

CONTINUOUS-TIME CAPTURE-RECAPTURE MODELS WITH COVARIATES

Wen-Han Hwang^{1,2} and Anne Chao¹

¹*National Tsing Hua University and* ²*Feng-Chia University*

Abstract: This paper develops a class of continuous-time closed capture-recapture models which incorporate the use of covariates such as environmental variables or an individual's characteristics. The capture intensity is allowed to vary with time, behavioural response and heterogeneity. The heterogeneity effect is modeled as a function of observable covariates but no assumptions regarding the time-varying function are made. The proposed hierarchy of models can be regarded as the continuous version of discrete-time models used in ecological applications. A unified likelihood-based approach is proposed to assess the effect of each possibly time-dependent covariate and to obtain population size estimators. Our model generalizes Yip, Huggins and Lin (1996) to incorporate an animal's behavioural response and to make use of all capture frequency data. The approach also extends Lin and Yip (1999) to a more general semi-parametric approach. Simulation results are presented to show the performance of the proposed estimation procedures. The estimators are applied to a set of capture data for house mouse (*Mus musculus*) discussed in the literature.

Key words and phrases: Conditional likelihood, Horvitz-Thompson estimator, population size, recurrent event.

1. Introduction

For a closed animal population, we assume there is no birth, death or migration so that the population size is a constant over the time period of a trapping experiment. There are two types of capture-recapture models: discrete-time and continuous-time models. In a typical discrete-time model, the target population is sampled several times (or over a certain number of occasions). For each trapping sample or occasion, traps are placed in the study area during the sampling time. After each sampling, one checks the traps and records "first-capture" or "recapture" for each capture. A unique tag or mark is attached to a first-capture, whereas for a recapture its tag number is recorded. The complete capture history for each animal is conveniently expressed as a sequence of 0's and 1's, where 0 denotes absence and 1 denotes presence. For example, in a five-sample capture-recapture experiment, each animal can be counted at most five times; a history (0 1 0 0 1) means that the animal was caught in the second and fifth sample,

but not in the others. The maximum count for each animal is the number of samples.

For a continuous-time model, in addition to the tagging process we also record the exact capture times for each animal. Any capture is regarded as a “trapping occasion” and the exact time for each occasion is recorded. Therefore, for each animal captured in the experiment, the complete capture history consists of a series of capture times. As an example, an individual’s capture history (1, 4, 6.5, 8, 9) means that the animal was caught in time unit of 1, 4, 6.5, 8 and 9. There is no limit on the capture frequency for each animal.

As indicated by Wilson and Anderson (1995), applications of continuous-time models include studies of sperm whales, grizzly bears, insects and butterflies. The model has also found application in software reliability theory; see Nayak (1988). In the latter cases involving a debugging process, each bug is regarded as an individual and the detection times for each bug can be considered as “capture times”.

Discrete-time capture-recapture models have been discussed extensively in biological and ecological literature. Pollock proposed a class of models where time, behavioural response or heterogeneity may affect the capture probabilities; see Otis, Burnham, White and Anderson (1978), Pollock (1991), Schwarz and Seber (1999) and references therein. These models are referred to as models \mathcal{M}_t , \mathcal{M}_b , \mathcal{M}_h , \mathcal{M}_{tb} , \mathcal{M}_{th} , \mathcal{M}_{bh} and \mathcal{M}_{tbb} where the subscript “ t ”, “ b ” and “ h ” denote “time-varying”, “behavioural-response” and “heterogeneity”, respectively. In this paper, we propose an analogous class of continuous-time models \mathcal{M}_t , \mathcal{M}_b , \mathcal{M}_h , \mathcal{M}_{tb} , \mathcal{M}_{th} , \mathcal{M}_{bh} and \mathcal{M}_{tbb} . There is relatively little literature on continuous-type models. Papers by Becker (1984), Becker and Heyde (1990), Chao and Lee (1993), Yip and Chao (1996), Yip, Huggins and Lin (1996) and Lin and Yip (1999) are exceptions.

In addition to capture records, some environmental variables (temperature, humidity and rainfall) as well as an individual’s characteristics (age, sex, body weight and wing length) are usually also recorded in an experiment. Pollock, Hines and Nichols (1984) were the first to propose a full likelihood approach using covariates in a discrete-time model. One difficulty with the full likelihood approach is that the covariates for uncaptured animals are not observable. Thus the unconditional method may not be theoretically applicable unless some assumptions regarding covariates are made. Huggins (1989, 1991) circumvented the foregoing difficulty by using a conditional likelihood so that the covariates of the uncaptured animals can be avoided in the analysis, but Huggins only treated discrete-time models.

For the continuous-time model \mathcal{M}_{th} , which assumes that capture intensity varies with time and with an individual’s covariates, Yip et al. (1996) developed a

partial likelihood approach to assess the effects of covariates, but their approach ignores the information of first captures. Lin and Yip (1999) applied a likelihood score function and a martingale method to a general class of regression models in which a parametric assumption for the time-varying function is made.

This paper deals with continuous-time models and assumes the intensity function varies with time, behavioural response and individual. The heterogeneity effect is modeled as a function of the individual’s covariates but no assumptions regarding the time-varying function are made. Our model generalizes Yip et al. (1996) to incorporate an animal’s behavioural response and to make use of all capture records; this paper also extends Lin and Yip (1999) to a semi-parametric approach.

Section 2 presents our models and estimation procedures. We focus on the most general model \mathcal{M}_{tbh} and only outline the results for its submodels. A unified conditional likelihood approach is developed to assess the effect of each covariate and to obtain population size estimators. The resulting estimating equations can also be justified from the full likelihood. A simulation study is reported in Section 3 to examine the performance of the proposed estimation procedure. Section 4 presents an illustrative example of the house mouse (*mus musculus*) provided in the software CAPTURE (Rexstad and Burnham (1991)).

2. Models and Estimators

Assume there are ν individuals, indexed by $1, \dots, \nu$. Also assume that the experiment period is relatively short so that the population size remains fixed in the study period. Suppose that the experiment terminates at the time τ and $N_i(t)$ denotes the number of times the i th animal has been caught in $[0, t]$. Each $\{N_i(t); 0 \leq t \leq \tau\}$ is a continuous-time counting process with intensity $\lambda_i(t)$. The intensity for the i th animal, $\lambda_i(t)$ is $\lambda_i(t)dt = P\{dN_i(t) = 1|F_{t-}\}$, where F_t is the capture history generated by $\{N_1(u), \dots, N_\nu(u); 0 \leq u \leq t\}$. Let the associated covariates for the i th individual be $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})'$. In the following, we only present the estimation procedure for time-independent covariates because an experiment’s duration is usually short for a closed model as a matter of practice. (The extension to the case with deterministic time-dependent covariates is parallel.) Let $\lambda_0(t)$ be any arbitrary non-negative time-varying function defined in $[0, \tau]$. The covariates are used to model individual heterogeneity. Let $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ be a vector of unknown parameters. We use $\lambda_0(t)$, $\exp(\boldsymbol{\beta}'\mathbf{Z}_i)$ and ϕ to model, respectively the time, heterogeneity and the behavioural response effects. Thus a multiplicative type of model \mathcal{M}_{tbh} is

$$\lambda_i(t) = \begin{cases} \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{Z}_i) & \text{until first capture,} \\ \phi \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{Z}_i) & \text{for any recapture.} \end{cases}$$

The proposed hierarchy of continuous-time models is summarized in Table 1.

