EFFECTIVE USE OF MULTIPLE ERROR-PRONE COVARIATE MEASUREMENTS IN CAPTURE-RECAPTURE MODELS

Kun Xu and Yanyuan Ma

Texas A&M University

Abstract: We study capture-recapture models in a closed population when multiple error-prone measurements of a covariate are available. Due to the identity between the number of captures and the number of measurements, no suitable complete and sufficient statistic exists, and the existing method no longer applies. The familiar strategy of generalized method of moments fails to resolve this issue satisfactorily, and complexity lies in the loss of the surrogacy assumption commonly assumed in measurement error problems. Our approach to this problem through a semi-parametric treatment overcomes these difficulties. The superior performance of the new method is demonstrated through numerical experiments in simulated and data examples.

Key words and phrases: Capture-recapture, complete sufficient statistic, generalized method of moments, measurement error, semiparametrics, surrogacy.

1. Introduction

Capture-recapture models are widely used to describe the abundance of a species of interest. Modeling the probability of different numbers of captures of a single animal as a function of the associated covariates enables use of the observed covariate and capture information to infer the total population size. We work in the closed population framework where there is no population flow such as mortality and immigration occurring during the experiment. For various reasons, covariate measurements are almost always prone to error. Hwang, Huang, and Wang (2007) studied the effect of covariate measurement error on estimating the population size in capture-recapture models. To find that ignoring the covariate measurement error leads to estimation bias, they proposed an effective method to correct this bias through accounting for the measurement error structure.

In capture-recapture models individual animals can be captured multiple times. Commonly at each capture event the same covariates are measured, especially if they are difficult to measure precisely or their values could fluctuate. If these are different measurements of an underlying true covariate that affects the capture probability, it is natural to use their average in the capture-recapture

model. If the measurements are systematic, one expects their average gains in precision with more captures. This different measurement error variabilities result for animals captured different number of times. This information is ignored in Hwang, Huang, and Wang (2007), who assume a single covariate measurement is available and that has a common error variance across different animals. Huggins and Hwang (2009) realized the advantage of multiple measurements and successfully utilized it. However, they made an additional normality assumption on the unobservable covariates. Then the problem is completely parametric and no longer in the functional measurement error model framework. It is our goal to take into account the multiple measurements in the estimation procedure, to retain consistency, and to improve estimation efficiency.

In the capture-recapture literature, the capture probability is typically assumed to relate to covariates through a linear logistic model (Huggins (1989)), see Pollock (2002) for a comprehensive review of this topic. Extension to a partially linear structure was studied in Huggins (2006), while Hwang and Huggins (2007) further incorporated categorical variables. When covariates are measured with error, Wang (2000) proposed a refined regression calibration estimator and Hwang and Huang (2003) proposed a conditional likelihood-based method to estimate the population size. A conditional score-based method was later proposed by Hwang, Huang, and Wang (2007); Huggins and Hwang (2009) extended the method to handle unknown measurement error variance. Our contribution in this work is to allow unknown, unequal measurement error variance that depends on the capture times, and to construct consistent and efficient estimators that benefit from the special error properties.

The rest of the paper is organized as follows. In Section 2, we investigate a generalized method of moments (GMM) procedure. In Section 3, we propose an effective way of using multiple measurements based on semiparametrics. Numerical experiments are reported for simulated examples in Section 4 and for capture-recapture data of a bird population in Section 5. We conclude with a discussion in Section 6, and relegate proofs to the Appendix.

2. Generalized Method of Moments Procedure

In the capture-recapture model we use N to denote the total population size of the animals under study. Interest lies in estimating and making inference about N. Let J be the total number of capture occasions. Using i to index the distinct animals and j to index the capture occasions, we denote the random event of the ith animal being captured on the jth occasion as Y_{ij}, with $Y_{ij} = 1$ for capture and $Y_{ij} = 0$ otherwise. Here $i = 1, \ldots, N$ and $j = 1, \ldots, J$. Assume there are a total of D distinct animals captured at least once. For convenience, we label these animals from 1 to D and the animals never captured from $D + 1$

to N. A widely used model to describe the probability mass function of a binary outcome is the logistic regression model

$$\text{logit}\{\text{pr}(Y_{ij} = 1 \mid \mathbf{X}_i = \mathbf{x}_i)\} = \alpha + \mathbf{x}_i^{\text{T}}\boldsymbol{\beta}$$

that relates the capture probability of an animal on one occasion to its covariates, see for example Pollock, Hines, and Nichols (1984), Huggins (1989), Alho (1990), Huggins (1991), and Pledger (2000). Let Y_i be the total number of captures of the ith animal, so $Y_i = \sum_{j=1}^{J} Y_{ij}$. If for each animal, conditional on its covariates, the different captures are independent of each other, we have

$$\text{pr}(Y_i = y_i \mid \mathbf{X}_i = \mathbf{x}_i) = \binom{J}{y_i} \frac{\exp\{(\alpha + \mathbf{x}_i^{\text{T}}\boldsymbol{\beta})y_i\}}{\{1 + \exp(\alpha + \mathbf{x}_i^{\text{T}}\boldsymbol{\beta})\}^J},$$

the logistic based capture-recapture model.

When the covariates $\mathbf{X}_i$'s are not directly observed, alternative information is usually collected. We consider that at each capture of an animal, its covariate is measured, subject to measurement error. Let $\mathbf{W}_{ij}$ be the measurement of $\mathbf{X}_i$ at the jth occasion if $Y_{ij} = 1$. We assume $\mathbf{W}_{ij} = \mathbf{X}_i + \mathbf{U}_{ij}$, where $\mathbf{U}_{ij}$ is a random measurement error, assumed to have a normal distribution with mean zero and variance-covariance matrix $\boldsymbol{\Sigma}$. Our goal is to estimate N based on $(Y_{ij}, Y_{ij}\mathbf{W}_{ij}), i = 1, \ldots, N, j = 1, \ldots, J$.

If only a single measurement $\mathbf{W}_i$ is available in place of $\mathbf{X}_i$, where $\mathbf{W}_i$ has the same relation to $\mathbf{X}_i$ as the $\mathbf{W}_{ij}$'s above, results have been established in the literature, see Hwang, Huang, and Wang (2007). If $(Y_i, \mathbf{W}_i)$ are fixed and the parameters $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})$ are known in the joint probability density function of $(Y_i, \mathbf{W}_i)$ given $\mathbf{X}_i$, then $\Delta_i = \mathbf{W}_i + Y_i\boldsymbol{\Sigma}\boldsymbol{\beta}$ is a complete sufficient statistic for $\mathbf{X}_i$. Thus, conditional on Δ_i, because of the sufficiency, $\mathbf{Y}_i$ does not depend on $\mathbf{X}_i$. Taking advantage of completeness, one can construct an estimating equation based on $(\mathbf{Y}_i, \Delta_i)$ alone; see Ma and Tsiatis (2006) for the details on how sufficiency and completeness contribute to the construction of the estimator. One might use $\overline{\mathbf{W}}_i = (\sum_{j=1}^{J} Y_{ij}\mathbf{W}_{i,j})/Y_i$ in place of $\mathbf{W}_i$. But for the different measurement error variances across different animals, it is no longer a sufficient or complete statistic.

