Received: 13 May 2022 | Revised: 18 August 2023 | Accepted: 30 September 2023

DOI: 10.1002/env.2827

RESEARCH ARTICLE

WILEY

On the identifiability of the trinomial model for mark-recapture-recovery studies

Simon J. Bonner[1] | **Wei Zhang**[2] | **Jiaqi Mu**[1]

[1]Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Canada

[2]School of Mathematics & Statistics, University of Glasgow, Glasgow, United Kingdom

Correspondence
Simon J. Bonner, Department of Statistical and Actuarial Sciences, University of Western Ontario, 1151 Richmond St., London, ON, N6A 5B7, Canada.
Email: simon.bonner@uwo.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: 43024-2016

Abstract

Continuous predictors of survival present a challenge in the analysis of data from studies of marked individuals if they vary over time and can only be observed when individuals are captured. Existing methods to study the effects of such variables have followed one of two approaches. The first is to model the joint distribution of the predictor and the observed capture histories, and the second is to draw inference from the likelihood conditional on events that depend only on observed predictor values, called the trinomial model. Previous comparison of these approaches found that joint modelling provided more precise inference about the effect of the covariate while the trinomial model was less prone to issues of model mis-specification. However, we believe that an important issue was missed. We show through mathematical analysis and numerical simulation that the trinomial model is not identifiable when the predictor has no effect on the survival probability. This also causes inferences from the trinomial model to be imprecise when the effect of the covariate on the survival probability is small. We also analyse data on the effect of body mass on the survival of meadow voles to demonstrate the importance of this issue in real applications.

KEYWORDS

continuous covariates, identifiability, limiting posterior distributions, mark-recapture, survival

1 | INTRODUCTION

Time-varying individual covariates present challenges in the study of marked animals in wild populations. Whether researchers are interested in the effects of breeding status (Pradel, 2005), location (Borchers & Efford, 2008; Diana et al., 2022; McLaughlin & Bar, 2021), or measures of fitness like body mass (Bonner & Schwarz, 2006) or parasite load (Argáez et al., 2020), these variables can only be observed when an animal is detected by physical capture or other means. The resulting data will have many missing values that need to be handled in some way. Our focus is on extensions of the Cormack-Jolly-Seber (CJS) model in which the survival probability is modeled as a function of a continuous, time-varying, individual covariate.

There have been two general approaches to model the effects of a continuous, time-varying, individual, covariate within the CJS model and its extensions. The first is to construct a complete data likelihood from the joint density of the covariates and the mark-recapture (-recovery) data and then to fit this model with a numerical method that

essentially removes the missing values through integration. Bonner and Schwarz (2006) approached this in the Bayesian framework via Markov chain Monte Carlo (MCMC) simulation working with mark-recapture (MR) data alone. The same authors also considered maximum likelihood estimation via the Expectation-Maximization (EM) algorithm, but abandoned this approach due to the difficulty in computing standard errors (Bonner, 2003). This approach was also discussed by Catchpole et al. (2004), but to our knowledge it was not followed up. More recently, Worthington et al. (2015) considered a multiple imputation approach in which many complete data sets are generated by modeling the covariate and imputing the missing values, fitting the CJS model to each complete data set, and combining the results to form a single set of estimates. The approach properly accounts for the uncertainty related to the missing data but may be prone to bias as the covariate values are not missing at random (the pattern of missingness is related to the missing values themselves). Langrock and King (2013) also proposed a classical approach in which the covariate is first discretized and the model is then fit with the efficient machinery for computing maximum likelihood estimates for hidden Markov models (HMM), assuming that the covariate is modelled through a first order Markov process.

The second approach is to conduct a conditional analysis excluding any events that depend on the missing covariate values. The first approach of this type was presented by Catchpole et al. (2004). The method is based on a two-stage approach in which the ratio of the capture and recovery probability is first estimated by analysing mark-recapture-recovery data ignoring the effect of the covariate. The ratio is then treated as known and used as an offset in a binomial generalized linear model of the number of individuals released on one capture occasion and captured the next, including the covariate as a predictor. The method is admittedly naïve (the authors' own word), but it has the advantage that it could be fit using software readily available at the time. Building on this, Catchpole et al. (2008) proposed a more formal likelihood function constructed by considering only the animals released on each occasion and modeling whether they are recaptured alive, recovered dead, or neither on the subsequent occasion. This is referred to as the trinomial model, since there are three possible outcomes for each trial. The resulting likelihood contains no missing values and maximum likelihood estimates and their standard errors can be computed easily. Bonner (2013) showed how the model can be fit in Program MARK (White & Burnham, 1999), making it easy for applied researchers to implement the method without writing their own code.

Bonner et al. (2010) compared the two approaches and found, not surprisingly, that the complete data approach is more efficient because the likelihood covers more of the data. They considered five different scenarios with different combinations of the capture and recovery probabilities in their simulation study and found that the average widths of the 95% credible intervals for the intercept parameter for survival on the logit scale (denoted α in Section 2) were between 1.11 and 1.76 times larger for the trinomial model than for the complete data model. The average widths of the 95% credible intervals for the slope parameter (denoted β in Section 2) were between 1.20 and 1.55 times wider for the trinomial model.

While these results imply that estimates from the complete data model are more precise, the differences are not great. However, we have realized that Bonner et al. (2010) missed a crucial issue that affects the identifiability of the trinomial model when the effect of the covariate on survival is small. In particular, the survival probability, capture probability, and recovery probability are completely confounded in the trinomial model if β is exactly equal to 0. The lack of identifiability when $\beta = 0$ also causes estimates to be highly imprecise when β is close to 0, meaning that results from the trinomial model are not reliable.

Bonner et al. (2010) did not encounter this problem because the value of β chosen for their simulations, 0.25, is actually quite large. Although the value appears small numerically, the effect on the survival also depends on the variance of the covariate. The model for simulating the covariate was parameterized with estimates obtained by analysing data from the population of Soay sheep on the Isle of Hirta considering body mass, measured in kilograms, as the predictor of survival. The mass at first capture for each sheep was generated from a normal random variable with mean 22.90 and standard deviation 2.90 and the changes in mass between subsequent years were generated from independent normal distributions with means between −2.00 and 2.05 and standard deviation 2.99. The resulting model produces masses between approximately 15 and 40 kg so that the survival probability ranges from about 0.64 to almost 1.00. If the covariate had been standardized, then the corresponding value of β would have been close to 1.50. This represents a relatively large effect on the logit scale.

The remainder of this paper is organized as follows. We define the notation in Section 2.1 and provide complete descriptions of the trinomial and complete data models in Sections 2.2 and 2.3. We present the theoretical results on the identifiability of the trinomial model in Section 2.4. Section 3 provides results from a simulation study exploring the impact of this problem across a range of parameter values, and in Section 4, we compare results from the analysis of data on the meadow vole, *Microtus pennsylvanicus*. Bonner et al. (2010) also analysed these data and found no effect of the covariate, body mass, on the voles' survival. We end by summarizing our findings in Section 5.

2 | MODELS AND METHODS

2.1 | Notation

Suppose that individuals from an open population are captured on T discrete occasions. We let n denote the number of individuals first captured and marked on the first $T-1$ occasions. Individuals first captured on the final occasion are ignored since CJS-based models condition on the first release and these animals do not contribute to the likelihoods. We let $\omega_{it}=1$ if individual i is captured alive on occasion t, $\omega_{it}=2$ if individual i is found dead between occasions $t-1$ and t, and $\omega_{it}=0$ otherwise. The capture history for individual i is denoted by $\boldsymbol{\omega}_i$ and the $n\times T$ matrix of histories for all individuals is denoted by Ω. We consider a single continuous covariate of the survival probability which is denoted by x_{it} for individual i on occasion t. The string of covariates for individual i is denoted by $\boldsymbol{x}_i$ and the $n\times T$ matrix containing all covariate values is denoted by $\boldsymbol{X}$. We assume that x_{it} is observed if and only if $\omega_{it}=1$.

We assume that the capture and recovery probabilities are the same for all individuals and constant over time and denote them by p and λ respectively. However, the results derived in the following sections can be extended to models in which the capture and recovery probabilities are time-dependent. We also impose the standard assumptions that no individuals are lost on capture, events are independent over time and between individuals, and dead animals can only be recovered in the interval in which they die.

The probability that individual i survives from occasion t to occasion $t+1$ is denoted by ϕ_{it}. We assume that variation in the survival probability depends only on the value of x_{it} with the standard logit link so that $\phi_{it}=\phi(x_{it};\alpha,\beta)$, where

$$\phi(x;\alpha,\beta)=\frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)} \tag{1}$$

for some coefficients $\alpha,\beta\in\mathbb{R}$.

2.2 | Trinomial model

Catchpole et al. (2008) constructed the likelihood of the trinomial model by considering only the events directly following releases for each of the marked animals. The likelihood function is

$$L_T(p,\lambda,\alpha,\beta|\Omega,X)=\prod_{i=1}^{n}\prod_{t\in\mathcal{T}_i}P(\omega_{i,t+1}|\omega_{i,t}=1,x_{it}),$$

where $\mathcal{T}_i=\{t|\omega_{i,t}=1,t<T\}$ represents the set of occasions prior to the final occasion on which individual i is captured and released. The individual likelihood contributions are

$$P(\omega_{i,t+1}=1|\omega_{it}=1,x_{it})=p\phi_{it}, \tag{2}$$
$$P(\omega_{i,t+1}=2|\omega_{it}=1,x_{it})=\lambda(1-\phi_{it}), \tag{3}$$

and

$$P(\omega_{i,t+1}=0|\omega_{it}=1,x_{it})=1-p\phi_{it}-\lambda(1-\phi_{it}). \tag{4}$$

The name for the trinomial model comes from the fact that this is equivalent to a multinomial model with three categories.

2.3 | Complete data model

For comparison, we also consider the complete data model of Bonner and Schwarz (2006) extended to allow for recoveries of dead individuals. Following Bonner and Schwarz (2006), we assume that the covariate follows a Markov process such that $X_{i,t+1}$ depends only on X_{it}. The capital letters highlight the fact that this model treats the covariate as a latent random variable whereas the trinomial treats the covariate as fixed and fully observed. We provide details of the exact

model in Section 3 and for now simply let $f(x_{i,t+1}|x_{it}, \boldsymbol{\theta}_x)$ represent the density of $X_{i,t+1}$ given X_{it} dependent on the parameters in $\boldsymbol{\theta}_x$.

To simplify the definition of the likelihood we let $a_i = \min_t\{t|\omega_{it} = 1\}$ denote the first occasion on which individual i is captured and introduce the latent variable D_i to represent the last occasion on which individual i was alive with $D_i = T$ if the individual survived past the final occasion. The complete data likelihood is

$$L_C(p, \lambda, \alpha, \beta, \boldsymbol{\theta}_x|X, \Omega, \boldsymbol{D}) = \prod_{\{i|a_i<D_i\}} \left[\prod_{t=a_i}^{D_i-1} \phi_{it} p^{\mathcal{I}(\omega_{i,t+1}=1)} (1-p)^{\mathcal{I}(\omega_{i,t+1}=0)} f(x_{i,t+1}|x_{it}, \boldsymbol{\theta}_x) \right] \times \prod_{\{i|D_i<T\}} \left[(1-\phi_{iD_i}) \lambda^{\mathcal{I}(\omega_{i,t+1}=2)} (1-\lambda)^{\mathcal{I}(\omega_{i,t+1}=0)} \right],$$

where $\mathcal{I}(\cdot)$ is the standard indicator function. The first line models the survival and recaptures of individuals that survive for at least one occasion. The second line models the death and recovery for individuals that do not survive to the end of the study. Evaluating the observed data likelihood requires integrating across the missing covariates and the unknown times of death, which we handle through MCMC simulation in the Bayesian framework as described in Section 3.

