The Canadian Journal of Statistics
Vol. 37, No. 4, 2009, Pages 645–658
La revue canadienne de statistique

Estimation in capture–recapture models when covariates are subject to measurement errors and missing data

Liqun XI[1]*, Ray WATSON[2], Ji-Ping WANG[1] and Paul S. F. YIP[3]*

[1]*Department of Statistics, Northwestern University, Evanston, IL 60208, USA*
[2]*Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Victoria, Australia*
[3]*Department of Social Work and Social Administration, The University of Hong Kong, Hong Kong, China*

Key words and phrases: Capture–recapture; EM algorithm; estimating function; measurement error; population size; missing covariate.

MSC 2000: Primary 62P10; secondary 62F99.

Abstract: For capture–recapture models when covariates are subject to measurement errors and missing data, a set of estimating equations is constructed to estimate population size and relevant parameters. These estimating equations can be solved by an algorithm similar to the EM algorithm. The proposed method is also applicable to the situation when covariates with no measurement errors have missing data. Simulation studies are used to assess the performance of the proposed estimator. The estimator is also applied to a capture–recapture experiment on the bird species *Prinia flaviventris* in Hong Kong. *The Canadian Journal of Statistics* 37: 645–658; 2009 © 2009 Statistical Society of Canada

Résumé: Pour les modèles de capture-recapture où les covariables sont sujettes à erreurs ou encore manquantes, un ensemble d'équations d'estimation est obtenu afin d'estimer la taille de la population et les paramètres pertinents. Ces équations d'estimation peuvent être résolues par un algorithme similaire à l'algorithme EM. Cette méthode est aussi applicable lorsqu'il y a des valeurs manquantes dans les covariables et que celles-ci sont mesurées exactement. Des études de simulations illustrent la performance de l'estimateur proposé. L'estimateur est aussi appliqué à une expérience de capture-recapture sur l'espèce aviaire *Prinia flaviventris* de Hong Kong. *La revue canadienne de statistique* 37: 645–658; 2009 © 2009 Société statistique du Canada

1. INTRODUCTION

Estimation of population size is one of the most important issues in ecology. Heterogeneity among individuals is the most difficult problem to deal with in population size estimation. Using covariates to explain heterogeneity in capture probability is one of the more promising approaches (Pollock, 2002). However, an individuals' covariates in capture–recapture may be subject to measurement error or may be missing.

Measurement error is a common problem in regression analysis and has received much attention in literature (Carroll et al., 2006). However, in capture–recapture studies, only a very limited body of literature has discussed what may happen when there are errors in measuring the covariates (Pollock, 2002). For analysis of closed capture–recapture data associated with covariates, the approach based on the conditional likelihood and the Horvitz–Thompson estimator proposed by Huggins (1989) has become a standard method. Following this approach, Hwang & Huang (2003) investigated how covariate measurement errors affect the estimation of population size for

* *Author to whom correspondence may be addressed.*
E-mail: lxiaa@graduate.hku.hk, sfpyip@hku.hk

the discrete-time capture–recapture model described in Huggins (1989). They found that ignoring the measurement error yields a biased estimator. They proposed a refined regression calibration estimator to improve estimation of the regression coefficients in the capture probabilities and then using them in an adjusted Horvitz–Thompson estimator.

For the model considered in Hwang & Huang (2003), the variances of the measurement errors are assumed known, and for each captured individual, only one observation of covariates is used. For the data relating the bird species *Prinia flaviventris*, which was used for illustration in their paper, there are replicates of measurement for the birds which were captured more than once. The information based on the replicates makes it possible to avoid the assumption of known variances of measurement errors. Also, it is found that some measurements were not recorded for each capture. In the example, for two of the captured birds, each was captured three times but the covariates were measured only twice; another bird was captured once but no measurement was recorded. The missing covariates may be due to observer oversight, or it may be that measurement could not be taken. For discrete-time capture–recapture experiments with missing covariates, even without measurement error, there is no available estimator for population size. For the method used by Hwang & Huang (2003), the bird with no measurement records has to be excluded from the analysis.

Yip, Lin & Xi (2005) considered a continuous-time capture–recapture model where measurement error and missing data are incorporated. A semiparametric method for population size estimation was proposed, and the capture time for each capture is essential for the method. However, most of the real data of capture–recapture studies exist in discrete-time form. When applying a continuous-time model to a discrete-time data, serious biases occur if the number of capture occasions is not sufficiently large; see Xi, Yip & Watson (2007). In this paper, we consider the discrete-time closed capture–recapture model proposed by Huggins (1989), but we assume that the covariates are subject to measurement error and may be missing completely at random (i.e., MCAR, see Little & Rubin, 2002) for each capture. For each captured individual, measurement is supposed to be taken for each capture (including recapture) but not always, that is, there are replicates of measurement (with errors) but missing data are allowed. The variances of measurement errors are assumed unknown. Using an approach similar to that of Yip, Lin & Xi (2005), a set of estimating equations is constructed to estimate the population size and the relevant parameters. These estimating equations can be solved using an algorithm similar to the EM algorithm. The proposed method is also applicable when covariates are subject to missing data only (with no measurement errors). The model considered by Hwang & Huang (2003) is a special case of the model proposed here. Section 2 specifies the model and the inference procedure. In Section 3, simulation studies are conducted to assess the performance of the proposed estimator, which is also compared with existing estimators. In Section 4, the proposed method is applied to the data of the bird species *P. flaviventris* in Hong Kong Mai Po Bird Sanctuary. A short discussion is given in Section 5.

2. THE ESTIMATING PROCEDURES

2.1. Model and Notations

Let $i = 1, \ldots, \nu$ index the individuals in a closed population; let $l = 1, \ldots, t$ index the capture occasions, and suppose that δ_{il} denotes the indicator function for whether or not the ith individual is captured in the lth capture occasion. Individuals are assumed to act independently. The probability of being captured is related to the covariates via a logistic function

$$P(\delta_{il} = 1|\mathbf{Z}_i) = \frac{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i^*\right)}{1 + \exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i^*\right)} \triangleq H(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i^*), \quad i = 1, \ldots, \nu, \quad l = 1, \ldots, t,$$

where $\mathbf{Z}_i$ is an $s \times 1$ vector corresponding to the covariates of the ith individual, $\mathbf{Z}_i^* = (1, \mathbf{Z}_i^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_s)^{\mathrm{T}}$ is the coefficient vector. This probability is independent of l, that is, the model considered is $\mathcal{M}_{\mathrm{h}}$ (see Otis et al., 1978). Since $\mathbf{Z}_i$ is measured with random errors, instead of observing the true $\mathbf{Z}_i$, only the surrogates $\mathbf{W}_{ij}$ are observable, where

$$\mathbf{W}_{ij} = \mathbf{Z}_i + \boldsymbol{\varepsilon}_{ij}, \quad j = 1, \ldots, m_i \quad (m_i \leqslant n_i),$$

where $\boldsymbol{\varepsilon}_{ij}$ is the measurement error vector, n_i is the total number of capture times for the ith individual among which measurement is taken m_i times. Since the model considered is $\mathcal{M}_{\mathrm{h}}$, we need not distinguish on which capture occasion the individual is captured: thus, we let j take values from 1 to m_i. When $n_i > 0$ but $m_i = 0$, this means the covariate is missing. For those individuals uncaptured during the whole experiment, $n_i = m_i = 0$.

