A Unified Likelihood-based Approach for Estimating Population Size in Continuous-Time Capture-Recapture Experiments with Frailty

Liqun Xi[1], Paul S. F. Yip[1,*], Ray Watson[2]

[1]Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong

[2]Department of Mathematics and Statistics, University of Melbourne, Australia

Abstract

A unified likelihood-based approach is proposed to estimate population size for a continuous-time closed capture-recapture experiment with frailty. The frailty model allows the capture intensity to vary with individual heterogeneity, time and behavioral response. The individual heterogeneity effect is modelled as being gamma distributed. The first-capture and recapture intensities are assumed to be in constant proportion but may otherwise vary arbitrarily through time. The approach is also extended to capture-recapture experiments with possible random removals. Simulation studies are conducted to examine the performance of the proposed estimators. By asymptotic efficiency comparison and simulation studies, the proposed estimators have been shown to be superior than their discrete-time model counterparts in genuine continuous time capture-recapture experiment.

Keywords: behavioral response, capture-recapture, continuous vs discrete, gamma distribution, frailty, heterogeneity, population size.

*Corresponding author: Dr. Paul Yip, Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, E-mail: sfpyip@hku.hk

1 Introduction

Population size estimation is an important procedure in many fields, such as ecology, epidemiology and software reliability (Pollock, 1991). Capture-recapture method is often used to estimate population size (Otis et al.,1978). In capture-recapture experiments, there are three major sources of variation in the catchability of individuals: heterogeneity among individuals; time variation; and behavioral response to capture. The models considered are classified according to these three factors, and are referred to as $\mathcal{M}_{\mathrm{h}}$, $\mathcal{M}_{\mathrm{t}}$, $\mathcal{M}_{\mathrm{b}}$, $\mathcal{M}_{\mathrm{ht}}$, $\mathcal{M}_{\mathrm{hb}}$, $\mathcal{M}_{\mathrm{tb}}$ and $\mathcal{M}_{\mathrm{htb}}$, where the subscript "h", "t", "b" denote "heterogeneity", "time-varying" and "behavioral-response" respectively. The models with heterogeneity are the most challenging.

In capture-recapture experiments, capture efforts may be made at a limited number of occasions or on a continuous basis. We refer to these two settings as discrete-time and continuous-time respectively. In discrete-time settings, the above models and corresponding estimation methods have been fully developed, and a wide variety of parametric and nonparametric approaches have been applied to estimate population size (Chao, 2001). Compared with discrete-time models, there has been relatively little published research for the continuous-time counterparts. Earlier work includes the papers by Craig (1953) and Darroch (1958) for a homogeneous population. Becker (1984) and Yip, Fong & Wilson (1993) have established a counting process framework to study capture-recapture experiment in continuous time. Assume there are ν independent individuals, indexed by $1, 2, \ldots, \nu$. We also assume that the capture-recapture experiment period $[0, \tau]$ (where τ denotes the duration of the experiment) is relatively short so that the population size remains constant over the course of the experiment. Let $N_i(t)$ denote the number of times the ith individual

2

has been caught in $[0, t]$. Each $\{N_i(t); 0 \le t \le \tau\}$ is a continuous-time counting process with an intensity $\lambda_i(t)$. The intensity for the ith individual, $\lambda_i(t)$, is defined by $\lambda_i(t)dt = \Pr(dN_i(t) = 1 \,|\, \mathcal{F}_{t-})$, where $\mathcal{F}_{t-}$ denotes the capture history up to time t but not including t.

A multiplicative form of the most general model $\mathcal{M}_{\mathrm{htb}}$ is given by

$$\lambda_i(t) = \beta \gamma_i \, \phi^{r_i(t)} \lambda_0(t) \tag{1}$$

where $r_i(t) = \mathrm{I}(N_i(t) > 0)$, i.e. a capture indicator: it is zero until the ith individual is first captured, and 1 after the ith individual is captured; $\lambda_0(t)$ is an arbitrary non-negative time-varying function in $[0, \tau]$; $\gamma_1, \ldots, \gamma_\nu$ and ϕ are positive values, representing the effects of heterogeneity and behavioral response respectively; and β represents an overall capture intensity.

To avoid identifiability problems, we assume that in each case the parameter reverts to unity in the case of homogeneity. Thus in the homogeneous case $\phi = 1$; and the capture intensity $\lambda_0(t)$ is assumed to have average value 1. Further, we assume the heterogeneity effects $\{\gamma_1, \ldots, \gamma_\nu\}$ to be sampled from a gamma distribution with mean 1: $\gamma_i \overset{\mathrm{d}}{=} \mathrm{Ga}(\alpha, \alpha)$. The case $\gamma = 1$ represents homogeneity; and so departure of γ_i from 1 indicates departure from the average population behavior. The extent of the population heterogeneity is indicated by α, since $\mathrm{var}(\gamma_i) = 1/\alpha$.

All sub-models $\mathcal{M}_{\mathrm{ht}}$, $\mathcal{M}_{\mathrm{hb}}$, $\mathcal{M}_{\mathrm{tb}}$, $\mathcal{M}_{\mathrm{h}}$, $\mathcal{M}_{\mathrm{t}}$ $\mathcal{M}_{\mathrm{b}}$ and $\mathcal{M}_0$ are seen to be a particular case of (1), i.e.