2.4 | Identifiability of the trinomial model

The following two theorems summarize the key results regarding the identifiability of the trinomial model.

Theorem 1. *The trinomial model is identifiable provided that (1) either $p > 0$ or $\lambda > 0$, (2) $\beta \neq 0$, and (3) $K > 2$ unique values of the covariate are observed.*

Proof. Consider the case in which $p > 0$ and let p_1, α_1, and β_1 and p_2, α_2, and β_2 denote two distinct sets of parameters. We show in Appendix A that the functions $p_1\phi(x; \alpha_1, \beta_1)$ and $p_2\phi(x; \alpha_2, \beta_2)$ can intersect more than two times only if $\beta_1 = \beta_2 = 0$ or if $p_1 = p_2$, $\alpha_1 = \alpha_2$, and $\beta_1 = \beta_2$. Given that $\beta = 0$, this means there is a unique solution to the system

$$p\phi(x_k; \alpha, \beta) = p\phi_k, \quad k = 1, \ldots, K,$$

for any $K > 2$. This further implies that changing any of the parameters α, β, or p will change at least some of the likelihood contributions in Equation (2), resulting in a different value for the final likelihood. The condition that $K > 2$ will always be satisfied in the limit as $n \to \infty$ for a continuous covariate, because its support must contain at least one open interval on the real line, and will also be satisfied if the covariate is discrete with support including at least three points. Note also that if ϕ is identifiable then λ is also identifiable through the likelihood contribution in Equation (3), and that the identifiability holds even if $\lambda = 0$.

The proof for the case in which $\lambda > 0$ proceeds in the same way by considering the functions $\lambda_1(1 - \phi(x; \alpha_1, \beta_1))$ and $\lambda_2(1 - \phi(x; \alpha_2, \beta_2))$. ∎

Theorem 2. *The parameters α, λ, and p are not identifiable from the trinomial model if $\beta = 0$.*

Proof. If $\beta = 0$ then the expression in (1) simplifies so that the survival probability becomes a constant,

$$\phi_{it} = \frac{\exp(\alpha)}{1 + \exp(\alpha)} = \phi,$$

for all i and t. The likelihood components in (2), (3), and(4) can then be reparameterized in terms of the two products, $\theta_{11} = \phi p$ and $\theta_{12} = (1 - \phi)\lambda$ (the double subscript match the notation for the transparent reparameterization defined in Appendix C). This shows that the model is intrinsically parameter redundant and hence non-identifiable (Cole, 2020, p. 36). More formally, let $\alpha^\dagger$, $p^\dagger$, and $\lambda^\dagger$ denote the true values of the parameters, $\phi^\dagger = \text{logit}^{-1}(\alpha^\dagger)$, $\theta_{11}^\dagger = \phi^\dagger p^\dagger$, and $\theta_{12}^\dagger = (1 - \phi^\dagger)\lambda^\dagger$. Then for any $\phi \in \left(\theta_{11}^\dagger, 1 - \theta_{12}^\dagger\right)$

$$L_T(p, \lambda, \alpha, 0|\Omega, X) = L_T(p^\dagger, \lambda^\dagger, \alpha^\dagger, 0|\Omega, X),$$

where

$$p = \frac{\theta_{11}^{\dagger}}{\phi^{\dagger}} = \frac{p\phi^{\dagger}}{\phi}, \quad \lambda = \frac{\theta_{12}^{\dagger}}{(1-\phi)} = \frac{\lambda^{\dagger}(1-\phi^{\dagger})}{(1-\phi)},$$

and $\alpha = \text{logit}(\phi)$. Note also that $\left(1-\theta_{12}^{\dagger}\right) - \theta_{11}^{\dagger} = (1-\lambda^{\dagger}) - \phi^{\dagger}(p^{\dagger}-\lambda^{\dagger}) > 0$, meaning that such a ϕ can always be found. ∎

The constraint that $\phi \in (\theta_{11}^{\dagger}, 1-\theta_{12}^{\dagger})$ is required in the proof so that the capture and recover probabilities remain within the unit interval, and the set $\left\{\phi, \lambda, p | \phi \in \left(\theta_{11}^{\dagger}, 1-\theta_{12}^{\dagger}\right), p = \theta_{11}^{\dagger}/\phi, \lambda = \theta_{12}^{\dagger}/(1-\phi)\right\}$ forms the identification region. Maximum likelihood estimates would be constrained to this set as the sample size increases to infinity, but would not converge to a point mass at the true value. As noted in the proof, the identification region for ϕ always has positive width, meaning that it never collapses to a single point. Moreover, the identification region becomes wider as either λ or p decrease, when ϕ is fixed, and as ϕ increases if $p < \lambda$ or as ϕ decreases if $p > \lambda$.

Considering the problem in the Bayesian framework can provide more information because it is possible to study the limiting behaviour of the posterior distribution, as opposed simply deriving the identification region. In particular, we apply the methods of Gustafson (2015) to derive the limiting posterior distributions for each parameter. Expressions for the marginal limiting posterior distributions of p, λ, α, and ϕ when $\beta = 0$ are provided in Appendix C. In our simulation study, we use the limiting posterior distributions as a check on the theoretical results and to compare with the the behavior of the posterior distributions when β is close, but not exactly equal, to 0.

The immediate implication of Theorem 2 is that it is not possible to estimate the actual survival probability from the trinomial model when there is no effect of the covariate on survival, no matter how much data are collected. Although β itself is identifiable, so that it is possible to determine whether the covariate affects survival, the actual survival probability can only be determined to lie within the range $\left(\theta_{11}^{\dagger}, 1-\theta_{12}^{\dagger}\right)$. Similar results also hold for more general versions of the trinomial model in which the survival, capture, and/or recovery probabilities are time dependent so that, for example, $\phi_{it} = \phi(x_{it}; \alpha_t, \beta)$.

Although the theoretical results apply when β is exactly equal to 0, they also raise concern about the use of the trinomial model when β is close to 0. In the following section, we provide the results of a simulation study to examine the performance of the trinomial model under a variety of parameter settings. We also compare the results from the trinomial model with the results from complete data model which is too complex to be studied analytically.

3 | SIMULATION STUDY

3.1 | Setup

The missing covariate values in the complete data models are easily handled through MCMC simulation, and so we conduct inference in the Bayesian framework. We selected non-informative priors for all parameters following the recommendations of Gelman et al. (2008) for logistic regression to define priors for α and β (see Appendix D). We then sampled from the joint posterior distribution of the parameters and missing data for each model via MCMC sampling implemented via the `R` package `nimble` (de Valpine et al., 2017). We generated a total of 60,000 samples from the posterior distribution for each simulated data set running three parallel chains following a burn-in of 2000 iterations per chain.

Simulations were conducted for all possible combinations of the parameter values provided in Table 1. This resulted in a total of 96 different parameter settings which we call scenarios. We simulated 50 data sets for each scenario. The occasion of first capture for individual i, a_i, was sampled uniformly from $\{1, \ldots, T-1\}$ and the value of the covariate at first capture was simulated as $X_{ia_i} \sim \text{N}(0.50, 0.50)$. The values of the remaining covariates were then generated sequentially such that

$$X_{i,t+1} | X_{it} \sim \text{N}\left(X_{it} + \mu_t, \sigma^2\right), \quad t = a_i, \ldots, T-1,$$

with $\mu_t \sim \text{N}(0, 0.25^2)$ and $\sigma^2 = 0.50$. The marginal distribution of the covariate in this scenario and the corresponding survival probabilities for the different combinations of α and β are illustrated in Figure 1. The first combination represents a case in which there is a strong effect of the covariate on survival, $\beta = 4$, so that there should be no issues with identifiability. In the second, the survival probability is high, $\phi = 0.90$, but the covariate has no effect, $\beta = 0$. In this case, the

TABLE 1 Parameters of the simulation study.

Parameter	Values
n	50, 100, 500, 1000
T	10
p	0.25, 0.75
λ	0, 0.25, 0.75
(α, β)	(logit(.45), 4), (logit(.90), 0), (logit(.45), .25), (logit(.90), .25)

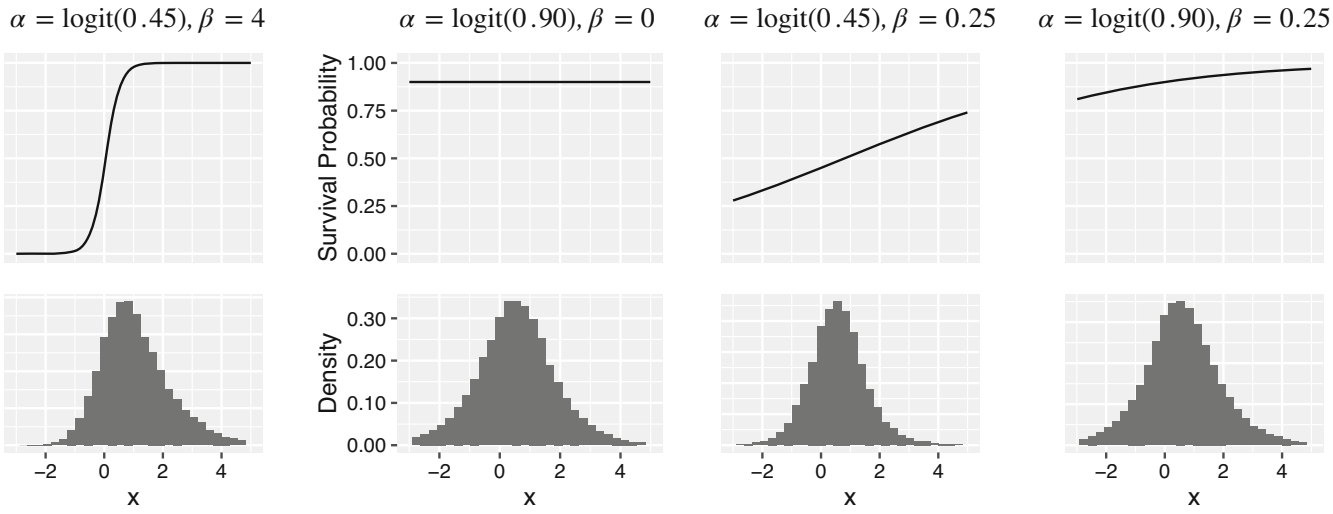

FIGURE 1 Survival probability curves (top) and distribution of covariate values (bottom) for the four combinations of α and β considered in the simulation study.

theoretical results about the identifiability of the trinomial should apply exactly. The final two combinations represent cases in which there is a weak effect of the covariate, $\beta = 0.25$, with either low or high survival probability overall. These represent cases in which the parameters are theoretically identifiable from the trinomial model, but the application in small samples raises concern.