Table 1. A Hierarchy of Continuous-Time Models.

Model	Assumption	Restriction in model \mathcal{M}_{tbh}
\mathcal{M}_{tbh}	$\lambda_i(t) = \begin{cases} \lambda_0(t) \exp(\beta' \mathbf{Z}_i) & \text{until first capture,} \\ \phi \lambda_0(t) \exp(\beta' \mathbf{Z}_i) & \text{for any recapture.} \end{cases}$	
\mathcal{M}_{bh}	$\lambda_i(t) = \begin{cases} \lambda \exp(\beta' \mathbf{Z}_i) & \text{until first capture,} \\ \phi \lambda \exp(\beta' \mathbf{Z}_i) & \text{for any recapture.} \end{cases}$	i.e., set $\lambda_0(t) \equiv \lambda$ in model \mathcal{M}_{tbh}
\mathcal{M}_{th}	$\lambda_i(t) = \lambda_0(t) \exp(\beta' \mathbf{Z}_i)$	i.e., set $\phi = 1$ in model \mathcal{M}_{tbh}
\mathcal{M}_{tb}	$\lambda_i(t) = \begin{cases} \lambda_0(t) & \text{until first capture,} \\ \phi \lambda_0(t) & \text{for any recapture.} \end{cases}$	i.e., $\beta = 0$ in model \mathcal{M}_{tbh}
\mathcal{M}_h	$\lambda_i(t) = \lambda \exp(\beta' \mathbf{Z}_i)$	i.e., set $\lambda_0(t) \equiv \lambda$, $\phi = 1$ in model \mathcal{M}_{tbh}
\mathcal{M}_b	$\lambda_i(t) = \begin{cases} \lambda & \text{until first capture,} \\ \phi \lambda & \text{for any recapture.} \end{cases}$	i.e., set $\beta = 0$, $\lambda_0(t) \equiv \lambda$ in model \mathcal{M}_{tbh}
\mathcal{M}_t	$\lambda_i(t) = \lambda_0(t)$	i.e., set $\beta = 0$, $\phi = 1$ in model \mathcal{M}_{tbh}

Lin and Yip (1999) discussed a similar type of regression model, but they adopted a parametric approach by assuming that the time-varying function is characterized by only one parameter. Their model \mathcal{M}_{tbh} is thus equivalent to our model \mathcal{M}_{bh} .

Let $\phi = \exp(\alpha)$ and $X_i(t) = I$ [the i th animal has been captured in $(0, t)$] denotes the prior capture history, where $I[\cdot]$ is the usual indicator function. For the most general model \mathcal{M}_{tbh} , the intensity of the i th individual can be rewritten as

$$\lambda_i(t) = \lambda_0(t) \exp(\beta' \mathbf{Z}_i + \alpha X_i(t)). \quad (2.1)$$

This is the Cox (1972) regression model for recurrent event analysis, where “event” means “capture” and $\lambda_0(t)$ denotes the baseline intensity function; see Andersen and Gill (1982) and Andersen, Borgan, Gill and Keiding (1993). In our approach, no assumptions regarding the form of this function are necessary. The Andersen-Gill family of models has been extensively discussed in the context of survival analysis, reliability theory and recurrent event analysis. However, the inference procedure under the Andersen-Gill model cannot be directly applied to our models because (1) the number of subjects is known in the Cox and Andersen-Gill models, while in our model the population size is the main parameter of interest, and (2) all the covariates are observable in the Cox model, but in our model the covariates for the uncaptured animals are missing. Therefore, the statistical methods used in the recurrent event analysis need to be modified to handle our models.

2.1. Model \mathcal{M}_{tbbh}

Let $N_t = \sum_{i=1}^{\nu} N_i(t)$ be the total number of captures by time t , and M_t be the total number of animals that are marked (i.e., first-captures) by time t . Hence M_{τ} represents the number of distinct animals captured in the whole experiment period. Let $K_t = N_t - M_t$ denote the number of re-captures by time t . Without loss of generality, label the captured individuals as $1, \dots, M_{\tau}$. Let the i th individual be captured $m_i = N_i(\tau)$ times and the capture times be t_{i1}, \dots, t_{im_i} . Thus $\sum_{i=1}^{M_{\tau}} m_i = N_{\tau}$ is the total number of captures.

For notational simplicity, let $\gamma_i = \exp(\beta' \mathbf{Z}_i)$ and denote the baseline cumulative intensity function by $\Lambda_t = \int_0^t \lambda_0(u) du$, $t \in [0, \tau]$. If $m_i > 0$, the likelihood function can be obtained by using a similar argument as in Crowder, Kimber, Smith and Sweeting (1991, p.165) as follows. Given a small increment Δt_{ij} at each point t_{ij} , $j = 1, \dots, m_i$, we consider the following independent consecutive events and their associated probabilities: no capture in $(0, t_{i1})$ with probability $\exp[-\int_0^{t_{i1}} \gamma_i \lambda_0(u) du]$, one capture in $(t_{i1}, t_{i1} + \Delta t_{i1})$ with probability $\gamma_i \lambda_0(t_{i1}) \Delta t_{i1}$, no capture in $(t_{i1} + \Delta t_{i1}, t_{i2})$ with probability $\exp[-\phi \int_{t_{i1} + \Delta t_{i1}}^{t_{i2}} \gamma_i \lambda_0(u) du]$, one capture in $(t_{i2}, t_{i2} + \Delta t_{i2})$ with probability $\phi \gamma_i \lambda_0(t_{i2}) \Delta t_{i2}, \dots$, and so on, up to no capture in $(t_{im_i} + \Delta t_{im_i}, \tau)$ with probability $\exp[-\phi \int_{t_{im_i} + \Delta t_{im_i}}^{\tau} \gamma_i \lambda_0(u) du]$. Let all the increments tend to 0, and a standard approach implies that the likelihood under model \mathcal{M}_{tbbh} for the i th captured animal is

$$\begin{aligned}
 L_i &\propto \exp(-\gamma_i \Lambda_{t_{i1}}) \gamma_i \lambda_0(t_{i1}) \\
 &\times \left[\prod_{j=2}^{m_i} \phi \lambda_0(t_{ij}) \gamma_i \exp[-\phi \gamma_i (\Lambda_{t_{ij}} - \Lambda_{t_{i,j-1}})] \right] \cdot \exp[-\phi \gamma_i (\Lambda_{\tau} - \Lambda_{t_{i,m_i}})] \\
 &= \phi^{m_i-1} \gamma_i^{m_i} \left[\prod_{j=1}^{m_i} \lambda_0(t_{ij}) \right] \exp[-\phi \gamma_i \Lambda_{\tau} + (\phi - 1) \gamma_i \Lambda_{t_{i1}}]. \tag{2.2}
 \end{aligned}$$

