Statistica Sinica **29** (2019), 589-610
doi:https://doi.org/10.5705/ss.202017.0088

CLOSED POPULATION CAPTURE–RECAPTURE MODELS WITH MEASUREMENT ERROR AND MISSING OBSERVATIONS IN COVARIATES

Jakub Stoklosa, Shen-Ming Lee and Wen-Han Hwang

The University of New South Wales, Feng Chia University and National Chung Hsing University

Abstract: In capture–recapture experiments, covariates collected on individuals, such as body weight and length, are often measured imprecisely or are missing at random. Furthermore, the number of recorded covariate measurements collected on each observed individual is usually equal to or less than the individual's capture frequency. Correcting for multiple error-prone covariates is seldom seen in capture–recapture models and even fewer researchers have considered cases where individual's have no measurements at all. In this paper, we develop an unbiased estimating equation using the conditional score within the capture–recapture framework. We then extend this approach to simultaneously account for both measurement error and missing data using two well-known missing data methods: (1) inverse probability weighting; and (2) multiple imputation. These new methods are shown to yield consistent and asymptotically normal estimators, and the two approaches are shown to be asymptotically equivalent. We evaluated these methods on simulated and real capture–recapture data. Our results show improvements in both precision and efficiency when using the proposed methods.

Key words and phrases: Conditional score, differential measurement errors, inverse probability weighting, missing at random, multiple imputation, population size estimation.

1. Introduction

Over the last several decades there has been growing development in enhancing population size estimation in capture–recapture studies through the use of covariates (McCrea and Morgan (2014)). For closed population models, covariates are often used to model capture probabilities in the form of a logistic regression (Huggins (1989); Alho (1990)). These models are commonly referred to as "observed heterogeneity" models since the heterogeneity is modelled via covariates. For models concerning unobserved heterogeneity, see Pledger (2005) and Farcomeni (2016). These models not only help explain variation and heterogeneity in capture probabilities but also reduce bias in the estimation of the population size (Pollock (2002); Hwang and Huggins (2005)). We consider observed heterogeneity models in this study. Ideally, the number of recorded covariate measurements is equal to the individual's capture frequency, although in practice some covariate measurements are not recorded on each capture occasion. Furthermore, covariates collected on individuals may be imprecisely measured. For example, in Section 5.2 we analyse recapture–recapture data collected on Eastern barred bandicoots in which we encounter both imprecise measurements and missing values for observed body weight. As in the general regression context, it is well-known that measurement errors and missing data may yield biased estimation for the regression parameters (Rubin (1987); Little (1992); Carroll et al. (2006)). The same issues may occur when estimating population sizes in capture-recapture models, which is the main focus of this study.

Correcting for measurement error in covariates has been well-established in closed populations capture–recapture studies over the last two decades. When the measurement error variance is constant across individual subjects, a variety of measurement error methods have been developed to address this problem, including: simulation–extrapolation (Gould, Stefanski and Pollock (1999); Stoklosa et al. (2011)), regression calibration (Hwang and Huang (2003)), and conditional score (Hwang, Huang and Wang (2007)). However, the measurement error variance of an observed covariate often depends on the capture frequency which is the response variable for modelling capture probabilities. This is referred to as a *differentiable measurement error* problem that has no general functional method, see Carroll et al. (2006). To overcome this difficulty, Huggins and Hwang (2010) used an approximated estimating equation approach via maximizing a partial likelihood, Xu and Ma (2014) proposed a semiparametric efficient score method, and Xi et al. (2009) developed a parametric likelihood approach that required making a parametric distributional assumption on the true underlying covariate.

Accounting for missing values in the observed covariates in closed capture–recapture studies is more challenging to develop because a "missing observation" is confounded with the fact that the individual was simply not observed on an occasion. However, several likelihood based methods have been developed. Wang (2005) considered a semi-parametric approach for continuous time capture–recapture models and Zwane and van der Heijden (2007) used capture–recapture log-linear models with missing categorical covariates. The methods of Xi et al. (2009) also accounted for missing data in covariates but they require a normality assumption on the true underlying covariate. More recently, Lee, Hwang and Jean (2016) showed that a naïve complete data analysis underestimates the population size in most common situations. To correct for this bias, Lee, Hwang and Jean (2016) proposed several methods that make no distributional assumptions on the missing covariates. Their techniques made use of regression calibration, inverse probability weighting or multiple imputation to handle missing at random data in the covariates.

With the exception of Xi et al. (2009), very few attempts have been made to account for both missing values and imprecise measurements of covariates. In this study, we develop several methods to address these issues. First, we examine the measurement error case, allowing for the number of recorded covariate measurements to be equal to or related to the individual's capture frequency. Our framework is built around the conditional score method (Carroll et al. (2006); Huang, Hwang and Chen (2011)) to develop an unbiased estimating equation. We then extend this approach to simultaneously account for both measurement error and missing data in covariates. Our proposed methods incorporate the aforementioned estimating equation and two typical missing data techniques: (1) inverse probability weighting; and (2) multiple imputation.

In Section 2 we give notation, review the naïve method and discuss the conditional score approach under the measurement error framework. We then discuss the missing data framework and present several existing methods along with proposed methods in Section 3. Simulations and examples are given in Sections 4 and 5, respectively, followed by a discussion in Section 6. Technical results are given in a Web Appendix.

2. Notation and the Measurement Error Framework

Consider a closed population of N individuals, labelled $i = 1, \ldots, N$, where a capture–recapture experiment has been conducted over capture occasions $j = 1, \ldots, \tau$. Let $Y_{ij} = 1$ if the ith individual is caught on the jth occasion and $Y_{ij} = 0$ otherwise. Let $\mathcal{Y}_i = \sum_{j=1}^{\tau} Y_{ij}$ be the capture frequency for the ith individual and D be the number of uniquely capture individuals. Assume that $\mathcal{Y}_i > 0$ for $i = 1, \ldots, D$, and let $\mathcal{C}_i$ denote the event of $\mathcal{Y}_i > 0$ for the ith individual.

For now, suppose that X_i is an observed continuous covariate (such as body weight or head-to-tail length) measured with no error or missing values. Further, let Z_i denote a covariate vector (which often consists of constant values – e.g., an intercept term and gender records) that is always correctly observed. These covariates are assumed to be constant across capture occasions in a closed population capture–recapture model. We consider heterogeneity type models where capture probabilities depend on an individual's covariate. Let $H(u) = \{1 + \exp(-u)\}^{-1}$ be the logistic function such that capture probabilities are written as $P(Y_{ij} = 1|X_i, Z_i) = H(\beta X_i + \gamma^{\mathsf{T}} Z_i)$ where β and γ are the unknown parameters associated with X_i and Z_i, respectively. For simplicity, we write $\boldsymbol{\theta} = (\beta, \gamma^{\mathsf{T}})^{\mathsf{T}}$ and $P_i(\boldsymbol{\theta}) = H(\beta X_i + \gamma^{\mathsf{T}} Z_i)$. We assume here that X_i is a univariate variable but note that this structure can be easily extended to the multiple covariate case.

Suppose the covariate X_i has been measured m_i times where m_i ranges from 0 to $\mathcal{Y}_i$. In capture–recapture experiments, it is common for $m_i = \mathcal{Y}_i$ since covariates can only be measured on each capture event. However, for various reasons, it is also plausible that no measurements have been collected on any capture event even if the individual has been observed. An extreme case is $m_i = 0$ (but $\mathcal{Y}_i > 0$) which is equivalent to X_i being a missing value. In this section we assume that $m_i > 0$ for $i = 1, \ldots, D$, although we revisit and address the missing data problem in Section 3.

Let W_{ik} denote the kth observed error-contaminated measurements for X_i. When $m_i > 0$, we assume the classical measurement error structure (Carroll et al. (2006)) $W_{ik} = X_i + \varepsilon_{ik}$ for $k = 1, 2, \ldots, m_i$ where ε_{ik} denotes the measurement error, with $\varepsilon_{ik} \sim \mathcal{N}(0, \sigma_u^2)$ independent of all other variables in the model. For a positive m_i, we denote $\overline{W}_i$ as the average of W_{ik} for $k = 1, \ldots, m_i$. In practice, $\overline{W}_i$ is viewed as a surrogate for X_i. Since σ_u^2 is usually unknown in practice, we can obtain as estimate of it using a pooled sample variance estimator: $\hat{\sigma}_u^2 = \sum_{i:m_i>1} \sum_{j=1}^{m_i} (W_{ij} - \overline{W}_i)^2 / \{\sum_{i:m_i>1} (m_i - 1)\}$.

