear populations in Colorado. I concentrate on radio telemetry because telemetry studies typically provide more information on survival than mark-recovery studies for a set level of trapping effort. In describing feasibility, I present estimator performance for a number of scenarios, including those with varying study area size, levels of marking effort, number of years of data, and depending upon whether or not there is a covariance between marking probability and subsequent probability of harvest. In contrast to previous chapters, my focus here is on making inferences

about population trend rather than absolute abundance. Recasting the problem in this manner required reformulation of model structure, a topic which I discuss first. Second, I describe the methods I used to simulate data. In particular, including the possibility of a covariance between initial marking probability and subsequent harvest probability required that I employ an individual-based simulation model. Next, I describe the population model fit to the data, and explore estimator performance and robustness. Finally, I offer my own opinions on the viability of using this approach for monitoring black bear populations in Colorado.

## 3.2 Model Development

In Chapter 1, I advocated basing inference on the likelihood  $L = L_1 \times L_2$ , where  $L_1$  gave a likelihood for age-at-harvest data, and  $L_2$  specified a likelihood for auxiliary data. In this section, I describe changes to both portions of the likelihood. In particular,  $L_1$  must be modified to incorporate population trend, and alternative versions of  $L_2$  must be specified to acknowledge that radio telemetry data are used.

### 3.2.1 Incorporating Population Trend into $L_1$

As written in Chapter 1, the population process conditions on the abundance vector  $[N_{10}, N_{11}, \dots, N_{1A}]$ , while future age-specific abundances are treated as latent variables. Thus, total annual abundance (across cohorts) does not enter into the likelihood function, although posterior samples of this quantity for year  $i$  can still be generated by computing  $\sum_{j=1}^{A+1} N_{ij}$  across Markov chains (setting  $N_{1,A+1} = 0$

since it is not a parameter). To introduce total abundance from year  $i$ ,  $N_i$ , into the likelihood, I suggest replacing  $N_{i0}$  with the quantity  $N_i - \sum_{j=2}^{A+1} N_{ij}$ . Using this approach, it is possible to incorporate population trend into the model. In this Chapter, I focus on a linear trend for simplicity, letting  $N_i = \beta_0 + \beta_1 i$ , although it is certainly possible to specify nonlinear trends as well. In addition, all applications in this example use the alternative Seber parameterization for  $L_2$  presented in Chapter 2 because it is the most compatible with known fate (telemetry) data.

### 3.2.2 Likelihood for Telemetry Data

Telemetry data typically consist of known fates of individuals (i.e., live or dead is known) at discrete sampling occasions. In the case where no censoring occurs, likelihoods are simple to write down. For instance, if we utilize the same notation as in Table 2.1, but redefine  $t_k'$  to be the time interval in which an animal dies (if it dies), and  $J_k$  to be an indicator function for whether animal  $k$  dies within the course of the study, we may write  $L_2$  as

$$L_2 \propto \prod_{k=1}^M \Pr(H_k),$$

where

$$\Pr(H_k) = \begin{cases} (1 - S_{t_k', a_{t_k'}}) r_{t_k', a_{t_k'}} \prod_{i=t_k}^{t_k'-1} S_{i, a_i}, & I_k = 1, \\ (1 - S_{t_k', a_{t_k'}}) (1 - r_{t_k', a_{t_k'}}) \prod_{i=t_k}^{t_k'-1} S_{i, a_i}, & I_k = 0 \text{ \& } J_k = 1, \\ \prod_{i=t_k}^{Y-1} S_{i, a_i} & J_k = 0. \end{cases}$$

When censoring occurs at random, other approaches typical of survival analyses (e.g., Cox and Oakes, 1984) are appropriate. However, for simplicity, I do not explicitly consider this type of censoring. At times in this chapter, I make use of a deterministic censoring model where animals are removed from the study after being marked for 3 years. In this case, the probability of the given history is only calculated for the first three time intervals after marking.

### 3.3 Analysis Methods

#### 3.3.1 Simulating Data

As shown in Appendix A, there is substantial potential for bias in the Lincoln-Petersen estimator (Seber, 1982) of abundance when there is a relationship between initial marking probability and subsequent recovery probability. A positive covariance between these quantities might exist, for instance, when trapping occurs near roads and hunters depend on the same roads to try to find bears. In this case, hunters may be more likely to “sample” a tagged individual than an untagged individual.

To explore consequences of such a scenario, it was necessary to simulate data

at the individual level. To do this, I followed the template of White, Gill, and Beck (*Unpublished manuscript*), who conceptualized the female segment of a bear population as consisting of a number of individuals, each of which is assigned a vector of variables which describe the current state of the individual. This list of individuals is updated every year to account for mortality and recruitment to the population, and each individual's state vector is also updated to account for changes in age, reproductive status, etc.

At the beginning of each simulation run, each population was started at its stable stage distribution conditional on pre-specified levels of abundance, survival, and recruitment (see section 3.3.2). State vectors were compiled for each animal in the population at the beginning of each simulation run, as well as for bears recruiting into the population at future time steps. State vectors included current age, whether the bear had been marked or not, time at first capture, age at first capture, reproductive status (whether the bear currently has cubs, had cubs the previous year, or neither), as well as individual random effects for marking probability and harvest probability. To start simulations, I randomly assigned one cub to each breeding age (5 years or older) female until the number of females with cubs equalled the number of cubs in the stable stage distribution.

Assuming that harvest rate equaled recovery rate (i.e., that all harvested bears were reported to wildlife personnel), I partitioned total annual survival ( $S_{ij}$ ) into harvest mortality ( $h_{ij}$ ) and natural mortality ( $M_{ij}$ ) for simulation purposes. The population census date was immediately prior to harvest, so harvest mortality was simulated first, and natural mortality second. If an animal died, it was removed from

the population. If it was previously marked, its encounter history was appended to an encounter history list. If its death was because of harvest, it contributed to the age-at-harvest summary.

For breeding age females that survived the year and did not have cubs the previous year, it was assumed that they had cubs in the current year. However, the number of female cubs could be 0, 1, 2, or 3, depending upon the total number of cubs as well as the sex ratio of the litter. For the total number of cubs, I assumed that females would have 1 cub 25% of the time, 2 cubs 50% of the time, and 3 cubs 25% of the time. These are biologically reasonable values given results of previous studies of black bear in Colorado (Beck, 1991). I then assumed that the sex of an individual cub was female with probability 0.5 in order to determine the total number of female cubs per breeding female. Young of year survival (i.e., from birth until just prior to the hunting season) was assumed to be 0.793 regardless of other simulation inputs. This number was chosen so as to achieve zero average population growth for the case where populations were stable.

### **3.3.2 Simulation Design**

#### **Study Areas**

Bear habitat in Colorado is highly heterogeneous, and different locations face different amounts of harvest pressure (Gill and Beck, 1990). I thus concentrated on conducting power analyses that reflected populations that could be expected in several smaller, more homogeneous areas of the state. In particular, two possible



“study areas” were selected: Colorado Division of Wildlife Game Management Unit (GMU) 62, which encompasses the Uncompahgre Plateau, and black bear Data Analysis Unit (DAU) B-5, which encompasses GMUs 40, 60-62, 64-65, and 70 (all in west-central Colorado). To calculate the approximate number of bears I expected to be in GMU 62, I used habitat composition statistics (Anderson et al., 1992) to subjectively stratify GMU 62 into a 2000 km<sup>2</sup> block of “good quality” habitat and a 1000 km<sup>2</sup> block of low quality habitat. Applying density estimates of 1 bear per 5 km<sup>2</sup> to high quality habitats and 1 bear per 13 km<sup>2</sup> on the low quality habitats (see e.g., Gill and Beck, 1990), I arrived at an approximate figure of 477 bears. Given that sex ratios typically favor females in exploited populations, I conservatively estimated there to be around 250 female black bears in GMU 62. This quantity was used as a potential input into simulations. For DAU B-5, I simply doubled this figure (i.e., 500 females) as a simulation input.

### **Population Trajectory**

In this chapter, I focused on the ability of proposed methodology for diagnosing population declines. As such, I considered several scenarios. First, I considered a stable population; assuming a cub survival rate (August-August) of 0.6, as well as the implied recruitment rate, I set survival at 0.85 for yearlings and 2-year-olds, and 0.88 for older age classes to obtain a roughly constant population (eigenvalue based  $\lambda = 0.998$ ). Following educated guesses by Gill and Beck (1990), I assumed that recovery rates were 0.0 for cubs, 0.06 for yearlings and 2-year-olds, and 0.05 for remaining age classes. These values implied that survival in absence of hunting

(assuming an additive harvest model) would have been 0.904 for yearlings and 2-year-olds, and 0.926 for older individuals. These values are slightly less than reported by Beck (1991), but were chosen so as to achieve a stable population.

In order to simulate decreasing populations, I assumed an additive harvest model and simply manipulated recovery rates. Two scenarios were considered for decreasing populations. In the first, I increased recovery rates for yearlings and 2-year-olds to 0.08 and recovery rates for older ages to 0.07. In the second, I increased recovery rates for yearlings and 2-year-olds to 0.104 and recovery rates for older ages to 0.095. These yielded decreasing population projections, with an eigenvalue based  $\lambda = 0.981$  in the first case, and  $\lambda = 0.958$  in the second. With a value of  $\lambda = 0.981$ , deterministic populations will decline by about 10.4% in 5 years, and by about 17.5% in 10 years. With a value of  $\lambda = 0.958$ , the same population would decline by about 19% over a 5 year periods and by about 35% over a 10 year period. Actual simulations were stochastic, however, so realized population trend should on average be slightly less than for deterministic projections.

### **Covariance in Encounter Probabilities**

Given the potential for bias of abundance estimators when marking and harvest probabilities covary (Appendix A), I wanted to consider such covariance as a design factor. For simulation runs that included covariance, I generated

$$\begin{bmatrix} \epsilon_i^p \\ \epsilon_i^h \end{bmatrix} \sim f \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.10 & 0.05 \\ 0.05 & 0.10 \end{bmatrix} \right),$$

were  $f()$  denotes the bivariate normal distribution. For designs where covariance was not a factor, I generated

$$\begin{bmatrix} \epsilon_i^p \\ \epsilon_i^h \end{bmatrix} \sim f \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.10 & 0 \\ 0 & 0.10 \end{bmatrix} \right).$$

Marking probability and recovery rate for animal  $i$  were then generated as  $p_{ij} = \text{expit}(\text{logit}(p_j) + \epsilon_i^p)$  and  $h_{ij} = \text{expit}(\text{logit}(h_j) + \epsilon_i^h)$ , where  $j$  indicates an age effect, and  $p_j$  and  $h_j$  denote expectations of marking probability and recovery rate. Individual random effects were set to zero for cubs since harvest rates were zero by definition. However, for older age classes, this formulation produced correlations between marking probability and recovery rate of 0.5 when covariance terms were equal to 0.05 and variance terms equal to 0.10.

### **Marking Effort, Study Duration, and Censoring**

In any power analysis, focus is on the relationship between sample size and estimator performance. There are several possibilities for manipulating the sample size of marked individuals in telemetry studies. First, one can simply mark more or less “new” animals per year. Second, one can change the duration of the study, either increasing or decreasing the total number of marked animals when a fixed number of marks are put out per year. Lastly, one can choose to be more or less vigilant about the effort put forth to change radio collars. The batteries of radio collars put on black bears typically do not last for more than 4 or 5 years, so that locating bears in their dens during the winter lethargy period to replace collars can increase

effective sample size.

## Response Surface Study Designs

I employed two response surface designs to quantify estimator performance. In the first ( $RSD_1$ ), it was assumed that no censoring occurred. In the second ( $RSD_2$ ), I assumed that a collar would provide useful information for three years after marking. I assumed that these animals would not be caught during marking operations again, as black bears are notoriously intelligent and many learn to avoid traps after one such negative experience.

In  $RSD_1$ , I performed a complete factorial simulation design, varying each of the following factors:

- Starting abundance ( $N_{1.}$ ) = 250 or 500
- Expected number of new females marked/year ( $ER$ )= 10, 20, or 30
- Duration of study ( $Y$ )= 5 years or 10 years
- Correlation between  $p$  and  $h$ , ( $\rho$ ) = 0 or 0.5
- $\lambda = 0.998, 0.981, \text{ or } 0.958$

The result was a complete factorial design with 72 design points; at each of these points I simulated 5 different datasets.

For  $RSD_2$ , my main focus was on the 10 year study, where I felt that having no form of censoring would be somewhat unrealistic. I also thought that  $RSD_1$  would

suffice for exploring issues associated with correlation between encounter probabilities. Thus, I focused on the relationship between sample size and the ability to detect population trends. Accordingly, I performed a complete factorial design varying the following factors:

- $N_1 = 250$  or  $500$
- $ER = 10, 20,$  or  $30$
- $\lambda = 0.998, 0.981,$  or  $0.958$

The result here was a complete factorial design with 18 design points; at each of these points I simulated 5 different data sets.

