Routledge
Taylor & Francis Group

Measuring density dependence in survival from mark-recapture data

RICHARD BARKER[1]**, DAVID FLETCHER**[1] **& PAUL SCOFIELD**[2]**,**
[1]*Department of Mathematics and Statistics, University of Otago, Box 56, Dunedin, New Zealand and* [2]*Department of Zoology, University of Otago, Box 56, Dunedin, New Zealand*

ABSTRACT   *We discuss the analysis of mark-recapture data when the aim is to quantify density dependence between survival rate and abundance. We describe an analysis for a random effects model that includes a linear relationship between abundance and survival using an errors-in-variables regression estimator with analytical adjustment for approximate bias. The analysis is illustrated using data from short-tailed shearwaters banded for 48 consecutive years at Fisher Island, Tasmania, and Hutton's shearwater banded at Kaikoura, New Zealand for nine consecutive years. The Fisher Island data provided no evidence of a density dependence relationship between abundance and survival, and confidence interval widths rule out anything but small density dependent effects. The Hutton's shearwater data were equivocal with the analysis unable to rule out anything but a very strong density dependent relationship between survival and abundance.*

1 Introduction

1.1 General background

Mark-recapture models are useful for summarizing encounter history data using population parameters such as survival and birth rates or abundance. The associated theory is well developed with models suitable for a wide variety of sampling situations. Computer software such as program MARK (White & Burnham, 1999) can be used to fit these models encompassing a large variety of mark-recapture

data types within a common analysis framework. Studies using these models tend to focus on survival rates. More recently, models have been developed that allow us to focus attention on recruitment rates (Pradel, 1996) and probabilities of moving between states (Brownie *et al.*, 1993).

A statistical model can be thought of as a way of summarizing data using a small number of parameters. In an example we discuss below, a mark-recapture study was carried out with 48 sampling occasions and 1045 unique encounter histories. The full Jolly-Seber model has 141 estimable parameters, and although this is considerably fewer than the 1045 encounter histories, it hardly represents a compact summary of the data. Moreover, the 46 identifiable survival rates in the Jolly-Seber model offer a simple description of the survival process operating over time, but provide little insight into the underlying biological processes. What is needed in this example is a model for the survival probabilities that incorporates key structural features of the biological processes.

1.2 Random effects models

To focus our attention on interesting biological processes, we can think of mark-recapture data as arising through a two-stage process. In stage I, the true abundances, birth rates and survival rates can be thought of as being sampled from some distribution that depends on a small set of parameters. The realized (but unobserved) values for these can be regarded as fixed parameters at stage II. In stage II, individual encounter histories are obtained by sampling the animal population that exists during sampling. There may be considerable interest in stage II sampling, but matters of real biological interest are in stage I.

This hierarchical sampling process places us in the realm of 'random effects models', in which we envisage a sampling distribution for the abundances, births and survivals at stage I. We can hypothesize that there is a relationship between some of these parameters, for example an effect of abundance on survival rate. We do not expect that variation in the survival rates over time will be wholly determined by abundance. Other influences such as weather are also likely to be important. If we have measured these other influences then we can specify their effect in the model. However, there will inevitably be unexplained sources of variation but if we are prepared to assume that these unexplained departures from the model are random, then their combined influence can be expressed by including a random term in the model.

Random effect mark-recapture models have been developed by Burnham (in press) and implemented in program MARK. These models are relatively simple and at present restricted to the case where the vector of parameters is sampled independently from a common distribution. These models are useful where a simple summary such as an average or trend is sought, or where the study has been conducted within the context of a simple experimental design. The method-of-moments estimators in MARK are simple and allow the user to avoid assuming a particular distribution at stage I. Random-effects models and Bayesian inference are closely related and share the same methodological problems. In recent years there has been a resurgence of interest in Bayesian methods, largely prompted by the development of new analysis methods such as Markov chain Monte Carlo (MCMC) (Gelfand & Smith, 1990; Zeger & Karim, 1991; Smith & Gelfand, 1992), that make fitting the necessary models feasible.