3.2 | Results

3.2.1 | Small sample size

We focus our discussion primarily on inference for α and β since p and λ are generally viewed as nuisance parameters. Figures 2 and 3 show how the mean posterior standard deviations (MPSD) of α and β vary with n and the true values of α, β, p, and λ for the scenarios in which recoveries are possible, $\lambda > 0$. The MPSD for a parameter, generically θ, for a given sample size, n, is

$$\text{MPSD}_n(\theta) = \frac{1}{50}\sum_{k=1}^{50}\text{SD}^{(k)}(\theta),$$

where $\text{SD}^{(k)}(\theta)$ denotes the posterior standard deviation of θ for the kth replicate of size n. Numerical values are provided in Tables S1 and S2 in the Supplementary Materials. If a parameter is identifiable, then the posterior distribution will converge to a point mass at the true value so that the MPSD approaches zero. If not, then the posterior standard deviation will converge to a strictly positive value determined by the limiting posterior distribution. Gustafson (2015, p. 37) called

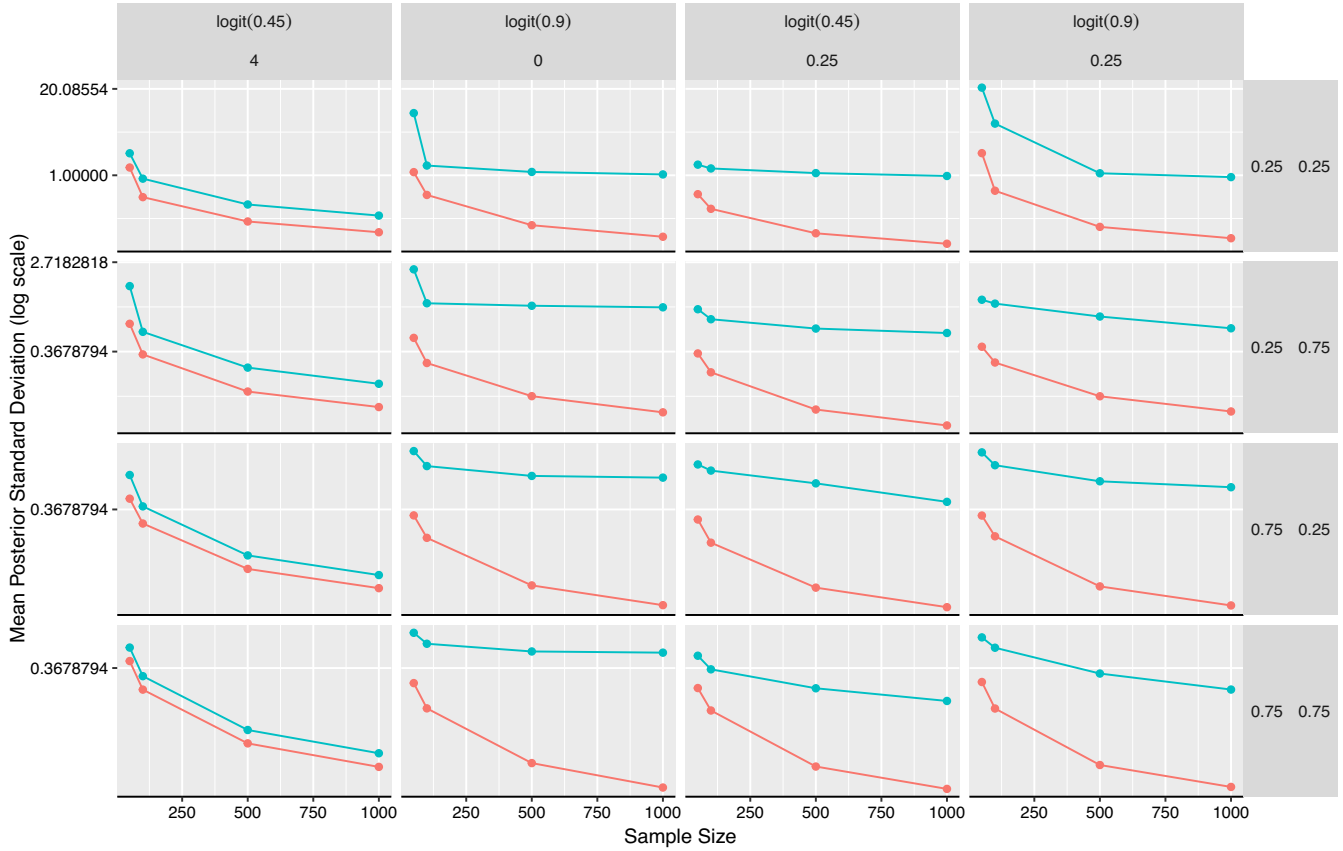

FIGURE 2 Mean posterior standard deviation of α versus the number of marked individuals for each simulation scenario with recoveries ($\lambda > 0$). Results for the trinomial model are shown in blue and for the complete data model are shown in red. The header identifies the values of α (top) and β (bottom) while the labels on the right identify the values of p (left) and λ (right).

this "the tell-tale signature of partial identification". As expected from the theoretical results, the MPSD for both α and β tend toward zero for both the trinomial and complete data models when $\beta = 4$ (shown in the left column of the two plots). The MPSD for β also show similar patterns for the remaining values of α and β (shown in columns 2 through 4 in Figure 3). However, the MPSD of α from the trinomial model does not seem to converge when $\beta = 0$ (second column from the left in Figure 2). Instead, the MPSD decreases slightly when n initially increases from 50 to 100 and then plateaus when n reaches 500 and 1000. In comparison, the MPSD from the complete data model still decreases toward zero. This supports the theoretical results showing that α is not identifiable from the trinomial model when $\beta = 0$ while β remains identifiable in all cases. The results also indicate that α and β are identifiable from the complete data model in all cases.

As noted above, our concern is not only with the lack of identifiability when $\beta = 0$ exactly, but also that this will cause the trinomial model to perform poorly when β is close to 0 as well. Note that the results when $\beta = 0.25$ (final two columns of Figure 2) are much more similar to the case in which $\beta = 0$ than $\beta = 4$. The mean MPSD of α from the trinomial model is much higher than that of the complete data model and converges toward zero more slowly. We know from the theoretical results that α is identifiable in these cases and so the MPSD must converge to 0. However, this happens so slowly in some cases, for example when $p = \lambda = 0.25$, that the difference when $n = 500$ and $n = 1000$ is imperceptible. Once again, the MPSD from the complete data model behaves in almost exactly the same way when $\beta = 0.25$ as when $\beta = 4$. In all cases, the MPSD of β decreases toward zero, even though the standard deviation from the trinomial model is slightly higher than that of the complete data model.

Figures 4 and 5 provide the corresponding results for the scenarios in which there are no recoveries, $\lambda = 0$. Numerical values are provided in Tables S3 and S4 in the Supplementary Materials. The results for α are very similar. The MPSD from the trinomial model is higher relative to that of the complete data model when $\beta = 4$ but still decreases as n increases. This supports our finding that the trinomial model is identifiable even when $\lambda = 0$ provided that β is far from 0. Once again, the MPSD from the trinomial was much higher than that of the complete data model when $\beta = 0$ and did not decrease toward 0 (second column in Figure 4). In fact the MPSD of α increased slightly with n for both values p, though

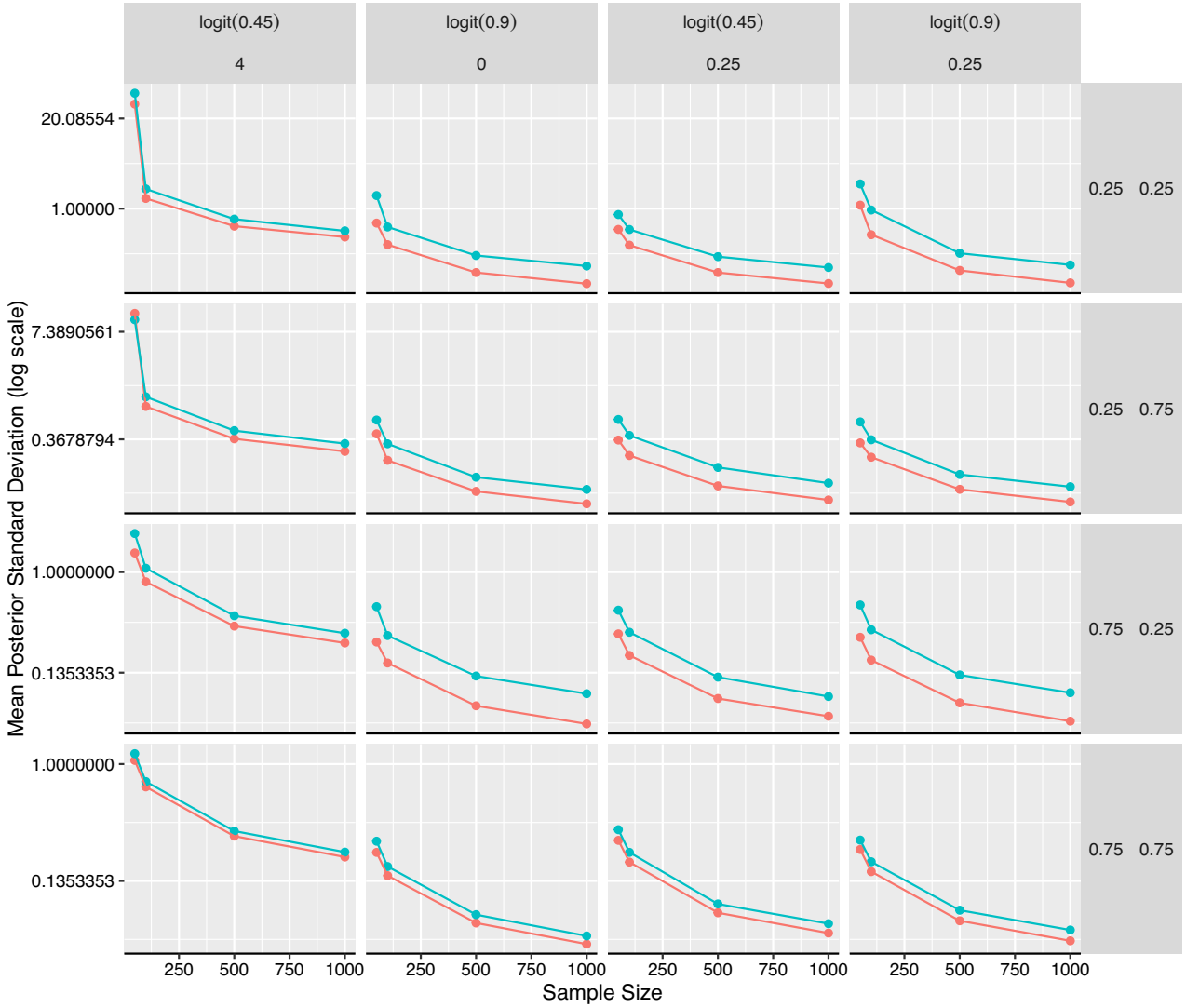

FIGURE 3 Mean posterior standard deviation of β versus the number of marked individuals for each simulation scenario with recoveries ($\lambda > 0$). Results for the trinomial model are shown in blue and for the complete data model are shown in red. The header identifies the values of α (top) and β (bottom) while the labels on the right identify the values of p (left) and λ (right).

this is likely due to simulation error. The MPSD behaves similarly when $\beta = 0.25$ except in the case that $\alpha = \text{logit}(0.45)$ and $p = 0.75$, but even then the MPSD from the trinomial model is several times higher than the MPSD from the complete data model. As noted above, the theory from Section 2 indicates that the MPSD must converge to 0 for these cases, but the simulations indicate that the required sample size will be very large.

What is different in the scenarios in which $\lambda = 0$ is the behavior of the MPSD for β obtained from the trinomial model. While the MPSD for the complete data model decreases as expected, the MPSD from the trinomial model decreases much more slowly, particularly in the scenarios in which $\beta = 0$ or $\alpha = \text{logit}(0.90)$ and $\beta = 0.25$ (second and fourth columns of Figure 5). While Theorem 2 indicates that β should remain identifiable, the simulation results indicate that the precision of this parameter's estimate is also affected.