### 3.3.3 Estimation

While the model for simulating data took on a degree of biological realism, I specified a relatively simple population model (Figure 3.1) for estimation purposes. When sample sizes are low, I did not expect that models with time- or age-varying components would be estimable. Similarly, one harvest rate was estimated for all individuals that were yearlings or older (the harvest rate of cubs was constrained to be 0). A simple linear trend on annual abundance was specified, such that

$$N_i = \beta_0 + \beta_1 i.$$

Estimation proceeded as in other chapters; MCMC was implemented according to the Metropolis-within-Gibbs hybrid update, with proposal standard errors chosen

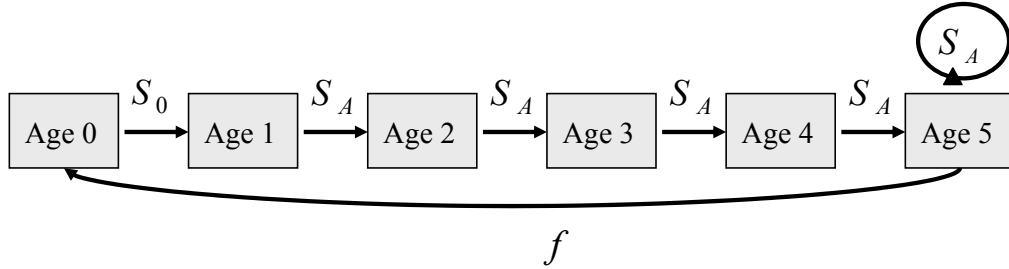


Figure 3.1: A pictorial depiction of the population process model for female black bears in Colorado.

so as to achieve a 35%-40% acceptance rate. Two chains with overdispersed starting values were run for 1,000,000 iterations each. If the Gelman-Rubin diagnostic (Gelman et al., 2004) indicated convergence, the second halves of each chain were combined to yield a sample of 1,000,000 from the posterior. If convergence was not indicated, the particular simulation was not included in response surface analyses. This sample was thinned to 200,000 by recording every fifth observation to save memory.

## 3.4 Estimator Performance

### 3.4.1 Measures of Estimator Performance

Since I was concerned with the ability to detect population trends, I focused on quantifying the performance of the linear population trend estimator  $\hat{\beta}_1$ . In particular, I was interested in bias, coefficient of variation (CV), and 90% credible interval coverage (BCOV). The latter was quantified by calculating whether or not the true value for  $\beta_1$  was within its estimated 90% credible interval, where the “true” value for  $\beta_1$  was determined by simple linear regression of the true values  $N_i$  on year. Thus, the “true” value for  $\beta_1$  varied from simulation to simulation due to stochasticity in realized population trajectories. As I was also interested in statistical power, I determined whether the 90% credible interval for  $\hat{\beta}_1$  included zero. The probability of a type I error could also be assessed in this manner. The number of simulations that resulted in converging Markov chains was also recorded. Since an improper prior was assumed for abundance, nonconverging chains may be indicative of one or more inestimable parameters.

### 3.4.2 Predictors of Markov Chain Convergence

If Markov chains did not converge after 500,000 iterations, it was a good sign that there were not enough data to estimate model parameters. Thus, predictions of whether or not chains would converge based on design factors provide a good indication of the utility of different levels of sampling effort. It is one thing to have an imprecise estimate, but another entirely to get no sort of “answer” at all.

Table 3.1: Summary of convergence rates (proportion of simulations converging within 500,000 iterations) by response surface design, duration of study (years), and expected number of marked releases per year. Response surface design 1 ( $RSD_1$ ), had a sample size of 60 per cell, while  $RSD_2$  had a sample size of 30 per cell.

Years	Design	Expected Releases		
		10	20	30
5	$RSD_1$	0.55	0.87	0.97
10	$RSD_1$	1.00	0.98	1.00
10	$RSD_2$	0.90	1.00	1.00

Simple tabulations of convergence rates for different combinations of years of data and number of expected markings per year are presented in Table 3.1. For  $RSD_1$ , convergence failures mainly occurred when there were five years of data and either 10 or 20 individuals marked per year. For  $RSD_2$ , there were some convergence failures, mainly associated with 10 releases per year.

### 3.4.3 Statistical Power for Detecting Population Declines

Conditional on Markov chain convergence, I examined statistical power for detecting population declines. I considered a population decline to be “detected” if the upper limit of the 90% Bayesian credible interval for the linear trend on abundance was less than zero. I conducted two analyses; one for  $RSD_1$  and one for  $RSD_2$ . In each case, I fit all combinations of predictor variables, including the possibility of two way interactions if comprising main effects were also included in the model structure. I also allowed for the possibility of the three way interaction  $\lambda \times ER \times Y$ . Model comparisons were made with  $AIC_c$ , and I selected one model within 2.0 units of the highest ranked  $AIC_c$  model for inference. This selection was made somewhat subjectively, and I typically selected a model that included highly visible effects,



but was simple enough to allow visual depictions of predictions.

For  $RSD_1$ , I selected the fifth ranked model ( $\Delta AIC_c = 0.50$ ), which included main effects for all predictor variables, as well as interactions  $\lambda \times Y$ ,  $\lambda \times ER$ , and  $\lambda \times N$ . This model included variables consistently appearing in all highly ranked models. Predictions of power to detect population trend could be made with the equation

$$\begin{aligned} \text{logit(Power)} = & 61.40 - 70.12\lambda + 10.578Y + .0146ER - .1973N_1. \\ & + 26.02I_\rho - 10.59\lambda Y + 0.210\lambda N_1. - 26.17\lambda I_\rho, \end{aligned}$$

where  $I_\rho$  is an indicator variable that takes on the value 1 if  $\rho > 0$  and 0 otherwise.

While there are too many predictor variables to visually portray power easily, predictions were heavily dependent upon the length of the study and the strength of population decline. When  $\lambda \approx 0.998$ , predicted power was typically less than 0.14, no matter what other input variables were considered. However, because  $\lambda$  was so close to one, this may be more indicative of type I error. When  $\lambda \approx 0.981$ , 5 year studies typically were predicted to have less than 20% power to detect population trends, while 10 year studies ranged from 5% to 44% depending upon input configuration. When  $\lambda \approx 0.958$ , power was predicted to be between 10% and 47% for 5 year studies, and between 56% and 91% for 10 year studies. Power was predicted to be higher for higher initial abundance, and for the case where there *was* correlation between encounter probabilities. Interestingly, expected releases per year appeared to be least important in detecting population trends in this design.

This was possibly due to the accumulation of radio collars over the course of the study when no censoring was employed.

For  $RSD_2$ , which included a more realistic approach to censoring for 10 year studies (but which did not include  $\rho$  as a design factor or allow for 5 year studies), I selected the 2nd highest ranked model ( $\Delta AIC_c = 0.49$ ) for inference. This model included effects for  $\lambda$ ,  $ER$ , and the interaction  $\lambda \times ER$ . In general, power predictions increased with decreasing values of  $\lambda$ , and increased when number of expected releases increased (Figure 3.2).

#### 3.4.4 Credible Interval Coverage

In order to investigate factors influencing 90% credible interval coverage for  $\hat{\beta}_1$ , I determined whether or not  $\beta_1^N$  fell within its 90% credible interval for each simulation. The true value of  $\beta_1$  was determined by regressing true abundance on time for each simulation (which could differ from simulation to simulation because of demographic stochasticity). As in the previous section, I compared the relative parsimony of a number of logistic regression models in order to investigate which factors had the greatest influence on 90% credible interval coverage.

For  $RSD_1$ , I selected the 6th ranked model ( $\Delta AIC_c = 1.4$ ) for inference, which included  $Y$  and  $N_1$ , the most consistent predictors with asymptotic 95% confidence intervals not including zero. Predictions indicated that coverage would be close to nominal (Figure 3.3). For  $RSD_2$ , I selected the highest ranked  $AIC_c$  model, which included the main effect  $N_1$  as its only predictor. According to this model, coverage

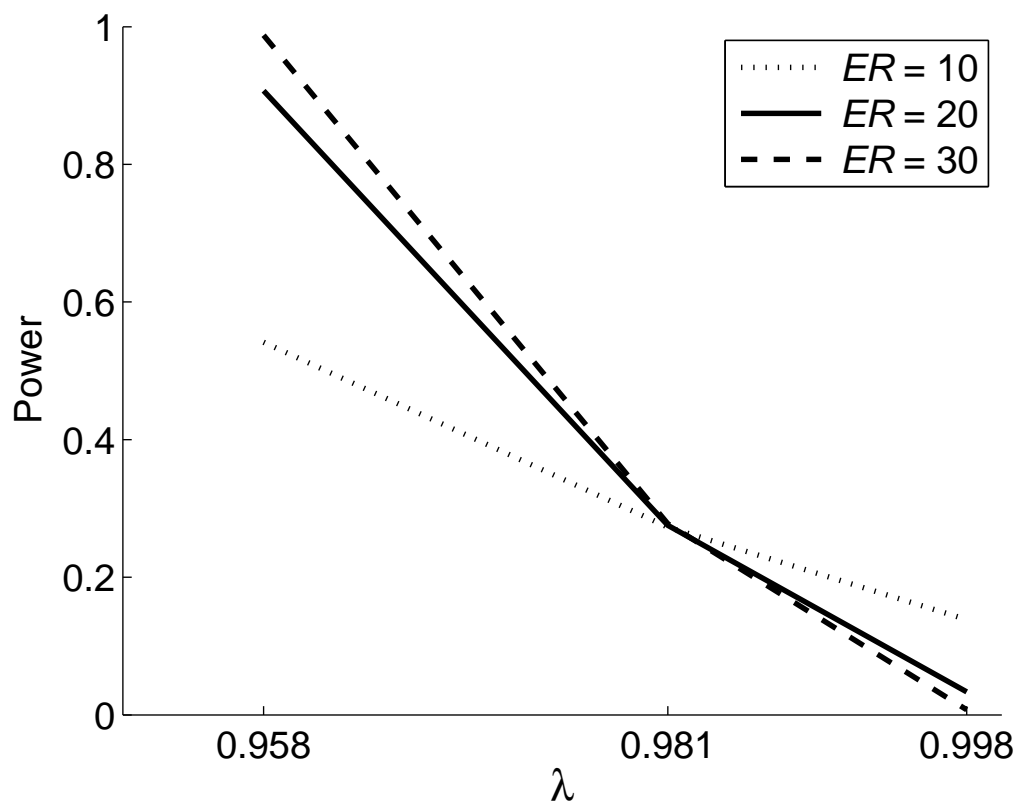


Figure 3.2: Power to detect population trends as predicted by response surface design 2. Important predictors included the finite rate of population increase ( $\lambda$ ), as well as the expected number of marked releases per year ( $ER$ ). Predictions of power at  $\lambda = 0.998$  are indicative of type I error rates, which should nominally be 0.10 at  $\lambda = 1.0$ .

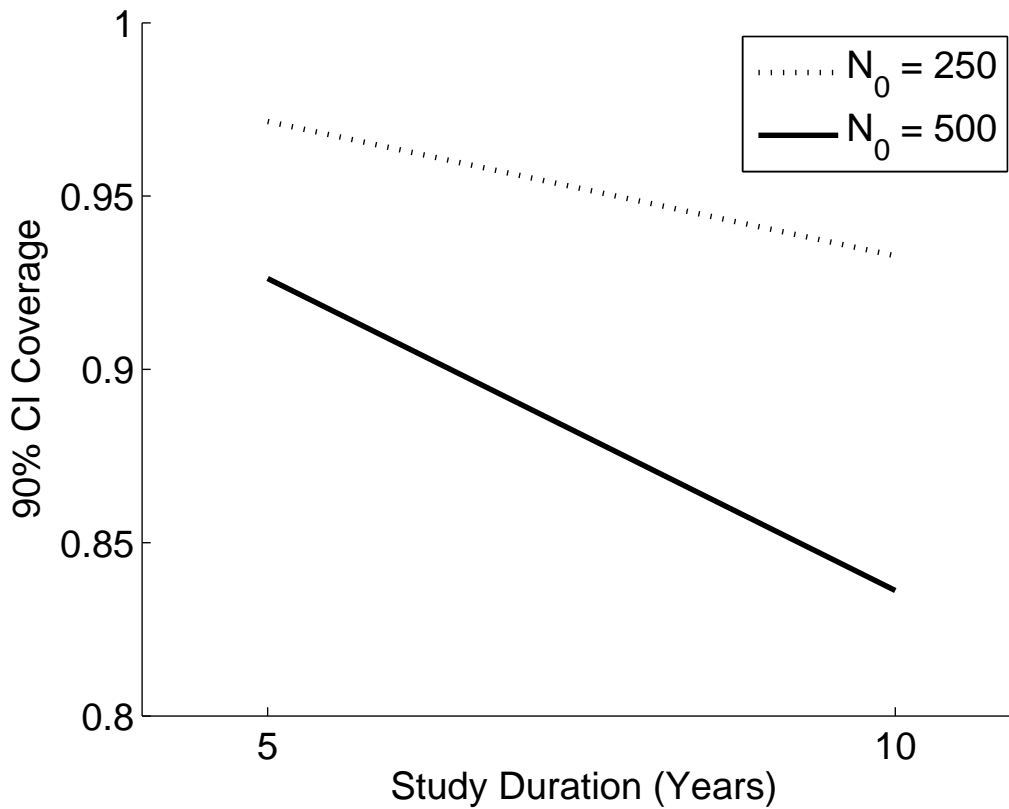


Figure 3.3: 90% Bayesian credible interval coverage on population trend as predicted by different levels of  $Y$ ,  $N_1$ , and  $I_\rho$  for response surface design 1.

was predicted to be 0.92 for an initial population size of  $N_1 = 250$ , and 0.91 for an initial population size of  $N_1 = 500$ . Thus it would seem that some of the variation in predicted coverage apparent in Figure 3.3 is due to sampling error, at least for the case where there is 10 years of data.

