

Model Selection For Closed-Population Capture-Recapture Studies with Two or Three Trapping Occasions using AIC and BIC

Brandon Lockhart*

University of Victoria

Abstract

Closed-population capture-recapture models are used for estimating a population's abundance. Eight commonly used models allow for the capture probabilities to vary due to temporal effects, behavioural effects after the first capture, and heterogeneity across members in the population. Although initially developed for ecology, these models have been used in many fields including epidemiology where it can be difficult to ascertain data from many trapping occasions. We performed a simulation study to determine whether AIC and BIC could be used effectively as a model selection technique in closed-population capture-recapture studies with two or three trapping occasions. We found that the AIC and BIC procedures were incapable of correctly selecting models with temporal, behavioural, or heterogeneity effects from two trapping occasions, but with three trapping occasions, these procedures generally selected the correct model.

1 Introduction

Closed-population capture-recapture models have been used to estimate population abundance since the 1950's [10]. Otis, Burnham, White, and Anderson (1978) proposed and provided a thorough analysis of closed-population capture-recapture models which allow variation in capture probabilities due to time, behaviour, and heterogeneity. Although initially developed for ecological studies, these methods have been used in epidemiology, amongst other fields. It is usually possible to attain multiple time ordered samples in ecological studies, whereas in most epidemiological studies only two to four lists are available [4]. The motivation for this paper is to help experimenters with model selection in a study where it is difficult to attain multiple samples (the samples being naturally time ordered) from their population of interest. The results from this study should

*brandonlockhart@uvic.ca

inform the experimenter about what to be aware of when performing model selection with AIC and BIC, which could help them get the most accurate estimate of abundance. Another motivating factor is meeting the closed-population assumptions: no births, deaths, immigration or emigration. These assumptions are often attempted to be met by having a capture-recapture study during a short time period. Hence with each additional trapping occasion, the longer the study takes, and so the closed-population assumptions are less likely to be met. Additionally, this study is motivated by reducing costs for the experimenter; if they can only afford to perform two or three trapping occasions, then the results from this research can guide them in model selection for their dataset.

We performed a comprehensive simulation study in order to assess model selection on data generated with two or three trapping occasions. Program MARK [5] was used to fit the data to the models. Parameter estimates in Program MARK are obtained by the method of maximum likelihood estimation [16]. In Program MARK, maximum likelihood estimation is performed via numerical methods (we used the default Newton-Raphson numerical method).

1.1 Related work

Otis et al. (1978) performed simulation studies to assess the estimates of N for the models $M_0, M_t, M_b, M_{th}, M_{tb}, M_{th},$ and M_{bh} ; in particular, they examined bias of the estimates, confidence interval coverage, and model robustness at estimating N . However, they only simulated data with five or more trapping occasions, and suggested using their methods in a study with at least five trapping occasions [10]. They do not analyze information-theoretic methods of model selection.

Stanley and Burnham (1998) performed a simulation study which evaluated in part AIC, AIC_c and CAIC model selection for the closed-population capture-recapture models $M_0, M_t, M_b,$ and M_{tb} . However, they analyzed only seven trapping occasions and use small population abundances of 50 or 200 members. They found that overall AIC and AIC_c outperformed CAIC, and are preferred to CAIC for selection among the closed-population capture-recapture models they investigated.

Burnham, White, and Anderson (1995) performed a simulation study to evaluate information-theoretic approaches (AIC, $AIC_c,$ and CAIC) to selection of a parsimonious model and compared them to the use of likelihood ratio tests. The models were open-population CJS models. They found that in general, the information-theoretic approaches consistently outperformed the likelihood ratio test approach. They also concluded because of its theoretical basis, and good performance in their study, AIC is recommended for general use in objective selection among CJS models.

2 Methods

2.1 Closed-population assumptions

Closure means the size of the population is constant over the period of investigation; we assume the population to be closed geographically - no movement on or off the study area - and demographically - no births or deaths [5].

2.2 Closed-population capture-recapture models

Otis et al. (1978) proposed eight closed-population capture-recapture models: M_0 , M_t , M_b , M_h , M_{tb} , M_{th} , M_{bh} , and M_{tbb} , where t , h , and b refer to time, heterogeneity, and behavioural effects. In this study, the first four models listed are investigated. For these models, there are three assumptions:

1. the population is closed,
2. members do not lose their marks during the period of the experiment, and
3. all marks are correctly noted and recorded at each trapping occasion.

2.2.1 Model M_0

The constant model M_0 assumes the capture probability of each member of the population is the same on every trapping occasion. This model has two parameters: the probability of capture p , and the population size N . With two trapping occasions, the maximum likelihood estimate of N is given by

$$\hat{N}_0 = \frac{2(n_1 + n_2)}{4m_2^2}$$

where n_i is the number of members captured during the i^{th} trapping occasion, and m_2 is the number of recaptures in the second sample.

2.2.2 Model M_t

The time model M_t assumes capture probabilities vary with time. This phenomenon may occur for example if there's a rainy day during a study (capture probabilities might be lower), or if different capture methods are used during the same study. In a dataset with t trapping occasions, this model has $t + 1$ parameters: capture probabilities for each trapping occasion p_1, p_2, \dots, p_t and population size N .

With two trapping occasions, the maximum likelihood estimate of N for model M_t is the Lincoln-Petersen estimator

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

where n_1 , n_2 , and m_2 have already been defined.

2.2.3 Model M_b

The behaviour model assumes capture probabilities vary by behavioural response to being captured. When a member of the population is more likely to be recaptured than initially captured, this is denoted as being “trap-happy” behaviour, and when a member of the population is less likely to be recaptured than it was to be captured initially, this is denoted as being “trap-shy” behaviour. An animal may become trap-happy if an experimenter is luring the animals to their capture location with food. Conversely, an animal may become trap-shy if the animal was frightened or hurt during initial capture and marking. All members in the population have the same probability of capture and the same probability of recapture. The three parameters of model M_b are population size N , capture probability p , and recapture probability c . A member’s capture probability is altered only once since multiple different recapture probabilities do not affect the estimate of population size [10].

2.2.4 Model M_h

The heterogeneity model assumes each member of a population has an inherent probability of being captured. In this study we investigate hidden heterogeneity, where hidden refers to unobserved variability in the data, rather than say explicit covariates denoting the heterogeneity. This study analyses the case when each member of the population is a member of one of two mixture groups (A and B). Pledger (2000) developed this idea of partitioning the population into two or more groups with relatively homogeneous capture probabilities. An example of two mixture groups in a population could be sex; perhaps males and females have different capture probabilities. In this case, the model has four parameters: the probability π of being in mixture group A ($1 - \pi$ is the probability of being in mixture group B), the capture probability p_A of a member in mixture group A , the capture probability p_B of a member of mixture group B , and population abundance N .

2.3 Likelihoods

Two likelihoods were used in this study: the full likelihood and Huggins (1989, 1991) conditional likelihood where N is conditioned out of the likelihood. The full likelihood \mathcal{L} is defined as

$$\mathcal{L}(N, \mathbf{p}, \mathbf{c} \mid \text{data}) \propto \frac{N!}{(N - M_{t+1})!} \prod_h \Pr[h]^{n_h} \Pr[\text{not encountered}]^{N - M_{t+1}}$$

where M_{t+1} is the number of unique individuals captured during the study, and n_h is the number of individuals with capture history h [5]. For example, the likelihood for model M_t with two trapping occasions is

$$\mathcal{L}_t(N, p_1, p_2 \mid \text{data}) \propto \frac{N!}{n_{00}!} (p_1 p_2)^{n_{11}} (p_1 (1 - p_2))^{n_{10}} ((1 - p_1) p_2)^{n_{01}} ((1 - p_1) (1 - p_2))^{n_{00}}$$

where n_{01} , n_{10} , n_{11} , and n_{00} are the number of individuals with capture history 01, 10, 11, and 00 respectively

Using the notation from Huggins (1991), Huggins conditional likelihood can be expressed as

$$\mathcal{L}_H(\mathbf{p}, \mathbf{c} \mid \text{data}) \propto \prod_{i=1}^{M_{t+1}} \prod_{j=1}^t \frac{p_{ij}^{x_{ij}} (1 - p_{ij})^{(1-x_{ij})}}{1 - \prod_{l=1}^t (1 - p_{il}^*)}$$

where there are t trapping occasions, p_{ij} denotes the probability that individual i is captured on trapping occasion j , x_{ij} equals 1 if individual i is captured on trapping occasion j and 0 otherwise, and p_{ij}^* is p_{ij} evaluated when individual i has not been captured before trapping occasion j [6, 7]. Notice this likelihood only depends on the encountered individuals. For example, the Huggins conditional likelihood of model M_t with two trapping occasions is

$$\mathcal{L}_{H,t}(p_1, p_2 \mid \text{data}) \propto \frac{(p_1(1-p_2))^{n_{10}} ((1-p_1)p_2)^{n_{01}} (p_1 p_2)^{n_{11}}}{(1 - (1-p_1)(1-p_2))^{M_{t+1}}}.$$

2.4 Information-theoretic methods of model selection

2.4.1 Akaike's Information Criterion AIC

Akaike (1974) introduced the minimum information theoretical criterion (AIC) estimate which is designed for the purpose of statistical identification [1]. The minimum information theoretical criterion estimate is defined by the model and the maximum likelihood estimates of the parameters. This gives the minimum of AIC defined as

$$\text{AIC} = -2 \ln(\mathcal{L}) + 2k \quad (1)$$

where \mathcal{L} is the maximum likelihood of the model and k is the number of independently adjusted parameters within the model [1]. In practice, AIC is computed for each model, and the model selected as best supported by the data is the model having the lowest AIC value [15].

2.4.2 Corrected Akaike's Information Criterion AIC_c

To correct for small sample bias, a variant of AIC, AIC_c, was proposed by Hurvich and Tsai (1989) and is defined as

$$\text{AIC}_c = \text{AIC} + \frac{2k^2 + 2k}{N - k - 1}$$

where N is the sample size [8]. One can see that for large N , $\text{AIC} \approx \text{AIC}_c$.

2.4.3 Bayesian Information Criteria BIC

Schwarz (1978) developed the Bayesian information criterion which for capture-recapture studies is defined as

$$\text{BIC} = -2 \ln(\mathcal{L}) + \ln(M_{t+1})k \quad (2)$$

where M_{t+1} is the number of members known to have been in the population [14].

2.5 95% confidence interval for \hat{N}

Program MARK calculates a 95% confidence interval for the estimate of N as

$$(M_{t+1} + (\hat{f}_0/C), M_{t+1} + (\hat{f}_0 \times C))$$

where $\hat{f}_0 = \hat{N} - M_{t+1}$ and

$$C = \exp \left\{ 1.96 \left[\ln \left(1 + \frac{\widehat{\text{var}}(\hat{N})}{\hat{f}_0^2} \right) \right]^{1/2} \right\}.$$

Note that the lower bound of this confidence interval is greater than M_{t+1} .

2.6 Simulation studies

The simulations were written in the R statistical programming language [13]. The R package RMark [9] was used to fit the models using the Program MARK software.

2.6.1 Generating the data

After specifying the model and its parameter values, a capture history for each member in the population is simulated. In closed-population studies, a capture history of an individual is represented as $s_1 s_2 \dots s_t$ where t equals the number of trapping occasions and

$$s_i = \begin{cases} 1 & \text{if a capture occurred on sampling occasion } i \\ 0 & \text{if a capture did not occur on sampling occasion } i \end{cases}.$$

Temporal effects, behavioural effects, and hidden heterogeneity were simulated in the data as follows.

1. *Time data.* Suppose the probability of capture on trapping occasion i is p_i , where $i \in \{1, 2, \dots, t\}$. To simulate the capture history of a member in the population, for each trapping occasion i , a Bernoulli random variable X_i was simulated using the R function `rbinom` such that

$$X_i = \begin{cases} 1 & \text{with probability } p_i \\ 0 & \text{with probability } 1 - p_i \end{cases}.$$

After simulating a capture history for each member in the population, the capture histories of members that were never captured (members with an all 0 capture history) are discarded, and the remaining data form the final dataset.