2.1. The naïve method

If measurement error is present in covariates but unaccounted for, the naïve method solves the estimating equation

$$U_n(\boldsymbol{\theta}) = \sum_{i=1}^{D} \Psi_i(\boldsymbol{\theta}) = 0, \tag{2.1}$$

where $\Psi_i(\boldsymbol{\theta}) = \big(\overline{W}_i, Z_i^{\mathsf{T}}\big)^{\mathsf{T}} \big\{\mathcal{Y}_i - \tau H\big(\beta \overline{W}_i + \gamma^{\mathsf{T}} Z_i\big) / P_i^*(\boldsymbol{\theta})\big\}$ with $P_i^*(\boldsymbol{\theta}) = 1 - \big\{1 - H\big(\beta \overline{W}_i + \gamma^{\mathsf{T}} Z_i\big)\big\}^{\tau}$. Given $\sigma_u^2 = 0$, we have $P_i^*(\boldsymbol{\theta}) = P(\mathcal{C}_i | X_i, Z_i)$ and (2.1) is the score function of the distribution $\mathcal{Y}_i$ conditional on $\mathcal{C}_i$. Let $\widehat{\boldsymbol{\theta}}_n$ denote the solution of (2.1). A Horvitz–Thompson type estimator $\widehat{N}_n = \sum_{i=1}^{D} 1/P_i^*\big(\hat{\boldsymbol{\theta}}_n\big)$ is used to estimate the population size. This analysis is referred to as the Huggins–Alho approach for closed populations when there are no measurement errors. Both $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{N}_n$ are biased if covariates are contaminated with measurement errors, and we demonstrate this in our simulations.

2.2. Conditional score estimation

In the context of measurement error analysis, a functional method treats the unknown X_i as parameters for all i. Under this setting, the number of model parameters significantly increases as the sample size grows. To accommodate for the large number of parameters, Stefanski and Carroll (1987) developed conditional score estimation where a novel surrogate for X_i is used rather than $\overline{W}_i$.

First, we take $\Delta_{ij} = \mathcal{Y}_i\beta\sigma_u^2 + W_{ij}$, for $j = 1, \ldots, m_i$, and let $\bar{\Delta}_i = \mathcal{Y}_i\beta\sigma_u^2/m_i + \overline{W}_i$. For now, suppose that each m_i is a non-random constant or independent of $(\mathcal{Y}_i, X_i, Z_i)$, for example, $m_i = 1$ for all $i \le D$. Following Stefanski and Carroll (1987), we treat each $\bar{\Delta}_i$ as observed variables. Huang, Hwang and Chen (2011) showed that for each $i \le D$ and $k = 1, \ldots, \tau$ we have

$$P(\mathcal{Y}_i = k|X_i, \bar{\Delta}_i, Z_i, \mathcal{C}_i) \propto \binom{\tau}{k} \exp\left(k\big(\beta\bar{\Delta}_i + \gamma^{\mathsf{T}} Z_i\big) - \frac{1}{2m_i}k^2\beta^2\sigma_u^2\right). \quad (2.2)$$

Importantly, the right-hand side of (2.2) does not involve X_i, so that the distribution is identical to $P(\mathcal{Y}_i = k|\bar{\Delta}_i, Z_i, \mathcal{C}_i)$ which allows us to calculate the conditional expectation $\mathrm{E}(\mathcal{Y}_i|\bar{\Delta}_i, Z_i, \mathcal{C}_i)$. In other words, $\bar{\Delta}_i$ is considered as a surrogate for X_i.

Estimates of $\boldsymbol{\theta}$ can be obtained by solving the estimating equation

$$\sum_{i=1}^{D} \big(\bar{\Delta}_i, Z_i^{\mathsf{T}}\big)^{\mathsf{T}} \big\{\mathcal{Y}_i - \mathrm{E}(\mathcal{Y}_i|\bar{\Delta}_i, Z_i, \mathcal{C}_i)\big\} = 0. \quad (2.3)$$

This approach is referred as the *naïve conditional score* (NCS) estimation (Huang, Hwang and Chen (2011)), since each m_i is, in general, related to $\mathcal{Y}_i$. To clarify its inappropriateness, consider $m_i = \mathcal{Y}_i$ so that $\bar{\Delta}_i = \beta\sigma_u^2 + \overline{W}_i$. This is just a translation of $\overline{W}_i$ and cannot serve as a valid surrogate for X_i. Particularly, the usual "surrogate assumption" (Carroll et al. (2006)) does not hold, since the response variable $\mathcal{Y}_i$ and the surrogate variable $\overline{W}_i$ are not independent given the condition of X_i. As a result, the NCS method generally yields biased estimates as each m_i is related to the individual's capture frequency.

To account for variation in measurement error with the capture frequency, we can use Δ_{i1} in place of $\bar{\Delta}_i$ at (2.3). This is called the CS1 method. As shown in Hwang, Huang and Wang (2007), the CS1 method is consistent for estimating $\boldsymbol{\theta}$. Nevertheless, CS1 uses only one covariate value W_{i1} and ignores the subsequent measurements, hence some efficiency is lost. To improve this, Huang, Hwang and Chen (2011) proposed an error augmentation CS method that is equivalent to a Rao–Blackwellized estimating function for CS1. Although this method works well, the Rao–Blackwellized procedure requires Monte-Carlo simulation. Furthermore, Huang, Hwang and Chen (2011) did not consider population size estimation.

We propose an alternative estimating function that fully utilizes all measurements of W_{ij} and does not require generating pseudo random variables. Consider the estimating equation

$$U_c(\boldsymbol{\theta}) = \sum_{i=1}^{D} \Phi_i(\boldsymbol{\theta}) = 0, \tag{2.4}$$

where $\Phi_i(\boldsymbol{\theta}) = (1/m_i) \sum_{j=1}^{m_i} \Phi_{ij}(\boldsymbol{\theta})$ with $\Phi_{ij}(\boldsymbol{\theta}) = (\Delta_{ij}, Z_i^{\mathsf{T}})^{\mathsf{T}} \{\mathcal{Y}_i - \mathrm{E}(\mathcal{Y}_i | \Delta_{ij}, Z_i, \mathcal{C}_i)\}$. It is straightforward to show that $\Phi_i(\boldsymbol{\theta})$ is zero unbiased by applying the double expectation law

$$\mathrm{E}\{\Phi_i(\boldsymbol{\theta})\} = \mathrm{E}[\mathrm{E}\{\Phi_i(\boldsymbol{\theta}) | m_i, Z_i, \mathcal{C}_i, \Delta_{i1}, \ldots, \Delta_{i m_i}\}] = \mathrm{E}[\mathrm{E}\{\Phi_{i1}(\boldsymbol{\theta}) | \Delta_{i1}, Z_i, \mathcal{C}_i\}] = 0.$$

As a result, the solution to (2.4), denoted by $\hat{\boldsymbol{\theta}}_c$, is a consistent estimator for $\boldsymbol{\theta}$.

Remark 1. Xu and Ma (2014) considered a modified CS1 method that similarly aims to use all information contained in the observed measurements. As shown in simulations conducted in Xu and Ma (2014), this approach does not improve performance, while our CS method outperformed CS1 in almost all of our simulations. The estimating function of the CS method used in Xu and Ma (2014) differs slightly from our approach.

In Theorem 1 we establish large sample properties for $\hat{\boldsymbol{\theta}}_c$. Let $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathsf{T}}$ for a vector $\boldsymbol{a}$ and let $A^{-\mathsf{T}}$ be the transpose for A^{-1}. Let $M_c(\boldsymbol{\theta}) = \mathrm{E}\left\{\mathrm{I}(\mathcal{C}_i)\Phi_i(\boldsymbol{\theta})^{\otimes 2}\right\}$, and $G_c(\boldsymbol{\theta}) = \mathrm{E}\left[-\mathrm{I}(\mathcal{C}_i)\{\partial \Phi_i(\boldsymbol{\theta})/\partial \boldsymbol{\theta}\}\right]$. Generally, when investigating large sample properties for capture–recapture models, the capture occasion τ is considered fixed. Here we also assume fixed τ, but the variability of estimators increases if τ is decreased.

Theorem 1. *Under regularity conditions* A1–A3 *(see Web Appendix S1),* $\hat{\boldsymbol{\theta}}_c$ *is a consistent estimator as* $N \to \infty$. *Moreover,* $\sqrt{N}\big(\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}\big)$ *converges in distribution to* $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_c)$ *where* $\boldsymbol{\Sigma}_c = G_c^{-1}(\boldsymbol{\theta}) M_c(\boldsymbol{\theta}) G_c^{-\mathsf{T}}(\boldsymbol{\theta})$.