Model $\mathcal{M}_{\mathrm{ht}}$: $\lambda_i(t) = \beta \gamma_i \lambda_0(t)$;

Model $\mathcal{M}_{\mathrm{hb}}$: $\lambda_i(t) = \beta \gamma_i \, \phi^{r_i(t)}$;

Model $\mathcal{M}_{\mathrm{tb}}$: $\lambda_i(t) = \beta \phi^{r_i(t)} \lambda_0(t)$;

Model $\mathcal{M}_{\mathrm{h}}$: $\lambda_i(t) = \beta\gamma_i$;

Model $\mathcal{M}_{\mathrm{t}}$: $\lambda_i(t) = \beta\lambda_0(t)$;

Model $\mathcal{M}_{\mathrm{b}}$: $\lambda_i(t) = \beta\phi^{r_i(t)}$;

Model $\mathcal{M}_0$: $\lambda_i(t) = \beta$.

For the models without heterogeneity (i.e. $\mathcal{M}_{\mathrm{tb}}$, $\mathcal{M}_{\mathrm{t}}$, $\mathcal{M}_{\mathrm{b}}$), the relevant estimation procedures are well developed; see Chao (2001). However, heterogeneity among the individuals remains the most difficult part in estimating population size. Becker (1984) developed a martingale estimator for model $\mathcal{M}_{\mathrm{ht}}$ where $\{\gamma_1, \ldots, \gamma_\nu\}$ are assumed to be sampled from a gamma distribution. However, this procedure can result in negative estimates, which motivated Chao & Lee (1993) and Yip & Chao (1996) to derive new estimators for model $\mathcal{M}_{\mathrm{ht}}$ using the sample coverage and estimating function approach in which $\{\gamma_1, \ldots, \gamma_\nu\}$ are regarded as fixed parameters. There have been no estimators given in the literature for models $\mathcal{M}_{\mathrm{hb}}$ and $\mathcal{M}_{\mathrm{htb}}$.

This paper proposes a unified likelihood-based approach to estimate the population size ν for models $\mathcal{M}_{\mathrm{h}}$, $\mathcal{M}_{\mathrm{ht}}$, $\mathcal{M}_{\mathrm{hb}}$, $\mathcal{M}_{\mathrm{htb}}$, where the heterogeneity effects $\{\gamma_1, \ldots, \gamma_\nu\}$ are assumed to be sampled from a gamma distribution, i.e. the so-called frailty model (Andersen et al., 1993, Chapter 9). The estimators of ν for model $\mathcal{M}_{\mathrm{h}}$ and model $\mathcal{M}_{\mathrm{ht}}$ are found to be the same and not depend on the capture times. The proposed approach is also easily extended to capture-recapture experiments with possible random removals. We compare the proposed models and approach with the corresponding discrete-time models and the existing discrete-type estimators. For a homogeneous population the gain on the continuous-time models is marginal, whereas for a heterogeneous population, the gain is significant.

In Section 2, the estimation procedures are presented. In Section 3, simulation

studies are conducted to examine the performance of the proposed estimators. In Section 4, the advantage of the proposed continuous-time models and approach are investigated by simulation study and asymptotic efficiency in comparison with the corresponding discrete-time models and existing discrete-type estimators. The circumstance in which the proposed methods have some application is recommended.

2 Estimation procedures

Let n denote the number of distinct individuals captured over the course of the experiment. Without loss of generality, label the captured individuals as $1, 2, \ldots, n$, and those uncaptured as $n+1, n+2, \ldots, \nu$. Let $m_i = N_i(\tau)$ denote the total number of times the ith individual has been caught, and if $m_i > 0$, denote the capture times as $t_{i1}, \ldots, t_{im_i}$. Let $N(t) = \sum_{i=1}^{\nu} N_i(t)$ denote the total number of captures by time t.

2.1 Model $\mathcal{M}_{\mathrm{htb}}$

For model $\mathcal{M}_{\mathrm{htb}}$, we have $\lambda_i(t) = \beta \gamma_i \, \phi^{r_i(t)} \lambda_0(t)$, where $\gamma_i \stackrel{\mathrm{d}}{=} \mathrm{Ga}(\alpha, \alpha)$. However, for convenience of derivation, we reparameterize the model as follows:

$$\lambda_i(t) = \rho_i \, \phi^{r_i(t)} \omega(t)$$

where $\rho_i \stackrel{\mathrm{d}}{=} \mathrm{Ga}(\alpha, 1)$ and $\omega(t)$ is an arbitrary non-negative time-varying function.

For the ith individual, given ρ_i, when $m_i > 0$, the likelihood function is proportional to

$$\rho_i \, \omega(t_{i1}) \, e^{-\rho_i \Omega(t_{i1})} \left(\prod_{j=2}^{m_i} \phi \rho_i \omega(t_{ij}) \right) e^{-\rho_i \phi (\Omega(\tau) - \Omega(t_{i1}))}$$

$$= \left(\prod_{j=1}^{m_i} \omega(t_{ij}) \right) \rho_i^{m_i} \, \phi^{m_i - 1} e^{-\rho_i \left(\Omega(t_{i1}) + \phi(\Omega(\tau) - \Omega(t_{i1})) \right)},$$

5

where $\Omega(t) = \int_0^t \omega(u)du$ denotes the cumulative intensity function; and $\Omega(\tau)$ is the value of Ω at termination τ; see Andersen et al. (1993). We define $\Omega_i^* = \phi\Omega(\tau)+(1-\phi)\Omega(t_{i1})$. Then, since $\rho_i \overset{d}{=} \mathrm{Ga}(\alpha, 1)$, the likelihood function for the ith individual is given by

$$
\begin{aligned}
L_i \quad &\propto \quad \int_0^\infty \left(\prod_{j=1}^{m_i} \omega(t_{ij})\right) \rho_i^{m_i}\, \phi^{m_i-1}\, e^{-\rho_i\Omega_i^*}\, \frac{\rho_i^{\alpha-1}e^{-\rho_i}}{\Gamma(\alpha)}\, d\rho_i \\
&= \quad \left(\prod_{j=1}^{m_i} \omega(t_{ij})\right) \phi^{m_i-1}\, \frac{\alpha^{[m_i]}}{(1+\Omega_i^*)^{\alpha+m_i}}, \qquad i = 1,\ldots,n; \qquad (2)
\end{aligned}
$$

where $x^{[k]} = x(x+1)\cdots(x+k-1)$. For those individuals uncaptured in the whole experiment, all the likelihood functions are same, equal to $(1+\Omega(\tau))^{-\alpha}$, and denoted by L_0.