2. *Behaviour data.* Suppose p is the probability of capture and c is the probability of recapture. The first trapping occasion for each member in the population is simulated by a Bernoulli trial with success probability p ; if the trial was a success, a 1 is added to the capture history, if not, a 0 is added to the capture history. For the subsequent trapping occasions, for each member, we first check whether it has been caught before; if the member has been caught before, a Bernoulli trial is performed with success probability c , and if the member has not been caught before, a Bernoulli trial is performed with success probability p . As with generating time data, the R function `rbinom` is used to perform the Bernoulli trials. Finally, the capture histories of members that were never captured are discarded, and the remaining data form the final simulated dataset.
3. *Hidden-heterogeneity data.* Let π be the probability of being a member of mixture group A , p_A be the capture probability for mixture group A , and p_B be the capture probability for mixture group B . First, for each member in the population, a uniform random number between 0 and 1 is generated using the R function `runif`; if this number is less than π , this individual is classified as mixture A , and if this number is greater than π , this individual is classified as mixture B . Next, for each individual in mixture group A , a Bernoulli trial is performed for each trapping occasion with success probability p_A , similarly, for each individual in mixture group B , a Bernoulli trial is performed for each trapping occasion with success probability p_B . If a Bernoulli trial is a success, a 1 is added to the capture history, and if not a 0 is added to the capture history. Finally, the all zero capture histories are discarded, and the remaining data form the final dataset.

2.6.2 Number of simulations

Stanley and Burnham (1998) generated one thousand capture histories for each of the cases investigated in their study. This study performed one thousand repetitions as well; for each data model (time, behaviour, heterogeneity), and choice of parameters to define that model, one thousand different capture histories were simulated and fit to the models of interest. In this study, one thousand repetitions was sufficient to achieve equilibrium in terms of the statistics averaged across all repetitions.

2.6.3 Probability values

The capture probabilities and mixture membership probabilities (used to simulate heterogeneity) examined in this study were chosen with the intention of getting a general idea of the entire range of possible probabilities. With two trapping occasions, time data was generated with $p_1, p_2 \in \{0.2, 0.5, 0.8\}$, and behaviour data was generated with $p, c \in \{0.2, 0.5, 0.8\}$, giving nine different combinations of capture probabilities each. Additionally, the range around each

combination of probabilities was investigated; this was done in order to determine how strong the effects of time or behaviour needed to be for the constant model not to be chosen as the best model, and to see how small changes in probabilities had an affect on model selection. So for each combination of $p_1 = x \in \{0.2, 0.5, 0.8\}$ and $p_2 = y \in \{0.2, 0.5, 0.8\}$, eight more combinations of probabilities were examined where $p_1 = x$ and $p_2 \in \{y - 0.04, y - 0.03, y - 0.02, y - 0.01, y + 0.01, y + 0.02, y + 0.03, y + 0.04\}$ (substitute $p = p_1$ and $c = p_2$ if generating behaviour data). Small capture probabilities were also investigated in the range 0.05 to 0.1.

To simulate time data with three trapping occasions, all combinations of $p_1, p_2, p_3 \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ were used, and to simulate behaviour data for three trapping occasions, all combination of $p, c \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ were examined.

To simulate heterogeneity data, 125 different combinations of capture probabilities were examined where $\pi, p_A, p_B \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ (for any number of trapping occasions). As above, these probabilities were chosen with the intention of getting a general idea of the whole range of possible probabilities.

2.6.4 Choosing N

The plan for choosing population abundance N was to select a number large enough such that the capture probabilities that were used to model the data were accurately represented in the data. For this reason, $N = 10000$ was chosen for all simulations. However, a subset of simulations were performed with $N = 1000$ and $N = 100000$ to analyze how differing N affected the results.

2.6.5 Statistics

I will use the term “experiment” going forwards which I am defining to mean one simulation study. One experiment is comprised of choosing the data model (time, behaviour, heterogeneity) and parameters (capture probabilities, population abundance), from which 1000 capture histories are generated and each fit to the models of interest, finally the statistics are calculated using the results of the models fit to 1000 different capture histories.

The following lists all of the metrics calculated for each experiment.

1. **AIC count.** From the 1000 simulationss, this is simply a count of how many times the AIC model selection procedure chooses each model.

To be precise, RMark used AIC_c not AIC. But since we are using $N \geq 1000$, $AIC_c = AIC + O(10^{-2})$, so the AIC and AIC_c values are very close if not the same (same with respect to the order of precision they are calculated to). I henceforth use the term AIC although technically AIC_c was used.

2. **BIC count.** From the 1000 simulations, this is simply a count of how many times the BIC model selection procedure chooses each model.

3. **Average difference in AIC values between models M_t and M_b .** If two models are such that the absolute value of the difference of their AIC values is less than 2, these models are considered indistinguishable using AIC [2]. For some simulations, we were interested in how confidently AIC chose between model M_t and model M_b , thus we calculated the average difference in AIC values between models M_t and M_b as

$$\frac{1}{n} \sum_{i=1}^n |\text{AIC}_{t_i} - \text{AIC}_{b_i}|$$

where AIC_{t_i} is the AIC value for model M_t on simulation i and AIC_{b_i} is the AIC value for model M_b on simulation i .

4. **Average population abundance estimate $\bar{\hat{N}}$.** Define the mean estimate for the true abundance of the population as

$$\bar{\hat{N}} = \frac{1}{n} \sum_{i=1}^n \hat{N}_i$$

where n is the number of simulations and \hat{N}_i is the estimate of the true abundance for simulation i .

5. **Confidence interval count.** Over the 1000 simulations, this is a count of how many times the true population abundance N is inside the 95% confidence interval for each model.
6. **Standard deviation of the estimates of population abundance.** The standard deviation s of the estimates of population abundance is calculated as

$$s = \sqrt{\frac{\sum_{i=1}^n (\hat{N}_i - \bar{\hat{N}})^2}{n - 1}}.$$

7. **Root-mean-square error of N .** The root-mean-square error (RMSE) of the true population abundance N is calculated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{N}_i - N)^2}{n}}.$$

8. **Average bias of N .** The average bias of the true population abundance is calculated as

$$\bar{\hat{N}} - N.$$

9. **Average relative bias of N .** The average relative bias of the true population abundance is calculated as

$$\frac{\bar{\hat{N}} - N}{N}.$$

10. **Average count of capture histories.** For two trapping occasions, an average of the number of individuals with capture history n_{11} was calculated as

$$\hat{n}_{11} = \frac{1}{n} \sum_{i=1}^n n_{11_i}$$

where n_{11_i} is the number of individuals with capture history 11 generated on simulation i . This calculation was performed for capture histories n_{10} and n_{01} as well.

2.6.6 Erroneous simulation results

For some combinations of capture probabilities, a model's estimate for population abundance is much larger than the known true value. It is not uncommon for studies that simulate large amounts of data to incur pathological results [15]. Stanley and Burnham (1998) dealt with erroneous results by deleting estimates of population abundance over three times as large as the true value. This approach was applied throughout this study, however in one situation it did not work. The situation was when time data was fit by the behaviour model. When time data was generated with p_1 small and $p_2 \gg p_1$, the behaviour model consistently output an estimate of N orders of magnitude larger than the true value. Upon inspection of this situation's estimates of N , it appeared as though the estimates were random numbers ranging from 8 to 12 digits when the true population size was $N = 10000$. These cases are handled carefully in the results section of this report.

3 Results

3.1 Full likelihood models

3.1.1 Accuracy of the AIC and BIC model selection procedures

With two trapping occasions, if constant data was generated, AIC and BIC chose the constant model. More precisely, in general, during any experiment when data was generated following the constant model, AIC chose the constant model over 70% of the time and BIC chose the constant model over 95% of the time. However, the converse did not hold. When heterogeneity data was generated with two trapping occasions, the AIC and BIC model selection procedures chose model M_0 generally over 80% of the time in each experiment. One general indicator for AIC and BIC choosing the constant model was when $\hat{n}_{01} \approx \hat{n}_{10}$ (which clearly should be the case for constant data), and this was often the case when heterogeneity data was generated with two trapping occasions.

With two trapping occasions, AIC and BIC cannot accurately detect temporal or behavioural variation in the data. For example, when time data was generated with a capture probability on the first trapping occasion of 0.8 and a capture probability on the second trapping occasion of 0.24, AIC and BIC

incorrectly chose model M_b every time (in experiments with $N = 1000$ and $N = 10000$). Moreover, the average difference in AIC values between model M_t and model M_b was 3.75; this is evidence that AIC can confidently choose the incorrect model. As another example, when behaviour data was generated with capture probability 0.5 and recapture probability 0.8, AIC and BIC chose model M_t every time (in experiments with $N = 1000$ and $N = 10000$). In this case, the average difference in AIC values between model M_t and model M_b was 1.18.

With two trapping occasions, the AIC and BIC model selection procedures were unable to detect heterogeneity in the data. Generally, for any experiment, AIC would choose the constant model (over 80% of the time) and the remaining times it would choose model M_b or model M_t (less than 10% of the time each). For BIC, generally in an experiment it would choose model M_0 almost 1000 times. On occasion, for example when $\pi = 0.5$, $p_A = 0.1$ and $p_B = 0.9$ (which defines very extreme heterogeneous data) AIC and BIC would choose model M_0 less and it would instead choose models M_b and M_t more, but it would not choose model M_h .

With three trapping occasions, AIC and BIC can accurately detect temporal and behavioural variation in the data. When time data was simulated, for any experiment, AIC and BIC would choose model M_t over 95% of the time in general. Only when $p_1 = 0.7$ and $p_2 = p_3 = 0.1$ or when $p_1 = 0.9$ and $p_2 = p_3 \in \{0.1, 0.3, 0.5, 0.9\}$ did AIC and BIC incorrectly choose model M_b a non-negligible amount of the time (greater than 30% of the time). When behaviour data was simulated, AIC and BIC chose the behaviour model often 100% of the time in an experiment. Table 1 displays how strong behavioural effects need to be in the data so that AIC and BIC can detect it.

When heterogeneity data was generated with three trapping occasions, AIC and BIC could start to detect the heterogeneity. When an experiment modeled data with parameters defining clear heterogeneity, such as when $\pi = 0.3$, $p_A = 0.3$ and $p_B = 0.7$ or when $\pi = 0.5$, $p_A = 0.9$ and $p_B = 0.5$ then AIC and BIC chose model M_h correctly every time. However when the heterogeneity was less pronounced, for example when $\pi = 0.7$, $p_A = 0.5$ and $p_B = 0.3$, AIC and BIC could not confidently detect the heterogeneity (in this experiment it chose the correct model (M_h) 361 out of 1000 times).

3.1.2 Estimates of N and model robustness

For time and behaviour data, on average, the model that corresponds to the data has a more accurate estimate of N . In terms of model robustness towards estimating N , model M_t is not robust to behaviour data, model M_b is not robust to time data, and model M_0 is not robust to time or behaviour data. Next I detail how each model performed when fit to each data model.

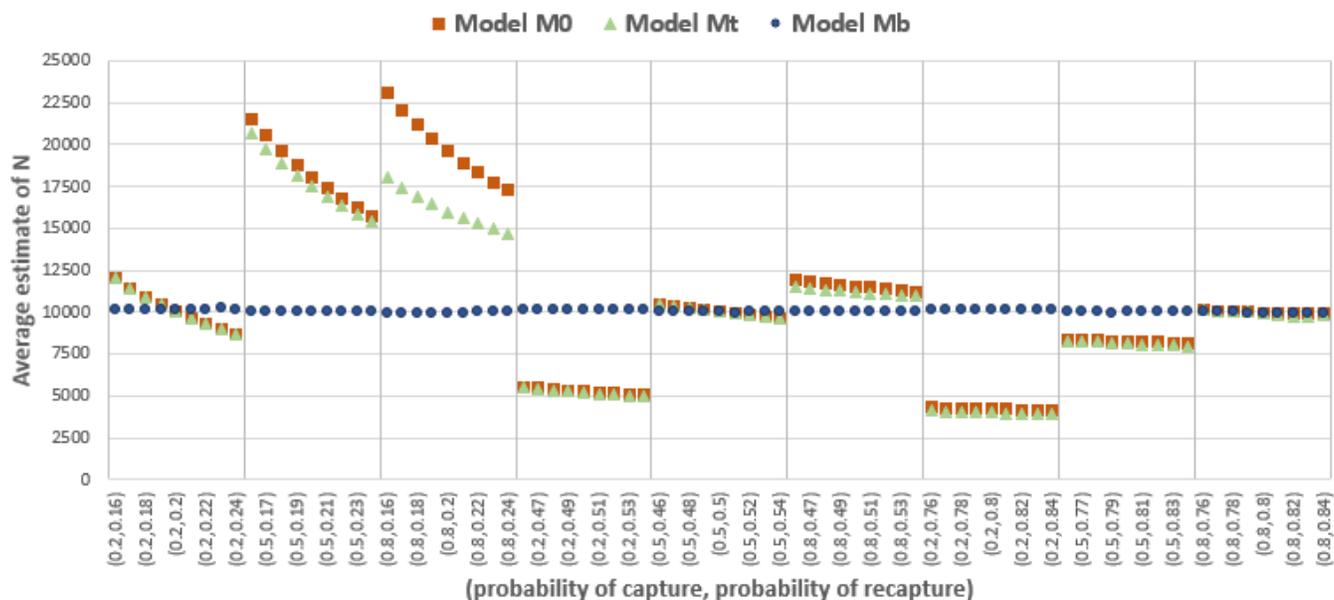
Table 1: This table lists for the 1000 simulations in an experiment, how many times the AIC and BIC model selection procedure chose each model. The data is simulated with behavioural variation over three trapping occasions and $N = 10000$. The purpose of this table is to show how strong behavioural effects in the data need to be for the AIC and BIC model selection procedures to confidently choose the behaviour model.