Let $\delta_i = I$ [the i th individual is captured at least once in the experiment] and take

$$P_i \equiv P_i(\beta, \Lambda_{\tau}) \equiv P(\delta_i = 1) = 1 - \exp(-\gamma_i \Lambda_{\tau}). \tag{2.3}$$

Therefore, the conditional likelihood that only considers the captured animals (i.e., $m_i > 0$) becomes

$$L_c = \prod_{i=1}^{M_{\tau}} \frac{L_i}{P_i} = \left[\prod_{i=1}^{M_{\tau}} \prod_{j=1}^{m_i} \lambda_0(t_{ij}) \right] \prod_{i=1}^{M_{\tau}} \frac{\gamma_i^{m_i} \phi^{m_i-1} \exp[-\phi \gamma_i \Lambda_{\tau} + (\phi - 1) \gamma_i \Lambda_{t_{i1}}]}{1 - \exp(-\gamma_i \Lambda_{\tau})}. \tag{2.4}$$

For the moment, if we assume that the function $\lambda_0(t)$ is known in the above likelihood, the resulting estimating equations are:

$$\frac{\partial \log L_c}{\partial \phi} = \sum_{i=1}^{M_{\tau}} [\phi^{-1} (m_i - 1) - \gamma_i (\Lambda_{\tau} - \Lambda_{t_{i1}})] = \frac{K_{\tau}}{\phi} - \sum_{i=1}^{M_{\tau}} \gamma_i (\Lambda_{\tau} - \Lambda_{t_{i1}}) = 0, \tag{2.5a}$$

$$\frac{\partial \log L_c}{\partial \beta} = \sum_{i=1}^{M_\tau} \left[m_i - \frac{\gamma_i \Lambda_\tau}{P_i} + (1 - \phi) \gamma_i (\Lambda_\tau - \Lambda_{t_{i1}}) \right] \mathbf{Z}_i = \mathbf{0}. \tag{2.5b}$$

These equations are given in terms of the baseline cumulative intensity functions. For model \mathcal{M}_{bh} , we have $\lambda_0(t) \equiv \lambda$, thus $\Lambda_t = \lambda t$. In this case, the estimating equations are directly obtained from (2.5) by use of the reparametrization $\lambda = \exp(\beta_0)$ and adding a constant covariate. See Section 2.2.

For models with an arbitrary time-varying intensity function, as will be theoretically justified later, the function Λ_t is estimated by a modified Nelson-Aalen estimator (Nelson (1972); Aalen (1978)). Here we first give an intuitive interpretation of the modified Nelson-Aalen estimator. Define M_t^* as the collection of all marked animals just before time t . Note that given the capture history before time t , there are $\nu - M_t$ unmarked animals with total intensities $[\sum_{i=1}^\nu \gamma_k - \sum_{i \in M_t^*} \gamma_i] \lambda_0(t)$, and M_t marked animals with total intensities $[\sum_{i \in M_t^*} \phi \gamma_i] \lambda_0(t)$. Then $E(dN_t | F_{t-}) = [\sum_{i=1}^\nu \gamma_i + (\phi - 1) \sum_{i \in M_t^*} \gamma_i] \lambda_0(t)$. From this, the Nelson-Aalen estimator $\hat{\Lambda}$ of Λ is a step function with jumps occurring at capture times, i.e.,

$$d\hat{\Lambda}_t = \left[\sum_{i=1}^\nu \gamma_i + (\phi - 1) \sum_{i \in M_t^*} \gamma_i \right]^{-1} dN_t, \quad 0 \leq t \leq \tau. \tag{2.6a}$$

However, in the above estimator, the population size ν and the covariate information for the uncaptured animals are not available, thus we replace $\sum_{i=1}^\nu \gamma_i$ by the Horvitz-Thompson (1952) estimator $\sum_{i=1}^\nu \gamma_i I[\delta_i = 1] / P_i = \sum_{i=1}^{M_\tau} \gamma_i / P_i$, where P_i is defined in (2.3). The modified estimator of the baseline cumulative intensity function is

$$d\hat{\Lambda}_t = \left[\sum_{i=1}^{M_\tau} (\gamma_i / P_i) + (\phi - 1) \sum_{i \in M_t^*} \gamma_i \right]^{-1} dN_t, \quad 0 \leq t \leq \tau. \tag{2.6b}$$

Combining (2.5), (2.6b) and the identities

$$\sum_{i=1}^{M_\tau} \gamma_i (\Lambda_\tau - \Lambda_{t_{i1}}) = \int_0^\tau \left(\sum_{i \in M_t^*} \gamma_i \right) d\Lambda_t, \tag{2.7a}$$

$$\sum_{i=1}^{M_\tau} \gamma_i (\Lambda_\tau - \Lambda_{t_{i1}}) \mathbf{Z}_i = \int_0^\tau \left(\sum_{i \in M_t^*} \gamma_i \mathbf{Z}_i \right) d\Lambda_t, \tag{2.7b}$$

we obtain Equations (2.8a) and (2.8b), in which only one additional parameter Λ_τ , the baseline cumulative intensity function for the whole experiment period,

is involved. Hence we write the resulting system of equations as follows:

$$\frac{K_\tau}{\phi} - \int_0^\tau \frac{\sum_{i \in M_t^*} \gamma_i}{[\sum_{i=1}^{M_\tau} (\gamma_i/P_i) + (\phi - 1) \sum_{i \in M_t^*} \gamma_i]} dN_t = 0, \tag{2.8a}$$

$$\sum_{i=1}^{M_\tau} (m_i - \frac{\gamma_i \Lambda_\tau}{P_i}) \mathbf{Z}_i + (1 - \phi) \int_0^\tau \frac{\sum_{i \in M_t^*} \gamma_i \mathbf{Z}_i}{[\sum_{i=1}^{M_\tau} (\gamma_i/P_i) + (\phi - 1) \sum_{i \in M_t^*} \gamma_i]} dN_t = \mathbf{0}, \tag{2.8b}$$

$$\Lambda_\tau = \int_0^\tau \frac{1}{[\sum_{i=1}^{M_\tau} (\gamma_i/P_i) + (\phi - 1) \sum_{i \in M_t^*} \gamma_i]} dN_t. \tag{2.8c}$$

