ABSTRACT

GOSKY, ROSS MATTHEW. Bayesian Analysis and Matching Errors in Closed Population Capture Recapture Models. (Under the direction of Leonard Stefanski and Sujit Ghosh.)

Capture-Recapture models are used to estimate the unknown sizes of animal populations. When the population is closed, with constant size, during the study, eight standard models exist for estimating population size. These models allow for variation in animal capture probabilities due to time effects, heterogeneity among animals, and behavioral effects after the first capture. Our research focuses on two areas:

1. Using Bayesian statistical modeling, we present versions of these eight models. We explore the use of Akaike's Information Criterion (AIC), and the Deviance Information Criterion (DIC) as tools for selecting the appropriate model for a given dataset. Through simulation, we show that AIC performs well in model selection.

2. A new, non-invasive method of capturing animals is to substitute captures of DNA profiles, through sources such as hair samples, for live animal captures. However, DNA profiles of close relatives may not be distinguishable from each other, and some animals in the population may not be uniquely identifiable. This problem leads to negative bias in estimating population size. We present a statistical model which accounts for this type of matching error, leading to more accurate estimation of population size.

Bayesian Analysis and Matching Errors in Closed Population Capture Recapture Models

by

Ross M. Gosky

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

STATISTICS

in the

GRADUATE SCHOOL
at
NC STATE UNIVERSITY
2004

Leonard Stefanski, Co-Chair

Sujit Ghosh, Co-Chair

Kenneth Pollock

Cavell Brownie

Bayesian Analysis and Matching Errors in Closed Population Capture Recapture Models

Copyright 2004

by

Ross M. Gosky

This is dedicated to Becky and Ryan.

Biography

Ross Matthew Gosky was born in Columbus, Ohio on September 12, 1971 to Garry and Patricia Gosky. Ross soon moved to Cleveland, Ohio and graduated from Saint Ignatius High School in Cleveland, Ohio in 1989. He entered Miami University in 1989, and graduated in 1993 with a B.S. degree in Mathematics and Statistics. On May 13, 1995 he married Becky Shippy, whom he had met while at Miami University. From 1993 to 1999 Ross worked as a consultant for Andersen Consulting. In 1999 Ross entered graduate school at North Carolina State University in Raleigh, earning a Master's degree in Statistics in May, 2001. In April, 2001 Ryan Matthew Gosky was born to Ross and Becky. Ross will continue his career at Bucknell University in Lewisburg, Pennsylvania starting in the fall of 2004.

Acknowledgements

I would like to thank my family for their love and support during the past five years. The last five years at NC State have been wonderful for me. Becky and Ryan have supported me so much during this time and have helped me immeasurably. Becky has been my best friend through this, and Ryan never fails to bring a smile to my face and helps me to keep everything in proper perspective.

My advisors, Len Stefanski and Sujit Ghosh, have been of immense help throughout the development of this thesis. They are excellent statisticians and people, and I am fortunate to have learned from them.

I would like to thank Harsha Rao and Terry Byron for their help with any and all computing questions (there were many) and for being good friends as well.

Many graduate students have also helped me very much with their friendship and advice. Although I am certain to miss many names, some of these people are Kevin Daly, Steve Hoffner, Jimmy Doi, Jared Lunceford, Jeff Burnett, Doug Robinson, Julie McIntyre, Clay Barker, Jing Wang, Hyeyoung Lee, Michael Crotty, Alvin Van Orden, Tae-Young Heo, Palani Ravindran, Shaw Yoshitani, Karen Chiswell, Michael Crotty, Joe Boyer, Prasheen Agarwal, Katja Remlinger, and Sheng Feng.

My committee members, Ken Pollock and Cavell Brownie, have been very helpful with this research and thesis. Ken was very instrumental in helping frame the topic, and his time and efforts are greatly appreciated.

Additionally, I express thanks to Sastry Pantula and Bill Swallow for their help and mentoring through my graduate career.

Contents

List of Figures

List of Tables

Chapter 1

Closed Population Capture

Recapture Models

Closed Population Capture-Recapture models (also often called Mark-Recapture Models) are used to estimate the size of a population in cases when enumeration or census is not a viable option (see Seber (2002) for an introduction). These methods typically apply to animal populations, but the methodology can also be applied to lists of people, as a means to estimate the (unknown) size of the population. The basic methodology name capture-recapture describes the sampling method used to obtain estimates of population size. The term closed refers to the assumption that the population size is constant during the study, and that there are no births, deaths, or migrations during the study. The sampling methodology is easiest to describe for animal populations. Several samples are taken from the population, and each time the captured animals are marked with an identifiable tag. After each sample is taken, the captured animals are re-released into the population. After

the first sample is taken, subsequent samples return a combination of marked and unmarked animals. When several samples are taken, different tags (or the same tag with a unique identifier) are used for each sample so that a distinct capture history is available for all the captured animals in the population. The goal is to use the percentages of recaptured animals as a basis for estimating the number of animals that were never captured during the study. This number of animals which are non-captured (which is estimated) plus the number that were captured during at least one of the captures forms an estimate $\hat{N}$ of the population size N. When this methodology is applied to lists of people, each separate list is considered a separate capture and the information from the separate lists is used to estimate the population size N.

Denoting the number of capture periods for the study as k, we can examine the simple case where there are $k = 2$ captures from the population. Denoting

$$n_1 = \text{number of captured animals in capture period 1}$$

$$n_2 = \text{number of captured animals in capture period 2}$$

$$m_2 = \text{number of recaptured animals in capture period 2}$$

the most common estimator for N is

$$\hat{N} = \frac{n_1 n_2}{m_2}. \tag{1.1}$$

This estimator is called the Petersen estimate, or the Lincoln-Petersen (LP) estimator. The development of the method is credited to Petersen (1896), but a more recent treatment of the probability model is given by Bailey (1951). The LP estimate is based upon hypergeo-

metric or multinomial likelihood functions which condition on n_1, and is a commonly-used estimator of population size.

A more general set of models with possibly $k > 2$ captures using the multinomial distribution are also available. A good introduction to these models is presented by Otis, Burnham, White, and Anderson (1978). We introduce the details of their general framework here, which is used throughout the rest of the thesis. Define indicator variables X_{ij} as

$$X_{ij} = 1 \text{ if animal } i \text{ is captured during capture period } j$$

for $i = 1, 2, ..., N$ and $j = 1, 2, ..., k$.

Also denote

$$p_{ij} = \Pr(X_{ij} = 1)$$

as the probability that animal i is captured during capture period j. Denote the capture matrix $\boldsymbol{X}$ with dimensions $N \times k$ with entry $X[i, j] = X_{ij}$ in row i and column j. Denoting $X[i, .]$ as the ith row of $\boldsymbol{X}$, we note that this vector has 2^k possible values, as each entry in the vector must be zero or one. For notational simplicity, these outcomes can be ordered as

Outcome 0: capture history $(0, 0, 0, ..., 0, 0, 0)$;

Outcome 1: capture history $(0, 0, 0, ..., 0, 0, 1)$;

Outcome 2: capture history $(0, 0, 0, ..., 0, 1, 0)$;

Outcome 3: capture history $(0, 0, 0, ..., 0, 1, 1)$;

Outcome 4: capture history $(0, 0, 0, ..., 1, 0, 0)$;

Outcome 5: capture history $(0, 0, 0, ..., 1, 0, 1)$;

through

Outcome $2^k - 1$: capture history $(1, 1, 1, ..., 1, 1, 1)$.

Each animal in the population has exactly one of the 2^k capture histories. A mathematical representation of the above outcomes is obtained through the notation $(X_{i1}, X_{i2}, ..., X_{ik})$ of k values to represent the observed capture history of any animal in the population. For $l = 0, 1, 2, ..., 2^k - 1$, we define Capture History L corresponding to the previous ordering of outcomes, where $L = \sum_{j=1}^{k} X_{ij} 2^{k-j}$. When necessary to refer to the capture history of animal i, we denote the capture history of animal i as capture history L_i, where $L_i = \sum_{j=1}^{k} X_{ij} 2^{k-j}$. Denote the matrix $\boldsymbol{Y}$ with dimensions $N \times 2^k$, with entry $Y[i, l] = 1$ if animal i has capture history outcome $L_i - 1$ from the list of outcomes above. That is, $Y[i, 1] = 1$ if animal i has capture history $(0, 0, ..., 0)$, $Y[i, 2] = 1$ if animal i has capture history $(0, 0, 0, ..., 0, 1)$, etc. Exactly one entry in the vector $Y[i, .]$ is equal to 1, the rest of the entries in the vector are zero. Denote $Z_L = \sum_{i=1}^{N} Y[i, l+1]$ as the number of animals with capture history L, for $L = 0, 1, ..., 2^k - 1$. Notice that Z_0 cannot be observed, as it represents the number of animals with capture history $(0, 0, ..., 0)$. Notice also that $\sum_{l=0}^{2^k-1} Z_l = N$. Denote $S = \sum_{l=1}^{2^k-1} Z_l = N - Z_0$ as the number of animals observed during at least one capture period. Note also that $Z_l = \sum_{i=1}^{N} I(l_i = l)$.

We can compute the probability of animal i having capture history l_i (denoted as P_{l_i}) as

$$P_{l_i} = \prod_{j=1}^{k} p_{ij}^{x_{ij}} (1 - p_{ij})^{1 - x_{ij}} \tag{1.2}$$

for capture history vector l_i. If P_{l_i} is equal for all i animals in the population, then defining

$L = 2^k - 1$ and $P_l = \prod_{i=1}^N P_{l_i}^{I(l_i = l)}$ then the likelihood function can be written as

$$L(N, P | Z_1 = z_1, ..., Z_L = z_L) = \frac{N!}{(N - S)! \prod_{l=1}^L Z_l!} \prod_{l=1}^L P_l^{Z_l} \left(1 - \sum_{l=1}^L P_l\right)^{N - S}. \qquad (1.3)$$

Notice that if N were known, this model would represent a multinomial likelihood function with counts $Z_0, ..., Z_{2^k - 1}$ and probabilities $P_0, ..., P_{2^k - 1}$. A set of assumptions underlying this modeling structure are (Otis et al. (1978)):

1. The population is closed (no births, deaths, or migrations during the study);

2. Animals retain their tags throughout the experiment;

3. All tags/marks are correctly noted after each capture.

Many contributions to the literature on Capture-Recapture models have dealt with violations of the above assumptions. Models by Jolly (1965) and Seber (1965) dealt with violation of the closure assumption with the addition of several more parameters to the models. Throughout this work, we assume that Assumption 1 is valid, and that the population is closed. Tag loss by animals during the study is mentioned in Seber (1973). Other treatments of tag loss have come from Kremers (1998), and Nelson, Anderson and Burnham (1980). Tag loss can be considered a type of matching error in capture-recapture studies.

1.0.1 Background for Our Approaches to Bayesian Modeling and Matching Errors

When the three listed assumptions are valid, we use Bayesian inference and model selection procedures. In Otis et al. (1978), a set of eight closed population models are proposed. We present Bayesian versions of these models and use model selection criteria to select the best model for a given dataset. Each model makes different assumptions about the complexity of the capture probabilities p_{ij}. Brief summaries of the model types follow.

Model M_0: p_{ij} is constant for all i, j.

Capture probabilities are constant for all individuals and all capture periods.

Model M_t: The subscript t denotes time effects in the model.

p_{ij} varies by capture period (that is, by j), but not by individuals.

Model M_h: The subscript h denotes heterogeneity in the population.

p_{ij} varies by individual but not by capture period. This is often a realistic assumption, as some animals may be easier to capture than others (for example, the probability of capturing an animal may be different for animals of different ages).

Model M_b: The subscript b denotes a behavioral effect in the population.

A trap-effect is assumed after the first capture, which changes the capture probability. An increase in the base capture probability is referred to as making an animal trap-happy. A decrease in the base capture probability is generally for situations when animals become trap-shy after first capture. Subsequent captures are assumed not to cause such an effect.

Model M_{th}: Use of two or more subscripts indicates the presence of both effects. Here, p_{ij} varies both by individual (by i) and by capture period (by j).

Model M_{tb}: As the subscript b indicates, a trap effect is assumed after the initial capture, but we also assume that different capture periods may have different capture probabilities (that is, vary by j).

Model M_{bh}: Capture probabilities vary by animal (by i). And, a trap effect is assumed after first capture.

Model M_{tbh}: Combines all three types of effects. Capture probabilities vary by capture period, by individual, and a trap effect is assumed.

In Model M_0, one assumes that p_{ij} are constant across all time periods and all animals. That is, $p_{ij} = p$, where p is constant. To incorporate time effects for Model M_t, one can assume $p_{ij} = p_j$. Similarly, behavioral effects for Model M_b can be incorporated by using of $p_{ij} = p_B$ if the animal has not been captured before capture period j, and $p_{ij} = p_A$ if the animal has been captured before capture period j.

Models with heterogeneity present a challenge to the form of the likelihood we have defined previously. Under heterogeneity models, p_{ij} may vary with i. This could cause P_{l_i} to vary by i as well. Our likelihood function assumes that P_{l_i} do not vary with i. If a finite mixture model is implemented to account for heterogeneity, the above likelihood function may be used. We describe this approach further in Chapter 2. Another approach to heterogeneity defined by Burnham and Overton, 1978, uses the Jackknife method to obtain an estimator for N in the presence of heterogeneity within the population. It should

be noted that in Link (2003), inference about N in heterogeneity models is shown to be highly model-dependent, suggesting that careful consideration in required when accounting for heterogeneity in capture-recapture models.

Program CAPTURE (described in Otis et al. (1978)), provides estimators of N for some the above models. Additionally, it provides model selection functionality based on Chi-Square goodness of fit tests. The models and estimators of N used in program CAPTURE are as follows:

M_0: basic M_0 model estimator

M_h: uses jackknife estimator (Burnham and Overton, 1978)

M_b: uses Zippin removal estimator (Zippin, 1956)

M_t: uses Darroch estimator (Darroch, 1958)

M_{bh}: uses a generalized removal estimator (see Seber, 1982)

M_{tb}: uses an estimator from Burnham (unpublished)

M_{th}: uses Chao estimator (Chao, et al, 1992)

M_{tbh}: not available in Program CAPTURE

However, it is noted (Pollock, Nichols, Brownie, and Hines (1990)) that the goodness-of-fit tests in Program CAPTURE have low power when the number of captures is relatively small. Pledger (2000) discussed the use of mixture models to fit heterogeneity models to capture-recapture data, and discussed use of Akaike's Information Criterion (AIC) as a model selection tool.

Another form of model selection occurs when experimenters have basic beliefs about the population that may logically eliminate one or more of the models from consideration. However, the removal of potentially appropriate models based upon assumptions about the population would be a danger to this approach. For these reasons, further work on model selection criteria is needed.

The work we present in this thesis is separated into two different topics:

Bayesian Models and Model Selection for Closed Population Models .

We present Bayesian versions of the above models, utilizing a mixture approach to modeling heterogeneity, and allowing for fixed time and behavioral effects in the appropriate models. Furthermore, through extensive simulations, we explore the use of Akaike's Information Criterion (AIC) and also the Deviance Information Criterion (DIC) as potential model selection criteria.

Effects of Matching Errors in Closed Population Models .

Matching errors for closed population models is considered a violation of the assumption that all tags or marks are correctly noted after each capture. Such matching errors can occur when these models are used with lists of people, where varying amounts of information (and possible mistakes) are available on each list. Within animal populations, use of DNA fingerprints as a replacement for live animal captures also gives rise to potential matching errors. Two types of matching errors are possible. Firstly, two individuals who are truly unique may not be recognized as such, and may be consolidated into one capture record. Secondly, a match from a particular sample (or list) may be corrupted so that the true match is not detected, and the corrupted in-

formation is falsely listed as a new, previously undetected, member of the population. We focus on a hierarchical model to deal with the first of these two errors, specifically to deal with the DNA sampling case, where animals may not be uniquely identifiable through their genetic fingerprints.

Chapter 2

Bayesian Estimation and Model Selection for Closed Population Models

2.1   Introduction and Motivation for Model Selection Procedures

Analysis of Capture-Recapture data for closed populations typically uses one of the eight standard models mentioned in Chapter 1. Some definitive methods surrounding selection of the correct model is necessary in such analyses. The eight models are generally (though not exclusively) nested, ranging from very simple models (M_0) where capture probabilities are equal for all animals across all capture periods, to complex models (M_{tbh}) allowing variation of capture probabilities by animal, capture period, and for behavioral

effects. In such situations, model selection criteria are critical to allow the best model of the eight to be fit to the data. Following the researcher's suspicions about which model to use may lead to choice of a model which is too simple to fit the data. Similarly, although it is tempting to choose the most complex model to analyze the data, this may lead to over-fitting and hence a higher predictive variance. The desire to strike a balance between choice of a model that neither under-fits nor over-fits the data is the motivation for model selection criteria (Burnham, 2002, chap.1).

Akaike's Information Criterion (AIC) is one such method of model selection, and seems to be the most commonly used measure for model selection. The intent of AIC is to measure the mathematical distance between the true population and the fitted model, by using the so-called Kullback-Leibler discrepancy. To differentiate between models with different number of parameters, AIC adds two times the number of model parameters to the estimated Kullback-Leibler discrepancy. In a case where there are two models, one simple and the other complex, that fit a data set equally well, this allows AIC to choose the simpler model by penalizing the complex models. Generally the rule of parsimony says that a researcher should choose the simplest model that describes the behavior of the population. Use of AIC generally supports this rule.

As our models are nested, we examine AIC as a tool for model selection. However it might be noticed that AIC is not an asymptotically consistent criterion in the sense that the probability that it chooses the correct model (given that the data has been generated from the correct model) does not converge to one as the sample size tends to infinity (Schwartz (1978)). We use a modified version of AIC within our Bayesian framework for parameter

estimation.

The Deviance Information Criterion (DIC) is a Bayesian model selection crite-
rion (Speigelhalter, Carlin, et al 2002). It is structured similarly to AIC, but is strictly a
Bayesian criterion. The main difference between AIC and DIC is in the penalty term added
to the estimated Kullback-Leibler discrepancy. DIC adds two times the effective number of
parameters to the estimated Kullback-Leibler discrepancy. The effective number of param-
eters is a Bayesian concept. It deals with the fact that the number of model parameters
in a Bayesian model is influenced by their prior distributions. The DIC criterion presents
a methodology to measure this number of parameters. The DIC is, then, the difference
between the estimated mean KL distance, and the KL distance estimated at the posterior
mean of each of the model parameters.

We also examine DIC as a model selection criterion for these models. Use of DIC
does not require the models to be nested. However, as we will see, our modeling uses a
mixture approach for some of the models. There are some questions about use of DIC for
mixture models.

We considered using the Bayesian Information Criterion (BIC) for model selection,
but, for capture-recapture models it is not clear what the sample size should be (as N itself
is a parameter and k, the number of capture periods, is usually much smaller than necessary
for asymptotic properties to work). Therefore, we chose to focus on AIC and DIC as the
two model selection criteria to be examined.

Specifically, our goals are to answer the following questions:

(i) Does AIC, choose the correct underlying model for a given data set?

(ii) Similarly, how does DIC perform as a model selection criterion for these models?

(iii) Are one or more of the eight closed population models particularly robust when fitting data generated under a different model structure?

To assess this, we develop and present Bayesian versions of the eight standard closed population capture-recapture models. We present simulation results to determine whether AIC and DIC work well in differentiating between the models.

2.2 Bayesian Closed Population Capture Recapture Models

We use the general framework outlined in Chapter 1. For each model, the following notation is common. N represents the (unknown) population size. We use p_{ij}, P_l, Z_l, and S as previously defined in Chapter 1.

The set of sufficient statistics is given by $\{Z_1, Z_2, \ldots, Z_L\}$ where $L = 2^k - 1$. Estimating N is equivalent to estimating Z_0, the number of individuals in the population which were undetected during the study. Using P_l as defined in Chapter 1, let $P = (P_1, P_2, \ldots, P_L)$. We can thus obtain the likelihood function as $L(N, P)$

$$L(N,P) \propto \binom{N}{S} \prod_{l=1}^{L} P_l^{z_l} \left(1 - \sum_{l=1}^{L} P_l\right)^{N-S}. \tag{2.1}$$

It is of interest to estimate N, treating P as a nuisance parameter. The capture probability vector P varies depending on the specific model. The main difference between the models is the modeling of capture probabilities. We take a Bayesian perspective to derive parameter

estimates. For each of the eight models we use the following prior distribution for N,

$$Pr\left(N=n\right) \propto \frac{1}{n^{\delta}}, n = 1, 2, ..., N_{max}$$

where $\delta > 0$ is fixed at a specific value and N_{max} is set at a realistic upper bound for N. The Jeffreys prior can be approximated with $\delta = \frac{1}{2}$, which we use for our simulations (results given in Section 2.3). Now we describe the basic modeling of these capture probabilities for each of the eight models and associated prior distributions for model parameters.

Model M_0: $p_{ij} = \alpha$ for all i, j.

We use a Beta distribution for α, specifically $\alpha \sim Beta(a, b)$. For our simulations (discussed in Section 2.3), we fixed $a = b = 0.5$.

Model M_t: $p_{ij} = \alpha_j$ for all i.

We use a sequence of independent and identically distributed (i.i.d.) Beta random variables for α_j's, specifically, $\alpha_j \overset{i.i.d.}{\sim} Beta(a, b)$. For our simulations (Section 2.3), we fixed $a = b = 0.5$.

Model M_h: $p_{ij} = p_i$ for all j, and $p_i \sim F(\cdot)$.

To account for heterogeneity, we assume a finite mixture distribution for univariate distribution function, F, specifically the probability mass function (p.m.f.) is given by $dF(p) = \sum_{m=1}^{r} \pi_m I(p = \theta_m)$, where π_m denotes the probability at support point θ_m with the restriction that $\sum_{m=1}^{r} \pi_m = 1$. It has been shown that when using mixture models to account for heterogeneity, the MLE of the number of support points is less than four for most data sets (Norris and Pollock, 1996), (Pledger, 2000). For

our simulations (Section 2.3), we fix $r = 2$ mass points representing possibly two distinct groups within the population. One group represents a relatively easier group to capture, and the second a relatively more difficult group to capture. As this is a mixture model, some extra restrictions are needed to allow all the parameters in the model to be identified. In our mixture model, we imposed an order restriction on θ_1, θ_2 to force $\theta_1 \leq \theta_2$. To do this, we created $\theta_1^*, \theta_2^*, \pi \overset{\text{i.i.d.}}{\sim} Beta(a, b)$. The probability at the $r = 2$ support points is given by $1 - \pi_2 = \pi_1 = \pi$. Our order restriction sets $\theta_1 = Min(\theta_1^*, \theta_2^*)$. Similarly, $\theta_2 = Max(\theta_1^*, \theta_2^*)$. For our simulations (Section 2.3), we fixed $a = b = 0.5$.

Model M_b: $p_{ij} = \alpha I(\tau_{ij} = 0) + \beta I(\tau_{ij} = 1)$

where $\tau_{ij} = 1$ if animal i has been captured before capture period j, and $\tau_{ij} = 0$ otherwise. This means that a trap-effect is assumed after the first capture. Subsequent captures are assumed not to cause such an effect. We use $\alpha, \beta \overset{\text{i.i.d.}}{\sim} Beta(a, b)$ as the prior distributions. For our simulations (Section 2.3), we fixed $a = b = 0.5$.

Model M_{th}: $p_i = (p_{i1}, p_{i2}, \ldots, p_{ik})^T \sim F(\cdot)$.

Similar to an M_h model, we assume a finite mixture distribution for the k-dimensional distribution function, F, specifically the p.m.f. is given by

$dF(p) = \sum_{m=1}^{r} \pi_m I(p = \theta_m)$, where π_m denotes the probability at support point $\theta_m = (\theta_{1m}, \ldots, \theta_{km})^T$ with the restriction that $\sum_{m=1}^{r} \pi_m = 1$. For our simulations (Section 2.3), we fix $r = 2$ mass points representing possibly two distinct groups within the population, which means that $1 - \pi_2 = \pi_1 = \pi$. We impose similar order restrictions as in M_h on the parameters θ_{1j}, θ_{2j}. That is, $\pi, \theta_{1j}^*, \theta_{2j}^* \overset{\text{i.i.d.}}{\sim} Beta(a, b)$,

$\theta_{1j} = Min\left(\theta_{1j}^*, \theta_{2j}^*\right)$, and $\theta_{2j} = Max\left(\theta_{1j}^*, \theta_{2j}^*\right)$ for $j = 1, \ldots, k$. For our simulations (Section 2.3), we fixed $a = b = 0.5$.

Model M_{tb}: $p_{ij} = \beta_j I(\tau_{ij} = 1) + \alpha_j I(\tau_{ij} = 0)$.

where τ_{ij} is as defined in model M_b. As in model M_b, the trap effect is assumed to occur only after initial capture but now allowed to vary over capture periods. For $j = 1, 2, ..., k$, define α_j as the capture probabilities for capture period j for animals that have not been previously captured. For $j = 2, 3, ..., k$, define β_j as the capture probabilities for animals that have been captured prior to capture period j. Note that β_1 is undefined, because subsequent capture cannot occur in the first capture period, so we fix $\beta_1 = 0$. In addition, we assume that $\beta_j = \alpha_j + c$, for $j = 2, \ldots, k$ We use $\alpha_j \overset{\text{i.i.d.}}{\sim} Beta(a, b)$ as the prior distribution for α's. The conditional prior distribution of c, given $\alpha_2, ..., \alpha_j$ is $U(-\min_{2 \leq j \leq k} \alpha_j, 1 - \max_{2 \leq j \leq k} \alpha_j)$, where $U(a, b)$ denotes a uniform distribution over (a, b).

Model M_{bh}: $p_{ij} = p_{1i} I(\tau_{ij} = 0) + p_{2i} I(\tau_{ij} = 1)$ and $p_i = (p_{1i}, p_{2i})^T \sim F(\cdot)$

where τ_{ij} is as defined in model M_b. Similar to the M_h model, we assume a finite mixture distribution for the 2-dimensional distribution function, F, specifically $dF(p) = \sum_{m=1}^r \pi_m I(p = \theta_m)$, where π_m denotes the probability at support point $\theta_m = (\theta_{1m}, \theta_{2m})^T$ with the restriction that $\sum_{m=1}^r \pi_m = 1$. For our simulations (Section 2.3), we fix $r = 2$ mass points representing possibly two distinct groups within the population, which means that $1 - \pi_2 = \pi_1 = \pi \sim Beta(a, b)$. Again, we impose order restrictions on the θ parameters for convergence and use $\pi, \theta_{1m}^*, \theta_{2m}^* \overset{\text{i.i.d.}}{\sim} Beta(a, b)$ for $m = 1, 2$ as a prior distribution. We set $\theta_{11} = Min\left(\theta_{11}^*, \theta_{12}^*\right)$, and $\theta_{12} = Max\left(\theta_{11}^*, \theta_{12}^*\right)$.

Similarly, we set $\theta_{21} = Min\,(\theta_{21}^*, \theta_{22}^*)$, and $\theta_{22} = Max\,(\theta_{21}^*, \theta_{22}^*)$. For our simulations (Section 2.3), we fixed $a = b = 0.5$.

Model M_{tbh}: $p_i = (p_{i1}, \ldots, p_{i2})^T = p_i^{(1)} I(\tau_{ij} = 0) + p_i^{(2)} I(\tau_{ij} = 1)$ and $(p_i^{(1)}, p_i^{(2)}) \sim F(\cdot)$.

This is the most general model that allows effects for individual heterogeneity, time, and behavior. Similar to M_{bh} and M_{th} models, we assume a finite mixture distribution for the $2k$-dimensional distribution function, F, specifically that

$dF(p) = \sum_{m=1}^{r} \pi_m I(p = \theta_m)$, where π_m denotes the probability at support point $\theta_m = (\theta_{11m}, \ldots, \theta_{1km}, \theta_{21m}, \ldots, \theta_{2km})^T$ with the restriction that $\sum_{m=1}^{r} \pi_m = 1$. As in model M_{tb}, we set the behavior effect to be additive. That is, $\theta_{2jm} = \theta_{1jm} + c$ for $j = 2, \ldots, k$ and $m = 1, 2, \ldots, r$. For our simulations (Section 2.3), we fix $r = 2$ mass points representing possibly two distinct groups within the population, which means that $1 - \pi_2 = \pi_1 = \pi$. Prior distributions for π and θ_{1jm} are $\pi, \theta_{1jm}^* \overset{\text{i.i.d.}}{\sim} Beta(a, b)$ for $j = 1, \ldots, k$ and $m = 1, 2$. A conditional prior distribution of c given θ_{1jm} for $j = 1, \ldots, k$ and $m = 1, 2$ is $U(-\min_{2 \leq j \leq k} \theta_{1jm}, 1 - \max_{2 \leq j \leq k} \theta_{1jm})$, where $U(a, b)$ denotes a uniform distribution over (a, b). Notice that this implies the behavioral effect c is constant across the m groups within the population. Also, we imposed similar order restrictions as in the previous heterogeneity models, so that $\theta_{1j1} \leq \theta_{2j1}$ and $\theta_{1j2} \leq \theta_{2j2}$. For our simulations (Section 2.3), we fixed $a = b = 0.5$.

The number of parameters in each model as a function of r, the number of support points assumed from the finite mixture distribution F and k, the number of capture periods, is given in Table 2.1. Using model M_{tbh} as an example, we have parameters $N, \pi_1, \ldots, \pi_{r-1}$, $\theta_{111}, \ldots, \theta_{1k1}, \theta_{112}, \ldots, \theta_{1k2}, \ldots, \theta_{11r}, \ldots, \theta_{1kr}$, and c. The number of parameters is thus $1 + (r -$

	Model Parameters
Model	Number of Parameters
M_0	2
M_t	1+k
M_b	3
M_h	2r
M_{tb}	k+2
M_{bh}	3r
M_{th}	r(k+1)
M_{tbh}	r(k+1) + 1

Table 2.1: Parameters in Bayesian Closed Population Capture-Recapture Models

$1) + kr + 1 = r(k+1) + 1$. The other entries in Table 2.1 can be established in a similar fashion.