1.3 Density dependence

Density dependence is an important ecological concept that has particular relevance to the management of exploited populations. Our interest in density-dependence is motivated by an exploited 'titi', or sooty shearwater, (*Puffinus griseus*) population in the southern part of the South Island, New Zealand. The Rakiura Maori have harvested titi nestlings for at least 300 years. Recently, a research programme has been established that aims to ensure that titi harvest is sustainable. Simple population models suggest that population growth rate is likely to be sensitive to changes in survival, age at first breeding and breeding success. However, if these parameters are density-dependent, the titi population may be able to compensate for losses due to harvest. Such a compensatory mechanism is a central theme of harvest literature (Robson & Youngs, 1971) and the main scientific justification for wildlife harvest.

In trying to measure the strength of any density dependent relationship between survival and abundance, the form of the relationship will be important. For example, a linear relationship between abundance and survival (or suitably trans-formed survival) might be hypothesized over the range of abundances being considered. Alternatively, density dependence might not operate until abundance exceeds a particular level, in which case a non-linear relationship would be appropriate.

Note that we are assuming that abundance of the population under study is a measure of density. It is possible that a more general measure of density could be more appropriate. For titi, the number of breeding birds per square metre might be more important, or the density of non-breeders.

As the term suggests, density dependence implies that population parameters are a function of animal density or abundance. Importantly, if the relationship is between abundance and survival then density-dependence implies that the density function for survival rate conditional on abundance is not the same as the marginal density for survival rate.

1.4 Fitting density dependence models

Given long-term data sets with relatively constant effort we argue that it is possible to investigate finer-scale population processes such as the effect of density on the population. This idea is not new. In their monograph Lebreton *et al.* (1992) briefly discuss the idea of using density as a covariate and give some examples.

A full likelihood approach to modelling density dependence requires a distribution to be specified for each of the two stages described above. Let N_i denote the abundance at time i, S_i the survival rate from time i until time $i+1$, and β_i the birth rate at time i. At stage I, we assume that the abundances, survival rates, and birth rates are sampled from a distribution $g(\{N_i\}, \{S_i\}, \{\beta_i\}; \mu)$ with parameters represented by μ and cumulative distribution function $G(\{N_i\}, \{S_i\}, \{\beta_i\}; \mu)$. Stage II involves modelling the mark-recapture data represented by $\mathbf{x}$ as being a sample from the distribution $f(\mathbf{x}|\{N_i\}, \{S_i\}, \{\beta_i\})$. Estimation is carried out using the distribution $h(\mathbf{x}; \mu)$ which is found by integration (Link, 1999):

$$h(\mathbf{x};\mu) = \int f(\mathbf{x}|\{N_i\}, \{S_i\}, \{\beta_i\}) \, \mathrm{d}\, G(\{N_i\}, \{S_i\}, \{\beta_i\}; \mu) \tag{1}$$

Analytical evaluation of the integral in (1) is problematic for mark-recapture models (Burnham, in press), although the relatively recent development of MCMC

methods means that finding $h(\mathbf{x}; \mu)$ by computer intensive means is becoming increasingly feasible. In addition to the structural elements specifying model density, a distribution must also be specified at the first stage of sampling in order to use MCMC methods.

In the analysis discussed below we concentrate on density-dependent survival and avoid specifying a full model for stage I sampling. Instead we only specify the structural form of the relationship between survival and abundance and estimate model parameters using the method of moments.

2 Methods

2.1 Estimating density dependence by linear regression

We assume that from the mark-recapture analysis we obtain estimates of $\hat{S}_i | S_i$ which are random variables with mean S_i and covariance given by $\text{Cov}(\hat{S}_i | S_i, \hat{S}_j | S_j) = \omega_{ij}^2$. From the stage I process we assume that $\mathbf{S} = (S_1, \ldots, S_k)'$ (or some transformation of $\mathbf{S}$) is a random vector with mean $\mathbf{X}\beta$ and variance-covariance matrix $\sigma^2 I$, where row i of the matrix $\mathbf{X}$ is given by $\mathbf{x}_i = (1 \ N_i)$. Unconditionally, $\hat{Y} = (\hat{S}_1, \ldots, \hat{S}_k)'$ is a random vector with mean vector $X\beta$ and variance-covariance matrix $D = \sigma^2 I + W$ where

$$W = \begin{bmatrix} \omega_{11}^2 & \omega_{12}^2 & \ldots & \omega_{k}^2 \\ \omega_{21}^2 & \omega_{22}^2 & \ldots & \omega_{2k}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{k1}^2 & \omega_{k2}^2 & \ldots & \omega_{kk}^2 \end{bmatrix}$$

