Penguin survival estimates from tag data using a multinomial likelihood

William Robinson[1] and Doug Butterworth

MARAM (Marine Resource Assessment and Management Group)
Department of Mathematics and Applied Mathematics,
University of Cape Town

October 2010

Abstract

A simple population model for Robben Island penguins is considered which incorporates fitting to both moult counts and tag data. The latter are now fit using a multinomial likelihood which is the method used in program MARK. Probability intervals on survival rates are more reliable than those obtained previously using an over-dispersed Poisson likelihood. Furthermore, the incorporation of a prior for relative undercounts of juveniles in the moult counts generally moves penalized likelihood estimates for adult survival rate away from an upper constraint boundary.

Introduction

This work is an extension of that previously presented (Robinson and Butterworth 2009, 2010, Butterworth and Robinson 2010a,b). Besides minor alterations to the model, this paper reports on the development of incorporating a multinomial likelihood for the tag data. Results are compared with the previous work where an over-dispersed Poisson approximation was used.

Population model

The number of female penguins $N_{y,a}$ in year y of age a at Robben Island is determined by the following equations:

$$N_{y+1,a} = \begin{cases} H_0 e^{\eta_y} \sum_{k=a^*}^{A} N_{y,k} & \text{if } a = 1 \\ N_{y,a-1} S_y & \text{if } 2 \leq a < A \\ \left(N_{y,a-1} + N_{y,a} \right) S_y & \text{if } a = A \end{cases}$$

where

H_0 is the average breeding success rate, incorporating fledging success and first year survival,

[1]william.robinson@uct.ac.za

η_y is the residual breeding success rate in year y,

a^* is the age at which the penguins first attempt to breed,

A is the plus-group age, and

$S_y = \dfrac{1}{1 + e^{-\tilde{S}_y}}$ is the adult survival rate in year y.

The initial population size and structure is

$$N_{y_0, a} = N_0 e^{-\lambda a}$$

where both N_0 and λ are parameters where values are estimated.

The population model is fitted to annual moult count data for both adult and juvenile birds using the negative log-likelihood functions:

$$-L_M = \sum_y \left[\ln \sigma_M + \frac{1}{2\sigma_M^2} \left(\ln \bar{N}_y - \ln \hat{N}_y \right)^2 \right]$$

$$-L_J = \sum_y \left[\ln \sigma_J + \frac{1}{2\sigma_J^2} \left(\ln J_y - \ln \hat{J}_y \right)^2 \right]$$

where

σ_M and σ_J are respectively the standard deviations of the logarithms of the adult moult counts and juvenile proportions about their true values,

$\bar{N}_y = \displaystyle\sum_{a=2}^{A} N_{y,a}$ is the number of female birds in adult plumage (aged 2 and over) in year y,

$J_y = \dfrac{p_J N_{y,1}}{N_{y,1} + \bar{N}_y}$ is the proportion of juvenile birds in the model, with p_J being the relative detectability of juvenile moulters,

$\hat{N}_y$ is the number of adult moulters observed in year y, and

$\hat{J}_y$ is the observed proportion of moulters in immature plumage counted in year y.

A penalty term for the estimated breeding success residuals

$$P_\eta = \frac{1}{2\sigma_\eta} \sum_y \eta_y^2$$

is added to the likelihood, where σ_η is the standard deviation of the η_y residuals.

The values of the quantities held constant in the model are given in

Table 1. A list of the parameters which are estimated is given in Table 2, along with their priors for Bayesian estimation.

In earlier model versions, the detectability of juvenile moulters and adult moulters was assumed to be equal. Including the estimable parameter p_J with uniform prior [0.5, 0.9] was found to have a strong influence on the key population parameters of interest here, namely adult survival and breeding success (see also Butterworth and Robinson 2010).

Tag-recapture analysis

Previously, annual penguin survival estimates were derived from tagging data assuming an over-dispersed Poisson error distribution. Here, instead, the multinomial likelihood of the encounter histories is calculated. This is the method used in program MARK. See Lebreton *et al*. (1992) for an overview of how survival is estimated in this manner using tag-recapture data.

The tag-recapture data used consisted of the live sighting histories for the period 2001–2008 of 6484 birds banded as adults and labelled as belonging to the Robben Island colony.

The log-likelihood function for the multinomial distribution, ignoring the constant multinomial coefficient, is:

$$\ln L_{\text{T-R}} = \frac{1}{\hat{c}} \sum_{j=1}^{m} n_j \ln \hat{p}_j$$

where $\hat{p}_j$ is the estimated probability of the j th encounter history occurring and n_j is the number of times which it has been observed. The number of unique encounter histories is m. In order to account for over-dispersion, the likelihood is scaled by a factor of $1/\hat{c}$, which was obtained from analysing the data using MARK.