The posterior distributions of the model parameters for all eight models can be closely approximated using Markov Chain Monte Carlo (MCMC) methods available in the WinBUGS V1.4 software package

(http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml).

2.3   A Simulation Study

2.3.1   Data Generation Process

The objective of the simulation is to determine whether standard model selection criteria, such as AIC and DIC, can correctly identify the true underlying model when all eight models are fit to a particular data set. Furthermore, we wish to assess if a particular model is robust when used to analyze data generated from a different model. To this end,

| Model | Data Generation Process | | |
	Capture Probabilities	Mean of p_{ij}	Std Dev of p_{ij}
M_0	0.2	0.2	0
M_b	1st capture: 0.2, after 1st capture: 0.41	0.27	0.10
M_t	(0.09, 0.13, 0.2, 0.29, 0.41)		
	for captures 1-5	0.39	0.27
M_h	$p_{ij} = p_i = F(\mu + \kappa z_i)$	0.25	0.2
M_{tb}	1st captures: same as M_t.		
	2nd captures: (0.29, 0.4, 0.53, 0.65)		
	for captures 2-5	0.28	0.18
M_{bh}	1st captures: $p_{ij} = p_i = F(\mu + \kappa z_i)$.		
	2nd captures: $p_{ij} = p_i = F(\mu + \kappa z_i + \eta)$	0.27	0.22
M_{th}	$p_{ij} = F(\mu + \kappa z_i + \phi_j)$	0.27	0.22
M_{tbh}	1st captures: $p_{ij} = F(\mu + \kappa z_i + \phi_j)$.		
	2nd captures: $p_{ij} = F(\mu + \kappa z_i + \phi_j + \eta)$	0.32	0.28

Table 2.2: Data generating assumptions for each of the 8 Bayesian models for simulation experiment one. F refers to the Logistic distribution function $F(x) = [1 + e^{-x}]^{-1}$, $\mu = F^{-1}(0.2) = -1.39$, $\kappa = 1.25$, $z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$, $\eta = 1$, and $\phi_j = \frac{j-3}{2}$, $j = 1, 2, ..., 5$

we generated 100 data sets under each modeling assumption

$(M_0, M_t, M_h, ..., M_{tbh})$. Using Markov Chain Monte Carlo (MCMC) methods, we fit each

data set using each of the eight models. A total of one-hundred data sets generated under

each of the eight modeling assumptions gives a total of eight-hundred data sets. Each data

set is a simulated capture-recapture study with $k = 5$ capture periods. For each simulated

data set discussed in this section, we set $N = 500$, and the p_{ij} values were generated

according to the design outlined in Table 2.2. Ultimately, as we discuss in Section 2.4, we

performed this experiment under different design conditions seven more times to verify our

conclusions about estimation of N and performance of the model selection criteria. Thus,

the design outlined and the results presented in this section refer to results from simulation

experiment one.

For each data set, and under each model, an estimate of the posterior density of N was constructed using WinBUGS Version 1.4 software, and the median of this posterior distribution was chosen as an estimate of N. We also computed the AIC (Section 2.3.3), and the coverage probability of a 95 % equal-tailed interval from the posterior distribution of N by computing the 2.5th and 97.5th percentile of N, and these results follow in Table 2.4. When fitting our models to the generated data sets, we made no assumptions about the capture probabilities being small or large. Using a large range of the prior distribution of N sufficiently allows for capture probabilities to be large or small. However, the range of N would be specific to each study.

We used a burn-in period of 3000 samples, and 2000 samples from each of three MCMC chains with dispersed starting values for the model parameters. Therefore, our posterior distribution estimates are based upon 6000 total samples.

Results of the first simulation experiment are given in Tables 2.3 - 2.12. Each table's columns correspond to the data generating process, that is, which model's assumptions were used to generate the data. Each table's rows correspond to the model fit to the data. For example, the first row and fourth column of any table correspond to data generated under model M_b, but analyzed using model M_0. Similarly, the fourth row and first column represents performance of M_0 data sets analyzed using Model M_b.

Table 2.3 contains results from the posterior density of N. The table lists means and standard errors, computed from the 100 MC estimates of the posterior median of N. Table 2.5 contains means and standard errors of the MCMC estimates of AIC. Table 2.6 gives the percentages of times each model was selected by AIC. Table 2.7 contains means

and standard errors of the MCMC estimates of DIC. Table 2.8 gives the percentages of times each model was selected by DIC, out of the 100 data sets generated under each set of modeling assumptions. Table 2.9 gives the mean values of the MCMC estimates of the penalty term in DIC, as a means to assess the effective number of parameters in DIC. Further description of the DIC criterion is given in Section 2.3.4. Tables 2.10, 2.11, and 2.12 give the same corresponding results as in Tables 2.7, 2.8, and 2.9, with the difference between tables arising because different estimators of the model parameters were used for each group of tables. These different estimators are discussed in Section 2.3.4.

2.3.2 Estimation of Population Size N

For estimating N, simulation results are given in Table 2.3. Estimated coverage probabilities for a 95% equal-tailed posterior interval from the posterior distribution of N is seen in Table 2.4.

Because we expect a fitted model to estimate N accurately when the data analyzed are sampled from the true model (for example, fitting Model M_{th} to data generated with time and heterogeneity effects), we expect the estimation of N to be accurate on the diagonals of each of the above tables, as the diagonals represent cases where the data generating assumptions match the model used to fit the data. Secondly, we expect that a fitted model would accurately estimate N when fit to data generated under simpler assumptions, specifically when the simpler model can be viewed as a special case of the more complex model. For example, data generated under model M_0 with constant capture probability across all captures, should fit well when analyzed with a model such as M_{th}, which allows the capture probabilities to vary by individual animal and by capture period. We expect

Table 2.3: Means and Standard Errors of Posterior Median as an estimator of N (True $N = 500$; Number of Simulations 100; Standard Errors given in parentheses)

Model Fit	Data Generating Process (DGP)							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	500.27	355.96	530.53	379.38	327.29	423.22	384.00	356.29
	(24.73)	(21.34)	(21.34)	(12.34)	(10.97)	(15.11)	(14.18)	(12.02)
M_h	526.68	433.77	551.99	401.07	361.06	441.44	456.18	394.18
	(29.72)	(54.83)	(25.38)	(19.58)	(22.30)	(22.94)	(51.12)	(25.06)
M_t	497.21	355.34	501.01	377.44	326.64	399.98	374.76	348.68
	(24.32)	(12.68)	(19.09)	(12.31)	(10.95)	(13.27)	(13.26)	(11.73)
M_b	506.99	368.65	1123.63	517.48	367.87	1119.58	1001.10	1005.10
	(48.92)	(17.59)	(11.14)	(56.90)	(18.32)	(16.10)	(49.35)	(38.26)
M_{bh}	545.28	434.80	1123.59	563.46	457.24	1119.24	1010.00	1012.92
	(54.66)	(63.35)	(11.01)	(64.48)	(81.52)	(16.30)	(41.72)	(32.82)
M_{tb}	518.10	460.70	506.86	513.00	461.33	469.78	458.63	468.12
	(71.00)	(55.58)	(44.99)	(70.24)	(60.07)	(45.13)	(53.63)	(54.06)
M_{th}	511.04	401.80	512.41	396.60	354.13	417.63	415.53	371.41
	(27.08)	(28.68)	(21.96)	(17.99)	(17.78)	(18.58)	(25.90)	(15.87)
M_{tbh}	502.74	398.70	501.59	488.59	413.88	460.00	425.26	421.18
	(55.06)	(42.94)	(32.80)	(65.05)	(53.06)	(41.98)	(36.04)	(45.55)

Table 2.4: Coverage Probability of $\hat{N}$ from 95 percent Equal Tail Posterior Interval (True $N = 500$)

Model Fit	Data Generating Process (DGP)							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	95	0	69	0	0	0	0	0
M_h	93	69	61	95	16	98	87	37
M_t	95	0	97	0	0	0	0	0
M_b	97	2	0	93	1	0	0	0
M_{bh}	99	83	0	94	92	0	0	0
M_{tb}	98	99	94	98	99	72	80	86
M_{th}	95	32	97	22	5	21	33	1
M_{tbh}	99	82	98	99	91	84	69	77

M_{th} is sufficiently robust to perform well in this situation because M_0 can be viewed as a special case of M_{th}.

Because the tables have been arranged with the complexity of the models in roughly increasing order, entries below the diagonal indicate situations where the model fit is more complex than the data generating assumptions. We generally expect to see accurate estimation of N in these situations. We expect less accuracy in estimating N when the column number is larger than the row number, as this (generally) represents situations where the data contain more effects than the model to which it was fit.

The results in Table 2.3 give some insights into estimation of N. Firstly, notice that when the data are generated from model M_0 (entries in column one of the table), that all the models except M_{bh} provide estimates of N with little to no bias. Model M_{bh} has a relative bias of +0.10. There is some slight bias to models M_{tb} and model M_h (with relative biases of +0.04 and +0.05, respectively), but overall performance is pretty good for simple data.

Data generated under model M_h (column two of Table 2.3) show that none of the eight models estimate N particularly well for these data sets. The smallest bias in estimating N came from Model M_{tb}, with a relative bias of -0.08. We expected a reasonable estimate of N would be provided by at least the four models which account for heterogeneity. In analyzing the behavior of the p_{ij} values generated in this simulation, their distribution follows a right skewed shape on the (0,1) interval. Thus it appears our mixture-modeling approach to heterogeneity does not always accurately estimate N. Although estimation of N from model M_h is poor (relative bias -0.13), it is better than that of M_{th} and M_{tbh},

showing us that addition of the extra effects in those higher models detracts from estimation of N, perhaps because those models may be finding effects in the data which are simply a result of random variation.

In column three of Table 2.3, data generated under model M_t have accurate estimates of N for models with a t subscript. The time effects in the data were somewhat pronounced, so this is not surprising. Estimation of N from each of these models performs well. What is surprising is the incredibly biased estimates of N that come from models with b but not t in the subscript. This effect occurs a few times in these simulation results, but can be explained somewhat by a result on the necessary conditions for the MLE of N to exist under model M_b (Seber and Whale, 1970) and page 29 of Otis et al. (1978). For the MLE of N to exist for model M_b, the number of captured animals in the early captures must be large enough to allow for estimation of capture probabilities and population size. However, our models had time effects that were increasing with the number of captures. Hence, many of our data sets had a relatively small number of early captures, violating the requirements for the MLE of N to exist. Even though we take a Bayesian approach to modeling and use the posterior density of N as the basis for inference about N, this feature of the likelihood causes our estimates of population size to be significantly biased. This problem recurs for almost all fitted models with subscript b where time effects are present in the data. Notice, though, that M_{tb} does estimate N reasonably well for these data sets (relative bias $+0.01$), suggesting that when both time and behavior effects are incorporated into the same model, the problem is resolved by the accounting for time effects.

Column four of Table 2.3 shows that data generated under model M_b have the

most accurate estimates of N when fit by model M_b, model M_{tb} (relative bias around +0.03 in both models), and by model M_{tbh} (relative bias of -0.02). Model M_{bh} has a larger relative bias (approximately +0.13) in estimating N, which is surprising since this model accounts for behavior effects. A behavior effect splits the population into two groups, namely those animals captured previously and those not captured previously. This is the type of effect we model with our heterogeneity models. However, the proportion of animals in each group is not constant when the effect is behavioral. The membership of the group of animals which have been previously caught grows with each sample taken. This causes problems in fitting model M_{bh}, even though it also accounts for behavioral effects. Also, we see somewhat the reverse effect of the problem encountered earlier. Due to the significant trap effect in capture period two, some animals have a higher capture probability than in capture period one. Still more animals have a higher capture probability in capture period three than in capture period two, and so on. If the model analyzing the data has only a time effect but not a trap effect, such as model M_{th}, this behavior effect is similar to a time effect. An overstated time effect causes, in these cases, a reduced estimate of N (relative biases of -0.20 to -0.25).

Column five of Table 2.3 shows that, while M_{bh} and M_{tb} (with relative biases of -0.08 and -0.09 percent) provide the best estimation of N when data are generated under model M_{bh} assumptions, estimates of N are still not very accurate. None of the eight models perform very well in estimating N, although the performance of M_{bh} relative to the other models tells us that estimation of N when using the correct model has smaller relative bias than all other models except model M_{tb}.

Column six of Table 2.3 gives results similar to those from column five. Model M_{tb} provides the most accurate estimates of N (relative bias is -0.06). Six of the eight models have negative bias in estimation of N. The two that are severely positively biased are again those models only incorporating the behavior effect, but not the time effect. This causes the same problem we saw when fitting behavioral models to data generated under model M_t. Model M_{tb} performs better than M_t, with an average bias of -30 versus -100, suggesting that incorporation of the behavioral effect in the model leads to smaller absolute bias in estimation of N.

Column seven of Table 2.3 gives results for data generated according to model M_{th}. Estimation of N is biased for all the models fit to this data. Curiously, model M_h has smaller absolute bias than model M_{th} when estimating N. The reason for this discrepancy is not known at this time.

Column eight gives results for data generated according to model M_{tbh}. Again, estimation of N is biased for all models. Model M_{tb} gives the best estimation of N (relative bias of -0.06), followed by model M_{tbh} (relative bias of -0.16). This may be due to the confounding of trap effects, and the way we've modeled the heterogeneity.

It should again be noted that we used relatively small capture probabilities for this study. It is known (Otis, et al, 1978) that smaller values of p_{ij} can lead to negative bias in estimating population size. This effect occurs in this simulation, particularly in the data sets for which we simulated a heterogeneity effect. It is possible that similar effects at higher mean values of p_{ij} would lead to different performance in estimating N. However, all models estimate N for data generated from model M_0 quite well, which gives some evidence

that the problem in estimating N arises largely from the variability of the populations, not simply in the relatively small mean values of p_{ij}. Perhaps a larger number of captures (we used $k = 5$ captures) would also allow the effects in the data to be more clearly manifested in the more complex models. Manifestation of time, behavioral, and heterogeneity effects (such as in M_{tbh}), for instance, may be a lot to ask of a data set containing only $k = 5$ captures. However, in a larger number of captures, perhaps those effects would be more clearly identifiable. Another possibility for improving performance with respect to the data sets containing heterogeneity is to re-examine all the data sets, using a version of Model M_h that has capture probabilities themselves generated from the beta distribution.

2.3.3 Analysis of AIC as a Model Selection Criterion

AIC (Akaike, 1973) has been used extensively as a model selection tool. Calculation of AIC adds a parameter penalty to the estimated Kullback-Leibler Discrepancy between the fitted model and the true model. Using θ as a general term to represent all the model parameters (e.g. $\theta = (N, P)$ as in (2.1)), X as a general term to represent the observed data (e.g. $X = (Z_1, ..., Z_L)$ as in (2.1)), p' as the number of model parameters (see Table 2.1), and $\mathrm{Log}L$ as the log likelihood function, a form for calculation of AIC is given by

$$AIC = -2LogL(\hat{\theta}|X) + 2p',$$

Table 2.5: Means and Standard Errors of AIC Posterior Mean

Model Fit	Data Generating Process							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	134.28	197.15	320.72	191.77	252.44	533.40	354.33	539.01
	(7.45)	(15.98)	(25.63)	(12.52)	(19.52)	(37.62)	(29.40)	(36.43)
M_h	138.41	159.21	325.05	194.40	196.56	537.15	328.54	501.36
	(7.28)	(7.03)	(25.65)	(11.99)	(13.51)	(37.66)	(27.55)	(34.94)
M_t	142.38	205.36	136.21	173.18	236.21	164.62	198.07	239.68
	(3.55)	(16.13)	(8.35)	(10.11)	(17.13)	(12.95)	(16.69)	(23.50)
M_b	136.37	197.93	207.97	157.92	201.64	227.15	193.56	181.33
	(7.20)	(15.90)	(14.08)	(6.95)	(14.40)	(15.65)	(11.87)	(13.03)
M_{bh}	142.51	163.02	214.79	163.58	165.45	234.12	184.46	174.53
	(7.05)	(6.75)	(14.13)	(6.85)	(7.12)	(15.59)	(10.57)	(10.42)
M_{tb}	144.21	203.75	137.92	165.81	207.14	151.14	186.13	174.49
	(6.91)	(15.68)	(8.27)	(6.92)	(14.06)	(7.32)	(11.74)	(12.97)
M_{th}	153.84	174.97	147.71	176.99	181.10	161.94	166.74	165.08
	(6.14)	(6.44)	(7.36)	(6.20)	(7.45)	(7.00)	(6.23)	(7.90)
M_{tbh}	155.64	176.64	149.40	177.14	179.82	162.12	168.48	163.48
	(5.94)	(6.30)	(7.50)	(6.14)	(6.16)	(6.55)	(6.15)	(6.57)

where $\hat{\theta}$ is the MLE of θ under the assumed model. However, our AIC calculation is different from the usual form of AIC. Defining

$$D(\theta) = -2LogL(\theta|X),$$

we use $AIC = E[D(\theta)|X] + 2p'$ where $E[D(\theta)|X]$ represents the mean of the posterior distribution of $D(\theta)$.

Analysis of Table 2.5, which gives the MCMC AIC means and standard errors, indicates that overall the AIC is effective in determining the correct model. For the first seven columns in the table, the smallest AIC mean value occurs when the fitted model

Table 2.6: AIC Model Selection: Percentage of times each model selected

Model Fit	Data Generating Process (DGP)							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	92	0	0	0	0	0	0	0
M_h	0	95	0	0	0	0	0	0
M_t	1	0	87	0	0	1	0	0
M_b	7	0	0	99	0	0	0	0
M_{bh}	0	5	0	1	99	0	2	12
M_{tb}	0	0	13	0	0	98	3	10
M_{th}	0	0	0	0	1	2	87	36
M_{tbh}	0	0	0	0	0	0	8	44
Model Selection Error	8	5	13	1	1	2	13	52

matches the data generating assumptions. This suggests that although estimation of N may be biased even in the correct model, the AIC is capable of identifying it correctly.

Perhaps more indicative of the performance of AIC is the summary in Table 2.6, which tells us the percentage of selections for each model using the AIC criterion. Ideally, the diagonal entries in the table should have the highest percentages of selections by AIC. The columns of the table represent the true model generating assumptions. When another model is selected by AIC, this may be called a model selection error, and the percentage of times AIC makes a model selection error is listed in Table 2.6. In this respect, for seven of the eight models, AIC performs quite well as a model selection tool. Among these seven models, for M_t and M_{th}, the percentage of selections is 87 percent, which is somewhat lower than for the other models. When M_t and M_{th} are not selected by AIC, though, AIC does select a similar model, albeit with more effects. This is better than the selection of an unrelated model. Model M_{tbh} does not perform as well. Data generated under the

assumptions of M_{tbh} only produced a 44 percent selection rate by AIC. When M_{tbh} was not selected in this column, the model selected was one of the sub-models containing two of the effects (M_{th}, M_{bh}, and M_{tb}). Some of this could be due to relative weighting of the time, behavioral, and heterogeneity effects within M_{tbh}, as AIC may be picking the model based on the most significant of these effects present in any particular data set. Furthermore, as previously stated, with five captures, model M_{tbh} may be somewhat over-parameterized. We observe 31 distinct capture histories, and model M_{tbh} includes 13 parameters for such data, which may lead to the estimation of effects due only to random chance. However, from an overall look at this table, we conclude that AIC performs well as a model selection tool.

2.3.4 Analysis of DIC as a Model Selection Criterion

The DIC criterion is a recent development in model selection. DIC can be expressed in a form similar to AIC. Given the common use of AIC, this feature allows users to quickly understand the form and use of DIC. Another significant benefit of DIC is that it is easy to calculate, as it is just a function of the posterior parameters and the model deviance (where deviance is related to the log-likelihood).

DIC can be expressed similarly to AIC. Using the same notation as in the definition of AIC, and again denoting

$$D(\theta) = -2LogL(\theta|X),$$

and defining

$$pD = E(D(\theta)|X) - D(\hat{\theta}),$$

where again $E(D(\theta)|X)$ represents the posterior mean of the deviance function $D(\theta)$, we denote

$$DIC = D(\hat{\theta}) + 2pD \qquad (2.2)$$

where $\hat{\theta}$ is a posterior estimate of θ, e.g., $\hat{\theta} = E[\theta|X]$ or $Median[\theta|X]$. As stated previously, the pD term in DIC represents an effective number of parameters. The pD term measures the decrease in the deviance (corresponding to increase in the likelihood) obtained by using posterior estimates of the parameters θ. It is important to note that although DIC is structured to look like AIC, the penalty term is actually a function of the fit of the model itself, not simply a discrete number of parameters.

Generally for computational purposes, defining $Dev(\theta)$ as the MCMC computed deviance for any particular data set and model combination, and defining $\bar{D}$ as the MCMC mean of the deviance statistic, pD is computed as

$$pD = \bar{D} - Dev(\hat{\theta})$$

and computationally, we have

$$DIC = Dev(\hat{\theta}) + 2pD.$$

In our simulations, DIC did not perform as well as AIC in model selection. A table of how often each of the models was selected using the DIC criterion is given in Table 2.8. For simple data sets (such as M_0, M_b, or M_t), DIC selects a more complex model in the majority of cases. Of particular concern was the disproportionately large

Table 2.7: Means and Standard Errors of DIC

Model Fit	Data Generating Process (DGP)							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	131.26	194.41	317.65	188.98	249.75	530.54	351.55	536.27
	(7.46)	(15.97)	(25.64)	(12.52)	(19.51)	(37.62)	(29.39)	(36.43)
M_h	123.25	152.09	311.45	178.64	191.05	522.87	320.33	494.86
	(13.16)	(7.47)	(26.20)	(14.59)	(13.22)	(38.75)	(28.15)	(34.75)
M_t	135.36	198.64	129.17	166.41	229.50	157.80	191.30	232.92
	(7.08)	(16.13)	(8.36)	(10.10)	(17.13)	(12.93)	(16.68)	(23.48)
M_b	131.75	194.10	201.89	153.23	197.82	221.08	187.70	175.48
	(7.28)	(15.90)	(14.06)	(7.13)	(14.40)	(15.63)	(11.86)	(13.03)
M_{bh}	100.64	150.71	140.52	127.81	149.81	173.37	162.60	156.78
	(16.40)	(8.46)	(24.48)	(14.50)	(9.68)	(24.84)	(20.69)	(15.65)
M_{tb}	129.33	182.62	129.50	150.70	186.95	142.88	177.15	165.39
	(7.74)	(21.30)	(8.23)	(6.93)	(16.50)	(7.27)	(11.51)	(12.70)
M_{th}	120.43	161.64	112.19	155.77	168.23	141.73	153.12	151.97
	(17.52)	(6.30)	(17.98)	(8.59)	(7.35)	(9.49)	(5.89)	(7.91)
M_{tbh}	126.15	157.44	125.47	145.35	157.55	131.03	151.99	145.74
	(13.44)	(7.39)	(9.33)	(8.52)	(8.05)	(11.40)	(5.84)	(7.26)

number of selections of model M_{bh} across all data sets. The problem with DIC arises in the computation of the penalty term pD. Although pD is positive in most cases, it is possible that pD can be negative for a particular model and data set, if the likelihood function is not log-concave. If pD is negative, instead of being penalized for the model complexity, the model is in fact rewarded by a negative pD value.

In Table 2.9, M_{bh} and other models have negative mean pD values in our simulations. In particular, model M_{bh} frequently has a negative pD. These negative values cause M_{bh} to be selected by DIC a significant number of times, even often in cases where the underlying data were generated under significantly different assumptions. Such performance is cause for concern. However, in Speigelhalter, Carlin, et al 2002, the authors state that

Table 2.8: DIC Model Selection: Percentage of times each model selected

Model Fit	Data Generating Process (DGP)							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	0	0	0	0	0	0	0	0
M_h	1	37	0	0	0	0	0	0
M_t	0	0	2	0	0	0	0	0
M_b	0	0	0	0	0	0	0	0
M_{bh}	82	51	13	95	78	5	15	13
M_{tb}	0	1	1	1	2	0	0	0
M_{th}	14	0	71	1	0	7	38	9
M_{tbh}	3	11	13	3	20	88	47	78
Model Selection Error	100	73	98	100	22	100	62	22

Table 2.9: mean values of pD: $\bar{D}(\theta)$ - $D(\hat{\theta})$

Model Fit	Data Generating Process (DGP)							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	0.97	1.26	0.93	1.22	1.30	1.14	1.22	1.26
M_h	-7.16	0.89	-5.60	-7.76	2.49	-6.28	-0.21	1.50
M_t	4.98	5.27	4.96	5.24	5.29	5.18	5.23	5.24
M_b	1.38	2.17	-0.08	1.31	2.18	-0.06	0.14	0.15
M_{bh}	-29.87	-0.31	-62.27	-23.77	-3.64	-48.75	-9.86	-5.75
M_{tb}	-0.88	-7.13	5.58	-1.11	-6.19	5.74	5.02	4.90
M_{th}	-9.40	10.67	-11.52	2.78	11.13	3.79	10.37	10.89
M_{tbh}	-3.49	6.81	2.07	-5.79	3.72	-5.09	9.51	8.25

Table 2.10: Means and Standard Errors of DIC: when posterior median used for $\hat{\theta}$

Model Fit	Data Generating Process							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	131.27	194.42	317.66	188.99	249.76	530.55	351.55	536.28
	(7.46)	(15.97)	(25.64)	(12.52)	(19.51)	(37.62)	(29.39)	(36.43)
M_h	129.61	153.97	316.62	181.07	191.80	525.07	323.13	496.52
	(7.19)	(7.00)	(25.78)	(15.41)	(13.45)	(39.52)	(27.63)	(34.91)
M_t	135.37	198.64	129.18	166.41	229.50	157.80	191.30	232.93
	(7.08)	(16.13)	(8.36)	(10.10)	(17.13)	(12.93)	(16.68)	(23.48)
M_b	132.35	194.19	202.31	153.85	197.91	221.50	188.07	175.83
	(7.21)	(15.90)	(14.07)	(6.98)	(14.39)	(15.64)	(11.86)	(13.03)
M_{bh}	130.07	155.50	202.91	150.43	157.58	221.69	174.89	164.59
	(8.04)	(6.75)	(14.10)	(7.81)	(7.04)	(15.84)	(10.70)	(10.44)
M_{tb}	135.60	195.10	129.80	157.25	198.49	143.13	178.10	166.43
	(6.89)	(15.61)	(8.26)	(6.87)	(14.09)	(7.30)	(11.71)	(12.92)
M_{th}	135.53	161.80	129.26	160.76	168.31	146.65	153.42	152.18
	(6.37)	(6.33)	(7.62)	(6.69)	(7.40)	(6.94)	(6.17)	(7.89)
M_{tbh}	135.72	161.69	129.90	152.07	164.69	142.00	153.51	148.56
	(6.83)	(6.17)	(7.77)	(8.90)	(5.99)	(7.01)	(6.09)	(6.54)

alternative choices for $\hat{\theta}$ could be the posterior median or posterior mode. So, pD can be calculated with these alternatives to the posterior mean of $\hat{\theta}$. The poor performance of DIC when $\hat{\theta}$ was the posterior mean called for examination of DIC when the posterior median was used for $\hat{\theta}$ instead.

In Tables 2.10, 2.11, and 2.12 the performance of DIC improves when the posterior median is used for $\hat{\theta}$. However, the model selection error rate is still quite high when compared with that of AIC in Table 2.6. However, in focusing on Tables 2.11 and 2.12, although the mean pD values are negative less frequently when using the posterior median, the mean pD value is still negative for the M_h and M_{bh} models. Secondly, in Table 2.12, DIC still tends to overfit each of the models. For instance, in column three, where the data

Table 2.11: mean values of pD: $\bar{D}(\theta)$ - $D(\hat{\theta})$ where $\hat{\theta}$ is the Posterior Median

Model Fit	Data Generating Process							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	0.99	1.27	0.94	1.23	1.32	1.14	1.23	1.27
M_h	-0.81	2.77	-0.43	-5.33	3.24	-4.08	2.59	3.16
M_t	4.99	5.27	4.97	5.24	5.29	5.18	5.23	5.24
M_b	1.98	2.26	0.34	1.94	2.27	0.35	0.51	0.50
M_{bh}	-0.44	4.48	0.12	-1.15	4.13	-0.43	2.43	2.06
M_{tb}	5.40	5.35	5.88	5.43	5.35	5.99	5.97	5.94
M_{th}	5.69	10.83	5.55	7.77	11.22	8.71	10.68	11.09
M_{tbh}	6.08	11.05	6.49	0.93	10.86	5.88	11.03	11.07

Table 2.12: DIC Model Selection: Percentage of times each model selected when posterior medians used for $\hat{\theta}$

Model Fit	Data Generating Process							
	M_0	M_h	M_t	M_b	M_{bh}	M_{tb}	M_{th}	M_{tbh}
M_0	8	0	0	0	0	0	0	0
M_h	44	79	0	2	0	0	0	0
M_t	2	0	52	0	0	0	0	0
M_b	1	0	0	10	0	0	0	0
M_{bh}	40	15	0	55	95	0	0	4
M_{tb}	3	0	15	0	0	31	0	1
M_{th}	0	2	25	1	1	17	71	21
M_{tbh}	2	4	8	32	4	52	29	74
Model Selection Error	92	21	48	90	5	69	29	26

was generated under the assumptions of model M_t, DIC chose model M_t for 52 of the 100

data sets. Alternative models chosen for the other 48 data sets were M_{tb}, M_{th}, and M_{tbh},

which are more complex versions of model M_t. Although all the chosen models account for

time effects, they also contain effects such as behavioral effects and heterogeneity effects

that were not present in the population. DIC also selects a more complex model than

necessary for data generated under models M_b and M_h. Overall, though, the performance

of DIC in model selection for these models is inferior to that of AIC, and based on this

simulation study, use of AIC as a model selection tool is recommended over DIC.