To estimate the variance of $\hat{\boldsymbol{\theta}}_c$, we use the sandwich estimator

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_c) = \left\{\sum_{i=1}^{D} \frac{\partial}{\partial \boldsymbol{\theta}} \Phi_i(\boldsymbol{\theta})\right\}^{-1} \left\{\sum_{i=1}^{D} \Phi_i(\boldsymbol{\theta}) \Phi_i^{\mathsf{T}}(\boldsymbol{\theta})\right\} \left\{\sum_{i=1}^{D} \frac{\partial}{\partial \boldsymbol{\theta}} \Phi_i(\boldsymbol{\theta})\right\}^{-\mathsf{T}},$$

where $\boldsymbol{\theta}$ is evaluated at $\hat{\boldsymbol{\theta}}_c$. In a simulation study, not reported here, the proposed CS method was shown to perform equally well compared to the error augmentation CS method of Huang, Hwang and Chen (2011).

To estimate the population size, Δ_{ij} serves as a surrogate for X_i so that a Horvitz–Thompson type estimator can be constructed based on the conditional distribution of $\mathcal{Y}_i$ given (Δ_{ij}, Z_i). It follows that $P(\mathcal{C}_i|\Delta_{ij}, Z_i) = C(\Delta_{ij}, Z_i)/\{1 + C(\Delta_{ij}, Z_i)\}$ where $C(\Delta_{ij}, Z_i) = \sum_{k=1}^{\tau} \binom{\tau}{k} \exp\left(k\left(\beta\Delta_{ij} + \gamma^{\mathsf{T}} Z_i\right) - k^2\beta^2\sigma_u^2/2\right)$. For $j = 1, \ldots, m_i$, we propose the population size estimator

$$\hat{N}_c = \sum_{i=1}^{D} \frac{1}{\overline{P}^*_{i\Delta}(\hat{\boldsymbol{\theta}}_c)}, \tag{2.5}$$

where

$$\overline{P}^*_{i\Delta}(\boldsymbol{\theta})^{-1} = \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{1}{P(\mathcal{C}_i|\Delta_{ij}, Z_i)} = 1 + \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{1}{C(\Delta_{ij}, Z_i)}.$$

Let $\hat{N}_c(\boldsymbol{\theta}) = \sum_{i=1}^{N} \mathrm{I}(\mathcal{C}_i)/\overline{P}^*_{i\Delta}(\boldsymbol{\theta})$, and write $\hat{N}_c = \hat{N}_c(\boldsymbol{\theta}) + \left\{\hat{N}_c\left(\hat{\boldsymbol{\theta}}_c\right) - \hat{N}_c(\boldsymbol{\theta})\right\}$. Then, we have $\mathrm{Var}\left\{\hat{N}_c(\boldsymbol{\theta})\right\} \approx \sum_{i=1}^{D} \left\{1 - \overline{P}^*_{i\Delta}(\boldsymbol{\theta})\right\} / \left\{\overline{P}^*_{i\Delta}(\boldsymbol{\theta})^2\right\}$. A further calculation shows that the covariance of $\hat{N}_c(\hat{\boldsymbol{\theta}}_c) - \hat{N}_c(\boldsymbol{\theta})$ and $\hat{N}_c(\boldsymbol{\theta})$ is negligible, so only the remaining variance terms are required for calculation. Consequently, we estimate the asymptotic variance of $\hat{N}_c$ by

$$\widehat{\mathrm{Var}}(\hat{N}_c) = \sum_{i=1}^{D} \frac{1 - \overline{P}^*_{i\Delta}(\boldsymbol{\theta})}{\overline{P}^*_{i\Delta}(\boldsymbol{\theta})^2} + \left(\frac{\partial \widehat{N}_c}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}} \widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_c) \left(\frac{\partial \widehat{N}_c}{\partial \boldsymbol{\theta}}\right),$$

where $\boldsymbol{\theta}$ is evaluated at $\hat{\boldsymbol{\theta}}_c$. Let $H_c(\boldsymbol{\theta}) = \mathrm{E}\left\{\dfrac{\partial}{\partial \boldsymbol{\theta}} \dfrac{\mathrm{I}(\mathcal{C}_i)}{\overline{P}^*_{i\Delta}(\boldsymbol{\theta})}\right\}$.

Theorem 2. *Under regularity conditions* A1–A3 *(see Web Appendix S1),* $\hat{N}_c/N$ *converges to one in probability as* $N \to \infty$. *Moreover, the limiting distribution of* $N^{-1/2}\left(\hat{N}_c - N\right)$ *is* $\mathcal{N}(\mathbf{0}, \boldsymbol{\nu}_c)$ *where* $\boldsymbol{\nu}_c$ *is the variance of* $\mathrm{I}(\mathcal{C}_i)/\bar{P}^*_{i\Delta}(\boldsymbol{\theta}) + H_c(\boldsymbol{\theta}) G_c^{-1}(\boldsymbol{\theta}) \mathrm{I}(\mathcal{C}_i) \Phi_i(\boldsymbol{\theta})$.

3. Missing Data Framework

In addition to imprecise measurements with error, suppose that some of these covariates are now missing. Let $\delta_i = 1$ be the indicator that covariate $\overline{W}_i$ is observed and 0 if it is missing. Under these settings, a naïve complete case method ignores the data for individuals with $\delta_i = 0$ and solves the estimating equation

$$U_{ncc}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) = \sum_{i=1}^{D} \delta_i \Psi_i(\boldsymbol{\theta}) = 0, \tag{3.1}$$

where $\Psi_i()$ is the function at (2.1). The solution to (3.1) is denoted as $\hat{\boldsymbol{\theta}}_{ncc}$ which we refer to as the *naïve complete case* estimator for $\boldsymbol{\theta}$. The corresponding naïve complete case population size estimator is $\hat{N}_{ncc} = \left(D/D^{\delta}\right) \sum_{i=1}^{D} \delta_i / P_i^*\left(\hat{\boldsymbol{\theta}}_{ncc}\right)$, where $D^{\delta} = \sum_{i=1}^{D} \delta_i$ is the number of captured individuals without missing covariates. Lee, Hwang and Jean (2016) showed that $\hat{N}_{ncc}$ generally underestimates N when $\sigma_u^2 = 0$. The bias is even worse when the measurement error is present.

Assume that covariates are missing at random, such that $P(\delta_i = 1 | \mathcal{Y}_i, X_i, Z_i) = \pi(\mathcal{Y}_i, Z_i)$. We take $\pi_i = \pi(\mathcal{Y}_i, Z_i)$, as the *selection probability* for δ_i. In practice, selection probabilities π_i can be estimated nonparametrically or parametrically. When Z_i is categorical, we can use the empirical probability $\hat{\pi}_i$, which is the percentage of $\delta_\ell = 1$ with $(\mathcal{Y}_\ell, Z_\ell) = (\mathcal{Y}_i, Z_i)$ for all $\ell \leq D$. When Z_i contains continuous variables, π_i can be estimated by kernel smoothing. If Z_i is of high dimension, it is more suitable to seek a binary regression model with the response δ_i and covariates $\mathcal{Y}_i$ and Z_i (Seaman and White (2013)).

3.1. Naïve inverse probability weighting estimation

The naïve *inverse probability weighting* (IPW) approach accounts for missing covariate values but ignores measurement error. These models were developed in Lee, Hwang and Jean (2016). The estimating equation is

$$U_{nw}(\boldsymbol{\theta}, \hat{\boldsymbol{\pi}}) = \sum_{i=1}^{D} \frac{\delta_i}{\hat{\pi}_i} \Psi_i(\boldsymbol{\theta}) = 0, \tag{3.2}$$

where $\hat{\boldsymbol{\pi}}$ denotes the set collection of $\hat{\pi}_i$. The solution to (3.2) is denoted by $\hat{\boldsymbol{\theta}}_{nw}$, which is a naïve IPW estimator for $\boldsymbol{\theta}$ since it does not account for the effects of measurement error. The naïve IPW estimator for the population size is $\hat{N}_{nw} = \sum_{i=1}^{D} \delta_i / \left\{\hat{\pi}_i P_i^*\left(\hat{\boldsymbol{\theta}}_{nw}\right)\right\}$. If the measurement error variance $\sigma_u^2 = 0$, $\hat{\boldsymbol{\theta}}_{nw}$ and $\hat{N}_{nw}$ are consistent and asymptotically normal, see Lee, Hwang and Jean (2016). These asymptotic properties are not valid if $\sigma_u^2 > 0$.