We consider the model in a parametric framework, and assume $\boldsymbol{\varepsilon}_{ij} \overset{\text{i.i.d.}}{\sim} f_\varepsilon(\boldsymbol{\varepsilon}_{ij}; \boldsymbol{\phi})$, $\mathbf{Z}_i \sim f_z(\mathbf{Z}_i; \boldsymbol{\alpha})$. We assume the forms of f_ε and f_z are known, $\boldsymbol{\phi}$ and $\boldsymbol{\alpha}$ are corresponding unknown parameter vectors. The model considered in Hwang & Huang (2003) assumes normality of f_ε and f_z, $m_i = 1$ for all the captured individuals, and the variance of $\boldsymbol{\varepsilon}_{ij}$ is known.

2.2. Estimating the Parameters and Population Size

Let $\Delta_i = \{\delta_{il}, l = 1, \ldots, t\}$ denote the capture history for the ith individual. Let $\mathcal{W}_i = \{\mathbf{W}_{ij},\ j = 1, \ldots, m_i\}$, and when $m_i = 0$, define $\mathcal{W}_i$ as a null-set. Under the MCAR assumption, the missingness of $\mathbf{W}_{ij}$ does not depend on the actual value of missing data $\mathbf{W}_{ij}$, nor on any observed data. Therefore, the inference regarding the probability of missingness can be made solely based on $\mathcal{W}_i$ and t. The likelihood function for $\boldsymbol{\rho} = (\boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\beta})$ based on data $\mathcal{D} = \{\Delta_i, \mathcal{W}_i, i = 1, \ldots, \nu\}$ is given by

$$L(\boldsymbol{\rho}) = \prod_{i=1}^{\nu} \int_{-\infty}^{+\infty} f(\Delta_i | \mathbf{Z}_i; \boldsymbol{\beta}) f(\mathcal{W}_i | \mathbf{Z}_i; \boldsymbol{\phi}) f(\mathbf{Z}_i; \boldsymbol{\alpha})\, \mathrm{d}\mathbf{Z}_i \tag{1}$$

where

$$f(\Delta_i | \mathbf{Z}_i; \boldsymbol{\beta}) = \prod_{l=1}^{t} H(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i^*)^{\delta_{il}} \{1 - H(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i^*)\}^{1-\delta_{il}},$$

$$f(\mathcal{W}_i | \mathbf{Z}_i; \boldsymbol{\phi}) = \prod_{j=1}^{m_i} f_\varepsilon(\mathbf{W}_{ij} - \mathbf{Z}_i; \boldsymbol{\phi}) \text{ and } f(\mathbf{Z}_i; \boldsymbol{\alpha}) = f_z(\mathbf{Z}_i; \boldsymbol{\alpha}).$$

For $m_i = 0$, let $f(\mathcal{W}_i | \mathbf{Z}_i; \boldsymbol{\phi}) = 1$.

The likelihood (1) is not a closed form and it has many of the computational difficulties found in mixed models (McCulloch & Searle, 2001). The EM algorithm provides a useful tool to solve such problems; see Dempster, Laird & Rubin (1977). If the population size ν is known, the EM algorithm can be used to estimate $(\boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\beta})$. However, in a capture–recapture experiment ν is also unknown and is of interest. Thus, the EM algorithm cannot be applied directly. Following the idea in Yip, Lin & Xi (2005), in each iteration step of the EM algorithm, we insert a Horvitz–Thompson type estimator for updating ν. Thus, we construct a set of estimating equations for parameters $(\boldsymbol{\rho}, \nu)$, and an algorithm similar to the EM algorithm is used.

For the complete data of the ith individual, $(\Delta_i, \mathcal{W}_i, \mathbf{Z}_i)$, the corresponding likelihood is the integrand of (1), denoted by $L_i(\boldsymbol{\rho})$. In the maximization step of the EM algorithm (if ν is known), the conditional expectation of $\sum_{i=1}^{\nu} \log L_i(\boldsymbol{\rho})$ (given the observed data) is maximized. All of the parameter estimates in the maximization step depend on the conditional expectation of some function of $\mathbf{Z}_i$, namely $E\{h(\mathbf{Z}_i)|\Delta_i, \mathcal{W}_i; \boldsymbol{\rho}\}$. We denote it by $E_i\{h(\mathbf{Z}_i)\}$. $h(\mathbf{Z}_i)$ can be derived from the given forms of the density functions. The maximum of the conditional expectation of $\sum_{i=1}^{\nu} \log L_i(\boldsymbol{\rho})$ with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\phi}$, $\boldsymbol{\beta}$ is obtained at

$$\sum_{i=1}^{\nu} E_i \left\{ \frac{\partial \log f_z(\mathbf{Z}_i; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right\} = 0, \tag{2}$$

$$\sum_{i=1}^{\nu} E_i \left\{ \sum_{j=1}^{m_i} \frac{\partial \log f_\varepsilon \left(\mathbf{W}_{ij} - \mathbf{Z}_i; \boldsymbol{\phi}\right)}{\partial \boldsymbol{\phi}} \right\} = 0, \tag{3}$$

$$\sum_{i=1}^{\nu} E_i \left\{ \frac{\partial \log f(\Delta_i|\mathbf{Z}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right\} = 0, \text{ that is, } \sum_{i=1}^{\nu} E_i[\{n_i - tH(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i^*)\}\mathbf{Z}_i^*] = 0. \tag{4}$$

In the case of $m_i = 0$, $\sum_{j=1}^{m_i}$ in (3) equals zero since $f(\mathcal{W}_i|\mathbf{Z}_i; \boldsymbol{\phi}) = 1$. We denote the left hand sides of (2), (3), (4) as $\sum_{i=1}^{\nu} \mathbf{u}_{\alpha i}(\boldsymbol{\rho})$, $\sum_{i=1}^{\nu} \mathbf{u}_{\phi i}(\boldsymbol{\rho})$, $\sum_{i=1}^{\nu} \mathbf{u}_{\beta i}(\boldsymbol{\rho})$, respectively. The means of these summations are zero because they are score functions. For the individuals which have never been captured during the whole experiment, their $L_i(\boldsymbol{\rho})$ are the same. Therefore, the conditional expectation of $\sum_{i=1}^{\nu} \log L_i(\boldsymbol{\rho})$ is a linear function of ν, thus, a score function for ν cannot be obtained by maximizing the conditional expectation. Based on the idea of the Horvitz–Thompson estimator (Horvitz & Thompson, 1952), we construct an estimating equation for ν as follows:

$$\sum_{i=1}^{\nu} \left\{ E_i \left(\frac{Y_i}{p_i^*} \right) - 1 \right\} = 0, \tag{5}$$

where Y_i is the indicator function for whether or not the ith individual is captured during the whole experiment, and