### 3.4.5 Bias

Percent relative bias is undefined for  $\beta_1$  when a population is stable ( $\lambda = 1$ ), so bias and absolute bias were quantified. However, in all cases I divided bias by  $N_{1,t}$  to put it on a relative scale similar to that of population trend ( $\lambda$ ). Ostensibly, an average change in abundance of 10 individuals per year would mean much more for a population of size 100 than for a population of size 10,000. Using this approach,  $\text{BIAS}(\hat{\beta}_1)/N_{1,t}$  was estimated to be approximately -0.008 (SE = 0.002) over all simulation inputs for  $RSD_1$ . For  $RSD_2$ , there was a more pronounced negative bias;  $\text{BIAS}(\hat{\beta}_1)/N_{1,t}$  was estimated to be -0.016 (SE = 0.002). Thus it would seem that there is an overall tendency to underestimate population trend, at least for the range of simulation inputs considered here.

To further examine factors affecting bias of population trend, I compared the relative parsimony of models expressing the response variable  $|\text{BIAS}(\hat{\beta}_1)/N_{1,t}|$  as a function of predictor variables. Here,  $\text{BIAS}(\hat{\beta}_1)$  was determined for each design point by averaging across results of all simulations at a given design point (maximum  $n = 5$ ). Models could include all combinations of simulation inputs as main effects, as well as all two way interactions and the three way interaction  $\lambda \times ER \times Y$  (provided that the main effects of these interactions were also included in the model). For  $RSD_1$ , a 0.3 power transformation of the predictor variable and removal of 1 outliers was necessary to meet model assumptions according to plots of quantiles and fitted values versus residuals (Fox, 2002). The top  $AIC_c$  model included effects for  $ER$  and  $Y$ , which were also included in all models within 2.0  $\Delta AIC_c$  units of

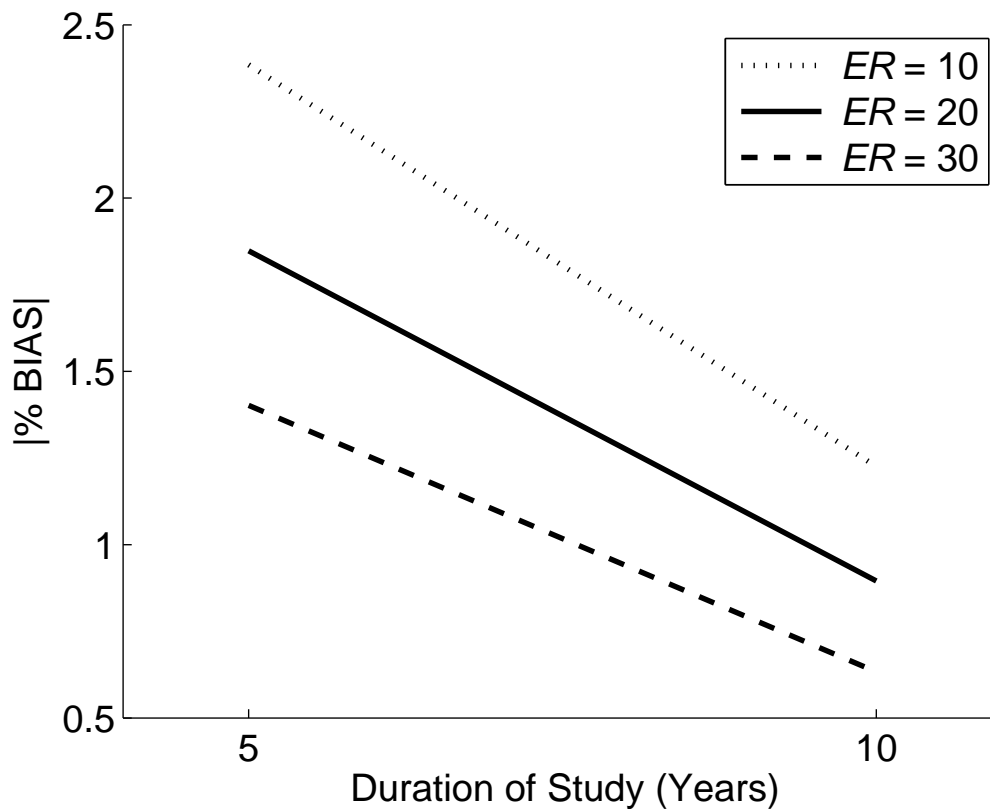


Figure 3.4: Average absolute percent relative bias for  $\beta_1/N_1$ , as predicted by different values of expected releases ( $ER$ ) and length of study for response surface design 1.

the top model. I thus used this model for inference, which predicted absolute bias to decrease as sample size (i.e., either  $ER$  or  $Y$ ) increased (Figure 3.4).

A power transformation of the response variable did not appear necessary for  $RSD_2$ . In this case, the top  $AIC_c$  model included effects for  $ER$ ,  $N$ , and the interaction  $N \times ER$ , predicting that absolute scaled bias would decrease with the number of expected releases at lower initial abundances, but would be increasingly similar at higher abundances (Figure 3.5).

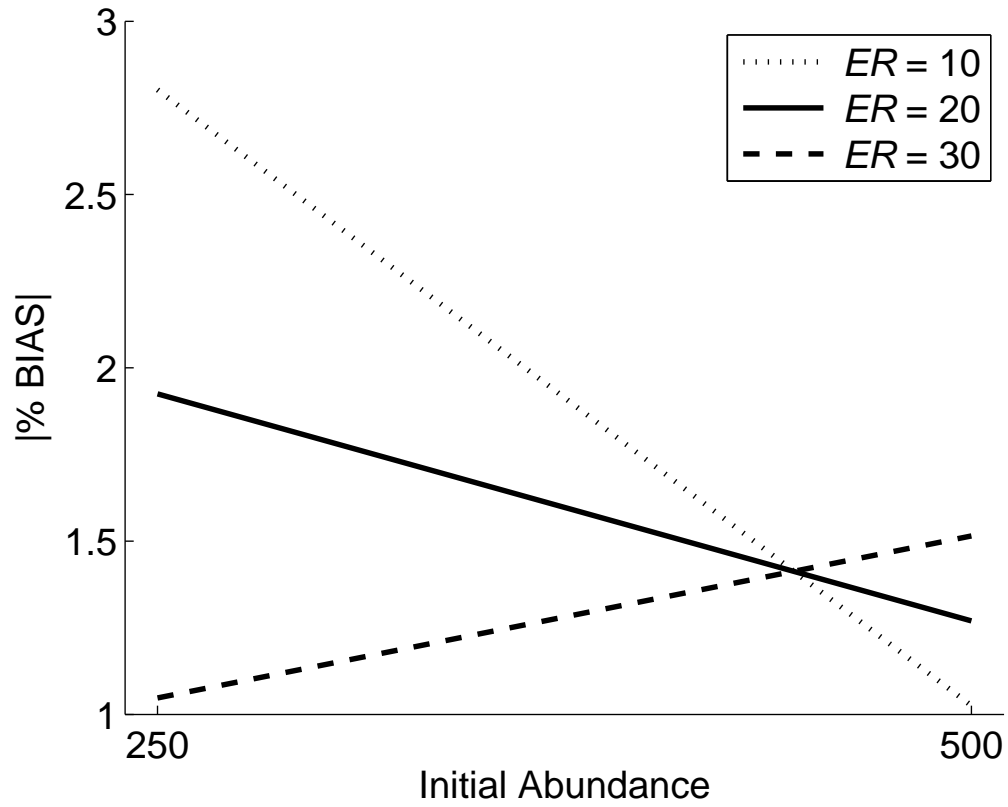


Figure 3.5: Average absolute percent relative bias for  $\beta_1^N/N_1$ , as predicted by different values of expected releases ( $ER$ ) and initial abundance for response surface design 2.

### 3.4.6 Coefficient of Variation

As was the case for percent relative bias, the CV of  $\beta_1$  is undefined for values of  $\beta_1$  approaching zero. However, converting it to the same scale as  $\lambda$  is one option; for instance,  $(N_1 + \hat{\beta}_1)/N_1$  is an estimate of the average annual per capita growth rate of the population over the course of the study (however, note that this is not rigorously justified; see section 3.5). An estimate of CV for this quantity is just

$$\hat{CV}(\hat{\lambda}) = \frac{\sqrt{\hat{\text{Var}}(\hat{\beta}_1)}}{N_1 + \hat{\beta}_1}.$$

I calculated this quantity for each simulation, and sought to identify factors explaining it's variation.

As par for the course, I compared  $AIC_c$  for a variety of linear models relating the response variable  $\hat{CV}(\hat{\lambda})$  to predictor variables. All variable simulation inputs were considered as candidate predictor variables, as were all two way interactions and the three way interaction  $\lambda \times Y \times ER$ . For  $RSD_1$ , a power transformation of -0.4 on the response variable seemed sufficient to meet linear model assumptions. The highest ranked  $AIC_c$  model included all main effects except for  $I_\rho$ , as well as the interactions  $N_1 \times ER$  and  $Y \times \lambda \times ER$ . Using estimated regression coefficients from this model, CV could be predicted as

$$\begin{aligned} (\hat{CV}(\hat{\lambda}))^{-0.4} = & 8.36 - 9.35\lambda + 0.4467Y - 0.00108N_1 + 0.057ER \\ & + 0.000037N_1 \times ER - 0.00447\lambda \times Y \times ER. \end{aligned}$$



For 5-year studies, CV was predicted to range from 0.03 to 0.18, decreasing with the number of marked releases and to a lesser extent  $\lambda$ . For 10 year studies, CV was predicted to range from 0.014 to 0.031.

For  $RSD_2$ , I removed six outliers for analysis; all but one of these was associated with 10 releases per year, and all had a CV several orders of magnitude greater than other simulation runs. Following their removal, a -0.5 power transformation of the response variable was necessary to better meet linear model assumptions. Model selection favored models incorporating all main effects and the interaction  $ER \times N_1$ . (the highest ranked  $AIC_c$  model included no other effects). Conditional on this model, predictions were that CV would decrease as the number of expected releases increased (Figure 3.6).

