Received: 1 February 2019 | Revised: 3 May 2020 | Accepted: 30 June 2020

DOI: 10.1111/biom.13334

BIOMETRIC PRACTICE

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Maximum likelihood abundance estimation from capture-recapture data when covariates are missing at random

Yang Liu[1] | **Yukun Liu**[1] | **Pengfei Li**[2] | **Lin Zhu**[3]

[1] KLATASDS - MOE, School of Statistics, East China Normal University, Shanghai, China

[2] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada

[3] Institute of Statistics and Big Data, Renmin University of China, Beijing, China

Correspondence
Yukun Liu, KLATASDS - MOE, School of Statistics, East China Normal University, Shanghai 200062, China.
Email: ykliu@sfs.ecnu.edu.cn

Funding information
the 111 project, Grant/Award Number: B14019; National Natural Science Foundation of China, Grant/Award Number: 11771144

Abstract

In capture-recapture experiments, individual covariates may be subject to missingness, especially when the number of captures is small. When the covariate information is missing at random, the inverse probability weighting method and the multiple imputation method are widely used to obtain point estimators of the abundance. These estimators are then used to construct Wald-type confidence intervals. However, such intervals may have seriously inaccurate coverage probabilities. In this paper, we propose a maximum empirical likelihood (EL) estimation approach for the abundance in the presence of missing covariates. We show that the maximum EL estimator is asymptotically normal, and that the EL ratio statistic for the abundance has a chi-square limiting distribution with one degree of freedom. Simulations indicate that the proposed estimator has a smaller mean square error than existing estimators, and the proposed EL ratio confidence interval usually has more accurate coverage probabilities than the existing Wald-type confidence intervals. We illustrate the proposed method by analyzing data collected in Hong Kong for the yellow-bellied prinia, a bird species.

KEYWORDS
abundance, capture-recapture data analysis, empirical likelihood, missing at random

1 | INTRODUCTION

The estimation of population size or abundance is a fundamental problem in many fields, such as conservation biology, demography, epidemiology, and software reliability (Tilling *et al.*, 2001; Borchers *et al.*, 2002; Barnard *et al.*, 2003; Boden and Ozonoff, 2008). The population in abundance estimation is usually assumed to be either closed or open. In a closed population, there is no birth, death, or migration and the abundance remains unchanged throughout the period of study. A population that is not closed is said to be open. To estimate the abundance, the capture-recapture technique has been widely used, since a general census is usually too expensive or impractical. In a capture-recapture experiment, individuals or animals from the population are captured; they are marked or their existing marks are noted if they have been previously marked; and then they are released. Depending on whether the capture efforts are made on separate occasions or continuously, capture-recapture experiments can be divided into two types: discrete time and continuous time. In this paper, we focus on discrete-time capture-recapture experiments, and we make a closed-population assumption so that the abundance can be regarded as a parameter to be inferred.

It is widely accepted that heterogeneity is usually present in capture-recapture experiments, and ignoring it may result in seriously biased estimates (Otis *et al.*, 1978;

Chao, 2001). To account for the heterogeneity, Huggins (1989) and Alho (1990) modeled the capture probability via a logistic regression model on covariates such as individual characteristics and environmental conditions. Since then, there have been many studies of abundance estimation from capture-recapture data when the individual covariates are completely observed; see Chao (2001), Huggins and Hwang (2011), Stoklosa *et al.* (2011), Liu *et al.* (2017), and the references therein.

In practice, covariate information is vulnerable to missingness. For example, in real-world data from the Mai Po Sanctuary in Hong Kong (Lin and Yip, 1999; Yip and Wang, 2002; Wang and Yip, 2003), the determinant covariate "gender" was missing for 74 of 132 captured birds (Wang, 2005). Another example can be found in Section 4. In this paper, we assume that the data are missing at random (MAR) (Rubin, 1976; Little and Rubin, 2014). In other words, the missingness does not depend on the missing data themselves. Under this assumption, Xi *et al.* (2009) assumed a parametric distribution on the covariates and developed an Expectation-Maximization (EM) algorithm to obtain an estimator of the population size. To weaken the distribution assumption, Lee *et al.* (2016) proposed three kinds of estimators through regression calibration, inverse probability weighting, and multiple imputation methods. The latter two are recommended because they are consistent.

The above estimation methods in the presence of missing covariates usually have three steps. First, the underlying parameters in the assumed models are estimated by solving estimating equations. Second, a Horvitz-Thompson type estimator is constructed for the abundance after the underlying parameters are replaced by the estimators obtained in the first step. Finally, Wald-type confidence intervals are constructed for interval estimation based on the asymptotic normality of the Horvitz-Thompson type estimator. Since the estimating equations are derived from the conditional likelihood instead of the full likelihood, these estimation methods often have potential efficiency loss. Our simulation results also indicate that the coverage accuracy of the Wald-type confidence interval is usually unsatisfactory when the sample size is small or moderate.

For the case where the individual covariates are all completely observed, Liu *et al.* (2017) proposed a full empirical likelihood (EL) estimation method for the abundance. The resulting maximum EL estimator and the EL ratio confidence interval were shown to outperform the traditional Horvitz-Thompson type estimators and Wald-type confidence intervals. Liu *et al.* (2018) extended the method to continuous-time capture-recapture data. However, if individual covariates are missing, this approach will produce biased estimators (see our simulation study) if we simply discard the subjects with missing covariates.

In this paper, we further extend Liu *et al.* (2017)'s full-EL abundance estimation method to the case with missing covariates under the MAR assumption. We show that the maximum EL estimator is asymptotically normal, and that the EL ratio statistic for the abundance has a chi-square limiting distribution with one degree of freedom. The resulting EL ratio confidence interval is one-step and free from variance estimation. Our simulations indicate that the proposed maximum EL estimator has a smaller mean square error than the existing Horvitz-Thompson-type estimators. Moreover, the EL ratio confidence interval usually has more accurate coverage probabilities than the Wald-type confidence intervals.

The rest of the paper is organized as follows. In Section 2, we first introduce the capture probability model of Huggins (1989) and Alho (1990) as well as the MAR mechanism. We then present the proposed estimation method and investigate its large-sample properties. Section 3 compares our method and several existing methods through simulation studies. Section 4 is devoted to a careful analysis of the yellow-bellied prinia bird data collected in Hong Kong. Section 5 concludes the paper with a short discussion.

2 | FULL LIKELIHOOD ESTIMATION

2.1 | Model and data

Let N be the abundance of a closed population and K the number of capture occasions in a discrete-time capture-recapture experiment. For a generic individual in the population and $k = 1, \dots, K$, let $D_{(k)} = 1$ if it was captured on the kth occasion and $D_{(k)} = 0$ otherwise. Then $(D_{(1)}, \dots, D_{(K)})^\top$ is the so-called capture history. Let $\mathbf{Z}$ be a p-dimensional covariate of this generic individual with its first component being 1 and cumulative distribution function being $F_{\mathbf{Z}}(\mathbf{z})$. Following Huggins (1989) and Alho (1990), we assume that the capture probability on each occasion follows the logistic regression or Huggins-Alho model:

$$\mathrm{pr}(D_{(k)} = 1 | \mathbf{Z} = \mathbf{z}) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{z})}{1 + \exp(\boldsymbol{\beta}^\top \mathbf{z})} =: \mathrm{g}(\mathbf{z}; \boldsymbol{\beta}). \quad (1)$$

Throughout this paper, pr is used to denote the probability of an event or the probability mass/density function of a discrete/continuous random variable. Denote the number of times that the generic individual is captured by $D = \sum_{k=1}^{K} D_{(k)}$. Under the Huggins-Alho model, D given $\mathbf{Z} = \mathbf{z}$ follows a binomial distribution with size K and success probability $\mathrm{g}(\mathbf{z}; \boldsymbol{\beta})$.