The encounter history probabilities $\hat{p}_j$ are made up of combinations of factors involving the probability p_y of sighting the bird in year y and the probability ϕ_{y-1} that a bird survives from year $y-1$ to year y. The logit link function relates these p_y and ϕ_y probabilities to the β parameters which are estimated. For example:

$$\phi_i = \frac{1}{1 + e^{-\beta_i}}$$

Results

The following five cases were considered:

- A. Poisson likelihood for tag data, no juvenile relative detectability factor.
- B. Poisson likelihood for tag data, juvenile relative detectability p_J estimated.
- C. Multinomial likelihood for tag data, no adjustment for over-dispersion, p_J estimated.

D. Multinomial likelihood for tag data, adjusted for over-dispersion, p_J estimated.

E. No tag data included, p_J estimated.

In Figure 1 to Figure 4, model output time series of posterior median values and Bayesian 90% probability intervals are plotted. Black squares indicate the corresponding maximum (penalized) likelihood estimates (posterior modes), except in Figure 4, where they indicate the observed counts. The figures are arranged as follows:

Figure 1: Reproductive success time series for models A, B and D.

Figure 2: Adult survival time series for models A and B.

Figure 3: Adult survival for models C, D and E.

Figure 4: Adult female population trajectories for models B and D.

Figure 5: Prior and posterior distributions of relative juvenile detectability in models B and D.

Figure 6: Time series of posterior medians of reproductive success, adult survival and female population numbers for all the models plotted together.

Discussion

Interesting aspects of the results are:

- Higher and less precise estimates of annual reproductive success when the prior on relative detectability of juveniles in the moult count is introduced (Figure 1). However, these do suggest high reproductive success – typically one female chick and hence two chicks overall each year.
- A lower adult survival rate given that prior, plus maximum (penalized) likelihood estimates generally no longer on the constraint boundary.
- Less precise estimates for that survival rate (and also population estimates) over the 2001–2006 period for the multinomial likelihood than the Poisson. (The latter is an approximation; the former is more reliable.)
- Little difference amongst Bayesian posterior median estimates for reproductive success, adult survival and the female population except the case without allowance for lesser detectability of juveniles in the moult count.

References

Butterworth DS, Robinson W. 2010a. Response to Comments on : Robben Island penguin survival rates from Bayesian analysis by William Robinson and Doug Butterworth (Altwegg: MCM/2010/SWG-PEL/08). Document MCM/2010/SQG-PEL/09.

Butterworth DS, Robinson W. 2010b. Further responses to comments in Altwegg MCM/2010/SWG-PEL/23 regarding penguin analyses. Document MCM/2010/SQG-PEL/27.

Lebreton J, Burnham KP, Clobert J, Anderson D. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Monographs 62, 67-118.

Robinson W, Butterworth DS. 2009. Fitting both moult counts and tagging data to a population model for Robben Island penguins. Document MCM/2009/SWG-PEL/33.

Robinson W, Butterworth DS. 2010. Robben Island penguin survival rates from Bayesian analysis. Document MCM/2010/SQG-PEL/07.

Tables and figures

Table 1: Penguin population model constants and values

Constant	Symbol	Value
Plus-group age	A	10
Age of first breeding attempt	a^*	4
Standard deviation of logged adult moult counts	σ_M	0.2
Standard deviation of logged juvenile proportions	σ_J	0.2
Standard deviation of breeding success residuals	σ_η	0.5

Table 2: Parameters which are estimated. Note that $\tilde{S}_y$ is not estimated for years with tag data, except for the final two years, since the survival estimates for these years come from the β parameters.