### 3.5 Discussion

Simulation-based performance of population trend estimators indicate that joint age-at-harvest and radio telemetry analysis may be a viable strategy for monitoring black bear in Colorado. However, ten-year studies are clearly preferable, as five-year studies had more bias, less precision, and scant power to detect population trends. Further, for many five-year datasets, Markov chains did not even converge, implying that there were too little data to estimate parameters. In contrast, ten-year studies had reasonable power to detect population trends, especially when the expected number of bears that were annually tagged was high. As quantified by CV, simulations also indicated that  $\lambda$  could be estimated with a reasonable degree

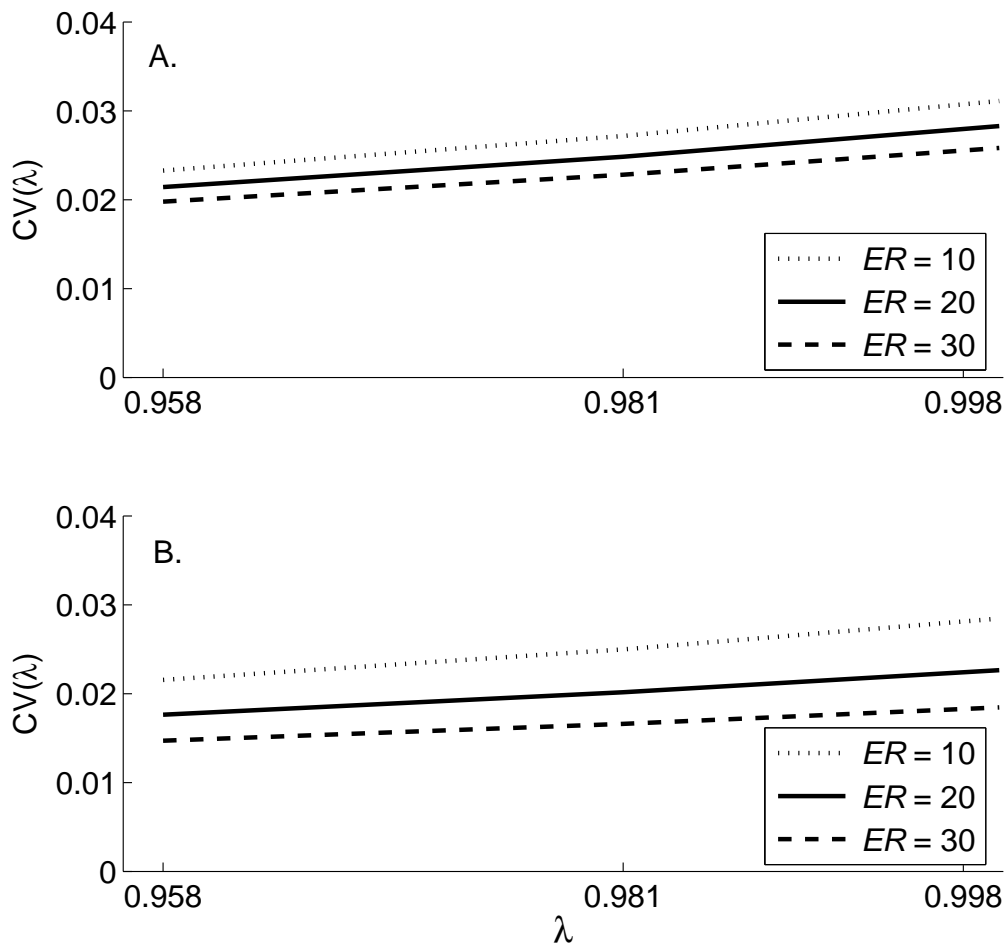


Figure 3.6: Predicted coefficient of variation ( $CV$ ) associated with an estimator of population growth rate ( $\hat{\lambda}$ ) for different levels of  $\lambda$ , expected number of releases ( $ER$ ), and initial population size. Panel A. gives predictions for  $N_1 = 250$ , while B. gives predictions for  $N_1 = 500$ .

of precision for ten-year studies.

Contrary to my expectations, it appeared that covariance between initial marking probability and subsequent harvest probability only served to enhance estimator performance. While it likely biased estimators of abundance, inference about population trend was seemingly unaffected. There was, however, an overall negative bias in estimators of  $\lambda$  that is somewhat worrisome. While this bias was small, and decreased with the number of marked releases each year, it could tend to lead to overly conservative management decisions. Whether this bias was due to misspecification of the estimation model relative to the model used to simulate data, to unmodeled heterogeneity, or to other factors is unknown. However, because these factors are likely operable in Colorado as well, there will likely be similar biases in field applications.

The estimation model employed in all simulations did not include time effects on demographic or harvest parameters. As such, it was assumed that survival, recruitment, and recovery rates remained constant over the course of the study. In practice, this is unlikely to be the case. For instance, the number of females recovered in Bear DAU 5 has varied considerably over the years (J. Apker, Colorado Division of Wildlife, *Unpublished data*). However, simple estimation models will likely be required to be able to estimate model parameters due to data sparseness. It may actually be possible to consider a time-varying survival model, but the model for recovery rates will likely need to be time-constant or at most to incorporate a linear trend effect.

Finally, I suggest several possible improvements to the model structure and

simulation design used here. First, the linear trend model used in simulations assumes that population trends are linear on the real scale. A more realistic expectation is that trends are linear on the log scale, although the difference between these two models is small over the range of population size and trends considered in this chapter. Still, from a demographic standpoint, a model such as  $\log(N_{i\cdot}) = \beta_0 + \beta_1 i$  would be a more coherent choice. Second, I note that scaling the trend estimator by abundance is a rather *ad hoc* way of arriving at  $\lambda$ . In theory, it should be possible to enter  $\lambda$  as a parameter in these models in a similar fashion to Pradel (1996). However, recruitment rate parameters would likely need to be removed from the model in this case. Another approach would simply be to keep track of  $\lambda_i = N_{i\cdot}/N_{i-1\cdot}$  throughout MCMC simulations to generate marginal posterior predictive distributions for these quantities.

## Chapter 4

# Adjusting Age and Stage

# Distributions for Misclassification

# Errors

## 4.1 Introduction

Ecologists often use data from the standing age- or stage-structure of a population to make inferences about population level processes. For instance, such data may be used to calculate survivorship and other demographic parameters using life table approaches (Caughley, 1977; Gotelli, 2001; Skalski et al., 2005), to estimate fertilities and stage-transition frequencies using stable stage distribution methods or inverse methods for time series (Caswell, 2001), and to model the status of fish stocks using quantitative stock assessment methods (cf. Fournier and Archibald, 1982; Megrey, 1989; Meyer and Millar, 1999). Even epidemiological force-of-infection models may

be fit with such data (Heisey et al., 2006).

When conducting analyses with age- or stage-structure data, several assumptions are typically required. For instance, an assumption of a stable age- or stage-distribution is often needed (Caswell, 2001). Another condition that frequently arises is when detection of animals is less than perfect, which may serve to bias estimators when detectability is related to age or stage class. Some assumptions may be relaxed to varying degrees by collecting ancillary data. For example, data from marked animals may be used to relax the assumption that detection probabilities do not vary by age- or stage-class.

One assumption common to most analyses involving age- or stage-structure data is that age- and stage-classes are determined without error. Otherwise, inferences based on analysis of such data can be compromised. For instance, an aging error probability of 0.2 for a single age class can bias survival estimators based on age ratios (cf. Ricklefs, 1997) by up to 20% (Conn et al., 2005). For many animal populations, positive biases in age determination are often observed for young individuals, while negative biases are observed for older animals (e.g., Coy and Garshelis, 1992; Harshyne et al., 1998; Hewison et al., 1999). If the standing age distribution is used for analysis in this case, one would typically arrive at overestimates of survival or stage-transition probabilities at younger ages, and underestimates at older ages. In a population modeling or population viability analysis context, these errors may lead to overly pessimistic projections of population trend for many populations since population growth rate is often heavily influenced by adult survival. In a disease context, ignoring errors in age-determination may serve to bias the shape of infection

hazard profiles (e.g., Heisey et al., 2006).

Ecologists have realized possible problems with interpreting raw age- and stage- distributions, and have sometimes focused energy on quantifying measurement error. For instance, in passerines, plumage characteristics may overlap between age classes, resulting in aging error (Green et al., 2001). In many mammal species, including black bear (*Ursus americanus*) and white-tailed deer (*Odocoileus virginianus*), characteristics of cementum annuli from teeth have been used to age individuals. However, investigations using data from known-age individuals have found that use of cementum annuli to age individuals is not error-free (e.g., Coy and Garshelis, 1992; Harshyne et al., 1998; Hewison et al., 1999). Similarly, in fisheries studies, otolith rings often are used to age fish, but concern has been raised about the possible accuracy of such data and implications of age-reading errors (cf. Beamish and McFarlane, 1983; Fournier and Archibald, 1982; Reeves, 2003).

Given the widespread use of age- and stage-based methods in ecology, and the effort that ecologists have sometimes dedicated to estimating error rates in age- and stage-determination, it is perhaps surprising that little attention has been paid to incorporating such errors. For the most part, the tendency has been for researchers to *a*) discard data that have been shown to contain a large amount of measurement error, or *b*) ignore measurement error, particularly if error rates are low. Several exceptions exist; for instance, Green (2004) corrected for aging misclassification in corncrakes (*Crex crex*) with a normal mixture model. However, his method was specific to two age classes. Similarly, Pella and Robertson (1979) and Richards et al. (1992) advocated use of an inverse classification matrix for estimating age distri-

butions. This latter approach acknowledges uncertainty about classification matrix values as estimated from auxiliary data on aging error (i.e., the  $\tau_{ij}$  described below), but ignores the additional stochasticity that should be associated with the observation of ages in the particular sample of interest. For example, even if classification probabilities were known exactly, we would still expect variation among samples with regard to the number of animals correctly or incorrectly classified. Thus, although this approach is reasonable when there are large sample sizes (as with many fisheries applications), there may be substantial negative bias in variance estimators in other situations.

The current paucity in available methods for correcting age- and stage- based data for measurement error suggests a need for a general statistical framework with which to address such problems. My current objective is to present a methodology for estimating the age- or stage-structure (hereafter, AS) of a sample by incorporating measurement error. My approach includes two conceptually distinct likelihood components,  $L_1$  and  $L_2$ . The first component,  $L_1$ , relates observed AS data to classification probabilities and the true AS structure and should be generally applicable to a wide class of problems. The second component,  $L_2$ , consists of a model relating observed measurement errors to classification probabilities, and will likely require tailoring to the specific problem at hand. This component also requires that the investigator has auxiliary data that can be used to estimate measurement error rates, as with data from known AS class. After describing an underlying statistical framework, including description of  $L_1$ , I apply this approach to age-structured data from harvests of black bears in Pennsylvania. In this example I formulate models for  $L_2$



that are specifically tailored to our focal study species. Nevertheless, I anticipate that this description will help to serve as a template for how one might formulate models for other taxa and types of misclassification.

## 4.2 Statistical Methods

To begin, assume that there are  $m$  AS classes, and that there are  $n$  samples (e.g., years) from which we wish to estimate an AS distribution. Further, let  $C'_{ij}$  be the number of individuals that are (perhaps erroneously) classified into AS class  $j$  in sample  $i$ ,  $C_{ij}$  be the number of individuals that are truly of AS class  $j$  in sample  $i$ , and  $\tau_{kj}$  be the probability of classifying an individual into AS class  $j$  given that it is truly of AS class  $k$  (note that  $\sum_j \tau_{kj} = 1$ ). For purposes of this chapter, I assume that the misclassification probabilities are constant across samples, although extensions to relax this assumption are relatively straightforward.

Conditioning on the total number of individuals caught in sample  $i$ ,  $C_i$ , it is possible to formulate a product multinomial likelihood for the observed data, given the true AS distribution and associated misclassification rates. In particular,

$$L(\mathbf{C}|\mathbf{C}', \boldsymbol{\tau}) = \prod_{i=1}^n \frac{C_i!}{C'_{i1}! C'_{i2}! \dots C'_{im}!} \pi_{i1}^{C'_{i1}} \pi_{i2}^{C'_{i2}} \dots \pi_{im}^{C'_{im}} \quad (4.1)$$

Here, bold symbols denote collections (vectors) of parameters, and  $\pi_{ij}$  is the unconditional probability that an individual in sample  $i$ , picked at random, is classified as a member of AS class  $j$ . This probability can be expressed as a function of the

true AS distribution and the number of individuals sampled:

$$\pi_{ij} = \frac{C_{i1}\tau_{1j} + C_{i2}\tau_{2j} + \dots + C_{im}\tau_{mj}}{C_i}.$$

Even if misclassification rates are known, this model contains one too many parameters for each sample; a possible solution is to replace  $C_{i1}$  with  $C_i - \sum_{j=2}^m C_{ij}$ .

If misclassification rates are known, maximum likelihood inference may be used to estimate the true AS distribution and its standard error directly from (1). Typically, however, the investigator will not know misclassification rates and must estimate them using auxiliary data, such as when a sample of known-age individuals can be used to estimate the  $\tau_{kj}$ . In this case, a solution is to perform maximum likelihood on the product

$$L = L_1 \times L_2,$$

where  $L_1$  is given by (1), and  $L_2$  is a likelihood for the auxiliary data set. This product likelihood requires an assumption of independence between  $L_1$  and  $L_2$ , so that data from known age individuals, for instance, should not be included directly into the sample for which AS distribution is to be estimated. In practice, a reasonable solution is to estimate AS distribution for those individuals whose true AS class is not known, and then to add in animals of known age as a posthoc correction. Despite this nuance, considerable flexibility is afforded by our formulation, because  $L_2$  can be specifically tailored to the focal study species and the type of auxiliary data available. In addition, issues of model complexity and fit may be explored by

comparing different non-parametric and parametric models for the  $\tau_{kj}$ . A specific example is provided below.