In practice, individual covariates are vulnerable to missingness if D is small. Let R be the missingness indicator: $R = 1$ if the covariate of a generic individual is not missing and $R = 0$ otherwise. In this paper, we adopt a MAR assumption, that is, the selection probability of covariate $\mathbf{Z}$ is conditionally independent of $\mathbf{Z}$ itself given D:

$$\mathrm{pr}(R = 1|\mathbf{Z} = \mathbf{z}, D = k) = \mathrm{pr}(R = 1|D = k). \tag{2}$$

Suppose n distinct individuals are captured at least once in the capture-recapture experiment. Let d_i and $\mathbf{z}_i$, respectively, be the number of captures and the covariate of the ith individual captured at least once. Without loss of generality, we assume that the covariates for the first m individuals are completely observed, and the covariates for the last $(n-m)$ individuals are missing. Denote by $r_i = 1$ the corresponding missingness indicator, which is 1 for $i = 1, \dots, m$ and 0 for $i = m+1, \dots, n$. We wish to perform inference on the abundance N.

2.2 | Likelihood

Let $\alpha_0 = \mathrm{pr}(D = 0)$ be the probability that a generic individual is never captured. Clearly, $\mathrm{pr}(D > 0) = 1 - \alpha_0$ and n follows the binomial distribution $\mathrm{Bi}(N, 1-\alpha_0)$, whose probability mass function is

$$\mathrm{pr}(n) = \binom{N}{n}(1-\alpha_0)^n \alpha_0^{N-n}. \tag{3}$$

Suppose that conditioning on n, the observations $\{(r_i = 1, d_i, \mathbf{z}_i) : i = 1, \dots, m\} \cup \{(r_i = 0, d_i) : i = m+1, \dots, n\}$ are independent of each other. Because the d_is satisfy $d_i > 0$, it follows that for $i = 1, \dots, m$

$$\begin{aligned}&\mathrm{pr}(R = 1, D = d_i, \mathbf{Z} = \mathbf{z}_i|D > 0)\\ &= \frac{\mathrm{pr}(R = 1|D = d_i)\mathrm{pr}(D = d_i|\mathbf{Z} = \mathbf{z}_i)\mathrm{pr}(\mathbf{Z} = \mathbf{z}_i)}{\mathrm{pr}(D > 0)},\end{aligned}$$

where we have used the MAR assumption in (2). Similarly, for $i = m+1, \dots, n$

$$\mathrm{pr}(R = 0, D = d_i|D > 0) = \frac{\mathrm{pr}(R = 0|D = d_i)\mathrm{pr}(D = d_i)}{\mathrm{pr}(D > 0)}.$$

With the above preparation, the full likelihood can be written as $L_0 \times L$, where $L_0 = \prod_{i=1}^m \mathrm{pr}(R=1|D=d_i) \cdot \prod_{i=m+1}^n \mathrm{pr}(R=0|D=d_i)$ is the likelihood contribution of $\{(r_i|d_i) : i = 1, \dots, n\}$, and $L = \mathrm{pr}(n) \times \prod_{i=1}^m \{\mathrm{pr}(D = d_i|\mathbf{Z} = \mathbf{z}_i)\mathrm{pr}(\mathbf{Z} = \mathbf{z}_i)\} \cdot \prod_{i=m+1}^n \mathrm{pr}(D = d_i) \times \{\mathrm{pr}(D > 0)\}^{-n}$. Since L_0 does not involve the main parameter of interest, N, we abandon it and proceed with the partial likelihood L.

For $k = 1, \dots, K$, let $\alpha_k = \mathrm{pr}(D = k)$ be the probability that a generic individual is caught exactly k times. Let m_k denote the number of individuals captured exactly k times whose covariates are missing. Then

$$\prod_{i=m+1}^{n} \mathrm{pr}(D = d_i) = \prod_{k=1}^{K} \alpha_k^{m_k}. \tag{4}$$

It follows from the Huggins-Alho model that

$$\begin{aligned}&\prod_{i=1}^{m} \mathrm{pr}(D = d_i|\mathbf{Z} = \mathbf{z}_i)\\ &= \prod_{i=1}^{m} \binom{K}{d_i}\{g(\mathbf{z}_i; \boldsymbol{\beta})\}^{d_i}\{1 - g(\mathbf{z}_i; \boldsymbol{\beta})\}^{K-d_i}.\end{aligned} \tag{5}$$

Substituting Equations (3) to (5) into L gives

$$\begin{aligned}L \propto \binom{N}{n}\alpha_0^{N-n} \times \prod_{i=1}^{m}\{g(\mathbf{z}_i; \boldsymbol{\beta})\}^{d_i}\{1 - g(\mathbf{z}_i; \boldsymbol{\beta})\}^{K-d_i}\\ \times \prod_{k=1}^{K} \alpha_k^{m_k} \times \prod_{i=1}^{m}\{\mathrm{pr}(\mathbf{Z} = \mathbf{z}_i)\mathrm{d}\mathbf{z}_i\}.\end{aligned}$$

Here we have introduced the $\prod_{i=1}^m \mathrm{d}\mathbf{z}_i$ term, which is independent of any parameter and has no likelihood contribution, to express the likelihood in terms of $F_{\mathbf{Z}}$.

Note that $\mathrm{pr}(\mathbf{Z} = \mathbf{z})\mathrm{d}\mathbf{z} = \mathrm{d}F_{\mathbf{Z}}(\mathbf{z})$. We propose to handle $F_{\mathbf{Z}}$ by EL (Owen, 1988, 1990), which in essence models $F_{\mathbf{Z}}$ by a discrete distribution assigning weight p_i to observed value $\mathbf{z}_i$, that is, $F_{\mathbf{Z}}(\mathbf{z}) = \sum_{i=1}^m p_i I(\mathbf{z}_i \le \mathbf{z})$. Since $F_{\mathbf{Z}}(\mathbf{z})$ is a distribution, the p_is satisfy the constraints

$$p_1 \ge 0, \dots, p_m \ge 0, \quad \text{and} \quad \sum_{i=1}^{m} p_i = 1. \tag{6}$$

Because $F_{\mathbf{Z}}(\mathbf{z})$ is the distribution of $\mathbf{Z}$, the p_is should also satisfy

$$\mathbf{0} = \int \mathbf{U}(\mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\alpha})\mathrm{d}F_{\mathbf{Z}}(\mathbf{z}) = \sum_{i=1}^{m} \mathbf{U}(\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\alpha})p_i, \tag{7}$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)^\top$ and $\mathbf{U}(\mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = (U_1(\mathbf{z}; \boldsymbol{\beta}, \alpha_1), \dots, U_K(\mathbf{z}; \boldsymbol{\beta}, \alpha_K))^\top$ with $U_k(\mathbf{z}; \boldsymbol{\beta}, \alpha_k) = \binom{K}{k}\{g(\mathbf{z}; \boldsymbol{\beta})\}^k\{1 - g(\mathbf{z}; \boldsymbol{\beta})\}^{K-k} - \alpha_k$ for $k = 1, \dots, K$. Equation (7) holds because $\alpha_k = \mathrm{pr}(D = k) = \mathbb{E}\{\mathrm{pr}(D = k|\mathbf{Z})\}$ and $\mathrm{pr}(D = k|\mathbf{Z} = \mathbf{z}) = \binom{K}{k}\{g(\mathbf{z}; \boldsymbol{\beta})\}^k\{1 - g(\mathbf{z}; \boldsymbol{\beta})\}^{K-k}$. It is worth noting that the equations in (7) are used to constrain $F_{\mathbf{Z}}$ or the p_is and to reduce their degrees of freedom. They cannot completely determine the p_is, even given the data, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$, when $\mathbf{Z}$ includes continuous measurements or when $\mathbf{Z}$ is a discrete variable taking more than $K+1$ distinct values, because there are $m-1$ unknown free parameters but many fewer ($K < m-1$) equations.