2.4   Further Simulation Experiments

To determine whether our conclusions reached in Section 2.3 hold in general, we

examined whether our conclusions held under data generated with the following factors:

1. True Population Size N

2. Amount of Heterogeneity in the Data (Small Amount, Large Amount)

3. Magnitude of Time Effects

4. Direction of Behavioral Effects

5. Average Capture Probability

Therefore, we performed the same analysis for 7 more combinations of data, gen-

erated with combinations of the above factors. Because even two levels of each factor above

would result in $2^5 = 32$ analyses, we chose a set of design parameters in a fashion similar

Experiment	N	Average p_{ij}	Time Effects	Behavioral Effects	Heterogeneity
1	500	0.2	Large	Positive	Large
2	500	0.2	Small	Positive	Small
3	500	0.4	Large	Negative	Large
4	100	0.4	Large	Positive	Small
5	100	0.4	Small	Positive	Large
6	100	0.2	Large	Negative	Small
7	500	0.4	Small	Negative	Small
8	100	0.2	Small	Negative	Large

Table 2.13: Data generating assumptions for simulation experiments 1 to 8

Model	p_{ij}
M_0	$F(\mu)$
M_b	$F(\mu + \eta\tau_{ij})$
M_t	$F(\mu + \beta_j)$
M_h	$F(\mu + \kappa Z_i)$
M_{bh}	$F(\mu + \eta\tau_{ij} + \kappa Z_i)$
M_{tb}	$F(\mu + \beta_j + \eta\tau_{ij})$
M_{th}	$F(\mu + \beta_j + \kappa Z_i)$
M_{tbh}	$F(\mu + \beta_j + \eta\tau_{ij} + \kappa Z_i)$

Table 2.14: Calculations of p_{ij} for simulation experiments 1 to 8; F refers to the Logistic distribution function $F(x) = [1 + e^{-x}]^{-1}$, $Z_i \stackrel{\text{i.i.d.}}{\sim} N(0,1)$, $\tau_{ij} = 1$ if the animal has been previously captured, and $\tau_{ij} = 0$ otherwise; values of μ, η, ϕ_j, and κ are given in Table 2.15

to a Fractional Factorial experimental design, with the goal of performing a total of eight analyses (including the one described in Section 2.3). Our method of data generation follows the same process as that outlined in Table 2.2, but is described in more detail in Tables 2.13 - 2.15.

For experiments two through eight, we analyzed a total of fifty data sets generated under each model's assumptions. That is, we generated fifty data sets with constant capture probability (Model M_0), fifty with time effects (Model M_t), etc. Because there are eight models, simulation experiments two through eight each analyzed a total of four-hundred data sets. From our data generating process, the means and standard deviations for the

Experiment	μ	ϕ_j	η	κ
1	$F^{-1}(0.2) = -1.385$	$\frac{j-3}{2}$ for $j = 1, ..., 5$	+1	1.25
2	-1.385	$\frac{j-3}{4}$ for $j = 1, ..., 5$	+1	0.25
3	$F^{-1}(0.4) = -0.405$	$\frac{j-3}{2}$ for $j = 1, ..., 5$	-1	1.25
4	-0.405	$\frac{j-3}{2}$ for $j = 1, ..., 5$	+1	1.25
5	-0.405	$\frac{j-3}{4}$ for $j = 1, ..., 5$	+1	0.25
6	-1.385	$\frac{j-3}{2}$ for $j = 1, ..., 5$	-1	1.25
7	-0.405	$\frac{j-3}{4}$ for $j = 1, ..., 5$	-1	0.25
8	-1.385	$\frac{j-3}{4}$ for $j = 1, ..., 5$	-1	0.25

Table 2.15: Design specifications for simulation experiments 1 to 8

capture probabilities p_{ij} for all simulation experiments are listed in Table 2.16. Results of simulation experiments two through eight are displayed in Figures 2.1 through 2.7. Recall that the primary observations from the first experiment were:

1. Models with behavioral effects, but not time effects, provide a poor estimate of N, on average, for data sets with time effects present.

2. Population sizes for data sets with a small number of effects (such as those generated under the assumptions of $M_0, M_b,$ or M_t) can be accurately estimated, on average, when the correct model is selected.

3. Heterogeneity in the population leads to negatively biased estimates of population size, even when the correct model is fit.

4. AIC selects the correct model a large percentage of the time.

5. DIC does not perform well in model selection due to a negative penalty term pD affecting some of the models.

Experiment	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
1	0.2	0.27	0.39	0.25	0.28	0.28	0.27	0.32
	(0)	(0.10)	(0.27)	(0.20)	(0.18)	(0.22)	(0.22)	(0.28)
2	0.2	0.27	0.22	0.20	0.27	0.27	0.21	0.27
	(0)	(0.10)	(0.12)	(0.04)	(0.14)	(0.11)	(0.07)	(0.15)
3	0.4	0.29	0.41	0.42	0.31	0.33	0.43	0.34
	(0)	(0.10)	(0.16)	(0.24)	(0.12)	(0.21)	(0.26)	(0.22)
4	0.4	0.53	0.41	0.40	0.50	0.53	0.41	0.50
	(0)	(0.12)	(0.16)	(0.06)	(0.24)	(0.14)	(0.17)	(0.24)
5	0.4	0.53	0.40	0.42	0.52	0.52	0.43	0.50
	(0)	(0.12)	(0.08)	(0.24)	(0.18)	(0.28)	(0.24)	(0.29)
6	0.2	0.16	0.22	0.20	0.18	0.16	0.22	0.18
	(0)	(0.05)	(0.11)	(0.04)	(0.10)	(0.06)	(0.12)	(0.10)
7	0.4	0.29	0.40	0.40	0.30	0.29	0.40	0.30
	(0)	(0.10)	(0.08)	(0.06)	(0.08)	(0.11)	(0.10)	(0.09)
8	0.2	0.16	0.21	0.26	0.17	0.19	0.26	0.20
	(0)	(0.05)	(0.06)	(0.20)	(0.06)	(0.16)	(0.21)	(0.16)

Table 2.16: Average Capture Probabilities, and their Standard Deviations, for Simulation Experiments 1 to 8

For simulation experiments two through eight, DIC continued to perform poorly as a model selection tool. This is again due to the occurrence of negative penalty terms pD. For this reason, the results of DIC for simulation experiments two to eight are omitted.

Figures 2.1 through 2.7 give results for both the average MCMC posterior median of N, and the AIC selection percentages, for the fifty data sets generated under each model's assumptions. Tables of the average posterior medians of N, and the average values of AIC are given in Section 2.6.

2.4.1   Simulation Experiment Two Summary

In the top panel of Figure 2.1, the horizontal axis is the true model used to generate the data. The vertical axis is the average value of the MCMC posterior median of N for the fifty data sets generated under the assumed model. Finally, because each data set was fitted under each of the eight models, there are eight separate plotted points, one for each model fit, for each data generating model. The posterior median of N for each fitted model is plotted according to symbols in the legend box at the right of the graph.

The bottom panel in Figure 2.1 gives AIC selection percentages for all the data sets generated for experiment two. The horizontal axis of this graph is the same as in the top panel of Figure 2.1. The vertical axis is the percentage of AIC selections out of the fifty data sets generated for each model. The bar shading indicates which fitted model (from the shading described in the legend box) is selected. The bars are stacked so that their total is one-hundred percent for each of the models.

The top panel of Figure 2.1 shows several similarities to the results from the first simulation experiment. Firstly, all models have little to no bias in estimating N. The models

Figure 2.1: Experiment 2 Posterior Median and AIC Results (True $N = 500$)
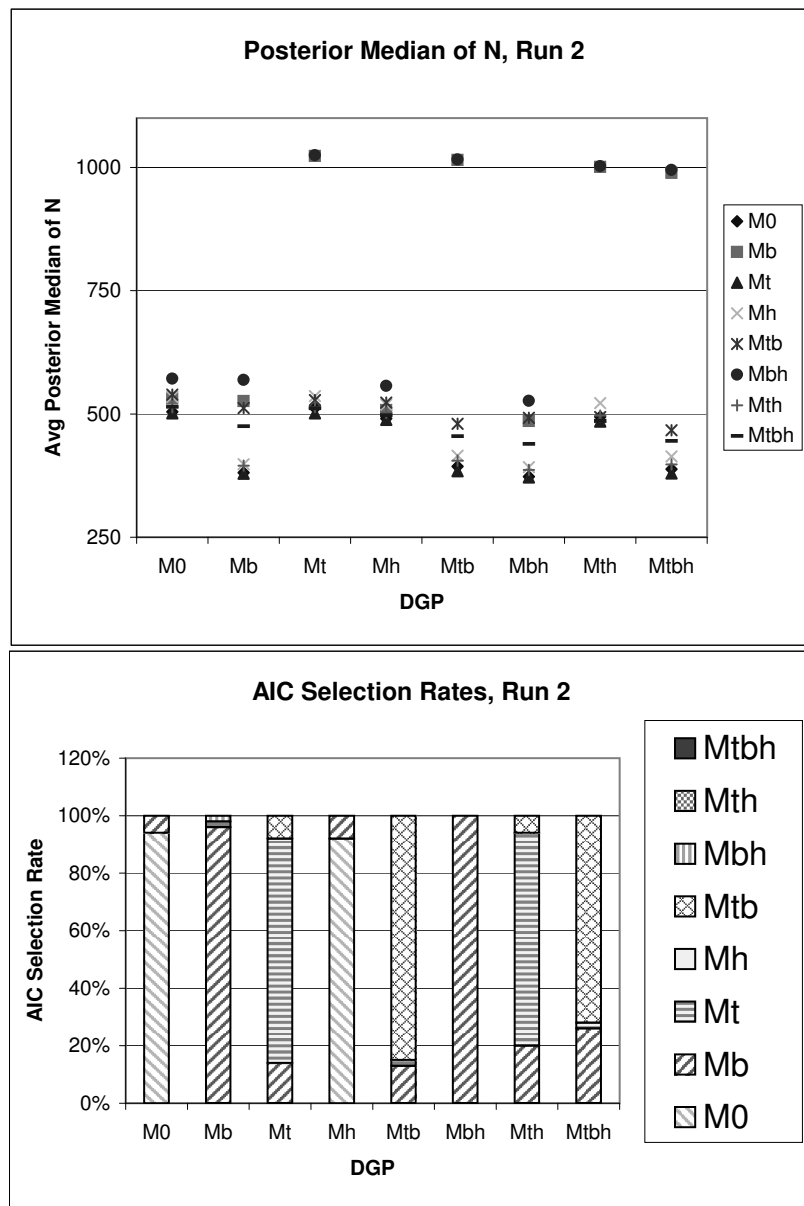
fit to the M_b data split into two groups. In the picture, models $M_0, M_t, M_h,$ and M_{th} show

negative bias. Models with behavioral effects (M_b, M_{tb}, M_{tbh} have average posterior median

values close to 500, ranging between 475 to 525). Model M_{bh} shows bias in estimating N,

with average posterior median of 568. Six models show at most a small bias in estimating

N when fit to data generated under Model M_t assumptions (because the time effects in experiment two are small in magnitude). However, as in the first simulation experiment, models M_b and M_{bh} are significantly positively biased when fitting data with time effects but not behavioral effects. The data sets generated under Model M_h show that all eight models show small to negligible bias in estimating N, due to the small heterogeneity effect in the data sets (see Table 2.13). When the amount of heterogeneity is small, the data sets come close to having a constant capture probability, and all eight models estimate N accurately when capture probabilities are constant. Model M_{tb} data shows that Model M_{tb} has the smallest bias in estimating N , which is expected. Models $M_b, M_{tb},$ and M_{bh} show at most small biases for estimating N when the data are from Model M_{bh}. All three of these models account for the behavior effects in the data. Model M_{th} data showed accurate estimation of N by all models except M_b and M_{bh}, because both the time effects and heterogeneity effects are small in magnitude, so most models are appropriate for these data sets, except for those with behavior effects and not time effects. Lastly, data sets generated with all three effects (time, behavior, and heterogeneity) have N estimated best by Models M_{tb} and M_{tbh}, although negative bias is seen in the estimates of N of these models.

The AIC results are favorable for data generated under models M_0, M_b, M_t and M_{tb}. We note from Figure 2.1 that AIC selects Model M_0 regularly for the M_h data sets. However, this reflects that using extra parameters to fit a small amount of heterogeneity does not provide a significantly better fit to the data, which is consistent with the purpose of AIC as a model selection criterion. In fact, for all data sets with heterogeneity, AIC selects the closest model without a heterogeneity effect at least 70% of the time. For example, for

M_{tbh} data, AIC selects Model M_{tb} most often. For M_{th} data, Model M_t is most commonly selected. It appears that the smaller magnitude of the heterogeneity effect for these data sets means that a heterogeneity effect in the model is not necessary to provide adequate fit to the data.

2.4.2 Experiment Three Summary

In Figure 2.2, all eight models have small, if any, bias in estimating N for the Model M_0 data sets, as all the average posterior medians of N are close to the true $N = 500$. For the M_b data sets, the Models M_b, M_{tb}, M_{bh} and M_{tbh} have posterior medians close to $N = 500$, while the other model estimates are positively biased, because the behavioral effect for the experiment three data is negative, and models without behavioral effects overestimate N. For the M_t data sets, most models accurately estimate N, except for Models M_b and M_{bh}, which are significantly positively biased, matching the results from experiment one. For the M_h data sets, the most accurate estimation of N occurs for a cluster of models including M_h, M_{th}, M_{bh}, and M_{tbh}, all of which show negative relative bias in estimating N of about seven percent. Of these four models, Model M_{bh} has the smallest absolute bias. This is similar to the results of experiment one, when the heterogeneity in the data was also large. Analysis of the M_{tb} data sets showed that the most accurate estimates of N occurred for Models M_{tb} and M_{tbh}, which both have average posterior median close to $N = 500$, with Model M_{tb} having the smallest relative bias, which is positive and near two percent. For the M_{bh} data sets, the best estimates of N are given by Models M_0, M_t, M_{tb}, and M_{tbh}. The Model M_{bh} posterior median is below the true $N = 500$, with a negative relative bias of about nine percent. For the M_{th} and M_{tbh} data sets, Models M_{bh} and M_b show positive

Figure 2.2: Experiment 3 Posterior Median and AIC Results (True $N = 500$)
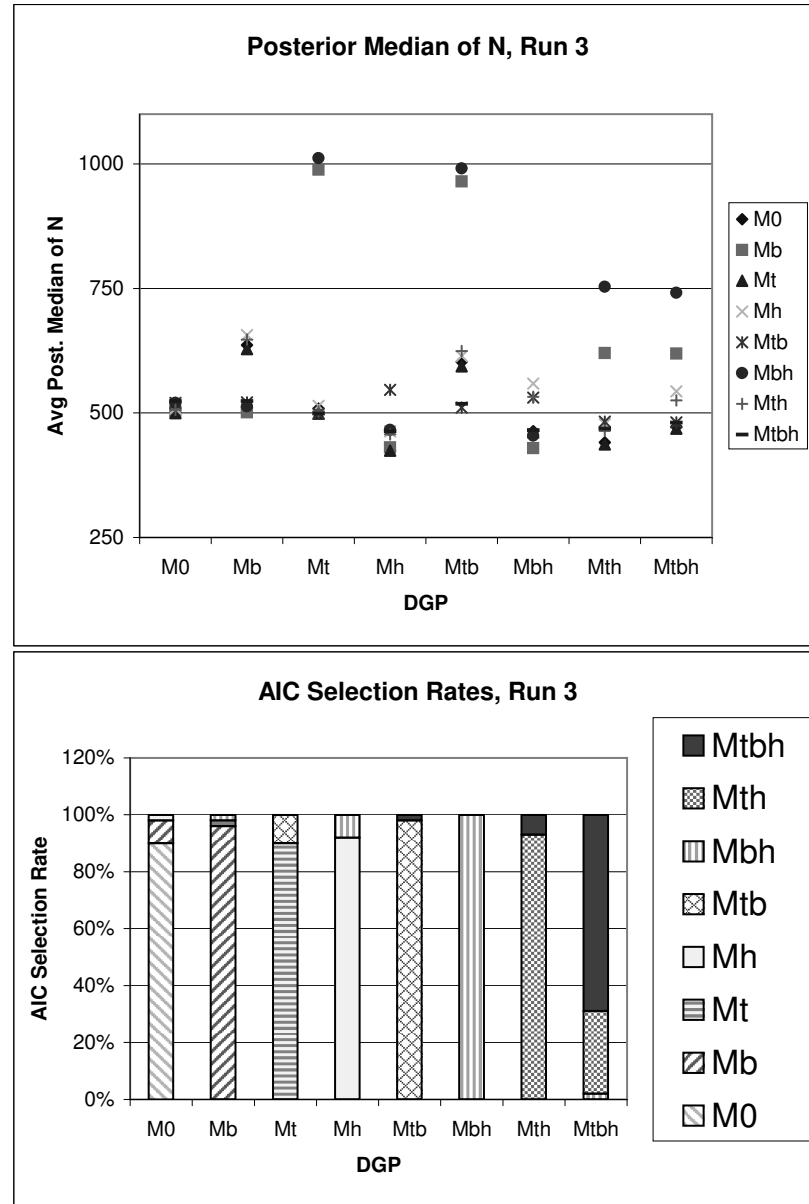
bias in estimating N, matching the results from simulation experiment one. For the M_{th}

data sets, the cluster of models M_h, M_{tb}, M_{th} and M_{tbh} are closest to the true $N = 500$,

but all show negative relative bias of between three and seven percent. For the M_{tbh} data

sets, Models M_{th}, M_{tb}, M_h, M_t, and even M_0 had average posterior medians of N near five-

hundred. The models M_h and M_{th} have averages above but close to the true $N = 500$. The other models mentioned show small negative bias in estimating N. Again, the behavioral models M_b and M_{bh} show large positive biases in estimating N.

The AIC analysis in Figure 2.2 shows that for the data sets in experiment three, AIC selects the true model the majority of the time for all data sets. This is due to the combination of the strong heterogeneity effects, large time effects, and large population size $N = 500$. AIC tells us that these data sets, given the strong underlying effects mentioned, are best fit by models that account for all the underlying sources of variability in capture probabilities.

2.4.3   Experiment Four Summary

Figure 2.3 gives results for simulation experiment four. In the first graph, all the points are very tightly clustered around the true value of $N = 100$, except when time effects are present in the data, and the model fit has behavior effects but not time effects. This tight clustering of the model results occurs because the average capture probability is high (near 40%) and the heterogeneity effects are small. Although the time effects are significant, because the average capture probability is high, even models which do not account for them, such as M_0, still provide an adequate fit, in the sense that on average, the posterior median of N is close to one-hundred.

Analysis of Figure 2.3 for the AIC results gives similar conclusions to those in experiment two. That is, $M_0, M_b, M_t,$ and M_{tb} data sets are selected most often by their corresponding model fit. However, data sets with heterogeneity are most commonly selected (by AIC) to be fit by the closest corresponding model which does not contain heterogeneity

Figure 2.3: Experiment 4 Posterior Median and AIC Results (True $N = 100$)
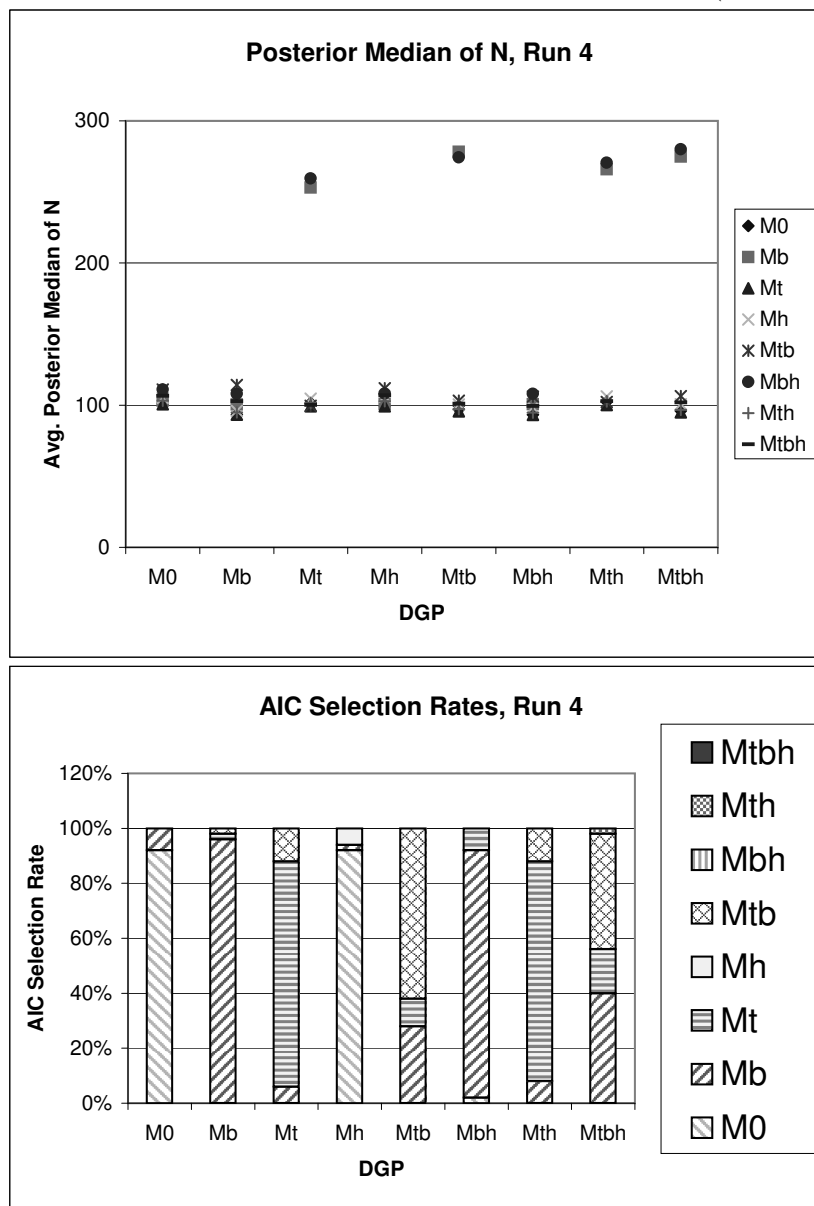
effects. It is noted, though, that the data sets generated under Model M_{tbh} do not specifically

follow this pattern, probably because the population size is small ($N = 100$), making the

clear discernment of all three sources of variability difficult in a small population.

2.4.4 Experiment Five Summary

Figure 2.4: Experiment 5 Posterior Median and AIC Results (True $N = 100$)
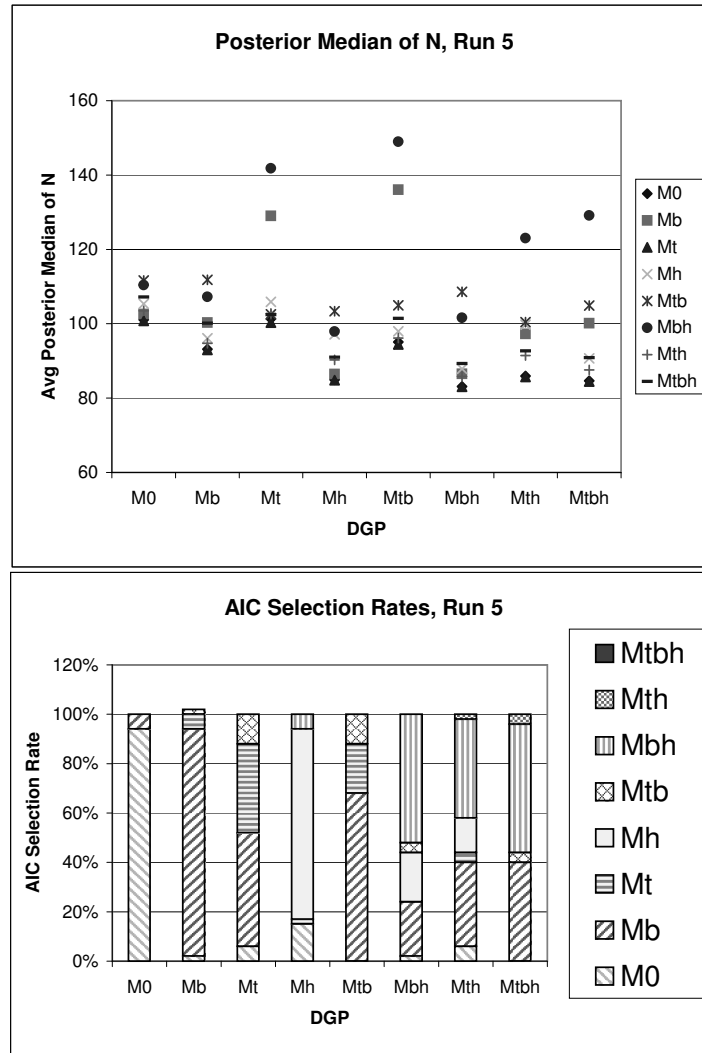
In Figure 2.4, the M_0 data sets are fit well by most models, with the highest observed relative bias in the posterior median of N being about 10%, for the Model M_{tb} fit. For the M_b data sets, the models M_b and M_{tbh} show negligible bias in estimating N, with both models showing average posterior median of N very close to the true $N = 100$, with the other models still having average posterior medians between ninety and one-hundred-

ten. The Model M_t data sets have the most accurate estimation of N by Model M_t, with most other models $(M_0, M_h, M_{tb}, M_{tbh})$ also clustered close to the true $N = 100$. Again the behavioral models M_b and M_{bh} are positively biased, although not as severely as in experiments one to three. For the M_h data sets, the true $N = 100$ is most accurately estimated by Models M_{bh}, M_h, and M_{tb}, which all show absolute relative biases of about three percent, while the rest of the eight models show larger negative relative bias. For the M_{tb} data sets, Models M_h, M_{tb}, and M_{tbh} show small absolute biases in estimating N, while models M_0 and M_t both show show a small negative bias. The same positive bias is noted in Models M_b and M_{bh}, for the same reasons noted previously.

The AIC results in Figure 2.4 show that for data sets of models M_0, M_b, and M_h, AIC chooses the true model in the significant majority of the cases. For data sets of model M_t, AIC chooses model M_t in 36% of the cases, compared with a 46% selection rate for Model M_b. Because the time effects for experiment five are small positive effects, the M_t data may not be clearly distinguishable from a trap-happy behavioral effect. Another reason for this may be that Model M_b has a smaller number of parameters than M_t (see Table 2.1), and that in many data sets the improved Model M_t fit does not justify the increase in the number of parameters. A similar result occurs for the M_{tb} data sets, where the most commonly selected model is M_b. For the M_{bh} data sets, AIC chooses Model M_{bh} most commonly, with Models M_b and M_h also being commonly chosen. These results make sense, and likely reflect AIC choosing the strongest effects present in the data when it chooses the sub-models M_b or M_h. For the M_{th} data AIC chooses M_b and M_{bh} most often as the best model. Due to the small magnitude of the time effects, these choices reflect that these

data sets can be best fit with simpler models with a smaller number of parameters. Some examination of the underlying results (not available in Figure 2.4) shows that the penalty term for the number of parameters is the reason that M_{th} has a higher AIC value for these data sets. Lastly, for the M_{tbh} data sets, AIC chooses Model M_{bh} most commonly, followed by model M_b. The choice of M_{bh} again reflects the small magnitude of the time effects in these data sets. The choice of model M_b is surprising given that the heterogeneity in the data is strong in experiment five. However, a behavioral effect and a heterogeneity effect are not completely unrelated. For capture periods two through k, the behavioral effect creates two distinct groups in the population: those which have been previously captured and those which have not been previously captured. Each group has separate capture probabilities. Although group membership is changing with each capture period, Model M_b fit could provide a reasonable fit to data with heterogeneity in some instances.