An alternative obvious attempt of taking advantage of the situation is to combine the procedures in Hwang, Huang, and Wang (2007) performed on each individual $\mathbf{W}_{ij}$. To this end, we resort to the GMM (Hansen (1982)) approach. Our consideration is the following. We first consider making use of the first measurement of each animal that is captured at least once, forming the complete sufficient statistic with the first measurement. This provides the first set of estimating equations. We then consider making use of the second measurement of each animal that is captured at least twice, forming the complete sufficient

statistic with the second measurement. This provides the second set of estimating equations. We continue this process and obtain a maximum of J sets of estimating equations in total. We then use GMM to take advantage of all these equations. Specifically, let $\mathcal{C}_{ik}$ denote the event $Y_i \geq k$ for $k = 1, \ldots, J$. Thus, $I(\mathcal{C}_{ik}) = 1$ if the ith animal is captured at least k times, and $I(\mathcal{C}_{ik}) = 0$ otherwise. For the ith animal with Y_i total captures, we denote its Y_i available measurements $\mathbf{W}_{i(l)}$, $l = 1, \ldots, Y_i$. Thus, the lth complete sufficient statistic is defined by $\Delta_{i(l)} = \mathbf{W}_{i(l)} + Y_i \mathbf{\Sigma} \boldsymbol{\beta}$, $l = 1, \ldots, Y_i$.

Using the above notation, the kth set of estimating equations can be written as

$$\sum_{i=1}^{N} \mathbf{g}_k(Y_i, \Delta_{i(k)}, \alpha, \boldsymbol{\beta})$$

$$= \sum_{i=1}^{N} I(\mathcal{C}_{ik}) \left[\begin{array}{c} Y_i - E(Y_i \mid \Delta_{i(k)}, \mathcal{C}_{ik}) \\ \{\Delta_{i(k)} - E(Y_i \mid \Delta_{i(k)}, \mathcal{C}_{ik}) \mathbf{\Sigma} \boldsymbol{\beta}\} \{Y_i - E(Y_i \mid \Delta_{i(k)}, \mathcal{C}_{ik})\} \end{array} \right] = 0,$$

where $k = 1, \ldots, K \leq J$. We emphasize here that when $k = 1$, the estimating equation is identical to the proposal in Hwang, Huang, and Wang (2007). Here we use K to denote the maximum k value where there are still data available to form the estimating equation, i.e. the largest possible k such that $\max_i I(\mathcal{C}_{ik}) = 1$. We now combine these K sets of equations via GMM. Specifically, Let $\mathbf{O}_i = (Y_{i1}, Y_{i1} \mathbf{W}_{i1}^{\mathrm{T}}, \ldots, Y_{iJ}, Y_{iJ} \mathbf{W}_{iJ}^{\mathrm{T}})^{\mathrm{T}}$ be the observations related to the ith animal, let $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$, write

$$\mathbf{g}(\mathbf{O}_i, \boldsymbol{\theta}) = \begin{pmatrix} \mathbf{g}_1(Y_i, \Delta_{i(1)}, \boldsymbol{\theta}) \\ \mathbf{g}_2(Y_i, \Delta_{i(2)}, \boldsymbol{\theta}) \\ \vdots \\ \mathbf{g}_K(Y_i, \Delta_{i(K)}, \boldsymbol{\theta}) \end{pmatrix},$$

and obtain the estimator of $\boldsymbol{\theta}$ through minimizing

$$\left\{\sum_{i=1}^{N} \mathbf{g}(\mathbf{O}_i, \boldsymbol{\theta})\right\}^{\mathrm{T}} \left\{\sum_{i=1}^{N} \mathbf{g}(\mathbf{O}_i, \boldsymbol{\theta}) \mathbf{g}^{\mathrm{T}}(\mathbf{O}_i, \boldsymbol{\theta})\right\}^{-1} \left\{\sum_{i=1}^{N} \mathbf{g}(\mathbf{O}_i, \boldsymbol{\theta})\right\}.$$

It is well known that the GMM estimator provides the optimal combination of the estimating equations in terms of the estimation efficiency, and the resulting estimator has the usual root-n consistency and asymptotic normality. Here, estimation efficiency is measured by the inverse of the variance of an estimator. Here, the estimation procedure has $\widehat{\boldsymbol{\theta}}$ consistent, with asymptotic variance

$$N^{-1} \left(E\left\{\frac{\partial \mathbf{g}^{\mathrm{T}}(\mathbf{O}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} \left[E\left\{\mathbf{g}(\mathbf{O}_i, \boldsymbol{\theta}) \mathbf{g}^{\mathrm{T}}(\mathbf{O}_i, \boldsymbol{\theta})\right\}\right]^{-1} E\left\{\frac{\partial \mathbf{g}(\mathbf{O}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\} \right)^{-1}.$$

The GMM uses each measurement $\mathbf{W}_{i(k)}$ separately. Alternatively, as suggested by a referee, one can also consider the averaged measurement $k^{-1}\sum_{l=1}^{k}\mathbf{W}_{i(l)}$ in the construction, with the estimating equation using all animals captured at least k times. Since the averaged quantity does not depend on Y_i, this results in a sufficient complete statistic. One obvious advantage of such a configuration is that it can stabilize the estimation procedure.

To take advantage of the multiple measurements fully, it is tempting to use the average of all the Y_i measurements, but the measurement error problem is now differential and this requires attention.

3. Semiparametric Method

Consider the animals captured at least once. Modeling the probability of the observed animal being captured y times given that it is captured at least once, we have

$$\mathrm{pr}(Y_i = y_i \mid \mathbf{X}_i = \mathbf{x}_i, \mathcal{C}_{i1})$$

$$= \binom{J}{y_i} \frac{\exp\{(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})y_i\}}{\{1 + \exp(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}^J} \left[1 - \left\{\frac{1}{1 + \exp(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right\}^J\right]^{-1}, \quad (3.1)$$

where $y_i = 1, \ldots J$. The averaged measurement $\overline{\mathbf{W}}_i$ satisfies $\overline{\mathbf{W}}_i = \mathbf{X}_i + \mathbf{U}_i$, where $\mathbf{U}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}/Y_i)$, with the convention that if $y_i = 0$, $\overline{\mathbf{W}}_i = 0$. If $\boldsymbol{\Sigma}$ is not known, we can estimate it by forming differences of the repeated observations, say $\mathbf{W}_{i(1)} - \mathbf{W}_{i(2)}$, and calculating the sample variance-covariance matrix, see Hall and Ma (2007) for details. Thus, unless we specifically point out otherwise, we assume $\boldsymbol{\Sigma}$ is known.