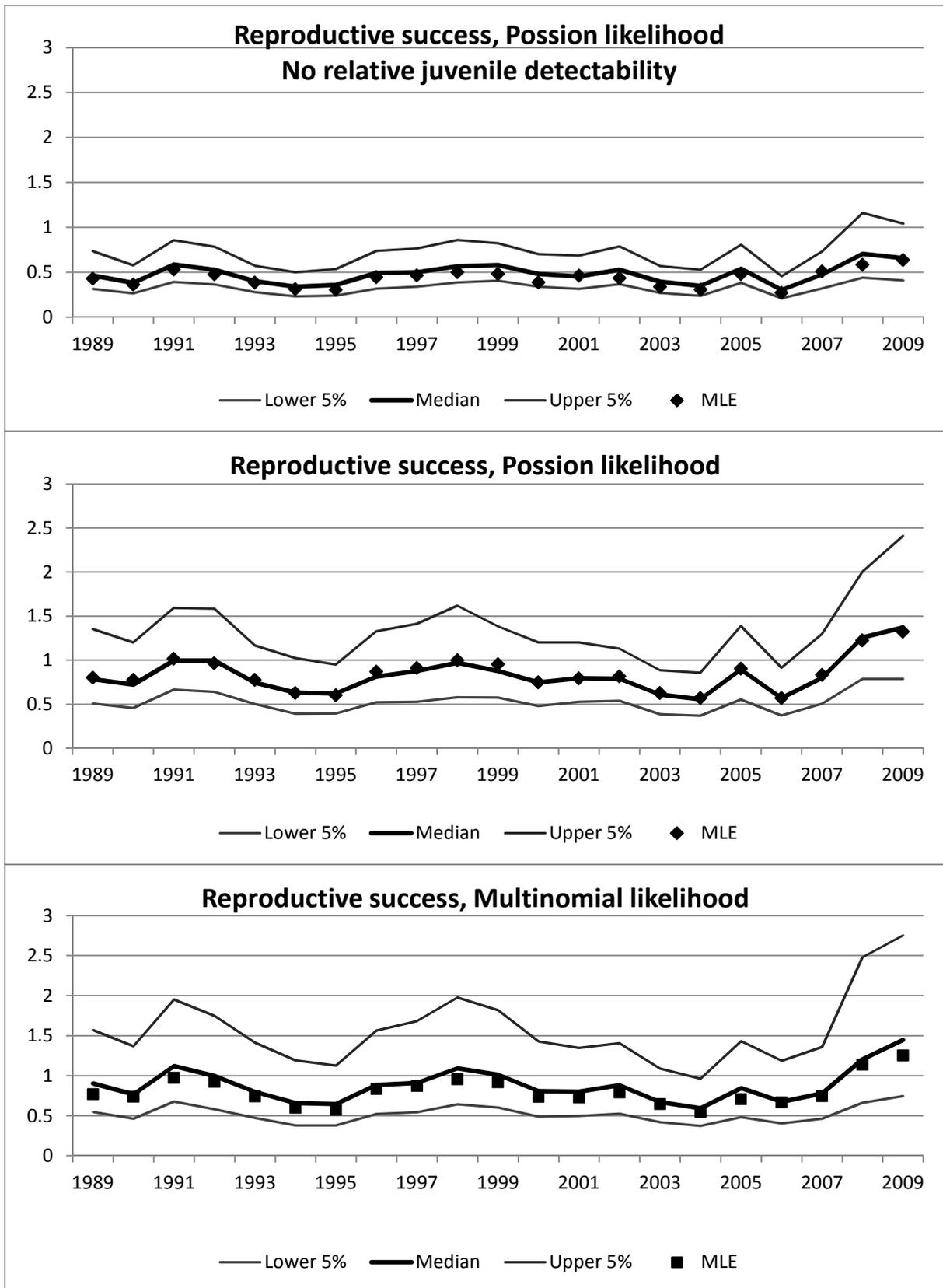| Parameter	Symbol	Prior
Initial population	$\ln N_0$	U[1, 10]
	λ	U[0, 3]
Average reproductive success	H_0	U[0, 2]
Relative juvenile detectability	p_J	U[0.5, 0.9]
Annual random effects in reproductive success	η_y	U[-4, 4]
Annual adult survival in logit space	$\tilde{S}_y$	$U\left[\ln\left(\frac{0.04}{1-0.04}\right), \ln\left(\frac{0.96}{1-0.96}\right)\right]$
Capture history probability components in logit space	β_i	$U\left[\ln\left(\frac{0.04}{1-0.04}\right), \ln\left(\frac{0.96}{1-0.96}\right)\right]$