With $p_i = dF_{\mathbf{Z}}(\mathbf{z}_i)$ in place of $\text{pr}(\mathbf{Z} = \mathbf{z}_i)d\mathbf{z}_i$ in the likelihood function L and taking logarithms, we have the empirical log-likelihood

$$\log \binom{N}{n} + \sum_{k=0}^{K} m_k \log(\alpha_k) + \sum_{i=1}^{m} \log(p_i) + \sum_{i=1}^{m} [d_i \log\{g(\mathbf{z}_i; \boldsymbol{\beta})\} + (K - d_i) \log\{1 - g(\mathbf{z}_i; \boldsymbol{\beta})\}], \tag{8}$$

where $m_0 = N - n$. Keep in mind that the feasible p_is satisfy constraints (6) and (7). Under EL, it is easier to perform inference based on the profile ELs after profiling out the p_is rather than directly. The empirical log-likelihood function of $N, \boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ can be obtained by maximizing (8) with respect to the p_is under constraints (6) and (7) given $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. By the method of Lagrange multipliers, we find that the maximum of (8) is attained at

$$p_i = \frac{1}{m} \cdot \frac{1}{1 + \boldsymbol{\lambda}^\top \mathbf{U}(\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\alpha})}, \quad i = 1, 2, \ldots, m, \tag{9}$$

where the Lagrange multiplier $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)^\top$ is the solution of

$$\sum_{i=1}^{m} \frac{\mathbf{U}(\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\alpha})}{1 + \boldsymbol{\lambda}^\top \mathbf{U}(\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\alpha})} = \mathbf{0}. \tag{10}$$

We refer to Owen (1990) for a detailed derivation of this maximizer. Substituting (9) into (8) gives the profile empirical log-likelihood of $(N, \boldsymbol{\beta}, \boldsymbol{\alpha})$:

$$\ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \log \binom{N}{n} + (N - n) \log(\alpha_0) + \sum_{k=1}^{K} m_k \log(\alpha_k) - \sum_{i=1}^{m} \log\{1 + \boldsymbol{\lambda}^\top \mathbf{U}(\mathbf{z}_i; \boldsymbol{\beta}, \boldsymbol{\alpha})\} + \sum_{i=1}^{m} [d_i \log\{g(\mathbf{z}_i; \boldsymbol{\beta})\} + (K - d_i) \log\{1 - g(\mathbf{z}_i; \boldsymbol{\beta})\}], \tag{11}$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ solves Equation (10) and $\alpha_0 = 1 - \sum_{k=1}^{K} \alpha_k$.

Remark 1. We may simply discard the missing data, assume that the capture-recapture data consists of only the m completely observed data, and apply Liu *et al.* (2017)'s approach to estimate N. However, because there are actually n distinct individuals captured at least once and m is only the number of individuals with no missing values ($m < n$), Liu *et al.* (2017)'s estimator of N must underestimate N, especially when there are many missing data.

2.3 | Estimation and asymptotics

Given the profile empirical log-likelihood $\ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha})$, we propose to estimate the parameters by their maximum EL estimator:

$$(\hat{N}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = \arg\max_{(N, \boldsymbol{\beta}, \boldsymbol{\alpha})} \{\ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha})\}.$$

Accordingly, we define the EL ratio functions of $(N, \boldsymbol{\beta}, \boldsymbol{\alpha})$ and N as

$$R(N, \boldsymbol{\beta}, \boldsymbol{\alpha}) = 2\{\max_{(N, \boldsymbol{\beta}, \boldsymbol{\alpha})} \ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha}) - \ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha})\} = 2\{\ell(\hat{N}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha})\},$$

$$R'(N) = 2\{\max_{(N, \boldsymbol{\beta}, \boldsymbol{\alpha})} \ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha}) - \max_{(\boldsymbol{\beta}, \boldsymbol{\alpha})} \ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha})\} = 2\{\ell(\hat{N}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) - \ell(N, \hat{\boldsymbol{\beta}}_N, \hat{\boldsymbol{\alpha}}_N)\},$$

where $(\hat{\boldsymbol{\beta}}_N, \hat{\boldsymbol{\alpha}}_N) = \arg\max_{(\boldsymbol{\beta}, \boldsymbol{\alpha})} \{\ell(N, \boldsymbol{\beta}, \boldsymbol{\alpha})\}$ given N.

Theorem 1 given below establishes the limiting distributions of the maximum EL estimator and the two EL ratio statistics. These results explicitly or implicitly depend on a matrix $\mathbf{W}$, which mimics the information matrix under a regular parametric likelihood. To ease the exposition of $\mathbf{W}$, we define the necessary notation. Let $(N_0, \boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0^\top)$ be the true value of $(N, \boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)$, $\boldsymbol{\alpha}_0 = (\alpha_{10}, \ldots, \alpha_{K0})$, and $\alpha_{00} = 1 - \sum_{k=1}^{K} \alpha_{k0}$. Define $h_k = \text{pr}(R = 1 | D = k)$, $\lambda_{00} = \sum_{k=1}^{K} h_k \alpha_{k0}$, $\mathbf{H}_1 = (h_1, \ldots, h_K)^\top$, $\mathbf{H}_2 = \text{diag}\{(1 - h_1)/\alpha_{10}, \ldots, (1 - h_K)/\alpha_{K0}\}$, and $\pi(\mathbf{z}; \boldsymbol{\beta}) = \sum_{k=1}^{K} \binom{K}{k} \{g(\mathbf{z}; \boldsymbol{\beta})\}^k \{1 - g(\mathbf{z}; \boldsymbol{\beta})\}^{K-k} h_k$. Then $\pi(\mathbf{z}; \boldsymbol{\beta}_0)$ is the probability that an individual has been captured at least once with no missing covariate given $\mathbf{Z} = \mathbf{z}$. Denote the first and second derivatives of $\pi(\mathbf{z}; \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ by $\dot{\pi}(\mathbf{z}; \boldsymbol{\beta})$ and $\ddot{\pi}(\mathbf{z}; \boldsymbol{\beta})$, respectively. Let $\mathbb{E}$ be the expectation operator with respect to $F_{\mathbf{Z}}$, and $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^\top$ for a vector or matrix $\mathbf{A}$. In addition, we use $\mathbf{0}_{K\times 1}$, $\mathbf{1}_{K\times 1}$, and $\mathbf{I}_{K\times K}$ to denote a $K \times 1$ vector of zeros, a $K \times 1$ vector of ones, and the $K \times K$ identity matrix. We define

$$\mathbf{W} = \begin{pmatrix} -\mathrm{V}_{11} & \mathbf{0}_{1\times p} & -\mathbf{V}_{13} \\ \mathbf{0}_{p\times 1} & -\mathbf{V}_{22} + \mathbf{V}_{24}\mathbf{V}_{44}^{-1}\mathbf{V}_{42} & -\mathbf{V}_{23} + \mathbf{V}_{24}\mathbf{V}_{44}^{-1}\mathbf{V}_{43} \\ -\mathbf{V}_{31} & -\mathbf{V}_{32} + \mathbf{V}_{34}\mathbf{V}_{44}^{-1}\mathbf{V}_{42} & -\mathbf{V}_{33} + \mathbf{V}_{34}\mathbf{V}_{44}^{-1}\mathbf{V}_{43} \end{pmatrix}, \tag{12}$$

where $V_{11} = 1 - \alpha_{00}^{-1}$, $\mathbf{V}_{31} = \mathbf{V}_{13}^\top = -\alpha_{00}^{-1}\mathbf{1}_{K\times 1}$, and

$$\mathbf{V}_{22} = \mathbb{E}\left[\frac{\{\dot{\pi}(\mathbf{Z};\boldsymbol{\beta}_0)\}^{\otimes 2}}{\pi(\mathbf{Z};\boldsymbol{\beta}_0)} - \ddot{\pi}(\mathbf{Z};\boldsymbol{\beta}_0) + Kg(\mathbf{Z};\boldsymbol{\beta}_0)\{g(\mathbf{Z};\boldsymbol{\beta}_0)-1\}\pi(\mathbf{Z};\boldsymbol{\beta}_0)\mathbf{Z}^{\otimes 2}\right],$$

$$\mathbf{V}_{23} = \mathbf{V}_{32}^\top = -\mathbb{E}\left\{\frac{\dot{\pi}(\mathbf{Z};\boldsymbol{\beta}_0)}{\pi(\mathbf{Z};\boldsymbol{\beta}_0)}\right\} \times \mathbf{H}_1^\top,$$

$$\mathbf{V}_{33} = -\alpha_{00}^{-1}\mathbf{1}_{K\times 1}^{\otimes 2} - \mathbf{H}_2 + \mathbb{E}\left\{\frac{1}{\pi(\mathbf{Z};\boldsymbol{\beta}_0)}\right\} \times \mathbf{H}_1^{\otimes 2},$$

$$\mathbf{V}_{43} = \mathbf{V}_{34}^\top = \lambda_{00}\mathbf{I}_{K\times K} - \lambda_{00}\mathbb{E}\left\{\frac{\mathbf{U}(\mathbf{Z};\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)}{\pi(\mathbf{Z};\boldsymbol{\beta}_0)}\right\} \times \mathbf{H}_1^\top,$$

$$\mathbf{V}_{24} = \mathbf{V}_{42}^\top = \lambda_{00}\mathbb{E}\left\{\frac{\dot{\pi}(\mathbf{Z};\boldsymbol{\beta}_0)\mathbf{U}^\top(\mathbf{Z};\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)}{\pi(\mathbf{Z};\boldsymbol{\beta}_0)} - \frac{\partial \mathbf{U}^\top(\mathbf{Z};\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\beta}}\right\},$$

$$\mathbf{V}_{44} = \lambda_{00}^2\mathbb{E}\left\{\frac{\mathbf{U}(\mathbf{Z};\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)\mathbf{U}^\top(\mathbf{Z};\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)}{\pi(\mathbf{Z};\boldsymbol{\beta}_0)}\right\}.$$

The matrix $\mathbf{W}$ is of the same form as $\mathbf{W}_s$ in Liu *et al.* (2017); the only difference is that the matrices $\mathbf{V}_{ij}$ are different.