### 4.3 Black Bear Example

The Pennsylvania Game Commission (PGC) collects harvest data for black bears at mandatory hunter check stations as part of its annual monitoring program (Diefenbach et al. (2004)). During this process, premolars are removed from dead black bears for age determination (Harshyne et al. (1998)). I used data from 1983-2004 to assemble an age-structured harvest matrix for females that was, ostensibly, subject to measurement error. I also compiled records from known-age females for this same time period, noting any errors made in age determination at the time of harvest (Table 4.1). Although interest focused on the number of bears harvested in each of six age classes (0-4 or 5+), I initially classified bears into one of 13 age classes (0-11 or 12+) for analysis to increase precision on parameters describing the aging error process. Known-age bears were those initially live captured as a cub or as a yearling still with its mother; blind cementum annuli analyses with recovered or recaptured known-age bears provided auxiliary information on aging error (for details, see Harshyne et al. 1998).

My goal was to estimate the age distribution of annual harvest samples, adjusted for misclassification rates. To do this, I needed to formulate models relating observed aging errors to misclassification probabilities (i.e., the  $\tau_{kj}$ ). My strategy was to compare the relative parsimony of several such models that represented dif-

Table 4.1: Aging errors by age class for known-age female black bears in Pennsylvania from 1983-2004.

True age	Classified age											
	1	2	3	4	5	6	7	8	9	10	11	12+
1	257	11	1	1	0	0	0	0	0	0	0	0
2	2	58	4	1	0	1	0	1	0	0	0	0
3	2	4	38	1	1	1	0	0	0	0	0	0
4	1	1	0	27	0	0	0	0	0	0	0	0
5	0	0	0	2	30	0	0	0	0	0	0	0
6	0	0	0	0	1	6	0	0	0	0	0	0
7	0	0	0	0	1	1	6	1	0	0	0	0
8	0	0	0	0	0	0	0	7	0	0	0	0
9	0	0	0	0	0	0	1	2	3	1	0	0
10	0	0	0	0	1	0	0	1	1	3	0	0
11	0	0	0	0	0	0	0	0	0	2	2	0
12+	0	0	1	0	0	0	0	0	0	0	1	10

ferent working hypotheses about how error rates changed with age. Because of the relative sparseness of known-age data (Table 4.1), I only considered parametric models for aging error. A non-parametric model based on aging error proportions, for instance, would in this case result in a positive probability of incorrectly aging a four-year-old as a one-year-old, but a zero probability of aging a four-year-old as a three-year-old. Following Richards et al. (1992), I considered models for aging error based on the normal density, but also included Laplace models that allowed for a more leptokurtic (i.e., fat-tailed) distribution for error rates. The Laplace (or double exponential) model is specified by the probability density function

$$f(x|\mu, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x - \mu|}{\sigma}\right).$$

Since atypical or broken teeth may result in larger errors than “normal” teeth

(Harshyne et al., 1998), I anticipated that the distribution for aging errors would have heavier tails than specified by the normal density.

In total, I compared the relative parsimony of six models for aging error (Table 4.2). Models could vary by parametric form (Normal or Laplace), and by the parameterizations for the mean and standard error of aging error distributions. I considered two parameterizations for mean error. In the first, the mode of the classified age was set to the true age (model  $\mu(a)$ ); in the second, the mode of classified age was allowed to change as a function of age, such that  $\mu = a + \beta a$  (model  $\mu(A)$ ). I included the second parameterization because negative biases in age determination are often observed for older age classes. Similarly, I allowed standard error to remain constant across age classes (model  $\sigma(\cdot)$ ) or to be a linear function of age class (model  $\sigma(A)$ ). I introduced the latter specification for standard error because I expected that the variance of aging error distributions would increase with age since older individuals have more cementum annuli than younger individuals, and thus more opportunity for irregularities in cementum deposition. Further, the premolars of older individuals are more susceptible to breaking during tooth extraction, a factor which can contribute to errors in age determination (Harshyne et al. 1998). Regardless of the specified model, a separate  $\sigma$  was always estimated for individuals in the 12+ age class since some individuals in the 12+ age category could be much greater than age 12, and thus have relatively small misclassification probabilities. For instance, an 18-year-old individual would have a small probability of being misclassified than an 11-year-old. The aging error model for this category could then be interpreted as applying to an individual from this group chosen at random.

Although the underlying, unconditional models for aging error were continuous, observed error rates were discrete as well as being right- and left-truncated. For instance, it is virtually impossible for a 2-year-old to be aged as a cub because cubs can be identified without error by the presence of milk teeth. In this case, errors could only be positive. In order to accommodate restrictions of this type, it was necessary to condition on permissible values for error rates. Rather than integrating over the relevant continuous probability distribution, I again followed Richards et al. (1992) and set discrete aging error probabilities based on values of the continuous probability density evaluated at discrete points, normalized to obtain a true probability mass function given the constraints. In addition, because the age 12+ category represents a pooling of age classes greater or equal to twelve, the probability of classifying an individual less than age twelve into this class could be given by summing probabilities of incorrectly aging an animal as age twelve or higher. Thus, the probability of classifying an individual into age class  $j$  given that it is of true age  $k$ ,  $k < 12$ , can be obtained as

$$\tau_{kj} = \begin{cases} c f(j|k, \sigma_k) & 0 < j < 12 \\ c \sum_{j=12}^Y f(j|k, \sigma_k) & 12 \leq j \leq Y \end{cases},$$

where  $Y$  gives a “large” age,  $f(x|\alpha, \beta)$  gives a continuous probability density function (e.g., Normal or Laplacian) with location parameter  $\alpha$  and scale parameter  $\beta$ , and

$$c = \frac{1}{\sum_{i=1}^Y f(i|k, \sigma_k)}.$$

In practice, it suffices to take  $Y$  sufficiently large so that  $\sum_{k=Y}^{\infty} f(Y|k, \sigma_k)$  is close to zero for  $k < 12$ . The aging error model for individuals that are truly of age category 12+ was similar, but here I simply considered the categorical model

$$\tau_{12,j} = \frac{f(j|k = 12, \sigma_{12})}{\sum_{i=1}^{12} f(i|k = 12, \sigma_{12})} \quad 0 < j \leq 12.$$

Following this parameterization, one still needs to formulate a likelihood for the auxiliary aging error data. Given a finite number of possible classes for aging error, it is once again natural to consider a multinomial form:

$$L_2 = \prod_{k=1}^{12} \frac{M_k!}{X_{k1}! X_{k2}! \dots X_{k,12+}!} \tau_{k1}^{X_{j1}} \tau_{k2}^{X_{k2}} \dots \tau_{k,12+}^{X_{k,12+}}.$$

Here,  $M_k$  gives the total number of individuals that were of known age  $k$  when aged at the time of harvest and  $X_{kj}$  gives the total number of individuals that were classified into age category  $j$  given they were of true age category  $k$ .

All model fitting was performed via maximum likelihood in SAS/IML (SAS Institute Inc, 2004); requisite code is available from the author. My approach was to first compare the parsimony of different models for aging error data (i.e.,  $L_2$ ), and then to use the highest ranked model to estimate the age distribution of harvest samples. For the second step, I used a log link on all  $C_{ij}$  parameters to ensure positive values for all estimates.

Comparing the relative parsimony of aging error models using  $AIC_c$  (Burnham and Anderson, 2002) provided convincing evidence that aging error was better

Table 4.2: Summary of model selection results for aging error models. Reported are  $K$ , the number of parameters in each model,  $\Delta\text{AIC}_c$ , the difference in  $\text{AIC}_c$  score from the top model, and  $\log L$ , the log likelihood. The effective sample size was 496, the total number of known-age bears.

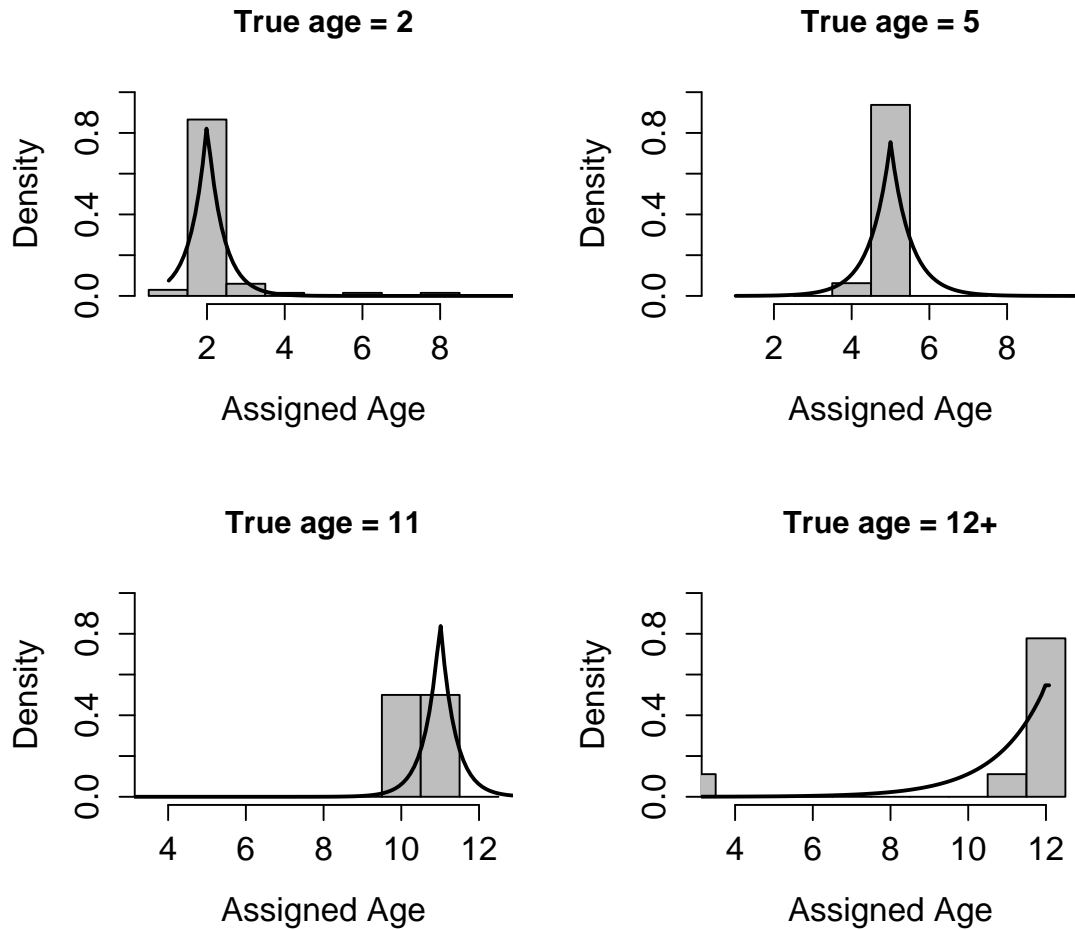
Model	$K$	$\Delta\text{AIC}_c$	$\log L$
Laplace $\mu(a) \sigma(A)$	3	0.0	-251.7
Laplace $\mu(A) \sigma(A)$	4	0.67	-251.0
Laplace $\mu(a) \sigma(\cdot)$	2	12.0	-258.7
Normal $\mu(A) \sigma(A)$	4	255.5	-380.4
Normal $\mu(a) \sigma(A)$	3	262.2	-383.8
Normal $\mu(a) \sigma(\cdot)$	2	319.7	-412.6

modeled using the leptokurtic Laplace distribution and strong evidence that the standard error of the underlying distribution for aging errors increased with age. However, there was no substantial evidence that the mode of aging error distributions changed with age (Figure 4.1, Table 4.2). The highest ranked model did not fit the aging error data according to a chi-square goodness-of-fit test ( $\chi^2_8=15.8$ ,  $p=0.05$ ). However, when one outlier (a 12+ individual incorrectly aged as a three year old) was removed the fit improved ( $\chi^2_7=12.0$ ,  $p=0.10$ ). In calculating the test statistic, I pooled cells with expectations less than 5.0 until the sum of expectations was greater than 5.0 to avoid problems caused by cells with small frequencies (cf. Steel, Torrie, and Dickey, 1997).

Using the  $\text{AIC}_c$  favored model, I estimated the age structure of harvest samples. While estimates were available for all 13 age classes (0-11 and 12+), interest has typically focused on 6 age classes (0-4 and 5+) for management purposes. Thus, I applied the delta method (Seber 1982) to get estimates of the numbers of individ-



Figure 4.1: A selection of observed and fitted distributions for aging error as estimated with data from known-age females. Histogram bars represent the observed relative frequency of aging classifications by age, while the solid line represents the estimated underlying, continuous Laplacian model for aging error.