2.4.5 Experiment Six Summary

In the top panel of Figure 2.5, for the the M_0 data sets, we see that the closest estimates of the true $N = 100$ are given by Models M_t and M_{tb}, each with small negative relative biases of less than one percent. Models M_0 and M_{tbh} have some small positive relative biases of about two to three percent. The remaining models show some positive bias in the posterior median of N. The M_b data sets have estimates of N that are most accurate under Models M_{tb} and M_b, with all eight models showing some positive bias, likely due to the behavioral effect being negative (a trap-shy effect). For example, in a model such as M_0, the reduced number of subsequent captures causes the estimate of the capture probability to be small, corresponding to a larger estimate of N. For the M_t data sets,

Figure 2.5: Experiment 6 Posterior Median and AIC Results (True $N = 100$)
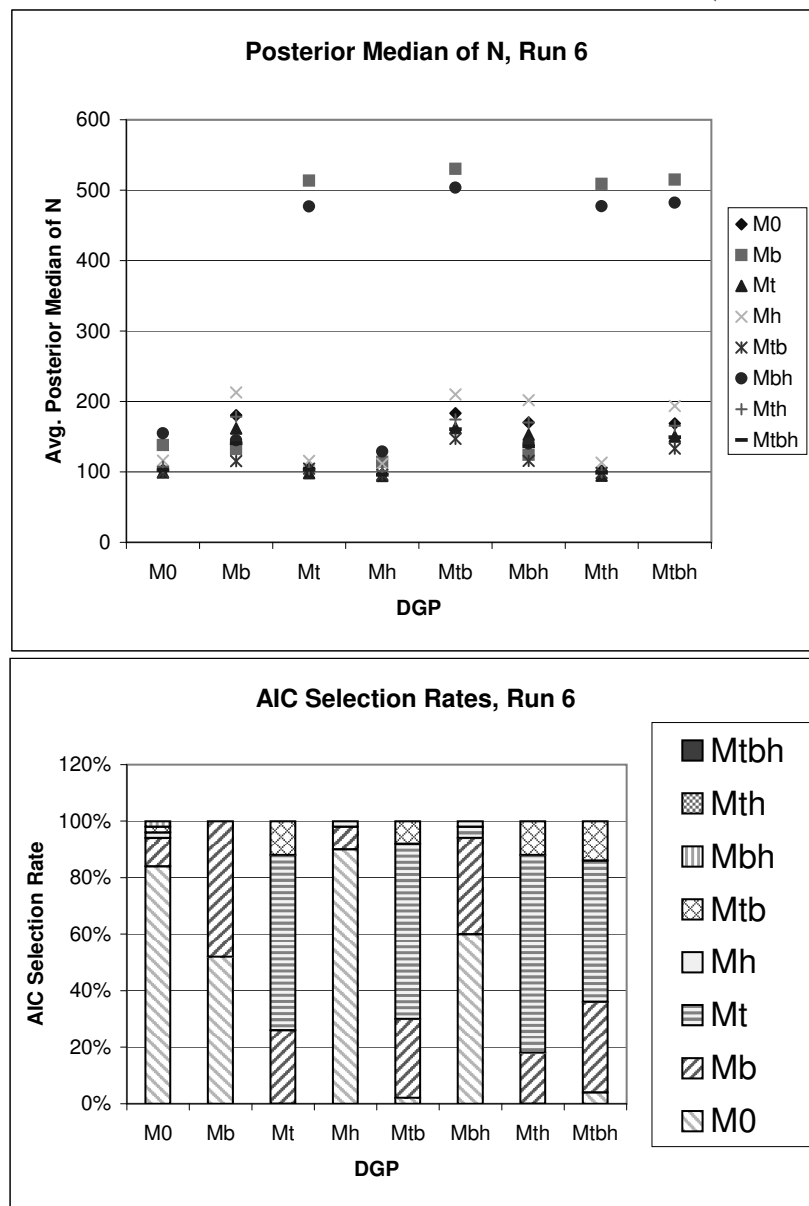
five of the models have small absolute relative biases between two and five percent. The models which show significant positive bias in estimating N are the behavioral models M_b and M_{bh}, due to the same reasons seen previously. For the M_h data sets, the most accurate estimation of N is given by Model M_0, with a negative relative bias of 3.1%. Several other

models have average posterior medians of N close to the true $N = 100$ as well. This occurs because the experiment six data have a small amount of heterogeneity, and most models (even those which do not model the heterogeneity effect) still produce estimates close to the correct level, although the M_{bh} model fit has a relative bias of approximately $+25\%$. For the M_{tb} data sets, all the models have positive bias in estimating N, with average values of the posterior median larger than one-hundred, but Model M_{tb} does provide the most accurate estimation of N. The best explanation for this is that the combination of large time effects and negative behavior effects for these data make estimation of N difficult. Positive bias under similar conditions for M_{tb} data sets occurred in Figure 2.2, although in that particular case the population sizes were five-hundred, and the average capture probability was high, which likely allowed for some increased accuracy in estimating N. For the M_{bh} data sets in Figure 2.5, all eight models fit show positive bias, with the most accurate fits given by Models M_{tb} and M_b. Model M_{bh}, the true model in this case, does not provide estimates of N as accurate as M_b and M_{tb}. In fact, Model M_{bh} shows a positive relative bias of approximately 40% for the M_{bh} data sets. This is better than from Models M_{th}, M_h, and M_0, but a more accurate estimate of N was expected. For the Model M_{th} data, we see two clusters of points, with one set all being very close to the true $N = 100$, and the other points showing high positive bias above $N = 100$. The outliers are again the models M_b and M_{bh} as has been previously noted. The relative accuracy of the other models for these data sets is largely due to the fact that the time effects are significant, but the heterogeneity effects are small. This combination allows Models such as M_t and M_{tb} to provide accurate estimation of N. The accuracy of Model M_0 in estimating N for these

data sets is surprising, given the large time effects in the data. This result (M_0 providing accurate estimates of M_{th} data) occurred also in experiment four results of Figure 2.3, and may reflect the ability of M_0 to estimate N accurately by using some measure of an average capture probability over all the samples. Lastly, for the M_{tbh} data sets, all eight models show positive bias, with average posterior medians above the true $N = 100$, due to the negative behavioral effect in these data sets, combined with the large time effects. The most accurate estimate of N is provided by Model M_{tb}, followed by Model M_{tbh}. Model M_{tb} provides an accurate estimate of N, because the time and behavioral effects are most significant in the M_{tbh} data sets. Still, the smallest absolute relative bias of these models (M_{tb}) still has positive bias of approximately 30%, suggesting that when all three sources of variation are affecting the capture probabilities, estimation of N is difficult.

In the bottom panel of Figure 2.5, for the M_0 data sets, AIC selects Model M_0 approximately 85% of the data sets. When the data were generated from Model M_b, AIC splits its selections evenly between Models M_0 and M_b, which is likely due to the trap-shy effect in these data sets. Although not a terrible result, this is unfortunate because the top panel of Figure 2.5 shows a clear advantage in accuracy using Model M_b to fit these data sets. This may be a consequence of the data reflecting a negative behavioral effect, making subsequent captures difficult when the population size is $N = 100$. If subsequent captures have small probability of occurrence, the data may consist of mostly initial captures (and fewer recaptures) which may lead to a data set that appears to be likely under Model M_0. It is noteworthy that in experiment three (see Figure 2.2) that this effect is not observed, and that AIC chooses Model M_b overwhelmingly. This suggests that the effect of the trap-shy

behavioral effect is not as significant in a population of larger size, such as $N = 500$. For the M_t data sets, Model M_t is chosen the majority of the time (over 60% of the time). The remaining data sets chose Models M_b or M_{tb}, both of which were seen to provide accurate estimates of N in the top panel of Figure 2.5. AIC chooses Model M_0 most often in experiment six for the M_h data sets, again due to the small amount of heterogeneity in the data. The M_{tb} data sets do reflect a problem in the data. Although AIC chooses Model M_{tb} or Model M_t a total of 70% of the time, Model M_b is chosen by AIC in 28% of the data sets. This is a problem because we have seen that Model M_b provides a heavily biased estimate of N for these data sets. Some examination of the underlying data shows that in the cases where M_b was chosen for these data sets, a more suitable model (such as M_{tb}) also had an AIC value close to the AIC value of Model M_b. This occurrence gives good reason to recommend using AIC as only part of the decision-making making process to identify suitable models, and not for it to be applied blindly. Finally, in Figure 2.5 we see that, for the M_{bh} data sets, AIC chooses Models M_0 or M_b most commonly. The choice of M_b is understandable because the heterogeneity in these data sets was small. The choice of Model M_0 appears to occur as noted above, when we saw AIC choose Model M_0 commonly for the M_b data sets. The M_{tbh} data sets generate a minimum AIC value commonly when fit by Models M_t and M_b in this simulation experiment, and Model M_{tb} is chosen a small percentage of the time. We believe this again reflects the small heterogeneity in these data sets.

Figure 2.6: Experiment 7 Posterior Median and AIC Results (True $N = 500$)
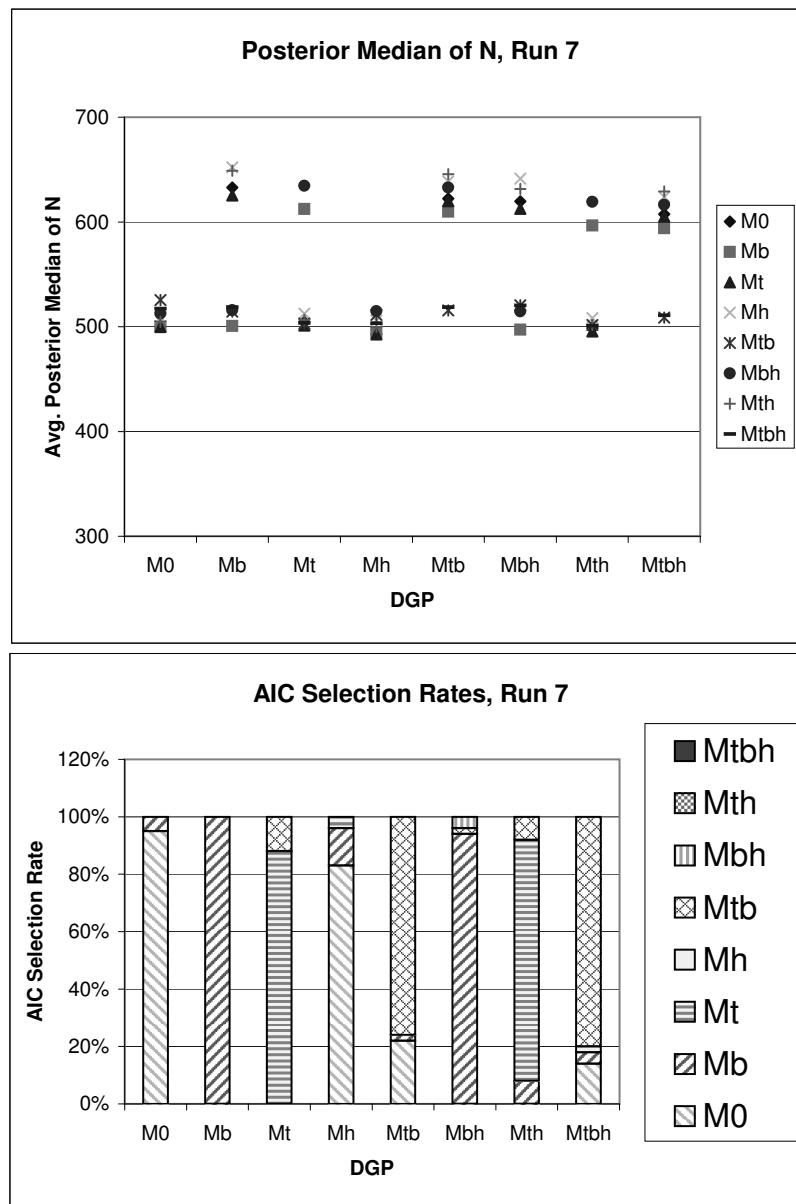
2.4.6 Experiment Seven Summary

In the top panel of of Figure 2.6, for the M_0 data sets, all eight models provide

an accurate estimate of the true $N = 500$. The most accurate estimation is given by the

posterior medians of N of Model M_0, and the absolute relative biases in estimating N for

models M_b and M_t are also very close to zero. For the M_b data sets, we see two clusters of points. One cluster of points are close to the true level of $N = 500$. Model M_b provides an approximately unbiased estimate of N, on average. Other points in the cluster close to $N = 500$ are Models M_{tb}, M_{bh}, and M_{tbh}. Models with positive bias are M_0, M_h, M_t, and M_{th}. These models do not account for the negative behavior effect in the data sets, which is similar to the results of experiments three (Figure 2.2) and six (Figure 2.5). Most of the eight models have average posterior medians of N that are close to $N = 500$ for the M_t data sets, as is shown in Figure 2.6, except for the previously noted outliers from Models M_b and M_{bh}. Due to the small amount of heterogeneity in the Model M_h data sets, all eight models show small absolute relative bias in estimating N. The M_{tb} data sets have the true N estimated most accurately by by Models M_{tb} and M_{tbh}, while all other models show positive bias in estimating N. Even though the time effects are of small magnitude, their inclusion in the chosen model to fit the data still provides a significant benefit, which can be seen by noting that Model M_{tb} is more accurate in estimating N than Model M_b for these data sets. For the M_{bh} data sets, the most accurate estimates of N are given by Models M_b, M_{bh}, followed by M_{tb} and M_{tbh}. That is, the models which contain a behavioral effect provide a relatively accurate estimate of N, the remaining models are positively biased in estimating N. The Model M_{th} data sets have a cluster of points near the true $N = 500$, indicating relatively accurate estimation of N, except for the previously noted outliers for Models M_b and M_{bh}. Finally, the M_{tbh} data sets have two points representing Models M_{tbh} and M_{tb} which have average values of the posterior median of N close to the true $N = 500$. The remaining models give an estimate of N that is positively biased.

The bottom panel of Figure 2.6 shows that the correct model is chosen in the vast majority of the M_0, M_b, M_t and M_{tb} data sets. AIC again predominantly selects Model M_0 for the M_h data sets because the amount of heterogeneity in the data is relatively small. Similarly, for the M_{th}, M_{bh}, and M_{tbh} data sets, AIC chooses the closest model without heterogeneity effects as the correct model, again reflecting that the improved fit from modeling heterogeneity is not significant enough in this case to account for the extra parameters required to fit these models.

2.4.7 Experiment Eight Summary

In the top panel of Figure 2.7, for the Model M_0 data sets, all eight models give estimates of N that are close to the true $N = 100$. The M_b data sets show some positive bias in estimating N, with the smallest absolute relative bias given by Model M_{tb}, followed by Model M_b. For the M_t data sets, the smallest absolute relative bias is given by Model M_t (positive relative bias of about 3%), and estimates of N are accurate, on average, for all models except the behavioral models M_b and M_{bh}, for the same reasons as seen earlier. For the M_h data sets, most models have some negative bias in estimating N, because all points lie below the true $N = 100$ in Figure 2.7. The most accurate estimates of N are provided by Models M_{bh} and M_h. Again, a large amount of heterogeneity in these populations causes some negative bias in estimating N. The M_{tb} data sets show positive bias for all eight fitted models, but the closest to the true $N = 100$ is Model M_{tb}. This effect likely occurs due to the negative behavioral effect in a smaller population, as we saw in simulation experiment six in Figure 2.5. For the M_{bh} data sets, the most accurate estimate of N is given by Model M_{th}, followed by Models M_{bh} and M_{tbh}. All eight models show some negative bias except

Figure 2.7: Experiment 8 Posterior Median and AIC Results (True $N = 100$)
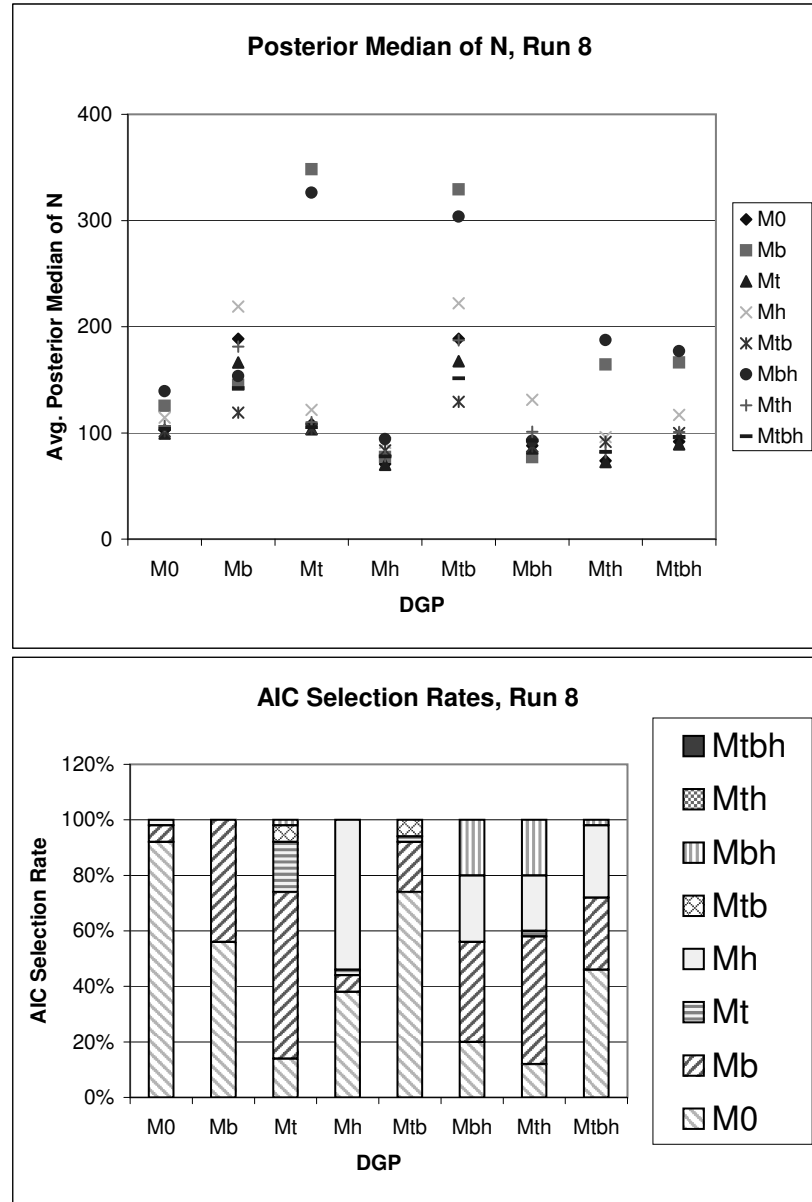
Model M_h. For the M_{th} data sets, the most accurate estimates of N are given by Model M_h, followed by M_{tb}, and then Models M_{tbh} and M_{th}. The behavioral models M_b and M_{bh} have the same bias problem noted previously. Finally, the M_{tbh} data sets have accurate estimates of N from Models M_{th}, M_{tb} and M_{tbh}. Again, the behavioral models M_b and M_{bh}

are heavily biased.

AIC selections in Figure 2.7 match the data generating assumptions regularly only for the M_0 data sets. For the M_b data sets, the Model M_0 is the most common choice, although M_b is the second most commonly chosen model for these data sets. Some analysis of the individual fifty data sets (detail not available in Figure 2.7) showed that when Model M_0 was chosen for the M_b data sets, the AIC values for M_b were very close to the M_0 AIC values, and should be studied as a suitable model for the data. Also in Figure 2.7, for the M_t data sets, AIC most commonly chooses M_b. This is again a concern because the M_b estimates of N are significantly biased for the M_t data sets. Again, some further analysis of the AIC values for each data set reveals that the AIC values of M_t and M_b are close in these cases, and the difference in AIC values is in fact due to the additional parameters required to fit Model M_t as opposed to Model M_b. Based on the AIC results for the M_b and M_t data sets, we reiterate that AIC should be used mostly as a guide to select a subset of suitable models for further analysis. For the M_h data sets, the Model M_h is the most commonly chosen model, but Model M_0 is chosen frequently as well. For the M_{tb} data sets, AIC most commonly chooses Model M_0. For the M_{bh}, M_{th}, and M_{tbh} data sets, no particular model is selected overwhelmingly, and the true model is also rarely selected for these data sets. We believe this reflects the combination of small population size of $N = 100$, the negative behavioral effects, the average capture probabilities being 20%, and the large degree of heterogeneity in the data. For a small population, it would be difficult to have one data set reflect all those sources of variation, and this is likely causing problems for a selection criterion such as AIC.

2.4.8 Conclusions on Estimation of Population Size and Model Selection

Based on the eight simulations, we make the following general conclusions regarding performance of the posterior median of N as an estimator of the true population size N:

- Model M_0 estimates N accurately for simple data sets, and handles small amounts of heterogeneity in the data, as well as moderate time effects. Behavioral effects in the data cause the most significant bias, with the direction of the bias being in the opposite direction of the behavioral effect. That is, when a trap-happy effect occurs in the population, Model M_0 shows negative bias in estimating N. Similarly, when a trap-shy effect occurs in the population, Model M_0 shows positive bias in estimating N.

- Model M_b performs well when the data have behavioral effects, and shows much smaller absolute relative bias in estimating N than models which do not account for behavioral effects. However, data with positive time effects cause significant positive bias problems for Model M_b, as has been noted in Seber and Whale, 1970.

- Model M_t performs well for simple data sets, and handles a minor amount of heterogeneity in the data, allowing for reasonably accurate estimates of N for data from M_{th}. When behavioral effects are present in the data, Model M_t estimates show the same bias described above for Model M_0.

- Model M_h: Our two-point mixture model for heterogeneity works well when handling minor amounts of heterogeneity in the data. Model M_h has negative bias in estimating

N for data sets with large amounts of heterogeneity, as has been noted previously by Otis et al. (1978). When behavioral effects are present in the population, M_h has bias in estimating N, in the direction opposite of the behavioral effect. Model M_h handles time effects in the data reasonably well.

- Model M_{tb} performed the best of all our models throughout the simulation. Overall, M_{tb} showed the most robustness in estimating N when fit to data sets from a model such as M_h, M_{th}, M_{bh}, and M_{tbh}, which cannot be expressed as special cases of M_{tb}. Model M_{tb} does not have the same bias problems as Model M_b when fit to data with time effects.

- Model M_{bh} did not predict N accurately in many of our simulated data sets. Even with relatively simple data sets such as M_0, or M_h data sets with a small amount of heterogeneity, Model M_{bh} did not perform as well as the other models. Furthermore, the estimates of Model M_{bh} were not always the most accurate even when fit to data sets that were generated under the model M_{bh} assumptions, which further shows that the Model M_{bh} did not consistently perform well even in cases where it should provide the best estimates of N.

- Model M_{th} performed well overall, providing accurate estimates for M_0, M_t, M_h, M_{th} and often M_{tbh} data sets. The one bias seen in this model was in fitting models with behavioral effects, and the direction of the bias was in the opposite direction of the behavioral effect.

- Model M_{tbh} performed reasonably well in our simulations, but did not provide the

most accurate estimates of N in many cases, even when the data were generated with time, behavior, and heterogeneity effects. This model large number of model parameters in M_{tbh} causes it to overfit many data sets. For the vast majority of data sets, a reduced model with fewer parameters seems to be sufficient to estimate N.

With regard to model selection criteria, we conclude that DIC does not perform well as a model selection tool for our Bayesian capture-recapture models, due to the problem with negative penalty terms described in Section 2.3.4.

We conclude that AIC performs well in choosing the correct model, or at least a suitable model, from a set of underlying data. There are some exceptions to this rule, as seen in experiment five when Model M_b is chosen a significant percentage of the time when the true data are from Model M_t. This is a situation where we know there is significant bias in estimating N, yet for many data sets, this model produced the minimum AIC value. In this specific case, several other models (with more accurate estimates of N) had AIC values that were close to the minimum, which leads to the question of how AIC should be interpreted when multiple models have values close to the minimum value of AIC.

Overall, we believe AIC provides a useful model selection criterion. We agree with the philosophy stated in Burnham and Anderson (2002) that AIC should be used to select a subset of candidate models worthy of further exploration from a larger set of models. At this model selection step, AIC should be used to select the subset of candidate models to explore further for a given data set. In reality, several models may have AIC values that are very close to the smallest one. Final decisions about the best model should be made from this subset of candidate models, allowing for other factors to impact the decision, such as

Selection Rates for data sets Generated via Listed Model Assumptions

Experiment	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
2	94%	96%	78%	0%	86%	0%	0%	0%
3	90%	96%	90%	92%	98%	100%	94%	70%
4	92%	96%	82%	6%	62%	0%	0%	0%
5	94%	92%	36%	78%	12%	52%	2%	0%
6	84%	48%	62%	2%	8%	0%	0%	0%
7	96%	100%	88%	0%	76%	4%	0%	0%
8	92%	44%	18%	54%	6%	20%	0%	0%

Table 2.17: Selection Rates for AIC for Simulation Experiments 2 to 8

the opinion of a subject area expert on the likely behavior of the underlying population.

A summary of the AIC selection rates from Figures 2.1 to 2.7 is given in Table 2.17. The most notable entries in Table 2.17 show low selection rates by AIC in Experiments two, four, six, seven, and eight. particularly when the true model had heterogeneity effects, as seen in columns four, six, seven, and eight. These low selection rates occur primarily for two reasons. Firstly, AIC does not select a heterogeneity model, such as M_h, when the amount of heterogeneity in the data is relatively small. Secondly, AIC selection rates are overall lower for our experiments where the population size was one-hundred, which occurred in experiments two, five, seven, and eight. Given the fractional factorial design used to generate our simulation parameters in Table 2.13, a more detailed analysis involving interactions of factors is not possible.

2.5   Analysis of Cottontail data set

In Edwards and Eberhardt (1967) a capture-recapture experiment involving one-hundred-thirty-five cottontail rabbits was performed. These rabbits were released into a forty acre rabbit-proof area, and eighteen capture periods followed after a four day period which allowed the rabbits to become familiar with their surroundings. Using just the first five capture periods (for computing purposes) we compare our Bayesian Models with those given in Program CAPTURE. The data are available through the Program CAPTURE website, http://www.mbr-pwrc.usgs.gov/software/, in the User's Manual. The first five capture periods produce thirty-one observable capture histories. Using our notation from Chapter 1, the data for the first five capture periods is: $(Z_1, ..., Z_{31}) =$

(4,11,0,3,0,0,0,2,1,1,0,0,1,1,0,4,1,0,0,1,1,0,0,0,0,0,0,1,0,1,0). A total of thirty-three rabbits were captured at least one time during the first five capture periods. We first fit the Program CAPTURE models to the data given above. Table 2.18 gives the estimate of N for each model, as well as the upper and lower limits of ninety-five percent confidence intervals for N. We also used the Model Selection tool in Program CAPTURE, which chose Model M_h as the best model.

We also fit our eight Bayesian models to the data, using the same non-informative prior distributions for the capture probabilities and for N described in Section 2.2. Table 2.19 lists the results of those analyses, listing the posterior median of N for each model, and the ninety-five percentm, equal-tailed posterior interval bounds from the MCMC posterior distribution of N, obtained via the WinBugs v. 1.4 software package. Figures 2.8 and 2.9 provide the MCMC posterior density of N for the M_0 and M_h models.

Program CAPTURE Results for Cottontail data set

Estimator	$\hat{N}$	LCL	UCL
M_0	50	40	74
M_h (jackknife)	76	59	107
M_b	75	39	374
M_t	49	40	72
M_t (Chao)	74	47	160
M_{bh}	78	39	374
M_{bh} (Pollock/Otto)	49	39	77
M_{th} (Chao)	95	56	201
M_{tb}	6120	95	602349

Table 2.18: Program CAPTURE Results for Cottontail data set (first 5 capture periods)

Bayesian Model Results for Cottontail data set

	N Posterior Percentiles			
Model	Median	2.5%	97.5%	AIC Mean
M_0	51	39	79	73.26
M_h	113	44	686	73.67
M_b	155	44	759	74.45
M_t	48	37	69	82.03
M_{tb}	72	39	324	82.32
M_{bh}	137	44	730	78.57
M_{th}	74	44	232	85.45
M_{tbh}	70	41	234	87.74

Table 2.19: Bayesian Model Results for Cottontail data set (first 5 capture periods)
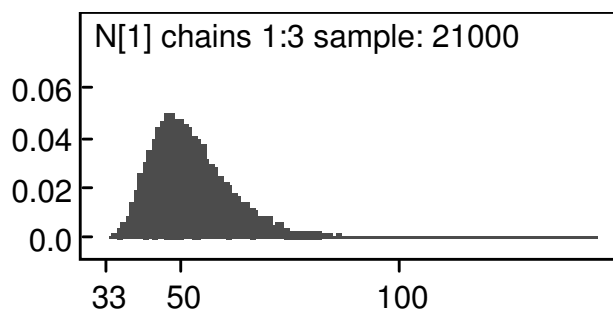
Figure 2.8: Posterior density of N for model M_0 for Cottontail dataset
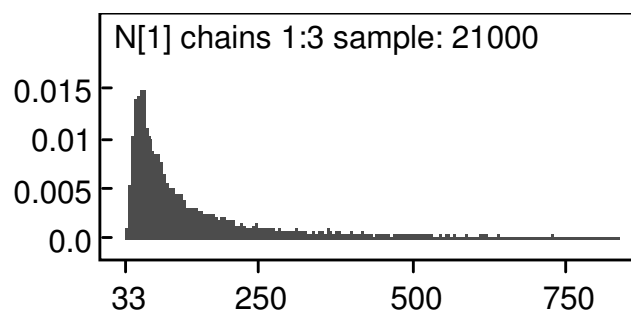
Figure 2.9: Posterior density of N for model M_h for Cottontail dataset

We first notice that the model selections are similar between the Bayesian models and those given in Program CAPTURE. Program CAPTURE chooses Model M_h, while using AIC as a model selection tool for the Bayesian Models would yield a set of candidate models including M_0, M_b, M_h, and perhaps M_{bh}, because all of these models have AIC values near the minimum value of 73.26 for Model M_0.

Using the posterior median of N as an estimator of population size, our Model M_0 underestimates the true population size for this population, as does the Model M_0 presented in Program CAPTURE. The interval estimates produced by the two methods are very comparable, which is not surprising given that we used relatively uninformative prior distributions for the parameters, and the likelihood functions of the two models are

the same.

The Bayesian M_h and M_b estimates differ from those in Program CAPTURE. Firstly, our method of modeling heterogeneity through the two-point mixture approach differs from the methodology used in the Jackknife M_h methods used by Program CAPTURE. In terms of evaluating the estimators, our posterior median is closer to the true population size $N = 135$ than the Jackknife Estimator. However, the posterior density of N for our Bayesian Model has a much higher 97.5th percentile than the upper bound for the confidence interval given for the Jackknife estimator in Program CAPTURE. Similarly, the posterior median of our Model M_b model is closer to the true population size, but again our 95% posterior interval is much wider than that given by Model M_b in Program CAPTURE. As can be seen from Figure 2.5, the posterior density for N is heavily right-skewed, which suggests the length of the confidence interval could be significantly shortened by choosing an interval other than an equal-tailed interval, or by lowering the confidence level. The end result is that our point estimate of N, via the posterior median, is more accurate than in Program CAPTURE estimates for M_h, and for M_b. However, the width of our posterior intervals offsets the benefit of a more accurate prediction. Because this is simply one data set, general conclusions cannot be made.