Therefore, the likelihood function based on the capture history is given by

$$
\begin{aligned}
L(\nu, \alpha, \phi, \Omega) &\propto \frac{\nu!}{(\nu - n)!}\left(\prod_{i=1}^n L_i\right) L_0^{\nu-n} \\
&= \frac{\nu!}{(\nu - n)!}\prod_{i=1}^n \left(\left(\prod_{j=1}^{m_i} \omega(t_{ij})\right) \phi^{m_i-1}\frac{\alpha^{[m_i]}}{(1+\Omega_i^*)^{\alpha+m_i}}\right)\left(\frac{1}{1+\Omega(\tau)}\right)^{\alpha(\nu-n)}.
\end{aligned}
$$

And the log-likelihood function is given by:

$$
\begin{aligned}
\log L(\nu, \alpha, \phi, \Omega) &= \kappa + \sum_{i=1}^n \log(\nu - i + 1) + \sum_{i=1}^n\sum_{j=1}^{m_i} \omega(t_{ij}) - \alpha(\nu - n)\log(1+\Omega(\tau)) \\
&\quad + \sum_{i=1}^n \left((m_i-1)\log\phi - (\alpha+m_i)\log(1+\Omega_i^*) + \sum_{j=1}^{m_i}\log(\alpha+j-1)\right).
\end{aligned}
$$

Taking derivatives of the log-likelihood function with respect to ν, α and ϕ and equating them to zero, gives

$$
\frac{\partial \log L}{\partial \nu} = \sum_{i=1}^n \frac{1}{\nu-i+1} - \alpha\log(1+\Omega(\tau)) = 0, \qquad (3)
$$

$$
\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^n\sum_{j=1}^{m_i} \frac{1}{\alpha+j-1} - \sum_{i=1}^n \log(1+\Omega_i^*) - (\nu-n)\log\left(1+\Omega(\tau)\right) = 0, \qquad (4)
$$

6

$$\frac{\partial \log L}{\partial \phi} = \frac{Z - n}{\phi} - \sum_{i=1}^{n} \frac{(\alpha + m_i)(\Omega(\tau) - \Omega(t_{i1}))}{1 + \Omega_i^*} = 0, \tag{5}$$

where $Z = \sum_{i=1}^{\nu} m_i$. As for Ω, following the approach of Andersen et al. (1993, Section IV.1.5), we use the MLE here in the broad sense of Kiefer & Wolfowitz (1956). To maximize the likelihood, Ω must be a jump function with jumps at the observed capture times only. Assume Ω has a jump θ_k at $t_{(k)}$, where $t_{(1)} < \cdots < t_{(d)}$ denotes the ordered distinct capture times. More precisely

$$\Omega(t) = \sum_{k=1}^{d} \theta_k \, \mathrm{I}(t_{(k)} \le t).$$

Substituting this expression into the log-likelihood function and taking derivatives with respect to θ_k $(k = 1, \ldots, d)$ gives

$$\frac{\partial \log L}{\partial \theta_k} = \frac{dN(t_{(k)})}{\theta_k} - \sum_{i=1}^{n} \frac{(\alpha + m_i)\, \phi^{\xi_i(t_{(k)})}}{1 + \Omega_i^*} - \frac{\alpha(\nu - n)}{1 + \Omega(\tau)}, \tag{6}$$

where $\xi_i(t) = \mathrm{I}(t_{i1} < t)$. Equating this derivative to zero, gives the equation

$$\theta_k = \left\{ \sum_{i=1}^{n} \frac{(\alpha + m_i)\, \phi^{\xi_i(t_{(k)})}}{1 + \Omega_i^*} + \frac{\alpha(\nu - n)}{1 + \Omega(\tau)} \right\}^{-1} dN(t_{(k)}) \tag{7}$$

for $k = 1, \ldots, d$. For uncaptured individuals, let $\xi_i(t) = 0$ and $\Omega_i^* = \Omega(\tau)$, then (7) can be written as

$$d\Omega(t) = \left\{ \sum_{i=1}^{\nu} \frac{(\alpha + m_i)\, \phi^{\xi_i(t)}}{1 + \Omega_i^*} \right\}^{-1} dN(t). \tag{8}$$

