Analysis of Capture–Recapture Models with Individual Covariates Using Data Augmentation

J. Andrew Royle

U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708, U.S.A.
email: aroyle@usgs.gov

SUMMARY. I consider the analysis of capture–recapture models with individual covariates that influence detection probability. Bayesian analysis of the joint likelihood is carried out using a flexible data augmentation scheme that facilitates analysis by Markov chain Monte Carlo methods, and a simple and straightforward implementation in freely available software. This approach is applied to a study of meadow voles (*Microtus pennsylvanicus*) in which auxiliary data on a continuous covariate (body mass) are recorded, and it is thought that detection probability is related to body mass. In a second example, the model is applied to an aerial waterfowl survey in which a double-observer protocol is used. The fundamental unit of observation is the cluster of individual birds, and the size of the cluster (a discrete covariate) is used as a covariate on detection probability.

KEY WORDS: Abundance estimation; Bayesian analysis; Capture–recapture; Cluster size; Data augmentation; Heterogeneity; Individual covariates; Markov chain Monte Carlo; Nonignorable missing data; Population size; `WinBUGS`.

1. Introduction

Capture–recapture models with heterogeneous detection probabilities among individuals are widely used in studies of animal populations. One class of heterogeneity models is that in which individual detection probabilities are viewed as latent parameters described by some probability distribution. Such models have received considerable attention in recent years, with many specific models having been developed that correspond to alternative probability distributions for the individual detection probability parameters (Norris and Pollock, 1996; Coull and Agresti, 1999; Dorazio and Royle, 2003; Royle and Nichols, 2003; Royle, 2006). The interpretation and utility of such models have been called into question by Link (2003), who showed that, for most practical purposes, N is not identifiable across classes of mixture distributions. One possible way to mitigate this problem is to attempt to model heterogeneity explicitly by, for example, identification of individual covariates that are thought to explain variation in detection probability. Obvious covariates include age, sex, size, or cluster size when individuals occur in groups.

Considerable attention has been paid to this problem of individual covariates in capture–recapture models (see Pollock, 2002; Williams, Nichols, and Conroy, 2002, p. 300–302, for reviews). Three basic strategies have been suggested for modeling individual covariates. One approach suggested by Pollock, Hines, and Nichols (1984) requires stratification of individuals into a finite number of discrete classes, yielding K strata with stratum population sizes $\{N_k\}_{k=1}^{K}$. Under this approach, the collection of N_k parameters is the object of inference. This approach is suited for discrete covariates, especially those that take on a small number of values. One shortcoming is that the parameter dimension increases with the cardinality of the covariate, or number of strata in the case of a continuous covariate. Avoiding this dimensionality increase by specification of fewer strata yields a poorer approximation of the covariate effect, and the issue is compounded as additional covariates are considered. A second strategy in widespread use is the Huggins–Alho procedure (Huggins, 1989; Alho, 1990). This is based on estimates of individual detectability derived from the so-called "conditional likelihood," the likelihood for the data conditional on capture. Under this approach, N is a derived parameter, its estimation is based on a generalized Horvitz–Thompson estimator (HTE), and classical methods of asymptotic inference are employed. A third approach, the so-called "joint likelihood" approach, specifies the joint distribution of the encounter data and the covariate x by prescribing a probability distribution for the covariates, say $g(x)$. This approach is consistent with the contemporary treatment of missing covariates in classical generalized linear models. See Ibrahim et al. (2005) for a comprehensive review. Pollock (2002) provides a good overview of different approaches.

In this article, I consider Bayesian analysis of individual covariate models. Analysis is based on the joint distribution of the encounter history observations and the individual covariate. Bayesian analysis by Markov chain Monte Carlo (MCMC)

is aided by the use of an efficient data augmentation procedure for multinomial models with unknown sample size (Royle, Dorazio, and Link, 2007). Data augmentation yields a flexible model reparameterization that permits a straightforward Bayesian analysis. Two examples are provided. In the first, a live trapping study of meadow voles in which body mass is thought to explain heterogeneity in detection probability. The second example involves waterfowl survey data in which the effect of a discrete covariate (cluster size) is evaluated.

2. The Joint Likelihood and Data Augmentation

Suppose that a closed population of size N is sampled on T occasions, yielding a sample of n unique individuals, where individuals $i = 1, 2, \ldots, n$ were captured $\{y_i\}_{i=1}^n$ times. For clarity, we assume that detection probability does not vary over the T occasions, but the following development extends directly to a formulation in terms of capture histories (an example is provided in Section 4). Suppose an auxiliary variable is measured on each individual, say x_i, and that this variable is thought to influence the detectability of individuals. We suppose that captures are independent and identically distributed (i.i.d.) Bernoulli trials with parameter $p(x_i; \theta_1) \equiv p_i$. For example,

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_i, \tag{1}$$

where, in this case, the parameter θ_1 is the vector $\theta_1 = (\alpha_0, \alpha_1)$.