Probabilities		AIC chose model			BIC chose model		
p	c	M_0	M_t	M_b	M_0	M_t	M_b
0.5	0.4	0	12	988	0	2	998
0.5	0.41	0	33	967	0	1	999
0.5	0.42	0	41	959	0	2	998
0.5	0.43	0	61	939	0	7	993
0.5	0.44	0	78	922	0	3	997
0.5	0.45	0	91	909	5	1	994
0.5	0.46	2	120	878	67	2	931
0.5	0.47	27	175	798	415	1	584
0.5	0.48	159	151	690	792	2	206
0.5	0.49	570	122	308	976	1	23
0.5	0.5	783	87	130	999	0	1
0.5	0.51	532	132	336	979	0	21
0.5	0.52	192	190	618	790	1	209
0.5	0.53	14	162	824	355	1	644
0.5	0.54	1	128	871	62	0	938
0.5	0.55	0	101	899	2	4	994
0.5	0.56	0	61	939	0	3	997
0.5	0.57	0	39	961	0	1	999
0.5	0.58	0	19	981	0	0	1000
0.5	0.59	0	11	989	0	1	999

3.1.3 Behaviour data

Model M_b on average accurately estimated N in each experiment. Moreover, the true population abundance was inside the confidence interval from model M_b generally over 90% of the time; the lowest confidence interval coverage seen was 716 out of 1000 simulations in an experiment. However, the only time N was in the confidence intervals of the other models was when the behavioural effects in the data were very minimal (when $p \approx c$) or nonexistent (ie when $p = c$, so constant data). Model M_0 and M_t have a positive bias estimating N (will overestimate N) when the data models trap-shy behaviour and have a negative bias estimating N (will underestimate N) when the data models trap-happy behaviour (see Figures 1, 6, and 7). With small capture probabilities (≤ 0.1), model M_b 's average estimates of N were slightly worse, but there is no clear bias (see Figure 8).

Figure 1: This plot has capture probabilities on the x -axis and the average estimate of N on the y -axis. These estimates are from behaviour data with two trapping occasions. The true population size is 10000. There are nine distinct subsets of the probability space presented in this plot, the regions for each subset are demarcated by a vertical line. The nine subsets of probabilities are all combinations of $p \in \{0.2, 0.5, 0.8\}$ and $c = \{z - 0.04, z - 0.03, z - 0.02, z - 0.01, z, z + 0.01, z + 0.02, z + 0.03, z + 0.04\}$ where $z \in \{0.2, 0.5, 0.8\}$. Although the labels along the x -axis are not exhaustive, they should make clear what region of the probability space is depicted above it. I emphasize that this plot is not increasing linearly along the x -axis, rather it is like nine plots side by side

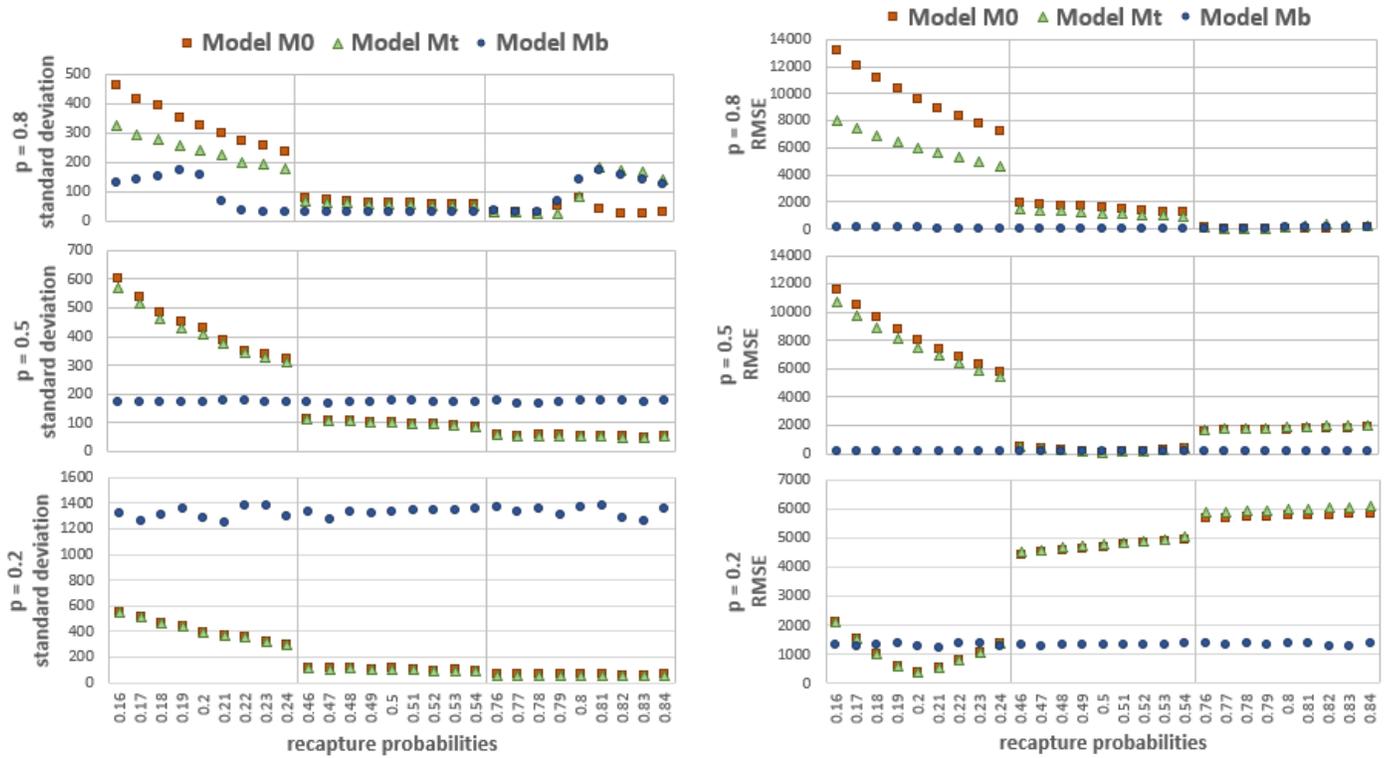


Despite model M_b having the most accurate estimate of N , it often had larger standard deviations than model M_0 and model M_t . This suggests that estimates from model M_0 and model M_t are inaccurate but precise, whereas estimates from model M_b are less precise but on average more accurate. As expected, model M_b had lower root-mean-square errors than the other models. See Figures 2, 9, and 10.

3.2 Time data

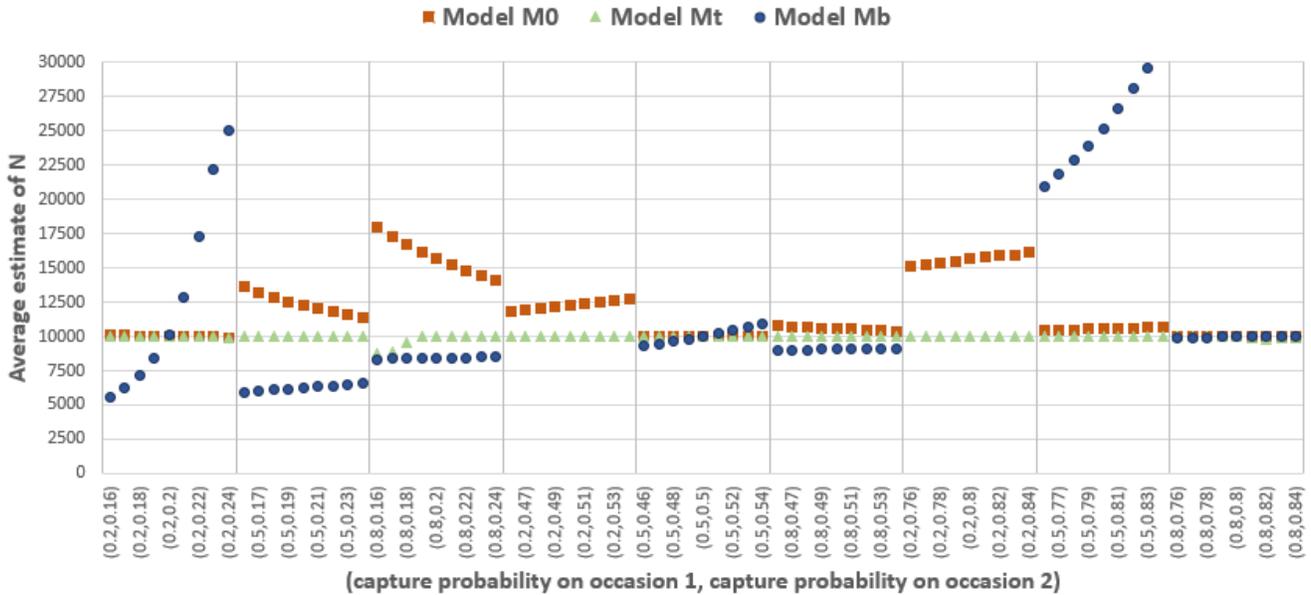
Model M_t generally accurately estimated N when it was fit to time data; it only underestimated N slightly when there were two trapping occasions with $p_1 = 0.8$ and $p_2 \in \{0.16, 0.17, 0.18\}$. In general, the confidence interval from model M_t contained the true population abundance over 90% of the time in an

Figure 2: These plots are built from behaviour data with $N = 10000$ and two trapping occasions. On the left is a chart matrix (of nine charts) of the standard deviation of the average estimates of N and on the right is a chart matrix (of nine charts) of the root-mean-square error (RMSE) of the estimates of N . Recapture probabilities are listed on the x -axis of both charts, and are plotted against standard deviation or RMSE on the y -axis. There were three initial capture probabilities $p = 0.2, 0.5, 0.8$, and these are listed to the left of the three charts to which each capture probability corresponds.



experiment, however in some instances, such as when data was generated with $p_1 = 0.1$, $p_2 = 0.5$ and $p_3 = 0.9$, the confidence interval never included the true abundance over any of the 1000 simulations (this situation only occurred twice). Also, the only time N was in the confidence intervals of the other models was when the temporal effects in the data were very minimal (when $p_1 \approx p_2$) or nonexistent (when $p_1 = p_2$, so constant data). Model M_0 has positive bias estimating N when fit to time data; the more extreme the temporal effects are in the data the more model M_0 overestimates N . Model M_b has positive bias estimating N when $p_1 < p_2$ and negative bias estimating N when $p_1 > p_2$. See Figures 3, 11, and 12. With small capture probabilities (≤ 0.1), model M_t 's average estimates of N were still accurate, but model M_b had very extreme estimates of N (see Figure 13).