3.2.2 | Large sample size

To further examine the posterior distributions with large samples, we consider the results for single data sets simulated with sample size $n = 5000$. We also ran the MCMC algorithm for longer to obtain better resolution of the posterior distribution. We generated 100,000 samples from each of three parallel chains after a burn-in period of 10,000 iterations

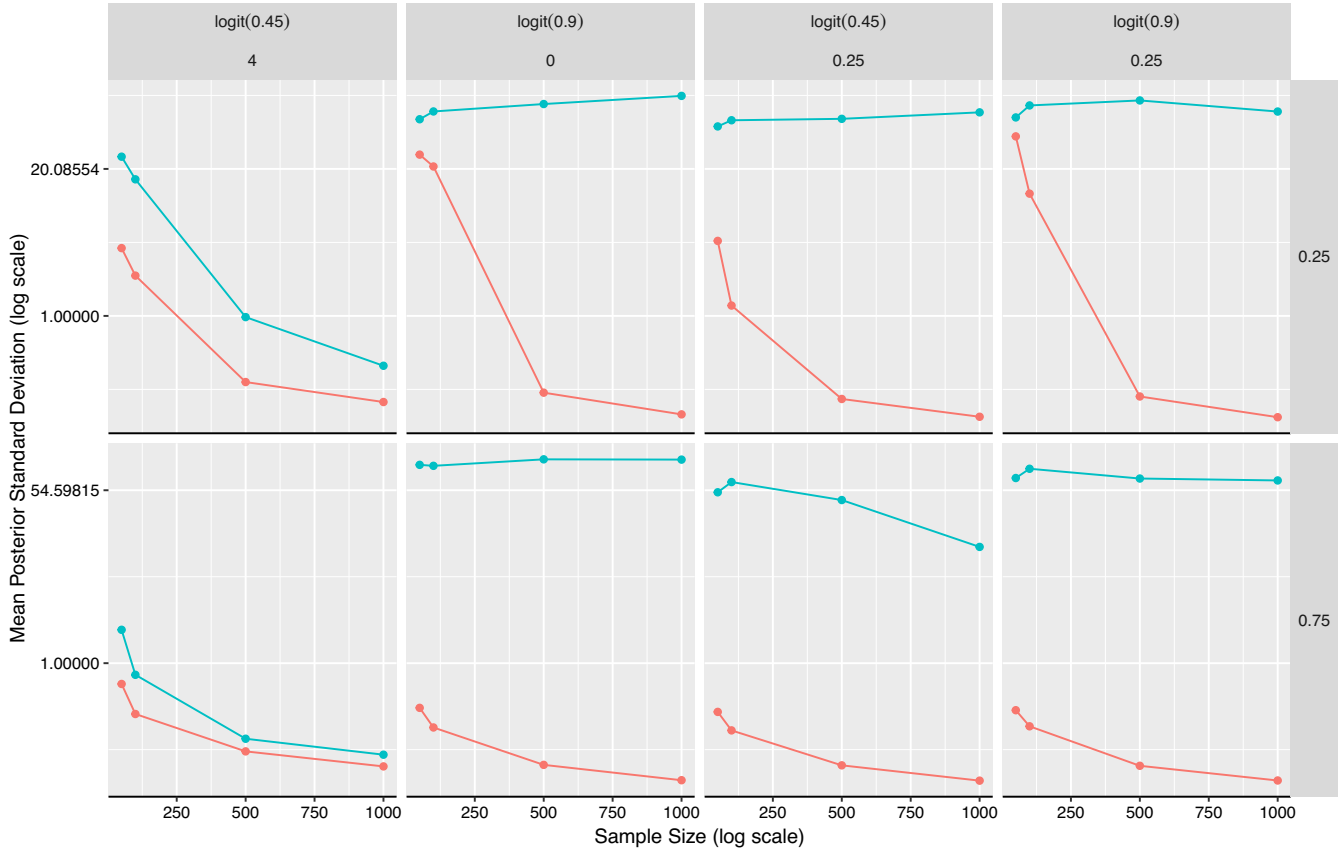

FIGURE 4 Mean posterior standard deviation of α versus the number of marked individuals for each simulation scenario without recoveries ($\lambda = 0$). Results for the trinomial model are shown in blue and for the complete data model are shown in red. The header identifies the values of α (top) and β (bottom) while the labels on the right identify the values of p.

for a total of 300,000 iterations (five times as many as before). Figures 6 and 7 compare smoothed approximations of the posterior densities of α and β, respectively, for the models with recoveries. Note that the differences between the maximum posterior densities are large, and so we scaled the densities to have maximum 1 to make visual comparison easier.

Throughout these plots, the posterior distributions of both α and β from the complete data models are all concentrated about the true parameter values, regardless the of the values of the other parameters. However, the shapes of the posterior distributions from the trinomial model depend heavily on the true parameter values. When $\beta = 4$, the posterior distributions of α and β from the trinomial model are very similar to those of the complete data model, both with and without recoveries. However, the posterior distributions for α are very different when $\beta = 0$. Most importantly, the posterior distributions remain diffuse, despite the large sample size (second column of Figure 6). Further to this, the posterior distribution is skewed and its mode lies well above the true value of α when λ is small, 0.25 (first and third panels in the second column of Figure 6).

When $\beta = 0.25$, the shape of the posterior distribution of α is determined by the values of p and λ. The posterior distribution is centred on the true value and relatively concentrated when both p and λ are large (bottom right panel in Figure 6). However, the posterior distribution becomes more diffuse as p and λ decrease, becoming much more similar to the posterior distribution from the scenarios in which $\beta = 0$. This is particularly true for the case in which $\alpha = \text{logit}(0.90)$ (top panel in the final column of Figure 6). We believe that the difference between the scenarios in which $\alpha = \text{logit}(0.90)$ and $\alpha = \text{logit}(0.45)$ occurs because the same value of β induces much less of a change in the survival probability over the range of the covariate in the first case, as shown in Figure 1.

Figures 6 and 7 present the same results for the scenarios with no recoveries, and the results for α are qualitatively similar. The posterior standard deviation for α from the trinomial model is broad and not centred over the true parameter when $\beta = 0$ or when $\beta = 0.25$. In comparison, the posterior standard deviations from the complete data model concentrate tightly about the true parameter values. Note that the posterior distributions for α from the trinomial model in the bottom right panels of Figure 8 may appear more concentrated than they really are because of the scaling of the x-axis. For

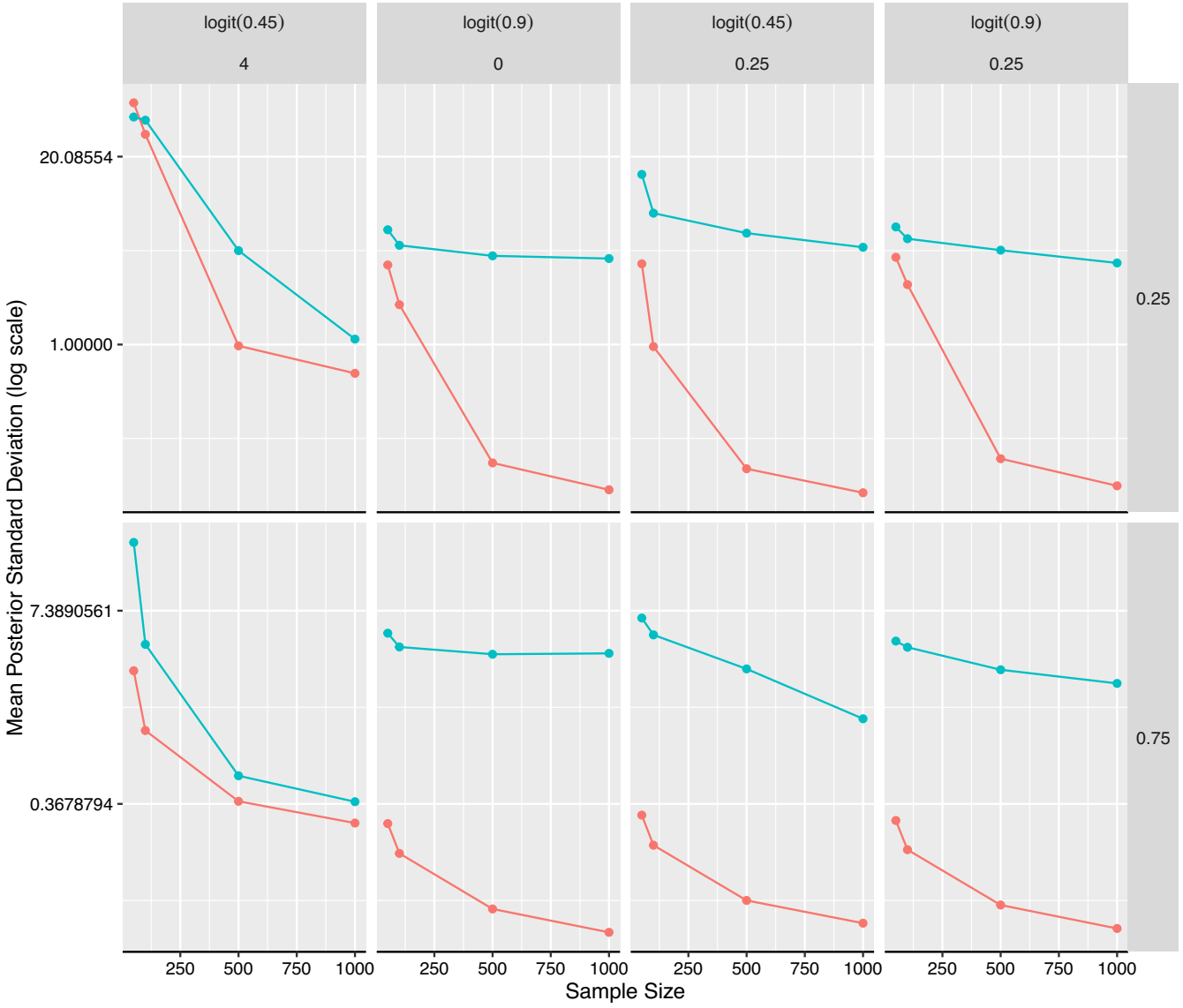

FIGURE 5 Mean posterior standard deviation of β versus the number of marked individuals for each simulation scenario without recoveries ($\lambda = 0$). Results for the trinomial model are shown in blue and for the complete data model are shown in red. The header identifies the values of α (top) and β (bottom) while the labels on the right identify the values of p.

example, the 95% CI for α extends from 0.96 to 3.98 in the case when $\alpha = \text{logit}(0.90)$, $\beta = 0.25$, $p = 0.75$, and $\lambda = 0$. This represents a large range on the logit scale and is much wider than the corresponding interval from the complete data model which runs from 2.14 to 2.27.

As noted in the results for the simulations with smaller sample sizes, the behavior of the posterior distribution for β differs in the scenarios without recoveries. When $\lambda > 0$, the posterior distributions from the trinomial model are almost indistinguishable from those of the complete data model (Figure 7). However, the posterior distributions become much more diffuse when $\lambda = 0$ (Figure 9). For the cases in which $\beta = 0$ (second column of Figure 9), the posterior distribution is broad. When $p = 0.25$, the 95% CI of β from the trinomial model ranges from -7.42 to 6.47 whereas the 95% CI from the complete data model ranges from 0.03 to 0.20. The posterior distribution for β also appears similar when $\alpha = \text{logit}(0.90)$ and $\beta = 0.25$ and is less broad, but still much more diffuse than the posterior distribution from the complete data model, when $\alpha = \text{logit}(0.45)$ and $\beta = 0.25$.