The variance expression for $\mathbf{U}_i$ indicates that Y_i and $\overline{\mathbf{W}}_i$ are no longer independent conditional on $\mathbf{X}_i$. This means that the standard surrogacy assumption in the measurement error literature is violated in this context. If p is the dimension of $\mathbf{X}_i$, the joint distribution of Y_i and $\overline{\mathbf{W}}_i$ conditional on $(\mathbf{X}_i, \mathcal{C}_{i1})$ is

$$f_{Y_i, \overline{\mathbf{W}}_i \mid \mathbf{X}_i, \mathcal{C}_{i1}}(y_i, \overline{\mathbf{w}}_i \mid \mathbf{x}_i, \mathcal{C}_{i1}) = \mathrm{pr}(Y_i = y_i \mid \mathbf{X}_i = \mathbf{x}_i, \mathcal{C}_{i1}) f_{\overline{\mathbf{W}}_i \mid Y_i, \mathbf{X}_i, \mathcal{C}_{i1}}(\overline{\mathbf{w}}_i \mid y_i, \mathbf{x}_i, \mathcal{C}_{i1})$$

$$= \binom{J}{y_i} \frac{\exp\{(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})y_i\}}{\{1 + \exp(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}^J} \left[1 - \left\{\frac{1}{1 + \exp(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right\}^J\right]^{-1}$$

$$(2\pi)^{-p/2} \left|\frac{\boldsymbol{\Sigma}}{y_i}\right|^{-1/2} \exp\left\{-\frac{y_i}{2}(\overline{\mathbf{w}}_i - \mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{w}}_i - \mathbf{x}_i)\right\}$$

$$= \binom{J}{y_i} (2\pi)^{-p/2} \left|\frac{\boldsymbol{\Sigma}}{y_i}\right|^{-1/2} \frac{\exp(y_i\alpha - y_i\overline{\mathbf{w}}_i^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{w}}_i/2)}{\{1 + \exp(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}^J - 1} \exp\left\{\mathbf{x}_i^{\mathrm{T}}(y_i\boldsymbol{\beta} + y_i\boldsymbol{\Sigma}^{-1}\overline{\mathbf{w}}_i)\right.$$

$$\left. - \frac{y_i\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}x_i}{2}\right\}.$$

Because it is in the form of the exponential family, the minimal complete statistic is $(Y_i\boldsymbol{\beta} + Y_i\boldsymbol{\Sigma}^{-1}\overline{\mathbf{W}}_i, Y_i)$, or equivalently $(\overline{\mathbf{W}}_i, Y_i)$. This statistic does not help simplify the problem.

Without the surrogacy property or a suitable sufficient and complete statistic, we turn to a semiparametric framework. The distribution of $\mathbf{X}_i$ has to be taken into account and we treat it as a nuisance parameter with infinite dimension, but we do not estimate it. Instead, we calculate its corresponding tangent space formed by the mean squared closure of the set of all score functions of its parametric submodels; the orthogonal complement of the tangent space subsequently contains the elements for building consistent estimating equations. This approach originates in Bickel et al. (1993), and discussion about such calculations can be found in Tsiatis (2006).

The joint distribution of the observed $(Y_i, \overline{\mathbf{W}}_i)$ is

$$
\begin{aligned}
&f_{Y_i, \overline{\mathbf{W}}_i | \mathcal{C}_{i1}}(y_i, \overline{\mathbf{w}}_i \mid \mathcal{C}_{i1}) \\
&\quad = \int f_{\overline{\mathbf{W}}_i | Y_i, \mathbf{X}_i, \mathcal{C}_{i1}}(\overline{\mathbf{w}}_i \mid y_i, \mathbf{x}_i, \mathcal{C}_{i1}) f_{Y_i | \mathbf{X}_i, \mathcal{C}_{i1}}(y_i \mid \mathbf{x}_i, \mathcal{C}_{i1}) f_{\mathbf{X}_i | \mathcal{C}_{i1}}(\mathbf{x}_i \mid \mathcal{C}_{i1}) d\mu(\mathbf{x}_i),
\end{aligned}
$$

where $d\mu(\mathbf{x}_i)$ is the dominating measure. Here $f_{\mathbf{X}_i | \mathcal{C}_{i1}}(\mathbf{x}_i \mid \mathcal{C}_{i1})$ is the unknown probability density function of $\mathbf{X}_i$ conditional on $\mathcal{C}_{i1}$, while the conditional distribution $f_{Y_i | \mathbf{X}_i, \mathcal{C}_{i1}}(y_i \mid \mathbf{x}_i, \mathcal{C}_{i1})$ is completely determined by $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$.

Our goal is to construct an estimating equation based on the conditional joint distribution of $(Y_i, \overline{\mathbf{W}}_i)$ through calculating the efficient score function. To calculate the score function with respect to $\boldsymbol{\theta}$ we have

$$
\mathbf{S}_\theta(Y_i, \overline{\mathbf{W}}_i) \equiv \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ \log f_{Y_i, \overline{\mathbf{W}}_i | \mathcal{C}_{i1}}(y_i, \overline{\mathbf{w}}_i \mid \mathcal{C}_{i1}) \right\} = E\left\{ \mathbf{S}_\theta^F(Y_i, \mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1} \right\},
$$

where $\mathbf{S}_\theta^F(Y_i, \mathbf{X}_i) \equiv \partial \log f_{Y_i | \mathbf{X}_i, \mathcal{C}_{i1}}(y_i \mid \mathbf{x}_i, \mathcal{C}_{i1})/\partial \boldsymbol{\theta}$ and $f_{Y_i | \mathbf{X}_i, \mathcal{C}_{i1}}(y_i \mid \mathbf{x}_i, \mathcal{C}_{i1})$ is given in (3.1). To find the nuisance tangent space Λ and its orthogonal complement with respect to the infinite dimensional parameter $f_{\mathbf{X}_i | \mathcal{C}_{i1}}(\mathbf{x}_i \mid \mathcal{C}_{i1})$, we consider the parametric submodels which lie in the family of the unknown conditional distributions and contain the true distribution. For each of them, the score function with respect to the nuisance parameter can be calculated directly. We take the mean squared closure of all these score functions corresponding to the different submodels to obtain Λ. Detailed calculation in the Appendix yields

$$
\Lambda = \left[E\left\{ \mathbf{h}(\mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1} \right\} : \ E\left\{ \mathbf{h}(\mathbf{X}_i) \mid \mathcal{C}_{i1} \right\} = \mathbf{0} \right].
$$

Here $\mathbf{h}(\mathbf{X}_i)$ is a random function in the Hilbert space $\mathcal{H}$. The corresponding orthogonal complement of Λ, denoted $\Lambda^\perp$ is given by

$$
\Lambda^\perp = \left[\mathbf{g}(Y_i, \overline{\mathbf{W}}_i) : E\left\{ \mathbf{g}(Y_i, \overline{\mathbf{W}}_i) \mid \mathbf{X}_i, \mathcal{C}_{i1} \right\} = \mathbf{0} \right].
$$