$$p_i^* = P\left(\sum_{l=1}^{t} \delta_{il} > 0 \,\middle|\, \mathbf{Z}_i; \boldsymbol{\beta} \right) = 1 - \{1 - H(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i^*)\}^t,$$

is the probability of being captured at least once for the ith individual. Denote the left hand side of (5) by $\sum_{i=1}^{\nu} u_{\nu i}(\boldsymbol{\rho})$. It can be shown that the mean of $\sum_{i=1}^{\nu} u_{\nu i}(\boldsymbol{\rho})$ is zero. Combining Equations (2), (3), (4), and (5), yields a set of estimating equations for $\boldsymbol{\theta} = (\boldsymbol{\rho}, \nu)$:

$$\mathbf{u}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{u}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) \\ \mathbf{u}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \\ \mathbf{u}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ u_\nu(\boldsymbol{\theta}) \end{bmatrix} = \sum_{i=1}^{\nu} \begin{bmatrix} \mathbf{u}_{\boldsymbol{\alpha} i}(\boldsymbol{\rho}) \\ \mathbf{u}_{\boldsymbol{\phi} i}(\boldsymbol{\rho}) \\ \mathbf{u}_{\boldsymbol{\beta} i}(\boldsymbol{\rho}) \\ u_{\nu i}(\boldsymbol{\rho}) \end{bmatrix} = \sum_{i=1}^{\nu} \mathbf{u}_i(\boldsymbol{\rho}) = \mathbf{0}.$$

Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\rho}}, \hat{\nu})$ be the solution of $\mathbf{u}(\boldsymbol{\theta}) = \mathbf{0}$. Since $E\{\mathbf{u}(\boldsymbol{\theta})\} = \mathbf{0}$ and $\mathbf{u}(\boldsymbol{\theta})$ is a summation of i.i.d. terms, it is believed, with some mild conditions on the distributions of $\mathbf{Z}_i$ and $\boldsymbol{\varepsilon}_{ij}$, $\nu^{1/2}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$ and $\nu^{-1/2}(\hat{\nu} - \nu)$ are asymptotically normal with zero mean. The consistency for $\hat{\nu}$ is $\hat{\nu}/\nu \xrightarrow{P} 1$, rather than the common $\hat{\nu} \xrightarrow{P} \nu$. With the normal assumptions for $\mathbf{Z}_i$ and $\boldsymbol{\varepsilon}_{ij}$, the simulation results in Section 3 suggest the asymptotic normality and consistency.

To solve the equations, we need to compute $E_i\{h(\mathbf{Z}_i)\}$: the conditional expectation of all functions needed in solving the equations. The conditional density of $\mathbf{Z}_i$ given the observed data and the current parameter estimates is given by

$$f(\mathbf{Z}_i|\mathcal{D};\boldsymbol{\rho}) = \frac{f(\Delta_i|\mathbf{Z}_i;\boldsymbol{\beta})f(\mathcal{W}_i|\mathbf{Z}_i;\boldsymbol{\phi})\,f(\mathbf{Z}_i;\boldsymbol{\alpha})}{\int_{-\infty}^{+\infty} f(\Delta_i|\mathbf{Z}_i;\boldsymbol{\beta})f(\mathcal{W}_i|\mathbf{Z}_i;\boldsymbol{\phi})\,f(\mathbf{Z}_i;\boldsymbol{\alpha})\,\mathrm{d}\mathbf{Z}_i}.$$

So

$$E_i\{h(\mathbf{Z}_i)\} = \int_{-\infty}^{+\infty} h(\mathbf{Z}_i)f(\mathbf{Z}_i|\mathcal{D};\boldsymbol{\rho})\,\mathrm{d}\mathbf{Z}_i.$$

An algorithm similar to the EM algorithm is outlined as follows. Let $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\rho}^{(k)}, \nu^{(k)})$ denote the value of the estimate of $\boldsymbol{\theta}$ after the kth iteration.

Step 1: Assume a starting value $\boldsymbol{\theta}^{(0)}$.
Step 2: Use $\boldsymbol{\rho}^{(k)}$ to estimate $f(\mathbf{Z}_i|\mathcal{D};\boldsymbol{\rho}^{(k)})$ and compute $E_i\{h(\mathbf{Z}_i)\}$.
Step 3: Solve $\mathbf{u}(\boldsymbol{\theta}^{(k+1)}|\boldsymbol{\theta}^{(k)}) = 0$ for $\boldsymbol{\theta}^{(k+1)}$.
Step 4: Return to Step 2 unless $\boldsymbol{\theta}^{(k)}$ and $\boldsymbol{\theta}^{(k+1)}$ differ insignificantly.

The above algorithm is the same as the EM algorithm except that the population size ν is updated by a Horvitz–Thompson type estimator in each iteration. The convergence of the proposed algorithm is checked using a simulation study in Section 3.

The approximate variance–covariance matrix of $\hat{\boldsymbol{\theta}}$ can be obtained by the sandwich method (Diggle, Liang & Zeger, 1994): estimated by $\mathbf{A}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{B}(\hat{\boldsymbol{\theta}})\mathbf{A}(\hat{\boldsymbol{\theta}})^{-\mathrm{T}}$ where

$$\mathbf{B}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{\hat{\nu}}\{\mathbf{u}_i(\hat{\boldsymbol{\rho}})\mathbf{u}_i(\hat{\boldsymbol{\rho}})^{\mathrm{T}}\} \text{ and } \mathbf{A}(\hat{\boldsymbol{\theta}}) = -\left(\frac{\partial \mathbf{u}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\hat{\boldsymbol{\theta}}}.$$

$\mathbf{A}^{-\mathrm{T}}$ denotes the transpose of the inverse of the matrix $\mathbf{A}$. For the calculation of the partial derivatives, for those with respect to ν, it is noted that the individuals which have never been captured during the whole experiment have the same $\mathbf{u}_i(\boldsymbol{\rho})$ (denoted by $\mathbf{u}_0(\boldsymbol{\rho})$). Therefore,

$$\mathbf{u}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\mathbf{u}_i(\boldsymbol{\rho}) + (\nu - n)\mathbf{u}_0(\boldsymbol{\rho}),$$

where n is the number of distinct captured individuals. The partial derivatives with respect to ν are the vector $\mathbf{u}_0(\boldsymbol{\rho})$, and the others can be obtained numerically. The variance of $\hat{\nu}$ can also be estimated by an easily implemented method: calculating numerically the first derivative of the conditional estimating equations with respect to ν and inverting it (see Herring & Ibrahim, 2001 for detailed discussion). Simulation studies show the two methods lead to very close results for the variance estimate of $\hat{\nu}$, but the sandwich method gives the estimate of the variance–covariance matrix of $\hat{\boldsymbol{\rho}}$. In the simulation studies and for the example given in Sections 3 and 4, we adopt the sandwich method to estimate the variance of $\hat{\nu}$.

2.3. Some Extensions

When there is a behavioral response to capture, that is, $p_{il} = H(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i^* + \gamma X_{il})$, where X_{il} is an indicator function for whether or not the ith individual has been captured before the lth occasion, the proposed inference procedure is still applicable. Since X_{il} is observable, we only need to add a score function for γ. In the same way, the proposed inference procedure is applicable for the model $\mathcal{M}_{\text{ht}}: p_{il} = H(\boldsymbol{\beta}^{\text{T}}\mathbf{Z}_i^* + \boldsymbol{\eta}^{\text{T}}\mathbf{Q}_l)$, where $\mathbf{Q}_l$ is the covariate vector for the lth capture occasion and is observable without measurement errors or missing data.