3.2. Inverse probability weighting with conditional score estimation

When measurement error is present in the covariates, the estimating function $\sum_{i=1}^{D} \delta_i \Psi_i(\boldsymbol{\theta}) \ / \pi_i$ is not zero unbiased, and $\hat{\boldsymbol{\theta}}_{nw}$ and $\hat{N}_{nw}$ do not preserve consistency. We use the conditional score estimating function $\Phi_i(\boldsymbol{\theta})$ in (2.4) to substitute for $\Psi_i(\boldsymbol{\theta})$ in (3.2), and call this method the *inverse probability weighting conditional score* (IPWCS) as it accounts for measurement error and missing covariates. Accordingly, it solves the estimating equation

$$U_{wc}(\boldsymbol{\theta},\hat{\boldsymbol{\pi}})=\sum_{i=1}^{D}\frac{\delta_i}{\hat{\pi}_i}\Phi_i(\boldsymbol{\theta})=0, \tag{3.3}$$

where $\Phi_i(\boldsymbol{\theta})$ is given in (2.4). Using the double expectation law, we find

$$\mathrm{E}\left\{\frac{\delta_i}{\pi_i}\Phi_i(\boldsymbol{\theta})\right\}=\mathrm{E}\left[\mathrm{E}\left\{\frac{\delta_i\Phi_i(\boldsymbol{\theta})}{\pi_i}\middle| m_i,\mathcal{Y}_i,Z_i,\Delta_{i1},\ldots,\Delta_{im_i}\right\}\right]=\mathrm{E}\{\Phi_i(\boldsymbol{\theta})\},$$

so the estimating function in (3.3) is zero unbiased if $\hat{\pi}_i$ is substituted by π_i. To estimate the population size we use

$$\hat{N}_{wc}=\sum_{i=1}^{D}\frac{\delta_i}{\hat{\pi}_i}\frac{1}{\overline{P}^*_{i\Delta}\big(\hat{\boldsymbol{\theta}}_{wc}\big)},$$

where $\hat{\boldsymbol{\theta}}_{wc}$ is the solution of (3.3) and $\overline{P}^*_{i\Delta}(\boldsymbol{\theta})$ is given in (2.5).

Let $\Phi^*_i(\boldsymbol{\theta})=\mathrm{E}\{\Phi_i(\boldsymbol{\theta})|\mathcal{Y}_i,Z_i\}$, $g^*_i(\boldsymbol{\theta})=\dfrac{\delta_i}{\pi_i}\Phi_i(\boldsymbol{\theta})-(\delta_i-\pi_i)/\pi_i\Phi^*_i(\boldsymbol{\theta})$, and M_{wc} $(\boldsymbol{\theta},\boldsymbol{\pi})=\mathrm{E}\big\{\mathrm{I}(\mathcal{C}_i)g^*_i(\boldsymbol{\theta})^{\otimes 2}\big\}$.

Theorem 3. *Under the regularity conditions* A1–A2 *and* B1–B3*, $\hat{\boldsymbol{\theta}}_{wc}$ is a consistent estimator as $N\to\infty$. Moreover, $\sqrt{N}\big(\hat{\boldsymbol{\theta}}_{wc}-\boldsymbol{\theta}\big)$ converges in distribution to $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{wc})$, where $\boldsymbol{\Sigma}_{wc}=G_c^{-1}(\boldsymbol{\theta})M_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})G_c^{-\mathsf{T}}(\boldsymbol{\theta})$.*

Theorem 4. *Under the regularity conditions* A1–A2 *and* B1–B3*, $\hat{N}_{wc}/N$ converges to one in probability as $N\to\infty$. If $\kappa^*_i(\boldsymbol{\theta})$ is the expectation of $\mathrm{I}(\mathcal{C}_i)/\overline{P}^*_{i\Delta}(\boldsymbol{\theta})$ conditional on $(\mathcal{Y}_i,Z_i)$, the limiting distribution of $N^{-1/2}\big(\hat{N}_{wc}-N\big)$ is $\mathcal{N}(\mathbf{0},\boldsymbol{\nu}_{wc})$, where $\boldsymbol{\nu}_{wc}$ is the variance of*

$$\mathrm{I}(\mathcal{C}_i)\left\{\frac{\delta_i}{\pi_i}\frac{1}{\overline{P}^*_{i\Delta}(\boldsymbol{\theta})}+H_c(\boldsymbol{\theta})G_c^{-1}(\boldsymbol{\theta})g^*_i(\boldsymbol{\theta})-\frac{\delta_i-\pi_i}{\pi_i}\kappa^*_i(\boldsymbol{\theta})\right\}.$$

Let $\tilde{\Phi}_i(\boldsymbol{\theta})$ be the average of $\Phi_\ell(\boldsymbol{\theta})$ with $(\mathcal{Y}_\ell,Z_\ell)=(\mathcal{Y}_i,Z_i)$ for all $\ell\le D$, and

$$\tilde{g}_i(\boldsymbol{\theta},\boldsymbol{\pi})=\frac{\delta_i}{\pi_i}\Phi_i(\boldsymbol{\theta})-\frac{\delta_i-\pi_i}{\pi_i}\tilde{\Phi}_i(\boldsymbol{\theta}),$$

for $i=1,\ldots,D$. Further, let $G_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})=-\partial U_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})/\partial\boldsymbol{\theta}$ and $M_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})=\sum_{i=1}^{D}\tilde{g}_i(\boldsymbol{\theta},\boldsymbol{\pi})\tilde{g}_i\big(\boldsymbol{\theta},\boldsymbol{\pi}\big)^{\mathsf{T}}$. According to Theorem 3, the variance estimator for $\hat{\boldsymbol{\theta}}_{wc}$ is

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_{wc})=\widehat{G}_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})^{-1}\widehat{M}_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})\widehat{G}_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})^{-\mathsf{T}},$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are evaluated at $\hat{\boldsymbol{\theta}}_{wc}$ and $\hat{\boldsymbol{\pi}}$, respectively. Moreover, to estimate the variance of $\hat{N}_{wc}$, let

$$\hat{\kappa}_i^*(\boldsymbol{\theta}) = \frac{\sum_{\ell=1}^D \delta_\ell \mathrm{I}(\mathcal{Y}_\ell = \mathcal{Y}_i, Z_\ell = Z_i)/\{\pi_\ell \overline{P}_{\ell\Delta}^*(\boldsymbol{\theta})\}}{\sum_{k=1}^D \delta_k \mathrm{I}(\mathcal{Y}_k = \mathcal{Y}_i, Z_k = Z_i)/\pi_k}$$

and $\widehat{A}(\boldsymbol{\theta},\boldsymbol{\pi}) = \{\partial \widehat{N}_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})/\partial\boldsymbol{\theta}\}^{\mathsf{T}} G_{wc}(\boldsymbol{\theta},\boldsymbol{\pi})^{-\mathsf{T}}$. A variance estimator of $\hat{N}_{wc}$ is

$$\sum_{i=1}^D \left[\frac{\delta_i}{\pi_i}\left\{\frac{1}{\overline{P}_{i\Delta}^*(\boldsymbol{\theta})} - \widehat{A}(\boldsymbol{\theta},\boldsymbol{\pi})\Phi_i(\boldsymbol{\theta})\right\} - \frac{\delta_i - \pi_i}{\pi_i}\left\{\hat{\kappa}^*(\boldsymbol{\theta}) + \widehat{A}(\boldsymbol{\theta},\boldsymbol{\pi})\tilde{\Phi}_i(\boldsymbol{\theta})\right\}\right]^2 - \hat{N}_{wc},$$

where again $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are evaluated at $\hat{\boldsymbol{\theta}}_{wc}$ and $\hat{\boldsymbol{\pi}}$, respectively.

3.3. Multiple imputation with conditional score estimation

We develop another approach to handling measurement error and missing data via multiple imputation (Rubin (1987)). When Z_i is a categorical variable, we consider the empirical distributions

$$\hat{F}_m(m|\mathcal{Y}_i, Z_i) = \sum_{\ell=1}^D \frac{\delta_\ell \mathrm{I}(\mathcal{Y}_\ell = \mathcal{Y}_i, Z_\ell = Z_i)}{\sum_{k=1}^D \delta_k \mathrm{I}(\mathcal{Y}_k = \mathcal{Y}_i, Z_k = Z_i)} \mathrm{I}(m_\ell \le m),$$

$$\hat{F}_w(w|\mathcal{Y}_i, Z_i) = \sum_{\ell=1}^D \sum_{j=1}^{m_\ell} \frac{\delta_\ell \mathrm{I}(\mathcal{Y}_\ell = \mathcal{Y}_i, Z_\ell = Z_i)}{\sum_{k=1}^D m_k \delta_k \mathrm{I}(\mathcal{Y}_k = \mathcal{Y}_i, Z_k = Z_i)} \mathrm{I}(W_{\ell j} \le w).$$

When covariate values $\overline{W}_i$ are missing, we impute m_i and W_{ij} by generating random observations from the empirical distributions $\hat{F}_m$ and $\hat{F}_w$. This imputation procedure is then replicated a fixed number of times, M. The imputed values are used to construct an estimating equation and a population size estimator similar to (2.4) and $\hat{N}_c$, respectively. We summarize the fitting procedure in the algorithm below.