We project the score function $\mathbf{S}_\theta(Y_i, \overline{\mathbf{W}}_i)$ onto $\Lambda^\perp$ to obtain the efficient score $\mathbf{S}_{\text{eff}}(Y_i, \overline{\mathbf{W}}_i)$. Any random function $\mathbf{g}(Y_i, \overline{\mathbf{W}}_i)$ in $\Lambda^\perp$ must satisfy that its conditional expectation on $\mathbf{X}_i$ and $\mathcal{C}_{i1}$ is zero, and any random function in Λ can be expressed as $E\left\{\mathbf{a}(\mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1}\right\}$. Thus, if we identify a function $\mathbf{a}(\mathbf{X}_i)$ such that

$$E\left[\mathbf{S}_\theta(Y_i, \overline{\mathbf{W}}_i) - E\left\{\mathbf{a}(\mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1}\right\} \mid \mathbf{X}_i, \mathcal{C}_{i1}\right] = \mathbf{0}, \qquad (3.2)$$

we have found the efficient score

$$\begin{aligned}
\mathbf{S}_{\text{eff}}(Y_i, \overline{\mathbf{W}}_i) &= \mathbf{S}_\theta(Y_i, \overline{\mathbf{W}}_i) - E\left\{\mathbf{a}(\mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1}\right\} \\
&= E\left\{\mathbf{S}_\theta^F(Y_i, \mathbf{X}_i) - \mathbf{a}(\mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1}\right\}.
\end{aligned}$$

The conditional expectation involved in the calculation of the efficient score relies on the unknown distribution $f_{\mathbf{X}_i|\mathcal{C}_{i1}}(\mathbf{x}_i \mid \mathcal{C}_{i1})$. In practice, we propose a candidate distribution $f_{\mathbf{X}_i|\mathcal{C}_{i1}}^*(\mathbf{x}_i \mid \mathcal{C}_{i1})$, and carry out the estimation procedure under $f_{\mathbf{X}_i|\mathcal{C}_{i1}}^*(\mathbf{x}_i \mid \mathcal{C}_{i1})$. We denote the resulting efficient score function $\mathbf{S}_{\text{eff}}^*(Y_i, \overline{\mathbf{W}}_i)$. Because our procedure in obtaining $\mathbf{a}(\mathbf{X}_i)$ from (3.2) calculated under $f_{\mathbf{X}_i|\mathcal{C}_{i1}}^*(\mathbf{x}_i \mid \mathcal{C}_{i1})$ ensures that $E\{\mathbf{S}_{\text{eff}}^*(Y_i, \overline{\mathbf{W}}_i)\} = \mathbf{0}$ regardless of $f_{\mathbf{X}_i|\mathcal{C}_{i1}}^*(\mathbf{x}_i \mid \mathcal{C}_{i1})$ being the true conditional distribution or not, we still have a consistent estimator when the candidate model is not the true model. This is usually referred to as a locally efficient estimator. To solve for $\mathbf{a}(\mathbf{X}_i)$, we use the technique in Tsiatis and Ma (2004). Although the context is different, the integral equation (3.2) shares the mathematical properties of the integral equation there. The estimating equation is

$$\sum_{i=1}^D \mathbf{S}_{\text{eff}}^*(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta}) = \mathbf{0}. \qquad (3.3)$$

Here only the animals captured at least once contribute to the estimation. Numerically, (3.3) is solved through the Newton-Raphson algorithm. And, in practice, especially when D is small, there can be multiple roots. In such a case, practical knowledge is typically used to select a suitable root.

Theorem 1. *Let $\widehat{\boldsymbol{\theta}}$ solve (3.3) and assume Σ is known. Then*

$$\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to N(\mathbf{0}, \mathbf{V})$$

in distribution when $N \to \infty$, where $\mathbf{V} = \mathbf{A}^{-1}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})\left\{\mathbf{A}^{-1}(\boldsymbol{\theta})\right\}^{\mathrm{T}}$,

$$\mathbf{A}(\boldsymbol{\theta}) = E\left\{\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \mathbf{S}_{\text{eff}}^*(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta})\right\},$$

$$\mathbf{B}(\boldsymbol{\theta}) = E\left\{\mathbf{S}_{\text{eff}}^*(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta})\mathbf{S}_{\text{eff}}^{*\mathrm{T}}(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta})\right\}.$$

In addition, when $f_{\mathbf{X}_i|\mathcal{C}_{i1}}^*(\mathbf{x}_i \mid \mathcal{C}_{i1}) = f_{\mathbf{X}_i|\mathcal{C}_{i1}}(\mathbf{x}_i \mid \mathcal{C}_{i1})$, the estimator achieves the optimal estimation variance $\mathbf{V}_{\text{opt}} = \mathbf{B}^{-1}(\boldsymbol{\theta})$.

When Σ is unknown and needs to be estimated, the asymptotic normality results in Theorem 1 still hold, but the variance $\mathbf{V}$ is different. The optimality result no longer holds when Σ is estimated. The proof of Theorem 1 is in the Appendix. For inference, we approximate $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$ with their sample versions evaluated at $\widehat{\boldsymbol{\theta}}$.

Given $\widehat{\boldsymbol{\theta}}$, we can use the procedure proposed in Hwang, Huang, and Wang (2007) to estimate the population size

$$\widehat{N}_C = \sum_{i=1}^{N} \frac{I(\mathcal{C}_{i1})}{\widehat{\mathrm{pr}}(\mathcal{C}_{i1} \mid \Delta_i)}, \tag{3.4}$$

with associated asymptotic variance

$$\mathrm{var}(\widehat{N}_C) = \sum_{i=1}^{N} I(\mathcal{C}_{i1}) \frac{1 - \mathrm{pr}(\mathcal{C}_{i1} \mid \Delta_i)}{\mathrm{pr}^2(\mathcal{C}_{i1} \mid \Delta_i)}$$
$$+ \Big\{ \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \sum_{i=1}^{N} \frac{I(\mathcal{C}_{i1})}{\mathrm{pr}(\mathcal{C}_{i1} \mid \Delta_i)} \Big\} \mathrm{var}(\widehat{\boldsymbol{\theta}}) \Big\{ \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^{N} \frac{I(\mathcal{C}_{i1})}{\mathrm{pr}(\mathcal{C}_{i1} \mid \Delta_i)} \Big\}.$$

Here

$$\Delta_i = \mathbf{W}_{i(1)} + Y_i \boldsymbol{\Sigma} \boldsymbol{\beta}, \mathrm{pr}(\mathcal{C}_{i1} \mid \Delta_i)$$
$$= 1 - \Big[\sum_{y_i=0}^{J} \binom{J}{y_i} \exp\Big\{ y_i(\alpha + \Delta_i^{\mathrm{T}} \boldsymbol{\beta}) - \frac{1}{2} y_i^2 \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\Sigma} \boldsymbol{\beta} \Big\} \Big]^{-1},$$

$\widehat{\mathrm{pr}}(\mathcal{C}_{i1} \mid \Delta_i)$ is $\mathrm{pr}(\mathcal{C}_{i1} \mid \Delta_i)$ calculated under $\widehat{\boldsymbol{\theta}}$, and $\mathrm{var}(\widehat{\boldsymbol{\theta}})$ is given in Theorem 1. We can obtain the approximation $\widehat{\mathrm{var}}(\widehat{N}_C)$ by replacing $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}$ in the expression of $\mathrm{var}(\widehat{N}_C)$.