In the case when there are missing covariates only (with no measurement errors), the proposed inference procedure is also applicable. The inference procedure is similar except that Equation (3) is ignored, and $E_i\{h(\mathbf{Z}_i)\} = h(\mathbf{Z}_i)$ for the captured individuals with recorded covariates $\mathbf{Z}_i$. For those captured but without recorded covariates (i.e., missing) and those uncaptured during the whole experiment, $f(\mathbf{Z}_i|\mathcal{D};\boldsymbol{\rho})$ which is used for the calculation of $E_i\{h(\mathbf{Z}_i)\}$ is as follows:

$$f(\mathbf{Z}_i|\mathcal{D};\boldsymbol{\rho}) = \frac{f(\Delta_i|\mathbf{Z}_i;\boldsymbol{\beta})f(\mathbf{Z}_i;\boldsymbol{\alpha})}{\int_{-\infty}^{+\infty} f(\Delta_i|\mathbf{Z}_i;\boldsymbol{\beta})f(\mathbf{Z}_i;\boldsymbol{\alpha})\,\mathrm{d}\mathbf{Z}_i},$$

where $\boldsymbol{\rho}$ does not include $\boldsymbol{\phi}$.

If measurement errors are not additive elements to the covariates, for example, the observation error for a gender, it is expected a similar method can be developed in the parametric framework. Furthermore, when individual covariates vary from occasion to occasion, there is no estimator available for ν. To deal with such a problem, some assumptions about variation of the covariates would be needed; see Bonner & Schwarz (2006) and King, Brooks & Coulson (2008).

2.4. Estimating With Normal Assumptions

We choose the normal assumptions for illustration, that is, assume $\mathbf{Z}_i \sim \mathbf{N}_s(\boldsymbol{\mu}, \mathbf{V})$ and $\boldsymbol{\varepsilon}_{ij} \sim \mathbf{N}_s(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a diagonal matrix. So $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \nu)$ and

$$f_z(\mathbf{Z}_i;\boldsymbol{\mu},\mathbf{V}) = \frac{|\mathbf{V}|^{-1/2}}{(\sqrt{2\pi})^s}\exp\left\{-\frac{1}{2}(\mathbf{Z}_i-\boldsymbol{\mu})^{\text{T}}\mathbf{V}^{-1}(\mathbf{Z}_i-\boldsymbol{\mu})\right\},$$

$$f_\varepsilon(\mathbf{W}_{ij}-\mathbf{Z}_i;\boldsymbol{\Sigma}) = \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(\sqrt{2\pi})^s}\exp\left\{-\frac{1}{2}(\mathbf{W}_{ij}-\mathbf{Z}_i)^{\text{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{W}_{ij}-\mathbf{Z}_i)\right\}.$$

The estimating equations for $(\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\Sigma})$ are obtained as follows:

$$\begin{aligned}
\mathbf{u}_{\boldsymbol{\mu}}(\boldsymbol{\theta}) &= \sum_{i=1}^{\nu}\{E_i(\mathbf{Z}_i)-\boldsymbol{\mu}\} = \mathbf{0},\\
\mathbf{u}_{\mathbf{V}}(\boldsymbol{\theta}) &= \sum_{i=1}^{\nu}\{E_i(\mathbf{Z}_i-\boldsymbol{\mu})(\mathbf{Z}_i-\mu)^{\text{T}}-\mathbf{V}\} = \mathbf{0},\\
\mathbf{u}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta}) &= \sum_{i=1}^{\nu}\sum_{j=1}^{m_i}\{E_i(\mathbf{W}_{ij}-\mathbf{Z}_i)(\mathbf{W}_{ij}-\mathbf{Z}_i)^{\text{T}}-\boldsymbol{\Sigma}\} = \mathbf{0}.
\end{aligned} \tag{6}$$

From the above equations, there are closed-forms for updating $(\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\Sigma})$ in each iteration step. For updating $\boldsymbol{\beta}$ at each iteration step, a one-step Newton–Raphson algorithm (Wulfsohn & Tsiatis, 1997) can be applied, that is,

$$\boldsymbol{\beta}^{(k+1)} = \beta^{(k)} - \{\mathbf{J}_{\boldsymbol{\beta}}(\boldsymbol{\theta}^{(k)})\}^{-1}\mathbf{u}_{\boldsymbol{\beta}}(\boldsymbol{\theta}^{(k)})$$

where $\mathbf{J}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ is the derivative matrix of $\mathbf{u}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$, having the following form:

$$\mathbf{J}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = -t\sum_{i=1}^{\nu}E_i\left[\mathbf{Z}_i^*\left\{\frac{\partial H(\boldsymbol{\beta}^{\text{T}}\mathbf{Z}_i^*)}{\partial\boldsymbol{\beta}}\right\}^{\text{T}}\right].$$

In applying the sandwich method to estimate the variance–covariance of $\hat{\boldsymbol{\theta}}$, the numerical method is applied to calculate the partial derivatives (except those with respect to ν). The numerical derivative of the xth estimating equation with respect to the yth parameter is obtained as follows:

$$\left(\frac{\partial u_x(\boldsymbol{\theta})}{\partial \theta_y}\right)_{\hat{\boldsymbol{\theta}}} \approx \frac{u_x(\hat{\boldsymbol{\theta}})|_{\hat{\theta}_y+d_y} - u_x(\hat{\boldsymbol{\theta}})|_{\hat{\theta}_y-d_y}}{2d_y}$$

where $u_x(\hat{\boldsymbol{\theta}})|_{\hat{\theta}_y \pm d_y}$ is $u_x(\boldsymbol{\theta})$ evaluated at $\hat{\boldsymbol{\theta}}$ but the yth element $\hat{\theta}_y$ is replaced by $\hat{\theta}_y \pm d_y$. Simulation studies indicate that choosing $d_y = 0.01\hat{\theta}_y$ is adequate for approximating the derivatives: the errors in the resulting estimates are negligible.

3. SIMULATION STUDIES

A simulation study was done to assess the performance of the proposed estimator (with the normal assumptions). We considered cases for which the number of covariates $s = 1$ and 2; $\nu = 200$ and 400. Without loss of generality, let $\boldsymbol{\mu} = \mathbf{0}$ and all the variances of the covariates be 1. The procedure for generating the data is as follows: for the ith individual with given $(\boldsymbol{\mu}, \mathbf{V})$, $\mathbf{Z}_i$ is generated first, then with given $\boldsymbol{\beta}$ the capture history is generated. Let P_{meas} be the probability of a measurement being taken for each capture. Then with given P_{meas}, for each capture, whether a measurement is taken or not is decided; if a measurement is taken, the corresponding measurement errors are generated with given $\boldsymbol{\Sigma}$. For convenience of comparison, for the different values of P_{meas} and $\boldsymbol{\Sigma}$, given $(\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\beta}, \nu)$, all the simulation results are based on the same data set including the capture histories described as above, except the missing cases and measurement errors. For each given $(\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\beta}, \nu)$, four different settings of $(P_{\text{meas}}, \boldsymbol{\Sigma})$ are chosen (see Table 1).