Figure 3: This plot has capture probabilities on the x -axis and the average estimate of N on the y -axis. These estimates are from time data with two trapping occasions. The true population size is 10000. The probability space is all combinations of $p \in \{0.2, 0.5, 0.8\}$ and $c = \{z - 0.04, z - 0.03, z - 0.02, z - 0.01, z, z + 0.01, z + 0.02, z + 0.03, z + 0.04\}$ where $z \in \{0.2, 0.5, 0.8\}$. I limited the maximum estimate of N in this plot to be 30000 so it omits the behaviour model's very high estimates of N (this occurs in two subplots below).



Model M_t in general had the smallest standard deviation for its estimates of N , but model M_0 had small standard deviations as well. Model M_b had in some cases very volatile estimates of N which resulted in large standard deviations and root-mean-square errors. Model M_t had very low root-mean-square errors

as expected, whereas the root-mean-square errors for model M_0 were generally much higher than model M_t but lower than model M_b . See Figures 4 and 14.

3.2.1 Heterogeneity data

All models (M_0 , M_t , M_b , and M_h) generally underestimated N when fit to heterogeneity data for two and three trapping occasions (see Figures 5a and 5b). With two trapping occasions, all models performed very similarly estimating N , but with three trapping occasions, model M_h sometimes had an average estimate of N more accurate than the other models. Since with two and three trapping occasions all models still on average underestimated N a non-negligible amount, the simulations were all repeated with four trapping occasions. With four trapping occasions, model M_h on average had accurate estimates of N (Figure 5c), but the other models still often underestimated N .

The confidence interval coverage of N for heterogeneity data with two and three trapping occasions was very poor for all models; generally the true population abundance was only inside the confidence intervals when the heterogeneity effects were very minimal (so it was almost constant data). However with four trapping occasions, for each experiment, the true population abundance was inside the confidence interval for model M_h at least 526 out of 1000 simulations, but generally it was inside over 900 out of 1000 simulations for an experiment. The confidence interval coverage for the other models was still poor with four trapping occasions.

In general, the standard deviation of the estimates of N were low for each model with two and three trapping occasions (see Figures 15a and 15c). Thus, despite each model estimating N often inaccurately on average, the models are consistent with their estimation. With four trapping occasions, model M_h had in general a higher standard deviation for its estimates of N (see Figure 15e), yet recall on average these estimates were very close to the true population abundance.

The root-mean-square error of models M_0 , M_t and M_b for their estimates of N were in general similar (see Figures 15b, 15d, and 15f), yet the root-mean-square error of model M_h was generally less than the other models for data with four trapping occasions as expected.

3.3 Conditional likelihood models

Every result stated for the full likelihood models holds for the conditional likelihood models with a few exceptions. I only detail these exceptions in this section.

With two trapping occasions, the likelihood function values for models M_t and M_b are exactly the same. In other words, when Huggins' time and behaviour models are fit to constant, time, behaviour, or heterogeneity data, Program MARK will derive parameter estimates for these models such that when the maximum likelihood is calculated with these estimates, the values of the likelihood functions are exactly the same. See Appendix B for an example. Since the

Figure 4: These plots are built from time data with $N = 10000$ and two trapping occasions. On the left is a chart matrix (of nine charts) of the standard deviation of the estimates of N and on the right is a chart matrix (of nine charts) of the root-mean-square error (RMSE) of the estimates of N . Capture probabilities for trapping occasion 2 are listed on the x -axis of both charts, and this is plotted against standard deviation or RMSE on the y -axis. There were three capture probabilities for trapping occasion 1, $p_1 = 0.2, 0.5, 0.8$, and these are listed to the left of the three charts to which each capture probability corresponds. Extremely high standard deviations and root-mean-square errors from model M_b are omitted from these plots.

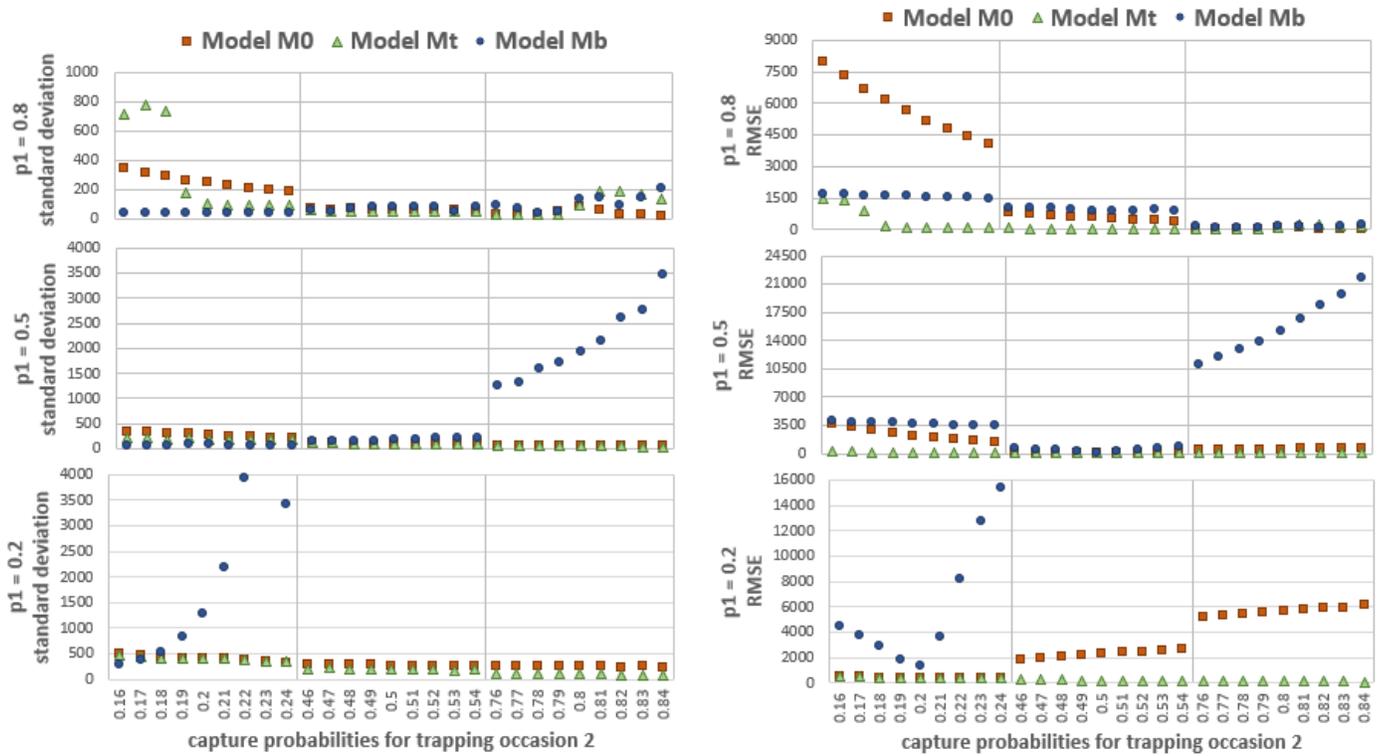
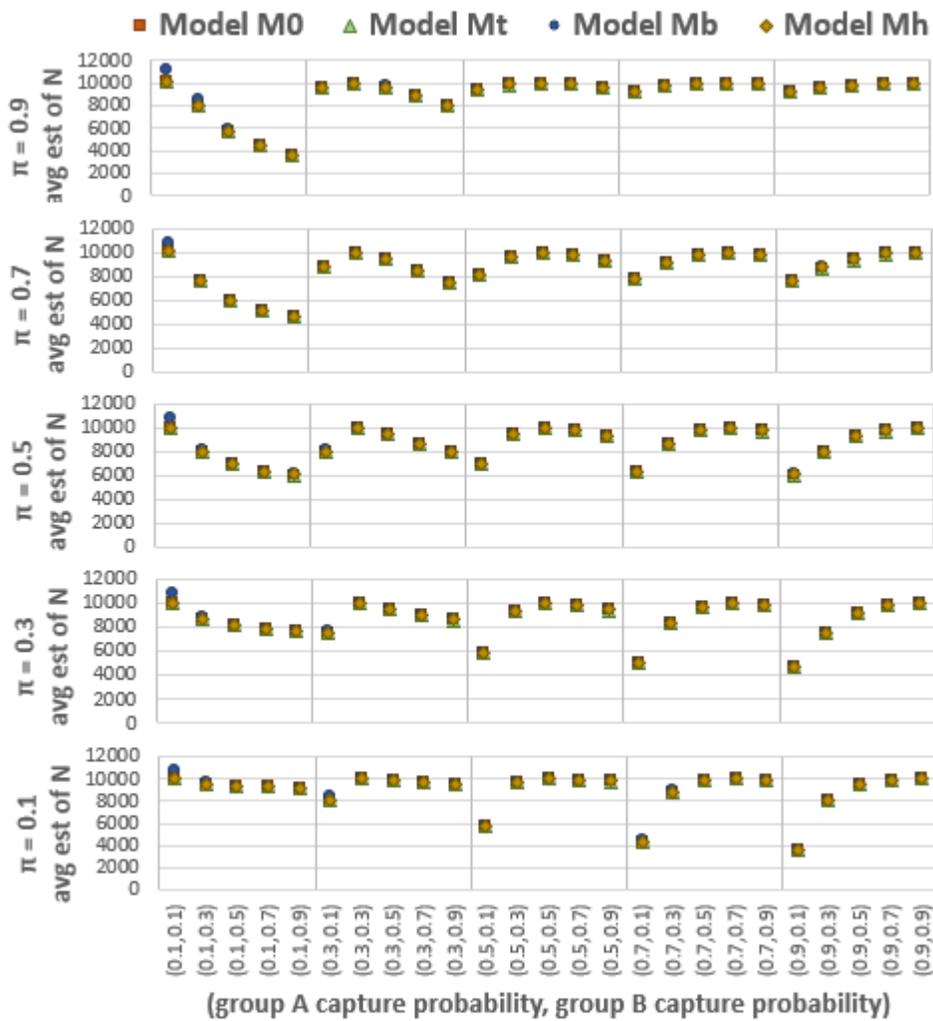
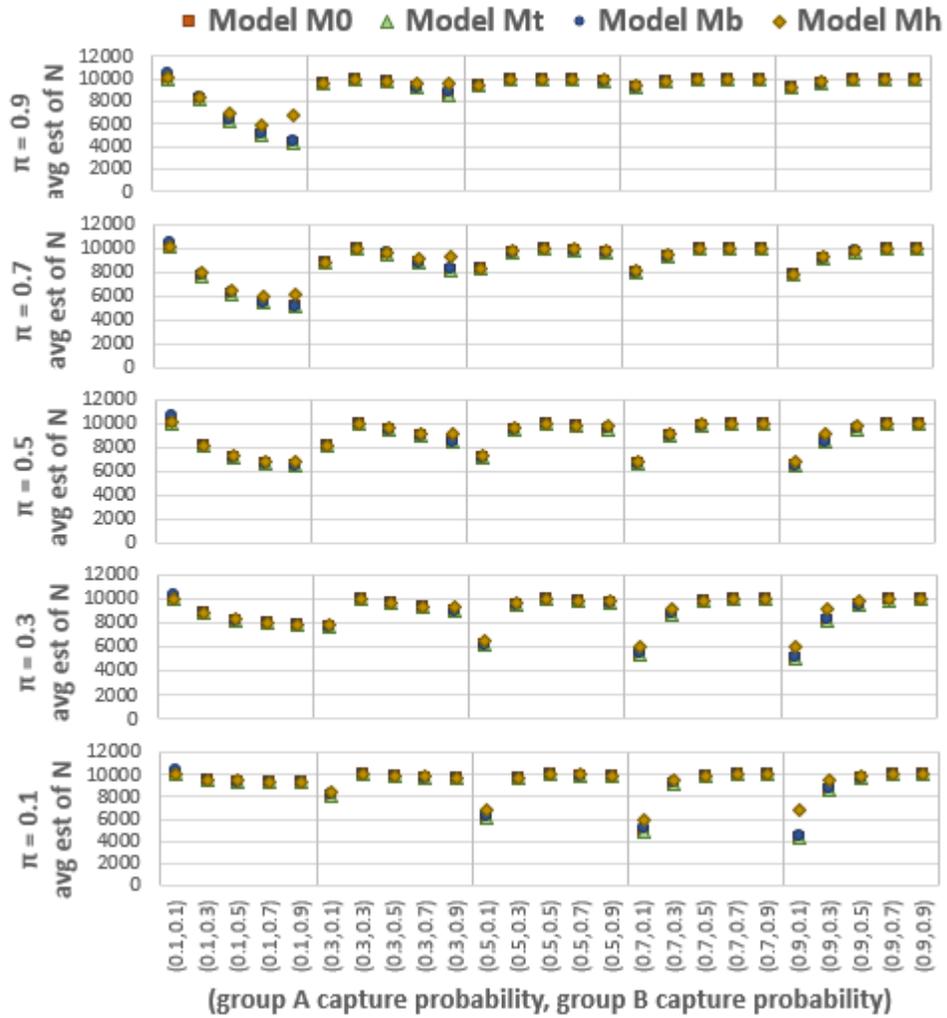