2.6 Simulation Results for Bayesian Models: Experiments Two to Eight Results

Table 2.20: Means and Standard Errors of Posterior Median: Experiment 2 (True $N = 500$)

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	504.52	381.2	511.76	490.88	393.66	372.78	494.1	387.8
	(29.89)	(18.22)	(23.40)	(22.98)	(12.48)	(12.63)	(22.84)	(12.53)
M_b	531.26	525.98	1023.18	507.76	1015	485.46	1000.82	988.82
	(62.80)	(73.92)	(33.73)	(50.88)	(45.07)	(40.93)	(41.35)	(54.73)
M_t	501.44	379.04	501.3	488.1	384.12	371.12	484.72	379.34
	(29.16)	(18.01)	(23.14)	(22.62)	(11.73)	(12.63)	(22.42)	(12.12)
M_h	534.9	398.22	536.44	519.04	415.42	392.12	522.18	413.64
	(41.45)	(21.76)	(27.56)	(31.71)	(26.83)	(20.86)	(26.36)	(30.10)
M_{tb}	539.46	511.74	528.16	523.14	480.3	492.28	493.88	466.78
	(76.03)	(74.55)	(72.85)	(79.50)	(53.46)	(57.38)	(54.43)	(48.54)
M_{bh}	571.42	568.92	1024.76	556.98	1016.26	526.56	1002.46	994.78
	(67.97)	(82.70)	(35.44)	(63.27)	(43.23)	(49.32)	(39.58)	(45.79)
M_{th}	513.6	395.1	513.86	502.14	405.06	386.3	496.98	397.76
	(29.55)	(22.45)	(25.98)	(27.46)	(16.11)	(14.39)	(23.48)	(16.58)
M_{tbh}	514.66	475.14	511.46	497.4	454.94	438.92	486.16	445.24
	(55.79)	(66.14)	(56.72)	(54.16)	(43.39)	(34.94)	(31.70)	(43.20)

Table 2.21: Means and Standard Errors of AIC: Experiment 2

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	131.89	189.71	181.93	135.24	318.72	191.61	185.05	316.25
	(7.25)	(12.91)	(14.19)	(6.41)	(23.46)	(14.15)	(16.96)	(26.62)
M_b	134.03	157.26	146.34	137.54	167.53	159.67	149.86	166.01
	(7.21)	(7.17)	(10.03)	(6.31)	(8.88)	(8.24)	(9.51)	(8.76)
M_t	140.72	170.33	140.11	143.32	172.05	174.20	143.31	173.00
	(7.20)	(8.54)	(7.04)	(5.40)	(9.94)	(8.59)	(6.73)	(10.94)
M_h	135.99	192.29	186.06	139.28	321.48	194.18	189.05	318.46
	(7.03)	(12.81)	(14.07)	(6.39)	(23.42)	(14.12)	(16.73)	(26.01)
M_{tb}	142.48	164.64	141.96	145.20	160.72	167.85	145.45	161.08
	(6.99)	(6.44)	(7.06)	(5.18)	(6.93)	(7.69)	(6.96)	(6.35)
M_{bh}	140.10	162.91	153.00	143.43	174.04	165.54	156.57	172.54
	(7.02)	(7.15)	(9.89)	(6.17)	(8.73)	(8.36)	(9.50)	(8.75)
M_{th}	152.67	175.42	151.72	154.85	171.79	179.10	154.74	172.02
	(6.66)	(6.01)	(6.41)	(4.84)	(7.21)	(7.03)	(6.07)	(6.51)
M_{tbh}	154.53	175.69	153.52	156.82	170.72	178.16	155.72	171.77
	(6.55)	(5.92)	(6.32)	(4.85)	(5.71)	(6.31)	(6.26)	(6.00)

Table 2.22: Means and Standard Errors of Posterior Median: Experiment 3 (True $N = 500$)

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	500.10	636.12	508.68	424.58	602.50	463.46	440.82	472.38
	(8.54)	(17.67)	(7.10)	(9.04)	(13.58)	(11.63)	(7.83)	(12.67)
M_b	503.62	501.38	988.32	430.70	965.04	429.24	620.04	619.16
	(11.36)	(10.84)	(104.85)	(10.19)	(103.54)	(10.44)	(45.08)	(60.91)
M_t	499.66	628.16	498.66	424.46	593.68	461.00	436.98	468.76
	(8.51)	(16.68)	(6.90)	(9.07)	(12.39)	(11.35)	(7.87)	(12.31)
M_h	508.52	656.50	514.06	462.78	614.36	558.88	477.50	543.32
	(9.88)	(21.93)	(7.78)	(14.57)	(14.49)	(43.18)	(18.61)	(37.03)
M_{tb}	520.14	521.50	498.54	546.46	510.52	531.00	482.62	481.08
	(41.93)	(87.47)	(10.83)	(106.83)	(9.48)	(94.06)	(21.55)	(24.98)
M_{bh}	520.14	513.20	1011.70	466.02	990.86	454.22	753.60	741.54
	(18.16)	(11.90)	(94.80)	(30.34)	(95.44)	(20.07)	(82.34)	(103.10)
M_{th}	506.16	647.22	503.04	457.08	624.16	532.32	463.90	525.24
	(10.06)	(28.10)	(6.77)	(13.56)	(32.15)	(31.52)	(13.16)	(29.82)
M_{tbh}	523.76	523.98	498.48	462.80	518.50	465.72	468.50	480.44
	(62.28)	(58.93)	(9.98)	(41.72)	(9.85)	(41.40)	(28.00)	(19.25)

Table 2.23: Means and Standard Errors of AIC: Experiment 3

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	168.74	204.58	432.69	244.30	245.36	252.53	427.62	281.64
	(7.30)	(14.17)	(31.11)	(17.76)	(20.05)	(22.36)	(25.59)	(20.34)
M_b	170.70	149.98	240.99	242.94	213.71	228.07	230.03	234.53
	(7.27)	(7.70)	(17.83)	(17.24)	(13.81)	(21.88)	(16.79)	(17.50)
M_t	176.92	167.67	167.08	252.87	177.12	228.46	247.29	224.38
	(6.18)	(10.38)	(6.20)	(17.58)	(8.25)	(20.35)	(21.32)	(16.76)
M_h	172.89	208.91	437.49	168.53	250.09	205.68	374.09	245.80
	(7.03)	(14.11)	(31.11)	(6.72)	(20.08)	(14.53)	(23.31)	(17.91)
M_{tb}	178.86	158.03	168.66	244.70	156.91	230.26	220.53	225.96
	(6.55)	(7.37)	(5.86)	(15.94)	(6.55)	(20.61)	(16.59)	(16.43)
M_{bh}	176.78	156.25	247.70	172.30	220.41	168.37	207.29	211.04
	(6.98)	(7.45)	(17.84)	(6.10)	(13.91)	(8.51)	(13.03)	(14.05)
M_{th}	187.98	178.23	178.79	184.50	186.10	184.10	178.38	186.13
	(5.96)	(9.40)	(6.09)	(5.00)	(7.54)	(6.85)	(5.95)	(5.83)
M_{tbh}	190.03	169.90	180.02	186.45	168.67	182.81	180.13	183.86
	(6.21)	(6.72)	(6.14)	(4.90)	(5.08)	(6.88)	(5.82)	(5.82)

Table 2.24: Means and Standard Errors of Posterior Median: Experiment 4 (True $N = 100$)

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	101.10	93.42	101.56	99.60	97.10	93.36	102.40	96.06
	(3.94)	(3.02)	(2.89)	(3.70)	(2.53)	(2.41)	(3.05)	(2.54)
M_b	103.16	100.56	253.14	100.78	278.20	100.58	265.96	274.90
	(6.75)	(4.45)	(92.50)	(4.82)	(106.18)	(5.64)	(98.78)	(93.93)
M_t	100.62	93.28	99.30	99.30	95.72	93.10	100.04	94.90
	(3.90)	(3.00)	(2.72)	(3.68)	(2.33)	(2.46)	(2.69)	(2.52)
M_h	105.06	96.10	104.62	105.96	100.42	98.26	106.24	100.64
	(5.31)	(3.42)	(3.96)	(15.23)	(6.28)	(7.70)	(5.50)	(8.39)
M_{tb}	110.92	114.24	99.56	111.96	103.20	106.02	102.22	106.30
	(14.27)	(24.16)	(4.42)	(17.80)	(6.37)	(12.90)	(8.74)	(10.38)
M_{bh}	111.02	107.90	259.54	108.12	274.22	108.02	270.60	280.08
	(11.49)	(7.82)	(85.44)	(9.11)	(84.78)	(9.90)	(87.64)	(81.99)
M_{th}	103.14	94.98	101.36	102.88	97.32	94.82	102.88	96.70
	(4.62)	(3.18)	(3.27)	(5.51)	(2.61)	(2.87)	(4.11)	(2.80)
M_{tbh}	106.60	102.72	100.14	106.96	101.00	99.32	102.86	101.84
	(10.49)	(8.93)	(3.39)	(11.74)	(5.12)	(6.31)	(6.34)	(6.24)

Table 2.25: Means and Standard Errors of AIC: Experiment 4

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	117.29	132.36	163.00	118.68	214.37	133.34	162.24	208.69
	(6.70)	(8.64)	(13.24)	(6.99)	(18.58)	(9.97)	(11.76)	(13.08)
M_b	119.29	116.47	128.42	120.94	113.17	117.50	127.31	111.14
	(6.60)	(6.10)	(10.14)	(6.91)	(8.77)	(5.98)	(9.35)	(9.65)
M_t	125.17	128.28	113.88	127.16	110.53	127.63	116.32	116.51
	(6.74)	(8.07)	(6.57)	(7.00)	(9.82)	(7.79)	(6.78)	(10.26)
M_h	121.83	136.09	167.78	122.72	218.86	136.51	166.93	212.08
	(6.78)	(8.21)	(13.35)	(6.65)	(18.42)	(8.84)	(11.94)	(12.95)
M_{tb}	127.46	124.55	115.65	129.43	104.60	125.71	118.09	107.67
	(6.88)	(6.30)	(6.64)	(6.98)	(6.88)	(6.16)	(7.03)	(6.41)
M_{bh}	125.91	123.05	134.85	126.96	119.71	123.88	133.72	117.35
	(6.66)	(5.97)	(10.21)	(6.50)	(8.73)	(5.67)	(9.40)	(9.86)
M_{th}	137.06	135.23	125.78	137.77	114.71	134.53	127.61	117.31
	(6.46)	(6.56)	(6.22)	(6.41)	(7.04)	(5.72)	(6.52)	(7.11)
M_{tbh}	139.03	136.23	127.15	139.69	115.85	136.30	129.20	118.52
	(6.44)	(6.24)	(6.38)	(6.10)	(6.61)	(5.46)	(6.81)	(6.76)

Table 2.26: Means and Standard Errors of Posterior Median: Experiment 5 (True $N = 100$)

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	101.10	93.18	101.30	84.96	95.08	83.12	85.96	84.66
	(4.06)	(2.92)	(2.91)	(4.55)	(3.32)	(4.82)	(4.18)	(4.26)
M_b	102.58	100.30	129.00	86.44	136.08	86.52	97.26	100.12
	(5.83)	(6.36)	(22.62)	(5.61)	(35.88)	(5.44)	(7.41)	(19.37)
M_t	100.76	93.00	100.34	84.82	94.44	83.08	85.68	84.50
	(3.97)	(2.89)	(2.77)	(4.50)	(3.19)	(4.77)	(4.25)	(4.18)
M_h	105.34	96.10	105.92	97.12	97.98	87.90	99.72	90.72
	(5.43)	(5.44)	(6.59)	(16.12)	(4.76)	(6.84)	(22.69)	(7.80)
M_{tb}	111.64	111.82	102.54	103.34	104.96	108.62	100.42	104.90
	(15.12)	(22.66)	(6.17)	(16.60)	(11.93)	(19.71)	(14.20)	(18.08)
M_{bh}	110.40	107.22	141.78	97.94	148.98	101.58	123.02	129.16
	(9.63)	(9.09)	(24.78)	(15.99)	(33.58)	(16.71)	(26.62)	(38.19)
M_{th}	103.46	94.72	102.78	90.26	96.16	85.82	91.46	87.60
	(4.59)	(3.30)	(3.11)	(6.39)	(3.51)	(5.38)	(6.23)	(4.78)
M_{tbh}	107.18	100.12	102.42	90.96	101.42	89.28	92.68	90.88
	(10.43)	(7.34)	(4.76)	(7.11)	(7.65)	(7.77)	(6.92)	(6.80)

Table 2.27: Means and Standard Errors of AIC: Experiment 5

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	116.38	131.18	129.49	126.03	166.75	137.07	135.80	159.65
	(5.44)	(10.12)	(9.32)	(9.78)	(11.87)	(12.96)	(9.57)	(14.00)
M_b	118.65	117.02	123.23	128.01	114.72	122.71	126.15	114.60
	(5.35)	(6.63)	(6.57)	(9.60)	(7.55)	(9.39)	(8.88)	(9.63)
M_t	124.61	127.47	122.66	134.48	121.92	138.86	134.00	137.60
	(5.20)	(7.41)	(5.51)	(9.58)	(9.30)	(12.15)	(10.02)	(11.63)
M_h	121.03	134.48	134.09	118.27	170.21	122.09	127.08	144.41
	(5.45)	(9.35)	(9.51)	(6.93)	(12.12)	(8.00)	(6.94)	(11.18)
M_{tb}	126.49	124.92	124.78	135.98	117.07	129.87	133.20	122.01
	(4.92)	(5.56)	(6.00)	(9.36)	(7.33)	(10.21)	(8.72)	(9.12)
M_{bh}	125.26	123.49	129.73	122.19	121.26	116.65	122.71	112.80
	(5.27)	(6.41)	(6.67)	(6.60)	(7.55)	(5.72)	(5.20)	(7.13)
M_{th}	136.26	134.05	134.52	132.92	126.84	127.31	130.46	120.69
	(4.89)	(5.32)	(5.48)	(5.34)	(7.02)	(6.79)	(4.99)	(6.66)
M_{tbh}	138.01	135.78	136.14	134.71	128.18	128.79	132.57	122.56
	(4.60)	(5.18)	(5.89)	(5.30)	(6.60)	(6.27)	(5.12)	(6.45)

Table 2.28: Means and Standard Errors of Posterior Median: Experiment 6 (True $N = 100$)

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	102.74	180.30	105.20	96.90	183.18	170.08	101.46	168.74
	(12.50)	(35.67)	(10.80)	(10.34)	(35.49)	(40.87)	(10.09)	(38.70)
M_b	137.98	132.22	513.64	114.22	530.20	124.08	509.06	514.86
	(59.48)	(63.54)	(67.10)	(39.64)	(61.69)	(43.56)	(59.71)	(65.55)
M_t	99.26	161.50	98.12	94.18	162.30	152.84	94.56	150.32
	(11.59)	(28.67)	(9.11)	(9.63)	(27.52)	(29.55)	(8.58)	(30.28)
M_h	115.94	212.82	115.52	111.30	209.84	201.66	113.26	193.66
	(15.75)	(45.03)	(14.16)	(24.56)	(42.87)	(45.16)	(16.55)	(48.86)
M_{tb}	99.42	115.44	104.48	96.10	146.92	115.50	98.60	132.94
	(20.16)	(28.11)	(21.21)	(14.88)	(36.05)	(27.39)	(16.83)	(28.85)
M_{bh}	154.98	144.80	476.94	128.80	503.66	140.10	477.08	482.20
	(56.75)	(63.99)	(64.83)	(42.43)	(71.08)	(42.31)	(63.14)	(61.85)
M_{th}	107.44	178.26	103.80	101.34	174.28	170.10	100.06	165.50
	(13.70)	(37.87)	(12.78)	(10.80)	(30.32)	(31.92)	(9.99)	(38.98)
M_{tbh}	102.38	140.48	102.82	95.80	160.42	136.30	98.46	148.98
	(16.39)	(28.81)	(15.07)	(11.71)	(32.52)	(23.27)	(11.28)	(30.90)

Table 2.29: Means and Standard Errors of AIC: Experiment 6

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	83.49	67.58	116.87	87.35	82.59	70.99	119.49	83.77
	(7.02)	(6.72)	(10.83)	(6.07)	(8.91)	(7.27)	(11.65)	(10.35)
M_b	85.59	66.89	94.76	89.48	76.92	70.67	95.99	77.76
	(7.14)	(6.74)	(5.91)	(6.16)	(6.05)	(7.27)	(6.38)	(7.23)
M_t	91.81	73.67	89.64	95.61	73.91	75.99	89.97	75.82
	(6.53)	(6.15)	(7.86)	(6.29)	(6.11)	(6.51)	(7.52)	(6.88)
M_h	87.52	71.63	121.25	91.24	86.75	74.86	123.70	87.92
	(6.58)	(6.58)	(10.86)	(5.71)	(8.93)	(7.08)	(11.58)	(10.37)
M_{tb}	93.40	74.73	91.12	97.23	75.20	76.98	91.75	77.07
	(6.55)	(6.75)	(8.10)	(6.44)	(6.13)	(6.64)	(7.83)	(6.98)
M_{bh}	91.30	72.86	101.65	95.21	83.83	76.34	102.67	84.55
	(6.28)	(6.44)	(6.11)	(5.78)	(6.12)	(6.82)	(6.32)	(7.01)
M_{th}	103.07	85.52	101.18	106.94	86.11	87.35	101.53	87.52
	(5.25)	(5.38)	(7.45)	(5.35)	(5.51)	(5.43)	(6.66)	(5.98)
M_{tbh}	104.86	87.09	102.93	108.58	87.74	88.90	103.43	89.22
	(5.42)	(5.73)	(7.64)	(5.49)	(5.63)	(5.52)	(6.88)	(6.03)

Table 2.30: Means and Standard Errors of Posterior Median: Experiment 7 (True $N = 500$)

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	500.02	632.96	504.46	493.00	622.18	619.58	498.60	607.34
	(8.01)	(17.07)	(7.40)	(8.59)	(13.93)	(19.36)	(5.80)	(15.77)
M_b	500.50	500.74	612.28	494.44	609.74	497.12	596.68	593.98
	(9.61)	(9.78)	(29.27)	(12.38)	(28.64)	(9.47)	(31.38)	(33.60)
M_t	499.70	625.28	501.12	492.62	619.98	612.68	495.62	605.12
	(7.86)	(16.22)	(7.30)	(8.65)	(13.92)	(18.50)	(5.88)	(15.61)
M_h	508.90	652.18	512.46	506.00	638.68	641.40	508.16	623.80
	(12.45)	(20.53)	(9.73)	(21.91)	(16.70)	(23.91)	(11.22)	(18.50)
M_{tb}	525.52	514.44	503.06	511.28	515.48	520.50	501.58	508.76
	(79.83)	(45.25)	(13.92)	(38.54)	(17.28)	(60.48)	(14.94)	(14.29)
M_{bh}	512.42	515.88	634.62	514.76	632.78	514.68	619.10	616.46
	(12.47)	(16.05)	(33.31)	(26.52)	(33.23)	(24.08)	(35.74)	(37.18)
M_{th}	504.78	648.86	506.54	499.68	645.58	631.40	501.76	629.12
	(8.51)	(32.60)	(7.35)	(9.88)	(40.56)	(34.10)	(8.13)	(30.35)
M_{tbh}	517.28	518.62	503.50	503.30	518.66	520.28	501.04	510.92
	(45.06)	(38.08)	(13.32)	(22.47)	(19.26)	(33.70)	(13.93)	(14.49)

Table 2.31: Means and Standard Errors of AIC: Experiment 7

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	170.01	204.71	245.42	170.63	165.96	199.02	239.02	164.93
	(6.43)	(15.35)	(16.39)	(7.77)	(9.58)	(14.53)	(18.48)	(9.51)
M_b	172.08	149.81	194.56	172.36	167.77	151.04	190.79	166.34
	(6.37)	(6.86)	(11.38)	(7.52)	(8.99)	(8.76)	(12.82)	(9.82)
M_t	178.44	167.63	175.53	178.39	172.21	165.72	176.69	170.84
	(6.63)	(9.21)	(6.14)	(7.07)	(8.36)	(10.57)	(7.60)	(8.72)
M_h	174.29	209.09	249.84	174.33	170.48	203.26	242.93	169.35
	(6.35)	(15.32)	(16.35)	(7.60)	(9.57)	(14.37)	(18.47)	(9.44)
M_{tb}	180.93	157.94	177.55	180.36	159.08	159.23	178.35	158.94
	(6.46)	(6.56)	(5.96)	(7.16)	(6.05)	(8.60)	(7.46)	(6.97)
M_{bh}	178.41	156.01	201.15	177.91	174.35	156.80	197.07	172.84
	(6.21)	(6.83)	(11.47)	(7.33)	(8.95)	(8.03)	(12.94)	(9.75)
M_{th}	189.97	177.38	186.86	188.51	182.04	176.10	187.02	180.83
	(6.05)	(8.11)	(6.09)	(6.32)	(6.85)	(8.17)	(7.37)	(7.40)
M_{tbh}	192.30	169.97	188.80	190.29	170.97	170.99	188.85	170.49
	(6.13)	(5.83)	(5.88)	(6.20)	(5.38)	(7.54)	(7.43)	(5.91)

Table 2.32: Means and Standard Errors of Posterior Median: Experiment 8 (True $N = 100$)

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	102.84	188.56	108.24	70.64	188.68	88.04	73.84	91.84
	(13.87)	(36.51)	(11.83)	(5.82)	(49.75)	(12.34)	(6.78)	(14.17)
M_b	125.54	145.20	348.34	77.32	329.22	77.04	164.50	166.26
	(47.77)	(75.08)	(103.45)	(15.83)	(111.01)	(13.70)	(87.41)	(93.10)
M_t	99.34	166.34	103.34	70.00	167.38	85.36	72.70	89.22
	(12.32)	(27.70)	(10.70)	(5.69)	(37.12)	(11.30)	(6.51)	(12.93)
M_h	114.26	219.08	121.72	90.64	222.18	131.20	96.12	116.98
	(16.94)	(44.94)	(16.56)	(20.56)	(60.03)	(52.84)	(26.61)	(26.86)
M_{tb}	101.60	119.18	103.82	83.54	129.26	87.30	91.56	99.84
	(19.37)	(30.37)	(20.53)	(12.87)	(39.79)	(16.16)	(17.75)	(22.59)
M_{bh}	139.28	153.48	326.16	94.24	303.56	92.42	187.54	176.98
	(44.65)	(65.48)	(94.68)	(22.55)	(93.06)	(32.12)	(78.11)	(77.90)
M_{th}	106.48	181.16	110.16	78.08	187.22	101.08	81.54	100.88
	(14.29)	(31.54)	(12.67)	(8.07)	(42.03)	(17.59)	(9.38)	(14.78)
M_{tbh}	103.74	142.16	105.34	77.74	151.38	92.66	82.02	96.10
	(17.14)	(27.26)	(15.22)	(9.00)	(38.09)	(16.39)	(9.85)	(17.70)

Table 2.33: Means and Standard Errors of AIC: Experiment 8

Model Fit	Data Generating Process (DGP)							
	M_0	M_b	M_t	M_h	M_{tb}	M_{bh}	M_{th}	M_{tbh}
M_0	85.38	66.87	93.09	106.88	68.21	97.85	113.31	93.26
	(8.32)	(7.11)	(8.86)	(6.86)	(8.11)	(10.35)	(8.79)	(9.53)
M_b	87.62	66.04	88.08	109.13	69.09	96.07	106.53	94.04
	(8.19)	(6.20)	(6.82)	(6.81)	(7.77)	(9.47)	(6.22)	(9.29)
M_t	93.01	72.64	91.58	114.85	74.43	102.54	113.54	100.72
	(7.47)	(6.38)	(6.49)	(6.95)	(7.40)	(9.84)	(8.00)	(9.25)
M_h	89.65	70.96	97.27	104.10	72.22	97.15	110.93	93.88
	(8.29)	(7.00)	(8.63)	(5.60)	(8.06)	(7.89)	(7.28)	(6.86)
M_{tb}	94.83	73.57	92.94	116.88	75.08	103.93	114.13	102.14
	(7.59)	(6.34)	(6.75)	(6.85)	(6.92)	(9.68)	(6.93)	(9.40)
M_{bh}	93.39	72.42	94.57	107.91	75.66	96.22	107.56	96.52
	(7.89)	(5.94)	(6.69)	(5.44)	(7.54)	(6.71)	(5.40)	(6.48)
M_{th}	104.70	84.76	103.41	118.76	85.82	108.26	116.86	108.07
	(6.91)	(5.59)	(6.16)	(4.13)	(6.79)	(7.03)	(5.73)	(6.54)
M_{tbh}	106.53	86.29	104.89	120.63	87.25	109.44	118.85	109.56
	(7.01)	(5.57)	(6.30)	(4.40)	(6.66)	(6.88)	(5.87)	(6.52)

Chapter 3

Matching Errors in Closed Population Capture Recapture Models

3.1 Introduction to Matching Errors

Matching errors in closed population capture-recapture models generally arise in one of two ways. In one way, two or more distinct individuals are not differentiable from each other through the matching process. Thus distinct individuals are mis-recorded as multiple captures of the same individual. Adopting terminology from Mills, Citta, Lair, Schwartz, and Tallmon (2000), we call this a *shadow-match* error, and we generally refer to a set of non-differentiable population members as a group. When this situation occurs, a researcher is only able to detect presence of at least one of the shadow members of a

particular group. Thus, the recorded capture history reflects the behavior of the group, not the capture history of an individual. In the second way, a recaptured individual's identifying information is corrupted to the extent that it cannot be matched to that individual's identifying information from any prior capture. Thus the same individual is recorded as two distinct individuals. We call this type of error an *identity-match error*.

Identity-match errors have been studied when lists of people are used as captures, and where a person's record on a particular list may be corrupted, such as by typographical errors, omissions or the use of different versions of a name, e.g., Jon and John, or Smith, Smyth, and Smythe. Some contributions to the literature have explored the effects of identity-match errors (Seber, Huakau, and Simmons (2000), Lee, Seber, Holden, and Huakau (2001), Ding and Fienberg (1994)). The first two articles focus on tag-loss from animal populations as a method to account for matching errors in the likelihood. The third article dealt with both types of errors, using information about error probabilities from a large-scale census review. They also make some key simplifying assumptions about possible matching errors for building their models.

Although identity-match errors are common, we focus exclusive on problems where shadow-match errors are problematic. Shadow-match errors arise when DNA fingerprints are used to identify animals in studies to estimate the size of an animal population. Such sampling techniques substitute the capture of DNA in hair or faecal samples, or other sources of DNA, for live animal captures. Such non-invasive sampling allows hard-to-capture animal populations to be sampled more thoroughly, potentially resulting in more accurate population estimates (Mills et al. (2000)). However, the use of DNA fingerprints to identify

animals creates a potential for shadow-match errors. Closely related, but distinct, animals may not be uniquely identified through the DNA fingerprinting process. This type of error is shown by Mills et al. (2000) to lead to estimates of population size that are biased low.

For a study with two capture periods, and a population in which each non-unique genotype is shared by exactly two population members, Mills et al. (2000) showed that

$$E(n_1) = up_1 + s(2p_1 - p_1^2),$$

$$E(m_2) = up_1p_2 + s(2p_1 - p_1^2)(2p_2 - p_2^2),$$

$$E(n_2) = E(m_2) + up_2(1 - p_1) + s(2p_2 - p_2^2)(1 - p_1)^2.$$

where

$n_1 = $ number of animals captured during capture period 1;

$n_2 = $ number of animals captured during capture period 2;

$m_2 = $ number of animals captured during both capture periods 1 and 2;

$u = $ number of uniquely identifiable genotypes in the population;

$s = $ number genotypes shared by at least two population members;

$p_j = $ probability an animal is detected in capture j, for $j = 1, 2$.

Mills et al. (2000) investigated the bias in the Lincoln-Petersen estimator in the case where each shared DNA profile is shared by exactly two members, and the case where each shared DNA profile is shared by exactly four members. They found that when each

shared DNA profile is shared by four population members, the bias in estimating N is more severe than when each shared profile is shared by two population members. Note that their analysis assumes that shared genotypes are shared by exactly the same number of population members, i.e., all shared genotypes are shared by exactly two (or four) population members. This assumption, though not realistic, leads to useful insights. A more realistic assumption for modeling purposes is to assume that the numbers of animals sharing a DNA profile varies from group to group.

As more DNA loci are used to verify the uniqueness of each genotype, the probability of shadow-match errors decreases (Mills et al. (2000)). But, while increasing the number of loci checked is favorable for error reduction, it is also increases costs. This is true especially when the source of the animal DNA is limited (Taberlet, Waits, and Luikart (1999)). The tradeoff between accuracy and expense means that errors will never be completely eliminated in all studies, thus indicating the need for statistical methods that explicitly account for shadow-match errors under realistic assumptions.

3.2 Modeling Shadow-Match Errors

We propose a probability model for shadow matches that allows for groups of varying size. The average group size is controlled by one additional parameter thus leading to a parsimonious statistical model for analyzing data with shadow-match errors. We start with some notation and definitions. Our motivation comes from DNA matching problems and we adopt terminology accordingly.

For a population of size N, let $A_1, A_2, ..., A_N$ represent the distinct animals in the

population, and let $D_1, D_2, ..., D_N$ denote their DNA profiles. Let T denote the number of recognizably unique DNA profiles in the population. Denote the set of uniquely identifiable DNA profiles by $(D_1^*, D_2^*, ..., D_T^*)$. Clearly, $T \leq N$ and the set of profiles $\{D_1^*, D_2^*, ..., D_T^*\}$ is a subset of $\{D_1, D_2, ..., D_N\}$. Let M_i denote the number of animals sharing DNA profile D_i^*, $i = 1, \ldots, T$. Note that $N = M_1 + \cdots + M_T$.

In the pioneering work by (Mills et al. (2000)) $M_1 = \cdots = M_T = 2$ (or 4) for the sake of mathematical convenience. We accommodate unequal M_i with the assumption that given T, $M_1, \ldots, M_T$ are independent and identically distributed with common probability mass function $g_M(m)$, i.e., $\Pr(M_i = m) = g_M(m)$. Although we derive some results for arbitrary mass functions, we specialize to the case wherein M_i has a shifted-Poisson distribution, i.e.,

$$\Pr(M_i = m) = \frac{e^{-\lambda}\lambda^{(m-1)}}{(m-1)!}; \qquad m = 1, 2, \ldots. \tag{3.1}$$

With the shifted-Poisson model the single additional parameter λ controls the extent of shadow matching. The average group size is $E(M_i) = 1 + \lambda$ and the variance of group size is λ. Thus $\lambda = 0$ corresponds to the case of no shadow matches (all $M_i = 1$) and the prevalence of shadow matches increases monotonically with λ.