Figures S1–S6 in the Supplementary Materials present the corresponding results for p and λ. In general, these present the same qualitative patterns that we see for α, the other non-identifiable parameter. The posterior distributions from the complete data model converge toward point masses under all scenarios, and the posterior standard deviation decreases

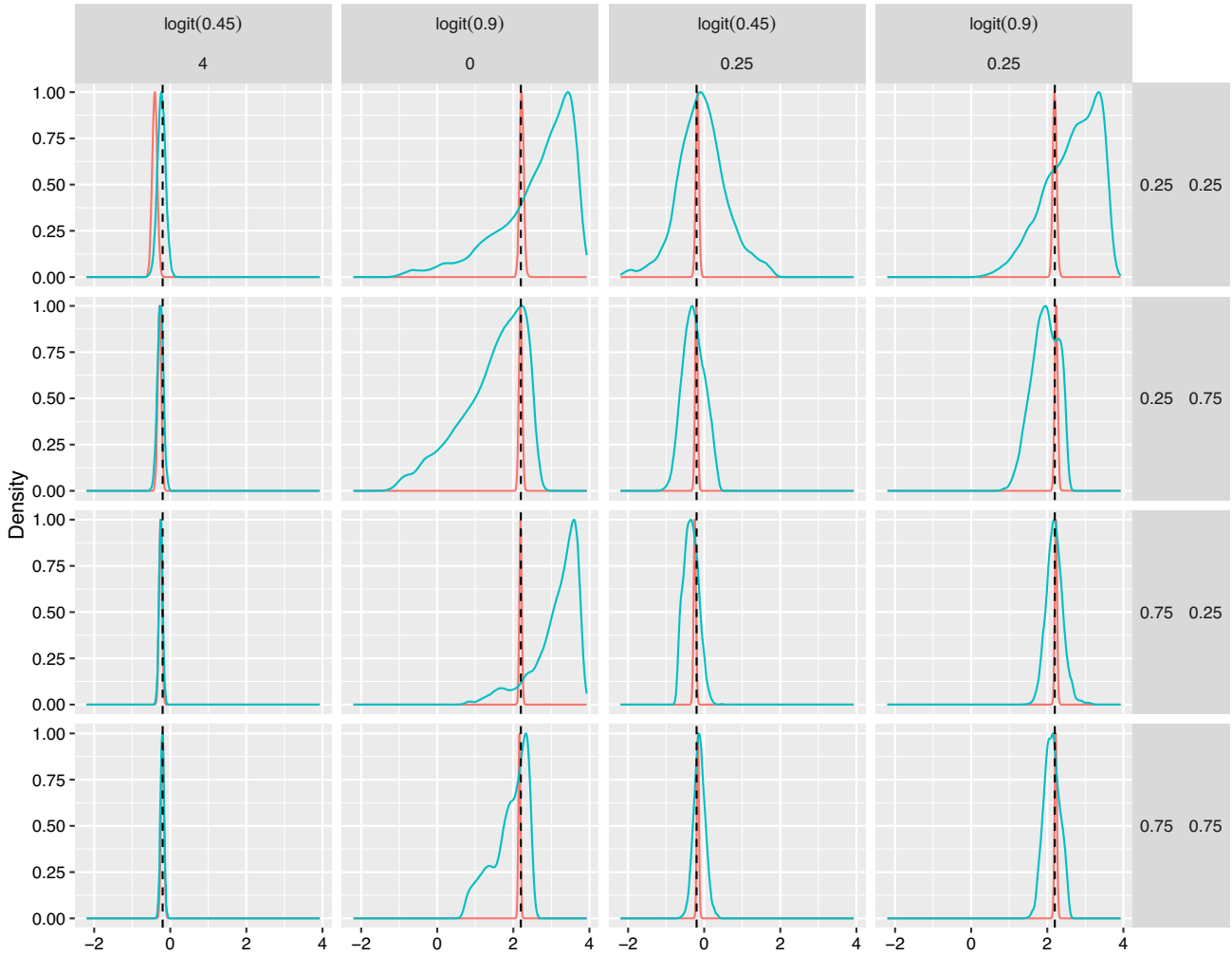

FIGURE 6 Smoothed estimates of the posterior density for α for a single replicate with 5000 individuals for each scenario with $\lambda > 0$. Individual densities have been scaled so that their maxima are at one. Results for the trinomial model are shown in blue and for the complete data model are shown in red. The header identifies the values of α (top) and β (bottom) while the labels on the right identify the values of p (left) and λ (right). The vertical dashed lines identify the true parameter values.

at a similar rate as the sample size increases. However, the posterior standard deviations from the trinomial model do not show the same decrease when $\beta = 0$ or when $\beta = 0.25$, remaining high even when $n = 1000$. One other thing to note is that the posterior distributions can fail to be centered on the true parameter values, even when $n = 5000$. This is most apparent in Figure S5 which shows the posterior distributions for p in the scenarios with no recoveries. The posterior density concentrates at the lower bound of the identification region, and in some cases (e.g., when $\beta = 0$ and $p = 0.75$) there is little overlap between the posterior density from the trinomial model and the posterior density of the complete data model.

4 | APPLICATION

As an example, we analyzed the data on survival of the meadow vole (*Microtus pennsylvanicus*) that was provided by Bonner and Schwarz (2006). These data were obtained from a mark-recapture study (without recoveries of dead individuals) conducted in the fall 1981 and spring of 1982 as originally described in Nichols et al. (1992). The study was organized according to the robust design of Pollock (1982) with four primary periods at monthly intervals each comprising five secondary occasions on subsequent days. A total of 515 voles were captured during the study, and each vole was weighed every time it was captured.

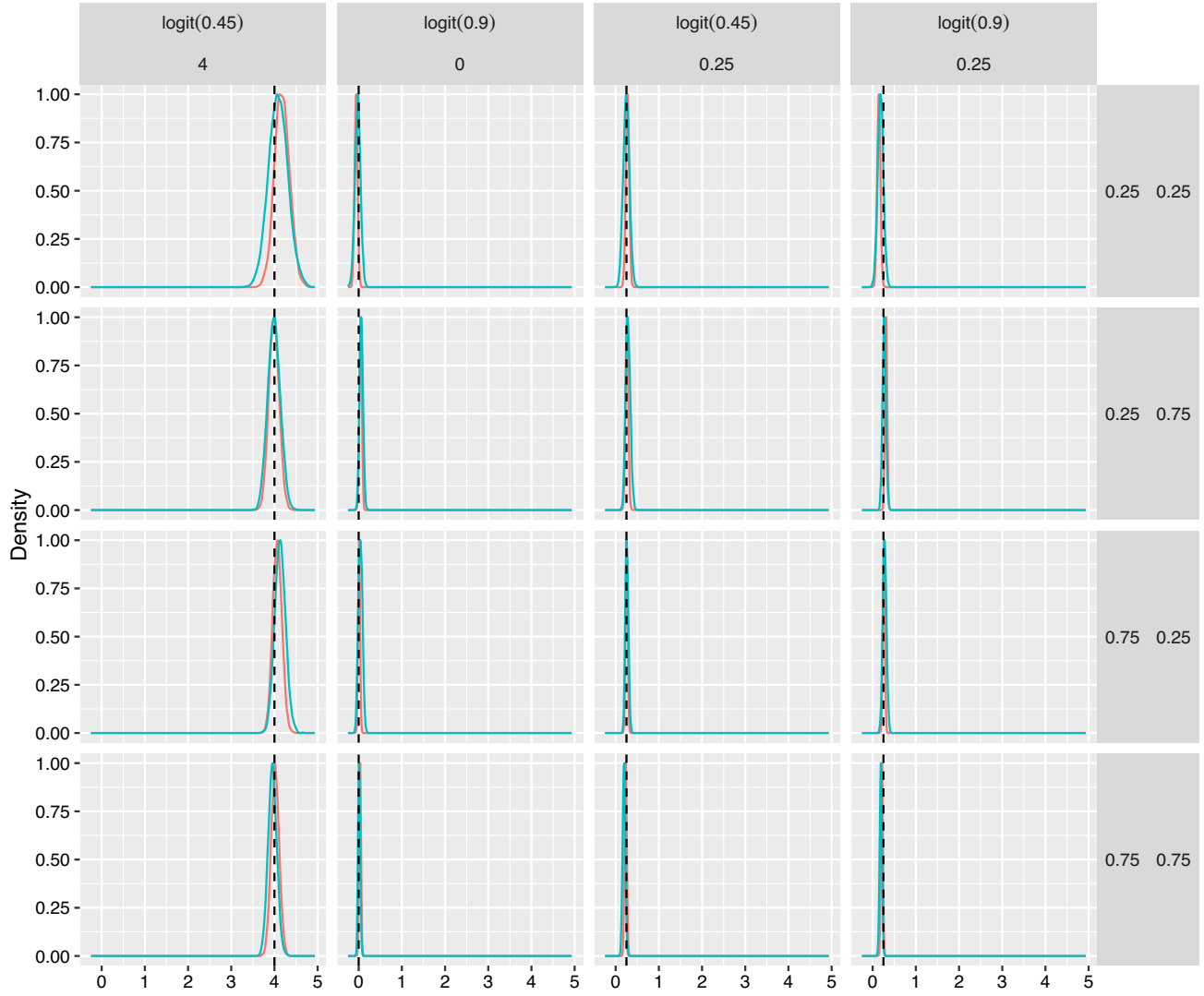

FIGURE 7 Smoothed estimates of the posterior density for β for a single replicate with 5000 individuals for each scenario with $\lambda > 0$. Individual densities have been scaled so that their maxima are at one. Results for the trinomial model are shown in blue and for the complete data model are shown in red. The header identifies the values of α (top) and β (bottom) while the labels on the right identify the values of p (left) and λ (right). The vertical dashed lines identify the true parameter values.

Bonner and Schwarz (2006) aggregated the data within each primary period for each individual to create capture histories of length $T = 4$, retaining the first observed mass within each period in the case of multiple captures. They also removed the records for 300 voles that were first captured and marked on the final occasion, and hence do not contribute to the likelihood, and captures of immature voles, defined to be those with a mass of 22 g or less. The entire history for one other individual with a recorded mass of 0 g was also removed. The final data set contained observations for 199 voles of which 59 (29.6%) were captured on one occasion, 66 (33.2%) on two, 37 (18.6%) on three, and 37 (18.6%) on all four. Body mass ranged from 23.0 to 60.0 g with a mean of 38.4 g.

We chose this example because the complete data model of Bonner and Schwarz (2006) found no evidence that body mass affected the survival probability, suggesting that the trinomial model may face the identifiability problem. The model Bonner and Schwarz (2006) fit to the data allowed for body mass to affect both the survival and capture probabilities within the CJS model, but they found no evidence to support either effect. The estimated coefficients of body mass in the logistic models of the survival and capture probabilities were 0.00 (95% CI = −0.03, 0.02) and 0.01 (95% CI = −0.07, 0.09), respectively. We analyzed the data by fitting the models described in Section 2 allowing only for the effect of body mass on the survival probability, since the trinomial model cannot accommodate effects on the capture probability. We also fixed $\lambda = 0$ since the study did not allow for the recovery of dead individuals.