If we could determine N_i exactly the obvious estimator would be: $\hat{\beta} = (X'\hat{D}^{-1}X)^{-1}X'\hat{D}^{-1}\hat{Y}$, obtained by conditioning on values of $\hat{\omega}_{ij}^2$ from the mark-recapture study and using an iterative procedure to estimate β and σ^2. There are several problems with this approach to modelling a density-dependent relationship between N_i and S_i. First, we cannot determine N_i, but instead must use $\hat{N}_i$, which we hope is a good estimate of N_i. If we replace X by $\hat{X}$ and use the estimator $\hat{\beta} = (\hat{X}'\hat{D}^{-1}\hat{X})^{-1}\hat{X}'\hat{D}^{-1}\hat{Y}$, we have an errors-in-variable problem. Secondly, in mark-recapture problems there is a sampling covariance between estimated abundances and survival rates that must be taken into account. Thirdly, the elements of the matrix W are not known exactly, but instead are estimated. Provided the mark-recapture study is based on a large sample size, the substitution of $\hat{\omega}_{ij}^2$ for ω_{ij}^2 should introduce little bias but it may be a problem in studies with small sample size. A final problem is that the parameter σ^2 must also be estimated, but is required in the matrix D to obtain $\hat{\beta}$.

2.2 Bias-adjusted errors-in-variables estimator

It is well known that errors-in-variables attenuate the estimate of β toward 0, and as a consequence the least-squares estimator will tend to underestimate the strength of the density-dependence relationship (Wittink, 1988). Fuller (1987, p. 14) provides a bias-adjusted estimator based on the method-of-moments for the case where $D = \sigma^2 I$. It is straightforward to extend Fuller's estimator to the case where $D = \sigma^2 I + W$, and where there are sampling covariances between $\hat{X} | X$ and $\hat{Y} | Y$ (see the appendix).

By fitting the Jolly-Seber model to mark-recapture data we can obtain maximum likelihood estimates of N_i, S_i, $\mathrm{Var}(\hat{N}_i|N_i)$, $\mathrm{Var}(\hat{S}_i|S_i)$, $\mathrm{Cov}(\hat{S}_i|S_i, \hat{N}_i|N_i)$, and $\mathrm{Cov}(\hat{S}_{i-1}|S_{i-1}, \hat{N}_i|N_i)$ for the sampling occasions $i = 2, \ldots, t-2$. Using the invariance property of maximum likelihood estimators (MLEs) it is relatively simple to find the MLEs for transformed abundances and survival probabilities denoted N_i' and S_i'.

Let

$$
M_{T'T} = \begin{bmatrix} 0 & 0 \\ 0 & \sum_{i=1}^{n} \mathrm{Var}(\hat{N}_i'|N_i) \end{bmatrix} \quad \text{and} \quad M_{T'e} = \begin{bmatrix} 0 \\ \sum_{i=1}^{n} \mathrm{Cov}(\hat{S}_i'|S_i, \hat{N}_i'|\hat{N}_i) \end{bmatrix}
$$

then $\tilde{\beta} = (\hat{X}'\hat{X} - M_{T'T})^{-1}(\hat{X}'Y - M_{T'e})$ is a bias-adjusted estimator for β (see appendix for derivation). Because $\tilde{\beta}$ is a complicated function of $\{\hat{N}_i\}$ and $\{\hat{S}_i^2\}$ we found $\mathrm{Var}(\tilde{\beta})$ by jackknifing the $\hat{N}_i$ and $\hat{S}_i$ pairs.

2.3 Examples

We examined density-dependence in shearwater population dynamics using two data sets. In each case we examined the relationship between the logit of the survival probability and untransformed abundance. Both data sets are from studies of birds closely related to *P. griseus*: (*P. tenuirostris*, see Bradley & Wooler, unpublished data; *P. huttoni*, see Cuthbert, unpublished data). The relationship between survival and abundance in these two species is expected to be similar to that in *P. griseus*. Specific details of the location and methods used in these studies is available in Bradley *et al.* (1991) for *P. tenuirostris* and Cuthbert (1999) for *P. huttoni*.