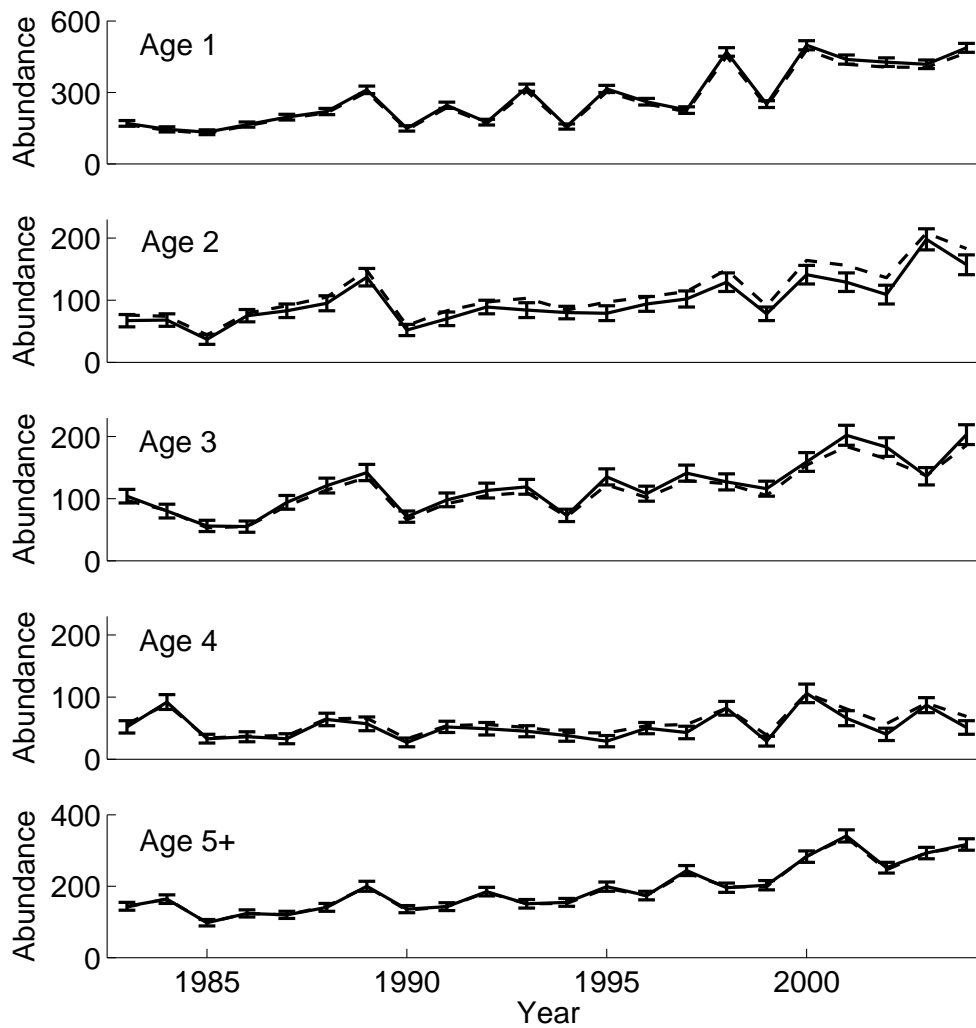
uals in age class 5+ and accompanying standard errors (Figure 4.2).

In general, the estimated age structure was similar to the observed data, but there were several differences. In particular, age classes 1, 3, and 5+ appeared to be under-represented in observed samples, whereas age classes 2 and 4 were over-represented. Further, the standard error for age classes 2-4 was often quite high (CV up to 33%), demonstrating considerable uncertainty about the age structure of the sample even under moderate levels of aging error, at least with the quantity of aging error data that were available. Uncertainty was even higher when all 13 age classes were estimated, with  $CV > 0.5$  for most older age classes.

## 4.4 Discussion

The process of assigning individuals into a given AS class is often subject to measurement error. As I have shown, the raw data (ignoring misclassification error) can include consistent biases in the number of individuals in each AS class, even when error rates are relatively low. In such situations, use of auxiliary data on error rates can be used to better estimate the AS structure of a sample and reflect uncertainty about this distribution. This is particularly relevant given the prevalence of age- and stage- based methods in ecology, where errors associated with AS distributions are often ignored. Methods for estimating population parameters  $\theta$  (especially population size) from AS data often are based on probability models which specify the likelihood of  $\theta$  given the data,  $P(\theta|\mathbf{C})$  (e.g., Gove et al., 2002; Skalski et al., 2005; Udevitz and Ballachey, 1998). A natural adjustment to acknowledge (and

Figure 4.2: Number of female black bears harvested in Pennsylvania from 1983-2004 classified by age category. Dotted lines represent raw harvest numbers (i.e., the  $C'_{ij}$ ), while solid lines represent estimated harvest numbers accounting for aging errors ( $\hat{C}_{ij}$ ). Also pictured are vertical bars representing 1 SE for estimated abundance. Note that the scale of the y-axis changes between panels to highlight differences between the observed and estimated age distribution. Not displayed here is the number of harvested cubs each year, which could be determined without error. Estimates of the number of age 5+ individuals are pooled from age classes 5-12+ of the original model, with standard error obtained via the delta method.



model) aging error is to reformulate the model as  $P(\boldsymbol{\theta}|\mathbf{C})P(\mathbf{C}|\mathbf{C}', \boldsymbol{\tau})$ .

In some cases, the performance of estimators for  $\boldsymbol{\theta}$  (as measured by mean squared error for example) could improve when uncertainty associated with AS determination is ignored. This would be the case when the additional variance imposed on an estimator outweighs improvements to bias (or, more specifically, bias squared). Researchers thus have reason to desire estimators of population parameters robust to errors in AS classification. Whatever estimation approach is adopted, it is important to quantify the effects of errors in AS determination on estimator performance. In this regard, Monte Carlo simulations may be used as a tool to explore the pros and cons of either ignoring or explicitly recognizing this type of uncertainty.

If the investigator chooses to use auxiliary data on aging error to help estimate the age structure of a sample, I recommend initially partitioning the data into a large number of age classes. Because the '+' group is an amalgamation of multiple age groups, model parameterizations that allow the mean or standard error to change as a function of age classes do not use data from the '+' group to inform these relationships. In my experience, precision on age structure estimates will tend to increase when using a large number of age classes because sample sizes are higher. If desired, older age classes may then be pooled back to a biologically meaningful '+' group, using the delta method to estimate an accompanying standard error.

My approach in this chapter was to use data from known age individuals to estimate misclassification rates. Although these data are preferable, logistical constraints may sometimes prevent the investigator from resampling known age in-

dividuals. Further, because known-age individuals are often followed from a young age, long studies may be needed to get data on error rates associated with older age classes. In these situations, it may be necessary to use the approaches of Richards et al. (1992) or Clark (2004), who proposed flexible methodologies for estimating misclassification probabilities ( $\tau_{ij}$ ) when there are multiple age readings for individual animals but none of known age. In this case, the investigator will be required to make more restrictive assumptions, such as the expected mode of raw age readings equalling the true age with the probability distribution for errors being symmetric about this mode. These assumptions may be unrealistic in some scenarios as the aging criterion may be biased towards younger or older ages depending on species and the true age of the individual (e.g., Beamish and McFarlane, 1983; Harshyne et al., 1998). In contrast, with sufficient known age data it is possible to entertain models in which measures of central tendency are permitted to change as a function of age class, as I have done here.

Demographic analyses using population models often depend critically on parameters such as survival, recruitment, and stage-transition probabilities. If AS data are used to estimate these parameters, persistent inferential biases may result when there are errors in AS determination and these errors go unnoticed or are ignored. Using auxiliary data on error rates, I have shown how to derive model-based estimates of AS distributions, together with a measure of precision. Under perfect detection of individuals, these estimates may be taken at face value and traditional life table approaches may be used to estimate population parameters. In other cases, more rigorous methods may be needed, such as those described at the beginning of

this section.

## Appendix A

# Bias of the Lincoln-Petersen Estimator Under Encounter Covariance

Seber(1982) described the 2-sample Lincoln-Petersen estimator of abundance. Ignoring a small sample size correction, the Lincoln-Petersen estimator is given as

$$\hat{N} = \frac{n_1 n_2}{m_2},$$

where  $\hat{N}$  denotes an estimator of population size,  $n_1$  denotes the number of individuals initially captured and marked during a first sample occasion,  $n_2$  denotes the

total number of individuals caught on a second sampling occasion, and  $m_2$  is the number of previously marked individuals caught in the second sample.

A first order Taylor series approximation for the expectation of the estimator may be written as

$$E(\hat{N}) \approx \frac{E(n_1)E(n_2)}{E(m_2)} = \frac{E(x_{11} + x_{10})E(x_{01} + x_{11})}{E(x_{11})},$$

where  $x_{11}$  gives the number of individuals caught in both samples,  $x_{10}$  gives the number caught in the first sample but not the second sample, and  $x_{01}$  gives the number caught in the second sample only. If  $p_{i1}$  denotes the probability individual  $i$  is caught in sample 1, and  $p_{i2}$  denotes the probability they are caught in sample 2, then the probability they are caught in both samples is  $p_{i1}p_{i2}$ . Similarly, the probability that they are caught in the first sample but not the second is  $p_{i1}(1 - p_{i2})$ , and the probability they are caught in the second sample but not the first is  $(1 - p_{i1})p_{i2}$ . Thus we equivalently have

$$E(\hat{N}) \approx \frac{\sum_{i=1}^N \{E[p_{i1}p_{i2}] + E[p_{i1}(1 - p_{i2})]\} \sum_{i=1}^N \{E[p_{i1}p_{i2}] + E[(1 - p_{i1})p_{i2}]\}}{\sum_{i=1}^N E[p_{i1}p_{i2}]}$$

Making the assumption that

$$\begin{bmatrix} p_{i1} \\ p_{i2} \end{bmatrix} \sim f \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right),$$



and using the well known property  $EXY = Cov(X, Y) + EXEY$ , we may write

$$E [p_{i1}p_{i2}] = \sigma_{12} + \mu_1\mu_2,$$

$$E [p_{i1} (1 - p_{i2})] = E [p_{i1}] - E [p_{i1}p_{i2}] = \mu_1 - \sigma_{12} - \mu_1\mu_2,$$

and

$$E [(1 - p_{i1}) p_{i2}] = \mu_2 - \sigma_{12} - \mu_1\mu_2.$$

We thus have

$$E (\hat{N}) \approx \frac{N\mu_1\mu_2}{\sigma_{12} + \mu_1\mu_2}$$

and

$$bias (\hat{N}) = E (\hat{N}) - N \approx N \left( \frac{\mu_1\mu_2}{\sigma_{12} + \mu_1\mu_2} - 1 \right).$$

This approximation works reasonably well when encounter probabilities are bounded away from 0 or 1, and confirms negative bias in the LP estimator whenever the covariance of encounter probabilities is positive. For a similar derivation in the context of mark-recapture distance sampling, see Borchers et al. (2006). For derivation of bias under individual heterogeneity, see Carothers (1973) and Gilbert (1973).

## Appendix B

# A User's Guide to Program

## AGEHARV

### B.1 Introduction

The incarnation of AGEHARV described here pertains to the example analysis in Chapter 1. Focus is thus on the case where a) the auxiliary dataset is one of mark-recovery, and b) the Brownie et al. (1985) parameterization is chosen for recoveries. Adaptations for other auxiliary datasets are relatively straightforward and typically only involve reprogramming functions for data input and log likelihood calculations. The current executable, as well as C++ source code are available from the author.

Program AGEHARV relies on a minimum of three user-created ASCII text files to provide data and model specification for analysis. The first is “inpfile.txt,” which gives AGEHARV information about the dimension of the problem, Markov Chain Monte Carlo (MCMC) options, and input values. The file “harvfile.txt” should contain the age-at-harvest matrix for the population to be analyzed. The file “histfile.txt” should include all mark-recovery encounter histories. Finally, the optional file “DMfile.txt” may be included if one wishes to use a design matrix to specify a user defined model. Below, I provide information on each of these components, together with implementation instructions and a description of program output.