Figure 1: Reproductive success time series of maximum likelihood estimates and Bayesian posterior medians and 90% probability intervals.
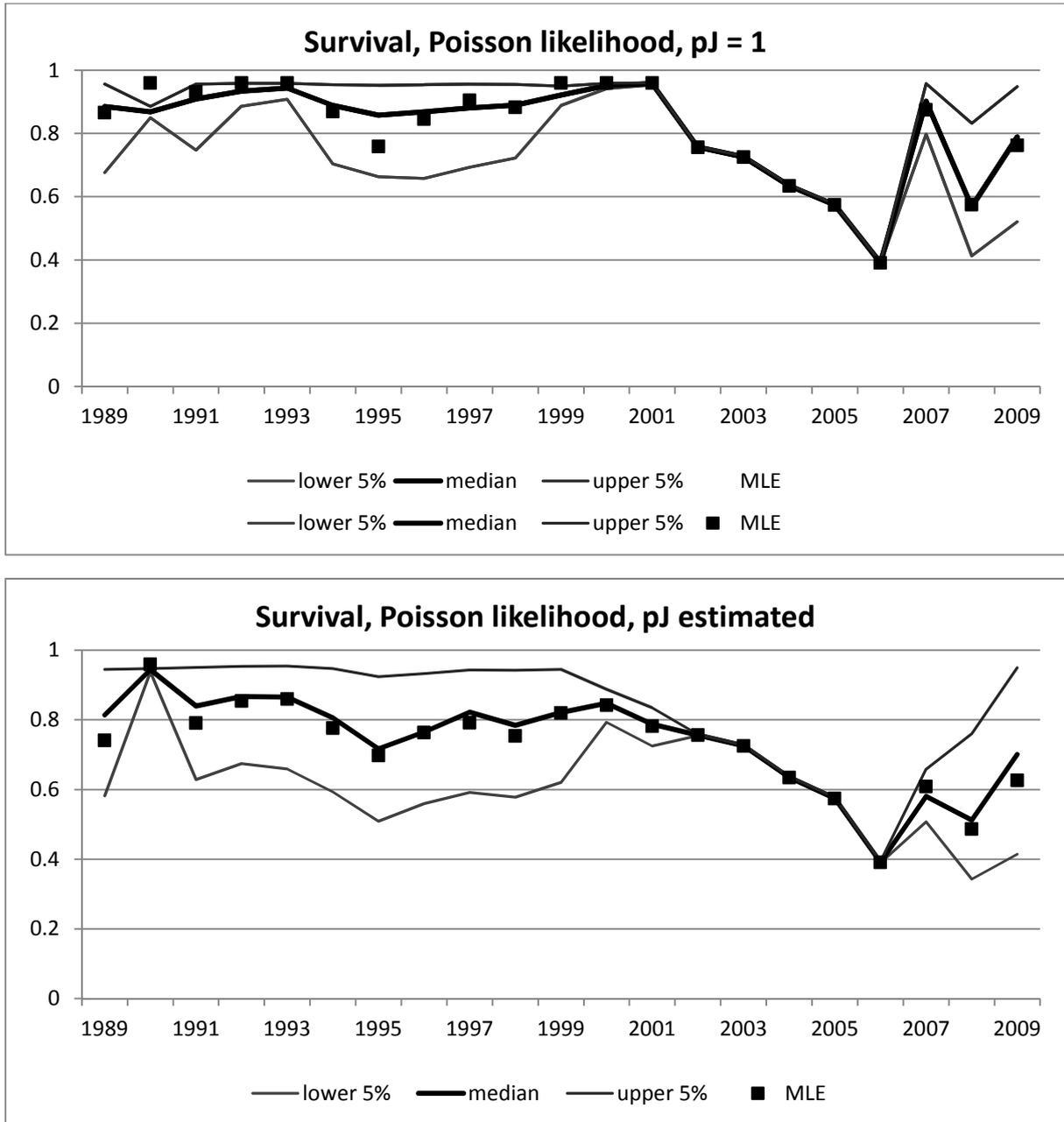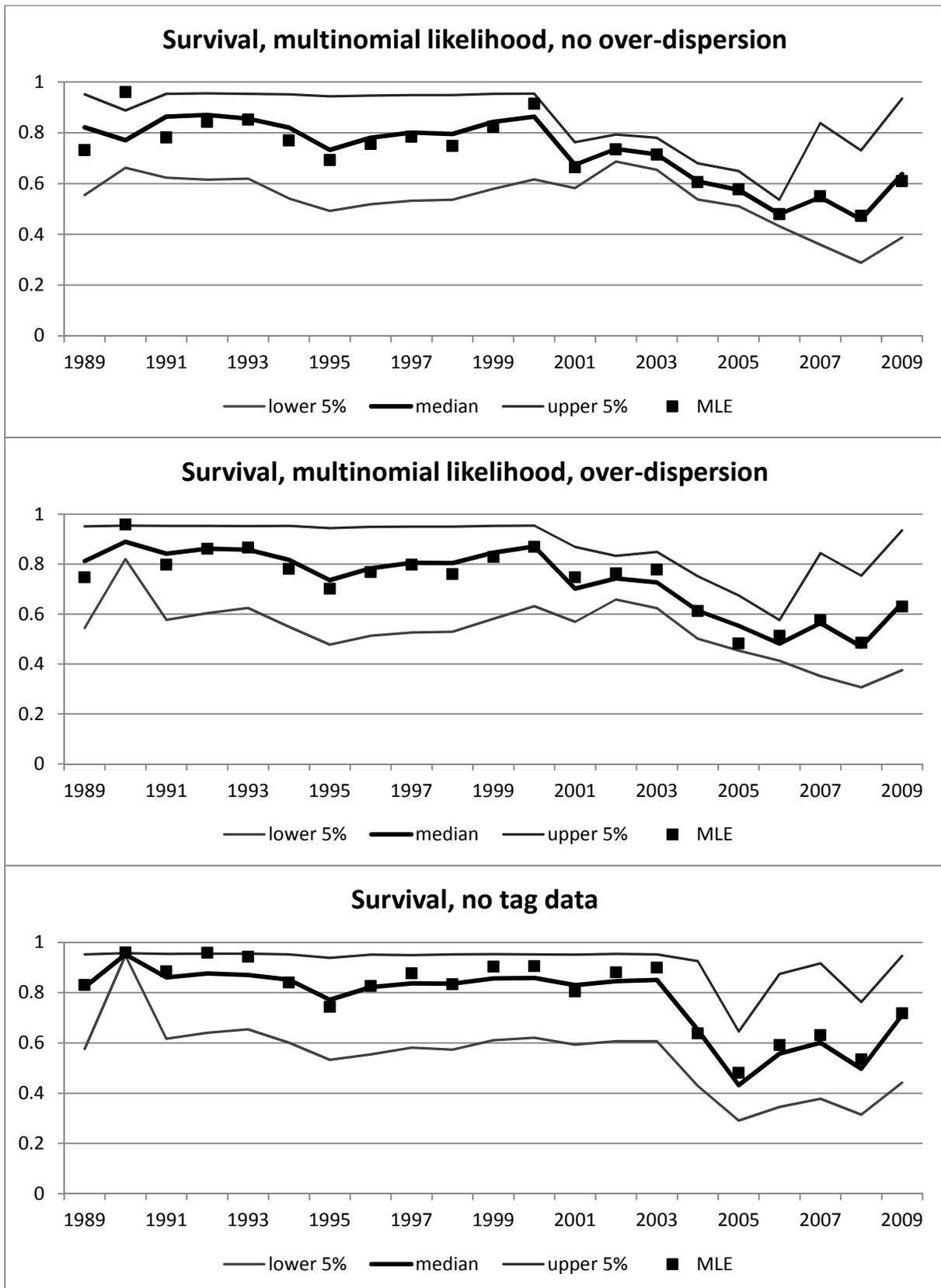
Figure 2: Time series of (penalized) maximum likelihood estimates and Bayesian posterior medians and 90% probability intervals of annual adult survival rates calculated using a Poisson likelihood formulation for the tag data. In the top plot, adult and juvenile moulters are assumed to be equally detectable, while below, the relative detectability is estimated.