The first term of $\mathrm{var}(\widehat{N}_C)$ captures the variability of estimating N_C from not observing all the animals, and the second term describes the additional variability due to the estimation of the related parameters. If we reduce the estimation variance of the parameters, we reduce the second term and the overall variance in estimating N. In the simulation section, we illustrate that the semiparametric method achieves this goal.

4. Simulation

We conducted a series of simulation experiments to investigate the performance of the semiparametric methods in comparison with GMM and the conditional score method Hwang, Huang, and Wang (2007). In each simulation experiment, we generated 1,000 data sets.

In the first simulation, the true population size was $N = 500$. We generated the true covariates X_i from Unif$[-3, 3]$, and set the measurement error standard deviation $\sigma_u = 0.6$. We then generated the observations $(Y_{ij}, W_{ij}Y_{ij}), j =$

$1, \cdots, 5$ from the model with the true parameter values $\alpha = 0.2$, $\beta = 1.0$. This yielded an average of 417 first time captures and 335 second time captures, corresponding to high capture probability. To implement different estimators, we replaced Σ with its estimate $\widehat{\Sigma}$, which has bias -0.009 and variance 0.0007. In the semiparametric estimation, we implemented the estimation when the true uniform distribution of X was used and when a false normal distribution was used. For the GMM methods, one used a single $\mathbf{W}_{i(k)}$ and the other used the average $k^{-1} \sum_{l=1}^{k} \mathbf{W}_{i(l)}$, called GMM1 and GMM2 respectively. The results of the various estimators are given in Table 1, where we report the mean and the standard error of the estimators as well as the average of estimated standard errors and the sample coverage of the 95% confidence intervals constructed using the asymptotic results. From Table 1, we can see that all five estimators have small biases in estimating the parameters α, β. Each of the five methods has positive bias for the population size N estimation, but it decreases toward zero as sample size increases.

The GMM methods performed similarly. Although the GMM estimators reduced the estimation standard error for the model parameters α and β, they did not reduce the estimation variability of the population size N. This is likely because the reduction of parameter estimation variability is not enough to mask out the first term of var$(\widehat{N}_C)$, but the semiparametric method yielded a larger variability reduction in parameter estimation that propagated to a more visible variability reduction of the population size estimate.

The second simulation used $\alpha = -1.0$, $\beta = 1.0$. This yielded an average of 298 first time captures and 125 second time captures. All the other settings are the same except that X_i was standard normal and $J = 3$. An estimated $\widehat{\Sigma}$ was used in the estimation procedures, where $\widehat{\Sigma}$ has bias -0.0019 and variance 0.0021. The semiparametric estimation also proposed both true and false distribution for X_i. The results are given in Table 2. From it, we can see that all five estimators have non-substantial biases in estimating the parameters α, β and the population size N. With the exception of GMM estimators, the sample standard error and the average of the estimated standard errors are close to each other, indicating satisfactory performance of the asymptotic results for relatively small N. This is further reflected in the close proximity of the observed 95% coverage to its nominal level. The poor performance of GMM is likely caused by the small number of animals captured more than once. Indeed, in simulations not reported here, when we increased the population size, the performance of the GMM estimators improved. The two GMM estimators performed similarly, with GMM2 having slightly smaller MSE, but both are clearly inferior to the semiparametric estimators.

Table 1. Simulation 1. Performance of five methods based on conditional score (CS), two types of GMM (GMM1, GMM2), in comparison with the semiparametric estimator using normal $f_X(x)$ (Semi-Nor) and using uniform $f_X(x)$ (Semi-Uni). The mean of the estimates (estimate), empirical standard error (emp se), mean square error (mse), average of estimated standard error (est se) and the sample coverage rate of the 95% confidence interval (95% cov) are reported.

		α	β	N
	true	0.2	1.0	500
CS	estimate	0.1986	1.0035	503.12
	emp se	0.0710	0.0607	23.79
	mse	0.0101	0.0074	1163.3
	est se	0.0712	0.0606	23.39
	95% cov	95.1%	95.5%	94.8%
GMM1	estimate	0.1970	1.0091	504.44
	emp se	0.0688	0.0591	23.96
	mse	0.0095	0.0070	1179.7
	est se	0.0690	0.0585	23.40
	95% cov	95.0%	95.8%	95.3%
GMM2	estimate	0.1977	1.0083	504.24
	emp se	0.0702	0.0598	24.06
	mse	0.0099	0.0071	1187.8
	est se	0.0701	0.0590	23.48
	95% cov	94.9%	95.0%	95.3%
Semi-Nor	estimate	0.2006	1.0033	502.14
	emp se	0.0640	0.0537	22.92
	mse	0.0084	0.0059	1056.5
	est se	0.0658	0.0546	22.28
	95% cov	95.4%	95.1%	93.3%
Semi-Uni	estimate	0.2005	1.0031	502.13
	emp se	0.0640	0.0536	22.91
	mse	0.0084	0.0059	1055.7
	est se	0.0658	0.0546	22.27
	95% cov	95.6%	95.0%	93.3%

Summarizing the first two simulation results, we find that the GMM did not improve drastically over the conditional score method in terms of the estimation efficiency. Its finite sample performance also heavily relies on the capture probability and the sample size, in that smaller sample sizes tend to inhibit the gain. Intuitively, the gain of the GMM comes mainly from the appropriate usage of additional measurements. When N is relatively small, there are very small amount of additional measurements available. This not only limits the source of additional information, but also adversely affects how such information is used, because the GMM weighting matrix relies on asymptotic results and is not a

Table 2. Simulation 2. Performance of five methods based on conditional score (CS), two types of GMM (GMM1, GMM2), in comparison with the semiparametric estimator using normal $f_X(x)$ (Semi-Nor) and using uniform $f_X(x)$ (Semi-Uni). The mean of the estimates (estimate), empirical standard error (emp se), mean square error (mse), average of estimated standard error (est se) and the sample coverage rate of the 95% confidence interval (95% cov) are reported.

		α	β	N
	true	-1	1	500
CS	estimate	-1.0229	1.0239	519.87
	emp se	0.1604	0.1829	85.55
	mse	0.0528	0.0651	15928
	est se	0.1612	0.1736	73.29
	95% cov	96.2%	94.2%	94.1%
GMM1	estimate	-1.0686	1.0692	544.22
	emp se	0.1782	0.2082	142.21
	mse	0.0682	0.0868	64989
	est se	0.1676	0.1800	94.34
	95% cov	95.6%	93.4%	96.2%
GMM2	estimate	-1.0673	1.0685	544.37
	emp se	0.1843	0.2107	147.12
	mse	0.0672	0.0827	53500
	est se	0.1663	0.1786	90.21
	95% cov	94.7%	93.5%	95.9%
Semi-Nor	estimate	-1.0101	1.0102	512.27
	emp se	0.1497	0.1571	68.16
	mse	0.0442	0.0494	10149
	est se	0.1465	0.1556	63.85
	95% cov	95.7%	95.0%	93.3%
Semi-Uni	estimate	-1.0103	1.0106	512.40
	emp se	0.1500	0.1577	68.36
	mse	0.0444	0.0497	10220
	est se	0.1468	0.1561	64.00
	95% cov	96.0%	95.0%	93.5%

suitable approximation to the true weights under small number of recaptures. In contrast, the semiparametric methods perform satisfactorily. The efficiency in estimating α, β is improved by 15% and 36% respectively, while that for the population estimation is improved by 58% in the worse capture scenario. The dramatic efficiency gain on the population estimation is caused by the multiplication of $\sum_{i=1}^{N} \partial\{I(\mathcal{C}_{i1})/\mathrm{pr}(\mathcal{C}_{i1} \mid \Delta_i)\}/\partial\boldsymbol{\theta}^{\mathrm{T}}$ in the second term of $\mathrm{var}(\widehat{N}_C)$, which amplifies the magnitude of the change in $\mathrm{var}(\widehat{\boldsymbol{\theta}})$ in this model.