If Z_i consists of continuous variables, we can estimate the conditional distributions $\hat{F}_m(m|\mathcal{Y}_i, Z_i)$ and $\hat{F}_w(w|\mathcal{Y}_i, Z_i)$ by using kernel smoothing techniques. Alternatively, a parametric distribution assumption (Wang and Robins (1998)) can be considered, especially when Z_i consists of many variables.

We propose the variance estimators of $\hat{\boldsymbol{\theta}}_{mc}$ and $\hat{N}_{mc}$. Let $\breve{g}_{vi}(\boldsymbol{\theta}) = \delta_i \Phi_i(\boldsymbol{\theta}) + (1-\delta_i)\Phi_{i,v}^{\dagger}(\boldsymbol{\theta})$, $\breve{g}_{v\cdot}(\boldsymbol{\theta}) = \sum_{i=1}^D \breve{g}_{vi}(\boldsymbol{\theta})$, so $U_{mc}(\boldsymbol{\theta}) = \sum_{v=1}^M \breve{g}_{v\cdot}(\boldsymbol{\theta})/M$. We estimate $\mathrm{Var}(\hat{\boldsymbol{\theta}}_{mc})$ using

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_{mc}) = G_{mc}(\boldsymbol{\theta})^{-1}\left\{\frac{1}{M}\sum_{v=1}^M\sum_{i=1}^D \breve{g}_{vi}(\boldsymbol{\theta})\breve{g}_{vi}(\boldsymbol{\theta})^{\mathsf{T}} + \left(1+\frac{1}{M}\right)\frac{\sum_{v=1}^M \breve{g}_{v\cdot}(\boldsymbol{\theta})\breve{g}_{v\cdot}(\boldsymbol{\theta})^{\mathsf{T}}}{M-1}\right\}G_{mc}(\boldsymbol{\theta})^{-\mathsf{T}},$$

where $\boldsymbol{\theta}$ is evaluated at $\hat{\boldsymbol{\theta}}_{mc}$, and $G_{mc}(\boldsymbol{\theta})$ is the gradient of $U_{mc}(\boldsymbol{\theta})$.

Algorithm: Multiple imputation with conditional score (MICS) estimation

{**Step 1 (Data imputation):**} First, generate $m_{i,v}$ from the empirical distribution $\hat{F}_m(m|\mathcal{Y}_i, Z_i)$ and $W^{\dagger}_{ij,v}$ from $\hat{F}_w(w|\mathcal{Y}_i, Z_i)$ for $v = 1, \ldots, M$ and $j = 1, \ldots, m_{i,v}$. For each missing value of $\delta_i = 0$ and $i \le D$, let $\tilde{\Phi}^{\dagger}_i(\boldsymbol{\theta}) = (1/M)\sum_{v=1}^{M} \Phi^{\dagger}_{i,v}(\boldsymbol{\theta})$, where

$$\Phi^{\dagger}_{i,v}(\boldsymbol{\theta}) = \frac{1}{m_{i,v}} \sum_{j=1}^{m_{i,v}} (\Delta^{\dagger}_{ij,v}, Z_i^{\mathsf{T}})^{\mathsf{T}} \left\{ \mathcal{Y}_i - \mathrm{E}(\mathcal{Y}_i|\Delta^{\dagger}_{ij,v}, Z_i, \mathcal{C}_i) \right\}$$

and $\Delta^{\dagger}_{ij,v} = \beta\sigma_u^2\mathcal{Y}_i + W^{\dagger}_{ij,v}$.

{**Step 2:**} Solve the estimating equation

$$U_{mc}(\boldsymbol{\theta}) = \sum_{i=1}^{D} \left\{ \delta_i\Phi_i(\boldsymbol{\theta}) + (1-\delta_i)\tilde{\Phi}^{\dagger}_i(\boldsymbol{\theta}) \right\} = 0$$

to find $\hat{\boldsymbol{\theta}}_{mc}$.

{**Step 3:**} For each missing value of $\delta_i = 0$ and $i \le D$, let $\tilde{P}^{\dagger}_{i\Delta}(\boldsymbol{\theta})$ be the harmonic average of $P(\mathcal{C}_i|\Delta^{\dagger}_{ij,v}, Z_i)$ for all $j = 1, \ldots, \mathcal{Y}_i$ and $v = 1, \ldots, M$. The MICS population size estimator is

$$\hat{N}_{mc} = \sum_{i=1}^{D} \left\{ \delta_i \frac{1}{\overline{P}^{*}_{i\Delta}(\hat{\boldsymbol{\theta}}_{mc})} + (1-\delta_i)\frac{1}{\tilde{P}^{\dagger}_{i\Delta}(\hat{\boldsymbol{\theta}}_{mc})} \right\}.$$

For the variance of $\hat{N}_{mc}$, take

$$\hat{N}_v(\boldsymbol{\theta}) = \sum_{i=1}^{D} \left\{ \delta_i \frac{1}{\overline{P}^{*}_{i\Delta}(\boldsymbol{\theta})} + (1-\delta_i)\frac{1}{\tilde{P}^{\dagger}_{vi}(\boldsymbol{\theta})} \right\},$$

where $\tilde{P}^{\dagger}_{vi}(\boldsymbol{\theta})$ is the harmonic average of $P(\mathcal{C}_i|\Delta^{\dagger}_{ij,v}, Z_i)$ for $j = 1, \ldots, \mathcal{Y}_i$. A variance estimator for $\hat{N}_{mc}$ is given by

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{N}_{mc}) = &\sum_{i=1}^{D} \left[\frac{\delta_i\{1-\overline{P}^{*}_{i\Delta}(\boldsymbol{\theta})\}}{\overline{P}^{*}_{i\Delta}(\boldsymbol{\theta})^2} + \sum_{v=1}^{M} \frac{(1-\delta_i)\left\{1-\tilde{P}^{\dagger}_{vi}(\boldsymbol{\theta})\right\}}{M\tilde{P}^{\dagger}_{vi}(\boldsymbol{\theta})^2} \right] \\ &+ \left(1+\frac{1}{M}\right)\frac{\sum_{v=1}^{M}\left(\hat{N}_v - \hat{N}_{mc}\right)^2}{M-1} + \left(\frac{\partial \hat{N}_{mc}}{\partial\boldsymbol{\theta}}\right)^{\mathsf{T}} \widehat{\mathrm{Var}}(\hat{\boldsymbol{\theta}}_{mc}) \left(\frac{\partial \hat{N}_{mc}}{\partial\boldsymbol{\theta}}\right),\end{aligned}$$

where $\boldsymbol{\theta}$ is evaluated at $\hat{\boldsymbol{\theta}}_{mc}$. Finally, we show that MICS and IPWCS are asymptotically equivalent.

Theorem 5. *Under regularity conditions* A1–A2 *and* B1–B3, $\sqrt{N}(\hat{\boldsymbol{\theta}}_{wc} - \hat{\boldsymbol{\theta}}_{mc})$ *converges to* $\mathbf{0}$ *in probability as both* N *and* M *increase without bound. Similarly,* $N^{-1/2}(\hat{N}_{wc} - \hat{N}_{mc})$ *converges to* 0 *in probability as* $N, M \longrightarrow \infty$.

4. Simulations

4.1. Simulation study 1: measurement error data

Our first simulation study examined the case when covariates are only subject to measurement error. We set the number of capture occasions to $\tau = 5, 7, 10$ and 14 and considered a moderate-sized data set, where we fixed the true population size as $N = 200$, and a large-sized data set with $N = 1{,}000$. We generated one covariate X_i according to (a) the standard normal, and (b) a uniform distribution with support $(-\sqrt{3}, \sqrt{3})$, also with mean 0 and variance 1. An error-free binary covariate Z_i was drawn from a Bernoulli (Bern) distribution with probability set to 0.4. Thus, the probability of being captured was set to $P_{ij} = H(\alpha + \beta X_i + \gamma Z_i)$ with $(\alpha, \beta, \gamma) = (-1, 1, -1)$ for $j = 1, \ldots, \tau$. The number of recorded covariate measurements was obtained as $m_i = \sum_{j=1}^{\tau} Y_{ij}$ for each i. Observed surrogates $W_{ij}, j = 1, \ldots, m_i$, measured only at $Y_{is} = 1$ for $s = 1, \ldots, \tau$ for each i and j, were generated as $W_{ij} = X_i + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_u^2)$. The measurement error variance σ_u^2 was set at 0.25 and 0.5. In each simulation study we generated 200 data sets.