3 Results

A plot of estimated survival probabilities against abundance estimates provides little suggestion of a functional relationship for the short-tailed shearwater population (Fig. 1). Formal analysis of these data provided no evidence of a linear
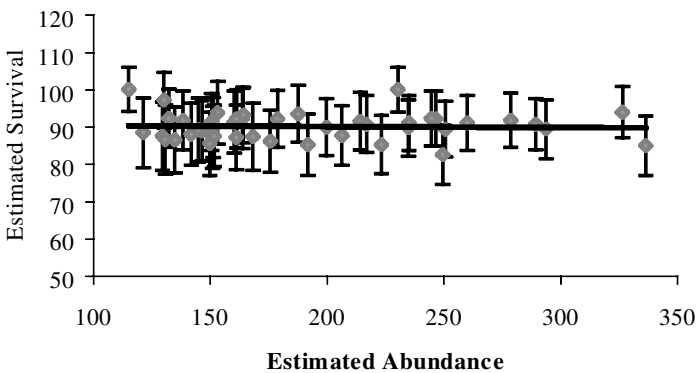

Fig. 1. Survival estimates plotted against abundance estimates from a sample of 1045 short-tailed shearwaters banded at Fisher Island, Tasmania, between 1947 and 1994. The error bars represent $+/-$ one standard error.

TABLE 1. Estimates of the slope of the effect of abundance on survival and associated standard errors for density for short-tailed shearwaters banded at Fisher Island 1947-1994 and Huttons shearwaters banded at Kaikoura 1991-1999

	Data		95% C.I	
	Slope	SE	Lower	Upper
Fisher Island	−0.0025	0.0055	−0.0134	0.0083
Kaikoura	−0.0612	0.0674	−0.1933	0.0709

functional relationship between abundance and survival probabilities (Table 1). A confidence interval for the odds ratio, indicates that a 20 bird change in the population (about 20% of the average estimated abundance) would lead to a change in the odds of survival by a factor of between 0.77 (a 23% reduction) and 1.181 (an 18% increase). Thus, although there is no evidence of a density dependent relationship between abundance and survival for short-tailed shearwaters on Fisher Island the data are sufficient to rule out a moderately strong relationship.

The Kaikoura data provided little evidence for density-dependence other than that the lowest survival estimate corresponds to the highest abundance estimate (Fig. 2). Even a naïve estimate of the relationship between survival and abundance, ignoring the uncertainty in $\hat{N}_i$, provides at best weak evidence of a negative relationship ($\hat{\beta}_1 = -0.0022$, SE $= 0.0012$). Our analysis using the bias adjusted estimator showed no evidence of a density-dependent relationship. Although the point estimate of the slope of the survival-abundance relationship was negative, the confidence interval for the odds ratio indicated that a change in abundance of approximately 20% would lead to a change in the odds of survival by a factor of between 0.38 (a 62% reduction) and 1.426 (a 43% increase). The correct inter-pretation of these results is that the Kaikoura data set is too sparse to rule out anything but a strong density dependent effect.
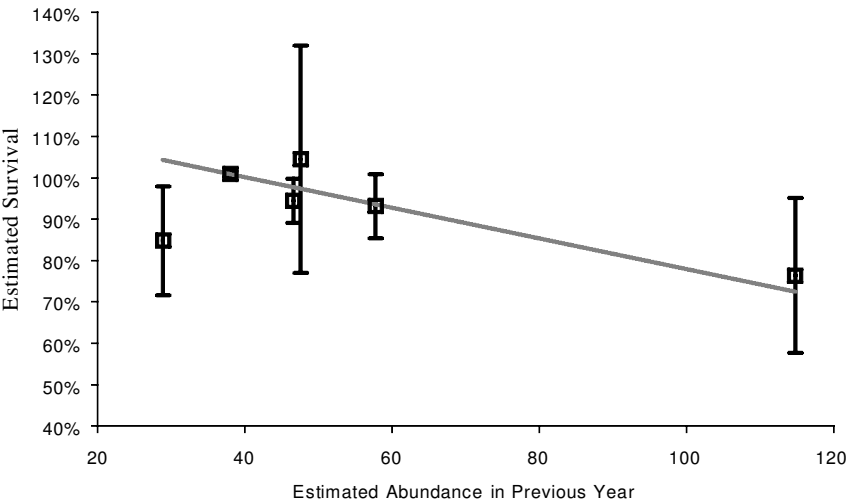

FIG. 2. Survival estimates plotted against abundance estimates from a sample of Hutton's Shearwaters banded at Kaikoura on the New Zealand mainland, 1991 to 1999. The error bars represent +/− one standard error.