Further, notice that if the covariate in (2.8b) is a scalar, then (2.8b) is equivalent to (2.8c). Thus in this case we can combine (2.8b) and (2.8c) into one by including an additional constant covariate. Defining $\mathbf{Z}_i^+ = (1, Z_{i1}, \dots, Z_{ip})'$, the combined equations become

$$\sum_{i=1}^{M_\tau} (m_i - \frac{\gamma_i \Lambda_\tau}{P_i}) \mathbf{Z}_i^+ + (1 - \phi) \int_0^\tau \frac{\sum_{i \in M_t^*} \gamma_i \mathbf{Z}_i^+}{[\sum_{i=1}^{M_\tau} (\gamma_i/P_i) + (\phi - 1) \sum_{i \in M_t^*} \gamma_i]} dN_t = \mathbf{0}. \tag{2.8d}$$

Here \mathbf{Z}_i^+ is a column vector of dimension $p + 1$, so (2.8d) contains $p + 1$ equations. Therefore, the above system of equations consists of $p + 2$ equations and there are $p + 2$ unknowns: $(\boldsymbol{\beta}, \phi, \Lambda_\tau) = (\beta_1, \dots, \beta_p, \phi, \Lambda_\tau)$. Numerical iteration is required to obtain the solution. An algorithm is the following. Given an initial value of $(\boldsymbol{\beta}, \phi)$, determine the value of Λ_τ by (2.8c); for this fixed value of Λ_τ , obtain a second value of $(\boldsymbol{\beta}, \phi)$ from (2.8a) and (2.8b); iterate to converge. Let the solution be denoted as $(\hat{\boldsymbol{\beta}}, \hat{\phi}, \hat{\Lambda}_\tau) = (\hat{\beta}_1, \dots, \hat{\beta}_p, \hat{\phi}, \hat{\Lambda}_\tau)$, then we subsequently get $\hat{\gamma}_i = \exp(\hat{\boldsymbol{\beta}}' \mathbf{Z}_i)$. The proposed Horvitz-Thompson type of population size estimator is

$$\hat{\nu} = \sum_{\delta_i=1} I(\delta_i = 1) / [1 - \exp(-\hat{\gamma}_i \hat{\Lambda}_\tau)] = \sum_{i=1}^{M_\tau} 1 / [1 - \exp(-\hat{\gamma}_i \hat{\Lambda}_\tau)]. \tag{2.9}$$

We now give a brief justification for the Nelson-Aalen estimator from the full likelihood. For the case of the i th animal who has never been caught, i.e., $m_i = 0$, the likelihood becomes $\exp(-\gamma_i \Lambda_\tau)$. It then follows from (2.2) that the full likelihood based on the capture history is

$$L(\boldsymbol{\beta}, \phi, \Lambda) \propto \left[\prod_{i=1}^{M_\tau} L_i \right] \exp\left(- \sum_{i=M_\tau+1}^\nu \gamma_i \Lambda_\tau\right). \tag{2.10}$$

We can factor the likelihood as $L(\boldsymbol{\beta}, \phi, \Lambda) = L_1(\boldsymbol{\beta}, \phi) L_2(\boldsymbol{\beta}, \phi, \Lambda)$, where

$$L_1(\boldsymbol{\beta}, \phi) = \phi^{K_\tau} \left[\prod_{i=1}^{M_\tau} \gamma_i^{m_i} \right] \left[\prod_{i=1}^{M_\tau} \prod_{j=1}^{m_i} \left[\sum_{k=1}^\nu \gamma_k + (\phi - 1) \sum_{k \in M_{t_{ij}}^*} \gamma_k \right] \right]^{-1},$$

$$L_2(\boldsymbol{\beta}, \phi, \Lambda) = \left[\prod_{i=1}^{M_\tau} \prod_{j=1}^{m_i} \left[\sum_{k=1}^{\nu} \gamma_k + (\phi - 1) \sum_{k \in M_{t_{ij}}^*} \gamma_k \right] \lambda_0(t_{ij}) \right] \exp \left[- \int_0^\tau \left[\sum_{k=1}^{\nu} \gamma_k + (\phi - 1) \sum_{k \in M_u^*} \gamma_k \right] d\Lambda_u \right].$$

The baseline cumulative intensity function appears only in L_2 . For any given ν and ϕ , L_2 corresponds to a probability measure based on the observed failure times $\{t_{ij}, j = 1, \dots, m_i, i = 1, \dots, M_\tau\}$ from a “multiplicative model” with intensity $\lambda^*(t) = [\sum_{k=1}^{\nu} \gamma_k + (\phi - 1) \sum_{k \in M_t^*} \gamma_k] d\Lambda_t$. Following the approach of Andersen et al. ((1993), Section IV.1.5), we use the MLE here in the broad sense of Kiefer and Wolfowitz (1956); see also Scholz (1980) for a unified approach. The approach is based on the pairwise comparison of probability measures. The non-parametric MLE of Λ in this sense turns out to be the Nelson-Aalen estimator; see Nelson (1972) and Aalen (1978). That is, the MLE $\hat{\Lambda}$ of Λ has jump $1/[\sum_{k=1}^{\nu} \gamma_k + (\phi - 1) \sum_{k \in M_t^*} \gamma_k]$ at $t = t_{ij}$. This yields $L_2(\boldsymbol{\beta}, \phi, \hat{\Lambda}) = \exp(-N_\tau)$, which is free of the parameters ν and ϕ . Thus the profile likelihood $L(\boldsymbol{\beta}, \phi, \hat{\Lambda})$ is proportional to $L_1(\boldsymbol{\beta}, \phi)$ and the estimation of $\boldsymbol{\beta}$ and ϕ is simply based on $L_1(\boldsymbol{\beta}, \phi)$. Then it can be seen that the resulting derivatives are identical to those given in (2.8) when the unknown quantities $\sum_{i=1}^{\nu} \gamma_i$ and $\sum_{i=1}^{\nu} \gamma_i \mathbf{Z}_i$ in the derivatives are, respectively, replaced by $\sum_{i=1}^{M_\tau} \gamma_i / P_i$ and $\sum_{i=1}^{M_\tau} \gamma_i \mathbf{Z}_i / P_i$.

A variance estimator and confidence interval associated with the estimator $\hat{\nu}_{tbh}$ under model \mathcal{M}_{tbh} can be constructed by using a nonparametric bootstrap procedure (Efron and Tibshirani (1993)). The “bootstrap population” consists of $\hat{\nu}_{tbh}$ animals with two groups. The first group consists of M_τ animals with capture history and covariates being the same as the observed ones in the data. The other group consists of animals that were not captured in the data, and there are $\hat{\nu}_{tbh} - M_\tau$ such animals. Assume that all animals in this “bootstrap population” have the same probability of being resampled. We draw $\hat{\nu}_{tbh}$ animals with replacement from this population to form a bootstrap sample. When an animal in the first group is selected, we record its associated capture times and covariates. (It is possible that in the bootstrap sample there are two animals with identical capture times and covariates.) Any animal selected from the second group represents a non-capture in the bootstrap sample and thus is excluded in the analysis. We remark that the number of animals selected from the first group varies with replications, but the expected value is equal to M_τ . For each bootstrap sample, a bootstrap estimate of ν can be obtained based on the data information of animals from the first group. After B replications, the bootstrap variance estimator of $\hat{\nu}_{tbh}$ is simply the sample variance of those B bootstrap estimates. The same replications can subsequently be used to construct a confidence interval using a

percentile method. For all submodels that will be discussed in Section 2.2, we are able to derive asymptotic variance formulas and thus a simple traditional approach with a log-transformation (Chao (1989)) can be used to construct a confidence interval for each submodel.