Theorem 1. *Suppose* $\alpha_{k0} \in (0,1)$, $k = 1,\dots,K$, $\sum_{k=1}^K \alpha_{k0} < 1$, *and* $\int\{\pi(\mathbf{z};\boldsymbol{\beta})\}^{-1}\mathrm{d}F_{\mathbf{Z}}(\mathbf{z}) < \infty$ *for* $\boldsymbol{\beta}$ *in a neighborhood of* $\boldsymbol{\beta}_0$. *If the matrix* $\mathbf{W}$ *defined in Equation (12) is nonsingular, then as* N_0 *goes to infinity, (a)* $\sqrt{N_0}\{\log(\widehat{N}/N_0), (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top, (\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^\top\}^\top \xrightarrow{d} N(0, \mathbf{W}^{-1})$, *where* $\xrightarrow{d}$ *indicates convergence in distribution; (b)* $R(N_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0) \xrightarrow{d} \chi^2_{K+p+1}$ *and* $R'(N_0) \xrightarrow{d} \chi^2_1$.

See Section 1 in the Supporting Information for a proof of Theorem 1. According to Theorem 1, the proposed EL ratio confidence interval for N is

$$\mathcal{I} = \{N : R'(N) \le \chi^2_1(1-a)\}, \tag{13}$$

where $\chi^2_1(1-a)$ is the $(1-a)$-quantile of χ^2_1. Theorem 1 implies that $\mathcal{I}$ has an asymptotically correct coverage probability when the confidence level is $1-a$. Compared with the usual Wald-type confidence interval, the proposed interval is clearly free from variance estimation. Moreover, its lower limit is never less than n, since the domain of the likelihood ratio function $R'(N)$ is $[n, \infty)$.

Remark 2. In practice, along with the point estimates $\widehat{N}$, $\widehat{\boldsymbol{\beta}}$, and $\widehat{\boldsymbol{\alpha}}$, reasonable estimates of their asymptotic variances should be provided to quantify their variabilities based on the data. According to Theorem 1, this is equivalent to the estimation of $\mathbf{W}$ or the $\mathbf{V}_{ij}$s, which necessitates reasonable estimates of λ_0, $\mathbf{H}_1$, and $\mathbf{H}_2$. We first estimate the selection probabilities h_k by the following nonparametric estimates:

$$\widehat{h}_k = \frac{\sum_{i=1}^n r_i I(d_i = k)}{\sum_{i=1}^n I(d_i = k)}, \quad k = 1,\dots,K, \tag{14}$$

which were used by Lee *et al.* (2016). Given the $\widehat{h}_k$s and the proposed estimate $\widehat{\boldsymbol{\alpha}}$, λ_{00}, $\mathbf{H}_1$, and $\mathbf{H}_2$ can be naturally estimated by $\widehat{\lambda}_0 = \sum_{k=1}^K \widehat{h}_k\widehat{\alpha}_k$, $\widehat{\mathbf{H}}_1 = (\widehat{h}_1,\dots,\widehat{h}_K)^\top$, and $\widehat{\mathbf{H}}_2 = \mathrm{diag}\{(1-\widehat{h}_1)/\widehat{\alpha}_1,\dots,(1-\widehat{h}_K)/\widehat{\alpha}_K\}$. Since the $\mathbf{V}_{ij}$s can be expressed as $\mathbb{E}\{\mathbf{J}(\mathbf{Z};\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\mathbf{H}_1,\mathbf{H}_2)\}$ for some function $\mathbf{J}$, we estimate them by $\sum_{i=1}^m \mathbf{J}(\mathbf{z}_i;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\alpha}},\widehat{\mathbf{H}}_1,\widehat{\mathbf{H}}_2)/\{\widehat{N}\pi(\mathbf{z}_i;\widehat{\boldsymbol{\beta}})\}$. Using the relationship between $\mathbf{W}$ and the $\mathbf{V}_{ij}$s, we can obtain reasonable estimates for $\mathbf{W}$ and the asymptotic variance estimates of $\widehat{N}$, $\widehat{\boldsymbol{\beta}}$, and $\widehat{\boldsymbol{\alpha}}$.

Remark 3. We have assumed that $\alpha_{k0} \in (0,1)$ for $k = 0,1,\dots,K$ or $\min_{0\le k\le K}\alpha_{k0} > 0$. It then follows that $\min_{0\le k\le K} m_k > 0$ holds with probability approaching 1 as $N_0 \to \infty$. This implies that the number of constraints in Equation (7) is exactly equal to K as N_0 becomes large. However, this is not always the case in practice, especially when K and α_{00} are large or moderate. If some m_k are zero, we suggest to redefine $\mathbf{U}$ by removing the estimating functions U_k whose corresponding m_ks equal zero and including the estimating function U_0. This does not change the maximum EL estimate of $(N, \boldsymbol{\beta}, \boldsymbol{\alpha}, \{p_i\})$ but can alleviate the computation burden. See Section 4 of the Supporting Information for more details.

Remark 4. In practice, there may be covariates that are always observed and have an impact on the selection probability even conditioning on D. The binary variable "fat index" in the prinia analysis in Section 4 is an example. Let X be a completely observed binary covariate and $\mathbf{Y}$ a general vector of covariates. Assuming $\mathbf{Y}$ is MAR, that is, $\mathrm{pr}(R = 1|\mathbf{Y} = \mathbf{y}, X = x, D = k) = \mathrm{pr}(R = 1|X = x, D = k)$, we have extended the proposed full likelihood approach to incorporate such a binary covariate: see Section 2 in the Supporting Information. The extension to a general vector-valued categorical covariate is straightforward and therefore omitted.

3 | SIMULATION

In this section, we investigate the finite-sample performance of our estimation methods (abundance estimator

$\widehat{N}$ and confidence interval $\mathcal{I}$ or their extension $\widehat{N}_e$ and $\mathcal{I}_e$ defined in Section 2 of the Supporting Information) by comparing them with the methods of Lee *et al.* (2016) and Liu *et al.* (2017). The numerical procedure for implementing our methods is discussed in Section 3 of the Supporting Information. Lee *et al.* (2016) proposed two estimators for N: the inverse probability weighting estimator and the multiple imputation estimator, which we now briefly review.