In the third simulation study, we generated the data by mimicking the bird data structure in Section 5. We set $N = 913$ and used $J = 12$ capture occa-

sions. We generated the covariates X_i from a normal with mean $\mu_X = 45.2$ and standard deviation $\sigma_X = 1.0$, with $\sigma_u = 0.8$ used to generate the measurement errors. The observations $(Y_{ij}, W_{ij}Y_{ij}), j = 1, \ldots, J$ were generated from the logistic model with $\alpha = -36.34$ and $\beta = 0.72$. The observed average number of first captures and second captures were respectively, 244 and 49. We conducted the five estimations procedures; the results are in Table 3. For all five estimators, an estimated $\widehat{\Sigma}$ was used, with bias -0.0136 and variance 0.1261. We saw small biases of $\widehat{\boldsymbol{\theta}}$ and the population estimate in all cases. The conditional score and the semi-parametric methods yielded close results between the sample estimation standard error and the empirical one, and between the observed coverage of the 95% confidence intervals and the nominal level, indicating the validity and relevance of the asymptotic results. Once more, the semiparametric methods provided the smallest estimation variability for both parameter and population estimation, with a gain of 59%, 63%, and 277% in terms of estimation efficiency in comparison with the conditional score method. Among the two semiparametric methods, the normal-based procedure was slightly better than the uniform-based procedure, indicating the optimality when the true f_X distribution is used, though the performance difference here was small. This is encouraging considering that the true f_X is not easy to obtain in practice. The GMM estimates of population were not competitive with the semiparametric methods due to the small number of recaptures. We conducted more extensive simulations, including the cases for two covariates, and provide the results in the supplementary material. In all simulations, the semiparametric methods outperformed CS and GMM.

5. Data Example

We implemented the conditional score method, the generalized method of moments, and the semiparametric method on a data set, regarding the bird species *Prinia flaviventris*, collected by the Hong Kong Bird Society from 1991 to 1995. In addition to the capture record, the data set contains the measurements of a bird's wing length, believed to be directly linked with a bird's capture probability and measured with error. We considered the data from 01/31/93 to 04/11/93. In this relatively short time period, the population size change is likely small and we treated it as a closed population. During this time period, 146 distinct birds were captured and measured on $J = 12$ occasions with 168 total captures. Among them, the average wing length was 45.20 and the variance was 1.64.

Under the normal additive measurement error assumption, and taking advantage of the multiple measurements of the wing length of recaptured birds, we formed the difference between the measurements and estimated the variance of the measurement error to be $\widehat{\sigma}_u^2 = 0.626$.

Table 3. Simulation 3. Performance of five methods based on conditional score (CS), two types of GMM (GMM1, GMM2), in comparison with the semiparametric estimator using normal $f_X(x)$ (Semi-Nor) and using uniform $f_X(x)$ (Semi-Uni). The mean of the estimates (estimate), empirical standard error (emp se), mean square error (mse), average of estimated standard error (est se) and the sample coverage rate of the 95% confidence interval (95% cov) are reported.

		α	β	N
	true	-36.34	0.72	913
CS	mean	-37.41	0.74	1031.44
	emp se	10.43	0.2251	527.56
	mse	353.43	0.1651	$1.994e^6$
	est se	10.21	0.2202	381.67
	95% cov	94.7%	94.7%	93.3%
GMM1	mean	-37.94	0.75	1051.96
	emp se	9.7499	0.2118	264.97
	mse	376.34	0.1769	$4.539e^5$
	est se	11.25	0.2427	365.93
	95% cov	97.6%	97.6%	97.7%
GMM2	mean	-37.93	0.75	1048.81
	emp se	9.3641	0.2033	269.47
	mse	252.54	0.1183	$2.833e^5$
	est se	10.53	0.2271	332.50
	95% cov	98.3%	98.3%	98.3%
Semi-Nor	mean	-36.74	0.73	976.73
	emp se	8.2814	0.1787	271.72
	mse	137.01	0.0638	$1.705e^5$
	est se	8.0443	0.1735	245.43
	95% cov	94.2%	94.3%	92.4%
Semi-Uni	mean	-36.91	0.73	982.07
	emp se	8.4262	0.1818	279.99
	mse	141.89	0.0661	$1.798e^5$
	est se	8.2185	0.1773	250.39
	95% cov	94.2%	94.3%	92.6%

The results of parameter and population size estimation based on the five methods are summarized in Table 4, with a normal and a uniform working model for the semiparametric estimation. For the GMM1 method, we incorporated only the first two captures while the maximum total recapture was five. Only four birds were captured three or more times, too small a sample to justify any analysis. Table 4 indicates that the GMM and the semiparametric methods resulted in more estimation variance reduction than the conditional score method, while the improvement from the two semiparametric methods was notable in terms of estimating the population size. The improvement in the semiparametric

Table 4. Estimation and the associated standard error (se) of bird data analysis based on conditional score (CS), generalized method of moments (GMM1) and semiparametric methods using normal (Semi-Nor) and uniform (Semi-Uni) candidate distributions for the wing lengths.

	$\widehat{\alpha}$(se)		$\widehat{\beta}$(se)		$\widehat{N}$(se)	
CS	-40.11	(31.742)	0.80	(0.694)	921.46	(713.767)
GMM1(First 2 captures)	-35.46	(9.274)	0.69	(0.201)	1117.31	(431.082)
Semi-Nor	-37.03	(14.451)	0.73	(0.318)	839.58	(293.993)
Semi-Uni	-32.75	(9.66)	0.64	(0.213)	770.82	(225.332)

estimators reflects the simulation results. But the improvement in GMM is less trustworthy. The inference results for GMM at this population size may not be sufficiently precise, it tends to under-estimate estimation variability, as in the simulation studies.

6. Discussion

We have investigated the issue of using multiple error-prone covariate measurements in the capture-recapture models. Although GMM is a possible way to take account of multiple measurements, its use of the information is somewhat superficial. Our semiparametric approach takes advantage of this information directly in the construction of the estimator. Its effectiveness is reflected in theoretical analysis and in numerical results. Our approach differs from that in Xi et al. (2009). In particular, the setting there is parametric, our is semi-parametric.