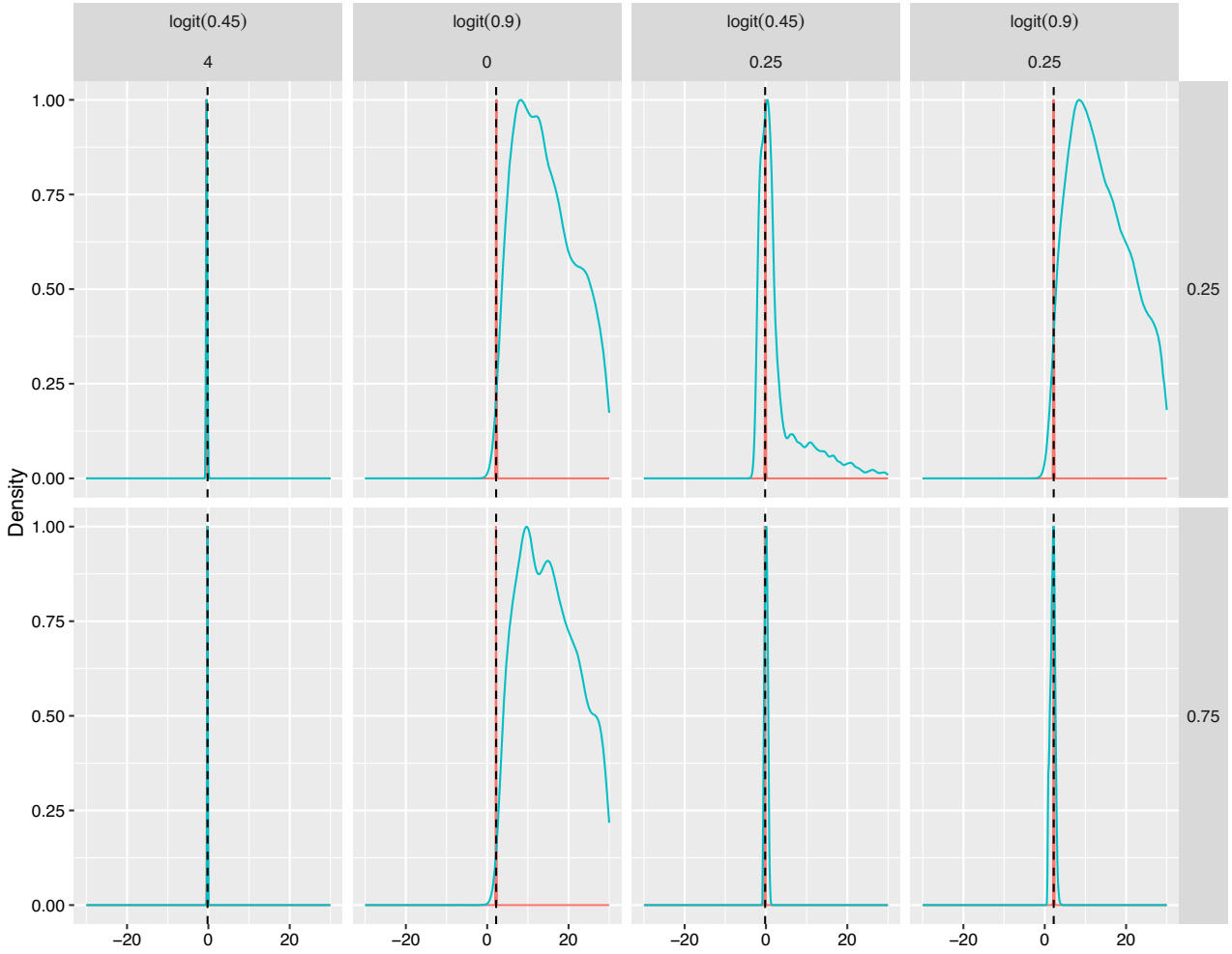

FIGURE 8 Smoothed estimates of the posterior density for α for a single replicate with 5000 individuals for each scenario with $\lambda = 0$. Individual densities have been scaled so that their maxima are at one. Results for the trinomial model are shown in blue and for the complete data model are shown in red. The header identifies the values of α (top) and β (bottom) while the labels on the right identify the values of p (left). The vertical dashed lines identify the true parameter values.

The posterior densities from the two models are compared in Figure 10 and summary statistics are provided in Table 2. The point estimates and credible intervals from the complete data model match the results from Bonner and Schwarz (2006). This is as expected since the models are essentially the same, given that Bonner and Schwarz (2006) found no effect of body mass on the probability of capture. However, the estimates from the trinomial model are significantly different. As in the simulation study when β is close to zero, the posterior variance of α is very large, with 95% CI $(-9.36, 288.85)$. Fixing $\beta = 0$, this would correspond to 95% CI for ϕ that covers the entire real line (to four significant digits). Further to this, the posterior distribution for β obtained from the trinomial model is much broader, with 95% CI covering $(-0.80, 0.98)$ versus only $(-0.03, 0.02)$ for the complete data model. This matches the behavior from the simulation study seen in Figure 9. The posterior distributions for p from the two models also show very little overlap, again matching the results from the simulation study shown in Figure S5 in the Supplementary Materials. Numerically, the 95% CIs do not overlap at all. In fact, the distance between the closest points in the two intervals is the same as the widths of the CIs themselves.

5 | DISCUSSION

The theoretical results in Section 2.4 and our simulation study in Section 3 show that the trinomial model is not identifiable when there is no effect of the covariate on the survival probability, $\beta = 0$. When this occurs, the capture, recovery,

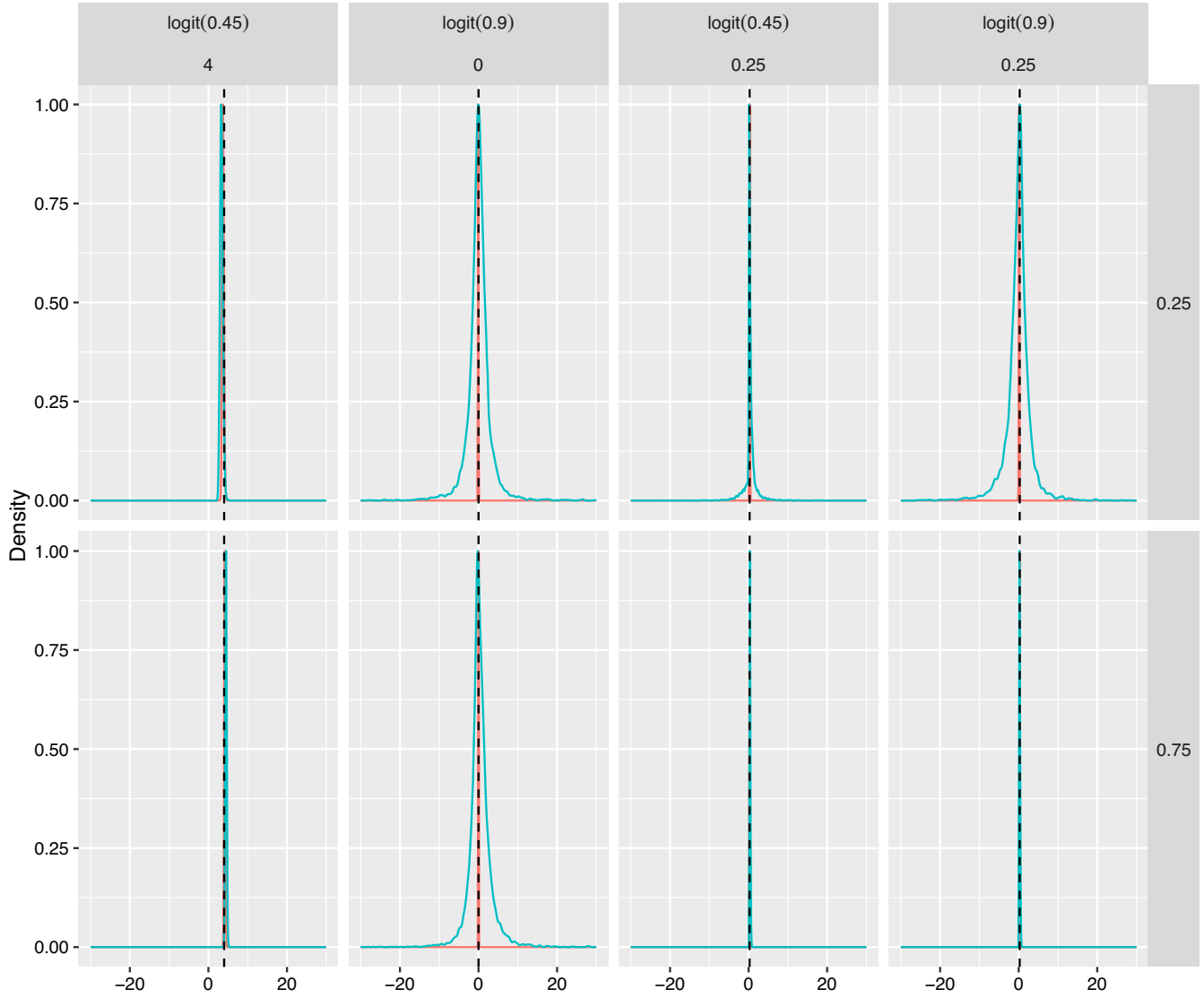

FIGURE 9 Smoothed estimates of the posterior density for β for a single replicate with 5000 individuals for each scenario with $\lambda = 0$. Individual densities have been scaled so that their maxima are at one. Results for the trinomial model are shown in blue and for the complete data model are shown in red. The header identifies the values of α (top) and β (bottom) while the labels on the right identify the values of p (left). The vertical dashed lines identify the true parameter values.

and survival probabilities become confounded, and the likelihood does not have a unique maximum, even as the sample size increases to infinity. Instead, the parameters are only identifiable within a certain range of values, the identification region. This region is defined by the two identifiable quantities in the transparent reparameterization defined in Appendix C, $\theta_{11} = \phi p$ and $\theta_{12} = (1 - \phi)\lambda$, representing the proportions of recaptures and recoveries that contribute to the likelihood of the trinomial model. In particular, the identification region of the survival probability is bounded by θ_{11} and $1 - \theta_{12}$, which get closer to 0 and 1, respectively, as the true capture and recovery probabilities decrease. In extreme cases, it may be possible to say little more than that the survival probability is between 0 and 1, no matter the sample size.

What is more concerning is that these results also impact inference from the trinomial model when the effect of the covariate on the survival probability is small, that is, β is close to zero. According to the theory we have provided, all parameters in the trinomial model are identifiable provided that β is not exactly equal to zero. However, the results of the simulation scenarios in which $\beta = 0.25$ are qualitatively similar to the case in which $\beta = 0$. The uncertainty about the parameters p, λ, and α is much higher for the trinomial model than for the complete data model, and the posterior distributions from the trinomial model converge slowly to point masses at the true parameter values. Quantifying exactly how close β needs to be to 0 for this to be a significant issue is a challenge because the effect depends on several factors.