2.2. Submodels

(1) Model \mathcal{M}_{th}

Letting $\phi = 1$ in the system of equations (2.8d), we obtain

$$\sum_{i=1}^{M_\tau} \left[m_i - \frac{\gamma_i \Lambda_\tau}{1 - \exp(-\gamma_i \Lambda_\tau)} \right] \mathbf{Z}_i^+ = \mathbf{0}. \tag{2.11}$$

The population size estimator is then $\hat{\nu}_{th} = \sum_{i=1}^{M_\tau} 1/[1 - \exp(1 - \hat{\gamma}_i \hat{\Lambda}_\tau)]$ using (2.9). The asymptotic variance of the estimator can be derived following an approach of Huggins (1989). For notational simplicity, we use a reparametrization of $\Lambda_\tau = \tau \exp(\beta_0)$ and define $\boldsymbol{\beta}^+ = (\beta_0, \dots, \beta_p)'$ and $\gamma_i^+ = \exp[(\boldsymbol{\beta}^+)'\mathbf{Z}_i^+]$. The resulting asymptotic variance formula is

$$\widehat{\text{Var}}(\hat{\nu}_{th}) \approx \sum_{i=1}^{M_\tau} \frac{1 - P_i(\hat{\boldsymbol{\beta}}^+)}{[P_i(\hat{\boldsymbol{\beta}}^+)]^2} + \hat{D}'[I(\hat{\boldsymbol{\beta}}^+)]^{-1}\hat{D},$$

where $P_i(\hat{\boldsymbol{\beta}}^+) = 1 - \exp(-\hat{\gamma}_i^+ \tau)$, $\hat{D} = \sum_{i=1}^{M_\tau} [1 - P_i(\hat{\boldsymbol{\beta}}^+)]\hat{\gamma}_i^+ \tau \mathbf{Z}_i^+ / [P_i(\hat{\boldsymbol{\beta}}^+)]^2$ is the estimated first derivative of $\nu(\boldsymbol{\beta}^+) = \sum_{i=1}^{M_\tau} [1 - \exp(-\gamma_i^+ \tau)]^{-1}$ with respect to $\boldsymbol{\beta}^+$, and $I(\boldsymbol{\beta}^+)$ is the negative matrix of the second derivative of the conditional likelihood, i.e.,

$$I(\boldsymbol{\beta}^+) = -\frac{\partial^2 \log L_c}{\partial \boldsymbol{\beta}^{+2}} = \sum_{i=1}^{M_\tau} \left[\frac{\gamma_i^+ \tau}{P_i(\boldsymbol{\beta}^+)} - \frac{[1 - P_i(\boldsymbol{\beta}^+)](\tau \gamma_i^+)^2}{[P_i(\boldsymbol{\beta}^+)]^2} \right] \mathbf{Z}_i^+ \mathbf{Z}_i^{+'}.$$

When an estimated variance is obtained, we can apply a log-transformation (Chao (1989)) to obtain a confidence interval.

(2) Model \mathcal{M}_{bh}

For this special model, $\lambda_0(t) = \lambda$ and $\Lambda_t = \lambda t$ for any t . Reparametrizing $\lambda = \exp(\beta_0)$ and adding a constant covariate, we obtain from (2.5) that the equations for model \mathcal{M}_{bh} are

$$\sum_{i=1}^{M_\tau} [(m_i - 1) - \phi \gamma_i \lambda (\tau - t_{i1})] = 0, \tag{2.12a}$$

$$\sum_{i=1}^{M_\tau} \left[m_i - \frac{\gamma_i \lambda \tau}{1 - \exp(-\gamma_i \lambda \tau)} + (1 - \phi) \gamma_i \lambda (\tau - t_{i1}) \right] \mathbf{Z}_i^+ = \mathbf{0}. \tag{2.12b}$$

After the estimator $\hat{\beta}^+$ is obtained, the resulting population size $\hat{\nu}_{bh}$ has the same form as that given in Equation (2.9). Let $\hat{\theta}' = [\hat{\alpha}(\hat{\beta}^+)]'$ and the asymptotic variance formula for the estimator can be shown to be

$$\widehat{\text{Var}}(\hat{\nu}_{bh}) \approx \sum_{i=1}^{M_\tau} \frac{1 - P_i(\hat{\beta}^+)}{[P_i(\hat{\beta}^+)]^2} + \hat{D}'[I(\hat{\theta})]^{-1}\hat{D}, \tag{2.13}$$

where $P_i(\hat{\beta}^+)$ is defined in model \mathcal{M}_{th} , and

$$\hat{D} = \left[0 \quad \sum \frac{1 - P_i(\hat{\beta}^+)}{[P_i(\hat{\beta}^+)]^2} \hat{\gamma}_i^+ \tau \mathbf{Z}_i^{+'} \right]',$$

$$I(\theta) = \begin{bmatrix} \phi \sum \gamma_i^+ (\tau - t_{i1}) & \phi \sum \gamma_i^+ (\tau - t_{i1}) (\mathbf{Z}_i^+)' \\ \phi \sum \gamma_i^+ (\tau - t_{i1}) \mathbf{Z}_i^+ & \sum \left(\frac{\gamma_i^+ \tau}{P_i(\beta^+)} - \frac{[1 - P_i(\beta^+)] (\gamma_i^+ \tau)^2}{[P_i(\beta^+)]^2} \right) \mathbf{Z}_i^+ (\mathbf{Z}_i^+)' \end{bmatrix},$$

and all the indices are from 1 to M_τ .

Remark. It can be verified (Hwang (1997)) that the system of estimating functions given in (2.12) is a likelihood score with the optimal weight. Lin and Yip (1999) adopted a likelihood score approach, but they used another weight function. One of their equations is identical to (2.12a) but the other equation turns out to be

$$\sum_{i=1}^{M_\tau} \left[m_i - \gamma_i t_{i1} - \log \frac{1 - \exp(-\gamma_i \lambda \tau)}{1 - \exp[-\gamma_i \lambda (\tau - t_{i1})]} + (1 - \phi) \gamma_i \lambda (\tau - t_{i1}) \right] \mathbf{Z}_i^+ = \mathbf{0},$$

which is slightly different from (2.12b). The two approaches will be compared numerically in Section 4.