I address now the problem of estimating the total size of the population, N, from which the n observed individuals were sampled, as well as the parameters of the model including those of the detection function in equation (1). The main difficulty in doing this is that the sample is a biased sample, in the sense that individuals that appear in the sample tend to have higher values of p_i than in the population at large. Thus, the probability that an individual appears in the sample is also a function of unknown model parameters. Analogous to conventional treatments of missing data, we proceed by specifying a probability distribution for the covariate x, say $g(x \mid \theta_2)$. For many continuous covariates, it might be reasonable to assume g to be a normal distribution with mean μ and standard deviation σ. The notation θ_2 will represent the parameters of g regardless of their number.

For inference under this model, I develop the likelihood for the observed values of y and x from the joint distribution of y and x conditional on the event that the individual appeared in the sample, i.e., that $y > 0$. Suppressing the dependency on parameters, the joint distribution of y and x is

$$[y, x] = [y, x \mid y > 0] \Pr(y > 0) + [y, x \mid y = 0] \Pr(y = 0).$$

Here we made use of conventional "bracket notation" in which, for example, $[u]$ represents the probability density function (pdf) of u, $[u \mid v]$ represents the conditional pdf of u given v, and so forth. The observed values of y are integers $k = 1, 2, \ldots, T$ and thus $[y, x \mid y = 0]$ is identically 0 for the observed encounter frequencies. Thus, the conditional likelihood is

$$[y, x \mid y > 0] = [y \mid x, y > 0][x \mid y > 0] = \frac{[y \mid x][x]}{\Pr(y > 0)}.$$

Under the assumption that captures are i.i.d. Bernoulli trials, so that $[y \mid x]$ is binomial with index T, then the marginal

probability of capture is

$$\Pr(y > 0) = 1 - f(0; \boldsymbol{\theta}) = 1 - \int_x \text{Bin}(0; T, p(x; \theta_1)) g(x \mid \theta_2) \, dx, \tag{2}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $\text{Bin}(0; T, p(x; \theta_1))$ denotes the binomial pdf with index T and parameter $p(x)$ according to equation (1). Thus, the specific form of the conditional likelihood is

$$L_c\big(\theta_1, \theta_2; \{y_i, x_i\}_{i=1}^n\big)$$

$$= \frac{\left\{\prod_{i=1}^n \text{Bin}(y_i; T, p(x_i; \theta_1))\right\} \left\{\prod_{i=1}^n g(x_i; \theta_2)\right\}}{(1 - f(0; \boldsymbol{\theta}))^n}, \tag{3}$$

which can be maximized using conventional methods to obtain the maximum likelihood estimates (MLEs) of θ_1 and θ_2. The MLEs can be used to obtain the so-called "conditional MLE" of N, i.e., $\hat{N}_c = n/\{1 - f(0; \hat{\theta})\}$.

The unconditional likelihood is the product of equation (3) and the conditional distribution of n given N, which is binomial with index N and parameter $1 - f(0; \boldsymbol{\theta})$. This yields

$$L\big(N, \theta_1, \theta_2; \{y_i, x_i\}_{i=1}^n\big)$$

$$= \left\{\prod_{i=1}^n \text{Bin}(y_i; T, p(x_i; \theta_1))\right\} \left\{\prod_{i=1}^n g(x_i; \theta_2)\right\}$$

$$\times \frac{N!}{n!(N-n)!} f(0; \boldsymbol{\theta})^{N-n}. \tag{4}$$

The conditional and unconditional MLEs of N are asymptotically equivalent (Sanathanan, 1972), and both are widely adopted in varying population size estimation contexts.

2.1 *Bayesian Estimation by Data Augmentation*

Analysis of either the unconditional or conditional likelihoods described in the previous section is relatively straightforward, the main difficulty being evaluation of the integral in equation (2). In most cases, this will be done numerically. For a model with r individual covariates, equation (2) is replaced by an r-dimensional analog. Subsequently, the integration required to evaluate the likelihood is avoided by adopting a Bayesian formulation of the individual covariate model under which the posterior distribution is characterized using conventional methods of MCMC. Analysis of the unconditional likelihood introduces an additional complication, that the dimension of the parameter space is itself a parameter (i.e., the number of unobserved x's and the number of "zero observations" is equal to $N - n$, N being unknown). This problem motivated the data augmentation scheme described by Royle et al. (2007) for modeling and estimation in multinomial models with unknown index, N. In the subsequent analyses, this data augmentation scheme is exploited to fix the dimension of the partially observed variables x and y.

The basic strategy is to augment the observed data set with a fixed, known number, say $M - n$, of all-zero capture histories and to model the augmented data set (of size M) as a zero-inflated version of the complete-data model using an

unknown, but estimable, zero-inflation parameter. In essence, the classical multinomial model for estimating population size is reparameterized to a site occupancy model of the form described by MacKenzie et al. (2002). Indeed, this duality between closed population size estimation models and site occupancy models largely motivated the data augmentation idea, but a more formal justification based on Bayesian arguments can also be made (Royle et al., 2007). Specifically, they noted that data augmentation is equivalent to specifying a uniform prior distribution for N on the integers up to M. Thus, for $M \gg N$, this yields a natural and common prior for N that is often viewed as being noninformative, or expressing little prior information about N as long as M is sufficiently large. The Bayesian construction establishes a reasonable mechanism for choice of M—it should be large enough so as to have no effect on posterior summaries, otherwise its use (as an informative prior) should require some justification as is customary. The main benefit of adopting data augmentation for the analysis of individual covariate models is that it yields a simple and efficient Bayesian implementation using MCMC methods.