Given $\rho_1, \rho_2, \ldots, \rho_\nu$, the Nelson-Aalen estimator of $d\Omega(t)$ in this case would be

$$d\Omega(t) = \left\{ \sum_{i=1}^{\nu} \rho_i \phi^{\xi_i(t)} \right\}^{-1} dN(t);$$

see Nelson (1972) and Aalen (1978). The expression (8) corresponds to this Nelson-Aalen estimating equation but with ρ_i replaced by the posterior expectation $\mathrm{E}(\rho_i \mid \mathcal{F}_\tau^{(i)})$, where $\mathcal{F}_\tau^{(i)}$ is the capture history up to time τ for the ith individual. The posterior

distribution for ρ_i is given by

$$\pi(\rho_i \mid \mathcal{F}_\tau^{(i)}) \propto L(\mathcal{F}_\tau^{(i)} \mid \rho_i) \, \pi(\rho_i)$$

$$\propto \left(\prod_{j=1}^{m_i} \omega(t_{ij})\right) \rho_i^{m_i} \, \phi^{m_i-1} \, e^{-\rho_i \Omega_i^*} \, \frac{\rho_i^{\alpha-1} e^{-\rho_i}}{\Gamma(\alpha)}$$

$$\propto \rho_i^{\alpha+m_i-1} e^{-(1+\Omega_i^*)\rho_i}$$

that is, $\rho_i \mid \mathcal{F}_\tau^{(i)} \overset{\mathrm{d}}{=} \mathrm{Ga}(\alpha + m_i, 1 + \Omega_i^*)$. Therefore

$$\mathrm{E}(\rho_i \mid \mathcal{F}_\tau^{(i)}) = \frac{\alpha + m_i}{1 + \Omega_i^*}.$$

The maximum likelihood estimates $\hat{\nu}$, $\hat{\alpha}$, $\hat{\phi}$, $\hat{\theta}_1, \ldots, \hat{\theta}_d$ can be obtained by solving the equations (3), (4), (5) and (7). We solve the equations by iteration: let $\boldsymbol{\Psi} = (\hat{\nu}, \hat{\alpha}, \hat{\phi}, \hat{\theta}_1, \ldots, \hat{\theta}_d)$; and write (3), (4), (5) and (7) in the form: $\boldsymbol{\Psi} = \mathbf{f}(\boldsymbol{\Psi})$, where $\mathbf{f} = (f_\nu, f_\alpha, f_\phi, f_1, \ldots, f_d)$. Here f_ν, f_α and f_ϕ are the implicit functions defined by the equations (3), (4) and (5) respectively; and $f_1, \ldots, f_d$ are specified by the right hand side of the equations (7). Numerical iteration is required to obtain the solution. Giving an initial value of $\boldsymbol{\Psi}^{(0)} = (\nu^{(0)}, \alpha^{(0)}, \phi^{(0)}, \theta_1^{(0)}, \ldots, \theta_d^{(0)})$, we obtain $\boldsymbol{\Psi}^{(1)} = \mathbf{f}(\boldsymbol{\Psi}^{(0)})$; and then iterate until it converges.

To estimate the variance of $\hat{\nu}$, we use the observed information matrix. As $\nu \to \infty$, under some mild conditions for $\lambda_0(t)$, $\nu^{-\frac{1}{2}}(\hat{\nu} - \nu)$ converges in distribution to a zero-mean normal random variable and $\hat{\nu}/\nu \overset{\mathrm{p}}{\to} 1$: see Parner (1998).

In capture-recapture experiments, some animals may die or be removed over the course of the experiments. They cannot be recaptured. Such experiments are called capture-recapture with possible random removals. The proposed estimation procedures can be easily extended to an experiment of this type by replacing $\Omega(\tau)$ by $\Omega(\tau_i)$ in the likelihood (2), where τ_i denotes the removal or termination time for the ith individual ($\tau_i = \tau$ if the individual is not removed during the whole experiment

period). With the likelihood of the ith individual L_i, the likelihood function based on the whole capture history is given by

$$L = \frac{\nu!}{(\nu - n)!} \prod_{i=1}^{n} L_i \, L_0^{\nu - n}.$$

With the same ideas and steps for the capture-recapture model $\mathcal{M}_{\text{htb}}$, the MLE of ν can be obtained. The details are not included here.

2.2 Models $\mathcal{M}_{\text{ht}}$ and $\mathcal{M}_{\text{h}}$

For model $\mathcal{M}_{\text{ht}}$, $\phi = 1$. Substituting this into (3), (4) and (7), gives the estimating equations:

$$\sum_{i=1}^{n} \frac{1}{\nu - i + 1} - \alpha \log(1 + \Omega(\tau)) = 0, \tag{9}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{1}{\alpha + j - 1} - \nu \log(1 + \Omega(\tau)) = 0, \tag{10}$$

$$\frac{n_k}{\theta_k} - \frac{\alpha \nu + Z}{1 + \Omega(\tau)} = 0, \quad (k = 1, \ldots, d); \tag{11}$$

where $n_k = dN(t_{(k)})$ denotes the number of individuals captured at time $t_{(k)}$. From (11), we have $\theta_k = n_k(1 + \Omega(\tau))/(\alpha \nu + Z)$, i.e.,

$$d\Omega(t) = \frac{1 + \Omega(\tau)}{\alpha \nu + Z} dN(t). \tag{12}$$

Given $\rho_1, \rho_2, \ldots, \rho_\nu$, the Nelson-Aalen estimator of $d\Omega(t)$ in this case is

$$\left(\sum_{i=1}^{\nu} \rho_i \right)^{-1} dN(t);$$

and again the estimator (12) is equivalent to the Nelson-Aalen estimator, with ρ_i replaced by the posterior expectation $\mathrm{E}(\rho_i \,|\, \mathcal{F}_\tau^{(i)})$ where $\mathcal{F}_\tau^{(i)}$ is the capture history up to time τ for the ith individual.