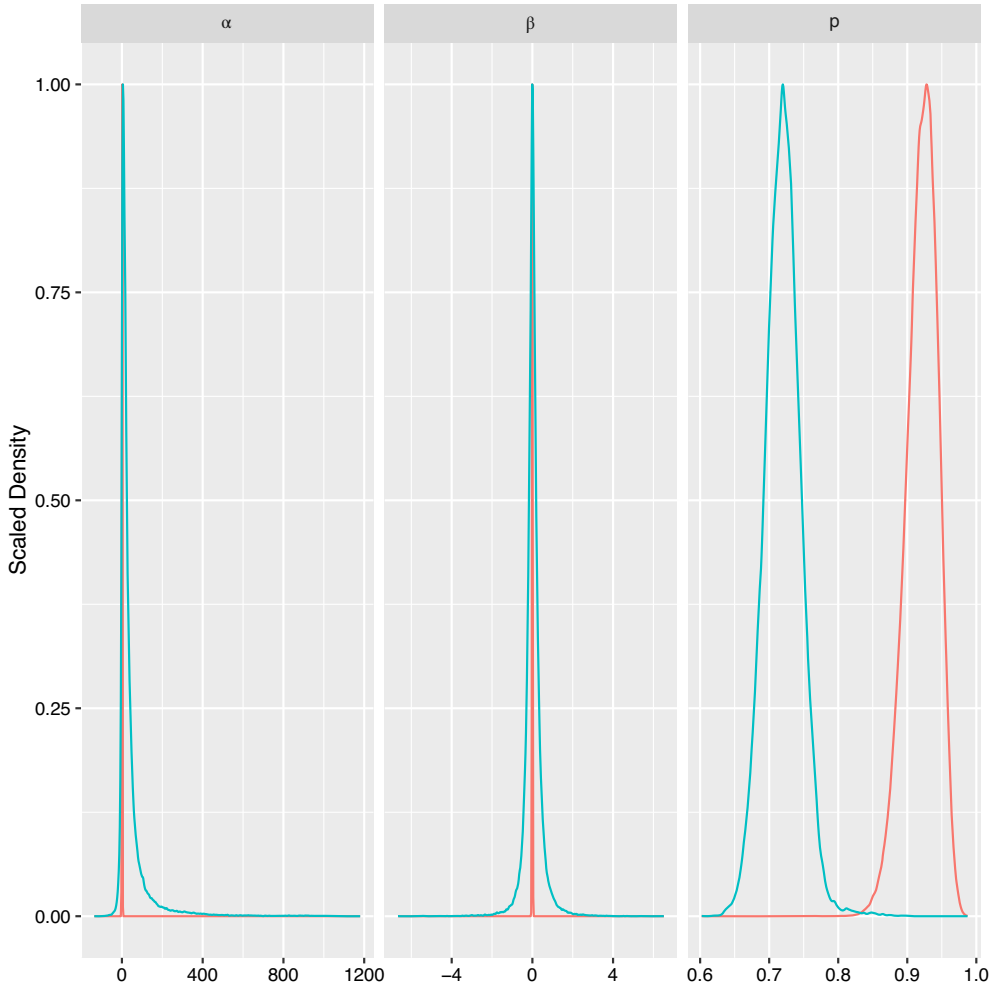

FIGURE 10 Posterior densities for the parameters α, β, and p for the analysis of the meadow vole data. Results for the trinomial model are shown in blue and for the complete data model are shown in red.

TABLE 2 Posterior summary statistics from the analysis of the meadow vole data.

Parameter	Trinomial	Complete
α	43.55 (−9.36, 288.85)	1.39 (0.32, 2.49)
β	0.03 (−0.80, 0.98)	0.00 (−0.03, 0.02)
p	0.72 (0.67, 0.77)	0.92 (0.87, 0.96)

Note: The table provides the posterior means and 95% credible intervals for the parameters modelling capture and survival from the trinomial model (left) and the complete data model (right).

As noted in the introduction, the range of the survival probability across the population depends on both the value of the coefficient and the scale of the covariate, so that the estimate of β is not informative by itself. Even if the covariate is standardized, the variation in the survival probability also depends on the value of α so that the range of the survival probability is less for a given value of β when α is either far below or far above 0. Finally, the magnitude of the problem increases when the capture and recovery probabilities decrease.

One important difference between the simulation scenarios in which there are and are not recoveries is the inference concerning the slope parameter, β. The theory shows that β remains identifiable in all cases. This means that the trinomial model should be able to estimate the size of this effect and to determine if the covariate affects the survival probability, provided that the sample size is large enough. However, the posterior distribution of β remains broad when β is close to zero, even when the sample size is large. We believe that this is due in large part to the prior distribution assigned to α. If

$\beta = 0$, then the Cauchy prior for α recommended by Gelman et al. (2008) places 70% of its mass on values of α bigger than 5 in absolute value. When constrained to the identification region, this pushes the survival probability close to 1 for all individuals. At this point, the value of β has almost no impact on the likelihood so that its posterior distribution is driven largely by the prior distribution.

We also believe that this causes the posterior distributions for p to concentrate around points well below the true value in these scenarios. The product $p\phi$ is constant across the identification, but the value of p is bounded below by $p^{\dagger}\phi^{\dagger}$ (since $\lambda = 0$). If the limiting posterior distribution of ϕ concentrates near 1 then the limiting posterior distribution of p will concentrate near this lower bound to maintain the balance. This suggests that inference from the trinomial model is sensitive to the prior distribution when the data do not include recovery of dead individuals, but also further highlights the challenges with identifiability when applying the trinomial model if the true value of β is close to 0. If the issue with identifiability is not acknowledged, then researchers may draw incorrect conclusions about the population under study.

The trinomial model does have the advantage that it does not require a model for the covariate, which makes modeling simpler and more robust in extreme cases, as shown in Bonner et al. (2010). The trinomial model can also be fit via maximum likelihood inference, avoiding the need for MCMC or other complex computational methods. Although specific algorithms have been developed for some models (for example, Hooten et al., 2023; King et al., 2016), MCMC remains computationally expensive for complex models with many latent variables and large data sets. However, the results we present raise significant concerns about the application of the trinomial model in cases where the covariate has a weak effect on survival. Overall, we recommend that complete data models be applied and that the model of the covariate be tested to ensure that the fitted model is appropriate for the data.

ACKNOWLEDGMENTS

We are grateful to Dr H. Sendov who provided key information on Descartes' rule. Thanks also to C. Araújo-Wang, A. Draghici, J. de Haan Ward, H. Kim, H. Steeves, B. Walker, and to the associate editor and two anonymous reviewers who provided comments on previous drafts. Our research was supported by the Natural Sciences and Engineering Research Council of Canada (grant number 43024-2016).

CONFLICT OF INTEREST STATEMENT

We declare that there are no sources of conflict of interest.

DATA AVAILABILITY STATEMENT

Code and data to replicate the simulation study in Section 3 and to run the analysis of the meadow vole data described in Section 4 are provided in the supplementary archive. We are happy for these files to be posted to figshare if the manuscript is accepted for publication. Data analyzed in this paper is included in the supplementary materials of the original source which are available from Wiley Online Library at https://onlinelibrary.wiley.com/doi/10.1111/j.1541-0420.2005.00399.x.

ORCID

Simon J. Bonner https://orcid.org/0000-0003-2063-4572
Wei Zhang https://orcid.org/0000-0002-7554-5115

REFERENCES

Argáez, V., Solano-Zavaleta, I., & Zúñiga-Vega, J. J. (2020). Do ectoparasites affect survival of three species of lizards of the genus sceloporus? *Zoology*, *138*, 125723.

Bonner, S. J. (2003). *Continuous, individual, time-dependent covariates in the Cormack-jolly-Seber model*. Master's thesis. Simon Fraser University.

Bonner, S. J. (2013). Implementing the trinomial MARK-recapture-recovery model in program MARK. *Methods in Ecology and Evolution*, *4*(1), 95–98.

Bonner, S. J., Morgan, B. J. T., & King, R. (2010). Continuous covariates in mark-recapture-recovery analysis: A comparison of methods. *Biometrics*, *66*(4), 1256–1265.

Bonner, S. J., & Schwarz, C. J. (2006). An extension of the Cormack-jolly-Seber model for continuous covariates with application to *Microtus pennsylvanicus*. *Biometrics*, *62*(1), 142–149.

Borchers, D. L., & Efford, M. G. (2008). Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*, *64*(2), 377–385.

Catchpole, E. A., Morgan, B. J. T., & Coulson, T. (2004). Conditional methodology for individual case history data. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *53*(1), 123–131.

Catchpole, E. A., Morgan, B. J. T., & Tavecchia, G. (2008). A new method for analysing discrete life history data with missing covariate values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(2), 445–460.
Cole, D. J. (2020). *Parameter redundancy and identifiability*. Chapman and Hall/CRC.
de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with nimble. *Journal of Computational and Graphical Statistics*, *26*, 403–413.
Diana, A., Matechou, E., Griffin, J. E., Jhala, Y., & Qureshi, Q. (2022). A vector of point processes for modeling interactions between and within species using capture-recapture data. *Environmetrics*, *33*(8), e2781.
Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383.
Gustafson, P. (2015). *Bayesian inference for partially identified models: Exploring the limits of limited data*. In *Monographs on statistics and applied probability* (Vol. *141*). Chapman and Hall/CRC Press.
Hooten, M. B., Schwob, M. R., Johnson, D. S., & Ivan, J. S. (2023). Multistage hierarchical capture–recapture models. *Environmetrics*, *34*(6), e2799.
King, R., McClintock, B. T., Kidney, D., & Borchers, D. (2016). Capture-recapture abundance estimation using a semi-complete data likelihood approach. *The Annals of Applied Statistics*, *10*(1), 264–285.
Langrock, R., & King, R. (2013). Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Annals of Applied Statistics*, *7*(3), 1709–1732.
McLaughlin, P., & Bar, H. (2021). A spatial capture–recapture model with attractions between individuals. *Environmetrics*, *32*(1), e2653.
Nichols, J. D., Sauer, J. R., Pollock, K. H., & Hestbeck, J. B. (1992). Estimating transition probabilities for stage-based population projection matrices using capture-recapture data. *Ecology*, *73*(1), 306–312.
Pollock, K. H. (1982). A capture-recapture design robust to unequal probability of capture. *The Journal of Wildlife Management*, *46*(3), 752–757.
Pradel, R. (2005). Multievent: An extension of multistate capture-recapture models to uncertain states. *Biometrics*, *61*(2), 442–447.
Rahman, Q. I., & Schmeisser, G. (2002). *Analytic theory of polynomials*. Volume 26 of *London Mathematical Society Monographs*. Oxford University Press Inc.
White, G. C., & Burnham, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study*, *46*, 120–139.
Worthington, H., King, R., & Buckland, S. T. (2015). Analysing mark-recapture-recovery data in the presence of missing covariate data via multiple imputation. *Journal of Agricultural, Biological and Environmental Statistics*, *20*(1), 28–46.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bonner, S. J., Zhang, W., & Mu, J. (2024). On the identifiability of the trinomial model for mark-recapture-recovery studies. *Environmetrics*, *35*(1), e2827. https://doi.org/10.1002/env.2827

APPENDIX A. IDENTIFIABILITY OF SCALED LOGISTIC FUNCTIONS

Consider the scaled logistic function $f(x;\alpha,\beta,p) = p\phi(x;\alpha,\beta)$ where $\text{logit}(\phi(x;\alpha,\beta)) = \alpha + \beta x$, $0 < p < 1$, and $\beta \neq 0$ so that $\phi(x;\alpha,\beta)$ is not a constant. The number of crossings of $f(x;\alpha_1,\beta_1,p_1)$ and $f(x;\alpha_2,\beta_2,p_2)$ is equal to the number of roots of

$$f(x;\alpha_1,\beta_1,p_1) - f(x;\alpha_2,\beta_2,p_2) = 0$$

which is equivalent to

$$p_1 e^{\alpha_1+\beta_1 x}\left(1 + e^{\alpha_2+\beta_2 x}\right) - p_2 e^{\alpha_2+\beta_2 x}\left(1 + e^{\alpha_1+\beta_1 x}\right) = 0. \tag{A1}$$

Letting $y = e^x$, we can rewrite Equation (A1) as

$$c_0 y^{\beta_1} + c_1 y^{\beta_2} + c_2 y^{\beta_1+\beta_2} = 0, \tag{A2}$$

where $c_0 = e^{\alpha_1}p_1$, $c_1 = -e^{\alpha_2}p_2$, and $c_2 = e^{\alpha_1}e^{\alpha_2}(p_1 - p_2)$. Dividing through by y^{β_1} then yields a generalized polynomial

$$c_0 + c_1 y^{\beta_2-\beta_1} + c_2 y^{\beta_2} = 0, \tag{A3}$$

where we assume, without loss of generality, that $\beta_2 > \beta_1$ and $\beta_2 > 0$ so that the exponents are all positive.