For models containing individual covariates, the data consist of the n observed detection frequencies $y_1, y_2, \ldots, y_n$, and associated covariate values $x_1, x_2, \ldots, x_n$. For estimation by data augmentation, introduce zero pseudo-operations ($y_{n+1} = 0, y_{n+2} = 0, \ldots, y_M = 0$), missing covariate values $\{x_i\}_{i=n+1}^M$, and a set of latent indicator variables $\{z_i\}_{i=1}^M$ that are observed ($z_i = 1$) for $i = 1, 2, \ldots, n$ and unobserved for $i = n + 1, \ldots, M$. Then, suppose that $z_i \sim \text{Bernoulli}(\psi)$. The parameter ψ has the interpretation of being the probability that an element of the augmented population (of size M) is a member of the sampled population (of size N), and the object of inference is the population size, $N = \sum_{i=1}^M z_i$.

The model for the augmented data consists of the following three components:

(1) $z_i \sim \text{Bernoulli}(\psi)$;
(2) $[y_i \,|\, p(x_i)] = \text{Binomial}(T, z_i p(x_i))$ with $\text{logit}(p_i) = \alpha_0 + \alpha_1 x_i$;
(3) $[x_i] = \text{Normal}(\mu_x, \sigma_x^2)$.

Estimation and inference are straightforward using conventional MCMC methods. Two examples are given in following sections. The MCMC algorithm requires sampling each "missing" x_i from its full conditional distribution, which is $[x_i \,|\, y_i = 0] \propto \{(1 - p(x_i))^T [x_i]\}$. This can be sampled from easily using generic methods such as the Metropolis–Hastings algorithm. The MCMC algorithm otherwise is straightforward to implement. The details can be avoided by use of existing software. The subsequent analyses were carried out in the software `WinBUGS` (Gilks, Thomas, and Spiegelhalter, 1994). A number of extensions of the basic covariate model described here are provided in the subsequent analyses.

3. Application: A Continuous Covariate

In this example, a population of meadow voles (*Microtus pennsylvanicus*) in Laurel, Maryland, was sampled on five consecutive days during 1981. The sampling was carried out using a 10×10 grid of traps baited with corn. The data considered here are encounter histories of 56 adult males, ob-

tained from five samples on consecutive days beginning June 27. The covariate body mass (grams) was measured on each individual at the time of first capture, and there is some a priori belief that there should be a positive effect of weight on detection probability. In the following analyses, the body mass covariate was standardized to have mean zero and unit variance. Further details of the study can be found in Nichols, Pollock, and Hines (1984) and the data were used extensively by Williams et al. (2002) and have also been analyzed by a number of other authors including Wolter (1990); Lee, Lee, and Gee (2003); Yang and Chao (2005); and Zhang, Liu, and You (2005).

For the microtus data, the encounter frequency distribution is $n_k = (12, 8, 9, 12, 15)$ for $k = 1, 2, 3, 4, 5$, respectively. Estimates obtained under the simple null model ("M_0") were $\hat{p} = 0.60$, and the MLE for the number of individuals not captured occurs on the boundary of the parameter space $n_0 = 0$. This is a sensible result, given the high detection probability of this sampling apparatus (baited traps arranged in a grid). The probability of an individual being captured at least one time is $1 - (1 - 0.60)^5 = 0.99$.

I extend the model to include the individual covariate "body mass." We suppose that body mass in the population of mice is normally distributed with mean μ_m and variance σ_m^2. Having augmented the data set with $M - n = 100$ observations of $y = 0$, the model for the augmented data is that described in Section 2. The analysis was repeated with augmentation based on 125 and 50 zeros and the results were consistent to within Monte Carlo error. To proceed with the analysis, prior distributions for the structural parameters are required and, for these, I adopt conventional default priors which, ostensibly, express little prior information about the model parameters. For mean parameters, normal priors with mean 0 and variance 1000 were used whereas for the inverse of variance parameters ($\tau = 1/\sigma^2$), gamma priors with shape and scale both equal to 0.001 were used.

The implementation of the MCMC algorithm is straightforward. Here I focus on implementation in the software `WinBUGS`. This requires little more than a "pseudocode" representation of the model structure. That is, specification of the probability distributions for the data, latent variables, and parameters, and any functional linkage among parameters. The `WinBUGS` model specification for the *Microtus* data is provided in Web Appendix A, and an `R` script to execute `WinBUGS` from within `R` is given in the supplement.