Let $\Psi(D,\mathbf{Z};\boldsymbol{\beta}) = \{D - Kg(\mathbf{Z};\boldsymbol{\beta})/\psi(\mathbf{Z};\boldsymbol{\beta})\}\mathbf{Z}$ be the score function with respect to $\boldsymbol{\beta}$ under the conditional distribution of D given $(\mathbf{Z}, D > 0)$, where $\psi(\mathbf{z};\boldsymbol{\beta}) = 1 - \{1 - g(\mathbf{z};\boldsymbol{\beta})\}^K$ denotes the probability of an individual being captured at least once given covariate $\mathbf{z}$. Then Lee *et al.* (2016)'s inverse probability weighting estimator is $\widetilde{N}_1 = \sum_{i=1}^m \{\widehat{h}_{d_i}\psi(\mathbf{z}_i;\widetilde{\boldsymbol{\beta}}_1)\}^{-1}$, where $\widetilde{\boldsymbol{\beta}}_1$ is the solution to $\sum_{i=1}^m \Psi(\mathbf{z}_i;\boldsymbol{\beta})/\widehat{h}_{d_i} = \mathbf{0}$, and $\widehat{h}_{d_i}$ is defined in (14). Lee *et al.* (2016)'s multiple imputation estimator is $\widetilde{N}_2 = \sum_{i=1}^m \{1/\psi(\mathbf{z}_i;\widetilde{\boldsymbol{\beta}}_2)\} + \sum_{j=m+1}^n \{M/\sum_{v=1}^M \psi(\mathbf{z}_{vj};\widetilde{\boldsymbol{\beta}}_2)\}$, where for each of the missing values $\mathbf{z}_j$ $(j = m+1, \dots, n)$, M imputed values $\{\mathbf{z}_{vj}, v = 1, \dots, M\}$ are generated from $\widehat{F}(\mathbf{z}|d_j) = \sum_{i=1}^m I(d_i = d_j, \mathbf{z}_i \le \mathbf{z})/\sum_{s=1}^m I(d_s = d_j)$, and $\widetilde{\boldsymbol{\beta}}_2$ is the solution to $\sum_{i=1}^m \Psi(\mathbf{z}_i;\boldsymbol{\beta}) + \sum_{j=m+1}^n \sum_{v=1}^M \Psi(\mathbf{z}_{vj};\boldsymbol{\beta})/M = \mathbf{0}$. Let $\mathcal{I}_1$ and $\mathcal{I}_2$ denote the corresponding Wald-type confidence intervals of N based on these two point estimators, that is,

$$\mathcal{I}_1 = \{N : (\widetilde{N}_1 - N)^2/\widetilde{\sigma}_1^2 \le \chi_1^2(1-a)\} \quad \text{and}$$
$$\mathcal{I}_2 = \{N : (\widetilde{N}_2 - N)^2/\widetilde{\sigma}_2^2 \le \chi_1^2(1-a)\},$$

where $\widetilde{\sigma}_1^2$ and $\widetilde{\sigma}_2^2$ are the variance estimates of $\widetilde{N}_1$ and $\widetilde{N}_2$ provided by Lee *et al.* (2016).

We may alternatively simply abandon the data with missing values and apply the maximum empirical likelihood estimation method of Liu *et al.* (2017) on the m completely observed data. This corresponds to the complete-case method in the usual missing data problem. Let $\widetilde{N}_3$ be the maximum EL abundance estimator $\widehat{N}_s$ and $\mathcal{I}_3$ the corresponding likelihood ratio confidence interval $\mathcal{I}_{1s}$ in Liu *et al.* (2017). As we have commented in Remark 1, $\widetilde{N}_3$ generally underestimates N.

For a generic abundance estimator $\breve{N}$, we use its absolute bias, $\text{Bias}(\breve{N}) = \mathbb{E}\breve{N} - N_0$, and relative mean square error, $\text{RMSE}(\breve{N}) = \mathbb{E}(\breve{N} - N_0)^2/N_0$, to evaluate its finite-sample performance. Corresponding to a generic two-sided confidence interval $[N_L, N_U]$ are two one-sided intervals, $[N_L, \infty]$ (lower limit) and $[n, N_U]$ (upper limit). We compare all the interval estimators by their coverage probabilities.

3.1 | Simulation setup

We generate data from the following four scenarios:

A. Set the total number of capture occasions $K = 2$ and individual covariate $\mathbf{Z} = (1, Y)^\top$, where Y follows the uniform distribution U(0, 3). Given $\mathbf{Z} = \mathbf{z}$, we generate D from a binomial distribution $\text{Bi}(K,\ g(\mathbf{z};\boldsymbol{\beta}_0))$ with $\boldsymbol{\beta}_0 = (-2, 1)^\top$. The selection probability defined in Equation (2) is set to $\text{pr}(R = 1|D = k) = \{1 + \exp(0.5 - 0.7k)\}^{-1}$. In this setting, both the overall capture probability and the selection probability are about 60%.
B. The same as Scenario A except $K = 16$ and $\boldsymbol{\beta}_0 = (-4.5, 1)^\top$. The overall capture probability and the selection probability become about 55% and 66%, respectively.
C. We consider the individual covariate $\mathbf{Z} = (1, X, Y)^\top$ with $X \sim \text{Bi}(1,\ 0.3)$ and $Y \sim \text{U}(0,\ 3)$. Given $\mathbf{Z} = \mathbf{z}$, we generate D from $\text{Bi}(K,\ g(\mathbf{z};\boldsymbol{\beta}_0))$ where $K = 2$ and $\boldsymbol{\beta}_0 = (-2, 1, 1)^\top$. Here the Xs are completely observed and the Ys are subject to missingness with the selection probability $\text{pr}(R = 1|X = j, D = k) = \{1 + \exp(0.5 - 0.7j - 0.7k)\}^{-1}$. In this setting, both the overall capture probability and the selection probability are about 66%.
D. The same as Scenario C except $K = 16$ and $\boldsymbol{\beta}_0 = (-5, 1, 1)^\top$. The overall capture probability and the selection probability become around 50% and 71%, respectively.

Scenarios A and B have only one covariate, and the covariate is subject to missingness. In contrast, Scenarios C and D have two covariates: One is subject to missingness, but the other is always observed. The proposed EL method ($\widehat{N}$, $\mathcal{I}$, and $R'(N)$) applies to Scenarios A and B, and its extension ($\widehat{N}_e$, $\mathcal{I}_e$, and $R_e'(N)$) applies to Scenarios C and D. Scenarios B and D are designed to have $K = 16$ captures, which mimics that of the prinia analysis in Section 4. In each of the four scenarios, we set the population size to $N_0 = 200$ and 400. All the results reported are obtained based on 5000 simulation replications.

3.2 | Simulation results

Comparison of point estimation. For the four abundance estimators being considered, Table 1 tabulates the biases (Bias), relative mean square errors (RMSE), simulated standard deviations (SD), and average standard deviation estimates or standard errors (ASDE) over 5000 simulation replications. We see that our estimator $\widehat{N}$ or $\widehat{N}_e$ uniformly outperforms Lee *et al.* (2016)'s estimators $\widetilde{N}_1$ and

TABLE 1 Finite-sample performance of the proposed estimator $\widehat{N}$ (or $\widehat{N}_e$) and three existing estimators $\widetilde{N}_1$, $\widetilde{N}_2$, and $\widetilde{N}_3$. Bias: absolute bias; RMSE: relative mean square error; SD: standard deviation; ASDE: average of standard deviation estimates

		Scenario A				Scenario B			
N_0		$\widehat{N}$	$\widetilde{N}_1$	$\widetilde{N}_2$	$\widetilde{N}_3$	$\widehat{N}$	$\widetilde{N}_1$	$\widetilde{N}_2$	$\widetilde{N}_3$
200	Bias	18.44	27.55	28.15	−86.39	6.70	12.56	11.74	−84.77
	RMSE	38.79	48.12	47.45	45.05	9.95	12.52	12.19	38.43
	SD	86.13	94.16	93.27	39.32	44.10	48.44	47.96	22.39
	ASDE	62.71	69.74	67.03	30.56	41.22	44.12	42.60	21.69
400	Bias	16.28	24.10	24.53	−181.60	5.95	11.19	10.78	−172.41
	RMSE	15.83	18.56	19.07	85.95	8.20	9.29	9.23	76.43
	SD	77.90	82.73	83.83	37.41	56.97	59.92	59.74	29.10
	ASDE	70.63	75.04	72.10	33.69	54.43	56.90	55.33	28.07
		Scenario C				Scenario D			
N_0		$\widehat{N}_e$	$\widetilde{N}_1$	$\widetilde{N}_2$	$\widetilde{N}_3$	$\widehat{N}_e$	$\widetilde{N}_1$	$\widetilde{N}_2$	$\widetilde{N}_3$
200	Bias	13.23	20.11	20.43	−75.52	9.17	15.71	12.66	−77.85
	RMSE	25.51	30.05	29.16	33.32	14.00	17.01	15.24	34.08
	SD	70.20	74.87	73.59	30.99	52.11	56.17	53.74	27.48
	ASDE	45.24	50.75	48.71	23.30	48.18	51.29	48.93	27.15
400	Bias	10.23	16.12	16.20	−158.08	7.63	13.28	11.22	−160.99
	RMSE	8.20	9.52	9.65	64.72	10.36	11.52	11.36	67.74
	SD	56.36	59.56	59.99	29.96	63.93	66.57	66.46	34.35
	ASDE	51.85	55.56	53.27	26.77	62.51	64.90	63.28	34.06

$\widetilde{N}_2$ in terms of Bias, RMSE, and SD. The ASDEs are always smaller than the SDs, but as N_0 increases from 200 to 400 the differences decrease.