Descartes' rule (see Rahman & Schmeisser, 2002, p. 391) states the number of positive roots of a polynomial of degree n, $f(x) = \sum_{v=0}^{n} c_v x^v$ is less than or equal to the number of sign changes in the sequence of coefficients, $c_0, c_1, c_2, \ldots, c_n$. For example, the polynomial $c_0 + c_1 x + c_2 x^2$ will have no positive roots if $\text{sign}(c_0) = \text{sign}(c_1) = \text{sign}(c_2)$, one positive root if either $\text{sign}(c_0) = -\text{sign}(c_1)$ or $\text{sign}(c_1) = -\text{sign}(c_2)$ but not both, and two positive roots if $\text{sign}(c_0) = -\text{sign}(c_1) = \text{sign}(c_2)$. This result also extends to generalized polynomials with non-integer exponents (Rahman & Schmeisser, 2002, p. 353).

Noting that a_1, a_2, p_1, and p_2 are all positive, the number of sign changes in c_0, c_1, and c_2 depends only on the relative values of p_1 and p_2. If $p_1 > p_2$ then $\text{sign}(c_0) = -\text{sign}(c_1) = -\text{sign}(c_2)$ and if $p_1 > p_2$ then $\text{sign}(c_0) = -\text{sign}(c_1) = \text{sign}(c_2)$. Either way, the number of sign changes is less than or equal to two. This means that Equation (A2) can have at most two roots in $\mathbb{R}^+$ or, equivalently, Equation (A1) can have at most two roots in all of $\mathbb{R}$, except in the trivial case that $p_1 = p_2$, $\alpha_1 = \alpha_2$, and $\beta_1 = \beta_2$ in which case the functions are exactly equal.

APPENDIX B. IDENTIFICATION REGION

The identification region of a model is defined by the subset of the parameter space for which the likelihood is identical to the likelihood given the true parameter values. For the trinomial model, this is the set of parameters such that

$$p\phi(x;\alpha,\beta) = p^\dagger\phi(x;\alpha^\dagger,\beta^\dagger)$$

and

$$\lambda(1-\phi(x;\alpha,\beta)) = \lambda^\dagger(1-\phi(x;\alpha^\dagger,\beta^\dagger)).$$

If $\beta = 0$ then $\phi(x;\alpha,\beta) = \exp(\alpha)/(1+\exp(\alpha)) = \phi$ for all x so that these equations become

$$p\phi = p^\dagger\phi^\dagger$$

and

$$\lambda(1-\phi) = \lambda^\dagger(1-\phi^\dagger).$$

Subject to the constraints that $0 < p < 1$ and $0 < \lambda < 1$, these equations are satisfied by any values ϕ, $p = p^\dagger\phi^\dagger/\phi$, and $\lambda = \lambda^\dagger(1-\phi^\dagger)/(1-\phi)$ where $p^\dagger\phi^\dagger < \phi < 1 - \lambda^\dagger(1-\phi^\dagger)$. Hence, the identification region is

$$\{(p,\lambda,\phi) : p^\dagger\phi^\dagger < \phi < 1-\lambda^\dagger(1-\phi^\dagger), p = p^\dagger\phi^\dagger/\phi, \lambda = \lambda^\dagger(1-\phi^\dagger)/(1-\phi)\}.$$

APPENDIX C. LIMITING POSTERIOR DISTRIBUTION

We initially consider the parameterization of the trinomial model in terms of p, λ, and $\phi = (1+\exp(-\alpha))^{-1}$ since this is simpler to work with when $\beta = 0$. We then transform the limiting posterior distribution (LPD) of ϕ to compute the LPD of α. To derive the LPD, we must identify what Gustafson (2015) calls a transparent reparameterization: a smooth and invertible function of the parameters, $(\theta_1, \theta_2) = h(p, \lambda, \phi)$, such that (1) the distribution of the data depends only on θ_1 and not on θ_2 and (2) standard asymptotic theory applies to the posterior distribution of θ_1. In particular, the posterior distribution of θ_1 must converge to a point mass at its true value, denoted by $\theta_1^\dagger$, as the sample size increases (assuming that the model is correctly specified) (Gustafson, 2015, p. 16). Given a transparent reparameterization, the LPD for θ_2 is equal to the conditional prior distribution for θ_2 given that $\theta_1 = \theta_1^\dagger$ (Gustafson, 2015, p. 16). The density is

$$\pi_\infty(\theta_2) = \pi(\theta_2|\theta_1 = \theta_1^\dagger) = \pi(h^{-1}(\theta_1^\dagger, \theta_2))|h^{-1}(\theta_1, \theta_2)|.$$

The transformation can then be inverted to construct the LPD for the original parameters.

Consider the reparameterization $\boldsymbol{\theta}_1 = (\phi p, (1 - \phi\lambda))$, $\theta_2 = \phi$. The likelihood of the trinomial model depends only on $\boldsymbol{\theta}_1$ and this quantity is constant across the identification region when $\beta = 0$. This means that $\boldsymbol{\theta}_1$ is identifiable and will satisfy the usual asymptotic results for Bayesian inference. Further to this, the transformation is invertible

$$h^{-1}(\boldsymbol{\theta}_1, \theta_2) = \left(\frac{\theta_{11}}{\theta_2}, \frac{\theta_{12}}{1-\theta_2}, \theta_2\right)$$

where θ_{11} and θ_{12} denote the individual elements of $\boldsymbol{\theta}_1$. Hence, $h(\cdot)$ constitutes a transparent reparameterization of the trinomial model when $\beta = 0$.

To find the marginal LPD of ϕ, note that

$$|h^{-1}(\boldsymbol{\theta}_1, \theta_2)| = \frac{1}{\theta_2(1-\theta_2)}.$$

The density of the conditional prior on θ_2 given $\boldsymbol{\theta}_1$ is then

$$\pi(\theta_2|\boldsymbol{\theta}_1) \propto \frac{\pi\left(\frac{\theta_{11}}{\theta_2}, \frac{\theta_{12}}{(1-\theta_2)}, \theta_2\right)}{\theta_1(1-\theta_2)},$$

subject to the constraints

$$0 < \frac{\theta_{11}}{\theta_2} < 1, \quad 0 < \frac{\theta_{12}}{(1-\theta_2)} < 1, \text{ and } 0 < \theta_2 < 1.$$

Here $\pi(p, \lambda, \phi)$ represents the joint prior density of p, λ, and ϕ. Substituting the true value of $\boldsymbol{\theta}_1$ we obtain

$$\pi(\theta_2|\boldsymbol{\theta}_1^\dagger) \propto \frac{\pi\left(\frac{\theta_{11}^\dagger}{\theta_2}, \frac{\theta_{12}^\dagger}{1-\theta_2}, \theta_2\right)}{\theta_2(1-\theta_2)},$$

with the single constraint $\theta_{11}^\dagger < \theta_2 < 1 - \theta_{12}^\dagger$.

The marginal LPDs of the original parameters can then be found by applying the correct transformations of $\boldsymbol{\theta}_1$ and θ_2, namely $\phi = \theta_2$, $p = \theta_{11}/\theta_2$, and $\lambda = \theta_{12}/(1-\theta_2)$. The resulting densities are:

$$\pi_\infty(\phi) \propto \frac{\pi\left(\frac{\theta_{11}^\dagger}{\phi}, \frac{\theta_{12}^\dagger}{(1-\phi)}, \phi\right)}{\phi(1-\phi)}, \quad \theta_{11}^\dagger < \phi < 1 - \theta_{12}^\dagger$$

$$\pi_\infty(p) \propto \frac{\pi\left(p, \frac{\theta_{12}^\dagger}{1-\theta_{11}^\dagger/p}, \frac{\theta_{11}^\dagger}{p}\right)}{p - \theta_{11}^\dagger}, \quad \frac{\theta_{11}^\dagger}{1-\theta_{12}^\dagger} < p < 1$$

$$\pi_\infty(\lambda) \propto \frac{\pi\left(\frac{\theta_{11}^\dagger}{1-\theta_{12}^\dagger/\lambda}, \lambda, 1 - \frac{\theta_{12}^\dagger}{\lambda}\right)}{\lambda - \theta_{12}^\dagger}, \quad \frac{\theta_{12}^\dagger}{1-\theta_{11}^\dagger} < \lambda < 1,$$

where $\theta_{11}^\dagger = \phi^\dagger p^\dagger$ and $\theta_{12}^\dagger = (1-\phi^\dagger)\lambda^\dagger$. In the specific case in which $p \sim \text{U}(0,1)$, $\lambda \sim \text{U}(0,1)$, and the prior on ϕ is induced by the prior on α, $\pi_\alpha(\cdot)$, so that

$$\pi(p, \lambda, \phi) \propto \frac{\pi_\alpha(\text{logit}\phi)}{\phi(1-\phi)},$$

these become

$$\pi_\infty(\phi) \propto \frac{\pi_\alpha\big(\text{logit}(\phi)\big)}{[\phi(1-\phi)]^2}, \quad \phi^\dagger p^\dagger < \phi < 1-(1-\phi^\dagger)\lambda^\dagger$$

$$\pi_\infty(p) \propto \frac{\pi_\alpha\left(\text{logit}\left(\frac{\phi^\dagger p^\dagger}{p}\right)\right)}{\left(1-\frac{\phi^\dagger p^\dagger}{p}\right)^2}, \quad \frac{\phi^\dagger p^\dagger}{1-(1-\phi^\dagger)\lambda^\dagger} < p < 1$$

$$\pi_\infty(\lambda) \propto \frac{\pi_\alpha\left(\text{logit}\left(1-\frac{(1-\phi^\dagger)\lambda^\dagger}{\lambda}\right)\right)}{\left(1-\frac{(1-\phi^\dagger)\lambda^\dagger}{\lambda}\right)^2}, \quad \frac{(1-\phi^\dagger)\lambda^\dagger}{1-\phi^\dagger p^\dagger} < \lambda < 1.$$

APPENDIX D. PRIOR DISTRIBUTIONS

Prior distributions for the parameters of each model in the simulation study were, where applicable,

$$p \sim \text{U}(0,1) \qquad \lambda \sim \text{U}(0,1)$$
$$\alpha \sim \text{Cauchy}(10) \qquad \beta \sim \text{Cauchy}(2.5)$$
$$\mu_t \sim \text{N}(0,1), \quad t = 1, \ldots, T-1 \qquad \sigma \sim \text{t}_4^+(.74).$$

The prior distribution for σ represents a half t-distribution (i.e., a t-distribution restricted to the positive real line) with four degrees of freedom scaled so that the prior median of σ is 1.