The posterior distribution of N, based on 50,000 Monte Carlo draws from the posterior distribution, is summarized in Table 1. A 95% posterior interval based on the estimated distribution is [56, 69]. The effect of body mass on detection probability appears to be very substantial, with 100% of the posterior mass of α_1 being above 0. Various other posterior summaries of the model parameters are given in Table 1.

3.1 *The Huggins–Alho Procedure*

The Huggins–Alho procedure (Huggins, 1989; Alho, 1990) is the de facto standard for analysis of individual covariate models. This is a heuristic procedure justified by analogy with the HTE. I applied this estimator to the *Microtus* data set, and it yielded $\hat{N} = 59.2$, which is consistent with the fully

Table 1

Posterior summaries of model parameters for the Microtus *data under a model containing (standardized) body mass as a covariate on detection probability. The model was fit using data augmentation in* WinBUGS. *The parameter ψ is the "zero-inflation" parameter associated with data augmentation. α_0 and α_1 are the intercept and coefficient on body mass, respectively.*

Parameter	Posterior mean	Posterior SD	2.5%	Median	97.5%
N	60.040	3.407	56.000	59.000	69.000
α_0	0.586	0.145	0.305	0.585	0.871
α_1	1.011	0.176	0.681	1.006	1.369
μ_x	-0.114	0.174	-0.482	-0.105	0.205
σ_x	1.087	0.129	0.878	1.072	1.380
ψ	0.293	0.035	0.228	0.292	0.367

model-based result. Extension of the Huggins–Alho procedure to more general situations requires a formulation of the model in which N is removed by conditioning on n.

3.2 *Model Extensions*

The data augmentation framework can be extended directly to accommodate more general models. For example, the *Microtus* traps were baited with corn, and this raises the possibility of what is usually referred to as a behavioral response where we might expect detection probability to increase after initial capture. I consider here the classic behavioral response model M_b, as well as an extension suggested by Yang and Chao (2005) allowing for a short-term behavioral response (what they referred to as an ephemeral response). This broader class of models has the form

$$\text{logit}(p_{ij}) = \alpha_0 + \alpha_1 x_{1,ij} + \alpha_2 y_{i,j-1}.$$

The covariate $x_{1,ij}$ is an indicator of previous capture. That is, $x_{1,ij} = 1$ if the individual was captured previous to sample j. The model contains a Markovian autoregressive-like structure, the regression on $y_{i,j-1}$, this is the short-term ephemeral response to being captured.

It is fairly straightforward to describe the model in WinBUGS (see Web Appendix A). The results are given in Table 2, where we see that there does not appear to be a very strong

Table 2

Posterior summaries of model parameters for the Microtus *data under a model containing persistent and ephemeral behavioral responses on detection probability. The model was described by Yang and Chao (2005) and fit using data augmentation in* WinBUGS. *The parameter ψ is the "zero-inflation" parameter associated with data augmentation. The parameters α_0, α_1, and α_2 are, respectively, the intercept, persistent, and ephemeral behavioral responses.*

Parameter	Mean	SD	2.5%	Median	97.5%
N	57.760	1.918	56.000	57.000	63.000
α_0	0.090	0.259	-0.447	0.096	0.570
α_1	0.241	0.346	-0.430	0.235	0.932
α_2	0.628	0.393	-0.160	0.630	1.385
ψ	0.321	0.036	0.254	0.320	0.394

behavioral response. There is some evidence for a slight ephemeral response, in the sense that most of the posterior mass is above 0 for that parameter.

I consider one further extension to include the body mass covariate:

$$\text{logit}(p_{ij}) = \alpha_0 + \alpha_1 x_{1,ij} + \alpha_2 y_{i,j-1} + \alpha_3 x_{2,i}.$$

The covariate $x_{1,ij}$ is an indicator of previous capture, as before, and $x_{2,i}$ is the body mass covariate. I assume that body mass in the population of mice is normally distributed with mean μ_x and variance σ_x^2, i.e., $x_{2,i} \sim \text{Normal}(\mu_x, \sigma_x^2)$. As with the previous models, analysis of the model for the augmented data by MCMC is straightforward. The results are summarized in Table 3, where we see (columns 2 and 3) a large effect of the individual covariate body mass, and the persistent behavioral response is very strong in the presence of the body mass covariate. On the other hand, the ephemeral response is weak. Therefore, this component of the model was removed and the resulting simpler model was refit (columns 4–8 in Table 3). We see that the posterior distributions of both α_1 (the behavioral effect) and also of α_3 (the body mass coefficient) are concentrated above 0. Thus, the results indicate a large increase in detection probability once an individual is captured, and also that heavier individuals have higher probabilities of a capture. Note also that when these two effects are included in the model, the posterior mass of N does shift away from zero. This is the intuitive effect, because the high apparent detection probability is due to the higher recapture rate of previously caught individuals, and the propensity for heavier individuals of being trapped.