Figure 3: Time series of (penalized) maximum likelihood estimates and Bayesian posterior medians and 90% probability intervals of annual adult survival rates. For the top and middle plots, a multinomial likelihood formulation is used for the tag data. The over-dispersion coefficient reduces the information content of the tag data in the middle plot.
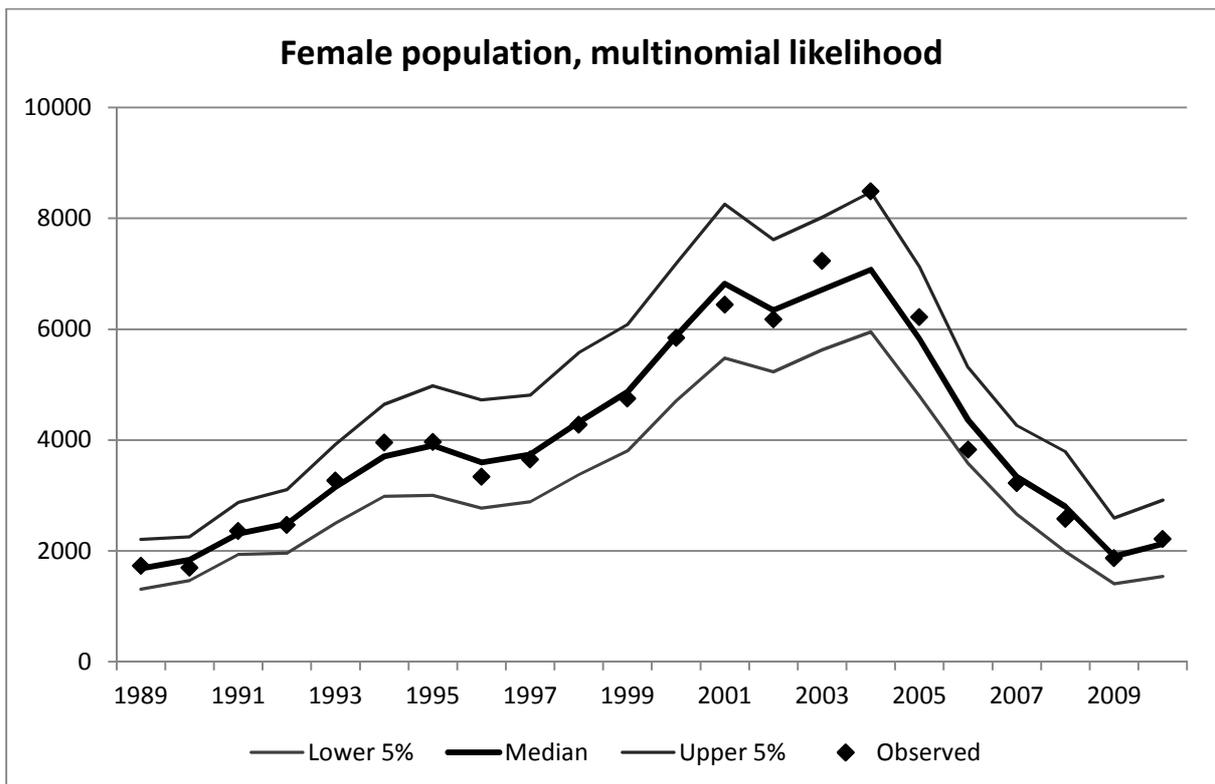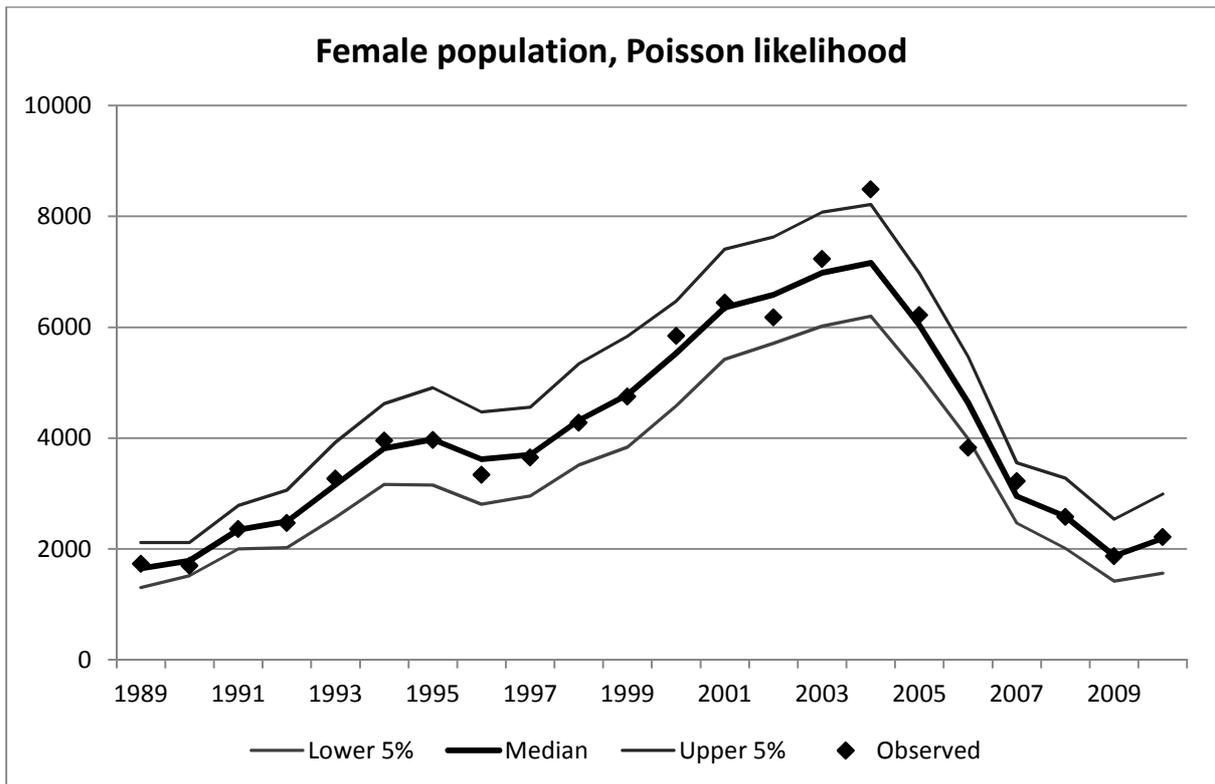