When all the missing data are ignored, the complete-case estimator $\widetilde{N}_3$ is severely downward biased, especially when the true abundance N_0 is large. This is different from the usual missing-data problem, in which the complete-case estimator is still consistent although it may not be efficient under the MAR assumption. A possible explanation for this observation is that the sample without the missing data is skewed toward individuals with higher numbers of captures D, which are associated with more catchable covariates $\mathbf{Z}$. This also explains why the empirical distribution of $\mathbf{Z}$ is estimated to be "more catchable" than it should be, and hence why the abundance estimates are downward biased.

Comparison of interval estimation. Table 2 reports the two-sided and one-sided coverage probabilities of the proposed interval ($\mathcal{I}$ or $\mathcal{I}_e$) and the two Wald-type intervals ($\mathcal{I}_1$ and $\mathcal{I}_2$) when the nominal levels are 90%, 95%, and 99%. Because the corresponding estimator $\widetilde{N}_3$ is severely biased, the complete-case interval estimator $\mathcal{I}_3$ is expected to have poor coverage probabilities and is hence omitted.

Table 2 shows that our confidence interval $\mathcal{I}$ or $\mathcal{I}_e$ always has the best performance among the three intervals in terms of both one-sided and two-sided coverage accuracy. For two-sided confidence intervals, although the two Wald-type confidence intervals $\mathcal{I}_1$ and $\mathcal{I}_2$ may produce desirable coverage probabilities at the 90% level, they usually have undercoverage at the 95% and 99% levels. For example, the undercoverage is as large as around 3% for Scenario A and $N_0 = 200$. For interval estimation, the lower limits of $\mathcal{I}_1$ and $\mathcal{I}_2$ often produce undercoverage, while their upper limits often produce overcoverage. When the population size increases from 200 to 400, the one-sided and two-sided coverage probabilities of all three interval estimators become more accurate. This is probably because more individuals are observed, and thus the distributions of the EL ratio statistic and the pivotal statistics are closer to their limiting distributions.

Comparison of QQ plots. Figure 1 shows QQ plots of the proposed likelihood ratio statistic $R'(N_0)$ or $R'_e(N_0)$ versus its limiting χ^2_1 distribution and the two pivotal statistics $(\widetilde{N}_1 - N_0)/\widetilde{\sigma}_1$ and $(\widetilde{N}_2 - N_0)/\widetilde{\sigma}_2$ versus their limiting distribution $N(0,1)$ in Scenarios (A) and (B) with $N_0 = 200$. The plots for the other two scenarios are similar and are omitted. The findings of our interval comparison can be well explained by these QQ plots.

We can clearly see that the finite-sample distributions of $R'(N_0)$ or $R'_e(N_0)$ are always much closer to the limiting χ^2_1 distribution than those of $(\widetilde{N}_1 - N_0)/\widetilde{\sigma}_1$ and $(\widetilde{N}_2 - N_0)/\widetilde{\sigma}_2$ are to their limiting distribution $N(0,1)$. This explains why the EL ratio confidence interval $\mathcal{I}$ or $\mathcal{I}_e$ has more accurate coverage probabilities than the two Wald-type confidence intervals $\mathcal{I}_1$ and $\mathcal{I}_2$. The coverage probabilities of $\mathcal{I}_1$ and $\mathcal{I}_2$ are very close to each other in all the sce-

TABLE 2 Simulated coverage probabilities of the proposed confidence interval ($\mathcal{I}$ or $\mathcal{I}_e$) and the two Wald-type confidence intervals ($\mathcal{I}_1$ and $\mathcal{I}_2$)

			Scenario A			Scenario B			Scenario C			Scenario D		
Type	**Level**	N_0	$\mathcal{I}$	$\mathcal{I}_1$	$\mathcal{I}_2$	$\mathcal{I}$	$\mathcal{I}_1$	$\mathcal{I}_2$	$\mathcal{I}_e$	$\mathcal{I}_1$	$\mathcal{I}_2$	$\mathcal{I}_e$	$\mathcal{I}_1$	$\mathcal{I}_2$
Two sided	90%	200	90	90	89	90	90	89	90	91	91	90	91	89
		400	90	91	91	90	91	90	90	90	91	90	90	89
	95%	200	95	93	92	95	93	92	95	94	93	96	94	93
		400	95	94	93	94	94	93	95	94	94	95	94	93
	99%	200	99	96	96	99	96	96	99	97	96	99	97	96
		400	99	97	96	99	97	97	99	98	97	99	97	97
Lower limit	90%	200	90	86	85	89	86	84	89	87	86	89	86	85
		400	90	87	86	89	87	86	89	88	86	89	87	86
	95%	200	95	90	89	94	90	89	95	91	91	95	91	90
		400	95	91	91	94	92	91	95	91	91	94	91	90
	99%	200	99	95	94	99	95	95	99	96	95	99	96	95
		400	99	96	96	99	96	96	99	97	96	99	97	96
Upper limit	90%	200	91	100	100	92	98	97	91	100	99	91	98	98
		400	90	99	98	91	95	94	90	98	96	91	96	96
	95%	200	96	100	100	96	100	100	95	100	100	96	100	100
		400	95	100	100	95	99	99	95	100	100	96	99	99
	99%	200	99	100	100	99	100	100	99	100	100	99	100	100
		400	99	100	100	99	100	100	99	100	100	99	100	100

narios because of the near-coincidence of the distributions of their corresponding pivotal statistics. In addition, the quantiles of these two pivotal statistics are generally smaller than the standard normal quantiles. This explains why the lower limits of the two Wald-type confidence intervals have severe undercoverage while their upper limits have severe overcoverage.

In summary, if we discard the missing data directly, the complete-case estimator $\widetilde{N}_3$ is unacceptably biased. However, our method corrects its bias and produces desirable point and interval estimators for the abundance. Our estimator $\widehat{N}$ or $\widehat{N}_e$ usually has a lower bias, RMSE, and SD than Lee *et al.* (2016)'s estimators $\widetilde{N}_1$ and $\widetilde{N}_2$. Our interval estimator always has better performance than the Wald-type interval estimators $\mathcal{I}_1$ and $\mathcal{I}_2$ in terms of one-sided and two-sided coverage accuracy.

4 | APPLICATION TO PRINIA DATA

In this section, we apply our estimation method to analyze data collected in Hong Kong for the yellow-bellied prinia, a bird species. There are 165 distinct birds captured at least once during the 17 weeks from January to April 1993. We consider three covariates: fat index X_1, wing length indicator X_2, and tail length Y. In the original dataset, the fat index ranges from 1 to 4 and two records are missing. Following Lee *et al.* (2016), we delete these two records, and we regard level 1 as not fat ($X_1 = 0$) and the other levels as fat ($X_1 = 1$). The wing lengths range from 43 to 49 mm. According to Zhao (2001), the wing lengths of yellow-bellied prinia generally range from 37 to 45.5 mm. Let X_2 indicate whether a wing length is too long: we set it to 1 if the wing length is above 45.5 mm, and 0 otherwise. The variables X_1 and X_2 are completely observed. The tail length Y has continuous values, and 25% of its values are missing.