The estimated population mean and standard deviation of the body mass covariate indicate that the measured covariate on the observed individuals represents only a slightly biased sample. Whereas the sample mean and standard deviation of body mass were 41.79 and 11.96, respectively, the posterior mean of $E[x]$ and $\text{Var}[x]$ (upon back-transforming the posterior means of the estimates for μ and σ given in Table 3) are obtained by solving $(x - 41.79)/11.96 = 1.122$, which yields the population value $E[x] = 39.45$ and also $\text{Var}((x - 41.79)/11.96) = 1.122^2$ yields the population standard deviation of $\text{SD}(x) = 13.42$. Thus, the population distribution of body mass is slightly lower and more variable than the sample, as expected. The probability of capture for an individual that is s standard deviations from the mean body mass was evaluated for $s = (-2, -1, 0, 1, 2)$, for a hypothetical individual that was not previously captured, resulting in (0.532, 0.848, 0.984, 0.999, 1.000). Individuals whose body mass is two standard deviations below the mean have only about a 50% chance of being captured during a study composed of $T = 5$ periods. Thus, considerable heterogeneity in p is induced by dependence of p on the covariate body mass. Now, consider the same calculation for an individual that *was* previously captured. This yields (0.766, 0.964, 0.998, 0.999, 1.000). The interpretation of this is as follows: The probability of capture for an individual two standard deviations below the mean body mass approximately doubles after initial capture. Due to the high mean capture probability (α_0), there is little effect of initial capture for individuals whose body mass is much above the mean.

Table 3
Posterior summaries of model parameters for the Microtus *data under a model with both behavioral effects and body mass (standardized), and then a model with only the persistent behavioral effect plus body mass. The model was fit using data augmentation in* WinBUGS. *The parameters* α_0, α_1, α_2, *and* α_3 *are, respectively, the intercept, persistent, and ephemeral behavioral responses, and the coefficient on body mass.*

	Full model		Reduced model				
Node	Mean	SD	Mean	SD	2.5%	Median	97.5%
N	64.06	7.585	64.110	7.532	56.000	62.000	83.000
α_0	0.242	0.289	0.243	0.291	-0.351	0.251	0.791
α_1	0.585	0.378	0.719	0.176	0.379	0.718	1.070
α_2	0.174	0.427					
(mass) α_3	1.016	0.187	1.025	0.183	0.682	1.021	1.398
μ_x	-0.196	0.223	-0.196	0.222	-0.704	-0.172	0.161
σ_x	1.122	0.149	1.121	0.148	0.891	1.102	1.464

3.3 *Model Selection*

The previous results appear reasonably unambiguous in support of an effect of body mass and, in the presence of body mass, a "permanent" behavioral response. Here I address the problem of formal model selection based on posterior model probabilities. The posterior probabilities of each model were computed by specifying a set of latent indicator variables, one for each model effect, say w_j for the jth effect, and imposing a Bernoulli prior on each w_j, say having parameter π_j. This notion was suggested by Kuo and Mallick (1998) (see also Congdon [2005, section 3.2]) . The latent indicator variables are:

$$w_k = \begin{cases} 1 & \text{linear predictor includes covariable k} \\ 0 & \text{linear predictor does not include covariable k,} \end{cases}$$

having prior distributions:

$$w_k \sim \text{Bernoulli}(0.5),$$

which I assume here to be mutually independent. The model is expanded by specifying the linear predictor as

$$\text{logit}(p_{ij}) = \alpha_0 + \alpha_1 w_1 x_{1,ij} + \alpha_2 w_2 y_{i,j-1} + \alpha_3 w_3 x_{2,i}.$$

Analysis of this expanded model within the data augmentation framework does not pose any additional difficulty. One broader, technical consideration is that posterior model probabilities are well known to be sensitive to priors on parameters (e.g., Aitkin, 1991; Link and Barker, 2006) and vague priors are not usually innocuous or "uninformative" when evaluating posterior model probabilities. The use of Akaike's information criterion seems to avoid this problem largely by imposing a specific and perhaps undesirable prior that is a function of the sample size (Kadane and Lazar, 2004). One solution to this problem is to compute posterior model probabilities under a model in which the prior for parameters is fixed at the posterior distribution under the full model (Aitkin, 1991). In the present case, the prior was the normal approximation to the four-dimensional posterior distribution of the regression coefficients (from which the summary results in Table 3 derive). Using this prior distribution, posterior model probabilities were obtained from 100,000 post-burnin Monte Carlo draws.

The eight possible models are indexed by the binary sequence (w_1, w_2, w_3), where w_1, w_2, and w_3 represent the persistent behavioral response, ephemeral response, and effect of body mass. The models $(1,1,1)$, $(1,0,1)$, $(0,1,1)$, and $(0,0,1)$ received posterior probabilities of 0.272, 0.347, 0.279, and 0.102, respectively. The other four models received 0 posterior probability. We see that the basic pattern favors models with body mass (all four models with positive probability have body mass in them). However, the posterior probability is about evenly distributed among the three models containing body mass and one or both behavioral responses.