When sample size is moderate or small, the asymptotic properties of the semiparametric estimator may not exhibit well. In addition, the superior performance of the semiparametric estimator comes at the cost of more intensive computation.

Throughout, we have worked under a closed population assumption in the capture-recapture framework. Over a short time period, this is reasonable, as in our data example (Hwang and Huang (2003), Hwang, Huang, and Wang (2007) and Xi et al. (2009)). It is certainly of interest to also study population sizes over the five year period that these data were collected. Here open population models, such as the Jolly-Seber open population model (Jolly (1965), Seber (1982) and Seber (1986)), are more appropriate. See also recent developments in open population studies in Schwarz and Arnason (1996) and Pledger, Pollock, and Norris (2003).

Acknowledgement

Appendix

A.1. Derivation of Λ

For the model

$$
f_{Y_i, \overline{\mathbf{W}}_i, \mathbf{X}_i | \mathcal{C}_{i1}} (y_i, \overline{\mathbf{w}}_i, \mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\theta})
$$
$$
= f_{\overline{\mathbf{W}}_i | Y_i, \mathbf{X}_i, \mathcal{C}_{i1}} (\overline{\mathbf{w}}_i \mid y_i, \mathbf{x}_i, \mathcal{C}_{i1}) f_{Y_i | \mathbf{X}_i, \mathcal{C}_{i1}} (y_i \mid \mathbf{x}_i, \mathcal{C}_{i1}; \boldsymbol{\theta}) f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}), \text{ (A.1)}
$$

where $f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1})$ is unknown, the nuisance tangent space is

$$
[\mathbf{h}(\mathbf{X}_i): \ E\left\{ \mathbf{h}(\mathbf{X}_i) \mid \mathcal{C}_{i1} \right\} = \mathbf{0}]. \tag{A.2}
$$

To see this, suppose the true model for $f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1})$ is $f_0(\mathbf{x}_i \mid \mathcal{C}_{i1})$, and let

$$
f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta}) = f_0(\mathbf{x}_i \mid \mathcal{C}_{i1})\{1 + \boldsymbol{\eta}^{\mathrm{T}} \mathbf{h}(\mathbf{X}_i)\},
$$

where $\mathbf{h}(\mathbf{X}_i)$ satisfies (A.2), is a bounded random function, and $\boldsymbol{\eta}_{r \times 1}$ is required to be sufficiently small that $1 + \boldsymbol{\eta}^{\mathrm{T}} \mathbf{h}(\mathbf{X}_i) \geq 0$. A simple calculation yields

$$
\int f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta}) d\mu(\mathbf{x}_i) = \int f_0(\mathbf{x}_i \mid \mathcal{C}_{i1}) d\mu(\mathbf{x}_i) + \int f_0(\mathbf{x}_i \mid \mathcal{C}_{i1}) \boldsymbol{\eta}^{\mathrm{T}} \mathbf{h}(\mathbf{X}_i) d\mu(\mathbf{x}_i)
$$
$$
= 1 + \boldsymbol{\eta}^{\mathrm{T}} E\left\{ \mathbf{h}(\mathbf{X}_i) \mid \mathcal{C}_{i1} \right\}
$$
$$
= 1,
$$

so $f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta})$ is a valid probability density function. When $\boldsymbol{\eta} = \mathbf{0}$, $f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta})$ is $f_0(\mathbf{x}_i \mid \mathcal{C}_{i1})$, the true model. Thus, one has a parametric submodel. Since $\partial f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta})/\partial \boldsymbol{\eta} = \mathbf{h}(\mathbf{X}_i)$, we have shown that any element in the set defined in (A.2) is indeed one element in the nuisance tangent space of model (A.1). On the other hand, for any parametric submodel of (A.1)

$$
f_{Y_i, \overline{\mathbf{W}}_i, \mathbf{X}_i | \mathcal{C}_{i1}} (y_i, \overline{\mathbf{w}}_i, \mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\theta}, \boldsymbol{\eta})
$$
$$
= f_{\overline{\mathbf{W}}_i | Y_i, \mathbf{X}_i, \mathcal{C}_{i1}} (\overline{\mathbf{w}}_i \mid y_i, \mathbf{x}_i, \mathcal{C}_{i1}) f_{Y_i | \mathbf{X}_i, \mathcal{C}_{i1}} (y_i \mid \mathbf{x}_i, \mathcal{C}_{i1}; \boldsymbol{\theta}) f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta}).
$$

Let

$$
\mathbf{S}_{\boldsymbol{\eta}}(\mathbf{x}_i; \boldsymbol{\eta}) = \frac{\partial}{\partial \boldsymbol{\eta}} \log \left\{ f_{Y_i, \overline{\mathbf{W}}_i, \mathbf{X}_i | \mathcal{C}_{i1}} (y_i, \overline{\mathbf{w}}_i, \mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\theta}, \boldsymbol{\eta}) \right\}
$$
$$
= \frac{\partial}{\partial \boldsymbol{\eta}} \log f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta}).
$$

Because $\int f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta}_0) d\mu(\mathbf{x}_i) = 1$ when evaluating at the true value $\boldsymbol{\eta}_0$,

$$
\frac{\partial}{\partial \boldsymbol{\eta}} \int f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta}_0) d\mu(\mathbf{x}_i)
$$
$$
= \int \frac{\partial}{\partial \boldsymbol{\eta}} \log f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta}_0) f_{\mathbf{X}_i | \mathcal{C}_{i1}} (\mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\eta}_0) d\mu(x_i)
$$
$$
= E\{ \mathbf{S}_{\boldsymbol{\eta}}(\mathbf{x}_i; \boldsymbol{\eta}_0) \mid \mathcal{C}_{i1} \} = \mathbf{0}.
$$

Thus, any element belongs to the nuisance tangent space of (A.1) must also belong to the set given in (A.2).

Finally, because the conditional joint distribution of $(Y_i, \overline{\mathbf{W}}_i)$ can be written as the conditional expectation

$$f_{Y_i, \overline{\mathbf{W}}_i | \mathcal{C}_{i1}}(y_i, \overline{\mathbf{w}}_i \mid \mathcal{C}_{i1}) = E\{f_{Y_i, \overline{\mathbf{W}}_i, \mathbf{X}_i | \mathcal{C}_{i1}}(y_i, \overline{\mathbf{w}}_i, \mathbf{x}_i \mid \mathcal{C}_{i1}; \boldsymbol{\theta}) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1}\},$$

the semiparametric nuisance tangent space is

$$\Lambda = \left[E\left\{ \mathbf{h}(\mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1} \right\} : \ E\left\{ \mathbf{h}(\mathbf{X}_i) \mid \mathcal{C}_{i1} \right\} = \mathbf{0} \right].$$