Figure 4: Time series of median population estimates and Bayesian posterior medians and 90% probability intervals when including tag data with a Poisson likelihood (top) and a multinomial likelihood (bottom). Relative juvenile detectability and over-dispersion are included.
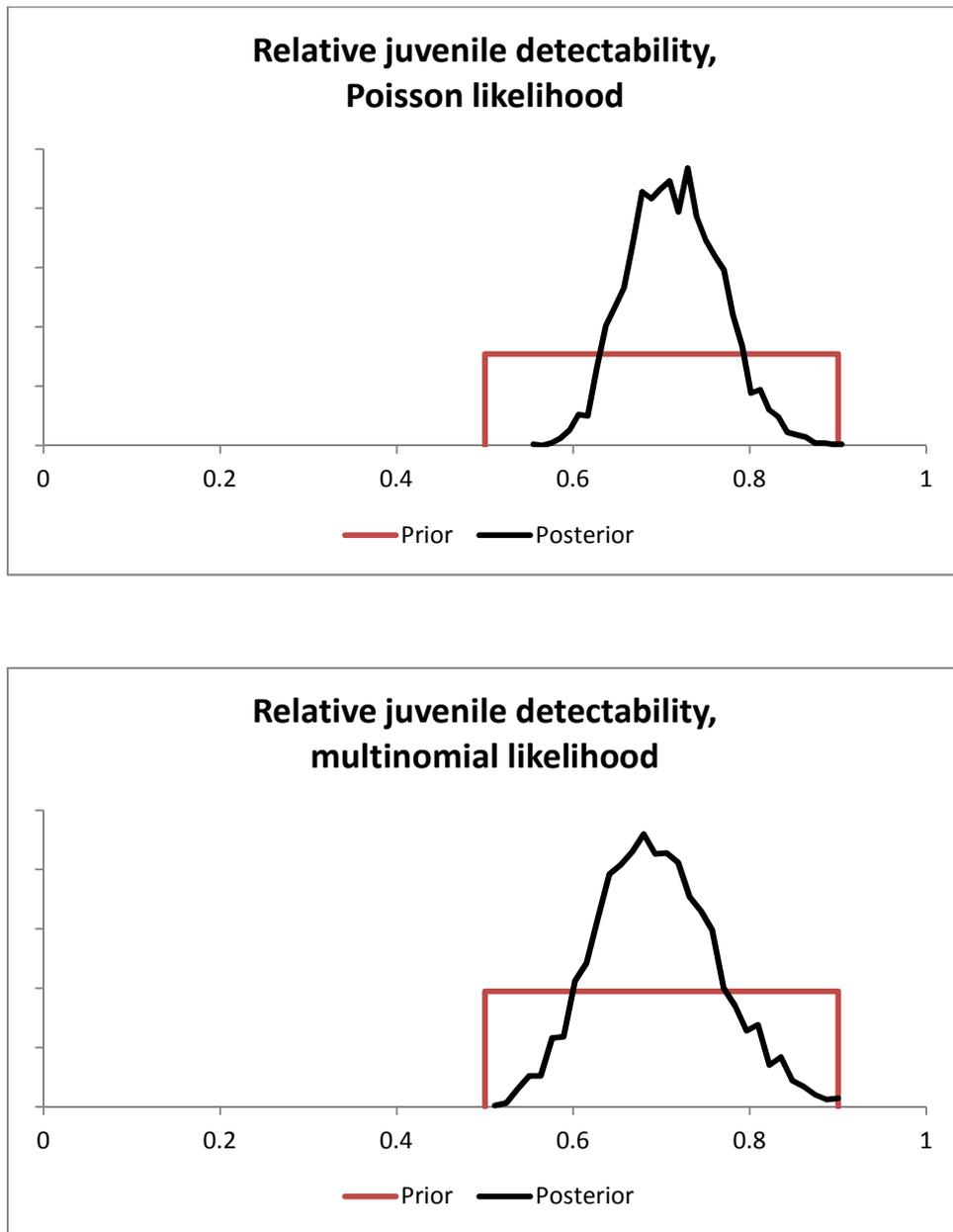
Figure 5: The [0.5, 0.9] prior on the relative detectability of juveniles is updated by the data in