3.4 *Model Sensitivity*

Formulation of the individual covariate model in terms of the joint distribution of y and x yields a procedure that is distinctly model based. Specifically, assumptions regarding both the functional form of the relationship between detection probability and the covariate(s), as well as the population distribution of the covariate, are required. The latter is distinct from classical methods that are based on the conditional likelihood, and thus it is natural to question whether this additional assumption is important, in the sense that results are overly sensitive to choice of $g(x)$. Link (2003) demonstrated, in the related class of models allowing for individual heterogeneity (so-called "Model M_h"), that different parametric choices of heterogeneity distributions yield substantially different inferences about N. It is natural to question whether this ambiguity arises in the present class of models. Note that in the models considered by Link (2003), p is regarded as an individual-specific latent parameter with distribution $g(p)$. In the present class of models, we have that $h(p_i) = \alpha_0 + \alpha_1 x_i$. The main difference between the class of models considered by Link (2003) and the individual covariate models being considered here is that, in the latter, we observe some direct information about $g(x)$ by virtue of observing x_i for those individuals that appear in the sample. Having direct information about these parameters may lead to some stability in this class of problems as compared to those considered by Link (2003).

The sensitivity to choice of g was evaluated under the full model (containing both behavioral effects and the body mass covariate) for several choices of g including the normal considered previously, the double exponential, logistic, log-normal,

and *t*-distribution with two degrees of freedom. The results under each model for the parameter of interest, *N*, are summarized in Web Table 1. The posterior mean deviance is also given. These results do suggest some sensitivity of the estimator of *N* to choice of $g(x)$. However, the effect on the posterior mean of *N* is mitigated to some extent by the fact that the overall detectability of individuals is high. The body mass coefficient appears relatively insensitive to choice of *g*.

4. Example: Cluster Size in Wildlife Surveys

In this example, I consider aerial survey data on mallard ducks (*Anas platyrhynchos*) collected by the U.S. Fish and Wildlife Service during the 2005 annual waterfowl population survey in the Northeastern United States and Eastern Canada (Koneff et al., 2008). Sample units in this fixed-wing survey are 18-mile linear segments (Smith, 1995). Sampling was conducted using two observers (front seat, back seat) in a double-observer sampling protocol (Magnusson, Caughley, and Grigg, 1978). This yields three observable encounter histories of the form $\{(0, 1), (1, 0), (1, 1)\}$, indicating whether the cluster was seen by the back-seat observer only, the front-seat observer only, or both observers. The model for such data is equivalent to a capture–recapture study with $T = 2$ periods, and detection probabilities p_1 (for the front-seat observer), and p_2 (back-seat observer). The front-seat observer is expected to have a higher detection probability than the back-seat observer. A total of 162 clusters of birds were observed during this survey. The detection history frequencies are given in Table 4. The observed cluster sizes ranged from 1 to 7, with very few observed cluster sizes greater than 2. In what follows, the population cluster size distribution is assumed to be $1 +$ Poisson(λ), a right-shifted Poisson distribution, with mean $1 + \lambda$. The model for the probability of detection for cluster *i* was

$$\text{logit}(p_{ik}) = \alpha_k + \beta x_i, \tag{5}$$

for observer $k = 1, 2$.

Data augmentation proceeds by adding a large number of $(0, 0)$ encounter histories. For this analysis, 125 zero encounter histories were added, which proved to be adequate in the sense that the posterior mass of *N* was concentrated away from the upper limit (see Figure 1). To implement the model in `WinBUGS`, note that each observation is the realization of a four-dimensional multinomial random variable with sample size 1. The cell probabilities of the multinomial, corresponding to the encounter histories $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$, are $\{(1 - p_{i1})(1 - p_{i2}), (1 - p_{i1})p_{i2}, p_{i1}(1 - p_{i2}),$

$p_{i1}p_{i2}\}$ for cluster *i* comprised of x_i individuals. The zero inflation of this multinomial model is analogous to that of the binomial model described in Section 3. The `WinBUGS` model specification is given in Web Appendix A.

The quantities of interest are the total number of sampled groups: $N_c = \sum_{i=1}^{M} z_i$, and the total number of individuals $N_{\text{ind}} = \sum_{i=1}^{M} z_i x_i$ (recall that z_i is a latent indicator of whether an element of the augmented population of size *M* is also an element of the sampled population). The covariate values x_i for $i = n + 1, n + 2, \ldots, M$ are missing values, and these are sampled from their posterior distribution as part of the MCMC algorithm. The actual focus of this study is to obtain an estimate of density over the survey area, and so the estimate of N_{ind} would require scaling by total sample area. For the present purposes, we neglect this detail.

Posterior summaries of model parameters based on 100,000 Monte Carlo draws from the posterior distribution are given in Table 5. The posterior mass of the cluster size coefficient, β, was concentrated in the vicinity of zero. Thus, there does not appear to be an effect of cluster size on detection probability. The estimated average detection probabilities are ordered as anticipated, $p_1 = 0.80$ and $p_2 = 0.51$, for front- and back-seat observers, respectively. Finally, the posterior distribution of the total number of individuals, N_{ind}, is shown in Figure 1, and posterior summaries of both N_c (the number of clusters) and N_{ind} are given in Table 5.