Note that the model assumptions are such that T is regarded as fixed (non-random), or equivalently, that the model is formulated conditioned on the value of T; but that $M_1, \ldots, M_T$ are random. The latter assumption implies that total population size $N = M_1 + \cdots + M_T$ is also random.

We assume a two-capture, capture-recapture study with capture probabilities, p_j

= probability that an animal is captured during capture period j, for $j = 1, 2$. Let $q_j = 1 - p_j$. Because we allow capture probabilities to vary by capture period only, this model is similar to the time-effects model M_t introduced in Chapter 2.

For DNA profile D_i^*, define

- $y_{i00} = 1$ if D_i^* is not captured during either capture period;

- $y_{i01} = 1$ if D_i^* is captured during capture period 2 but not capture period 1;

- $y_{i10} = 1$ if D_i^* is captured during capture period 1 but not capture period 2;

- $y_{i11} = 1$ if D_i^* is captured during both capture periods.

The following sums hold the numbers of each type of capture history:

$$Z_0 = \sum_{i=1}^{T} y_{i00} = \text{number of DNA profiles not captured in the study;} \quad (3.2)$$

$$Z_1 = \sum_{i=1}^{T} y_{i01} = \text{number of DNA profiles with capture history (0,1);} \quad (3.3)$$

$$Z_2 = \sum_{i=1}^{T} y_{i10} = \text{number of DNA profiles with capture history (1,0);} \quad (3.4)$$

$$Z_3 = \sum_{i=1}^{T} y_{i11} = \text{number of DNA profiles with capture history (1,1).} \quad (3.5)$$

Finally define the number of DNA profiles observed at least once during the study,

$$S = Z_1 + Z_2 + Z_3. \quad (3.6)$$

In Section **??** we show that the probability mass function of the observed counts (Z_1, Z_2, Z_3)

is given by

$$f_Z(z_1, z_2, z_3 | q_1, q_2, T, g_M) = \frac{T!}{z_1! z_2! z_3! (T - S)!} \theta_{00}^{T-S} \theta_{01}^{z_1} \theta_{10}^{z_2} \theta_{11}^{z_3} \qquad (3.7)$$

where

$$
\begin{aligned}
\theta_{00} &= E((q_1 q_2)^{M_i}), \\[2mm]
\theta_{01} &= E(q_1^{M_i}(1 - q_2^{M_i})), \\[2mm]
\theta_{10} &= E(q_2^{M_i}(1 - q_1^{M_i})), \\[2mm]
\theta_{11} &= E((1 - q_1^{M_i})(1 - q_2^{M_i})).
\end{aligned}
\qquad (3.8)
$$

Note that $\theta_{00} + \theta_{01} + \theta_{10} + \theta_{11} = 1$. Also note that each θ_{ij} can be determined from the probability generating function of M_i. Under the shifted-Poisson assumption (3.1),

$$
\begin{aligned}
\theta_{00} &= q_1 q_2 e^{\lambda(q_1 q_2 - 1)}, \\[2mm]
\theta_{01} &= q_1 e^{\lambda(q_1 - 1)} - q_1 q_2 e^{\lambda(q_1 q_2 - 1)}, \\[2mm]
\theta_{10} &= q_2 e^{\lambda(q_2 - 1)} - q_1 q_2 e^{\lambda(q_1 q_2 - 1)}, \\[2mm]
\theta_{11} &= 1 - q_1 e^{\lambda(q_1 - 1)} - q_2 e^{\lambda(q_2 - 1)} + q_1 q_2 e^{\lambda(q_1 q_2 - 1)}.
\end{aligned}
\qquad (3.9)
$$

In this case we emphasize the dependence on λ of the probability mass function in (3.7) by writing $f_Z(z_1, z_2, z_3 | q_1, q_2, T, \lambda)$. Thus given observed counts (Z_1, Z_2, Z_3) we obtain the

likelihood

$$L(q_1, q_2, T, \lambda | Z_1, Z_2, Z_3) = f_Z(Z_1, Z_2, Z_3 | q_1, q_2, T, \lambda).$$

Estimation of λ with only two capture periods is problematic. Thus we assume that λ is known or independently estimated and consider only estimation of q_1, q_2, T. Our assumption that λ is known is comparable to the assumptions in Mills et al. (2000) that average group size is known or that the probability of a DNA match is known. In our model the average group size is $E(M_i) = 1 + \lambda$. Thus if we want to study the effect of matching errors for a specified average group size (e.g., 2 or 4 as done in Mills et al. (2000)), we need only choose λ accordingly. Alternatively, suppose that π is the probability that a given animal has at least one DNA match in the population. Mills et al. (2000) argue that this probability can be roughly estimated from available data for certain populations, and we take this as a starting point for our argument. In our model an animal has a match if its group size exceeds one. According to our model this event occurs with probability $\Pr(M_i > 1) = 1 - e^{\lambda}$, and π and λ are related via $\pi = 1 - e^{\lambda}$, or equivalently $\lambda = -\ln(1 - \pi)$. Thus, while our assumption that λ is known is somewhat limiting, there are defensible strategies for approximating λ in applications. Regardless, sensitivity to specification of λ is of concern and we investigate this problem in the simulation study reported on in Section 3.4.

3.3 Model Inference

3.3.1 Parameter Estimation

Under the assumption that λ is known there are three unknown parameters in the likelihood (3.10), q_1, q_2, and T, that can be estimated either by maximum likelihood or method-of-moments. The likelihood is nonlinear in the parameters and thus for maximum likelihood estimation numerical optimization is required. The particular numerical algorithm we used in the simulations and example in Sections 3.4 and 3.5 is described in detail in Section 3.6.3.

The moments required for method-of-moments estimation are readily derived using moment generating function identities for the Poisson distribution. For example,

$$
\begin{aligned}
E(Z_1) &= E\left(\sum_{i=1}^{T} Y_{i01}\right) = \sum_{i=1}^{T} E\left(Y_{i01}\right) \\
&= \sum_{i=1}^{T} \Pr(Y_{i01} = 1) \\
&= \sum_{i=1}^{T} \sum_{m_i=1}^{\infty} \Pr(Y_{i01} = 1, M_i = m_i) \\
&= \sum_{i=1}^{T} \sum_{m_i=1}^{\infty} \Pr(Y_{i01} = 1 | M_i = m_i) g_M(m_i) \\
&= \sum_{i=1}^{T} \sum_{m_i=1}^{\infty} q_1^{m_i}(1 - q_2^{m_i}) g_M(m_i) \\
&= \sum_{i=1}^{T} E(e^{t_1 M} - e^{t_3 M}), \text{ where } t_1 = \ln q_1 \text{ and } t_3 = \ln q_1 q_2 \\
&= \sum_{i=1}^{T} q_1 e^{\lambda(q_1-1)} - q_1 q_2 e^{\lambda(q_1 q_2-1)} \\
&= T\left[q_1 e^{\lambda(q_1-1)} - q_1 q_2 e^{\lambda(q_1 q_2-1)}\right].
\end{aligned}
\tag{3.10}
$$

Similar derivations result in

$$E(Z_2) = T[q_2 e^{\lambda(q_2-1)} - q_1 q_2 e^{\lambda(q_1 q_2-1)}], \tag{3.11}$$

$$E(Z_3) = T[1 - q_1 e^{\lambda(q_1-1)} - q_2 e^{\lambda(q_2-1)} + q_1 q_2 e^{\lambda(q_1 q_2-1)}]. \tag{3.12}$$

The resulting moment equations are

$$Z_1 = T(q_1 e^{\lambda(q_1-1)} - q_1 q_2 e^{\lambda(q_1 q_2-1)}); \tag{3.13}$$

$$Z_2 = T(q_2 e^{\lambda(q_2-1)} - q_1 q_2 e^{\lambda(q_1 q_2-1)}); \tag{3.14}$$

$$Z_3 = T(1 - q_1 e^{\lambda(q_1-1)} - q_2 e^{\lambda(q_2-1)} + q_1 q_2 e^{\lambda(q_1 q_2-1)}). \tag{3.15}$$

From Equation 3.13 we obtain

$$q_1 q_2 e^{\lambda(q_1 q_2-1)} = q_1 e^{\lambda(q_1-1)} - \frac{Z_1}{T} \tag{3.16}$$

Substituting (3.16) into (3.15) we obtain:

$$q_2 e^{\lambda(q_2-1)} = 1 - \frac{Z_1 + Z_3}{T} \tag{3.17}$$

Lastly, substitute (3.17) into (3.14) giving

$$q_1 e^{\lambda(q_1-1)} = 1 - \frac{Z_2 + Z_3}{T}. \tag{3.18}$$

Then the method-of-moments estimators $\tilde{q}_1$ solves Equation 3.18, $\tilde{q}_2$ solves Equation 3.17,

with solutions as a function of T. Lastly, $\tilde{T}$ solves Equation 3.16 for T. Solutions are found numerically. Our simulation results (Section 3.4) showed little difference between the maximum likelihood and method-of-moments approaches.

3.3.2 Predicting Population Size

Point Prediction

Primary interest lies not in the parameter estimates per se, but rather in the population size N. In our model N is a random variable, not a parameter. Thus we formulate the inference problem for N as a prediction problem. First we determine the form of the best predictor of N given the observed counts (Z_1, Z_2, Z_3) assuming parameters are known. Then we obtain an estimated predictor, $\widetilde{N}$, by replacing the unknown parameters with their estimators (maximum likelihood or method-of- moments).

The best predictor of N based on the observed counts is $E(N|Z_1, Z_2, Z_3)$. Under the assumptions of our model the best predictor of N given Z_1, Z_2, Z_3 is $\tilde{N} = E[M_1 + ... + M_T|Z_1, Z_2, Z_3]$. In Section 3.6.2 we derive the required expression, showing that

$$\tilde{N} = (T - S)(1 + \lambda q_1 q_2) + Z_1 E_1 + Z_2 E_2 + Z_3 E_3 \tag{3.19}$$

where

$$E_1 = 1 + \frac{\lambda q_1(\phi_1 - q_2\phi_3)}{\phi_1 - \phi_3},$$
$$E_2 = 1 + \frac{\lambda q_2(\phi_2 - q_1\phi_3)}{\phi_2 - \phi_3},$$

Stage	Parameters	Observed Data	Output Quantities Estimated	Predicted
Population	T, q_1, q_2	Z_1, Z_2, Z_3	$\hat{T}, \hat{q}_1, \hat{q}_2$	$\tilde{N}$
Stage 1 Bootstrap	$\hat{T}, \hat{q}_1, \hat{q}_2$	$Z_{1,B}, Z_{2,B}, Z_{3,B}$	$\hat{T}_B, \hat{q}_{1,B}, \hat{q}_{2,B}$	$\tilde{N}_B$
Stage 2 Bootstrap	$\hat{T}_B, \hat{q}_{1,B}, \hat{q}_{2,B}$	$Z_{1,B2}, Z_{2,B2}, Z_{3,B2}$	$\hat{T}_{B2}, \hat{q}_{1,B2}, \hat{q}_{2,B2}$	$\tilde{N}_{B2}$

Table 3.1: Notation for the two-stage bootstrap procedure. Two-stage bootstrap parameters, observed data, and estimated and predicted quantities for each stage.

$$E_3 = \frac{1 + \lambda - \phi_1(1 + \lambda q_1) - \phi_2(1 + \lambda q_2) + \phi_3(1 + \lambda q_1 q_2)}{1 - \phi_1 - \phi_2 + \phi_3},$$

$$\phi_1 = q_1 e^{\lambda(q_1 - 1)},$$

$$\phi_2 = q_2 e^{\lambda(q_2 - 1)},$$

$$\phi_3 = q_1 q_2 e^{\lambda(q_1 q_2 - 1)}.$$

An estimated predictor of N is obtained by replacing the unknown parameters in (3.19) by their estimates (maximum likelihood or method-of-moments).

Interval Prediction

Because of the complexity of the point predictor of N in Equation (3.19), analytic derivation of a prediction interval is problematic. We develop a two-stage parametric bootstrap prediction interval. The two-stage bootstrap method is designed to address variation due to both estimation of the model parameters and to the variation in N for a given set of parameters (recall that N is random in our model).

Table 3.1 lists the notation for the two-stage bootstrap procedure. The estimates $\hat{q}_1, \hat{q}_2, \hat{T}$ derived from the observed data are regarded as population quantities in the first stage of the bootstrap. In the first-stage bootstrap sampling, five-thousand bootstrap data sets $(Z_{1,B}, Z_{2,B}, Z_{3,B})$ from a bootstrap population with parameters $\hat{q}_1, \hat{q}_2, \hat{T}$, and λ are simulated. For each stage-one data set, estimates $\hat{q}_{1,B}, \hat{q}_{2,B}, \hat{T}_B$, and the predictor of population size $\tilde{N}_B$ are calculated. Next generate one stage-two bootstrap data set for each of the five-thousand stage-one bootstrap data sets. That is, generate one set of counts $(Z_{1,B2}, Z_{2,B2}, Z_{3,B2})$ from a bootstrap population with parameters $\hat{q}_{1,B}, \hat{q}_{2,B}, \hat{T}_B$, and λ for each of the five-thousand stage-one data sets. From the values of $Z_{1,B2}, Z_{2,B2}, Z_{3,B2}$ we estimate $\hat{q}_{1,B2}, \hat{q}_{2,B2}, \hat{T}_{B2}$, and calculate the estimated predictor $\tilde{N}_{B2}$, resulting in five-thousand stage-two values of $\hat{q}_{1,B2}, \hat{q}_{2,B2}, \hat{T}_{B2}$, and $\tilde{N}_{B2}$. Note that in our bootstrap procedures, the prediction of N at each stage is not a bootstrap parameter in the next stage. Once the bootstrap parameters T, q_1, q_2, and λ are fixed, the population size N is a randomly generated quantity according to the model assumptions that $M_1, ..., M_T$ are random.

For the first stage of the bootstrap procedure, define $M_{1,B}, ..., M_{\hat{T},B}$ as the randomly generated group sizes for groups $1, 2, ..., \hat{T}$. Similarly, for the second stage of the bootstrap procedure, define $M_{1,B2}, ..., M_{\hat{T}_B,B2}$ as the randomly generated group sizes for groups $1, 2, ..., \hat{T}_B$. Then, denote $N_B = M_{1,B} + M_{2,B} + ... + M_{\hat{T},B}$ and $N_{B2} = M_{1,B2} + M_{2,B2} + ... + M_{\hat{T}_B,B2}$. The values N_B and N_{B2} represent the simulated true population size for stages one and two of the bootstrap procedure, respectively.

The intent of the two-stage bootstrap is to approximate the joint distribution of $(N, \tilde{N})$ by the joint bootstrap distribution of $(N_{B2}, \tilde{N}_{B2})$. In Section 3.4.4 we report results

from a simulation study designed to compare three strategies for calculating prediction intervals for N based on the five-thousand pairs $(N_{B2}, \tilde{N}_{B2})$.

1. Marginal Data Method: Using all five-thousand combinations of $(N_{B2}, \tilde{N}_{B2})$, generate a 95% prediction interval for N by computing the 2.5th percentile and 97.5th percentile of the five-thousand observed values of N_{B2}.

2. Conditional Data Method 1: Again compute the 2.5th percentile and 97.5th percentile of N_{B2}, but analyze only those stage-two bootstrap data sets where $\tilde{N}_{B2}$ is suitably close to $\tilde{N}$ from the original data set. For this method, we analyze the stage-two bootstrap data sets where $|N_{B2} - \tilde{N}| \leq 1$.

3. Conditional Data Method 2: Compute the 2.5th percentile and 97.5th percentile of N_{B2} for the one-thousand stage-two data sets where $\tilde{N}_{B2}$ is closest to $\tilde{N}$. The initial reason for examining this method was that Conditional Data Method 1 often produced a small number of points (50 - 200, typically) where $\tilde{N}$ was sufficiently close to $\tilde{N}_{B2}$, which may be too small a number of data points to produce accurate prediction intervals.

3.4   Simulation Study

3.4.1   Data Generation Process

To study the performance of $\tilde{N}$ as a predictor of N, we generated random data sets with counts Z_1, Z_2, and Z_3 for a capture-recapture experiment with $k = 2$ capture periods. The factors in the experiment are:

- $T = 25, 50, 100;$

- $q_1 = 0.1, 0.5, 0.9;$

- $q_2 = 0.1, 0.5, 0.9;$

- $\lambda = 0, 0.1, 0.2, 0.3, 0.4, 0.5.$

We generated one-hundred data sets under each unique factor level combination, of which there are 162. Although we assume λ to be known (as discussed in Section 3.2 , realistically we accept that the assumed value of λ may differ from the true value. To assess performance of $\tilde{N}$ when λ is mis-specified, we fit the model using several values of λ to each data set. Thus we distinguish between the value of λ used to generate the data set and the value of λ used to fit the data via the likelihood function of Section 3.2. We use λ_t for the value of λ used to generate the data set, and to λ_{fit} as the value of λ assumed in the model.

For each of the 16200 data sets, we computed $\tilde{N}$ for each data set using $\lambda_{fit} = 0, 0.1, 0.2, 0.3, 0.4, 0.5$. Thus, we analyze each data set from the above 162 factor combinations under six values of λ_{fit}, giving a total of 97200 analyses for the simulation.

For each data set, we estimated q_1, q_2, and T using both maximum likelihood and method-of-moments. For maximum likelihood analysis, we used SAS version 8.2 software to numerically maximize the likelihood function given in Section 3.2. For method-of-moments analysis, we also used the SAS software package to numerically solve the method-of-moments equations given in Equations 3.16, 3.17, and 3.18. Computational methods for our analyses are described in Section 3.6.3.

It should be noted that we did not analyze any data set for which the number of recaptures in the data turned out to be zero. This phenomenon occurred most commonly when $q_1 = q_2 = 0.9$, corresponding to small capture probabilities, and when T was twenty-five. This requirement means that our results are conditional on the presence of at least one recapture during the study. Practically speaking, a researcher would not end a capture-recapture experiment with zero recaptures. So we believe discarding these data sets is consistent with practice. These discarded data sets are not part of the 16200 data sets described above. For each factor level combination, we generated random data sets until we obtained one-hundred data sets with at least one recapture.

3.4.2   Simulation Results

The simulation experiment investigated the effects of five factors: T, q_1, q_2, λ_t, and λ_{fit}. For each simulated data set, although T, q_1, and q_2 are fixed, the true value of N is random. One way of determining the performance of our predictor is to measure

$$B_N = \frac{\tilde{N} - N}{T},$$

which is similar to a relative bias measurement. Dividing by T allows us to determine whether any increase in the magnitude of the bias $\tilde{N} - N$ when T increases is simply due to the increased value of T (and correspondingly N) that occurs when the population has a larger number of distinct DNA profiles. Intuitively, the average value of B_N should be zero when $\lambda_t = \lambda_{fit}$, greater than zero when $\lambda_{fit} > \lambda_t$, and less than zero when $\lambda_{fit} < \lambda_t$. A preliminary analysis of the results with response variable B_N indicated that a full analysis of

variance of a model with explanatory variables T, q_1, q_2, λ_t, and λ_{fit} and all their interactions explained about 50% of the variation in the simulation ($R^2 = 0.502$). The model with main and interaction effects for T, q_1, q_2 and the quantitative covariate $\lambda_{diff} = \lambda_{fit} - \lambda_t$ explained about 49% of the variation ($R^2 = 0.486$) in the simulation. The simple linear model with explanatory predictor λ_{diff} explains about 35% of the variation ($R^2 = 0.347$) in the simulation results. Thus, much of the explainable variation is explained by λ_{diff} and the three factors T, q_1, and q_2, and we focus attention on these in the discussion of the results. We expected that $\tilde{N}$ would be approximately unbiased when $\lambda_t = \lambda_{fit}$, that $\tilde{N} > N$ on average when $\lambda_{fit} > \lambda_t$, and that $\tilde{N} < N$ on average when $\lambda_{fit} < \lambda_t$, and these expectations were confirmed by the simulation results.

Figures 3.1 through 3.6 show the relative bias of $\tilde{N}$ when maximum likelihood was used to estimate q_1, q_2, and T. Results for method-of-moments estimation are very similar to the maximum likelihood results, and are omitted. Each plotted point represents the average value of B_N for the one-hundred data sets analyzed under a specific combination of T, q_1, q_2, λ_t, and λ_{fit}.

Results when $T = 25$:

Focusing first on the graphs in Figure 3.1, we see that when capture probabilities are very high for both capture periods (corresponding to $q_1 = q_2 = 0.1$), when $\lambda_t = 0$, the best fit, on average, is provided by the model where $\lambda_{fit} = \lambda_t$. All the other models produce values of $\bar{B}_N$ that are greater than zero, which is consistent with expectations. That is, picking $\lambda_{fit} = \lambda_t = 0$ leads to $\bar{B}_N \sim 0$, picking $\lambda_{fit} > \lambda_t$ leads to $\bar{B}_N > 0$. When the value of $\lambda_t = \lambda_{fit} = 0.1$, then $\bar{B}_N$ is approximately zero. We see $\bar{B}_N < 0$ when $\lambda_{fit} = 0$,

Figure 3.1: Simulation Results for $T = 25$ when $q_1 = q_2$
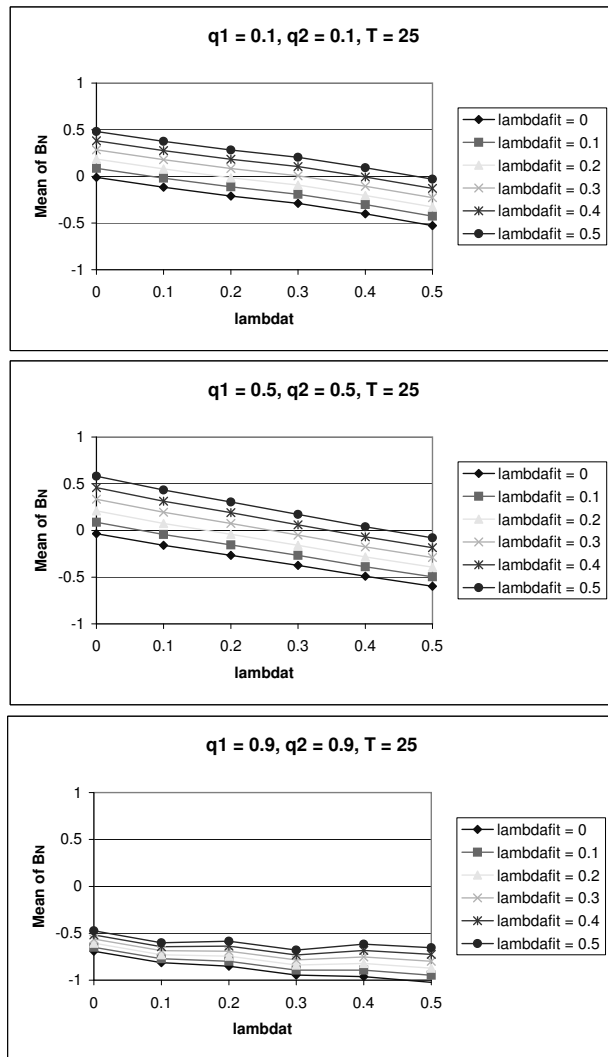
and $\bar{B}_N > 0$ when $\lambda_{fit} > \lambda_t$. Similar results occur in Figure 3.1 at all values of λ_t from 0

through 0.5. Thus, from the first graph in Figure 3.1, we conclude that $\tilde{N}$ is approximately

unbiased when $\lambda_{fit} = \lambda_t$, positively biased when $\lambda_{fit} > \lambda_t$, and negatively biased when

$\lambda_{fit} < \lambda_t$. Furthermore, relative biases as large as 0.5 and as small as -0.5 may occur

when λ_{fit} is mis-specified. Referring again to Figure 3.1, when capture probabilities are

moderately high ($q_1 = q_2 = 0.5$), the results are very similar to those when $q_1 = q_2 = 0.1$ in

Figure 3.2: Simulation Results for $T = 25$ when $q_1 \neq q_2$
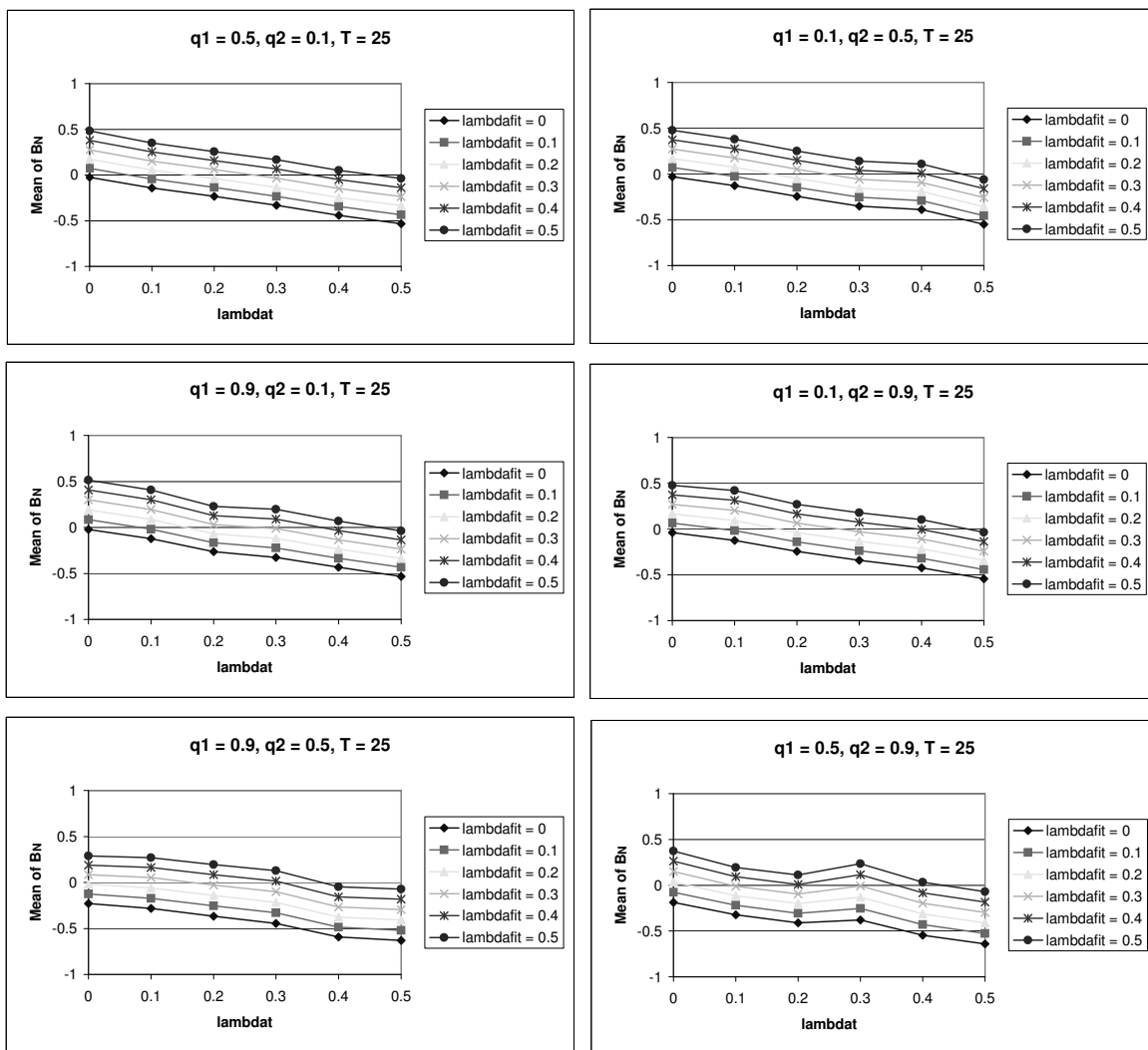
Figure 3.1.

A striking difference is noticed in Figure 3.1 when $q_1 = q_2 = 0.9$. We see negative bias in predicting N ($\bar{B}_N < 0$) for all levels of λ_t and λ_{fit}. This factor combination represents reasonably small capture probabilities and a small number of groups in the population. We focus on the simple case where $\lambda_{fit} = \lambda_t = 0$ to explain why $\bar{B}_N$ is negative in this picture. Recall that we only included data sets in our analysis where $Z_3 > 0$

(requiring at least one recapture). $E(Z_3)$ can be computed via Equation 3.12. In this case, we can find $E(Z_3) = 25 \times 0.1 \times 0.1 = 0.25$. Thus our data sets had $Z_3 > E(Z_3)$. Using the Lincoln-Petersen estimator from Equation 1.1 as a rough approximation of $\tilde{N}$ in the case where $\lambda_{fit} = 0$, and using the likelihood function in Equation 3.10 to compute $f(Z_1, Z_2, Z_3 | q_1 = q_2 = 0.9; T = 25, \lambda = 0)$ we compute the conditional expectation of the LP estimator of N conditional upon $Z_3 > 0$. We omit the detailed calculation, but the conditional expectation of the LP estimator of N, given $Z_3 > 0$, is approximately ten for this factor combination (well below the true value of twenty-five). This negative bias occurs for all data sets where $T = 25$ and $q_1 = q_2 = 0.9$, indicating that $\tilde{N}$ may be significantly negatively biased when T is small and q_1 and q_2 are both large. The severity of this problem is due somewhat to the small value of T in this combination of factors. A larger value of T would lead to $E(Z_3) = (T \times 0.1 \times 0.1) \geq 1$, meaning one or more recaptures would occur in a higher percentage of the generated data sets. Thus, the effect of conditioning on $Z_3 \geq 1$ would have smaller impact on the conditional expectation of the LP estimator given $Z_3 > 0$.

Figure 3.2 shows simulation results for data sets generated when $T = 25$ and when $q_1 \neq q_2$. The graphs placed next to each other are probabilistically symmetric, in that the second graph shows results for design points where the values of q_1 and q_2 are simply switched. We expect the simulation results to be very similar in these cases.