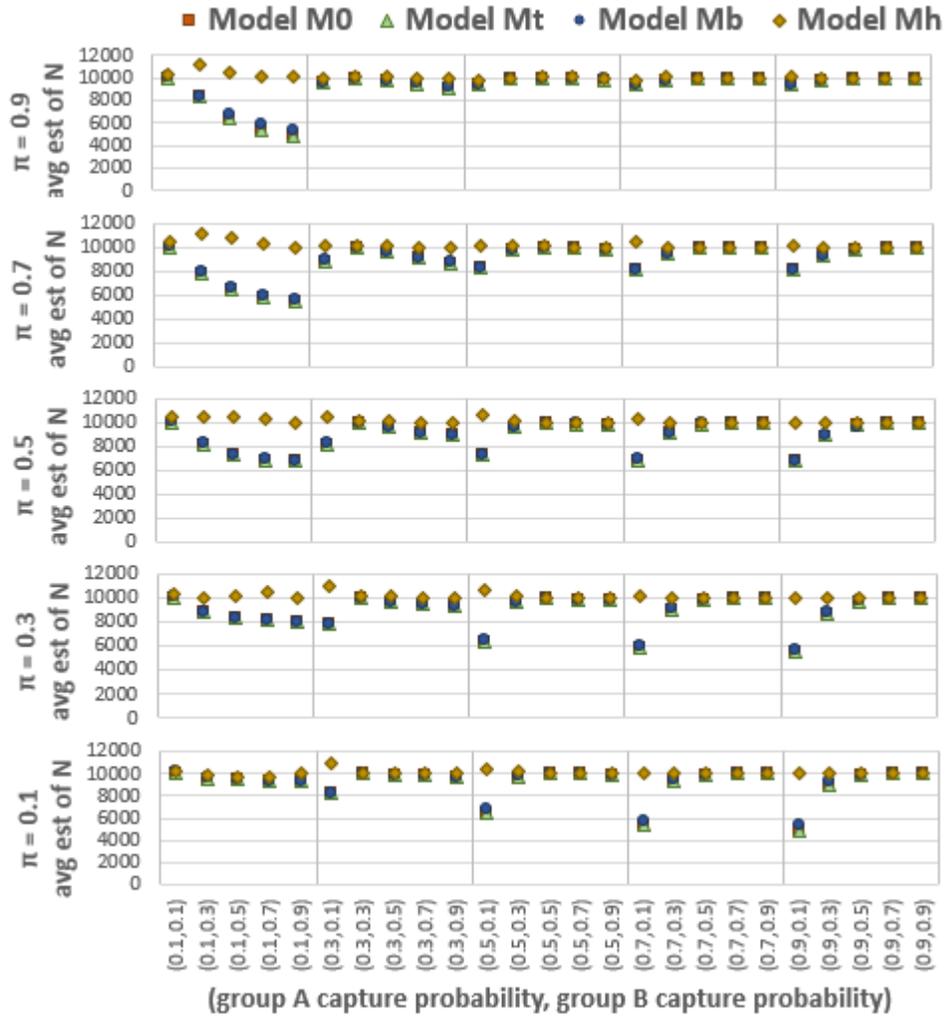
Figure 5: The following three plots are 5×5 chart matrices displaying the average estimates of N from each model for two, three, and four trapping occasions (one trapping occasion for each chart matrix). The data is heterogeneous and $N = 10000$. Each chart matrix plots all combinations of $\pi, p_A, p_B \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. On the x -axis are the capture probabilities p_A and p_B listed as an ordered pair, and they are plotted against the average estimate of N on the y -axis. Also, the π probability is listed to the left of the five subplots to which it corresponds.



(a) Two trapping occasions.



(b) Three trapping occasions.



(c) Four trapping occasions.

likelihood for the time and behaviour models are the same, and each model has two parameters for two trapping occasions (p_1 and p_2 for the time model and p and c for the behaviour model, since N is conditioned out of the likelihood), the AIC and BIC values for these models are the same. Thus, obviously the AIC and BIC model selection procedures cannot be used for model selection with time and behaviour models using Huggins' conditional likelihood with two trapping occasions. Note that with three trapping occasions, this phenomenon did not occur, and AIC and BIC performed model selection accurately with Huggins' conditional models.

One other difference was that model M_h was slightly better at estimating N with the conditional likelihood compared to the full likelihood. The true population abundance was inside the confidence interval of an estimate of N at least 275 out of 1000 simulations for all experiments, but it was generally inside over 500 out of 1000 simulations (whereas with the full likelihood the true population abundance was often never inside a single confidence interval during an experiment). Moreover, the estimates of N from the conditional likelihood were generally slightly more accurate compared to the full likelihood.

4 Discussion

This study assessed the efficacy of the information-theoretic methods of model selection AIC and BIC at detecting temporal variation, behavioural variation, and heterogeneity in the capture probabilities from capture-recapture data with two or three trapping occasions. It was found that AIC and BIC were not capable of distinguishing the correct model with only two trapping occasions. With three trapping occasions, AIC and BIC generally chose the correct model when constant, time, and behaviour data was generated; when significant heterogeneity was modeled in the data, AIC and BIC also generally chose the correct model. In terms of the accuracy of estimates for N , when constant data was generated, all models (M_0 , M_t , M_b , and M_h) estimated N accurately on average. When time and behaviour data was generated, the model corresponding to the data estimated N accurately on average, whereas the models not corresponding to the data inaccurately estimated N (sometimes significantly). For heterogeneous data, it took four trapping occasions for model M_h to accurately estimate N , and the other models often did not accurately estimate N when fit to heterogeneous data. Therefore, in terms of estimating N , each model is robust to constant data (as expected since models M_t , M_b , and M_h are nested under model M_0), but no model is robust estimating N when fit to data generated by a different model.

Some work was done looking for a reason why AIC and BIC were incapable of distinguishing the time and behaviour full likelihood models for two trapping occasions. Recall that for these two models, they each have 3 parameters: for the time model the parameters are the capture probabilities on each trapping occasion and population size, and for the behaviour model the parameters are the capture and recapture probabilities as well as population size. So for both

AIC and BIC in this case, it is only the term $-2\ln(\mathcal{L})$ from equations (1) and (2) that differentiate the time and behavioural models, and so clearly it is only \mathcal{L} that differentiate them. From section 2.3, the two trapping occasion maximum likelihood for model M_t is

$$\mathcal{L}_t(N, p_1, p_2 \mid \text{data}) \propto \frac{N!}{n_{00}!} (p_1 p_2)^{n_{11}} (p_1 (1-p_2))^{n_{10}} ((1-p_1) p_2)^{n_{01}} ((1-p_1)(1-p_2))^{n_{00}}$$

and for model M_b the maximum likelihood is

$$\mathcal{L}_b(N, p, c \mid \text{data}) \propto \frac{N!}{n_{00}!} (pc)^{n_{11}} (p(1-c))^{n_{10}} ((1-p)p)^{n_{01}} ((1-p)(1-p))^{n_{00}}.$$

It might simply be the case that the likelihoods for these models are too similar to perform the AIC and BIC procedure for model selection.

One general trend seen was that when time data was generated with $p_1 < p_2$ or when behaviour data was generated with $p < c$, AIC and BIC would choose model M_t , and conversely when time data was generated with $p_1 > p_2$ or when behaviour data was generated with $p > c$, AIC and BIC would choose model M_b . Note that when time data is generated with $p_1 < p_2$ or when behaviour data is generated with $p < c$, this implies that $n_{10} < n_{01}$, conversely when time data is generated with $p_1 > p_2$ or when behaviour data is generated with $p > c$, this implies $n_{10} > n_{01}$. So this trend can be expressed as when $n_{10} < n_{01}$, AIC and BIC chose model M_t and when $n_{10} > n_{01}$, AIC and BIC chose model M_b . However this result did not always hold: when time data was generated with $p_1 = 0.8$, $p_2 = 0.46$, and $N = 10000$, AIC and BIC chose model M_t over 80% of the time. Also, it was not the case that when n_{10} was much smaller than n_{01} , AIC and BIC chose model M_t very confidently which might be expected if the proportion of n_{01} and n_{10} determined which model AIC and BIC chose. Also, I was unable to theoretically explain this phenomenon.

Since AIC and BIC are insufficient for performing model selection with two trapping occasions, we would like to give guidance for the end-users. I have a few points based on the results of this study.

1. *Think about the population and how the study was performed.* The experimenter should consider whether trap-happy or trap-shy behaviour could manifest in their experiment. Pollock et al. (1990) stated that if possible, biological information should be used to reduce the number of models considered in a study [12]. They give the example that evidence may exist, based on behavior of the population's member concerned, that trap response is unlikely with the trapping method used. In this case, the experimenter should eliminate any model containing behavioural variation. Additionally, White (2008) stated that behavioural response may not be an attribute of the animals, but of the survey configuration [16]; the experimenter should consider this before fitting the data to the behaviour model. This notion also applies for heterogeneity: think about what mixture groups might be present in your population before fitting the heterogeneity model. Also, as displayed in Figures 5a, 5b, and 5c, it

took four trapping occasions for model M_h to perform significantly better at estimating N from heterogeneous data on average compared to the other models; one should keep that in mind when fitting the heterogeneity model to data with two or three trapping occasions.

2. *Have three trapping occasions.* Although this may be impossible in a particular study, my results indicate that with three trapping occasions, generally if there are time or behaviour effects, or clear heterogeneity in the data, the AIC and BIC procedure of model selection will choose the model corresponding to these effects.
3. *Using the Lincoln-Petersen estimator.* If the experimenter performs a capture-recapture study with two trapping occasions and can rule out behavioural effects from the capture-recapture process as well as heterogeneity in their population of interest, it is appropriate to use the Lincoln-Petersen estimator. From my results, the time model is robust at estimating N for the constant model, and since you have ruled out the other models, the Lincoln-Petersen estimator should give you an accurate estimate of population abundance.
4. *Perform other methods of model selection.* Suppose you have capture-recapture data from two trapping occasions. If the time and behaviour models are giving significantly different estimates of N , the AIC and BIC procedures cannot be trusted to correctly determine which model is correct. So you would need to investigate other methods of model selection to determine the correct estimate.

5 Future work

More research is needed to find out how to accurately perform model selection with closed-population capture-recapture data with two trapping occasions. Especially since model M_t accurately estimates N when fit to time data but the other models do not, and since model M_b accurately estimates N when fit to behaviour data but the other models do not, if the correct model is detected, you can get an accurate estimate of N from a capture-recapture study with two trapping occasions. It would also be interesting to look into whether you can rule out candidate models which you can in theory fit to a dataset with two trapping occasions. More clearly, it might be unrealistic to assume that a set of 11s, 10s and 01s can encode such information as heterogeneity with two or more mixture groups, thus a researcher should never fit those models to a capture-recapture study with two trapping occasions.

Some research could be performed to determine how many trapping occasions are necessary for accurate model selection using AIC and BIC for the models M_{th} , M_{bh} , M_{tb} and M_{tbb} . Since from my research you need three

trapping occasions to accurately select models M_0 , M_t , M_b and M_h , the analysis of these other models would complete the set of eight standard closed-population capture-recapture models commonly used.