4 Discussion

As emphasised by Link (1999), a particularly important role for random effects models is in providing a useful summary for a large number of parameters. Random effects models can also be useful for representing important functional relationships between parameters that have biological relevance. This is particularly relevant to mark-recapture data sets where relatively lengthy studies lead to a large number of parameter estimates.

We have modelled mark-recapture data as arising from a 2-stage process in which the mark-recapture parameters, the abundances survival rates and birth rates are first sampled from a 'hyperdistribution'. Conceptualizing mark-recapture parameters as random variables is a natural and logical step in mark-recapture models. Density dependent survival implies the presence of a functional relationship between abundance and survival. However it is unreasonable to believe that survival rates are determined wholly by abundance. Obvious environmental influences should also be included in the model at this stage. In addition to the structural components, the random term in the random effects model is particularly important as it is needed to account for influences that we are unable to explain or model explicitly.

In order to have a realistic chance of identifying major structural features of the data it is important to have a large number of observations. It is unreasonable to expect much to be learned from a regression based on a small number of observations. However, the interpretation of 'small' may be different for a biologist designing a study than for a statistician analysing the data. This emphasizes the need for biologists to commit resources to long-term monitoring programmes.

In contrast to the development of methods for modelling mark-recapture data conditional on the parameters, higher-order modelling is relatively undeveloped and there are few examples where such a procedure is followed (for an exception see Pledger, 1999). With the exception of Burnham's random effects models incorporated in MARK (White & Burnham, 1999), random effects models are not available in accessible software for mark-recapture data. For biologists to extract the full information from mark-recapture studies it is important that software for random effects models is developed and widely distributed. We anticipate that MCMC methods for Bayesian analyses will figure strongly in this development.

Our results give little support for density dependence in adult survival in both data sets examined. In the case of the Fisher Island study, data were sufficient for us to rule out anything but a small density dependent relationship. In this case, we would expect to see little compensation in survival rate as breeding densities were reduced by harvest. The Kaikoura data have too few years of observation for us to learn much about density dependence.

The methods presented here can also be used to assess density dependence in recruitment, which might arise as a consequence of intra-specific competition. With multi-state models (Brownie *et al*. 1993) becoming more prevalent, there is also potential to use them to measure density dependence in the rates at which birds miss breeding.

In the type of analysis we have considered here, 'density' was measured by population abundance. However, an inappropriate measure of density may result in density dependence being overlooked when it is present. This could occur if the abundance of breeder influences survival or reproduction, or both, but the data do not allow us to discriminate between breeders and non-breeders.

REFERENCES

BRADLEY, J. S., SKIRA, I. J., & WOOLER, R. D. (1991) A long-term study of short-tailed shearwaters *Puffinus tenuirostris*, *Ibis*, 133, pp. 55-61.
BROWNIE, C., HINES, J. E., NICHOLS, J. D., POLLOCK, K. H., & HESTBECK, J. B. (1993) Capture-recapture studies for multiple strata including non-Markovian transitions, *Biometrics*, 49, pp. 1173-1187.
CUTHBERT, R. (1999) Huttons Shearwaters in the Kaikoura Mountains. PhD thesis, University of Otago, Dunedin.
FULLER, W. A. (1987). *Measurement Error Models* (New York, Wiley).
GELFAND, A. E. & SMITH, A. F. M. (1990) Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, pp. 398-409.
LEBRETON, J. D., BURNHAM, K. P., CLOBERT, J. & ANDERSON, D. R. (1992) Modeling survival and testing biological hypotheses using marked animals—a unified approach with case-studies, *Ecological Monographs*, 62, pp. 67-118.
LINK, W. A. (1999) Modeling pattern in collections of parameters, *Journal of Wildlife Management*, 63, pp. 1017-1027.
PLEDGER, S. (1999) Finite mixtures in capture-recapture models. PhD thesis, Victoria University, Wellington.
PRADEL, R. (1996) Utilization of capture-mark-recapture for the study of recruitment and population growth rate, *Biometrics*, 52, pp. 703-709.
ROBSON, D. S., & YOUNGS, W. D. (1971) Statistical analysis of reported tag-recaptures in the harvest from an exploited population, Biometrics Unit, Cornell University, Ithaca, New York. BU-369-M.
SEBER, G. A. F. (1965) A note of the multiple recapture census, *Biometrika*, 52, 249-259.
WHITE, G. C. & BURNHAM, K. P. (1999) Program MARK: survival estimation from populations of marked individuals, *Bird Study*, 46 (supplement), pp. 120-139.
WITTINK, D. R. (1988) *The Application of Regression Analysis* (Boston, MA, Allyn and Bacon, Boston).
ZEGER, S. L., & KARIM, M. R. (1991) Generalized linear models with random effects: a Gibbs sampling approach, *Journal of the American Statistical Association*, 86, pp. 79-86.