In the simulations, for convenience of comparison, we only keep the repetitions for which the proposed estimating procedure is successful (convergent) on all the four data sets with the same capture histories but different missing cases and measurement errors due to the four different values of $(P_{\text{meas}}, \boldsymbol{\Sigma})$. This will not affect the true performances of the proposed estimator since there are very few failures among the repetitions as long as the capture proportion is fairly large. We generate the random data until 1,000 such repetitions are obtained. The proportion of failures (denoted by pf) for each setting is listed, $pf = a/b$: the denominator b is the total number of repetitions among which there are 1,000 successful repetitions on the four settings of $(P_{\text{meas}}, \boldsymbol{\Sigma})$, given $(\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\beta}, \nu)$; the numerator a is the number of failures for the corresponding setting among the b repetitions. All simulation results are based on 1,000 such repetitions, av denotes the average of the 1,000 values of $\hat{\nu}$; avse denotes the average of 1,000 values of $\widehat{\text{se}}(\hat{\nu})$; SD denotes the standard deviation of the 1,000 values of $\hat{\nu}$; CP denotes the coverage of the 95% confidence intervals for ν (i.e., $\hat{\nu} \pm 1.96 \times \widehat{\text{se}}(\hat{\nu})$); $\bar{n}_{\text{miss}}$ denotes the average number of the individuals which are captured during the whole experiment but with missing covariates; and P_{T} denotes the average total capture proportion by the end of experiment. Some simulation results are presented in Table 1, in which $\hat{\nu}^{(\text{prop})}$ denotes the proposed estimator.

In the settings A and B of Table 1, $\sqrt{\Sigma/V}$ is 0.5 and 1, this means the measurement errors are large and very large; β_1 is 0.5 and 1, which means the capture probability is sensitive and very sensitive to the covariate (referring to the properties of the logistic function); for setting C, for which $s = 2$, the superposition of the measurement errors means larger measurement errors on the whole. Simulation studies show the proposed estimator $\hat{\nu}^{(\text{prop})}$ is slightly positively biased, especially for large measurement errors along with a sensitive capture probability. As ν increases or the total capture proportion increases, $(\hat{\nu}^{(\text{prop})} - \nu)/\nu$ decreases and becomes very small. Further, SD$(\hat{\nu})$ and avse$(\hat{\nu})$ are quite close, indicating that the sandwich method for the variance estimate is satisfactory. When the measurement errors are large and the capture probability is sensitive to

TABLE 1: Simulation results for the proposed estimating procedure based on 1,000 successful repetitions.

(A) $s = 1$, $\nu = 200$ ($P_T = 77.2\%$)		$\mu = 0$, $V = 1$, $\boldsymbol{\beta} = (-1.0, 0.5)^T$, $t = 5$									
		$P_{\text{meas}} = 1.0$ ($\bar{n}_{\text{miss}} = 0$)					$P_{\text{meas}} = 0.9$ ($\bar{n}_{\text{miss}}$=7.4)				
		av	SD	avse	CP	pf	av	SD	avse	CP	pf
$\Sigma = 0.5^2$	$\hat{\nu}^{(\text{prop})}$	202	14	14	95	$\frac{0}{1,001}$	202	14	14	95	$\frac{0}{1,001}$
	$\hat{\nu}^{(\text{CL})}$	198	13	12	91		184	12	11	60	
	$\hat{\nu}^{(\text{CLm})}$	199	13	13	93		195	12	12	87	
$\Sigma = 1.0^2$	$\hat{\nu}^{(\text{prop})}$	203	15	15	95	$\frac{0}{1,001}$	203	16	15	95	$\frac{1}{1,001}$
	$\hat{\nu}^{(\text{CL})}$	193	11	11	82		180	11	9	43	
	$\hat{\nu}^{(\text{CLm})}$	195	12	11	86		191	11	11	79	

(B) $s = 1$, $\nu = 200$ ($P_T = 83.1\%$)		$\mu = 0$, $V = 1$, $\boldsymbol{\beta} = (-0.5, 1.0)^T$, $t = 5$									
		$P_{\text{meas}} = 1.0$ ($\bar{n}_{\text{miss}} = 0$)					$P_{\text{meas}} = 0.9$ ($\bar{n}_{\text{miss}} = 5.4$)				
		av	SD	avse	CP	pf	av	SD	avse	CP	pf
$\Sigma = 0.5^2$	$\hat{\nu}^{(\text{prop})}$	203	15	15	92	$\frac{3}{1,005}$	203	16	14	93	$\frac{2}{1,005}$
	$\hat{\nu}^{(\text{CL})}$	193	11	9	75		183	10	8	44	
	$\hat{\nu}^{(\text{CLm})}$	198	13	11	87		194	11	10	79	
$\Sigma = 1.0^2$	$\hat{\nu}^{(\text{prop})}$	203	16	15	93	$\frac{2}{1,005}$	203	16	15	93	$\frac{1}{1,005}$
	$\hat{\nu}^{(\text{CL})}$	184	8	6	30		176	8	6	8	
	$\hat{\nu}^{(\text{CLm})}$	190	10	9	65		186	9	8	51	

(C) $s = 2$, $\nu = 400$ ($P_T = 75.9\%$)		$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\mathbf{V} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$, $\boldsymbol{\beta} = (-1.0, 0.5, 1.0)^T$, $t = 6$									
		$P_{\text{meas}} = 1.0$ ($\bar{n}_{\text{miss}} = 0$)					$P_{\text{meas}} = 0.8$ ($\bar{n}_{\text{miss}} = 22.8$)				
		av	SD	avse	CP	pf	av	SD	avse	CP	pf
$\boldsymbol{\Sigma} = \mathbf{D}(0.2)$	$\hat{\nu}^{(\text{prop})}$	403	28	27	94	$\frac{6}{1,009}$	403	28	27	93	$\frac{5}{1,009}$
$\boldsymbol{\Sigma} = \mathbf{D}(0.5)$	$\hat{\nu}^{(\text{prop})}$	403	29	28	94	$\frac{6}{1,009}$	404	29	28	94	$\frac{7}{1,009}$

$\mathbf{D}(\sigma) = \text{diag}(\sigma^2, \sigma^2)$.

the corresponding covariates, the coverages of the 95% confidence intervals are a little smaller than 95%. Using the 95% confidence interval proposed by Chao (1987) which is calculated using a log transformation, the coverage increases by 0.5–1.5%. For a fairly large capture proportion, there are very few failures for the proposed algorithm. As measurement errors become large or the number of missing covariates increases, the performance of the proposed estimator $\hat{\nu}^{(\mathrm{prop})}$ does not change much. It might be that $\hat{\nu}^{(\mathrm{prop})}$ makes full use of the normal assumptions and the replicates. In the case of no missing covariates, the estimator proposed by Hwang & Huang (2003), denoted by $\hat{\nu}^{(\mathrm{HH})}$ was applied. This estimator is obtained by assuming the variances of the measurement errors are known and using only the first covariate observations. Simulation results (not reported) show, for moderately large measurement errors, $\hat{\nu}^{(\mathrm{HH})}$ is better than $\hat{\nu}^{(\mathrm{prop})}$ in terms of RMSE; but for large measurement errors, $\hat{\nu}^{(\mathrm{prop})}$ is better.