9

Integrating both sides of (12), and re-arranging, we obtain

$$\Omega(\tau) = \frac{Z}{\alpha\nu}.$$

Putting this into (9) and (10), gives

$$\sum_{i=1}^{n} \frac{1}{\nu - i + 1} - \alpha \log(1 + \frac{Z}{\alpha\nu}) = 0, \tag{13}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m_j} \frac{1}{\alpha + j - 1} - \nu \log(1 + \frac{Z}{\alpha\nu}) = 0. \tag{14}$$

Solving (13) and (14) using the Newton-Raphson algorithm, the MLEs of ν and α are obtained. Also, (13) and (14) can be regarded as the estimating equations from a profile likelihood in which $\Omega(t)$ is profiled out. Here, there is no difference between MLE and MPLE (maximum profile likelihood estimate); see Murphy & VanDerVaart (2000). To estimate the variance-covariance matrix of $(\hat{\nu}, \hat{\alpha})$, we use the observed profile information matrix.

From (13) and (14), the estimators of ν and α depend only on $(n, m_1, \ldots, m_n)$; they do not depend on the capture times and the form of $\omega(t)$. In fact, for the model $\mathcal{M}_{\mathrm{ht}}$, with $\phi = 1$, the likelihood function is as follows:

$$L = \frac{\nu!}{(\nu - n)!} \left(\prod_{i=1}^{n} \prod_{j=1}^{m_i} \omega(t_{ij}) \right) \left(\prod_{i=1}^{\nu} \frac{\alpha^{[m_i]}}{(1 + \Omega(\tau))^{\alpha + m_i}} \right).$$

Recall that for uncaptured individuals, $m_i = 0 \quad (i = n+1, \ldots, \nu)$. It is seen from the expression for L that the sufficient statistic for (ν, α) is $(n, m_1, \ldots, m_n)$. A more detailed explanation is given in the next section.

For model $\mathcal{M}_{\mathrm{h}}$ (for which $\phi = 1$ and $\omega(t) = \omega$), it is shown similarly that the MLEs of ν and α are also obtained by solving (13) and (14). Since $\omega(t)$ is profiled out in model $\mathcal{M}_{\mathrm{ht}}$, it does not provide extra information for estimating ν or α. Therefore, the estimators of ν and α are exactly the same as for the model $\mathcal{M}_{\mathrm{ht}}$.

2.3 Model $\mathcal{M}_{hb}$

For model $\mathcal{M}_{hb}$, $\lambda_0(t) = 1$. Similarly, for convenience of derivation, we reparameterize the model as $\lambda_i(t) = \rho_i \phi^{r_i(t)} \omega$, where $\rho_i \overset{\mathrm{d}}{=} \mathrm{Ga}(\alpha, 1)$ and ω is an arbitrary non-negative value. The likelihood function is as follows:

$$
\begin{aligned}
\log L \;=\; & \kappa + \sum_{i=1}^{n} \log(\nu - i + 1) + \omega Z - \sum_{i=1}^{n} (\alpha + m_i) \log(1 + \omega t_i^*(\phi)) \\
& - (\nu - n)\alpha \log(1 + \omega\tau) + (Z - n)\log\phi + \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log(\alpha + j - 1),
\end{aligned}
$$

where $t_i^* = t_i^*(\phi) = t_{i1} + \phi(\tau - t_{i1})$. Differentiating gives the estimating equations:

$$
\begin{aligned}
\frac{\partial \log L}{\partial \nu} \;=\; & \sum_{i=1}^{n} \frac{1}{\nu - i + 1} - \alpha \log(1 + \omega\tau) = 0, \\
\frac{\partial \log L}{\partial \alpha} \;=\; & \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{1}{\alpha + j - 1} - \sum_{i=1}^{\nu} \log(1 + \omega t_i^*) = 0, \\
\frac{\partial \log L}{\partial \phi} \;=\; & \frac{Z - n}{\phi} - \sum_{i=1}^{n} \frac{\omega(\alpha + m_i)(\tau - t_{i1})}{1 + \omega t_i^*} = 0, \\
\frac{\partial \log L}{\partial \omega} \;=\; & Z - \sum_{i=1}^{\nu} \frac{(\alpha + m_i)t_i^*}{1 + \omega t_i^*} = 0,
\end{aligned}
$$

in which $t_i^* = \tau$ for $i = n + 1, \ldots, \nu$. Solving these four equations, using Newton-Raphson, the maximum likelihood estimates $\hat{\nu}$, $\hat{\alpha}$, $\hat{\phi}$ and $\hat{\omega}$ can be obtained. The estimates depend on the capture times but only the first capture times of the captured individuals, and the termination time. To estimate the variance of $\hat{\nu}$, we use the observed information matrix.

2.4 Models $\mathcal{M}_{tb}$, $\mathcal{M}_t$, $\mathcal{M}_b$, $\mathcal{M}_0$

Without heterogeneity among individuals, we obtain the homogeneous models $\mathcal{M}_{tb}$, $\mathcal{M}_t$, $\mathcal{M}_b$, $\mathcal{M}_0$. The likelihood-based estimation procedures for which have been given by Hwang, Chao & Yip (2002).

3 Simulation studies

Recall that the model $\mathcal{M}_{\text{htb}}$ is such that

$$\lambda_i(t) = \beta\gamma_i\phi^{r_i(t)}\lambda_0(t) \qquad (0 \leqslant t \leqslant \tau).$$

For purposes of identifiability, the average value of $\lambda_0(t)$ $(0 \leqslant t \leqslant \tau)$ is assumed to be 1; and the mean value of γ_i $(k = 1, 2, \ldots, \nu)$ is 1; so that β represents an overall capture intensity.