We assume that Y is MAR, so $\text{pr}(R = 1|X_1 = x_1, X_2 = x_2, Y = y, D = k) = \text{pr}(R = 1|X_1 = x_1, X_2 = x_2, D = k)$. In our EL approach, the specific expression of $\text{pr}(R = 1|X_1 = x_1, X_2 = x_2, D = k)$ does not matter, but the covariates X_1 and X_2 must take discrete values. This is why we discretize the continuous wing length to obtain the categorical covariate X_2. Suppose that the capture probability of a yellow-bellied prinia follows the Huggins-Alho model in (1). The seven nonempty subsets of $\{X_1, X_2, Y\}$ correspond to seven choices of $\mathbf{Z}$, and these lead to seven Huggins-Alho models; see Table 3. When Y does not appear in $\mathbf{Z}$, there are no missing data and we apply only the complete-case-based method (CC) of Liu *et al.* (2017) to analyze the data. Otherwise, we analyze the data by the extension of our EL method (EL), the inverse probability weighting method (IPW), the multiple imputation method (MI), and the CC method. Table 3 tabulates the point estimates of $\boldsymbol{\beta}$ and N, their standard errors (SE), and 95% confidence intervals for N based

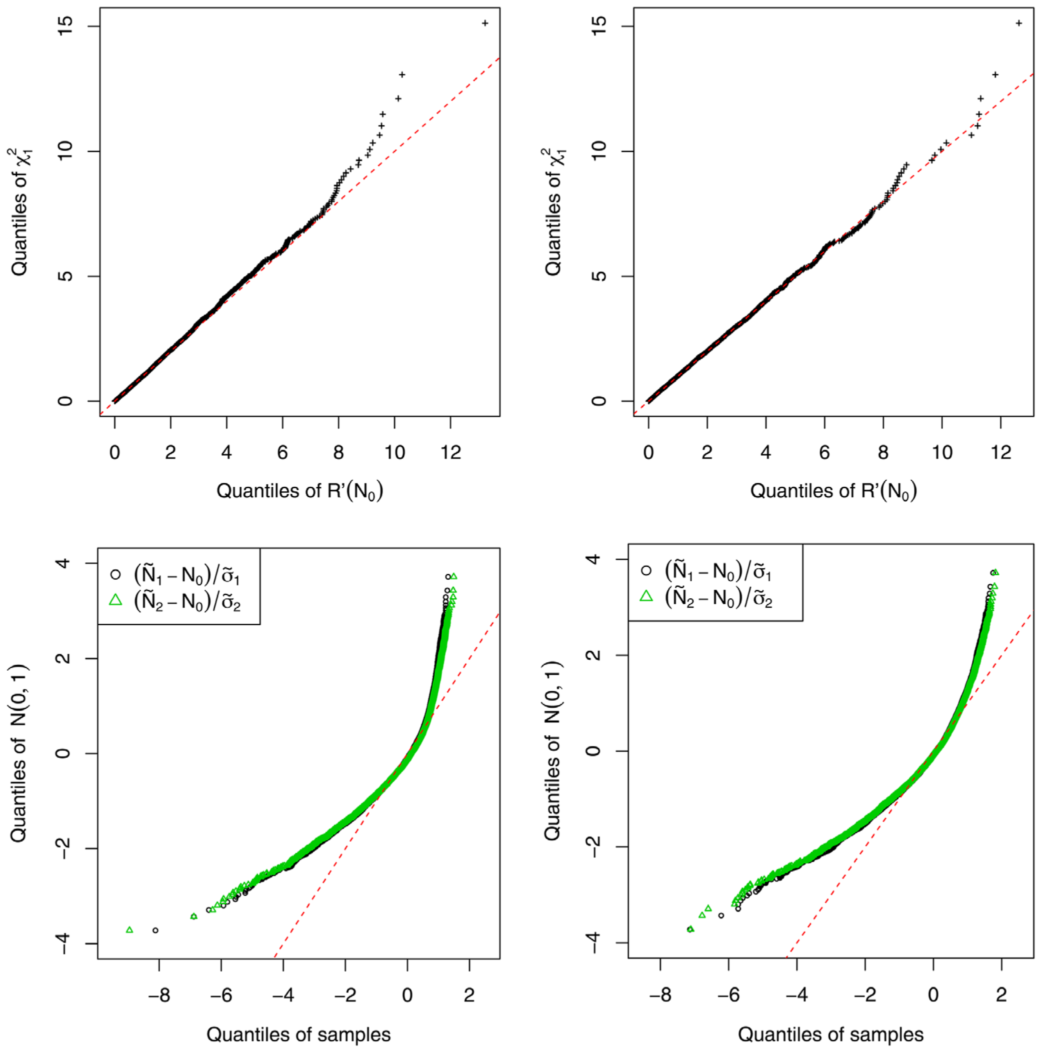

FIGURE 1 QQ plots of the proposed likelihood ratio $R'(N_0)$ versus χ_1^2 (Row 1) and the Studentized statistics based on Lee *et al.* (2016)'s two abundance estimators versus $N(0,1)$ (Row 2) in Scenarios A (Column 1) and B (Column 2) with $N_0 = 200$. Circle: the IPW estimate $\widetilde{N}_1$; Triangle: the MI estimate $\widetilde{N}_2$; Dashed line: the identity line. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

on these methods. Specifically, the point and interval estimators corresponding to EL, IPW, MI, and CC are $\{\widehat{N}, \mathcal{I}\}$, $\{\widetilde{N}_1, \mathcal{I}_1\}$, $\{\widetilde{N}_2, \mathcal{I}_2\}$, and $\{\widetilde{N}_3, \mathcal{I}_3\}$, respectively.

With different covariates, the estimation results vary considerably. Under Models 4-7, which include Y as a covariate, EL, IPW, and MI produce very similar point estimates for N, which are much larger than the CC estimates. According to our simulations, CC usually gives biased estimates when the missing data are ignored. To compare these models, we calculate their EL-based Akaike information criterion (AIC) values, and report the difference ΔAIC between each AIC value and the smallest AIC value in the final column of Table 3. The smaller the AIC, the better the model. We see that models with more covariates have smaller AIC values. In particular, adding Y to each of Models 1-3 always leads to a decreased AIC, and Model 7, the largest model, is the best.

Based on Model 7, we test whether each of the coefficients of X_1, X_2, and Y is zero via the proposed EL ratio test calibrated by the χ_1^2 limiting distribution. The P-values of these tests are 0.35%, 5.1%, and 3.9%, respectively. Roughly speaking, in addition to the wing length indicator X_2, both the fat index X_1 and the tail length Y are important for explaining the heterogeneity among the capture probabilities, and Model 7 is the most appropriate Huggins-Alho model. Under this model, we conclude that there are around 740 (with standard error 217) yellow-bellied prinia in total, with a 95% confidence interval of [452, 1652].

Previously, Hwang and Huang (2003), Yip *et al.* (2005), and Xi *et al.* (2009) analyzed this dataset under different model assumptions. They took the exact measurement of wing length as the only covariate. Their abundance estimates are 529 (107), 578 (153), and 542 (105). To understand the differences between their estimates and our estimate

TABLE 3 Results for the prinia data

Model	$Z^\top$	Method	N	SE(N)	CI	β	SE(β)	ΔAIC
1	$(1, X_1)$	CC	520	105	[369, 881]	(−4.26, 1.04)	(0.33, 0.36)	614.14
2	$(1, X_2)$	CC	502	94	[367, 763]	(−4.02, 0.97)	(0.26, 0.31)	613.33
3	$(1, X_1, X_2)$	CC	637	178	[420, 1177]	(−4.80, 1.09, 1.01)	(0.44, 0.38, 0.32)	605.61
4	$(1, Y)$	EL	606	156	[393, 1327]	(−10.03, 0.08)	(1.98, 0.03)	7.66
		IPW	602	233	[145, 1060]	(−10.04, 0.08)	(3.31, 0.04)	
		MI	598	231	[145, 1051]	(−9.96, 0.08)	(3.29, 0.04)	
		CC	362	75	[244, 729]	(−9.62, 0.08)	(1.36, 0.02)	
5	$(1, X_1, Y)$	EL	751	249	[443, 1890]	(−10.41, 1.00, 0.08)	(2.03, 0.39, 0.03)	1.80
		IPW	770	346	[91, 1449]	(−10.58, 1.01, 0.08)	(3.05, 0.42, 0.04)	
		MI	764	337	[103, 1425]	(−10.51, 1.01, 0.08)	(3.03, 0.41, 0.04)	
		CC	406	104	[258, 916]	(−10.04, 0.71, 0.08)	(1.45, 0.39, 0.02)	
6	$(1, X_2, Y)$	EL	600	140	[400, 1185]	(−8.64, 0.62, 0.06)	(2.09, 0.36, 0.03)	6.51
		IPW	607	203	[208, 1006]	(−8.74, 0.63, 0.06)	(3.18, 0.34, 0.04)	
		MI	604	202	[208, 1001]	(−8.67, 0.63, 0.06)	(3.16, 0.32, 0.04)	
		CC	372	79	[252, 690]	(−8.10, 0.72, 0.06)	(1.55, 0.42, 0.02)	
7	$(1, X_1, X_2, Y)$	EL	740	217	[452, 1652]	(−8.83, 1.03, 0.68, 0.05)	(2.10, 0.38, 0.36, 0.03)	0
		IPW	760	286	[199, 1321]	(−9.08, 1.04, 0.68, 0.06)	(2.95, 0.41, 0.33, 0.04)	
		MI	757	278	[213, 1302]	(−9.01, 1.04, 0.68, 0.06)	(2.92, 0.39, 0.31, 0.04)	
		CC	416	112	[269, 841]	(−8.32, 0.78, 0.79, 0.05)	(1.65, 0.41, 0.44, 0.02)	