For the settings without missing covariates (i.e., $P_{\mathrm{meas}} = 1.0$) in Table 1, the conditional likelihood with the Horvitz–Thompson estimator, denoted by $\hat{\nu}^{(\mathrm{CL})}$ was applied. This estimator is obtained by regarding the first observations as the exact values of the covariates. It is found for $\Sigma = 1.0^2$, $\hat{\nu}^{(\mathrm{CL})}$ is seriously negatively biased, especially for $\beta_1 = 1.0$ (in the setting B); for $\Sigma = 0.5^2$, the biases are not so serious, especially for $\beta_1 = 0.5$ (in the setting A). Our simulation studies show, when measurement errors are large and the capture probability is sensitive to the corresponding covariate, then $\hat{\nu}^{(\mathrm{CL})}$ is substantially negatively biased. With missing covariates, by excluding the captured individuals with missing covariates, the negative bias becomes greater. Further, we compute the modified estimate of $\hat{\nu}^{(\mathrm{CL})}$, denoted by $\hat{\nu}^{(\mathrm{CLm})}$. This estimate is obtained using the average of observations, $\sum_j \mathbf{W}_{ij}$, as the exact covariate value for the ith individual; the resulting estimate is adjusted by multiplying by $(1 - n_{\mathrm{miss}}/n)^{-1}$, where n is the number of distinct captured individuals, n_{miss} is the number of the distinct captured individuals with missing covariate. Simulation studies show that $\hat{\nu}^{(\mathrm{CLm})}$ is a big improvement on $\hat{\nu}^{(\mathrm{CL})}$. In the case that measurement error is not large and there are very few missing covariates, the performance of $\hat{\nu}^{(\mathrm{CLm})}$ is quite good.

We consider two special cases with the proposed estimator.

(i) *Missing covariates only, no covariate measurement errors.* For this case, measurement only needs to be taken once for each captured individual. The proposed estimating procedure for this case is described in Section 2.3. For the ith individual, the probability of no observed covariate values is $(1 - P_{\mathrm{meas}})^{\sum_l \delta_{il}}$. For comparison in the simulation study, we apply the conditional likelihood with the Horvitz–Thompson estimator $\hat{\nu}^{(\mathrm{CL})}$ which excludes the individuals captured but with missing covariates. Also, we compute the modified estimate $\hat{\nu}^{(\mathrm{CLm})}$ (with the adjusted factor $(1 - n_{\mathrm{miss}}/n)^{-1}$, but without the use of the average of observations). Some simulation results are presented in Table 2. The proportion of failures $pf = a/b$: the denominator b is the total number of repetitions among which there are 1,000 repetitions for which both $\hat{\nu}^{(\mathrm{prop})}$ and $\hat{\nu}^{(\mathrm{CL})}$ are successful; the numerator a is the number of failures for the corresponding estimator among the b repetitions. The simulation results are based on the 1,000 successful repetitions.

There is a little positive bias for $\hat{\nu}^{(\mathrm{prop})}$, but as the total capture proportion increases, the bias will diminish. In whole, $\hat{\nu}^{(\mathrm{prop})}$ performs well. $\hat{\nu}^{(\mathrm{CL})}$ seriously underestimates ν, and the 95% confidence interval has very poor coverage. $\hat{\nu}^{(\mathrm{CLm})}$ considerably improves on the results of $\hat{\nu}^{(CL)}$ but still worse than $\hat{\nu}^{(\mathrm{prop})}$.

(ii) $m_i = 1$ *for all captured individuals.* In this case, for each captured individual there is only one observation for the covariates without replicates and none is missing. For such a data set, the proposed algorithm does not converge. When the covariates are observed only once for each captured individual, it is impossible to estimate $\mathbf{V}$ and $\boldsymbol{\Sigma}$ separately; there is an overparameterization problem. For this case, we must assume $\boldsymbol{\Sigma}$ is known. This is the model

TABLE 2: Simulation results for comparison of the proposed estimator $\hat{\nu}^{(\text{prop})}$ with $\hat{\nu}^{(\text{CL})}$ and $\hat{\nu}^{(\text{CLm})}$ when there are missing covariates only $s = 1$, $Z_i \sim N(0, 1)$, $\boldsymbol{\beta} = (-1.5,\ 0.5)^{\text{T}}$.