both the Poisson likelihood model B (top) and the multinomial likelihood model D (bottom).
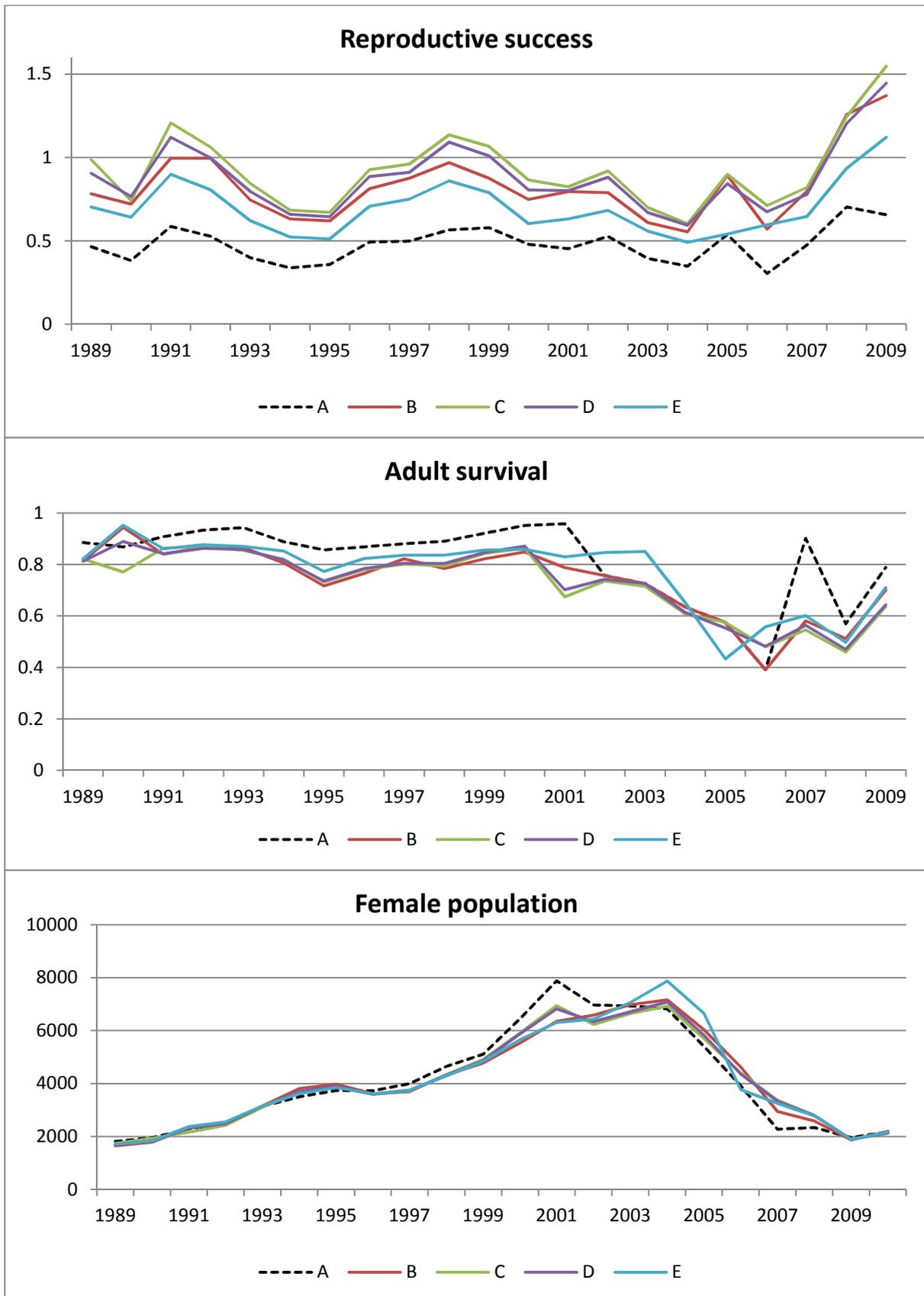
Figure 6: Comparison of posterior median reproductive success, adult survival and female population size among the five model variants. The dashed lines indicate the trajectories of the model without the relative juvenile detectability factor. The Model A trajectory is traced with a dashed line.