## **B.2 User Defined Files**

### **B.2.1 The Input File**

The input file specifies information about problem dimensions, MCMC options, and initial values (Figure B.1). The first block of input (i.e., NYR to OVERD) are exchangeable as far as relative position. Where defaults are provided, specific entries may also be omitted. The second block of the input file (NAME to FPRIORS) must be included in the exact order specified, although WATCH, SPRIORS, HPRIORS, and FPRIORS may be omitted. The following is a more complete description of what these values mean:

*NYR*– Number of years of the study. This should be equal to the number of

```

NYR 3
NCOH 3
NITER 1000000
NHIST 1000
NOBREED 2
NWATCH 2
ADAPT 1
NADAPT 50000
EVERY 5
USERDM 0
SPRIORS 1
HPRIORS 1
FPRIORS 1
GOF 1
REPSGOF 100
OVERD 0

NAME StdHddFdd
INIT 10000 8000 20000
      10000 8000 6400 14000
      10000 8000 6400 14000
      .4 .4 -2 .4
INIT 3000 1500 2000
      3000 1500 1200 1400
      3000 1500 1200 1400
      .4 .4 -2 .4
SIGMA 100 70 60
       90 45 50 60
       95 45 50 60
       .06 .06 .06 .05
WATCH 1 12
SPRIORS N 0 1.75 N .4 1
HPRIORS N -2 .5
FPRIORS N .3 .5

```

Figure B.1: Example input file for program AGEHARV, which would be saved as “inpfile.txt.”

rows in the age-at-harvest matrix. The current version of AGEHARV was compiled to allow a maximum of 15 years of data.

*NCOH*– Number of cohorts. It is assumed that the last distinguishable cohort is a '+' category (e.g., age  $A$  and older). The current version of AGEHARV allows a maximum of 8 cohorts.

*NITER*– Number of Markov chain iterations per chain. Typically this will need to be quite large for inference to be reliable (see Chapter 2). However, the maximum number of values that can be stored is 200,000 so if a value greater than 200,000 is used, the user must “thin” the chain (see the subsequent entry for *EVERY*).

*NHIST*– Total number of mark-recovery histories contained in encounter history file. The maximum is currently set to 10,000.

*NOBREED*– The number of age classes which do not contribute to recruitment the following year. For instance, in Pennsylvania, cubs and yearling black bears typically do not become pregnant, so this value is set to 2.

*NWATCH*– The number of beta parameters (transformed scale) to monitor throughout MCMC simulation. Note that the default is zero, so this entry can be omitted from the input file if desired.

*ADAPT*– If set to 1, AGEHARV uses an algorithm described in Chapter 2 to adjust proposal standard deviations to achieve “optimal” acceptance rates. The default is 1.

*NADAPT*– If  $ADAPT = 1$ , this value specifies how many iterations the adapt algorithm should continue for. Since a window of 1000 is used before an adjustment

is made, this value should often be in the 30,000-60,000 range. The default is 30,000.

*EVERY*– This value allows one to “thin” Markov chains by only recording “every”  $x$  iterations. This is useful for long chains that would otherwise exceed the program image’s maximum dimensions. The default is one.

*USERDM*– If set to zero (the default), AGEHARV will use the prespecified model named later in the input file. If set to 1, AGEHARV will look for the file “DMfile.txt” which contains a user defined model structure (see section B.2.4).

*SPRIORS*, *HPRIORS*, and *FPRIORS*– If set to zero, AGEHARV uses default prior distributions for these fixed effect parameters. Default priors are Normal(0,3) on the logit scale for survival and recovery rates, and Normal(.25, 1) on the log scale for recruitment. If set to one, these options allow the user to specify their own prior distributions later in the input file. Note that this only makes sense for fixed effect models. The default is zero.

*GOF*– If set to one, AGEHARV employs a Bayesian goodness-of-fit test as described in Chapter 2. The default is zero.

*REPSGOF*– If  $GOF = 1$ , the number of iterations for which to calculate a test statistic. These are spread evenly across the Markov chain after the burnin period is passed. The default is 100.

*OVERD*– If  $OVERD = 1$ , overdispersion random effects are included on all recovery rates as described in Chapter 1. The default is 0.

*NAME*– This gives one of a number of models that AGEHARV can build on it’s own. This is always in the form of  $SxaHybFzc$ , where  $x$  denotes a submodel for survival,  $y$  denotes a submodel for recovery rate, and  $z$  denotes a submodel for

recruitment rate. One of the following codes should be substituted for each:

- d – time- and age-constant
- t – time-specific fixed effects
- T – time-specific random effects
- a – fixed effects for each age class, time constant
- + – time- and age-specific additive fixed effects
- P – age-specific fixed effects with additive time-specific random effects

Additionally, it is possible to specify a linear trend on each of these submodels by specifying a 'T' for *a*, *b*, or *c* (however don't try it with the '+' or 't' submodels because these aren't identifiable!). Even if  $USERDM = 1$ , a model still needs to be specified here for consistency even though it will be meaningless.

*INIT*– User defined initial values for the start of simulation. Two sets of values are needed since two Markov chains are run for every invocation of AGEHARV. It is paramount that the correct number of values be entered here, and that values for abundance are chosen so that they are internally consistent (e.g., the size of a cohort in one year is less than or equal to the size the previous year minus harvest). The order in which they are entered is also important. Here are some guidelines:

- Cohort specific abundance is entered first, with rows representing time and columns representing age class. There are *NCOH* age classes in the first year, and *NCOH* + 1 age classes in remaining years, for a total of  $NYR \times (NCOH + 1) - 1$  entries

- Survival parameters are entered after abundance. Fixed effects on the logit of survival are entered first, and random effects second. For instance, with submodel specification 'T' there would be one fixed effect parameter and  $Y - 1$  random effects terms (recall that for the Brownie parameterization there are  $Y - 1$  possible time periods for survival).
- Recovery rate parameters are entered after survival, with the order following the same pattern. However, in this case there are  $Y$  possible time periods. Also, if  $OVERD = 1$ , one must also include  $NYR \times NCOH$  overdispersion random effect terms at the end. As with survival, all effects are on the logit scale.
- Recruitment rate parameters are entered after recovery rate. As with survival there are a  $Y - 1$  possible time periods. Also, if an age effect is desired (this will frequently be inestimable), there will be  $A - NOBREED$  fixed effects.
- Initial values for the precision parameters of random effects models are entered last. If they occur at all, they occur in the following order: 1) precision for survival random effects, 2) precision for time-specific recovery rate random effects, 3) precision for overdispersion random effects on recovery rate, and 4) precision of random effects on recruitment rate.

*SIGMA* – With the exception of random effect precision terms, an initial proposal standard deviation must be specified for each parameter. They are specified in the same order as for initial values.



*WATCH* – If *NWATCH* > 0 in the preamble, this is where you would specify which beta parameters you want to follow throughout the estimation procedure. For instance, in Figure B.1, parameters 1 and 12 are followed. These correspond to cub abundance in year 1 and year 1 survival. Current limitations allow you to “watch” a maximum of 20 parameters per run. Note that the *WATCH* line should be omitted if you do not desire to watch any parameters. See section B.3.2 for information about output related to watched parameters.

*SPRIORS*, *HPRIORS*, and *FPRIORS* – If you allowed user defined priors in the preamble, this is where you would specify them (otherwise omit these lines). Choices for distributions of fixed effects on the transformed scale are limited to normal and gamma, where an ‘N’ specifies a normal distribution and a ‘G’ specifies a gamma distribution. For the normal distribution you must specify a mean and variance, and for the gamma you must specify a shape and scale parameter.

## **B.2.2 The Encounter History File**

The encounter history file must always be saved as “histfile.txt.” Each line should contain an encounter history, analogous to the format for live-dead encounter histories in Program MARK (White and Burnham, 1999), the age of the animal or group of animals in question, and a frequency. Unlike MARK, AGEHARV does not calculate frequencies before estimation, so tabulating frequencies beforehand can substantially improve on computing time. Figure B.2 gives an example encounter history file for the case of two age groups and a three year study.

```
100000 0 65
100001 0 3
110000 0 8
100100 0 8
100000 1 171
110000 1 18
100001 1 10
100100 1 17
001000 0 72
001001 0 4
001100 0 8
001001 1 17
001000 1 176
001100 1 23
000010 0 71
000011 0 13
000010 1 192
000011 1 24
```

Figure B.2: An example encounter history file, which would be saved as “histfile.txt.”

### **B.2.3 The Age-at-harvest File**

The age-at-harvest file contains a summary of age-specific harvest mortality by year.

Rows delineate years, while age is delineated by column. This file should be saved as “harvfile.txt.” An example for six age classes and 14 years of data is provided in

Figure B.3.

### **B.2.4 The Design Matrix File**

If one wishes to fit a particular model that is not one of the pre-specified models, one must include a design matrix (DM) file (“DMfile.txt”) in the same directory as the other input files. Figure B.4 shows an example DM file for a case where there are 3 years of data and 3 age classes. For survival and recovery rate there are fixed

```

144 165 76 56 34 124
164 189 89 85 34 128
156 216 105 111 63 146
269 303 146 130 63 202
144 145 53 67 34 125
176 235 82 89 53 147
191 175 94 102 55 171
157 306 103 98 49 148
148 157 82 67 42 148
200 301 93 111 41 184
166 249 101 97 51 175
247 227 110 124 49 221
234 454 143 118 77 191
168 251 87 101 35 176

```

Figure B.3: An example age-at-harvest file.

effects for each age class, but for survival there are also random effects for year (note that random effects are of limited utility when there are so few years so this is for illustrative purposes only).

The first line of the preamble tells AGEHARV that the design matrix is 18 rows by 10 columns. The number of rows should always be the number of real parameters for survival, recovery rate, and recruitment (abundance parameters are omitted because they are all treated as fixed effects in this version of AGEHARV), plus one extra row for each parameter type that requires a random effect. The total number of such real parameters is  $(Y - 1) * A + Y * A + (Y - 1) * (A - NOBREED)$ , where  $Y$  gives the number of years of data,  $A$  gives the number of age classes.

The second row of the preamble tells AGEHARV that the first column having to do with recovery rates ( $h$ ) is the sixth, and the first column having to do with recruitment rate ( $f$ ) is the ninth. By process of elimination, AGEHARV can deduce that columns 1-5 apply to survival ( $S$ ), columns 6-8 apply to  $h$ , and column 9

```

NROW 18 NCOL 10
firsth 6 firstf 9
RES 1 REH 0 REO 0 REF 0
1 0 0 1 0 0 0 0 0 0
0 1 0 1 0 0 0 0 0 0
0 0 1 1 0 0 0 0 0 0
1 0 0 0 1 0 0 0 0 0
0 1 0 0 1 0 0 0 0 0
0 0 1 0 1 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 1

```

Figure B.4: An example design matrix file.

applies to  $f$ , with the final column associated with the random effects for survival (see below).

The 3rd line of the preamble tells AGEHARV how many parameter types require random effects. For instance, in the above file 'RES 1' occurs in the 3rd line of the preamble, indicating that there are to be random effects for survival. By default, random effect terms always occur at the end of a given parameter's columns; one must always insert enough random effect columns for the parameter type in question. For instance, in a dataset that includes  $Y$  years of data, there are  $Y - 1$  years that survival is estimated for,  $Y$  years that recovery rate is estimated for, and  $Y - 1$  years that recruitment rate is estimated for. Additionally, if there are overdispersion random effects on recovery rate (the REO entry is set to one), there must be  $A \times Y$  random effects. Finally, whenever including random effects, one must add an additional row and column to the end of the design matrix in order to allocate an additional parameter corresponding to the precision of the random effects model. In the above file, the entry in the 18th row and 10th column corresponds to the precision of survival's random effects. Note that if  $OVERD = 1$  in the main input file (i.e., there are overdispersion random effects on recovery rate), then the entry for REO in the DMfile should also be 1.

Let's consider an example of the models specified by the preceding design matrix. First, isolating the components of the design matrix having to do with survival, we have

1 0 0 1 0

Table B.1: The survival submodel as formulated in Figure B.4 in terms of model components.

<u>Real</u> <u>Parameter</u>	<u>Fixed Effects</u>			<u>Random Effects</u>		<u>Model</u>
	$\beta_1$	$\beta_2$	$\beta_3$	$\epsilon_1$	$\epsilon_2$	
$S_{11}$	1	0	0	1	0	$\beta_1 + \epsilon_1$
$S_{12}$	0	1	0	1	0	$\beta_2 + \epsilon_1$
$S_{13+}$	0	0	1	1	0	$\beta_3 + \epsilon_1$
$S_{21}$	1	0	0	0	1	$\beta_1 + \epsilon_2$
$S_{22}$	0	1	0	0	1	$\beta_2 + \epsilon_2$
$S_{23+}$	0	0	1	0	1	$\beta_3 + \epsilon_2$

```

0 1 0 1 0
0 0 1 1 0
1 0 0 0 1
0 1 0 0 1
0 0 1 0 1

```

Here, there are 3 columns for fixed effects and two columns for random effects, with each row corresponding to a real parameter. Rows proceed by first advancing in age class, and then in time, so that we might consider the schematic presented in Table B.1.

## B.3 Implementation

### B.3.1 Command Line Instructions

After formatting user defined files, one invokes AGEHARV by opening a DOS window, navigating to the directory and folder that AGEHARV.exe is stored, typing AGEHARV *path name* at the prompt. Here, *path name* tells AGEHARV the directory in which user defined input files are stored. If no warning message occurs, every-

thing appears to be in good shape. If there are errors associated with input values or proposal standard deviations, it is likely that an incorrect number of entries have been provided in the master input file. If a message appears saying that initial values are internally inconsistent, then the initial abundance values and harvest data are not compatible. In either case, one must correct the problem or you will obtain nonsensical results!