The results when $q_1 = 0.1$ and $q_2 = 0.5$ (or vice versa) follow a pattern close to those when $q_1 = q_2 = 0.1$ or $q_1 = q_2 = 0.5$, in that $\bar{B}_N$ is close to zero when $\lambda_{fit} = \lambda_t$. Over-fitting λ causes positive bias, and under-fitting λ causes negative bias, on average. The magnitude of the relative bias (measured by $\bar{B}_N$) when $T = 25$, $q_1 = 0.9$ and $q_2 = 0.5$

Figure 3.3: Simulation Results for $T = 50$ when $q_1 = q_2$
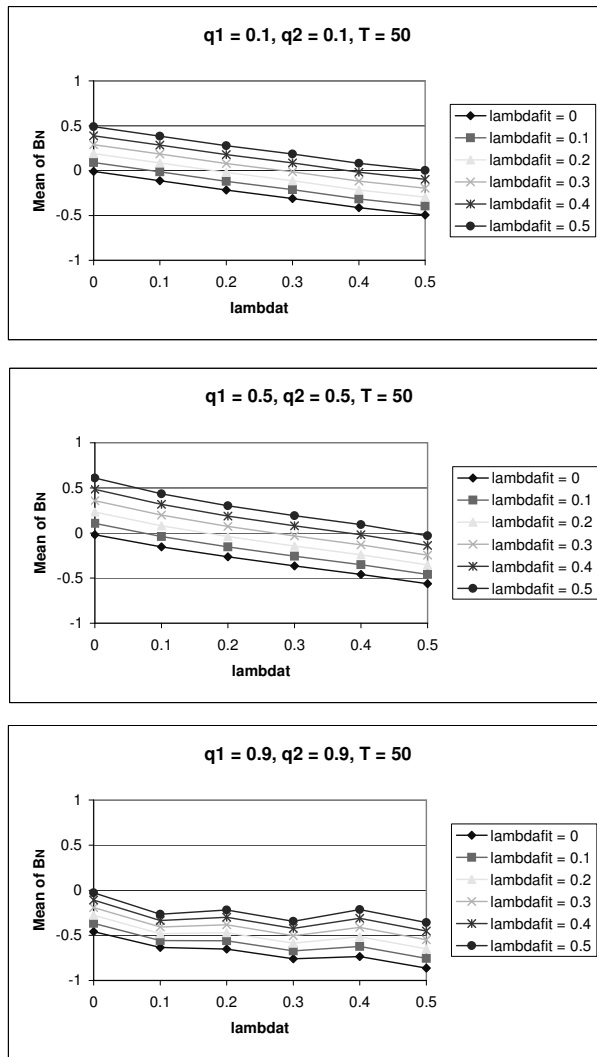
(and in the symmetric case) is not as large as when $T = 25, q_1 = 0.1$, and $q_2 = 0.5$. When

$q_1 = 0.9$ and $q_2 = 0.5$, the average value of B_N is negative and closer to -0.25 when

$\lambda_{fit} = \lambda_t = 0$. This is due, again, to the effect of ignoring data sets where the number of

recaptures is zero. This effect is reduced in data sets where $\lambda_t > 0$. That is, $\bar{B}_N$ is closer

to zero when $\lambda_t = \lambda_{fit} > 0$.

Figure 3.4: Simulation Results for $T = 50$ when $q_1 \neq q_2$
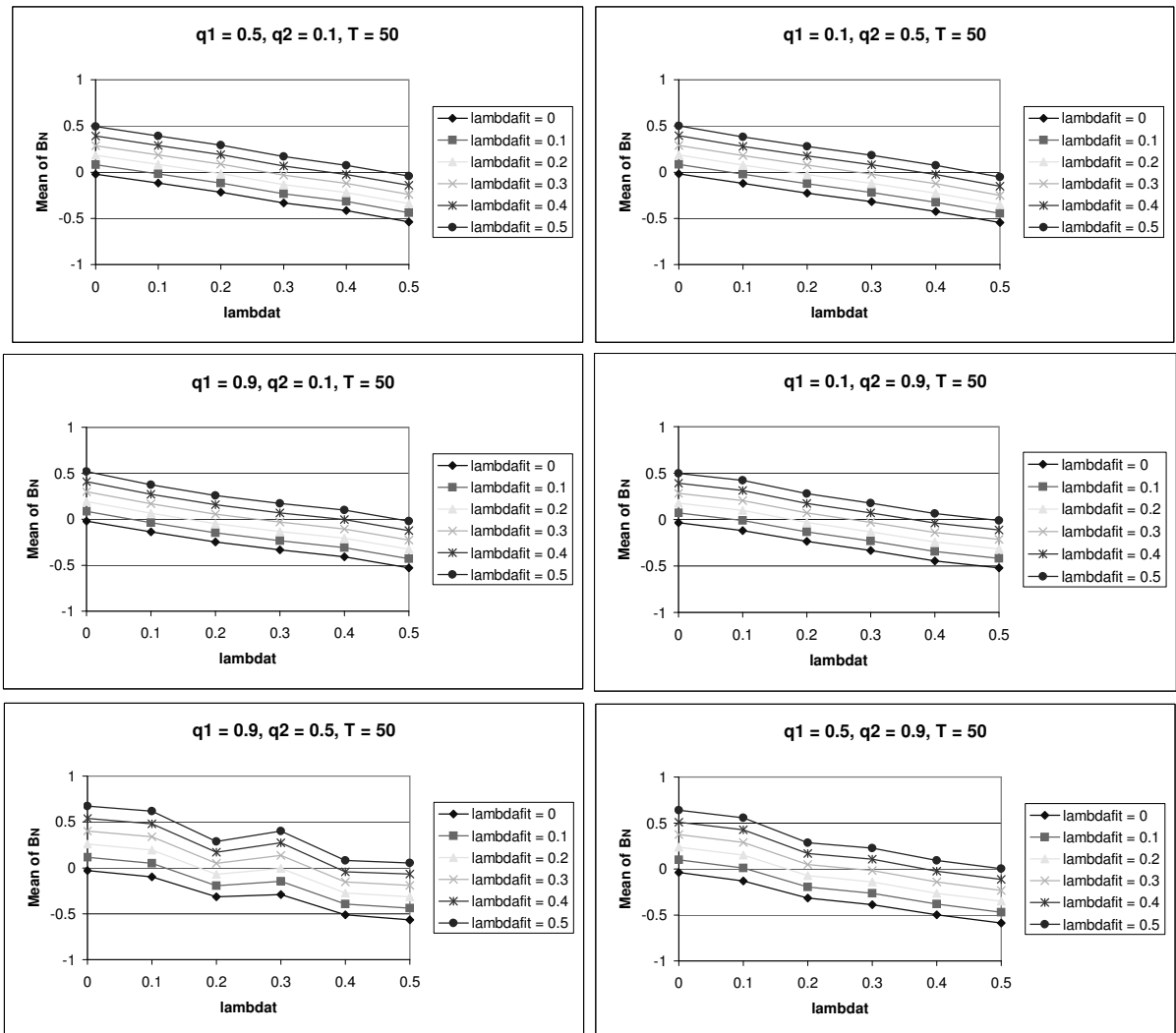
Results when $T = 50$:

For Figure 3.3, we see pattern similar to Figure 3.1. The first two graphs in Figure 3.3 again show that choosing a value of $\lambda_{fit} = \lambda_t$ gives a value of $\bar{B}_N$ near zero, that choosing $\lambda_{fit} > \lambda_t$ yields a value of $\bar{B}_N$ greater than zero, and that choosing $\lambda_{fit} < \lambda_t$ gives a value of $\bar{B}_N$ less than zero. One difference between Figures 3.1 and 3.3 is in the case where $q_1 = q_2 = 0.9$. The negative bias effect due to the requirement that $Z_3 > 0$ is reduced from

Figure 3.5: Simulation Results for $T = 100$ when $q_1 = q_2$
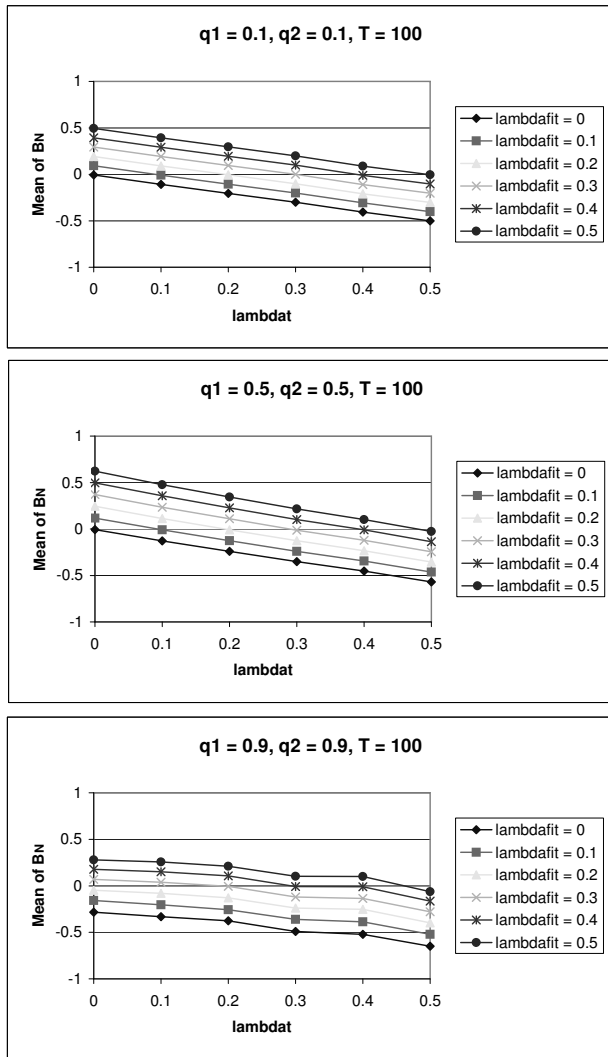
the effect in Figure 3.1. However, the negative bias in B_N is still present, and suggests that $\tilde{N}$ would be negatively biased in real data sets with small capture probabilities and a small to moderate number of group sizes.

In Figure 3.4, we see a pattern similar to that of Figure 3.2. It is noteworthy that the performance of $\tilde{N}$ is better than when $T = 25$, in the case where $q_1 = 0.9$ and $q_2 = 0.5$ (or vice versa). In this case when $\lambda_{fit} = \lambda_t$, $\bar{B}_N$ is closer to zero than when $T = 25$,

Figure 3.6: Simulation Results for $T = 100$ when $q_1 \neq q_2$
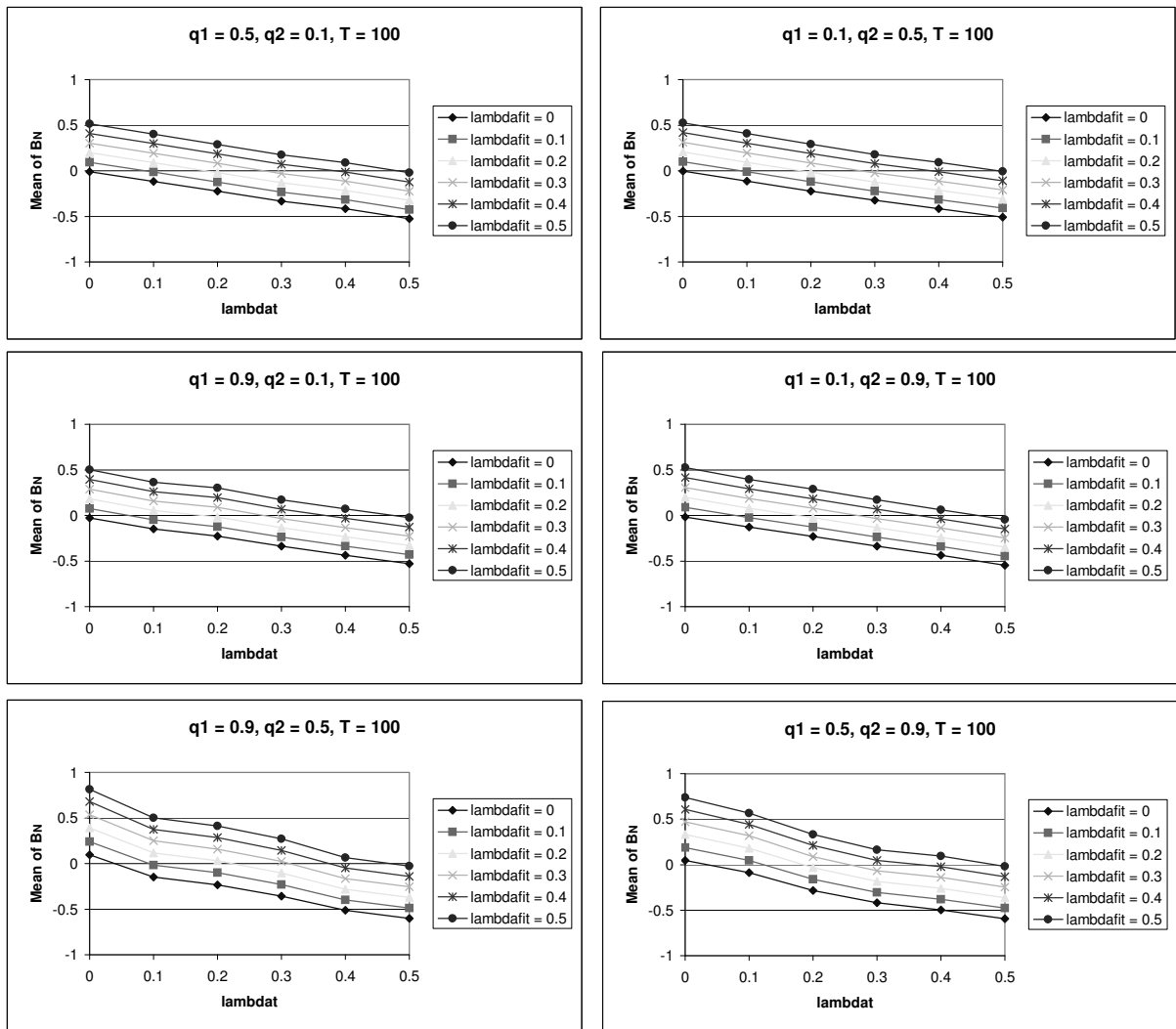
suggesting that the effects of including only data sets with at least one recapture in our analysis is not as significant when $T = 50$ and at least one capture period has a moderately high capture probability.

Results when $T = 100$:

Figures 3.5 and 3.6 show similar results as the previous graphs. Here, however, we ignored a smaller number of data sets due to the higher probability of the presence of at

least one recapture in most data sets. The effect of ignoring data sets with zero recaptures is still evident in Figure 3.5 when $q_1 = q_2 = 0.9$, but its effect is reduced with the larger value of T. Again, choosing $\lambda_t = \lambda_{fit}$ means $\bar{B}_N$ is near zero. Choosing λ_{fit} greater than (less than) λ_t gives values of $\bar{B}_N$ that are greater than (less than) zero.

3.4.3 Sensitivity of Prediction of N to Choice of λ

We refer back to the issue of assuming λ is known, which was raised in Section 3.2. Assuming that $\lambda_{fit} \neq \lambda_t$, we examine how close λ_{fit} needs to be to λ_t to improve prediction of N over classic capture-recapture methods (such as a standard Model M_t).

In our simulation results, a significant percentage of the explainable variation of B_N in can be explained by a simple linear model with one explanatory variable, the quantitative variable $\lambda_{diff} = \lambda_{fit} - \lambda_t$. Fitting a standard Model M_t is roughly the same as choosing $\lambda_{fit} = 0$. Accounting for shadow-match errors means choosing $\lambda_{fit} > 0$. Denote approach S_1 as choosing $\lambda_{fit} = 0$. Denote approach S_2 as choosing $\lambda_{fit} > 0$. To determine when S_2 is a better approach than S_1, recall that our ANOVA simple linear model in Section 3.4.2 implies that $B_N \sim \lambda_{diff}$. Thus $|B_N| \sim |\lambda_{diff}|$. In approach S_1 this means $|B_N| \sim \lambda_t$. In approach S_2 this means $|B_N| \sim |\lambda_{fit} - \lambda_t|$. Therefore, approach S_2 gives a smaller value of $|B_N|$ than S_1 when

$$|\lambda_{fit} - \lambda_t| \leq \lambda_t \tag{3.20}$$

which occurs when $-\lambda_t \leq (\lambda_{fit} - \lambda_t) \leq \lambda_t$, or when

$$0 \leq \lambda_{fit} \leq 2\lambda_t. \tag{3.21}$$

Therefore, when λ is mis-specified, our method should have smaller relative bias for the population size than classic capture-recapture methods when the chosen value of λ is no larger than twice the true value of λ. This encourages a bit of caution in choosing λ. Underestimating the true value of λ still improves performance over that of classic capture-recapture methods. Choosing too large a value of λ may lead create a situation where the bias in predicting N is more severe than if the shadow-match error problem had been ignored altogether. For this reason, a reasonable recommendation is to be conservative in choosing λ if a precise value of λ is not available.

3.4.4   Bootstrap Results

We tested the bootstrap method for four parameter combinations. Each parameter combination we examined had $q_1 = q_2 = 0.5$ and $\lambda_t = 0.3$. The difference among the parameter combinations was the different levels of T. We chose $T = 50, 100, 120, 150$ in the four parameter combinations. We chose a relatively high value of λ_t to ensure that our method works well even when λ_t is not close to zero. In each case, for each of the five-thousand stage-two bootstrap data sets, we computed prediction intervals for N under both the Marginal Data Method and the two Conditional Data Methods described in Section 3.3.2.

Results of the bootstrap methods are listed in Table 3.2. For each of the methods, we examined performance in terms of the observed coverage probability of a nominal 95% prediction interval for N. We also noted the average length of the prediction intervals. In Table 3.2, the coverage probabilities for all methods are below the nominal 95% level when $T < 120$. For $T \geq 120$, the Marginal Data Method produces a coverage probability that is

| | Marginal Data Method | | Conditional Data Method 1 | | Conditional Data Method 2 | |
T	Coverage %	Avg Length	Coverage %	Avg Length	Coverage %	Avg Length
50	82%	33.3	81%	25.4	80%	25.9
100	87%	47.4	80%	37.7	80%	37.4
120	97%	52.3	92%	41.5	93%	42
150	97%	45.4	92%	45.4	91%	46.5

Table 3.2: Bootstrap Prediction Interval Results: Simulated Coverage Probabilities and Average Interval Lengths of Nominal 95% Prediction Intervals for N (with Std. Error of 3% when the true coverage level is 90%

above the nominal 95% level (although the difference is not statistically significant), and also has the largest interval length. The average interval length for the Marginal Data Method is typically about 10 units longer than the typical interval length for either of the Conditional Data Methods. Another point to notice is that the performance of the Conditional Data Method One does not appear to be due to the relatively small number of data points analyzed under this method. This is apparent due to the similarity in performance between Conditional Data Methods One and Two, and because Method Two uses one-thousand data points to compute its prediction intervals.

Because our methods were analyzed for one-hundred randomly generated data sets, the coverage probabilities listed in Table 3.2 estimate the true coverage probability for the listed parameters. For one-hundred data sets, the standard error of our estimates at the $T = 150$ level (97% and 92%) and would be $\sqrt{(.97)(.03)/100} = 0.017$ and $\sqrt{(.92)(.02)/100} = 0.027$, respectively. This means that both estimates are within approximately one standard error from their nominal coverage levels. Based on these results, we believe both the Marginal and Conditional bootstrap methods are effective in generating prediction intervals for N when $T > 100$. More extensive simulation study is needed to

| | | Interval Estimate | |
Model	$\hat{N}$	Lower	Upper
Model M_t	57	37	120

Table 3.3: Confidence Interval for N for Bear Data of Woods, Paetkau, et al.

confirm these conjectures.

3.5   Example

In the article from Woods, et al. (1999) the authors discuss a data set with captures of bears over four capture periods. Because our method is currently developed only for two capture periods, we compared our predictor of population size $\tilde{N}$ with the results of Model M_t in Program CAPTURE, using only the data for the first two capture periods of the data set. The data are $Z_1 = 11, Z_2 = 12, Z_3 = 4$. The models fit to the data were:

- The Model M_t estimator, $\hat{N}$, from Program CAPTURE

- Our predictor, $\tilde{N}$, at $\lambda = 0, 0.1, 0.2, 0.3, 0.4, 0.5$

Results of the analysis are given in Tables 3.3 and 3.4.   Comparing Tables 3.3 and 3.4, when $\lambda_{fit} = 0$, $\tilde{N}$ is equal to $\hat{N}$ obtained from Model M_t from Program CAP-TURE. Our prediction interval is wider using the Marginal Data Bootstrap Method, and is comparable to Model M_t when using the Conditional Data Bootstrap Method. Using our model, as λ_{fit} increases, the value of $\tilde{N}$ increases, and the lower and upper bounds for the prediction interval for N both increase. The width of the prediction interval increases as well. This is to be expected, because higher values of λ_{fit} imply increases in both the mean and variance of group sizes M_i.

Matching Error Models:	$\tilde{N}$	Interval Estimates			
		Marginal Data Method		Conditional Data Method 2	
		Lower	Upper	Lower	Upper
with $\lambda_{fit} = 0$	56.64	30	167	38	121
with $\lambda_{fit} = 0.1$	65.43	33	182.5	42	132
with $\lambda_{fit} = 0.2$	74.09	37	216	47.5	164
with $\lambda_{fit} = 0.3$	82.52	40	235	49	181
with $\lambda_{fit} = 0.4$	90.71	44	274	57	211.5
with $\lambda_{fit} = 0.5$	98.63	48	302	61	227

Table 3.4: Prediction intervals for N for Bear Data of Woods, Paetkau, et al.

The main rationale for using our method, as opposed to Model M_t, (or other models) from Program CAPTURE is that our model should provide a more accurate prediction of N than a standard Model M_t in the presence of the shadow-match errors we have discussed. For each level of λ_{fit} we measured, $\tilde{N}$ was still within the confidence bands given for Model M_t. However, at levels of $\lambda_{fit} > 0$, the coverage probability of the Model M_t confidence bands would be reduced from the nominal 95% level. Suitable levels of λ_{fit} should be chosen based on the approach discussed in Section 3.2.

3.6   Technical Details

In this section we derive the likelihood function and predictor $\tilde{N}$ of population size N, for the shadow match errors capture-recapture model. We also include descriptions of the numerical methods techniques used to maximize the likelihood function and to solve the method-of-moments equations in Chapter 3.

3.6.1 Probability Model and Likelihood

Theorem 3.1: Let D_i^* for $i = 1, 2, ..., T$ represent the T uniquely identifiable DNA

profiles within a population. Let M_i represent the number of animals sharing profile D_i^* for

$i = 1, 2, ..., T$. Let p_j represent the probability that an animal is captured during capture

period j, for $j = 1, 2$, and let $q_j = 1 - p_j$. For DNA profile D_i^*, define $y_{i00} = 1$ if D_i^* is not

captured during either capture period; $y_{i01} = 1$ if D_i^* is captured during capture period 2

but not capture period 1; $y_{i10} = 1$ if D_i^* is captured during capture period 1 but not capture

period 2; and $y_{i11} = 1$ if D_i^* is captured during both capture periods. Then denote the counts

of DNA profiles with each capture history as $Z_0 = \sum_{i=1}^{T} y_{i00}$; $Z_1 = \sum_{i=1}^{T} y_{i01}$; $Z_2 = \sum_{i=1}^{T} y_{i10}$;

and $Z_3 = \sum_{i=1}^{T} y_{i11}$. Finally, denote the number of DNA profiles captured at least once

during the study as $S = Z_1 + Z_2 + Z_3$. Then, the probability mass function of (Z_1, Z_2, Z_3)

is

$$f_Z(z_1, z_2, z_3) = \frac{T!}{z_1! z_2! z_3! (T - S)!} \theta_{00}^{T-S} \theta_{01}^{z_1} \theta_{10}^{z_2} \theta_{11}^{z_3},$$

where

$$\theta_{00} = E\left((q_1 q_2)^{M_i}\right),$$

$$\theta_{01} = E\left(q_1^{M_i}\left(1 - q_2^{M_i}\right)\right),$$

$$\theta_{10} = E\left(\left(1 - q_1^{M_i}\right) q_2^{M_i}\right),$$

$$\theta_{11} = E\left(\left(1 - q_1^{M_i}\right)\left(1 - q_2^{M_i}\right)\right).$$

Proof: For DNA profile D_i^*, define :

$Y(i, 1) = 1$ if D_i^* is captured during neither capture period,

$Y(i, 2) = 1$ if D_i^* is captured only during capture period 2,

$Y(i, 3) = 1$ if D_i^* is captured only during capture period 1,

$Y(i, 4) = 1$ if D_i^* is captured during both capture periods.

Note that for D_i^*, $Y(i, 1) + Y(i, 2) + Y(i, 3) + Y(i, 4) = 1$.

Define the matrix $\boldsymbol{Y}$ with dimensions $T \times 4$ as

$$
\boldsymbol{Y} =
\begin{pmatrix}
y(1, 1) & y(1, 2) & y(1, 3) & y(1, 4) \\
y(2, 1) & y(2, 2) & y(2, 3) & y(2, 4) \\
. & . & . & . \\
. & . & . & . \\
. & . & . & . \\
y(T, 1) & y(T, 2) & y(T, 3) & y(T, 4)
\end{pmatrix}
$$

There are 4^T possible values of $\boldsymbol{Y}$, because $\boldsymbol{Y}(i, .) = (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)$, or $(0, 0, 0, 1)$ for all i. Define $\boldsymbol{y_r}$ with dimensions $T \times 4$ for $r = 1, 2, ..., 4^T$ as an observed outcome from the 4^T possible outcomes of $\boldsymbol{Y}$. Then the support of $\boldsymbol{Y}$ can be written as $\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_{4^T}}$. Furthermore, define the matrix $\boldsymbol{P}$ with entry $p(i, l)$ in its ith row and lth column, such that $\boldsymbol{P} = \mathrm{E}(\boldsymbol{Y})$. That is, because all entries in $\boldsymbol{Y}$ are zero or one, $p(i, l)$ is the probability that $y(i, l) = 1$. Later we describe the structure of $\boldsymbol{P}$ in more detail.

Assuming animal captures are independent of each other, conditioning on the group sizes M_i, the distribution of $\boldsymbol{Y}$ given $M_1, M_2, ..., M_T$ is

$$f(\mathbf{y_r}|\mathbf{P} = \mathbf{p}, M_1, ..., M_T) = Pr(\mathbf{Y_{T \times 4}} = \mathbf{y_{r_{T \times 4}}}|\mathbf{P}, M_1, ..., M_T) =$$

$$\prod_{i=1}^{T} p(i,1)^{y_r(i,1)} p(i,2)^{y_r(i,2)} p(i,3)^{y_r(i,3)} p(i,4)^{y_r(i,4)}$$

for $\mathbf{y_r}$ in $(\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_{4T}})$.

Define

$$Z_l = \sum_{i=1}^{T} y(i, l+1)$$

for $l = 0, 1, 2, 3$, and define the event A as

$A = \{\sum_{i=1}^{T} Y(i, 2) = z_1, \sum_{i=1}^{T} Y(i, 3) = z_2, \sum_{i=1}^{T} Y(i, 4) = z_3\}$. Then

$$Pr(Z_1 = z_1, Z_2 = z_2, Z_3 = z_3) = \sum_{r=1}^{4^T} f(\mathbf{y_r}|\mathbf{P}, M_1, ..., M_T) I(A). \tag{3.22}$$

The number of non-zero values of the indicator variable in the above expression can be determined as follows:

1. z_1 values of 1 can be placed into any of the T rows of the matrix, in column two. There are $\binom{T}{z_1}$ ways to set these values of 1 within column two.

2. z_2 values of 1 can be placed into any of the remaining $T - Z_1$ rows, in column three. There are $\binom{T - z_1}{z_2}$ ways to set these values of 1 within column three.

3. z_3 values of 1can be placed into any of the remaining $T - z_1 - z_2$ rows, in column four. There are $\binom{T - z_1 - z_2}{z_3}$ ways to set these values of 1 within column four.