5. Discussion

Capture–recapture models with structured heterogeneity, or "individual covariates" to explain variation in detection probability arise naturally in many problems. A generic and flexible framework for modeling and inference of individual covariate models can be achieved by adopting a Bayesian formulation of the model. A straightforward implementation can be achieved using MCMC methods in conjunction with a data augmentation scheme.

For the models considered here, the reparameterization that is induced by data augmentation yields a model which is equivalent to a site occupancy model (e.g., MacKenzie et al., 2002), with missing covariate values for some sites. The resulting estimation problem is analogous to that of "nonignorable" missingness in generalized linear models. The missingness is nonignorable "...if the failure to observe a value depends on the value that would have been observed" (Ibrahim et al., 2005). In the present context, the probability that the covariate is missing (i.e., that the individual is not captured) depends on the value of the covariate through the detection probability.

Two variations of the individual covariate models were applied to data from animal surveys. In the *Microtus* example, *p* was assumed constant among samples, but varied among individuals in response to a continuous covariate. In this case, the complete data likelihood is a zero-inflated binomial. In the aerial waterfowl survey example using a double-observer sampling protocol, *p* was assumed to vary for sample (corresponding to observer), and in response to a discrete covariate, the size of the cluster. In this case, the complete data likelihood is a zero-inflated multinomial likelihood.

Conventional treatments of individual covariate models are based on the conditional likelihood, in which inclusion

Table 4
Mallard cluster size encounter history frequencies. In the detection history (f, b) position f represents the front-seat observer and b represents the back-seat observer.

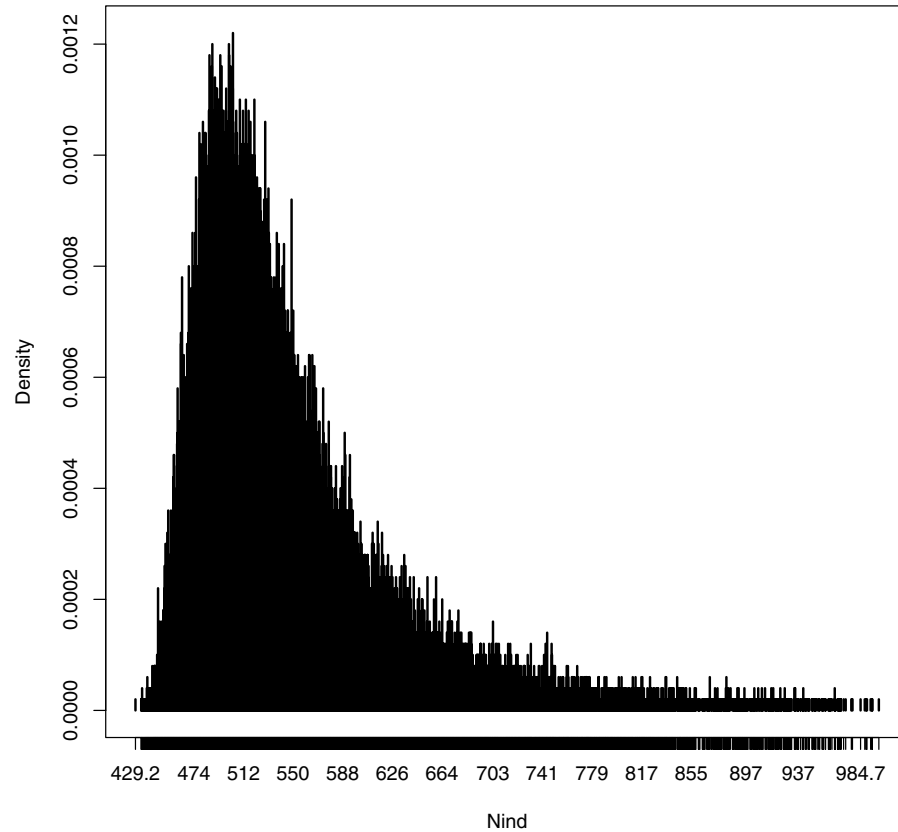| Detection history	Cluster size						
	1	2	3	4	5	6	7
$i = 1$ (1, 1)	42	13	3	0	4	0	2
$i = 2$ (0, 1)	10	6	1	0	2	0	0
$i = 3$ (1, 0)	52	19	4	1	1	2	0

Figure 1. Posterior distribution of the total number of mallards on the surveyed sample units.

Table 5
Posterior summaries of model parameters for the mallard aerial survey data. The parameter β is the effect of cluster size.