(3) Model \mathcal{M}_h

We can either let $\Lambda_\tau = \lambda \tau$ in model \mathcal{M}_{th} or let $\phi = 1$ in model \mathcal{M}_{bh} . Both approaches lead to the following equations:

$$\sum_{i=1}^{M_\tau} \left[m_i - \frac{\gamma_i \lambda \tau}{1 - \exp(-\gamma_i \lambda \tau)} \right] \mathbf{Z}_i^+ = \mathbf{0}.$$

The asymptotic variance can be similarly obtained as in model \mathcal{M}_{th} .

(4) Model \mathcal{M}_{tb}

Substituting $\gamma_i = 1$ (i.e., $\beta = \mathbf{0}$) and $\sum_{i \in M_t^*} \gamma_i = M_t$ in Equation (2.8), we have the following equations for model \mathcal{M}_{tb} :

$$\frac{K_\tau}{\phi} - \int_0^\tau \frac{M_t}{M_\tau/[1 - \exp(-\Lambda_\tau)] + (\phi - 1)M_t} dN_t = 0,$$

$$\Lambda_\tau = \int_0^\tau \frac{1}{M_\tau/[1 - \exp(-\Lambda_\tau)] + (\phi - 1)M_t} dN_t.$$

The population size estimator for homogeneous model is reduced to $\hat{\nu}_{tb} = M_\tau/[1 - \exp(-\hat{\Lambda}_\tau)]$.

(5) Model \mathcal{M}_b

Setting $\gamma_i = 1$ (i.e., $\beta = \mathbf{0}$) in model \mathcal{M}_{bh} , we have

$$\sum_{i=1}^{M_\tau} \left[m_i - \frac{\lambda\tau}{1 - \exp(-\lambda\tau)} + (1 - \phi)\lambda(\tau - t_{i1}) \right] = 0,$$

$$\sum_{i=1}^{M_\tau} [(m_i - 1) - \phi\lambda(\tau - t_{i1})] = 0.$$

(6) Model \mathcal{M}_t

We can either use the equations for model \mathcal{M}_{tb} by setting $\phi = 1$ or use those for model \mathcal{M}_{th} by setting $\gamma_i = 1$. Then the equation for model \mathcal{M}_t reduces to

$$A_\tau = \frac{N_\tau}{M_\tau/[1 - \exp(-\Lambda_\tau)]}.$$

3. Simulation

A limited simulation study was carried out to investigate the performance of the proposed estimators. Here we focus on the heterogeneous models \mathcal{M}_{th} and \mathcal{M}_{tbh} . The true population size ν was fixed to be 400. The heterogeneity effect was modeled by two covariates: Z_1 and Z_2 , where Z_1 corresponds to a discrete covariate (e.g., sex) and Z_2 corresponds to a continuous covariate (e.g., weight). The intensity function for the most general model \mathcal{M}_{tbh} was assumed to be $\lambda_i(t) = \lambda_0(t) \exp(\beta_1 Z_{i1} + \beta_2 Z_{i2} + \alpha X_i(t))$, where $Z_{i1} = 0$ for $i = 1$ to 200, and $Z_{i1} = 1$ for $i = 201$ to 400, $\{Z_{i2}, i = 1, \dots, \nu\}$ is a random sample from a normal distribution with mean 20 and variance 4, and $X_i(t)$ denotes the prior capture history (yes or no). The true parameters are $(\beta_1, \beta_2) = (1, -0.02)$. A positive β_1 implies that the sex taking value 1 is more catchable, whereas a negative β_2 means that the catchability decreases with weight. We considered the following six combinations of two time-varying functions of $\lambda_0(t)$ and three values of ϕ . The two types of time-varying functions of $\lambda_0(t)$ are $\lambda_0(t) = 1/(t + 0.5)$ and $\lambda_0(t) = 1 + \sin(2\pi t)$; three values of ϕ are 1, 1.2 and 0.8. The reader is referred to Hwang (1997) for more simulation results.

The results for three stopping times ($\tau = 1, 2, 4$) are shown in Table 2. For each combination of ϕ and $\lambda_0(t)$, capture-recapture data were generated. Then for each generated data set, the estimates under four models \mathcal{M}_{th} , \mathcal{M}_{bh} , \mathcal{M}_{tb}

and \mathcal{M}_{tbh} were calculated, along with their estimated s.e.'s and 95% confidence intervals. The four estimates are designated as $\hat{\nu}_{th}$, $\hat{\nu}_{bh}$, $\hat{\nu}_{tb}$ and $\hat{\nu}_{tbh}$, respectively, in Table 2. For model \mathcal{M}_{tbh} , the estimated s.e. was calculated by the bootstrap method with 500 replications, and the confidence interval was constructed by a percentile method. For the other estimators, the estimated s.e.'s were obtained by their asymptotic formulas. The iterative steps for computing $\hat{\nu}_{tbh}$ failed to converge for some trials. We excluded those data sets until 200 data sets had been generated. The resulting 200 estimates and their estimated s.e.'s were averaged to give the results of "average estimate" and "average estimated s.e." in Table 2. The proportion of divergence is given in the last column of the table. Here "divergence" means that either the estimates increased without a limit or the iterations failed to converge within 1000 steps. Divergence might be due to insufficient capture information or improper choices of initial values. These failure rates are generally negligible as more data become available. Based on the resulting 200 estimates, the sample s.e. and the sample root mean squared error (RMSE) for each estimator were calculated. The percentage of the 200 simulated data sets in which the 95% confidence intervals covered the true value was also recorded.

Table 2 shows the simulation results for the case of $\lambda_0(t) = 1/(t + 0.5)$ for $\phi = 1, 1.2$ and 0.8 , respectively. Results for $\lambda_0(t) = 1 + \sin(2\pi t)$ are generally consistent. In the table, we also list the averages of M_τ (the number of distinct individuals captured in the experiment) and N_τ (the total number of captures).

(a) Model $\mathcal{M}_{th}(\phi = 1)$:

Since the underlying model is \mathcal{M}_{th} , the estimator $\hat{\nu}_{th}$ derived under this model has the smallest bias and RMSE. The coverage probabilities of the 95% confidence interval associated with $\hat{\nu}_{th}$ are close to the anticipated nominal level. The estimator $\hat{\nu}_{bh}$ yields severe negative bias and most confidence intervals do not cover the true parameter. The estimator $\hat{\nu}_{tb}$ that does not account for heterogeneity is also negatively biased and the coverage probabilities are between 70% and 80%. When the capture rate is low, as in the case of $\tau = 1$, for which only a small fraction of individuals is captured, the estimator $\hat{\nu}_{tbh}$ has the largest variation due to estimating all effects. For this case, we feel the data information is insufficient for making a stable estimation of all effects. When data improve, this estimator $\hat{\nu}_{tbh}$ works well, although its variation is unavoidably the largest.

The s.e. estimates (with the heading "average estimated s.e." in Table 2) for all four estimators are generally comparable to sample s.e. (with the heading "sample s.e."). The percentile-based confidence interval associated with $\hat{\nu}_{tbh}$ also performs satisfactorily as regards coverage probability.