For the models considered here, this is simply related to the capture effort, defined as $\varepsilon = \beta\tau$. This relation is derived as follows.

For the model $\mathcal{M}_{\text{ht}}$, given γ_i, $N_i(t)$ is a nonhomogeneous Poisson process with intensity $\beta\gamma_i\lambda_0(t)$. And, since $\Lambda_0(\tau) = \int_0^\tau \lambda_0(t)dt = \tau$, we have

$$\Pr(N_i(\tau) = k \,|\, \gamma_i) = \frac{(\beta\gamma_i\tau)^k}{k!}e^{-\beta\gamma_i\tau}, \quad (k = 0, 1, 2, \ldots).$$

Since $\gamma_i \overset{\text{d}}{=} \text{Ga}(\alpha, \alpha)$, it follows that

$$\begin{aligned}
\Pr(N_i(\tau) &= k\,) \\
&= \int_0^\infty \frac{(\beta\gamma_i\tau)^k}{k!}e^{-\beta\gamma_i\tau}\frac{\alpha^\alpha}{\Gamma(\alpha)}\gamma_i^{\alpha-1}e^{-\alpha\gamma_i}\,d\gamma_i, \\
&= \frac{\alpha^\alpha}{k!}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{(\beta\tau)^k}{(\alpha+\beta\tau)^{\alpha+k}}, \quad (k = 0, 1, 2, \ldots).
\end{aligned} \tag{15}$$

A useful element in describing the process is the capture proportion, denoted by π. Hence

$$\pi = 1 - \Pr(N_i(\tau) = 0) = 1 - \left(1 + \frac{\beta\tau}{\alpha}\right)^{-\alpha}.$$

For model $\mathcal{M}_{\text{htb}}$, π is the same because for uncaptured individuals there is no behavioral modification, and so $\Pr(N_i(\tau) = 0)$ is unchanged. The models $\mathcal{M}_{\text{h}}$ and $\mathcal{M}_{\text{hb}}$ are special cases. Therefore, for the proposed models, π is given by

$$\pi = 1 - \left(1 + \frac{\varepsilon}{\alpha}\right)^{-\alpha}. \tag{16}$$

Thus the capture proportion depends only on the capture effort, ε and the homogeneity of the population, indicated by α. In the homogeneous case $(\alpha = \infty)$ $\pi = 1 - e^{-\varepsilon}$, while in the case $\alpha = 1$, which represents a quite heterogeneous population (the γ_i are exponentially distributed), $\pi = \varepsilon/(1 + \varepsilon)$.

In simulating the models $\mathcal{M}_{ht}$ and $\mathcal{M}_h$, it appears that we should set ν, β, α, τ and the form of $\lambda_0(t)$ ($\lambda_0(t) = 1$ for model $\mathcal{M}_h$). However, only ν, α and ε need to be set; or equivalently ν, α and π. Given ν, α and ε, the outcome is independent of the form of $\lambda_0(t)$, given that its average value is 1. The reason is as follows.

For models $\mathcal{M}_{ht}$ and $\mathcal{M}_h$, where the same estimators are obtained, $\hat{\nu}$ and $\hat{\alpha}$ depend only on the sufficient statistic $(n, m_1, \ldots, m_n)$, which is equivalent to $(f_1, f_2, \ldots)$ where f_k denotes the number of individuals captured exactly k times in the whole experiment. The distribution of $(f_1, f_2, \ldots)$ depends on ν and $\Pr(N_i(\tau) = k)$ $(k = 0, 1, 2, \ldots)$. From (15), $\Pr(N_i(\tau) = k)$ depends only on α and $\varepsilon = \beta\tau$, i.e. only on α and π. Therefore, the distribution of $(f_1, f_2, \ldots)$ is determined completely by ν, α and π.

A different form for $\lambda_0(t)$ would change the progress of the experiment. The effect would be to warp the time-scale: the end result would be unchanged. So, for simulation, we only need to specify ν, α and π: for $\lambda_0(t)$, we can select $\lambda_0(t) = 1$.

Here all simulation results are based on 1000 repetitions: $\text{av}(\hat{\nu})$ denotes the average of the 1000 values of $\hat{\nu}$; $\text{av.se}(\hat{\nu})$ the average of 1000 values of $\text{se}(\hat{\nu})$; and $\text{sd}(\hat{\nu})$ the standard deviation of the 1000 values of $\hat{\nu}$; C denotes the coverage of the 95% confidence intervals for ν, which are calculated using a log transformation presented in Chao (1987). Let ℓ denote the number of failures among the 1000 repetitions. A failure means that the iteration for the estimating equations fails to converge. When

a failure appears among the 1000 repetitions, we say that ν lies outside the 95% confidence interval in estimating C; while the average values $\mathrm{av}(\hat{\nu})$, $\mathrm{sd}(\hat{\nu})$, $\mathrm{av.se}(\hat{\nu})$ are based only on the successful repetitions. With $\nu = 400$, a set of simulation results for model $\mathcal{M}_{\mathrm{ht}}$ (and $\mathcal{M}_{\mathrm{h}}$) is presented in Table 1.