We fit the naïve conditional likelihood model of Section 2.1, a refined regression calibration (RRC) approach (see Web Appendix S2), a naïve conditional score (NCS) model, the conditional score 1 (CS1) model, and the proposed conditional score approach (CS-new) of Section 2.2. The RRC approach is a simple approximation method proposed by Hwang and Huang (2003) under the restriction that $m_i = 1$. In Web Appendix S2, we modify the RRC approach to allow for the general case of $m_i > 1$. For each estimating equation, we used the Newton–Raphson method (via the `nleqslv` `R`-package) for obtaining solutions.

We only present the results for $\tau = 5$. In Web Tables 1–4, we report the sample average, root mean squared error (RMSE) and 95% coverage probability (CP) for $\widehat{\boldsymbol{\theta}}$ and $\widehat{N}$. To examine the performance of the variance estimators, we report the mean of the standard error (SE) estimates using the respective method's standard error estimator and compared this with the standard deviations (SD) of $\widehat{\beta}$ and $\widehat{N}$. We also report sample averages for D and $\overline{\mathcal{Y}} = \sum_{i=1}^{D} \mathcal{Y}_i / D$. For each method and the two distributions (a)–(b), we plot boxplots in Figure 1 for

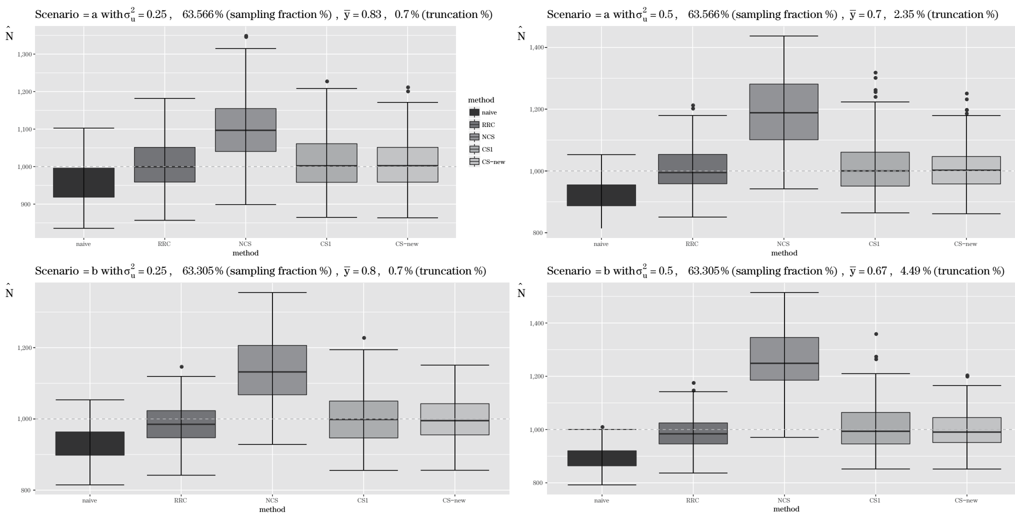

Figure 1. Simulation study 1. Comparative boxplots for $\widehat{N}$ with two measurement error variances for two scenarios (i.e., different distributions for the true covariate). The left-hand side column gives the results for $\sigma_u^2 = 0.25$ and the right-hand side column for $\sigma_u^2 = 0.5$. The top row is scenario (a) and bottom row is scenario (b). In this simulation study we used a large sized data set with $N = 1{,}000$ and $\tau = 5$.

population size estimates $\left(\widehat{N}\right)$ for $N = 1{,}000$ and $\tau = 5$ only. All other plots $\left(\text{including those for } \widehat{\beta}\right)$ are given in Web Appendix S4. For each boxplot we truncated several estimates as these were too large, and had skewed/distorted the plot. Although this rarely occurred, we removed the estimates above the third quartile plus 2.5 times inter-quartile range. The truncation percentages are shown above each comparative boxplot.

As seen in the figures and Web Tables 1–4, the naïve approach showed strong attenuation effects when estimating β, that subsequently resulted in underestimating N. Although the RMSE here was smallest, particularly for $N = 200$, its CPs were too small at the nominal 95% level. Furthermore, it performed worse as σ_u^2 increased. RRC performed well for the moderate-sized data sets with the normally distributed covariates, but it resulted in biased estimates and lower coverage for the uniform data. The poorer performance for RCC in the later case was expected as it is not a consistent method; in particular, its performance was sensitive to the normality assumption on X. The NCS method tended to "over adjust" the bias, and so that both β and N were overestimated. The resulting positive bias for N was at times severe (e.g., see Web Tables 1–4).

Both CS1 and CS-new performed as expected for the regression parameters

(Web Tables 1–4), outperforming all other methods regardless of the distribution of the true covariate and degree of measurement error. For the population size estimates, CS1 only marginally improved when $\sigma_u^2 = 0.25$ resulting in some positive bias. On the other hand, CS-new was almost unbiased for all considered cases and the length of the boxplots were generally shorter in length for each setting. It had smaller SDs and so its efficiency was better. In almost all cases, the standard error estimates for the proposed CS-new method were similar to the sample SD, see Web Tables 1–4. The relative performance for our proposed estimators was similar for large τ; we also observed that as more individuals entered the study, the results for CS1 and CS-new, both consistent, had greatly improved. In addition, all population size estimators approached the true N and their differences (in terms of bias and RMSE) were minor when τ was increased.

In reference to other studies that have considered similar problems, we compared our results with the generalized method of moments (GMM1) and the semiparametric efficient score (Semi-Nor) approaches from Xu and Ma (2014), see Remark 1. To compare model performances, the same simulation set-up was used as in Tables 1 and 2 of Xu and Ma (2014). Here, $\sigma_u = 0.6$, $N = 500$ with $\gamma = 0$ (no covariate Z_i) and the X_i were standard normal. We considered $\tau = 5$, $(\alpha, \beta) = (0.2, 1)$ and $\tau = 3$, $(\alpha, \beta) = (-1, 1)$. We generated 1,000 data sets to be consistent with the results of Xu and Ma (2014).

In Web Tables 5 and 6 we report means of the estimates and SE, SD, MSE and 95% CP for parameters and N. The results of GMM1 and the Semi-Nor methods were taken directly from Table 1 of Xu and Ma (2014). We calculated the MSE of each parameter to be MSE = bias^2 + SD^2; the "mse" values of Xu and Ma (2014) are incorrect. We also report a relative efficiency measure to compare the performance of Semi-Nor with all other methods in the study: RE = (MSE of Semi-Nor)/(MSE of model) for each setting. The relative efficiency of CS-new was over 90% for all parameters, in particular, the proposed CS-new yielded similar MSEs for $\widehat{N}$ compared to the Semi-Nor method. In Web Tables 5 and 6 the generalized method of moments approach showed little advantage over CS1 and CS-new, hence is not recommended.

4.2. Simulation study 2: measurement error with missing data

We extended this simulation by including missing data in the covariates. We considered the same settings as in Section 4.1, but now set $\sigma_u^2 = 0.1$, $\sigma_u^2 = 0.25$ and $\sigma_u^2 = 0.5$. We generated missing covariate values for different missing data cases. For each i, we used (1) $m_i \sim \sum_{j=1}^{\tau} \text{Bern}(Y_{ij}, P)$ for $P = 0.8$, (2) $m_i \sim$

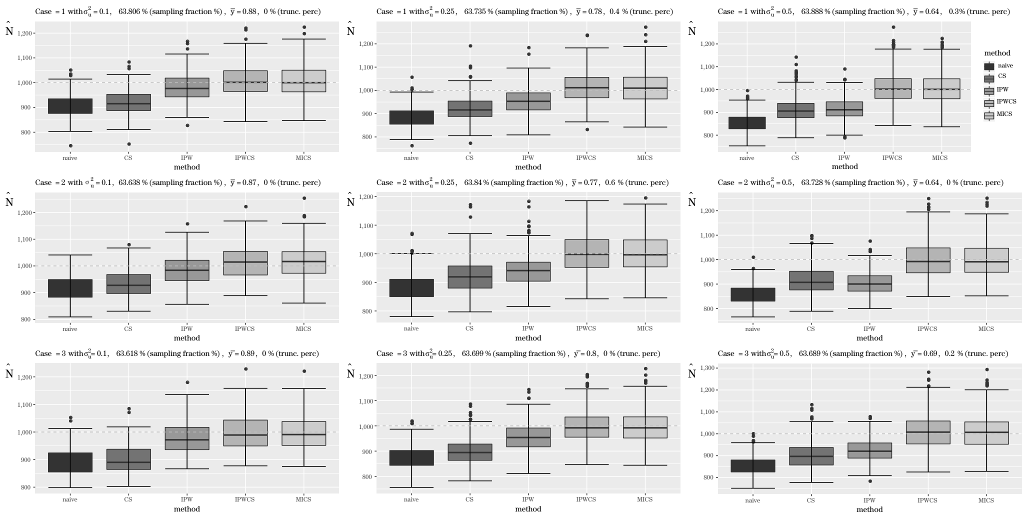

Figure 2. Simulation study 2. Comparative boxplots for $\widehat{N}$ with three measurement error variances (by columns) for three missing data cases (by rows). In this simulation study we used a large sized data set with $N = 1{,}000$ and $\tau = 5$ for scenario (a).