### B.3.2 Output

There are two default pieces of output that one gets after successful completion of AGEHARV. The file “outfile.txt” gives model selection results (DIC for the mode and mean, respectively), a Bayesian p-value, estimates of beta parameters, real parameters, total abundance, and Gelman-Rubin statistics. Second, the ASCII text file “nfile.txt” provides marginal posterior samples of total abundance by year. Samples are in sequential order with breaks indicated by carriage returns. In practice, I find it useful to pull these results into a statistical programming language for visualization. For example, sample paths for the first 2 annual abundance values can be plotted in *R* via code similar to the following:

```
npar=14
niter=200000
file="c:/conn/bear/PennsylvaniaBBdata/Nfile.txt"
efile=scan(file,what="double",sep='\n')
datamat=matrix(0,nrow=npar,ncol=niter)

curline=1
for(ipar in 1:npar){
  datamat[ipar,]=as.double(efile[curline:(curline+niter-1)])
}
```

```
    curline=curline+niter
}

par(mfrow=c(2,2))
for(ipar in 1:2){
  plot(datamat[ipar,2:niter],type="l")
}
```

If the *WATCH* option has been specified in the master input file, posterior samples for these parameters will be output in the same manner to the file “mat-file.txt.”



## BIBLIOGRAPHY

- Alt, G. L., 1989. Reproductive biology of female black bears and early growth and development of cubs in northeastern Pennsylvania. Ph.D. thesis, West Virginia University, Morgantown, WV, USA.
- Anderson, A. E., D. C. Bowden, and D. M. Kattner, 1992. The puma on Uncompahgre Plateau, Colorado. Technical Report DOW-R-T-40-92.
- Barry, S. C., S. P. Brooks, E. A. Catchpole, and B. J. T. Morgan. 2003. The analysis of ring-recovery data using random effects. *Biometrics* **59**:54–65.
- Beamish, R. J., and G. A. McFarlane. 1983. The forgotten requirement for age validation in fisheries biology. *Transactions of the American Fisheries Society* **112**:735–743.
- Beck, T. D. I., 1991. Black bears in west-central Colorado. Technical Report Tech. Publ. 39 DOW-R-T-39-91, Colorado Division of Wildlife, Fort Collins, Colorado.
- Bernardo, J. M., and M. A. Juárez, 2003. Intrinsic estimation. Pages 465–476 *in* J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman,

- A. F. M. Smith, and M. West, editors. Bayesian Statistics. Oxford University Press, Oxford.
- Borchers, D. L., J. L. Laake, C. Southwell, and C. G. M. Paxton. 2006. Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics* **62**:372–378.
- Box, G. E. P., and N. R. Draper. 1987. Empirical Model-Building and Response Surfaces. John Wiley & Sons, New York, NY, USA.
- Brooks, S. P., 1999. Bayesian analysis of animal abundance data via MCMC. Pages 723–731 *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors. Bayesian Statistics 6. Oxford University Press.
- Brooks, S. P., E. A. Catchpole, and B. J. T. Morgan. 2000. Bayesian animal survival estimation. *Statistical Science* **15**:357–376.
- Brownie, C., D. R. Anderson, K. P. Burnham, and D. S. Robson, 1985. Statistical Inference From Band Recovery Data: A Handbook. US Fish and Wildlife Resource Publication (156), Washington, DC.
- Burnham, K. P., 1993. A theory for combined analysis of ring recovery and recapture data. Pages 199–213 *in* J. D. Lebreton and P. M. North, editors. Marked Individuals in the Study of Bird Populations. Birkhauser, Switzerland.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd Edition. Springer-Verlag, New York.

- Carlin, B., and T. Louis. 2000. Bayes and Empirical Bayes Methods for Data Analysis, 2nd Edition. Chapman & Hall / CRC Press, Boca Raton, FL.
- Carothers, A. D. 1973. Effects of unequal catchability on Jolly-Seber estimates. *Biometrics* **29**:79–100.
- Casella, G., and R. L. Berger. 1990. *Statistical Inference*. Duxbury Press, Belmont, CA.
- Caswell, H. 2001. *Matrix Population Models*, 2nd Edition. Sinauer, Sunderland, MA.
- Catchpole, E. A., and B. J. T. Morgan. 1997. Detecting parameter redundancy. *Biometrics* **84**:187–196.
- Caughley, G. 1977. *Analysis of vertebrate populations*. Wiley, London.
- Chen, M. H., Q. M. Shao, and J. G. Ibrahim. 2000. *Monte Carlo Methods in Bayesian Computation*. Springer, New York.
- Clark, W. G. 2004. Nonparametric estimates of age misclassification from paired readings. *Canadian Journal of Fisheries and Aquatic Sciences* **61**:1881–1889.
- Conn, P. B., P. F. Doherty Jr., and J. D. Nichols. 2005. Comparative Demography of New World Populations of Thrushes (*Turdus* Spp.): Comment. *Ecology* **86**:2536–2541.
- Cooch, E. G., and G. C. White. 2005. *Program MARK: A Gentle Introduction*. <http://www.phidot.org/software/mark/docs/book/>.

- Costello, C. M., K. H. Inman, D. E. Jones, R. M. Inman, B. C. Thompson, and H. B. Quigley. 2004. Reliability of the cementum annuli technique for estimating age of black bears in New Mexico. *Wildlife Society Bulletin* **32**:169–176.
- Cox, D. R., and D. Oakes. 1984. *Analysis of Survival Data*. Chapman and Hall, London.
- Coy, P. L., and D. L. Garshelis. 1992. Reconstructing reproductive histories of black bears from incremental layering in dental cementum. *Canadian Journal of Zoology* **70**:2150–2160.
- Diefenbach, D. R., and G. L. Alt. 1998. Modeling and evaluation of ear tag loss in black bears. *Journal of Wildlife Management* **62**:1292–1300.
- Diefenbach, D. R., J. L. Laake, and G. L. Alt. 2004. Spatio-temporal and demographic variation in the harvest of black bears: Implications for population estimation. *Journal of Wildlife Management* **68**:947–959.
- Dupont, W. D. 1983. A stochastic catch-effort method for estimating animal abundance. *Biometrics* **39**:1021–1033.
- Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. *Canadian Journal of Fisheries and Aquatic Sciences* **39**:1195–1207.
- Fox, J. 2002. *An R and S-Plus Companion to Applied Regression*. Sage Publications, Inc., Thousand Oaks, California.

- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian Data Analysis, 2nd Edition. Chapman and Hall, Boca Ration.
- Gilbert, R. O. 1973. Approximations of bias in Jolly-Seber capture recapture model. *Biometrics* **29**:501–526.
- Gill, R. B., and T. D. I. Beck, 1990. Black bear management plan 1990-1995. Technical Report Div. Rep. No. 15.
- Giminez, O., R. Choquet, and J. D. Lebreton. 2004. Parameter redundancy in multistate capture-recapture models. *Biometric Journal* **45**:704–722.
- Gotelli, N. J. 2001. A Primer of Ecology, Third Edition. Sinauer Associates, Inc., Sunderland, MA.
- Gove, N. E., 1997. Using age-harvest data to estimate demographic parameters for wildlife populations. Ph.D. thesis, University of Washington, Seattle, WA, USA.
- Gove, N. E., J. R. Skalski, P. Zager, and R. L. Townsend. 2002. Statistical models for population reconstruction using age-at-harvest data. *Journal of Wildlife Management* **66**:310–320.
- Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**:711–732.
- Green, R. E. 2004. A new method for estimating the adult survival rate of the Corncrake *Crex crex* and comparison with estimates from ring-recovery and ring-recapture data. *Ibis* **146**:501–508.

- Green, R. E., N. Shaffer, and D. Wend. 2001. A method for ageing adult Corncrakes *Crex crex*. Ringing and Migration **20**:352–357.
- Gustafson, P. 2005. On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. Statistical Science **20**:111–119.
- Harshyne, W. A., D. R. Diefenbach, G. L. Alt, and G. M. Matson. 1998. Analysis of error from cementum-annuli age estimates of known-age Pennsylvania black bears. Journal of Wildlife Management **62**:1281–1291.
- Heisey, D. M., D. O. Joly, and F. Messier. 2006. The fitting of general force-of-infection models to wildlife disease prevalence data. Ecology **87**:2356–2365.
- Hewison, A. J. M., J. P. Vincent, J. M. Angibault, D. Delorme, G. Van Laere, and J. M. Gaillard. 1999. Tests of estimation of age from tooth wear on roe deer of known age: variation within and among populations. Canadian Journal of Zoology **77**:58–67.
- Jeffreys, H. 1961. Theory of Probability, 3rd Edition. Oxford University Press, Oxford, England.
- Keay, J. A. 1995. Accuracy of cementum age assignments for black bears. California Fish and Game **81**:113–121.
- Kendall, W. L., and J. D. Nichols. 2002. Estimating state-transition probabilities for unobservable states using capture-recapture/resighting data. Ecology **83**:3276–3284.

- Kordek, W. S., and J. S. Lindzey. 1980. Preliminary analysis of female reproductive tracts from Pennsylvania black bears. *International Conference on Bear Research and Management* **4**:159–161.
- Laake, J. L., 1992. Catch-effort models and their application to elk in Colorado. Ph.D. thesis, Colorado State University, Fort Collins, CO, USA.
- Lebreton, J. D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* **62**:67–118.
- Link, W. A., J. A. Royle, and J. S. Hatfield. 2003. Demographic analysis from summaries of an age-structured population. *Biometrics* **59**:778–785.
- Madigan, D., and J. C. York. 1997. Bayesian methods for estimation of the size of a closed population. *Biometrika* **84**:19–31.
- McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. Chapman and Hall, New York.
- Megrey, B. A. 1989. Review and comparison of age-structured stock assessment models from theoretical and applied points of view. *American Fisheries Science Symposium* **6**:8–48.
- Meyer, R., and R. B. Millar. 1999. Bayesian stock assessment using a state-space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences* **56**:37–52.

- Newman, K. B., S. T. Buckland, S. T. Lindley, L. Thomas, and C. Fernández. 2006. Hidden process models for animal population dynamics. *Ecological Applications* **16**:74–86.
- Pella, J. J., and T. L. Robertson. 1979. Assessment of composition of stock mixtures. *U.S. Fisheries Bulletin* **77**:387–398.
- Pollock, K. H. 2002. The use of auxiliary variables in capture-recapture modeling: an overview. *Journal of Applied Statistics* **29**:85–102.
- Pradel, R. 1996. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics* **52**:703–709.
- Reeves, S. A. 2003. A simulation study of the implications of age-reading errors for stock assessment and management advice. *ICES Journal of Marine Science* **60**:314–328.
- Ricklefs, R. 1997. Comparative demography of new world populations of thrushes (*Turdus* spp.). *Ecological Monographs* **67**:23–43.
- Rosenberry, C. S., D. R. Diefenbach, and B. D. Wallingford. 2004. Reporting-rate variability and precision of white-tailed deer harvest estimates in Pennsylvania. *Journal of Wildlife Management* **68**:860–869.
- Royle, J. A., and W. A. Link. 2002. Random effects and shrinkage estimation in capture-recapture models. *Journal of Applied Statistics* **29**:329–351.
- SAS Institute Inc. 2004. SAS/IML 9.1 User's Guide. SAS Institute Inc., Cary, NC.



- Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics* **6**:461–464.
- Seber, G. A. F. 1982. *The Estimation of Animal Abundance and Related Parameters*, 2nd Edition. Blackburn Press, Caldwell, NJ, USA.
- Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.
- Skalski, J. R., K. E. Ryding, and J. J. Millspaugh. 2005. *Wildlife Demography: Analysis of sex, age, and count data*. Elsevier Academic Press, San Diego, CA, USA.
- Smith, P. J. 1991. Bayesian analyses for a multiple capture-recapture model. *Biometrika* **78**:399–407.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B* **64**:583–639.
- Steel, R. G. D., J. H. Torrie, and D. A. Dickey. 1997. *Principles and Procedures of Statistics: A Biometrical Approach*, Third Edition. McGraw Hill, Boston, MA, USA.
- Udevitz, M. S., and B. E. Ballachey. 1998. Estimating survival rates with age-structure data. *Journal of Wildlife Management* **62**:779–792.

White, G. C., and K. P. Burnham. 1999. Program MARK: Survival estimation from populations of marked animals. *Bird Study* **46 Supplement**:120–138.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. *Analysis and Management of Animal Populations*. Academic Press, San Diego, CA, USA.