502 (94) under Model 2, we further analyze this dataset by using the exact measurement of wing length as the only covariate. The CC method of Liu *et al.* (2017) gives the estimate 510 (108), which is quite close to our estimate under Model 2. This may suggest that the difference between our estimate and the estimates of Hwang and Huang (2003), Yip *et al.* (2005), and Xi *et al.* (2009) are mainly because they consider the measurement error in wing length in their data analyses, but our method did not consider that. As Xi *et al.* (2009) indicated, if some major covariates affecting the capture probability are not included in the model, the final results might not be reliable. Our evidence suggests that X_1 and Y are such covariates, and therefore the abundance estimate of 740 under Model 7 is more reliable.

5 | DISCUSSION

This paper is a further development of the full-likelihood approach to abundance estimation, which was proposed by Liu *et al.* (2017) for discrete-time capture-recapture data in the case of fully observed covariates and then extended by Liu *et al.* (2018) to continuous-time capture-recapture data. However, their methods are not directly applicable to data with missing covariates. This paper extends the full-likelihood approach to discrete-time capture-recapture data with missing covariates.

Our EL method can be extended, with care, to analyze continuous-time capture-recapture data with missing covariates. In discrete-time capture-recapture data, the largest possible number of captures of an individual is the number of capture occasions, which is finite. Therefore, a finite-dimensional estimating equation **U** can be constructed in the development of our EL method. In contrast, in continuous-time capture-recapture data, the largest possible number of captures is infinite, and our EL method fails. To overcome this difficulty, one could truncate the number of captures and proceed as before. We leave this to future research.

Recently, Stoklosa *et al.* (2019) have proposed estimating-equation-based inference procedures that can incorporate both measurement errors and missing data. These methods work because the bias of the estimating equations can be removed by proper weighting. However, it is difficult to directly incorporate measurement errors in our EL method. One potential way to overcome this difficulty is to postulate a parametric model on the covariate distribution (Xi *et al.*, 2009).

Both the current paper and Liu *et al.* (2017) focus on M_h models, where the capture probability is affected only by heterogeneity between individuals (h). Another possible extension is to M_{th}, M_{bh}, and M_{tbh} models (Otis *et al.*, 1978), where the capture probability may also be influenced by either capture time (t) or behavioral response (b).

ACKNOWLEDGMENTS

The authors thank Dr. W. H. Hwang for providing the Mai Po bird data, and the editor, the associate editor, and two referees for helpful comments and suggestions that have led to significant improvements in the paper. Dr. Liu's research was supported by the National Natural Science Foundation of China (11771144, 11971300, 11871287), the State Key Program of the National Natural Science Foundation of China (71931004), the Natural Science Foundation of Shanghai (19ZR1420900,17ZR1409000), the development fund for Shanghai talents, the 111 project (B14019), and the Fundamental Research Funds for the Central Universities. Dr. Li was supported in part by the Natural Sciences and Engineering Research Council of Canada (grant number RGPIN-2020-04964).

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available from the corresponding author, Dr. W. H. Hwang, wenhan@nchu.edu.tw, of Lee *et al.* (2016).

ORCID

Yukun Liu https://orcid.org/0000-0002-9743-9276
Pengfei Li https://orcid.org/0000-0003-2165-9157

REFERENCES

Alho, J.M. (1990) Logistic regression in capture-recapture models. *Biometrics*, 43, 623–635.

Barnard, J., El Emam, K. and Zubrow, D. (2003) Using capture-recapture models for the reinspection decision. *Software Quality Professional*, 5, 11–20.

Boden, L.I. and Ozonoff, A. (2008) Capture–recapture estimates of nonfatal workplace injuries and illnesses. *Annals of Epidemiology*, 18, 500–506.

Borchers, D.L., Buckland, S.T. and Zucchini, W. (2002) *Estimating Animal Abundance: Closed Populations*. Dordrecht, the Netherlands: Springer Science & Business Media.

Chao, A. (2001) An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, 6, 158–175.

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

Huggins, R. and Hwang, W.-H. (2011) A review of the use of conditional likelihood in capture-recapture experiments. *International Statistical Review*, 79, 385–400.

Hwang, W.-H. and Huang, S.Y. (2003) Estimation in capture-recapture models when covariates are subject to measurement errors. *Biometrics*, 59, 1113–1122.

Lee, S.-M., Hwang, W.-H. and de Dieu Tapsoba, J. (2016) Estimation in closed capture–recapture models when covariates are missing at random. *Biometrics*, 72, 1294–1304.

Lin, D. and Yip, P.S. (1999) Parametric regression models for continuous time removal and recapture studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61, 401–411.

Little, R.J. and Rubin, D.B. (2014) *Statistical Analysis with Missing Data*, Vol. 333. Hoboken, NJ: John Wiley & Sons.

Liu, Y., Li, P. and Qin, J. (2017) Maximum empirical likelihood estimation for abundance in a closed population from capture-recapture data. *Biometrika*, 104, 527–543.

Liu, Y., Liu, Y., Li, P. and Qin, J. (2018) Full likelihood inference for abundance from continuous time capture–recapture data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80, 995–1014.

Otis, D.L., Burnham, K.P., White, G.C. and Anderson, D.R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, 62, 1–135.

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75, 237–249.

Owen, A. (1990) Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18, 90–120.

Rubin, D.B. (1976) Inference and missing data. *Biometrika*, 63, 581–592.

Stoklosa, J., Hwang, W.-H., Wu, S.-H. and Huggins, R. (2011) Heterogeneous capture–recapture models with covariates: A partial likelihood approach for closed populations. *Biometrics*, 67, 1659–1665.

Stoklosa, J., Lee, S.-M. and Hwang, W.-H. (2019) Closed-population capture–recapture models with measurement error and missing observations in covariates. *Statistica Sinica*, 29, 589–610.

Tilling, K., Sterne, J.A. and Wolfe, C.D. (2001) Multilevel growth curve models with covariate effects: Application to recovery after stroke. *Statistics in Medicine*, 20, 685–704.

Wang, Y. (2005) A semiparametric regression model with missing covariates in continuous-time capture-recapture studies. *Australian & New Zealand Journal of Statistics*, 47, 287–297.

Wang, Y. and Yip, P.S. (2003) A semiparametric model for recapture experiments. *Scandinavian Journal of Statistics*, 30, 667–676.

Xi, L., Watson, R., Wang, J.-P. and Yip, P.S. (2009) Estimation in capture–recapture models when covariates are subject to measurement errors and missing data. *Canadian Journal of Statistics*, 37, 645–658.

Yip, P.S., Lin, H.-Z. and Xi, L. (2005) A semiparametric method for estimating population size for capture–recapture experiments with random covariates in continuous time. *Biometrics*, 61, 1085–1092.

Yip, P.S. and Wang, Y. (2002) A unified parametric regression model for recapture studies with random removals in continuous time. *Biometrics*, 58, 192–199.

Zhao, Z.-J. (2001). *A Handbook of the Birds of China: Passerines*. Changchun, China: Jilin Science and Technology Press.

SUPPORTING INFORMATION

Web Appendices referenced in Sections 2 and 3, and the R code and a simulated data, are available with this paper at the *Biometrics* website on Wiley Online Library.

How to cite this article: Liu Y, Liu Y, Li P, Zhu L. Maximum likelihood abundance estimation from capture-recapture data when covariates are missing at random. *Biometrics*. 2021;77:1050–1060. https://doi.org/10.1111/biom.13334