$\sum_{j=1}^{\tau} \text{Bern}(Y_{ij}, P(Z_i))$ for $P(Z_i) = H(1+Z_i)$, and (3) $m_i = \mathcal{Y}_i \cdot \text{Bern}(1, P(Z_i, \mathcal{Y}_i))$ for $P(Z_i, \mathcal{Y}_i) = H(-0.5 + 0.7\mathcal{Y}_i + 0.7Z_i)$. Cases (1) and (2) are situations where $m_i > 0$ and are not equal to $\mathcal{Y}_i$, while case (3) yielded either $m_i = \mathcal{Y}_i$ or 0.

We fit and compared all models presented in Section 3. For MICS we used $M = 10$ replications (Lee, Hwang and Jean (2016)) for each simulation study. For $N = 1{,}000$ with $\tau = 5$, we give comparative boxplots (see Figures 2–3) for $\hat{N}$ for each method with missing data cases (1)–(3), measurement error variances and the two settings (a)–(b). All other plots, including those for $\hat{\beta}$, for $\tau = 5$ are given in Web Appendix S4.

Here the naïve complete case approach resulted in biased estimates due to the effects of missing data when the measurement error variance was $\sigma_u^2 = 0.10$. When σ_u^2 was increased, this bias further increased under the same settings, thus the performance for β and N worsened due to the additional effects of measurement error. When comparing results with Figure 1 (given the same σ_u^2 and scenario), both the naïve and CS underestimated N due to the additional effects of missing data. Similar phenomena were observed for estimating β (Web Figures 3–6). The naïve IPW approach performed well but only for the case where the measurement error was small, see the left columns of Figures 2 and 3, this is because the effect was mainly due to the missing mechanism. As expected,

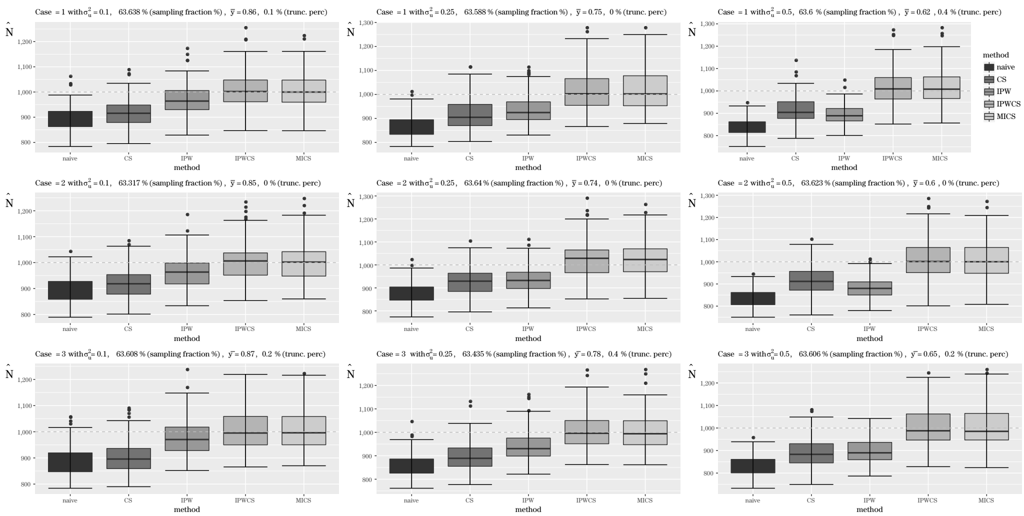

Figure 3. Simulation study 2. Comparative boxplots for $\widehat{N}$ with three measurement error variances (by columns) for three missing data cases (by rows). In this simulation study we used a large sized data set with $N = 1{,}000$ and $\tau = 5$ for scenario (b).

the naïve IPW performed poorly when the measurement error variance could not be ignored, see middle and right columns of Figures 2–3. Both IPWCS and MICS outperformed all models in terms of bias and coverage. When compared to each other, they gave similar results that improved as the sample size increased, resulting in less bias and better coverage. This suggests that these methods are asymptotically equivalent, see Figures 2–3 and Web Figures 3–6. The standard error estimates for IPWCS and MICS were comparable to the sample standard deviation of the estimates.

Finally, we compared our results with the parametric maximum likelihood method given in Xi et al. (2009). To compare model performances we followed the same set-up as given in Section 3 of Xi et al. (2009), and used results from their Table 1. Here, $\sigma_u^2 = 0.5$ and 1, $N = 200$, $\tau = 5$ with true parameter values set to $(\alpha, \beta) = (-1, 0.5)$ and $(-0.5, 1)$ for no missing data, $P_{\text{meas}} = 1$, and some missing data, $P_{\text{meas}} = 0.9$. Covariates X_i were standard normal and we generated 1,000 data sets for each scenario. We used the same measures as in Section 4.1 and reported the results in Web Tables 7 and 8 for each case, respectively. Here RE = (MSE of Xi et al. (2009))/(MSE of model). The proposed IPWCS and MICS were comparable with the maximum likelihood method. In particular, IPWCS and MICS gave appreciable results when $\sigma_u^2 = 0.5$.

5. Examples

We give two data examples where measurement error and missing values are prevalent in the observed covariates. In the first example we only considered the measurement error problem, there were no missing covariates, and in the second we simultaneously accounted for measurement error and missing values in covariates.

5.1. Example 1: Harvest mouse data

We use capture–recapture data collected on the Harvest mouse in Taiwan. These data have been analysed in Huang, Hwang and Chen (2011) and Stoklosa et al. (2011). Captures of mice were collected across 14 sampling occasions where, upon capture, additional measurements, such as body weight, gender, head-to-tail length, etc. were collected on individuals. These data consist of $D = 142$ uniquely captured individuals across $\tau = 14$ capture occasions.

Previous studies that have used similar data have identified *body weight* measured in grams (g) as a potential covariate to model heterogeneity in capture probabilities, thus the body weight measurements were subject to uncertainty. We used this covariate to model capture probability and correct for measurement error using the methods given in Section 2. The average number of times the body weight covariate was observed across the capture occasions was $\overline{m} = 2.18$. The average body weight was $\overline{W} = 8.24$g with a sample variance of $S_W^2 = 4.58\text{g}^2$. The measurement error variance for the body weight covariate was estimated to be $\sigma_u^2 = 0.64\text{g}^2$, which gives an estimate of the reliability percentage as $(1 - 0.64/4.58) \times 100 = 86.1\%$ with respect to a single measurement. We initially considered gender in the analysis but found no statistical significance and therefore decided to exclude it.

As in Section 4.1, we fit the naïve conditional likelihood, RRC, NCS, CS1 (the first observed measurement was used) and the proposed CS method. In Table 1 we give coefficient and population size estimates (standard errors are in parentheses) for each method. The naïve and NCS gave similar estimates for the intercept and slope parameters, although the population size estimate for NCS was slightly smaller. CS1, RRC and the proposed CS had very similar estimates for the slope that were quite different from those of the naïve method, but all three gave larger standard errors. RRC and the proposed CS gave similar population size estimates but the standard error for RRC was unrealistically large – this was unusual since we did not obtain such large standard errors in the simulation study. Finally, the population size for the proposed CS method was

Table 1. Coefficient and population size estimates for each method fitted to the harvest mouse capture–recapture data. We fit the naïve conditional likelihood model of Section 2.1, the NCS model, the CS1 model, an RRC approach (see Web Appendix S3), and CS-new.

Method	β_0 (Intercept)	β_1 (Body weight)	N (Population size)
naïve	−4.08 (0.37)	0.27 (0.04)	175.98 (10.61)
NCS	−4.05 (0.36)	0.27 (0.04)	173.62 (9.23)
CS1	−4.37 (0.44)	0.30 (0.05)	180.17 (11.49)
RRC	−4.26 (0.40)	0.29 (0.04)	178.18 (49.31)
CS-new	−4.48 (0.40)	0.31 (0.05)	179.90 (11.75)

similar to CS1, as observed in the first simulation study.