Appendix

Let $Y = X\beta + \mathbf{e}$, where Y is an $(n \times 1)$ vector, X is an $(n \times p)$ matrix of constants, $\mathbf{e}$ is an $(n \times 1)$ random vector with mean zero, and $\Sigma_{wz} = \text{Cov}(\mathbf{w}, \mathbf{z})$ is the covariance matrix for the random vectors $\mathbf{w}$ and $\mathbf{z}$. Suppose we observe Y and $\hat{X} = X + T$ where T is a $(n \times p)$ random measurement error matrix made up of column vectors $\mathbf{t}_i (i = 1, \ldots, p)$ each with mean $\mathbf{0}$. From least-squares estimation theory $E[X'X)^{-1} X'Y] = (X'X)^{-1} X'X\beta = \beta$. Now:

(1) $E[\hat{X}'Y] = E[(X + T)'Y] = X'X\beta + E[T'Y]$.

(2) $E[T'Y] = E[T'(X\beta + \mathbf{e})] = E[T'X\beta] + E[T'\mathbf{e}] = \mathbf{0} + (\text{tr}(\Sigma_{t_1\mathbf{e}}), \text{tr}(\Sigma_{t_2\mathbf{e}}), \ldots, \text{tr}(\Sigma_{t_p\mathbf{e}}))'$, where $\text{tr}(Z)$ is the trace of the matrix Z.

(3) $E[\hat{X}'\hat{X}] = E[(X + T)'(X + T)'] = E[X'X + X'T + T'X + T'T] = X'X + E[T'T]$

(4) $E[T'T] = \begin{bmatrix} \text{tr}(\Sigma_{t_1 t_1}) & \text{tr}(\Sigma_{t_1 t_2}) & \ldots & \text{tr}(\Sigma_{t_1 t_p}) \\ \text{tr}(\Sigma_{t_2 t_1}) & \text{tr}(\Sigma_{t_2 t_2}) & \ldots & \text{tr}(\Sigma_{t_2 t_p}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{tr}(\Sigma_{t_p t_1}) & \text{tr}(\Sigma_{t_p t_2}) & \ldots & \text{tr}(\Sigma_{t_p t_p}) \end{bmatrix}$ where $\text{tr}(\Sigma_{t_j t_k}) = \sum_{i=1}^{n} \text{Cov}(t_{ij}, t_{ik})$.

If the covariances in $E[T'T]$ are known, then $\hat{X}'\hat{X} - E[T'T]$ is a consistent estimator for $X'X$, and if the covariances in $E[T'Y]$ are also known, then

$$\tilde{\beta} = (\hat{X}'\hat{X} - E[T'T])^{-1}(\hat{X}'Y - E[T'Y])$$

should also be a consistent estimator for β. Assuming $E[T'Y] = \mathbf{0}$ gives us the matrix analogue of Feller's estimator. Note that we can generalize the above to weighted least squares estimators of the form $(\hat{X}'V^{-1}\hat{X})^{-1}\hat{X}'V^{-1}Y$ however, if the weight matrix has elements that need to be estimated we lack a theory that can be used as a basis for deriving an estimator.