		$\nu = 1,000$					$\nu = 200$				
$P_{\text{meas}} = 0.7$		av	SD	avse	CP	pf	av	SD	avse	CP	pf
$t = 4$	$\hat{\nu}^{(\text{prop})}$	1,012	71	68	95	$\frac{0}{1,000}$	207	34	34	94	$\frac{12}{1,012}$
$(P_{\text{T}} = 55.7\%)$	$\hat{\nu}^{(\text{CL})}$	689	47	42	0	$\frac{0}{1,000}$	141	22	20	22	$\frac{0}{1,012}$
$(\bar{p}_{\text{miss}} = 12.7\%)$	$\hat{\nu}^{(\text{CLm})}$	893	56	55	45		182	27	26	77	
$t = 8$	$\hat{\nu}^{(\text{prop})}$	1,003	28	28	95	$\frac{0}{1,000}$	202	13	13	95	$\frac{0}{1,000}$
$(P_{\text{T}} = 78.0\%)$	$\hat{\nu}^{(\text{CL})}$	796	24	19	0	$\frac{0}{1,000}$	160	10	9	6	$\frac{0}{1,000}$
$(\bar{p}_{\text{miss}} = 12.3\%)$	$\hat{\nu}^{(\text{CLm})}$	945	23	23	34		190	11	11	78	
$t = 12$	$\hat{\nu}^{(\text{prop})}$	1,001	16	16	96	$\frac{0}{1,000}$	201	7	7	94	$\frac{0}{1,000}$
$(P_{\text{T}} = 88.2\%)$	$\hat{\nu}^{(\text{CL})}$	865	16	12	0	$\frac{0}{1,000}$	173	8	6	5	$\frac{0}{1,000}$
$(\bar{p}_{\text{miss}} = 9.7\%)$	$\hat{\nu}^{(\text{CLm})}$	972	14	14	44		195	7	6	80	
$P_{\text{meas}} = 0.8$		av	SD	avse	CP	pf	av	SD	avse	CP	pf
$t = 4$	$\hat{\nu}^{(\text{prop})}$	1,012	70	67	94	$\frac{0}{1,000}$	207	32	33	94	$\frac{14}{1,014}$
$(P_{\text{T}} = 55.7\%)$	$\hat{\nu}^{(\text{CL})}$	792	53	49	6	$\frac{0}{1,000}$	161	24	23	48	$\frac{0}{1,014}$
$(\bar{p}_{\text{miss}} = 8.1\%)$	$\hat{\nu}^{(\text{CLm})}$	927	59	57	67		189	28	27	84	
$t = 8$	$\hat{\nu}^{(\text{prop})}$	1,003	27	28	95	$\frac{0}{1,000}$	202	13	13	95	$\frac{0}{1,000}$
$(P_{\text{T}} = 78.0\%)$	$\hat{\nu}^{(\text{CL})}$	869	24	22	0	$\frac{0}{1,000}$	175	11	10	32	$\frac{0}{1,000}$
$(\bar{p}_{\text{miss}} = 7.6\%)$	$\hat{\nu}^{(\text{CLm})}$	962	24	24	63		194	11	11	86	
$t = 12$	$\hat{\nu}^{(\text{prop})}$	1,001	16	16	96	$\frac{0}{1,000}$	201	7	7	94	$\frac{0}{1,000}$
$(P_{\text{T}} = 88.2\%)$	$\hat{\nu}^{(\text{CL})}$	918	16	14	0	$\frac{0}{1,000}$	184	7	6	30	$\frac{0}{1,000}$
$(\bar{p}_{\text{miss}} = 5.8\%)$	$\hat{\nu}^{(\text{CLm})}$	981	14	14	70		197	7	7	87	
$P_{\text{meas}} = 0.9$		av	SD	avse	CP	pf	av	SD	avse	CP	pf
$t = 4$	$\hat{\nu}^{(\text{prop})}$	1,011	70	66	95	$\frac{0}{1,000}$	206	32	33	94	$\frac{15}{1,015}$
$(P_{\text{T}} = 55.7\%)$	$\hat{\nu}^{(\text{CL})}$	898	60	59	51	$\frac{0}{1,000}$	183	28	27	78	$\frac{0}{1,015}$
$(\bar{p}_{\text{miss}} = 3.9\%)$	$\hat{\nu}^{(\text{CLm})}$	966	64	61	86		197	29	29	90	
$t = 8$	$\hat{\nu}^{(\text{prop})}$	1,003	27	27	95	$\frac{0}{1,000}$	202	13	13	95	$\frac{0}{1,000}$
$(P_{\text{T}} = 78.0\%)$	$\hat{\nu}^{(\text{CL})}$	938	25	24	30	$\frac{0}{1,000}$	189	12	11	75	$\frac{0}{1,000}$
$(\bar{p}_{\text{miss}} = 3.5\%)$	$\hat{\nu}^{(\text{CLm})}$	982	26	26	86		198	12	12	92	
$t = 12$	$\hat{\nu}^{(\text{prop})}$	1,001	16	16	96	$\frac{0}{1,000}$	201	7	7	94	$\frac{0}{1,000}$
$(P_{\text{T}} = 88.2\%)$	$\hat{\nu}^{(\text{CL})}$	962	16	15	30	$\frac{0}{1,000}$	193	7	7	74	$\frac{0}{1,000}$
$(\bar{p}_{\text{miss}} = 2.6\%)$	$\hat{\nu}^{(\text{CLm})}$	991	15	15	90		199	7	7	93	

$\bar{p}_{\text{miss}} = \bar{n}_{\text{miss}}/\nu$.

considered by Hwang & Huang (2003). In this case we apply the proposed algorithm but ignoring (6) since $\mathbf{\Sigma}$ is known. Simulation studies show that the performance is very close to the case when we allow for replicates but without knowing the variances of the measurement errors. With large measurement errors, the performance of the proposed estimator is better than that of Hwang & Huang (2003), since the proposed estimator makes full use of the normal assumptions. Also, if missing covariates occur in this case, that is, some $m_i = 0$, with the assumption of the known variances of the measurement errors, the proposed inference procedure is still applicable.

For capture–recapture models with heterogeneity in capture probability, Link (2003) shows that different assumptions for the distribution of the capture probability may lead to vastly different estimates of the population size. For the proposed model with the normal assumption, if the true distributions of the covariates are not normal, a simulation study is conducted to check the sensitivity of the estimate of the population size. For each of the settings in Tables 1 and 2 (except the setting C in Table 1), instead of the normal distributions, the random data for the covariates are generated with Gamma distributions having the same variances as the normal distributions. The covariates are obtained by subtracting the means of Gamma distributions from the generated data. Therefore, the obtained covariates have the same means and variances as the normal distributions, but with gamma shapes. The shape parameter is chosen as 1.5, 2, 3, and 6, respectively. Based on 1,000 repetitions for each of the shape parameters in each of the settings, the simulation results show that the estimates of population sizes are quite close to the corresponding results in Tables 1 and 2. Generally speaking, *CP* varies from 88% to 94%, a little smaller than the corresponding values in Tables 1 and 2. The RMSEs are almost the same or even a little better/smaller than the corresponding values in Tables 1 and 2. Besides the Gamma distributions, the covariates are also generated with truncated normal distributions, including left truncated, right truncated, and two-sided truncated. The probability of the truncated part is selected from 0.05 to 0.15. The simulation results show similar results: *CP* is a little worse but RMSE is almost the same or even a little better, when compared with the corresponding values in Tables 1 and 2. This simulation study suggests that the estimate of population size for the proposed model with the normal assumption is reasonably robust.

4. MAI PO EXAMPLE

We re-examined the capture–recapture data studied by Hwang & Huang (2003) on bird species *P. flaviventris* in Hong Kong Mai Po Bird Sanctuary. The bird's wing length is shown to be an important covariate associated with capture probability but it is subject to measurement error. Ignoring the measurement error leads to a serious underestimation (see Hwang & Huang, 2003). In this data set, there are 165 birds caught in 207 captures on 17 trapping occasions. For each capture (including recapture), the wing length of the captured bird was measured, but not always. It is found that $m_i < n_i$ for three of the 165 captured birds, that is, the number of measuring times is less than the number of capture times. For one of the captured birds, $m_i = 0$, that is, the measurement of the wing length is missing. For the estimator proposed by Hwang & Huang (2003), the captured bird with no record of wing length has to be excluded, the variance of the measurement error is assumed known and the first observation of wing length for each captured individuals is used.

We apply the proposed estimator which incorporates the replicates and allows for missing covariates. The capture probability is modelled with a logistic function $H(\beta_0 + \beta_1 Z_i)$, where Z_i is the exact wing length of the ith individual and is a realization of a normal distribution $N(\mu, V)$. The observed wing length is the true wing length plus a measurement error ε_{ij}, where $\varepsilon_{ij} \sim N(0, \sigma^2)$ and σ^2 is unknown. We list the results in Table 3 where $\hat{\nu}^{(\mathrm{HH})}$ and $\hat{\nu}^{(\mathrm{CL})}$ are applied to the data in which the captured bird with missing covariate is excluded. For $\hat{\nu}^{(\mathrm{HH})}$, σ^2

TABLE 3: Summary results of the proposed estimator and the others for the *P. flaviventris* data in Hong Kong *Mai Po*.