4. All remaining rows of the matrix must have a 1 placed in column one. There is only one way to do this.

Thus, the number of non-zero indicator variables is given by

$$\binom{T}{z_1, z_2, z_3} = \binom{T}{z_1}\binom{T-z_1}{z_2}\binom{T-z_1-z_2}{z_3}$$
$$= \frac{T!}{z_1! z_2! z_3! (T-z_1-z_2-z_3)!}.$$

Now we explain the structure of $\mathbf{P}$. Let $p_j =$ probability that an animal is captured during capture period j, for $j = 1, 2$, and $q_j = 1 - p_j$. For D_i^* to not be captured in a particular capture period, all of the M_i animals sharing profile D_i^* must not be captured. Hence, the probability that D_i^* is not captured in capture period j is $q_j^{M_i}$. Otherwise, if at least one of the M_i animals is captured, the profile is considered captured. Corresponding to our definitions of $y(i, 1), ..., y(i, 4)$, we define the probabilities of each capture history (conditional on M_i), as

$$p(i, 1) = \Pr(D_i^* \text{ has capture history } (0,0)|M_i = m_i),$$

$$p(i, 2) = \Pr(D_i^* \text{ has capture history } (0,1)|M_i = m_i),$$

$$p(i, 3) = \Pr(D_i^* \text{ has capture history } (1,0)|M_i = m_i),$$

$$p(i, 4) = \Pr(D_i^* \text{ has capture history } (1,1)|M_i = m_i).$$

Consequently,

$$p(i, 1) = q_1^{m_i} q_2^{m_i},$$

$$p(i, 2) = q_1^{m_i} (1 - q_2^{m_i}),$$

$$p(i, 3) = (1 - q_1^{m_i}) q_2^{m_i},$$

$$p(i,4) \quad = \quad (1 - q_1^{m_i})(1 - q_2^{m_i}). \tag{3.23}$$

Defining the probability of the capture history of D_i^* as

$$a_i(q_1, q_2, m_i, y_r(i, .)) = \prod_{l=1}^{4} p(i, l)^{y_r(i,l)},$$

we write the probability in Equation 3.22 as

$$\Pr(Z_1 = z_1, Z_2 = z_2, Z_3 = z_3 | \mathbf{P}, M_1, ... M_T) =$$

$$\sum_{r=1}^{4^T} \prod_{i=1}^{T} a_i(q_1, q_2, m_i, y_r(i, .)) I(A).$$

However, this probability is conditional on the unknown group sizes M_i for $i = 1, 2, ..., T$. Under the assumption that $M_1, M_2, ..., M_T$ are independent and identically distributed with common density $g_M(m)$, we obtain

$$\Pr(Z_1 = z_1, Z_2 = z_2, Z_3 = z_3) =$$

$$\sum_{M_1=1}^{\infty} .... \sum_{M_T=1}^{\infty} \sum_{r=1}^{4^T} \prod_{i=1}^{T} a_i(q_1, q_2, m_i, \mathbf{y_r(i, .)}) I(A) \prod_{l=1}^{T} g_M(m_l)$$

With some rearranging and factoring, we obtain

$\Pr\left(Z_1 = z_1, Z_2 = z_2, Z_3 = z_3\right)$

$$
\begin{aligned}
&= \sum_{M_1=1}^{\infty} \dots \sum_{M_T=1}^{\infty} \sum_{r=1}^{4^T} \prod_{i=1}^{T} a_i\left(q_1, q_2, m_i, \mathbf{y_r}(i,.)\right) I\left(A\right) \prod_{l=1}^{T} g_M\left(m_l\right) \\
&= \sum_{r=1}^{4^T} \sum_{M_1=1}^{\infty} \dots \sum_{M_T=1}^{\infty} \prod_{i=1}^{T} a_i\left(q_1, q_2, m_i, \mathbf{y_r}(i,.)\right) I\left(A\right) \prod_{l=1}^{T} g_M\left(m_l\right) \\
&= \sum_{r=1}^{4^T} \sum_{M_1=1}^{\infty} a_1\left(q_1, q_2, m_1, y_r(1,.)\right) g_M\left(m_1\right) \\
&\qquad \qquad \dots \sum_{M_T=1}^{\infty} a_T\left(q_1, q_2, m_T, y_r(T,.)\right) g_M\left(m_T\right) I\left(A\right) \\
&= \sum_{r=1}^{4^T} \prod_{i=1}^{T} \left[\sum_{m_i=1}^{\infty} a_i\left(q_1, q_2, m_i, \mathbf{y_r}(i,.)\right) g_M\left(m_i\right) \right] I\left(A\right) \\
&= \sum_{r=1}^{4^T} \prod_{i=1}^{T} E\left[a_i\left(q_1, q_2, m_i, \mathbf{y_r}(i,.)\right)\right] I(A).
\end{aligned}
$$

Now, notice that

$$
a_i\left(q_1, q_2, m_i, \mathbf{y_r}(i,.)\right) = \prod_{l=1}^{4} p(i,l)^{y_r(i,l)} = \sum_{l=1}^{4} y_r(i,l) p(i,l)
$$

because all values of $y_r(.,.)$ are zero or one, and $\sum_{l=1}^{4} y_r(i,l) = 1$.

Define

$$
\theta_{00} = E\left(\left(q_1 q_2\right)^{M_i}\right),
$$

$$
\theta_{01} = E\left(q_1^{M_i}\left(1-q_2\right)^{M_i}\right),
$$

$$
\theta_{10} = E\left(\left(1-q_1\right)^{M_i} q_2^{M_i}\right),
$$

$$
\theta_{11} = E\left(\left(1-q_1\right)^{M_i}\left(1-q_2\right)^{M_i}\right).
$$

Then,

$$E\left[a_i\left(q_1, q_2, m_i, \mathbf{y_r}(i, .)\right)\right] = y_r(i, 1)\theta_{00} + y_r(i, 2)\theta_{01} + y_r(i, 3)\theta_{10} + y_r(i, 4)\theta_{11}.$$

And then we have (unconditionally),

$$\Pr\left(Z_1 = z_1, Z_2 = z_2, Z_3 = z_3\right) = \sum_{r=1}^{4^T} \prod_{i=1}^{T} \theta_{00}^{y_r(i,1)} \theta_{01}^{y_r(i,2)} \theta_{10}^{y_r(i,3)} \theta_{11}^{y_r(i,4)} I(A).$$

Because all values of $y_r(., .) = 0$ or 1, each term in the product above is one of $(\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})$,

and thus we obtain

$$\begin{aligned}
\Pr\left(Z_1 = z_1, Z_2 = z_2, Z_3 = z_3\right) &= \sum_{r=1}^{4^T} \theta_{00}^{\sum_{i=1}^{T} y_r(i,1)} \theta_{01}^{\sum_{i=1}^{T} y_r(i,2)} \theta_{10}^{\sum_{i=1}^{T} y_r(i,3)} \theta_{11}^{\sum_{i=1}^{T} y_r(i,4)} I(A) \\
&= \sum_{r=1}^{4^T} \theta_{00}^{z_0} \theta_{01}^{z_1} \theta_{10}^{z_2} \theta_{11}^{z_3} I(A). \qquad (3.24)
\end{aligned}$$

Now Z_0 is unobserved, because Z_0 represents the number of DNA profiles with capture

history $(0, 0)$. But, because

$$y_r(i, 1) = 1 - y_r(i, 2) - y_r(i, 3) - y_r(i, 4)$$

for all i, summing over i gives

$$Z_0 = T - Z_1 - Z_2 - Z_3.$$

Denote $S = Z_1 + Z_2 + Z_3$ and note that that $T - S = Z_0$. The θ values do not depend on

r, so we can factor them out front giving

$$\Pr\left(Z_1 = z_1, Z_2 = z_2, Z_3 = z_3\right) = \theta_{00}^{T-S} \theta_{01}^{Z_1} \theta_{10}^{Z_2} \theta_{11}^{Z_3} \sum_{r=1}^{4^T} I(A) =$$

$$\binom{N}{Z_1, Z_2, Z_3} \theta_{00}^{T-S} \theta_{01}^{Z_1} \theta_{10}^{Z_2} \theta_{11}^{Z_3}, \tag{3.25}$$

and the result is established.

3.6.2 Prediction of N

Theorem 3.2: Suppose that $M_1, ..., M_T$ are independent and identically distributed with common shifted-Poisson density

$$\Pr(M_i = m) = \frac{e^{-\lambda} \lambda^{(m-1)}}{(m-1)!}, \qquad m = 1, 2, ...\infty.$$

Then the best predictor of N given $Z_1 = z_1, Z_2 = z_2, Z_3 = z_3$ is

$$\hat{N} = E[M_1 + ... + M_T | Z_1 = z_1, Z_2 = z_2, Z_3 = z_3] =$$

$$(T - S)(1 + \lambda q_1 q_2) + z_1 E_1 + z_2 E_2 + z_3 E_3$$

where

$$E_1 = 1 + \frac{\lambda q_1 (\phi_1 - q_2 \phi_3)}{\phi_1 - \phi_3},$$

$$E_2 = 1 + \frac{\lambda q_2 (\phi_2 - q_1 \phi_3)}{\phi_2 - \phi_3},$$

$$E_3 = \frac{1 + \lambda - \phi_1 (1 + \lambda q_1) - \phi_2 (1 + \lambda q_2) + \phi_3 (1 + \lambda q_1 q_2)}{1 - \phi_1 - \phi_2 + \phi_3},$$

$$\phi_1 = q_1 e^{\lambda(q_1 - 1)},$$

$$\phi_2 = q_2 e^{\lambda(q_2 - 1)},$$

$$\phi_3 = q_1 q_2 e^{\lambda(q_1 q_2 - 1)}.$$

Proof: In the proof of Theorem 3.1 we denoted

$$f(\mathbf{y_r}|\mathbf{P}, M_1, ..., M_T) = Pr(\mathbf{Y} = \mathbf{y_r}|\mathbf{P}, M_1, ..., M_T)$$

$$= \prod_{i=1}^{T} p(i,1)^{y_r(i,1)} p(i,2)^{y_r(i,2)} p(i,3)^{y_r(i,3)} p(i,4)^{y_r(i,4)}$$

for $\mathbf{y_r}$ in $(\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_{4^T}})$. Recall that $p(i,1), ..., p(i,4)$ were defined in Equations 3.23.

For D_i^*, we can write

$$f(\mathbf{y_r}(i,.), m_i|\mathbf{P}) = f(\mathbf{y_r}(i,.)|M_i, \mathbf{P})g_M(m_i) =$$

$$p(i,1)^{y_r(i,1)} p(i,2)^{y_r(i,2)} p(i,3)^{y_r(i,3)} p(i,4)^{y_r(i,4)} g_M(m) .$$

Then, we sum the above equation over the range of M_i to obtain

$$f(\mathbf{y_r}(i,.)|\mathbf{P}) = \sum_{m_i=1}^{\infty} f(\mathbf{y_r}(i,.), m_i|\mathbf{P}) = \theta_{00}^{y_r(i,1)} \theta_{01}^{y_r(i,2)} \theta_{10}^{y_r(i,3)} \theta_{11}^{y_r(i,4)}.$$

Now, by definition of conditional probability,

$$f(m_i|\mathbf{y_r}(i,.), \mathbf{P}) = \frac{f(\mathbf{y_r}(i,.), m_i|\mathbf{P})}{f(\mathbf{y_r}(i,.)|\mathbf{P})}.$$

From there, we can compute

$$E(M_i|\mathbf{y_r}(i,.)) = \frac{\sum_{m_i=1}^{\infty} m_i f(\mathbf{y_r}(i,.), m_i|\mathbf{P})}{f(\mathbf{y_r}(i,.)|\mathbf{P})} = \frac{\sum_{m_i=1}^{\infty} m_i f(\mathbf{y_r}(i,.), m_i|\mathbf{P})}{\sum_{m_i=1}^{\infty} f(\mathbf{y_r}(i,.), m_i|\mathbf{P})} =$$

$$\frac{\sum_{m_i=1}^{\infty} m_i a_i(q_1, q_2, m_i, \mathbf{y_r}(i,.)) g_M(m_i)}{\sum_{m_i=1}^{\infty} a_i(q_1, q_2, m_i, \mathbf{y_r}(i,.)) g_M(m_i)} = \frac{E(M_i a_i(q_1, q_2, m_i, \mathbf{y_r}(i,.)))}{E(a_i(q_1, q_2, m_i, \mathbf{y_r}(i,.)))}.$$

So, we have shown

$$E(M_i|\mathbf{y_r}(i,.)) = \frac{E(M_i a_i(q_1, q_2, m_i, \mathbf{y_r}(i,.)))}{E(a_i(q_1, q_2, m_i, \mathbf{y_r}(i,.)))}. \tag{3.26}$$

For any D_i^*, only one of $y_r(i,1), ..., y_r(i,4)$ are equal to one, and the rest are zero. That is, there are four possible outcomes for each group. We treat each case separately.

In our derivations, we need certain expressions related to the moment generating function of M. Noting that $M = R + 1$ where $R \sim \text{Poisson}(\lambda)$, for any real t,

$$E(e^{tM}) = E(e^{t(R+1)}) = e^t E(e^{tR}) = e^t e^{\lambda(e^t - 1)};$$

$$E(Me^{tM}) = \frac{d}{dt} E(e^{tM}) = \frac{d}{dt} e^t e^{\lambda(e^t - 1)} = e^t e^{\lambda(e^t - 1)}(1 + \lambda e^t);$$

and

$$\frac{E(Me^{tM})}{E(e^{tM})} = \frac{d}{dt} \ln(E(e^{tM})) = \frac{d}{dt}(t + \lambda(e^t - 1)) = 1 + \lambda e^t.$$

Define the following set of simplifying expressions

$$t_1 = \ln(q_1), t_2 = \ln(q_2), t_3 = \ln(q_1 q_2),$$

$$\phi_1 = q_1 e^{\lambda(q_1 - 1)}, \phi_2 = q_2 e^{\lambda(q_2 - 1)}, \phi_3 = q_1 q_2 e^{\lambda(q_1 q_2 - 1)}.$$

Case 1: If $Y_r(i,1) = 1$ then from Equation 3.26 we have

$$E(M_i|Y_r(i,1) = 1) = \frac{E[M_i(q_1 q_2)^{M_i}]}{E[(q_1 q_2)^{M_i}]}.$$

But, note that

$$E(M_i|Y_r(i,1)=1) = \frac{E[Me^{t_3 M}]}{E[e^{t_3 M}]} = \frac{d}{dt}\ln[E(e^{t_3 M})].$$

Using this result we obtain

$$\frac{d}{dt}\ln[E(e^{t_3 M})] = \frac{d}{dt}\ln(e^{t_3}e^{\lambda(e^{t_3}-1)}) = \frac{d}{dt}(t_3 + \lambda(e^{t_3}-1)) = 1 + \lambda e^{t_3} = 1 + \lambda q_1 q_2.$$

Thus, we have established

$$E[M_i|Y_r(i,1)=1] = 1 + \lambda q_1 q_2. \tag{3.27}$$

Case 2: If $Y_r(i,2)=1$ then from Equation 3.26 we obtain

$$
\begin{aligned}
E[M_i|Y_r(i,2)=1] &= \frac{E[Mq_1^M(1-q_2^M)]}{E[q_1^M(1-q_2^M)]} \\
&= \frac{E[M(q_1^M - (q_1 q_2)^M]}{E[q_1^M - (q_1 q_2)^M]} \\
&= \frac{E[M(e^{t_1 M} - e^{t_3 M})]}{E[e^{t_1 M} - e^{t_3 M}]} \\
&= \frac{\phi_1(1+\lambda q_1) - \phi_3(1+\lambda q_1 q_2)}{\phi_1 - \phi_3} \\
&= 1 + \frac{\lambda q_1(\phi_1 - q_2\phi_3)}{\phi_1 - \phi_3}.
\end{aligned}
$$

Thus,

$$E[M_i|Y_r(i,2)=1] = 1 + \frac{\lambda q_1(\phi_1 - q_2\phi_3)}{\phi_1 - \phi_3}. \tag{3.28}$$

Case 3: Using a similar argument to the above, we can show

$$E[M_i|Y_r(i,3) = 1] = 1 + \frac{\lambda q_2(\phi_2 - q_1\phi_3)}{\phi_2 - \phi_3}. \qquad (3.29)$$

Case 4: Lastly, if $Y_r(i,4) = 1$ from Equation 3.26 we have

$$
\begin{aligned}
E[M_i|Y_r(i,4) = 1] &= \frac{E[M(1 - q_1^M)(1 - q_2^M)]}{E[(1 - q_1^M)(1 - q_2^M)]} \\
&= \frac{E[M - Me^{t_1 M} - Me^{t_2 M} + Me^{t_3 M}]}{E[1 - e^{t_1 M} - e^{t_2 M} + e^{t_3 M}]} \\
&= \frac{1 + \lambda - \phi_1(1 + \lambda q_1) - \phi_2(1 + \lambda q_2) + \phi_3(1 + \lambda q_1 q_2)}{1 - \phi_1 - \phi_2 + \phi_3} \\
&= 1 + \frac{\lambda(1 - \phi_1 q_1 - \phi_2 q_2 + \phi_3 q_1 q_2)}{1 - \phi_1 - \phi_2 + \phi_3}. \qquad (3.30)
\end{aligned}
$$

Finally, to determine the best predictor of N, recall that $N = M_1 + ... + M_T$. The best predictor of M_i is the conditional expectation of M_i given the observed capture history for the group. But, because $N = \sum_{i=1}^{T} M_i$, the best predictor of N is

$$
\begin{aligned}
E[N|Z_1 = z_1, Z_2 = z_2, Z_3 = z_3] &= E[\sum_{i=1}^{T} M_i|Z_1 = z_1, Z_2 = z_2, Z_3 = z_3] \\
&= \sum_{i=1}^{T} E[M_i|\text{given all}Y(i,.) = y_i]. \qquad (3.31)
\end{aligned}
$$

Then,

$$E[M_1 + ... + M_T|Y_r] =$$

$$(T - S)(1 + \lambda q_1 q_2) + z_1 \left(E[M_i|Y_r(i,2) = 1]\right) +$$

$$z_2 \left(E[M_i|Y_r(i,3) = 1]\right) + z_3 \left(E[M_i|Y_r(i,4) = 1]\right).$$

Now, using some simplifying notation

$$E_1 = (E[M_i|Y_r(i,2)=1]) = 1 + \frac{\lambda q_1(\phi_1 - q_2\phi_3)}{\phi_1 - \phi_3},$$

$$E_2 = (E[M_i|Y_r(i,3)=1]) = 1 + \frac{\lambda q_2(\phi_2 - q_1\phi_3)}{\phi_2 - \phi_3},$$

$$E_3 = (E[M_i|Y_r(i,4)=1]) = \frac{1 + \lambda - \phi_1(1 + \lambda q_1) - \phi_2(1 + \lambda q_2) + \phi_3(1 + \lambda q_1 q_2)}{1 - \phi_1 - \phi_2 + \phi_3},$$

we express the best predictor of the population size N as

$$\hat{N} = (T - S)(1 + \lambda q_1 q_2) + Z_1 E_1 + Z_2 E_2 + Z_3 E_3. \tag{3.32}$$

The expression in Equation 3.32 is a function of the unknown parameters T, q_1, and q_2, and the observed counts Z_1, Z_2, and Z_3. The unknown parameters are estimable from the observed data, either via maximum likelihood or method-of-moments equations as described in Chapter 3. Using estimates of T, q_1, and q_2 in Equation 3.32 gives

$$\tilde{N} = (\hat{T} - S)(1 + \lambda \hat{q}_1 \hat{q}_2) + Z_1 \hat{E}_1 + Z_2 \hat{E}_2 + Z_3 \hat{E}_3, \tag{3.33}$$

where estimates of E_1, E_2, E_3 are required because the expressions are functions of q_1, q_2, and T.

3.6.3   Numerical Methods

Maximum Likelihood

A numeric maximization technique for the likelihood function of Equation 3.10 is given here using Proc NLIN in SAS version 8.2. Our approach takes advantage of the fact that the likelihood function $L(q_1, q_2, T, \lambda | Z_1, Z_2, Z_3)$ is bounded between zero and one because it is a discrete probability mass function. Then $\ln(L(q_1, q_2, T, \lambda | Z_1, Z_2, Z_3) \leq 0$, and $-\ln(L(q_1, q_2, T, \lambda | Z_1, Z_2, Z_3) \geq 0$. For convenience denote $\ln(L) = \ln(L(q_1, q_2, T, \lambda | Z_1, Z_2, Z_3)$ Maximizing $\ln(L)$ as a function of q_1, q_2, and T is equivalent to minimizing $-\ln(L)$. Because SAS efficiently minimizes sums-of-squares for nonlinear regression functions, we use the NLIN procedure within SAS to minimize $-\ln(L)$.

For any nonlinear mean function $Y_i = f(\mathbf{X_i}; \boldsymbol{\beta})$ where $\mathbf{X_i}$ is a $p \times 1$ vector of covariates and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, proc NLIN in SAS finds the value of $\boldsymbol{\beta}$ that minimizes $\sum_{i=1}^{n} (Y_i - f(\mathbf{X_i}; \boldsymbol{\beta}))^2$. Our maximization routine uses $n = 1, Y_1 = 0, \mathbf{X_i} = (Z_1, Z_2, Z_3)$, $\boldsymbol{\beta} = (q_1, q_2, T)$ and $f(\mathbf{X_i}; \boldsymbol{\beta}) = \sqrt{(-\ln(L))}$. Given these inputs, SAS finds values of q_1, q_2, and T that minimize $(0 - (\sqrt{-\ln(L)}))^2 = -\ln(L)$.

A set of sample SAS Version 8.2 code is listed below, for a capture-recapture experiment with hypothetical values $Z_1 = 5, Z_2 = 8, Z_3 = 3$ and finds the maximum likelihood values of q_1, q_2 and T for $\lambda = 0.1$.

SAMPLE SAS CODE:

data test;

z1 = 5; z2 = 8; z3 = 3; S = z1+z2+z3;

input y;

```
cards;

0

;

data test; set test;

const = log(gamma(z1+1)) + log(gamma(z2+1)) + log(gamma(z3 + 1));

lambda=0.1; /* INPUT THE VALUE OF LAMBDA HERE */


proc print; run;

proc nlin data=test noitprint convergeparm=.0000001 convergeobj=.0000001

method=newton;

parms th1 = -3 to 3 by .50

th2 = -3 to 3 by .50

th3 = -2 to 5 by .50;

q1 = 1/(1+exp(-th1)); q2 = 1/(1+exp(-th2)); T = s + exp(th3);

pr00 = q1*q2*exp(lambda*(q1*q2-1)); pr10 = q2*exp(lambda*(q2-1)) - pr00;

pr01 = q1*exp(lambda*(q1-1)) - pr00; pr11 = 1 - pr00 - pr01 - pr10;

LogL = log(gamma(T + 1)) - log(gamma(T - S + 1)) - const

+ z3*log(pr11) + z2*log(pr10) + z1*log(pr01) + (T - S)*log(pr00);

L2 = -LogL;

mu = L2**(0.5);

model y = mu;

output out=nlinout2 parms = th1 th2 th3 sse = sse;
```

data retrans2; set nlinout2;

q1 = 1/(1+exp(-th1)); q2 = 1/(1+exp(-th2)); T=s+exp(th3);

negloglik=sse;

keep lambda q1 q2 T negloglik;

proc print; run;

Method of Moments

To solve the method-of-moments equations (3.16), (3.17), and (3.18) we also use
the SAS Version 8.2 software and the NLIN procedure to solve the equations numerically.
We re-express Equations (3.16), (3.17), and (3.18) as a difference of terms set equal to zero.
For instance, we express Equation (3.16) as

$$q_1 q_2 e^{\lambda(q_1 q_2 - 1)} - q_1 e^{\lambda(q_1 - 1)} - \frac{Z_1}{T} = 0. \tag{3.34}$$

Denote the left-hand sides of rearranged expressions of Equations (3.16), (3.17), and (3.18)
as EQ_1, EQ_2, and EQ_3. Choosing $n = 1, Y_1 = 0, \mathbf{X_i} = (Z_1, Z_2, Z_3)$, $\boldsymbol{\beta} = (q_1, q_2, T)$, we use
$L = (EQ_1)^2 + (EQ_2)^2 + (EQ_3)^2$ and $f(\mathbf{X_i}; \boldsymbol{\beta}) = \sqrt{L}$. Notice that L is never negative, and
that the method-of-moments solutions occur when EQ_1, EQ_2, and EQ_3 are all zero. Given
these inputs, proc NLIN finds values of q_1, q_2, and T which minimize L. Since the goal is
to find combinations of q_1, q_2, and T so that $L = 0$, convergence of the routine is checked
checked to ensure that a true solution has been reached. In this sample program, a true
solution where L is suitably close to zero is reached.

A set of sample SAS code is included here as an example:

```
data test;

z1 = 5; z2 = 8; z3 = 3; S = z1+z2+z3; lambda = 0;

input y;

cards;

0

;

data test; set test;

const = log(gamma(z1+1)) + log(gamma(z2+1)) + log(gamma(z3 + 1));

lambda=0.1;


proc print; run;

proc nlin data=test noitprint convergeparm=.0000001 convergeobj=.0000001

method=newton;


parms th1 = -2 to 2 by .500

th2 = -2 to 2 by .500

th3 = -2 to 2 by .500;

q1 = 1/(1+exp(-th1)); q2 = 1/(1+exp(-th2)); T = s + exp(th3);


eq1 = q1*q2*exp(lambda*(q1*q2-1)) - q1*exp(lambda*(q1-1)) + (z1/T);

eq2 = q2*exp(lambda*(q2-1)) - 1 + ((z1+z3)/T);

eq3 = q1*exp(lambda*(q1-1)) - 1 + ((z2+z3)/T);
```

```
L2 = eq1*eq1 + eq2*eq2 + eq3*eq3;

mu = L2**(0.5);

model y = mu;

output out=nlinout2 parms = th1 th2 th3 sse = sse;

data retrans2; set nlinout2;

q1 = 1/(1+exp(-th1)); q2 = 1/(1+exp(-th2)); T=s+exp(th3); negloglik=sse;

keep lambda q1 q2 T negloglik;

proc print; run;
```

Chapter 4

Summary and Future Work

In summary, our work has established useful findings for closed population capture-recapture models. For our Bayesian models, we developed eight Bayesian capture-recapture models accounting for the known sources of variability in the capture probabilities of closed animal populations. Using the WinBUGS v.1.4 software, these models are easy to fit to capture-recapture data sets, and MCMC estimates of the posterior density of N are easily obtained from the output. Additionally, we found that our modified version of AIC works well as a model selection tool for capture-recapture data sets. We believe AIC is very useful as a preliminary method of reducing the set of candidate models from eight down to a smaller set of candidate models worthy of further exploration to determine the best fitting model. We found that the DIC criterion did not perform as well as AIC for capture-recapture data sets and we recommend the use of AIC over DIC.

Further areas for exploration within the framework of Bayesian modeling and model selection include examining whether informative priors improve estimation of N

when capture probabilities are small. Negative bias in estimating N is common when there is heterogeneity in the population and a significant percentage of the population has small capture probabilities. Another area of exploration entails examination of the performance of our heterogeneity models $(M_h, M_{th}, M_{bh}, M_{tbh})$ when the finite mixture distribution for modeling heterogeneity has $r > 2$ mass points.

For our work on matching errors, our findings include development of a model accounting for shadow-match errors in the observed data of a capture-recapture experiment. Using our hierarchical modeling approach, we developed a model which reduces the bias in predicting N with the incorporation of additional knowledge about the population (such as knowledge of the average number of animals sharing a DNA profile). This improvement was verified by simulation results presented in Chapter 3. We established a predictor $\tilde{N}$ of N when there are shadow-match errors present in a data set, and have given a parametric bootstrap method for obtaining a prediction interval for N.

For matching error models, directions for future work include generalization of our model and predictor $\tilde{N}$ for $k > 2$ capture periods. A portion of this generalization includes determination of the estimability of λ from capture-recapture data sets with shadow-match errors for $k > 2$ capture periods.

Another direction for future work on matching error modes involves adapting our model for a capture-recapture experiment where one live animal capture is followed by several subsequent DNA captures. During the one live capture, DNA information could be extracted from each animal. From DNA fingerprints from this capture, information regarding group sizes would be available. Thus, some information on λ should be available

from the live capture. This methodology may be of interest in cases where the value of λ is believed to be high in the population.

Finally, a direction for future work is incorporation of Identity Match errors into our matching-error model. In general, Identity Match errors cause a positive bias in making inference about N in capture-recapture experiments (Lee, Seber, et al., 2001). Adding Identity Match errors to our model may require simplifying assumptions due to the complexity of accounting for both Shadow Match and Identity Match errors. Some proposed approaches for handling this error on census or epidemiological lists (such as Ding and Fienberg, 1994) will be examined to determine their applicability to our problem.

Chapter 5

Bibliography

Bibliography

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. *Proc. 2nd International Symposiom Information Theory* .

Bailey, N.T.J. 1951. On estimating the size of mobile populations from recapture data. *Biometrika* 38:293–306.

Burnham, K.P., and D.A. Anderson. 2002. *Model selection and multimodel inference, a practical information-theoretic approach.* 2nd ed. Springer-Verlag.

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

Chao, A., S-M Lee, and S-L Jeng. 1992. Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* 48(1): 201–216.

Darroch, J.N. 1958. The multiple-recapture census: I. estimation of a closed population. *Biometrika* 45:343–359.

Ding, Y., and S.E. Fienberg. 1994. Dual system estimation of census undercount in the presence of matching error. *Survey Methodology* 20(2):149–158.

Edwards, W.R., and L. Eberhardt. 1967. Estimating cottontail abundance from live trapping data. *Journal of Wildlife Management* 31(1):87–96.

Jolly, G.M. 1965. Explicit estimates from capture-recapture data with both death and immigration-sthochastic model. *Biometrika* 52:225–247.

Kremers, W.K. 1988. Estimation of survival rates from a mark-recapture study with tag loss. *Biometrics* 44(1):117–130.

Lee, A.J., G.A. Seber, J.K. Holden, and J.T. Huakau. 2001. Capture-recapture, epidemiology, and list mismatches: Several lists. *Biometrics* 57(3):707–713.

Link, W.A. 2003. Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics* 59:1123–1130.

Mills, L.S., J.J. Citta, K. Lair, M. Schwartz, and D. Tallmon. 2000. Estimating animal abundance using non-invasive dna sampling: Promise and pitfalls. *Ecological Applications* 10(1):283–294.

Nelson, L.J., D.R. Anderson, and K.P. Burnham. 1980. The effect of band loss on estimates of animal survival. *Journal of Field Ornithology* 51(1):30–38.

Norris, J.L, and K.H. Pollock. 1996. Nonparametric mle under two closed capture recapture models with heterogeneity. *Biometrics* 52(2):639–649.

Otis, D.L., K.P. Burnham, G.C. White, and D.R. Anderson. 1978. Statistical inference for capture data on closed animal populations. *Wildlife Monographs* 62.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* 56(2):434–442.

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

Schwartz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6: 497–511.

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

Seber, G.A., J.T. Huakau, and D. Simmons. 2000. Capture-recapture, epidemiology, and list mismatches: Two lists. *Biometrics* 56(4):1227–1232.

Seber, G.A.F. 2002. *The estimation of animal abundance and related parameters.* 2nd ed. Blackburn Press.

Seber, G.A.F., and J.F. Whale. 1970. The removal method for two and three samples. *Biometrics* 26(3):393–400.

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, Methodological* 64(4):583–639.

Taberlet, P., L.P. Waits, and G. Luikart. 1999. Noninvasive genetic sampling: look before you leap. *Trends in Ecological Evolution* 14(8):323–327.

Zippin, C. 1956. An evaluation of the removal method of estimating animal populations. *Biometrics* 12(2):163–189.