References

- [1] Hirotugu Akaike. A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike*, pages 215–222. Springer, 1974.
- [2] David R Anderson and Kenneth P Burnham. Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management*, pages 912–918, 2002.
- [3] Kenneth P Burnham, Gary C White, and David R Anderson. Model selection strategy in the analysis of capture-recapture data. *Biometrics*, pages 888–898, 1995.
- [4] Anne Chao, PK Tsay, Sheng-Hsiang Lin, Wen-Yi Shau, and Day-Yu Chao. The applications of capture-recapture models to epidemiological data. *Statistics in medicine*, 20(20):3123–3157, 2001.
- [5] E COOCH. Program mark, "a gentle introduction". <http://www.phidot.org/software/mark/docs/book/>, 2008.
- [6] RM Huggins. On the statistical analysis of capture experiments. *Biometrika*, 76(1):133–140, 1989.
- [7] RM Huggins. Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, pages 725–732, 1991.
- [8] Clifford M Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.
- [9] J.L. Laake. RMark: An r interface for analysis of capture-recapture data with MARK. AFSC Processed Rep. 2013-01, Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., Seattle, WA, 2013.
- [10] David L Otis, Kenneth P Burnham, Gary C White, and David R Anderson. Statistical inference from capture data on closed animal populations. *Wildlife monographs*, (62):3–135, 1978.
- [11] Shirley Pledger. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, 56(2):434–442, 2000.
- [12] Kenneth H Pollock, James D Nichols, Cavell Brownie, and James E Hines. Statistical inference for capture-recapture experiments. *Wildlife monographs*, pages 3–97, 1990.
- [13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [14] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

- [15] Thomas R Stanley and Kenneth P Burnham. Information-theoretic model selection and model averaging for closed-population capture-recapture studies. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 40(4):475–494, 1998.
- [16] Gary C White. Closed population estimation models and their extensions in program mark. *Environmental and Ecological Statistics*, 15(1):89–99, 2008.

A Additional plots

Figure 6: This plot has capture probabilities on the x -axis and the average estimate of N on the y -axis. The true population size is 1000 (compared to Figure 1 where the true abundance is 10000), and the estimates are from behaviour data with two trapping occasions. The probability space is all combinations of $p \in \{0.2, 0.5, 0.8\}$ and $c = \{z - 0.04, z - 0.03, z - 0.02, z - 0.01, z, z + 0.01, z + 0.02, z + 0.03, z + 0.04\}$ where $z \in \{0.2, 0.5, 0.8\}$.

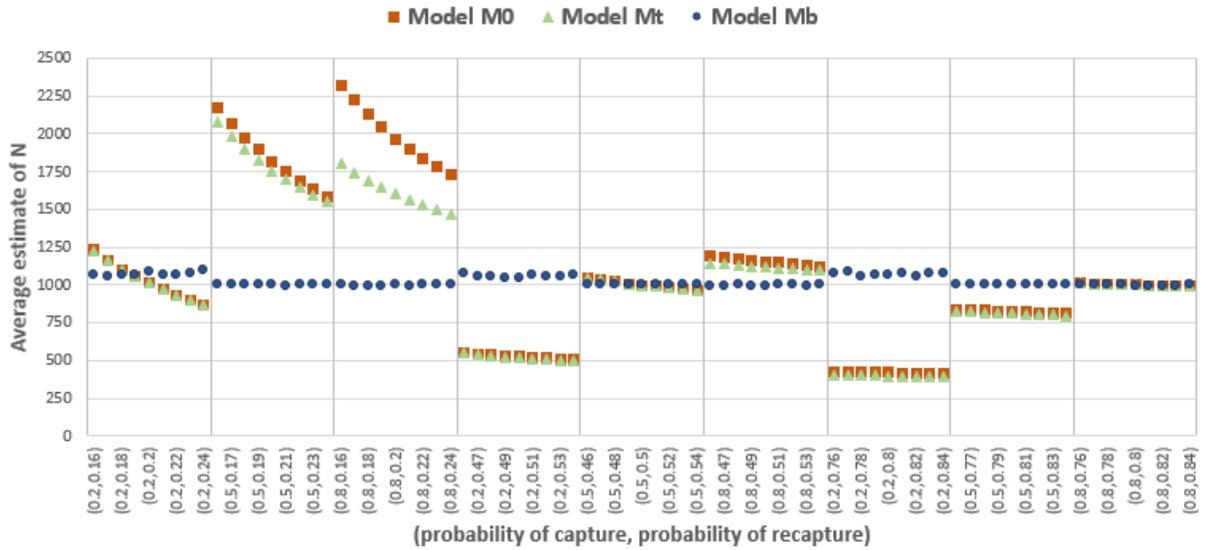


Figure 7: This plot has capture probabilities on the x -axis and the average estimate of N on the y -axis. The true population abundance is 10000, and the estimates are from behaviour data with three trapping occasions. The probability space is all combinations of $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and $c \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

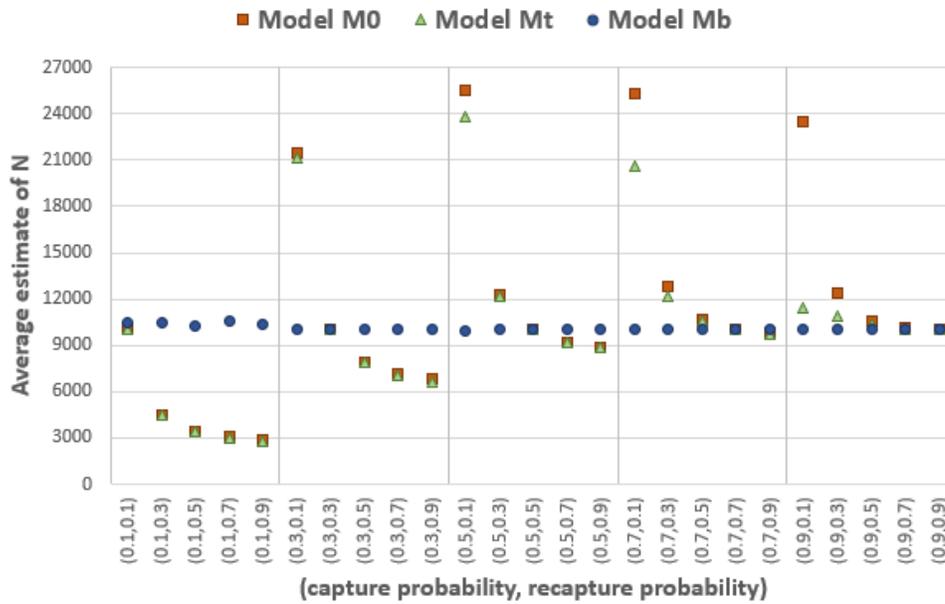


Figure 8: This plot has capture probabilities on the x -axis and the average estimate of N on the y -axis. The true population abundance is 10000, and the estimates are from behaviour data with two trapping occasions. The probability space is all combinations of $p, c \in \{0.05, 0.06, 0.07, 0.08, 0.09, 0.1\}$.

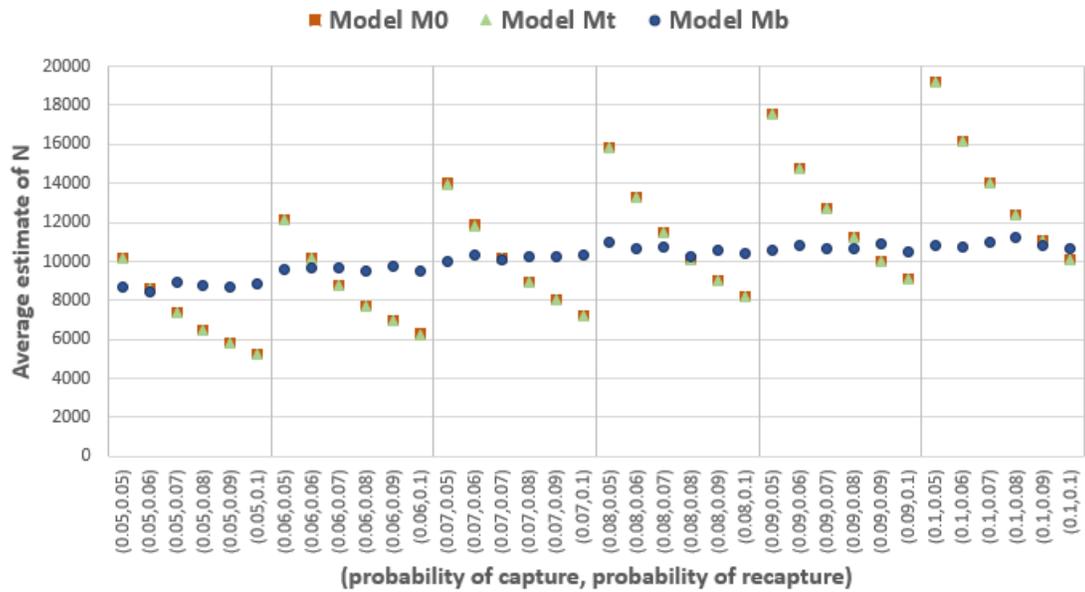


Figure 10: These plots are built from behaviour data with $N = 10000$ and three trapping occasions. The top plot displays the standard deviation of the estimates of N for each model, and the bottom plot displays the root-mean-square error of the estimates of N for each model. The probability space is all combinations of $p, c \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

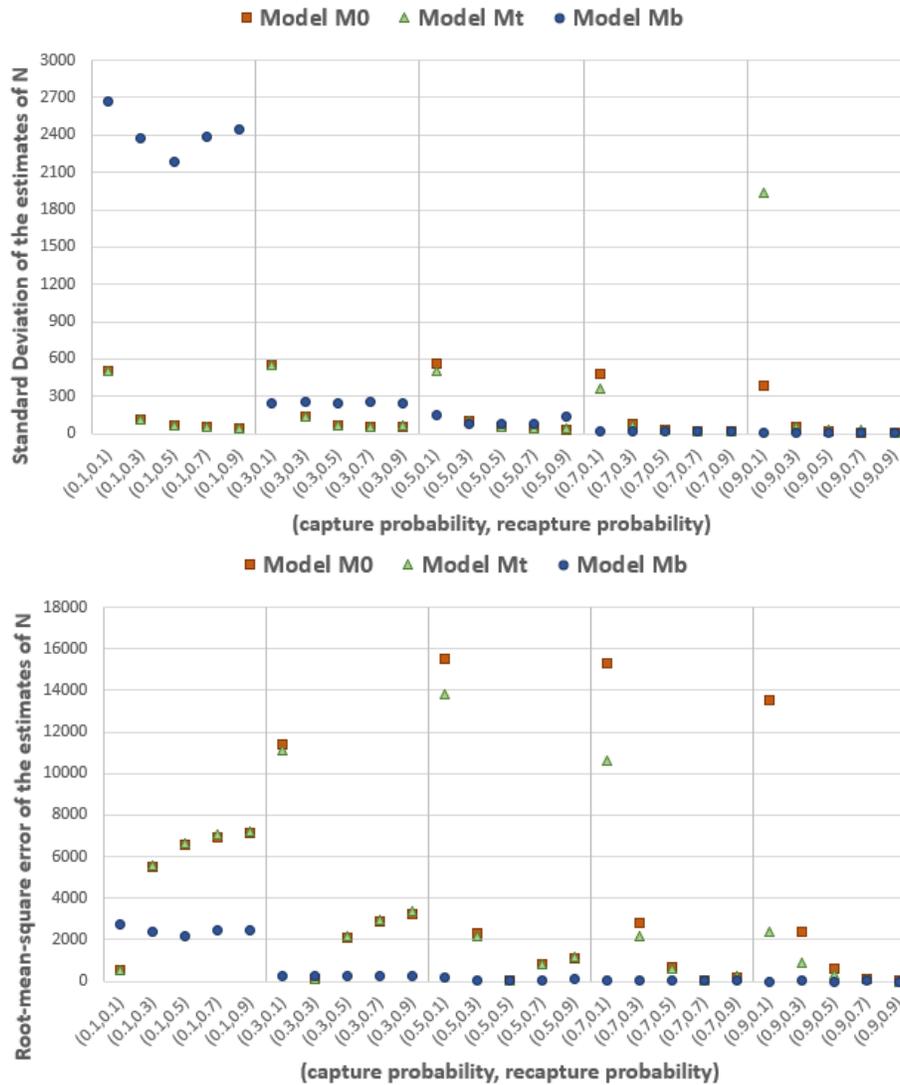


Figure 11: This plot has capture probabilities on the x -axis and the average estimate of N on the y -axis. These estimates are from time data with two trapping occasions. The true population size is 1000 (compared to Figure 3 where $N = 10000$). The probability space is all combinations of $p \in \{0.2, 0.5, 0.8\}$ and $c = \{z - 0.04, z - 0.03, z - 0.02, z - 0.01, z, z + 0.01, z + 0.02, z + 0.03, z + 0.04\}$ where $z \in \{0.2, 0.5, 0.8\}$. I limited the maximum estimate of N in this plot to be 3000 so it omits the behaviour model's very high estimates of N .