Parameter	Mean	SD	$q_{0.025}$	$q_{0.50}$	$q_{0.975}$
N_c	195.57	16.514	174.00	192.00	239.00
N_{ind}	540.72	71.160	459.60	522.30	735.90
α_1	1.58	0.461	0.68	1.58	2.47
α_2	0.21	0.478	-0.70	0.20	1.15
β	-0.18	0.193	-0.58	-0.17	0.17
λ	1.82	0.148	1.57	1.81	2.16
p_1	0.80	0.051	0.68	0.80	0.88
p_2	0.51	0.075	0.36	0.51	0.65

probabilities are estimated and then used in a HTE for population size. The most conceptually unappealing aspect to this approach is that N (the object of inference in many problems) is a derived parameter, formulated explicitly as a function of nuisance parameters (that are usually not of direct interest). The Bayesian formulation of the model based on data augmentation has several advantages that include its general extensibility. As an example, there are no additional complications if covariate values were not recorded for some of the observed individuals. Secondly, inference is valid in small samples using this Bayesian formulation. Finally, the implementation is readily accessible to practitioners, as the repa-

rameterized model may be implemented directly in the freely available software `WinBUGS`.

The fully model-based formulation of individual covariate models is contrary to the standard method of fitting such models based on the so-called Huggins–Alho procedure. This estimator is motivated by analogy with the HTE and seems to be favored in practice primarily because it is "design based," and therefore is, it might be hoped, robust to model assumptions (in particular, the choice of distribution for the individual covariate). However, Little (2004) provides an insightful discussion of this point, noting that the HTE has a model-based justification and the performance of the HTE can be sensitive to departures from that model. Thus, sensitivity to model choice should be a consideration regardless of whether one adopts ostensibly model-free procedures.

6. Supplementary Materials

The `WinBUGS` model specifications referenced in Sections 3 and 4, and the table referenced in Section 3.4, are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

References

Aitkin, M. (1991). Posterior Bayes factors. *Journal of the Royal Statistical Society, Series B* **53,** 111–142.

Alho, J. M. (1990). Logistic regression in capture-recapture models. *Biometrics* **46,** 623–635.

Congdon, P. (2005). *Bayesian Models for Categorical Data.* Hoboken, New Jersey: Wiley.

Cook, R. D. and Jacobson, J. O. (1979). A design for estimating visibility bias in aerial surveys. *Biometrics* **35,** 735–742.

Coull, B. A. and Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics* **55,** 294–301.

Dorazio, R. M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59,** 351–364.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *The Statistician* **43,** 169–178.

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76,** 133–140.

Ibrahim, J. G., Chen, M.-H., Lipsitz, S. R., and Herring, A. H. (2005). Missing-data methods for generalized linear models: A comparative review. *Journal of the American Statistical Association* **100,** 332–346.

Kadane, J. B. and Lazar, N. A. (2004). Methods and criteria for model selection. *Journal of the American Statistical Association* **99,** 279–291.

Koneff, M. D., Royle, J. A., Otto, M. C., Wortham, J. S., and Bidwell, J. K. (2008). A double-observer approach to estimating detection rates in aerial waterfowl surveys. Submitted to *Journal of Wildlife Management.*

Kuo, L. and Mallick, B. (1998). Variable selection for regression models. *Sankhya* **60B,** 65–81.

Lee, C. Y., Lee, S. M., and Gee, M. J. (2003). Estimating survival rates using an extended ricker's two-release method. *Computational Statistics and Data Analysis* **42,** 123–137.

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

Link, W. A. and Barker, R. J. (2006). Model weights and the foundations of multimodel inference. *Ecology* **87,** 2626–2635.

Little, R. J. (2004). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association* **99,** 546–557.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83,** 2248–2255.

Magnusson, W. E., Caughley, C. G., and Grigg, G. C. (1978). A double-survey estimate of population size from incomplete counts. *Journal of Wildlife Management* **42,** 174–175.

Nichols, J. D., Pollock, K. H., and Hines, J. E. (1984). The use of a robust capture-recapture design in small mammal population studies: A field example with Microtus pennsylvanicus. *Acta Theriologica* **29,** 357–365.

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

Pollock, K. H. (2002). The use of auxiliary variables in capture-recapture modelling: An overview. *Journal of Applied Statistics* **29,** 85–102.

Pollock, K. H., Hines, J. E., and Nichols, J. D. (1984). The use of auxiliary variables in capture-recapture and removal experiments. *Biometrics* **40,** 329–340.

Royle, J. A. (2006). Site occupancy models with heterogeneous detection probability. *Biometrics* **62,** 97–102.

Royle, J. A. and Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology* **84,** 777–790.

Royle, J. A., Dorazio, R. M., and Link, W. A. (2007). Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics* **16,** 67–85.

Sanathanan, L. (1972). Estimating the size of a multinomial population. *The Annals of Mathematical Statistics* **43,** 142–152.

Smith, G. W. (1995). A critical review of the aerial and ground surveys of breeding waterfowl in North America. U.S. Department of the Interior Biological Science Report 5.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations.* San Diego, California: Academic Press.

Wolter, K. M. (1990). Capture-recapture estimation in the presence of a known sex ratio. *Biometrics* **46,** 157–162.

Yang, H. and Chao, A. (2005). Modeling animals' behavioral response by Markov chain models for capture-recapture experiments. *Biometrics* **61,** 1010–1017.

Zhang, L., Liu, L., and You, N. (2005). Estimating population size in logistic capture-recapture models with a known sex ratio. *Communications in Statistics—Theory and Methods* **34,** 37–44.