(b) Model $\mathcal{M}_{tbb}(\phi = 1.2 \text{ and } 0.8)$

Except for the correct estimator $\hat{\nu}_{tbb}$, all other estimators that do not consider all effects are biased. The estimator that ignores the behavioural response effect (i.e., $\hat{\nu}_{th}$) severely underestimates in the trap-happy cases ($\phi = 1.2$), but overestimates in the trap-shy cases ($\phi = 0.8$). The two estimators $\hat{\nu}_{bh}$ and $\hat{\nu}_{tb}$ are biased downwards in both trap-happy and trap-shy cases. The estimator $\hat{\nu}_{bh}$ that does not account for time variation performs worst because it has the largest bias and RMSE as well as the lowest coverage probabilities. When there are enough captures to estimate all the parameters well, the estimator $\hat{\nu}_{tbb}$ is preferable in terms of bias, RMSE and coverage probability.

The s.e. estimates based on a bootstrap method tend to overestimate in most cases. A possible explanation is that there were some bootstrap samples with few captures. Therefore, extremely large population size estimates were obtained due to little capture information in those samples. The coverage probabilities of the associated confidence interval are thus slightly higher than the nominal level.

In summary, when there is enough capture and recapture information to allow for the stable estimation of the effects of various covariates, the most general model \mathcal{M}_{tbb} that considers time dependence, behavioural response and heterogeneous covariates can be recommended.

4. Numerical Example

The house mouse data considered in this section were originally collected by Coulombe in his 1965 master's thesis, University of California, Los Angeles, and given as an illustrative example in the program CAPTURE (Rexstad and Burnham (1991)). Otis et al. (1978, pp.62-64) and Huggins (1989) provided some discussions about this data set. Totally 173 mice were caught and associated with two covariates: age (juvenile, semi-adult or adult) and sex (male or female). We excluded two records in the analysis because the covariates for the two mice were missing. These 171 distinct mice were caught out of 581 captures in 10 trapping occasions. Since there were only 8 juveniles, we combined the groups of juveniles and semi-adults into a "non-adult" class. The data consist of 77 non-adults (45 males, 32 females) and 94 adults (41 males, and 53 females). On average, the capture frequencies for males and females were 3.08 and 3.72, respectively; the capture frequencies for adults and non-adults were 3.81 and 2.90, respectively. Since there were ten trapping occasions, it would be interesting to compare the results under discrete-time and continuous-time models with time units $1, \dots, 10$.

We first discuss the approaches that do not consider the covariates. Otis et al. (1978) adopted discrete-time models and indicated that the most likely model

is model \mathcal{M}_{th} and the next most likely model is model \mathcal{M}_{tbh} , using a model selection procedure. At that time, no estimators had yet been available for these two models and they concluded that the whole population was caught based on a detailed investigation of the data. The discrete-time sample coverage method for model \mathcal{M}_{th} (Chao et al. (1992)) yields a population size estimate of 174 with an estimated s.e. 1.0; a similar continuous-time approach (Chao and Lee (1993)) gives an estimate of 172 with an estimated s.e. 3.3. All these procedures without using covariates imply that almost no or only few individuals were missed in the experiment.

If the heterogeneity can be modeled via observable covariates of the individuals in a continuous-time framework, then model \mathcal{M}_{tbh} can be implemented with an intensity function $\lambda_i(t) = \lambda_0(t) \exp(\beta_1 Z_{i1} + \beta_2 Z_{i2} + \alpha X_i(t))$, where $Z_{i1} = I$ [the i th individual is a male], $Z_{i2} = I$ [the i th individual is an adult], and $X_i(t) = I$ [the i th individual has been captured in $(0, t)$]. We remark that for this data set five samples were conducted in the mornings and the others in the evenings. This covariate was not included in our analysis because it can be explained in the time-varying effect in model \mathcal{M}_{tbh} . The parameter estimates are $\hat{\beta}_1 = -0.155$ (s.e. 0.08), $\hat{\beta}_2 = 0.329$ (s.e. 0.08) and $\hat{\alpha} = 0.687$ (s.e. 0.253). Hence it implies that females are more catchable and adults have higher intensity than do non-adults, which is consistent with the data. The resulting population size estimate is $\hat{\nu}_{tbh} = 206$ (s.e. 14.8 based on 500 bootstrap replications). The variation is inevitably larger because of estimating more parameters. For this data set, a large portion of individuals was caught and thus the results for the most general model \mathcal{M}_{tbh} merit consideration.

For this data set, Otis et al. (1978) selected model \mathcal{M}_{th} as the most parsimonious model. Dropping the behavioural response effect and fitting model \mathcal{M}_{th} with age and sex gives an estimate of $\hat{\nu}_{th} = 179$ with an estimated s.e. of 3.3. Model \mathcal{M}_h shows the same result as that for model \mathcal{M}_{th} .

Huggins (1989) considered a linear logistic relationship to the covariates and obtained an estimate of 177 (s.e. 3.6). Applying the likelihood score method proposed in Lin and Yip (1999) to this data for model \mathcal{M}_{bh} , we have $\hat{\beta}_1 = -0.170$ (s.e. 0.075), $\hat{\beta}_2 = 0.272$ (s.e. 0.073) and $\hat{\alpha} = -0.067$ (s.e. 0.09), which subsequently gives a population size estimate of 176 (s.e. 5.0). Using our approach for fitting a model \mathcal{M}_{bh} , we have $\hat{\beta}_1 = -0.175$ (s.e. 0.091), $\hat{\beta}_2 = 0.293$ (s.e. 0.094) and $\hat{\alpha} = -0.022$ (s.e. 0.113). The resulting estimate is $\hat{\nu}_{bh} = 179$ (s.e. 3.9). Therefore, the results for continuous-time models \mathcal{M}_h , \mathcal{M}_{th} and \mathcal{M}_{bh} are very close to the previous estimates given in Huggins (1989) and Lin and Yip (1999). All these results incorporating the observable heterogeneous covariates conclude that there were some mice uncaptured in the experiment.

Table 2. Comparison of Various Estimators for Models \mathcal{M}_{th} and \mathcal{M}_{tbb} , 200 Trials, $\nu = 400$, $\lambda_0(t) = 1/(t + 0.5)$.