– Table 1 –

Table 1 shows that, as π increases, the bias of $\hat{\nu}$ and the standard deviation, $\mathrm{sd}(\hat{\nu})$ both decrease. Further, $\mathrm{sd}(\hat{\nu})$ and $\mathrm{av.se}(\hat{\nu})$ are quite close, indicating that the asymptotic standard error is satisfactory. As α increases, $\mathrm{sd}(\hat{\nu})$ decreases because of the decrease in heterogeneity $(\mathrm{var}(\gamma_i) = 1/\alpha)$. When $\alpha > 2$, the decrease in $\mathrm{sd}(\hat{\nu})$ is negligible. For large α, ℓ (the number of failures) increases, the estimate $\hat{\alpha}$ is positively biased (not reported here), however, the effect on $\hat{\nu}$ is minimal: since the equation for ν is

$$\sum_{i=1}^{n} \frac{1}{\nu - i + 1} - \log \left(1 + \frac{Z}{\alpha \nu}\right)^{\alpha} = 0,$$

and $\log(1 + \frac{Z}{\alpha \nu})^{\alpha} \to Z/\nu$ as $\alpha \to \infty$, the positively biased $\hat{\alpha}$ for large α changes the value little.

When the estimators for models $\mathcal{M}_{\mathrm{hb}}$ and $\mathcal{M}_{\mathrm{htb}}$ are applied to the setting of model $\mathcal{M}_{\mathrm{h}}$ (regarding $\phi = 1$ and $\lambda_0(t) = 1$), it is expected that there is some loss of accuracy through using more general models. Some simulation results are presented in Table 2.

– Table 2 –

In the setting of model $\mathcal{M}_{\mathrm{h}}$, changing the value of ϕ but keeping the other parameters the same, model $\mathcal{M}_{\mathrm{h}}$ becomes $\mathcal{M}_{\mathrm{hb}}$. Letting $\phi = 0.5$ and 2, applying the estimator for $\mathcal{M}_{\mathrm{hb}}$, it is seen that as ϕ increases, the performance of $\hat{\nu}$ improves because of more revisiting, and conversely, as ϕ decreases, the performance of $\hat{\nu}$ is worse because

14

there is less revisiting; see Table 2.

For model $\mathcal{M}_{\mathrm{htb}}$, let $\lambda_0(t) = 1 + \sin 9t$ and different values of α, ϕ and π are selected. A set of simulation results are presented in Table 3.

– Table 3 –

Simulation was also carried out for some different forms of $\lambda_0(t)$, including polynomial and exponential functions. The performance of $\hat{\nu}$ was found to be similar, mainly depending on α, ϕ and π. In general, for $\nu = 400$, as long as π is larger than about 0.7, there are very few failures; and for $\nu = 1000$, π should be larger than about 0.5.

4   Continuous-time vs discrete-time

There is little real data from continuous-time capture-recapture experiments available. Almost all the published research for developing continuous-time models and estimators used capture-recapture data in discrete-time form for illustration, by regarding the time length of each occasion as 1 and assuming that capture only occurs at the end of each occasion. It is not very genuine continuous time data. Simulation studies (not reported here) suggest that the advantages of continuous-time estimators are not revealed by such manufactured data and serious biases would appear if the number of capture occasion is not large enough (say, smaller than 20). Continuous-type estimators are best applied to continuous-time data.

The unavailability of continuous time capture-recapture data could attributable to the non-existence of the estimators. It makes the researchers to discretize the continuous data. Possible circumstances in which continuous-time data can be obtained

15

are (i) camera capture, which has been widely used to observe animals in the wild, see da Silva et al. (2000); (ii) data from electronically tagged traps, specifying time of capture.

Comparing with discrete-time capture-recapture experiments, more efforts are usually needed for continuous-time experiments. Naturally, we would ask, are they worth the effort? We attempt to answer that question by simulation studies and asymptotic efficiency comparison.

Continuous time data are generated. We can either use the proposed continuous-time estimators; or discretize the data and use the existing discrete-type estimators. The advantage of the continuous-time estimators can then be weighed against the effort of collecting the continuous-time data.

4.1   Models $\mathcal{M}_0$ and $\mathcal{M}_t$

For the continuous-time models $\mathcal{M}_0$ and $\mathcal{M}_t$, the MLEs of the population size are the same, and do not depend on capture time or the form of capture intensity, see Hwang et al. (2002). We assume an intensity function $\lambda(t)$ and cumulative intensity function $\Lambda(t)$.

We divide the capture duration $[0, \tau]$ into K subintervals (t_{j-1}, t_j), $(j = 1, 2, \ldots, K,$ with $t_0 = 0$, $t_K = \tau)$, regarding each subinterval as a capture occasion. Let $c_j = (\Lambda(t_j) - \Lambda(t_{j-1}))/\Lambda(\tau)$, so that c_j denotes the fraction of the capture effort that occurs in the jth capture occasion. The extent of the "discretization" of the time-scale is thus indicated by K, with the continuous time-scale corresponding to $K = \infty$. A number of simulation studies were conducted to compare $\hat{\nu}^{(K)}$ (the discrete-time MLE based on K capture occasions) with $\hat{\nu}^{(\infty)}$ (the continuous-time MLE).

For the case of equivalent capture occasions ($c_j = 1/K$, $j = 1, 2, \ldots, K$), we compare $\hat{\nu}_0^{(K)}$ (of the discrete-time $\mathcal{M}_0$) with $\hat{\nu}_t^{(\infty)}$ (of the continuous-time $\mathcal{M}_t$). Some simulation results are given in Table 4.