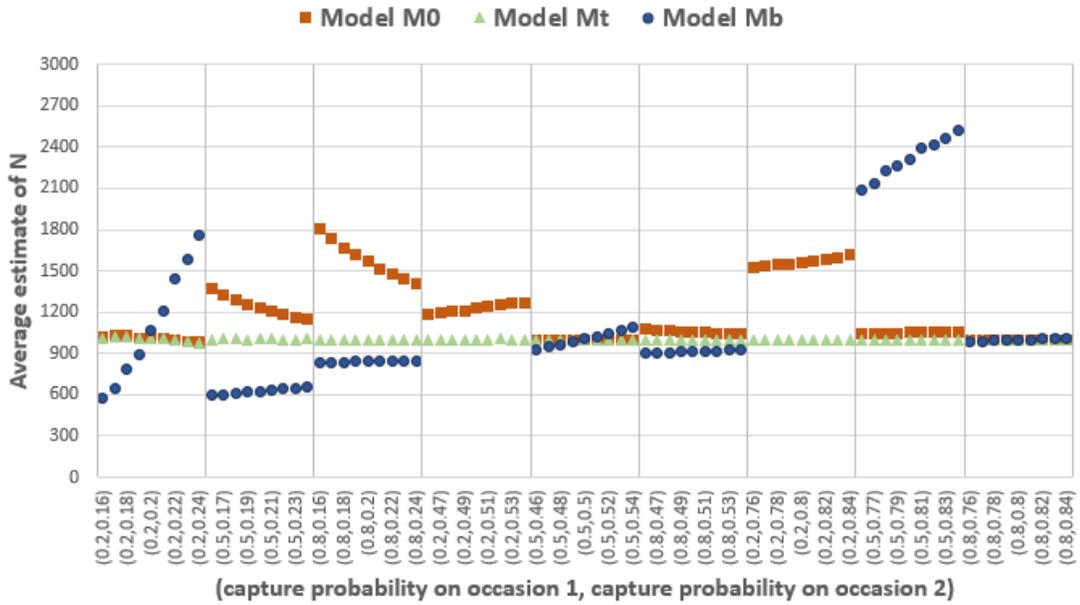


Figure 12: This plot has capture probabilities on the x -axis and the average estimate of N on the y -axis. These estimates are from time data with three trapping occasions. The true population size is 10000. The probability space is all combinations of $p_1, p_2, p_3 \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. I limited the maximum estimate of N in this plot to be 30000 so it omits the behaviour model's very high estimates of N . This plot is essentially the combination of twenty-five subplots, each subplot displays the results of five experiments. The probabilities p_1 and p_2 are fixed for each subplot (their values should be clear by looking at the x -axis labels), and probability p_3 ranges from 0.1 to 0.9 from left to right in each subplot.

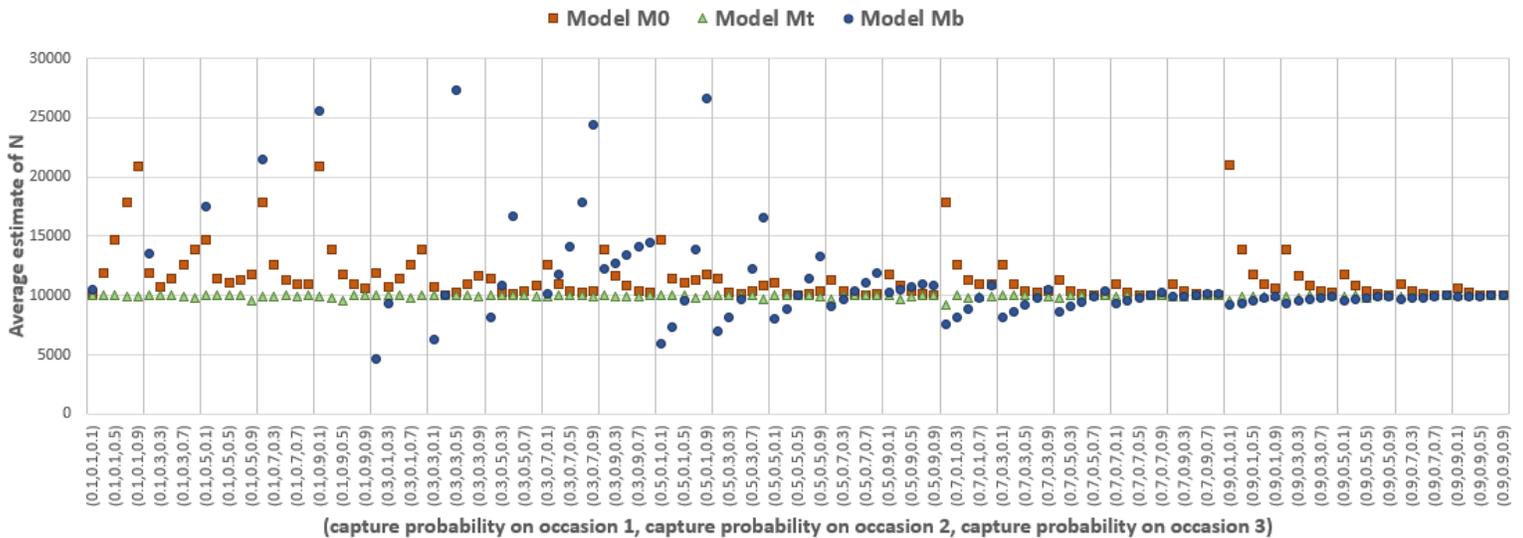


Figure 13: This plot has capture probabilities on the x -axis and the average estimate of N on the y -axis. The true population abundance is 10000, and the estimates are from time data with two trapping occasions. The probability space is all combinations of $p_1, p_2 \in \{0.05, 0.06, 0.07, 0.08, 0.09, 0.1\}$. I limited the maximum estimate of N to be 30000, so model M_b 's very high estimates are omitted from this plot.

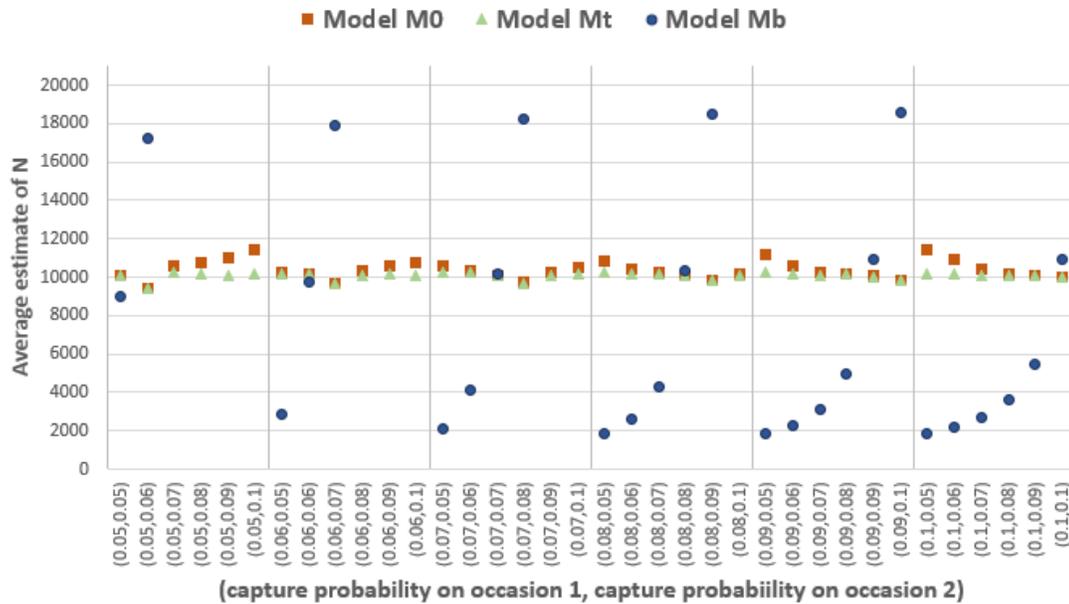


Figure 14: These plots are built from time data with $N = 1000$ and two trapping occasions (compared to Figure 4 where $N = 10000$). On the left is a chart matrix of the standard deviation of the estimates of N and on the right is a chart matrix of the root-mean-square error (RMSE) of the estimates of N . Capture probabilities for the second trapping occasion are listed on the x -axis of both charts, and this is plotted against standard deviation or RMSE on the y -axis. There were three capture probabilities for trapping occasion 1, $p_1 = 0.2, 0.5, 0.8$, and these are listed to the left of the three charts to which each capture probability corresponds. The extremely high standard deviations and root-mean-square errors from model M_b are omitted from these plots.

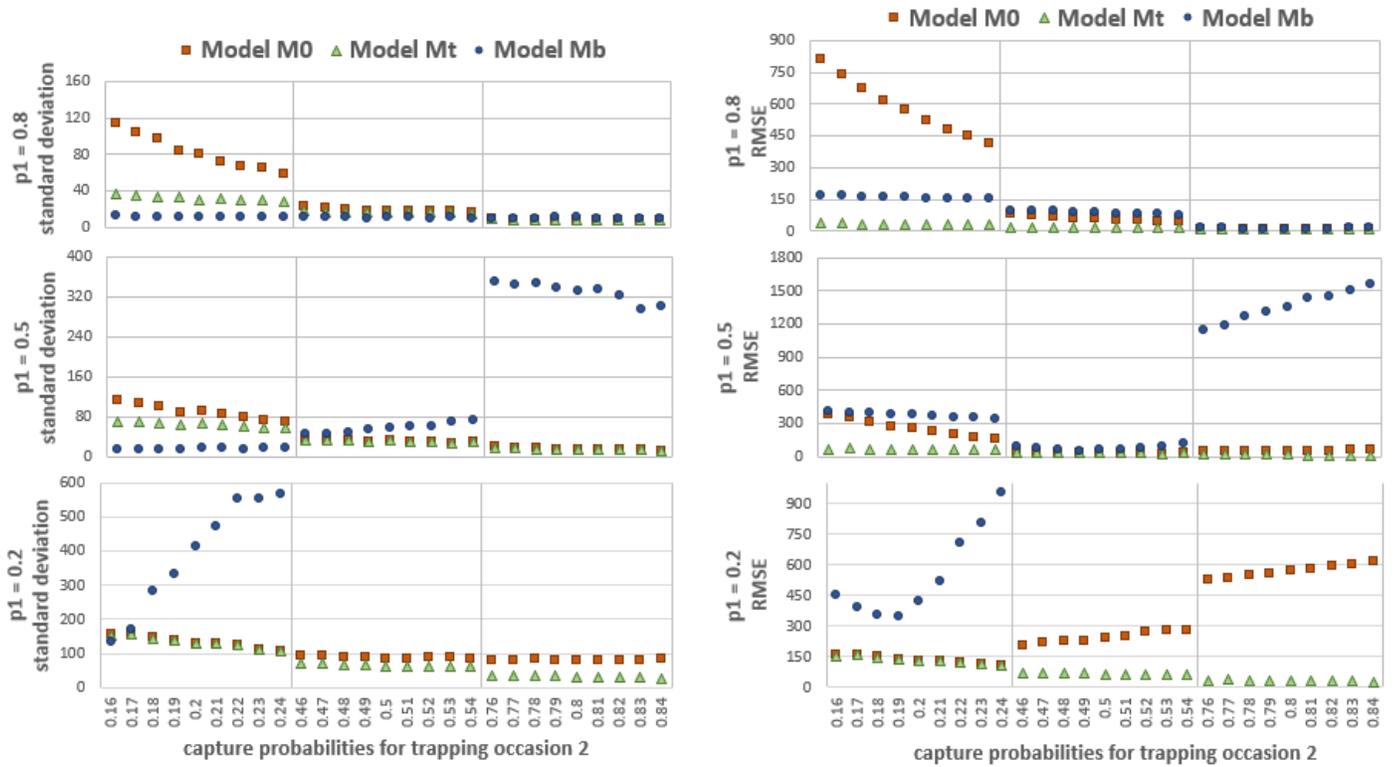
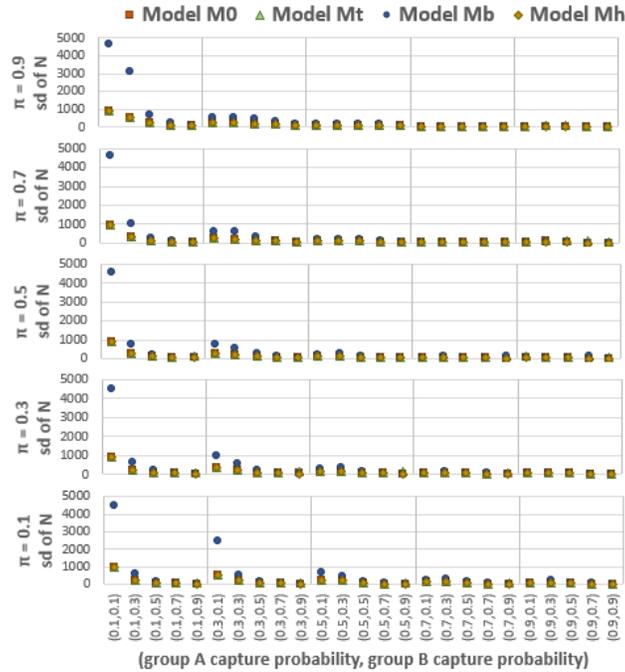
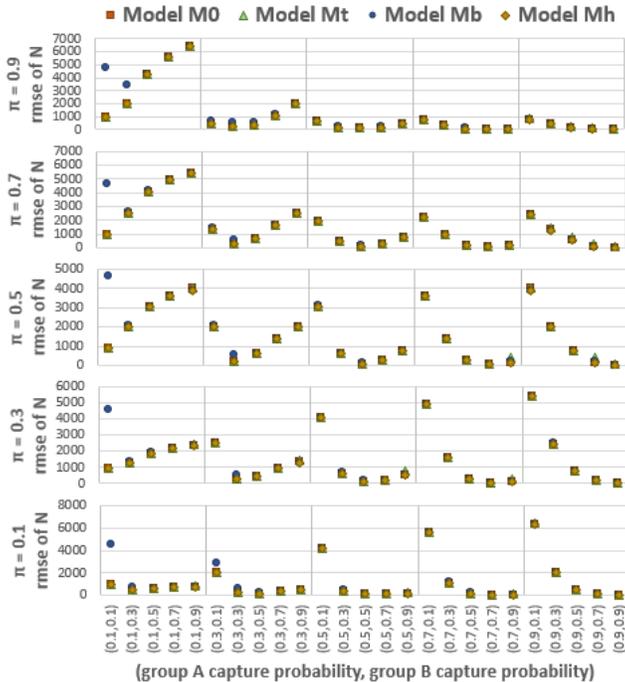