Estimator	$\hat{\nu}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\mu}$	$\hat{V}$	$\hat{\sigma}^2$
$\hat{\nu}^{(\mathrm{prop})}$	542 (104.9)	−26.0 (5.15)	0.493 (0.112)	44.7 (0.173)	1.32 (0.206)	0.325 (0.083)
	513 (94.8)	−21.9 (5.52)	0.403 (0.120)	Giving $\sigma^2 = 0.0625$ (i.e., 0.25^2)		
$\hat{\nu}^{(\mathrm{HH})}$	529 (106.7)	−24.5 (6.40)	0.461 (0.139)	Giving $\sigma^2 = 0.2500$ (i.e., 0.50^2)		
	570 (141.1)	−31.2 (8.69)	0.608 (0.189)	Giving $\sigma^2 = 0.5625$ (i.e., 0.75^2)		
$\hat{\nu}^{(\mathrm{CL})}$	512 (95.3)	−21.1 (5.28)	0.386 (0.115)			
$\hat{\nu}^{(\mathrm{CLm})}$	524 (100.6)					
$\hat{\nu}^{(\mathrm{prop})\mathrm{a}}$	539 (104.8)	−26.1 (5.30)	0.497 (0.116)	44.7 (0.172)	1.26 (0.207)	0.320 (0.085)
$\hat{\nu}^{(\mathrm{CL})\mathrm{a}}$	497 (92.1)					
$\hat{\nu}^{(\mathrm{CLm})\mathrm{a}}$	526 (101.8)					

[a] Estimator is applied to an artificial data in which four more captured birds are assumed with missing records of wing length.

is given and uses only the first observation of wing length for each captured bird. For $\hat{\nu}^{(\mathrm{CL})}$, the first observation of wing length is regarded as the exact length. For $\hat{\nu}^{(\mathrm{CLm})}$, the average of the observations of wing length is regarded as the exact length, the resulting estimate is adjusted by multiplying $(1 - n_{\mathrm{miss}}/n)^{-1}$.

For $\hat{\nu}^{(\mathrm{HH})}$, the different given values of σ^2 lead to quite different estimates for ν. The proposed estimator $\hat{\nu}^{(\mathrm{prop})}$ which uses the replicates and allows for missing covariates gives $\hat{\nu} = 542(104.9)$, where the number in brackets is the standard error, (the Fortran code with an explanation file can be obtained from the first author). Since $\hat{\sigma}^2 = 0.325(0.083)$ and $\hat{V} = 1.322(0.206)$, the measurement errors are large; $\hat{\beta}_1 = 0.493(0.112)$, this means the capture probability is sensitive to the wing length. Therefore, applying $\hat{\nu}^{(\mathrm{CL})}$ to the data will seriously underestimate ν. Using $\hat{\nu}^{(\mathrm{CLm})}$ significantly improves the estimate. Yip, Lin & Xi (2005) used this data set to illustrate the continuous-time model by treating the data as continuous-time data, the result is $\hat{\nu} = 578(153.4)$, quite different with that of $\hat{\nu}^{(\mathrm{prop})}$.

In addition, we randomly select four other individuals among the 164 captured individuals with no missing covariates, and assume the corresponding records of wing length are missing. Applying the proposed estimator to these artificial data sets that each have a total of five captured individuals with missing covariates, we found that the estimate of ν changes little or is almost unchanged, a typical result is presented in Table 3. Applying $\hat{\nu}^{(\mathrm{CL})}$ by excluding the five individuals yields a much smaller estimate for ν, $\hat{\nu}^{(\mathrm{CLm})}$ gives a significantly improved estimate.

5. DISCUSSION

When there are large measurement errors on a covariate to which the capture probability is sensitive or there are a large number of captured individuals with missing covariates, the conditional likelihood with the Horvitz–Thompson estimator leads to an estimate which is seriously negatively biased. The method proposed here provides an improved tool for such situations, provided appropriate distributions can be found for the covariates and measurement errors. For capture–recapture experiments, since only part of the population of interest can be observed, it is not easy to identify these distributions, (if only observed data are used). Link (2003) shows that different assumptions regarding capture probability distribution can fit equally well with observed capture frequency data, while the different assumptions may lead to vastly different estimates of the population size. Therefore, the identification of the distributions of the covariates in the proposed model is an important issue. To better identify the covariate distributions, besides the observed data, it is necessary to incorporate some other information, for example, historical observation (if possible). Because of the shortcomings of the proposed parametric model, it is worthwhile to develop a semi-parametric or nonparametric method for the case with replicates of mismeasured covariates and possibly missing covariate data.

For the proposed model with the normal assumption, the simulation study suggests that the estimate of population size is reasonably robust. This suggestion is different from that of Link (2003). In Link (2003), the population sizes are estimated with the number of distinct captured individual divided by the overall probability of being captured at least once that directly depends on the distribution of capture probability. The proposed estimate is based on the Horvitz–Thomson type estimate which might be more appropriate for heterogeneous populations when the capture probabilities are mainly decided by some individual's covariates.

If some major covariates affecting the capture probability are not included in the model, the final results may be quite unreliable. With the proposed model, it is likely that a much more homogeneous population would be suggested and the population size would be underestimated. We should be very careful to investigate which covariates really affect the capture probability and how they affect it.

ACKNOWLEDGEMENTS

The authors are grateful for the many constructive comments and suggestions from the Editor, the Associate Editor and two referees which have been incorporated in the paper. The research was supported by the Research Grant Committee of the Hong Kong Government.

BIBLIOGRAPHY

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

R. J. Carroll, D. Ruppert, L. A. Stefanski & C. M. Crainiceanu (2006). "*Measurement Error in Nonlinear Models: A Modern Perspective*," 2nd ed., Chapman & Hall/CRC Press, Taylor & Francis Group, Boca Raton, FL.

A. Chao. (1987). Estimating the population size for capture–recapture data with unequal catchability. *Biometrics*, 43, 783–791.

A. P. Dempster, N. M. Laird & D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society B*, 39, 1–38.

P. J. Diggle, K. Y. Liang & S. L. Zeger (1994). "*Analysis of Longitudinal Data*," Oxford University Press, Oxford.

A. H. Herring & G. Ibrahim (2001). Likelihood-based methods for missing covariates in the Cox proportional hazards model. *Journal of the American Statistical Association*, 96, 292–302.

D. G. Horvitz & D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663–685.

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

W. H. Hwang & S. Y. H. Huang (2003). Estimation in capture–recapture models when covariates are subject to measurement errors. *Biometrics*, 59, 1115–1124.

R. King, S. P. Brooks & T. Coulson (2008). Analysing complex capture–recapture data: Incorporating time-varying covariate information. *Biometrics*, 64, 1187–1195.

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

R. J. A. Little & D. B. Rubin (2002). "*Statistical Analysis With Missing Data*," 2nd ed., John Wiley, New York.

C. E. McCulloch & S. R. Searle (2001). "*Generalized, Linear, and Mixed Models*," Wiley, New York.

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

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

M. S. Wulfsohn & A. A. Tsiatis (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53, 330–339.

L. Xi, P. S. F. Yip & R. Watson (2007). A unified likelihood-based approach for estimating population size in continuous-time capture–recapture experiments with frailty. *Biometrics*, 63, 228–236.

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

Received 18 September 2007
Accepted 10 May 2009