Stopping Time(τ)	Estimator	Average Estimate	Sample s.e.	Average Estimated s.e.	Sample RMSE	95% C. I. Coverage (%)	Divergence Rate (%)
Model $\mathcal{M}_{th}(\phi = 1)$							
$\tau = 1$ $M_\tau = 277$ $N_\tau = 550$	$\hat{\nu}_{th}$	405	24.3	26.1	24.7	97.5	
	$\hat{\nu}_{bh}$	318	15.2	13.0	83.6	5.0	
	$\hat{\nu}_{tb}$	375	54.0	51.6	59.4	86.0	
	$\hat{\nu}_{tbb}$	418	60.6	68.4	63.2	98.5	3.4
$\tau = 2$ $M_\tau = 321$ $N_\tau = 802$	$\hat{\nu}_{th}$	400	9.2	15.8	15.5	95.0	
	$\hat{\nu}_{bh}$	336	9.2	5.7	64.7	0	
	$\hat{\nu}_{tb}$	375	23.4	21.8	34.0	80.0	
	$\hat{\nu}_{tbb}$	403	26.4	28.4	26.5	97.5	1.0
$\tau = 4$ $M_\tau = 350$ $N_\tau = 1093$	$\hat{\nu}_{th}$	399	10.3	10.5	10.3	95.5	
	$\hat{\nu}_{bh}$	354	6.1	2.4	46.4	0	
	$\hat{\nu}_{tb}$	381	12.5	11.8	22.7	71.5	
	$\hat{\nu}_{tbb}$	400	14.8	15.6	14.8	95.0	
Model $\mathcal{M}_{tbb}(\phi = 1.2)$							
$\tau = 1$ $M_\tau = 278$ $N_\tau = 601$	$\hat{\nu}_{th}$	375	21.2	20.4	32.6	75.0	
	$\hat{\nu}_{bh}$	318	16.1	12.9	83.2	4.5	
	$\hat{\nu}_{tb}$	371	43.4	45.5	52.3	87.5	
	$\hat{\nu}_{tbb}$	410	51.3	58.4	52.2	96.0	0.5
$\tau = 2$ $M_\tau = 321$ $N_\tau = 898$	$\hat{\nu}_{th}$	380	13.0	12.5	23.8	70.0	
	$\hat{\nu}_{bh}$	337	9.1	5.9	63.7	0	
	$\hat{\nu}_{tb}$	377	22.2	21.8	32.2	82.0	
	$\hat{\nu}_{tbb}$	404	25.8	27.0	26.0	97.0	0.5
$\tau = 4$ $M_\tau = 350$ $N_\tau = 1243$	$\hat{\nu}_{th}$	387	8.9	8.4	15.3	79.0	
	$\hat{\nu}_{bh}$	355	6.4	2.5	45.0	0	
	$\hat{\nu}_{tb}$	383	12.4	11.8	21.0	76.5	
	$\hat{\nu}_{tbb}$	402	14.0	14.9	14.1	96.0	
Model $\mathcal{M}_{tbb}(\phi = 0.8)$							
$\tau = 1$ $M_\tau = 278$ $N_\tau = 497$	$\hat{\nu}_{th}$	452	35.3	35.5	62.8	59.0	
	$\hat{\nu}_{bh}$	319	17.1	13.6	82.4	6.0	
	$\hat{\nu}_{tb}$	391	67.4	69.3	80.1	92.5	
	$\hat{\nu}_{tbb}$	428	79.8	79.0	73.0	98.5	13.4
$\tau = 2$ $M_\tau = 321$ $N_\tau = 710$	$\hat{\nu}_{th}$	434	22.6	21.7	41.0	52.5	
	$\hat{\nu}_{bh}$	336	9.4	5.7	64.3	0	
	$\hat{\nu}_{tb}$	377	26.6	23.1	35.3	83.5	
	$\hat{\nu}_{tbb}$	405	30.3	30.9	30.6	98.0	2.9
$\tau = 4$ $M_\tau = 351$ $N_\tau = 949$	$\hat{\nu}_{th}$	425	14.3	14.5	28.7	45.5	
	$\hat{\nu}_{bh}$	355	6.8	2.4	45.6	0	
	$\hat{\nu}_{tb}$	382	13.0	12.1	22.2	76.0	
	$\hat{\nu}_{tbb}$	402	15.8	17.2	15.9	94.5	0.5

Acknowledgements

The authors would like to thank two reviewers for helpful comments and suggestions. This work is based in part on the Ph.D. Thesis of the first author under supervision of the second author. Part of this research was supported by the National Science Council of Taiwan under contracts NSC 89-2118-M035-007 (for Hwang) and NSC 85-2121-M007-004 and NSC 86-2115-M007-033 (for Chao).

References

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-726.
- Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.
- Becker, N. G. (1984). Estimating population size from capture-recapture experiments in continuous time. *Austral. J. Statist.* **26**, 1-7.
- Becker, N. G. and Heyde, C. C. (1990). Estimating population size from multiple recapture experiments. *Stochastics Process. Appl.* **36**, 77-83.
- Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics* **45**, 427-438.
- Chao, A., Lee, S.-M. and Jeng S.-L. (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* **48**, 201-216.
- Chao, A. and Lee, S.-M. (1993). Estimating population size for continuous time capture-recapture models via sample coverage. *Biometrical J.* **35**, 29-45.
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.
- Crowder, M. J., Kimber, A. C., Smith, R. L. and Sweeting, T. J. (1991). *Statistical Analysis of Reliability Data*. Chapman and Hall, London.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47**, 663-685.
- 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.
- Hwang, W.-H. (1997). Estimating population size for continuous time capture-recapture experiments. Ph.D. Thesis, Institute Statistics, National Tsing Hua University, Taiwan.
- Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Statist.* **27**, 887-906.
- Lin, D. Y. and Yip, P. (1999). Parametric regression models for continuous time removal and recapture studies. *J. Roy. Statist. Soc. Ser. B* **61**, 401-411.
- Nayak, T. K. (1988). Estimating population size by recapture sampling. *Biometrika* **75**, 113-120.
- Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945-965.

- Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildl. Monog.* **62**, 1-135.
- Pollock, K. H. (1991). Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future. *J. Amer. Statist. Assoc.* **86**, 225-238.
- 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.
- Rexstad, E. and Burnham, K. P. (1991). *User's Guide for Interactive Program CAPTURE*. Colorado Cooperative Fish and Wildlife Research Unit, Fort Collins.
- Scholz, F. W. (1980). Towards a unified definition of maximum likelihood. *Canad. J. Statist.* **8**, 193-203.
- Schwarz C. J. and Seber, G. A. F. (1999). A review of estimating animal abundance III. *Statist. Sci.* **14**, 427-456.
- Wilson, K. R. and Anderson, D. R. (1995). Continuous-time capture-recapture population estimation when capture probabilities vary over time. *Environ. Ecol. Statist.* **2**, 55-69.
- Yip, P. and Chao, A. (1996). Estimating population size from capture-recapture studies via sample coverage and estimating functions. *Stoch. Models* **12**, 17-35.
- Yip, P., Huggins, R. M. and Lin, D. Y. (1996). Inference for capture-recapture experiments in continuous time with variable capture rates. *Biometrika* **83**, 477-483.

Department of Statistics, Feng-Chia University, Tai-Chung, Taiwan 40724.

E-mail: whhwang@fcu.edu.tw

Institute of Statistics, National Tsing Hua University, Hsin-Chu, Taiwan 30043.

E-mail: chao@stat.nthu.edu.tw

(Received February 2001; accepted April 2002)