A.2. Derivation of $\Lambda^\perp$

Suppose $\mathbf{g}(Y_i, \overline{\mathbf{W}}_i)$ is an element in $\Lambda^\perp$ and $E\left\{ \mathbf{h}(\mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1} \right\} \in \Lambda$. Since the two spaces are orthogonal, by using the conditional expectations iteratively, we have

$$\begin{aligned} \mathbf{0} &= E\left[\mathbf{g}(Y_i, \overline{\mathbf{W}}_i) E\left\{ \mathbf{h}(\mathbf{X}_i) \mid Y_i, \overline{\mathbf{W}}_i, \mathcal{C}_{i1} \right\} \right] \\ &= E\left[\mathbf{h}(\mathbf{X}_i) E\left\{ \mathbf{g}(Y_i, \overline{\mathbf{W}}_i) \mid \mathbf{X}_i, \mathcal{C}_{i1} \right\} \right]. \end{aligned}$$

This is true for any random function $\mathbf{h}(\mathbf{X}_i)$ in the Hilbert space. Thus, $E\{\mathbf{g}(Y_i, \overline{\mathbf{W}}_i) \mid \mathbf{X}_i, \mathcal{C}_{i1}\} = \mathbf{0}$, and the orthogonal complement of the nuisance tangent space is

$$\Lambda^\perp = \left[\mathbf{g}(Y_i, \overline{\mathbf{W}}_i) : E\left\{ \mathbf{g}(Y_i, \overline{\mathbf{W}}_i) \mid \mathbf{X}_i, \mathcal{C}_{i1} \right\} = \mathbf{0} \right].$$

A.3. Proof of Theorem 1

A standard Taylor expansion of the estimating equation yields

$$\begin{aligned} \mathbf{0} &= N^{-1/2} \sum_{i=1}^{N} \mathbf{S}^*_{\text{eff}}(Y_i, \overline{\mathbf{W}}_i; \widehat{\boldsymbol{\theta}}) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{S}^*_{\text{eff}}(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta}) + \frac{1}{\sqrt{N}} \left\{ \sum_{i=1}^{N} \frac{\partial \mathbf{S}^*_{\text{eff}}(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right\} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1). \end{aligned}$$

This implies

$$\begin{aligned} \sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &= \left\{ -\frac{1}{N} \sum_{i=1}^{N} \frac{\partial \mathbf{S}^*_{\text{eff}}(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right\}^{-1} \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{S}^*_{\text{eff}}(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta}) \right\} + o_p(1) \\ &= -\mathbf{A}(\boldsymbol{\theta}) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{S}^*_{\text{eff}}(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta}) + o_p(1), \end{aligned}$$

and the Central Limit Theorem then yields the asymptotic result in Theorem 1.

When the true distribution model $f_{\mathbf{X}_i | \mathcal{C}_{i1}}$ is used, all asterisks can be eliminated. Using integration by parts and observing that $\mathbf{S}_{\text{eff}}$ is the orthogonal

projection of the score function $\mathbf{S}_{\boldsymbol{\theta}}$ onto $\Lambda^{\perp}$, we have

$$
\begin{aligned}
\mathbf{A}(\boldsymbol{\theta}) &= E\left\{\frac{\partial \mathbf{S}_{\mathrm{eff}}(Y_i, \overline{\mathbf{W}}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\} \\
&= \int \frac{\partial \mathbf{S}_{\mathrm{eff}}(y_i, \overline{\mathbf{w}}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} f_{Y_i, \overline{\mathbf{W}}_i}(y_i, \overline{\mathbf{w}}_i) d\mu(y_i, \overline{\mathbf{w}}_i) \\
&= \mathbf{0} - \int \mathbf{S}_{\mathrm{eff}}(y_i, \overline{\mathbf{w}}_i; \boldsymbol{\theta}) \frac{\partial \log f_{Y_i, \overline{\mathbf{W}}_i}(y_i, \overline{\mathbf{w}}_i)}{\partial \boldsymbol{\theta}^{\mathrm{T}}} f_{Y_i, \overline{\mathbf{W}}_i}(y_i, \overline{\mathbf{w}}_i) d\mu(y_i, \overline{\mathbf{w}}_i) \\
&= -E\left\{\mathbf{S}_{\mathrm{eff}}(y_i, \overline{\mathbf{w}}_i; \boldsymbol{\theta}) \mathbf{S}_{\theta}^{\mathrm{T}}(y_i, \overline{\mathbf{w}}_i; \boldsymbol{\theta})\right\} \\
&= -E\left\{\mathbf{S}_{\mathrm{eff}}(y_i, \overline{\mathbf{w}}_i; \boldsymbol{\theta}) \mathbf{S}_{\mathrm{eff}}^{\mathrm{T}}(y_i, \overline{\mathbf{w}}_i; \boldsymbol{\theta})\right\} \\
&= -\mathbf{B}(\boldsymbol{\theta}).
\end{aligned}
$$

The general expression of $\mathbf{V}$ indicates that the variance is $\mathbf{B}^{-1}(\boldsymbol{\theta})$. Finally, it is the optimal variance because it is the variance of the efficient influence function.

References

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

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* The Johns Hopkins University Press. Baltimore.

Hall, P. and Ma, Y. (2007). Measurement error models with unknown error structure. *J. Roy. Statist. Soc. B.*, **69**, 429-446.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029-1054.

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

Huggins, R. M. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* **47**, 725-732.

Huggins, R. M. (2006). Semiparametric estimation of animal abundance using capture-recapture data from open populations. *Biometrics* **62**, 684-690.

Huggins, R. M. and Hwang, W. H. (2009). A measurement error model for heterogeneous capture probabilities in mark-recapture experiments: An estimating equation approach. *J. Agricultural, Biological, and Environmental Statist.* **15**, 198-208.

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

Hwang, W. H., Huang, S. Y. H. and Wang, C. Y. (2007). Effects of measurement error and conditional score estimation in capture-recapture models. *Statist. Sinica* **17**, 301-316.

Hwang, W. H. and Huggins, R. (2007). Application of semiparametric regression models the analysis of capture-recapture experiments. *Australian and New Zealand Journal of Statistics* **49**, 191-202.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* **52**, 225-247.

Ma, Y. and Tsiatis, A. A. (2006). Closed form semiparametric estimators for measurement error models. *Statist. Sinica* **16**, 183-193.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**, 434-442.

Pledger, S., Pollock, K. H. and Norris, J. L. (2003). Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics* **59**, 786-794.

Pollock, K. H. (2002). The use of auxiliary variables in capture-recapture modelling: An overview. *J. Appl. Statist.* **29**, 85-102.

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

Schwarz, C. J. and Arnason, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics* **52**, 860-873.

Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters.* The Blackburn Press, 2nd ed, London.

Seber, G. A. F. (1986). A review of estimating animal abundance. *Biometrics* **42**, 267-292.

Tsiatis, A. (2006). *Semiparametric Theory and Missing Data.* Springer, New York.

Tsiatis, A. A. and Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika* **91**, 835-848.

Wang, C. Y. (2000). Flexible regression calibration for covariate measurement error with longitudinal surrogate variables. *Statist. Sinica* **10**, 905-921.

Xi, L. Q., Watson, R., Wang, J. P. and Yip, P. S. F. (2009). Estimation in capture-recapture models when covariates are subject to measurement errors and missing data. *Canad. J. Statist.* **37**, 645-658.

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA.

E-mail: kunxu@stat.tamu.edu

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA.

E-mail: ma@stat.tamu.edu