5.2. Example 2: eastern barred bandicoots data

Next, we analysed capture–recapture data collected on the Eastern barred bandicoots *Perameles gunnii* in Hamilton, South eastern Victoria, Australia. In the experiment, $D = 77$ uniquely tagged bandicoots were trapped across $\tau = 5$ sampling occasions in November, 2012. We considered two covariates collected during trapping: *gender*, which was correctly identified each time an individual was seen; and *body weight*, which was missing on some occasions upon capture and consisted of imprecise measurements. There were 50 unique females and 27 males captured in this study period. The covariates of gender and body weight were used to model capture probabilities.

The observed average body weight was $\overline{W} = 0.67$kg with a sample variance of 0.019kg^2. The average number of times the body weight covariate was observed across the capture occasions was $\overline{m} = 1.39$, and the average number of times an individual was captured across the capture occasions was $\overline{Y} = 2.13$. The measurement error variance for the body weight covariate was estimated to be $\sigma_u^2 = 0.01$kg^2, giving a low reliability percentage estimate of 45.2%. There were 14 individuals without a record of body weight, hence the missing data rate was 18.2%.

We fit the naïve complete case method, naïve IPW, CS, IPWCS and MICS (using $M = 200$ replications). In Table 2 we give the coefficient and population size estimates (standard errors are in parentheses) for each method fit to the data. The naïve complete case method yielded a population size estimate around 84. The naïve IPW and CS methods gave similar population size estimates around 90, although their regression parameters estimates were quite different. The naïve IPW method gave large standard errors for both regression

Table 2. Coefficient and population size estimates for each method (discussed in Section 5.2) fitted to the bandicoot capture–recapture data. We fit the naïve complete case method, the naïve IPW model using the surrogate $\overline{W}_i$, the CS approach (2.4)–(2.5) with complete case only, the IPWCS model and MICS.

Method	β_0 (Intercept)	β_1 (Body weight)	γ_0 (Gender)	N (Population size)
naïve	−2.38 (0.62)	2.95 (0.90)	0.58 (0.26)	83.9 (4.07)
IPW	−2.98 (7.00)	3.49 (10.0)	0.60 (0.37)	89.8 (19.2)
CS	−5.06 (2.23)	7.06 (3.39)	0.37 (0.37)	88.9 (11.0)
IPWCS	−7.67 (3.76)	10.6 (5.71)	0.32 (0.48)	117 (51.9)
MICS	−7.69 (4.41)	10.7 (6.68)	0.28 (0.54)	110 (59.3)

parameters and population size.

IPWCS and MICS gave similar estimates that were distinctively different from those found by the other methods. These differences suggest that both measurement error and missingness are present in the body weight covariate. If both of these are ignored then conclusions can be misleading. For both methods, the estimated population sizes were around 110–117 with a high standard error estimated near 55. To confirm these high standard errors, we conducted a non-parametric bootstrap (with 100 bootstrap replications) to estimate them. We found that the bootstrap standard errors estimates yielded similar results. We caution that these large standard errors may have occurred due to the poor reliability of measurements.

6. Discussion

Through several simulation studies we showed that the proposed methods outperformed the naïve approach, and results were comparable (if not better) with other established methods.

Generally, IPWCS and MICS give the same results, but each method has its own advantages, briefly discussed here. The MICS approach can be easily generalized to impute values that are partial missing – e.g., if X is a bivariate vector that is partially missing. In such cases, MICS is more efficient than IPWCS. IPWCS is computationally faster than the MICS approach, and in simpler cases would be recommended in practice. For the MICS approach, instead of imputing m_i and W_{ij}, we can also impute the estimating function $\Phi_i(\boldsymbol{\theta})$. In Web Appendix S2, we provide an alternative algorithm that gives asymptotically equivalent results.

A general problem with the IPW method is that small $\hat{\pi}_i$ may result in large inflated estimates (Seaman and White (2013)). In simulations and examples we did not encounter these issues. We suspect that if the number of categories for Z is large, then this problem may be unavoidable, and some further adjustments would be necessary such as weight truncation and semi-parametric modelling with logistic regression, see Section 5.3 in Seaman and White (2013). In addition, we could consider an approach given in Stoklosa and Huggins (2012), where lower-bound methods were implemented on estimated capture probabilities to enhance robustness.

Our methods differ from those given by Xi et al. (2009). Their approach requires making distributional assumptions on the true underlying covariate, that we relaxed while maintaining good estimator performance. Still, we also required the assumption that the measurement error is normally distributed. To relax this assumption, we could consider a corrected score approach (Carroll et al. (2006)), but this is still not possible under the conditional likelihood framework (Huggins (1989)) – neither the logistic function $H(\cdot)$ nor the probability $P_i^*(\boldsymbol{\theta})$ is analytically tractable. An alternative approach uses an approximated correct score, or other models given in Stoklosa et al. (2011), to help solve this problem. An extension for the proposed study could consider more general capture–recapture models where both capture and recapture probabilities are related to temporal and behavioral effects (Huggins (1989)). Incorporating "unobserved heterogeneity" models (Pledger (2005); Farcomeni (2016)) into the measurement error and missing data model framework could be a very challenging problem. These extensions will be explored elsewhere.

Supplementary Materials

Web Appendices S1 contains proofs of Theorems 1–5, Web Appendices S2 contains a refined regression calibration method, Web Appendices S3 contains an alternative MICS algorithm, and Web Appendices S4 contains Web Figures 1–6 and Web Tables 1–8 consisting of results for simulation studies 1 and 2.

Acknowledgment

The authors would like to thank Dr Andrew Weeks (The University of Melbourne), Richard Hill (Department of Environment, Land, Water and Planning Victoria) and the Department of Environment, Land, Water and Planning Victoria for collecting and providing the Eastern barred bandicoots capture–recapture data.

References

Alho, J. M. (1990). Logistic regression in capture–recapture models. *Biometrics* **54**, 623–635.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective.* 2nd edition. Chapman & Hall/CRC, London.

Farcomeni, A. (2016). A general class of recapture models based on the conditional capture probabilities. *Biometrics* **72**, 116–124.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* **72**, 147–148.

Gould, W. R., Stefanski, L. A. and Pollock, K. H. (1999). Use of simulation–extrapolation estimation in catch–effort analyses. *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 1234–1240.

Huang, Y. H., Hwang, W. H. and Chen, F. Y. (2011). Differential measurement errors in zero-truncated regression models for count data. *Biometrics* **67**, 1471–1480.

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 113–140.

Huggins, R. M. and Hwang, W. H. (2010). A measurement error model for heterogeneous capture probabilities in mark-recapture experiments: An estimating equation approach. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 198–208.

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

Hwang, W. H. and Huggins, R. M. (2005). An examination of the effect of heterogeneity on the estimation of population size using capture–recapture data. *Biometrika* **55**, 387–395.

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

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

Little, R. J. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association* **87**, 1227–1237.

McCrea, R. S. and Morgan, B. J. (2014). *Analysis of Capture-Recapture Data.* Chapman and Hall/CRC Press.

Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics* **61**, 868–876.

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

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley: New York.

Seaman, S. R. and White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research* **22**, 278–295.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**, 703–716.

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

Stoklosa, J. and Huggins, R. M. (2012). A robust P-spline approach to closed population capture–recapture models with time dependence and heterogeneity. *Computational Statis-*

tics & Data Analysis **56**, 408–417.

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

Wang, N. and Robins, J. M. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika* **85**, 935–948.

Xi, L., 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. *The Canadian Journal of Statistics* **37**, 645–658.

Xu, K. and Ma, Y. (2014). Effective use of multiple error-prone covariate measurements in capture–recapture models. *Statistica Sinica* **24**, 1529–1546.

Zwane, E. N. and van der Heijden, P. G. M. (2007). Analysing capture–recapture data when some variables of heterogeneous catchability are not collected or asked in all registrations. *Statistics in Medicine* **26**, 1069–1089.

School of Mathematics and Statistics and Evolution & Ecology Research Centre, The University of New South Wales, Sydney NSW 2052, Australia.

E-mail: j.stoklosa@unsw.edu.au

Department of Statistics, Feng Chia University, Taichung City, 407, Taiwan.

E-mail: smlee@fcu.edu.tw

Institute of Statistics, National Chung Hsing University, Taichung City, 402, Taiwan.

E-mail: wenhan@nchu.edu.tw

(Received February 2017; accepted October 2017)