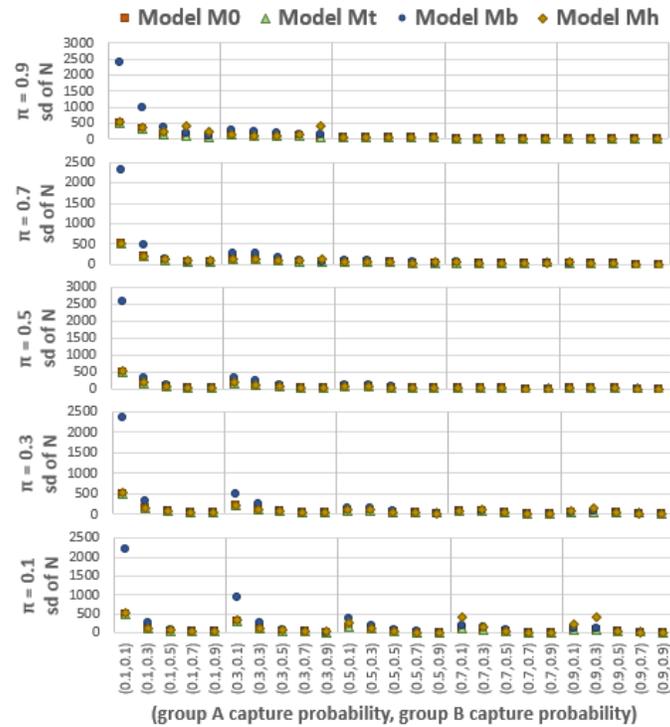
Figure 15: The following plots are from heterogeneous data with $N = 10000$. The plot on the top displays the standard deviation (sd) of the estimates of N and the plot on the bottom displays the root-mean-square error (RMSE) of the estimates of N . Each chart matrix plots all combinations of $\pi, p_A, p_B \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. On the x -axis are the capture probabilities p_A and p_B listed as an ordered pair, and they are plotted against standard deviation or root-mean-square error on the y -axis. Also, the π probability is listed to the left of the five subplots to which it corresponds.



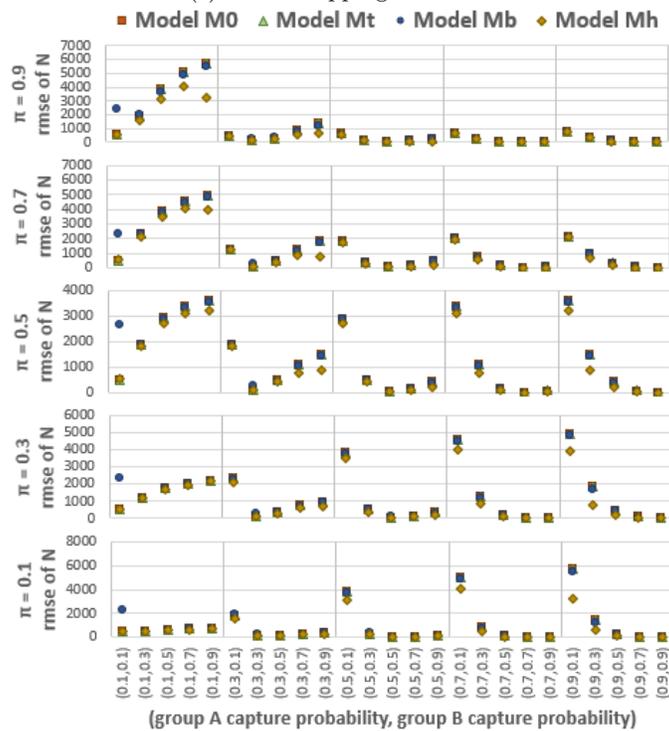
(a) Two trapping occasions.



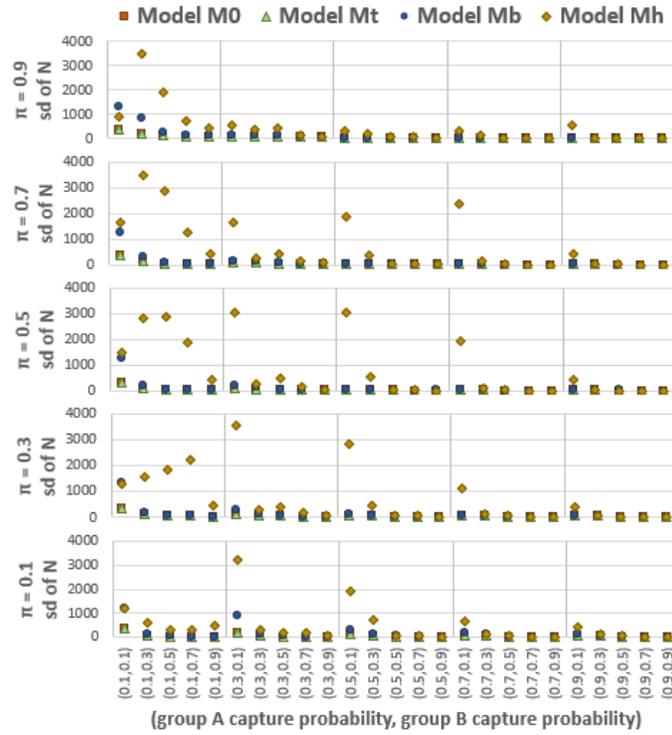
(b) Two trapping occasions.



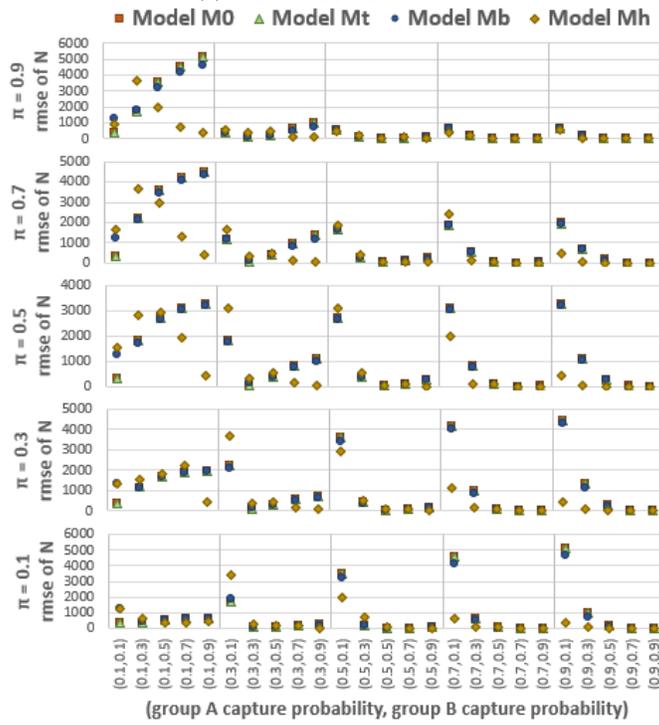
(c) Three trapping occasions.



(d) Three trapping occasions.



(e) Four trapping occasions.



(f) Four trapping occasions.

B Huggins same likelihood values example

This example comes from one simulation of behaviour data generated for two trapping occasions with $p = 0.8$, $c = 0.2$, and $N = 10000$. When the data was fit to the behaviour model, it estimated $\hat{p} = 0.79675$ and $\hat{c} = 0.202375$, and when the data was fit to the time model, it estimated $\hat{p}_1 = 0.4989214$ and $\hat{p}_2 = 0.202375$. Also, for this simulation $n_{11} = 1619$, $n_{10} = 6381$ and $n_{01} = 1626$. Section 2.3 introduced the Huggins likelihood model when capture probabilities could vary by time as

$$\mathcal{L}_{H,t}(p_1, p_2 \mid \text{data}) \propto \frac{(p_1(1-p_2))^{n_{10}}((1-p_1)p_2)^{n_{01}}(p_1p_2)^{n_{11}}}{(1-(1-p_1)(1-p_2))^{n_{11}+n_{10}+n_{01}}},$$

and the behaviour likelihood is defined similarly as

$$\mathcal{L}_{H,b}(p, c \mid \text{data}) \propto \frac{(p(1-c))^{n_{10}}((1-p)p)^{n_{01}}(pc)^{n_{11}}}{(1-(1-p)(1-p))^{n_{11}+n_{10}+n_{01}}}.$$

Plugging in the estimates and counts into the time model's likelihood gives

$$\begin{aligned} \mathcal{L}_{H,t} &\propto \frac{(0.4989214(1-0.202375))^{6381}((1-0.4989214)0.202375)^{1626}(0.4989214 * 0.202375)^{1619}}{(1-(1-0.4989214)(1-0.202375))^{1619+6381+1626}} \\ &= 2.499\,797\,820\,312\,541\,424\,742 \times 10^{-3649}, \end{aligned}$$

and plugging in the estimates and counts into the behaviour model's likelihood gives

$$\begin{aligned} \mathcal{L}_{H,b} &\propto \frac{(0.79675(1-0.202375))^{6381}((1-0.79675)0.79675)^{1626}(0.79675 * 0.202375)^{1619}}{(1-(1-0.79675)(1-0.79675))^{1619+6381+1626}} \\ &= 2.499\,797\,820\,320\,879\,847\,32 \times 10^{-3649}. \end{aligned}$$

These calculations were performed with a high-precision calculator. So obviously the log likelihood of these models will be the same, and since each model has two estimated parameters, the AIC and BIC values will be the same.