– Table 4 –

As $\nu \to \infty$, it can be shown that the asymptotic efficiency of $\hat{\nu}_0^{(K)}$ relative to $\hat{\nu}_t^{(\infty)}$ is given by

$$\text{asy.eff}\left(\hat{\nu}_0^{(K)}, \hat{\nu}_t^{(\infty)}\right) = \left\{1 - \frac{K\left[e^{\Lambda(\tau)/K} - \Lambda(\tau)/K - 1\right]}{e^{\Lambda(\tau)} - \Lambda(\tau) - 1}\right\}^{1/2},$$

and as $K \to \infty$, given $\Lambda(\tau)$, $\text{asy.eff}\left(\hat{\nu}_0^{(K)}, \hat{\nu}_t^{(\infty)}\right) \to 1$ increasingly. Controlling π ($= 1 - e^{-\Lambda(\tau)}$) as 40%, 60% and 80% respectively, $\text{asy.eff}\left(\hat{\nu}_0^{(K)}, \hat{\nu}_t^{(\infty)}\right)$ is plotted against K in Figure 1.

– Figure 1 –

The simulation results and the asymptotic efficiency show that, in the case of equivalent capture occasions, when the number of capture occasions is smaller than 5, there are considerable advantages to be gained by using the continuous-time estimator, especially for a low capture proportion. When the number of capture occasions is larger than 5, the gain is marginal.

In the case that the c_j are unequal, the MLE for the discrete-time $\mathcal{M}_t$, $\hat{\nu}_t^{(K)}$ is considered. We choose $\underset{\sim}{c} = (\frac{3}{4}, \frac{1}{4})$ for $K = 2$, $\underset{\sim}{c} = (\frac{3}{8}, \frac{1}{8}, \frac{3}{8}, \frac{1}{8})$ for $K = 4$ and $\underset{\sim}{c} = (\frac{3}{12}, \frac{1}{12}, \frac{3}{12}, \frac{1}{12}, \frac{3}{12}, \frac{1}{12})$ for $K = 6$. The simulation results are given in Table 5.

– Table 5 –

As $\nu \to \infty$, it can be shown that the asymptotic efficiency of $\hat{\nu}_t^{(K)}$ relative to $\hat{\nu}_t^{(\infty)}$ is given by

$$\text{asy.eff}\left(\hat{\nu}_t^{(K)}, \hat{\nu}_t^{(\infty)}\right) = \left\{1 - \sum_{j=1}^{K} \frac{e^{c_j \Lambda(\tau)} - c_j \Lambda(\tau) - 1}{e^{\Lambda(\tau)} - \Lambda(\tau) - 1}\right\}^{1/2}.$$

17

This efficiency is maximised when all the c_j are equal, i.e. $c_j = 1/K$, $j = 1, \ldots, K$, the maximum is equal to asy.eff$\left(\hat{\nu}_0^{(K)}, \hat{\nu}_t^{(\infty)}\right)$. When one of c_j tends to1 (so the others tend to 0, i.e. only one capture occasion), asy.eff$\left(\hat{\nu}_t^{(K)}, \hat{\nu}_t^{(\infty)}\right) \to 0$, but this is not the case considered. In fact, with no c_j being close to 1, the asymptotic efficiency of $\hat{\nu}_t^{(K)}$ relative to $\hat{\nu}_t^{(\infty)}$ is found to be quite close to the asymptotic efficiency of $\hat{\nu}_0^{(K)}$ relative to $\hat{\nu}_t^{(\infty)}$. For the settings in Table 9, the asymptotic efficiency is plotted in Figure 2.

– Figure 2 –

Extensive simulation studies and the asymptotic efficiency show that, in the case of unequal c_j, when the number of capture occasions is smaller than 5, there are advantages to be gained by using the continuous-time estimator. When the number of capture occasions is larger than 5 and there is no extreme dominated capture occasion, the gain is less.

4.2 Models $\mathcal{M}_h$ and $\mathcal{M}_{ht}$

For the proposed continuous-time models $\mathcal{M}_h$ and $\mathcal{M}_{ht}$, the MLEs of ν are the same. We assume that the intensity for individual i is given by $\lambda_j(t) = \beta \gamma_i \lambda_0(t)$, where $\gamma_i \stackrel{\mathrm{d}}{=} \mathrm{Ga}(\alpha, \alpha)$. Subdividing the capture duration $[0, \tau]$ into K sub-intervals (t_{j-1}, t_j), $(j = 1, 2, \ldots, K$, with $t_0 = 0$, $t_K = \tau)$ with equal $\Delta\Lambda_0(t_j)$ $(=\Lambda_0(t_j) - \Lambda_0(t_{j-1}))$, regarding each subinterval as a capture occasion, then the corresponding discrete-time model is $\mathcal{M}_h$.

The likelihood for the sub-divided data is given by

$$L^{(K)} = \frac{\nu!}{(\nu - n)! f_1! \cdots f_K!} \prod_{j=1}^{K} (P_{Kj})^{f_j} \, (P_{K0})^{\nu - n}$$

where P_{Kj} denotes the probability of being captured j times in the K capture occasions for each individual, given by

$$
\begin{aligned}
P_{Kj} &= \binom{K}{j} \int_0^\infty [1 - \exp(-\gamma\varepsilon/K)]^j \exp(-j\gamma\varepsilon/K) \frac{\alpha^\alpha}{\Gamma(\alpha)} \gamma^{\alpha-1} \exp(-\alpha\gamma) d\gamma \\
&= \binom{K}{j} \sum_{i=0}^j (-1)^i \binom{j}{i} \left[\frac{\alpha}{\alpha + (K - j + i)\varepsilon/K} \right]^\alpha, \quad j = 0, 1, \ldots, K,
\end{aligned}
$$

The MLE $\hat{\nu}_{\mathrm{h}}^{(K)}$ can be obtained by maximizing the likelihood, and the standard error obtained from the observed information matrix.

A series of simulation studies were conducted to compare $\hat{\nu}_{\mathrm{h}}^{(K)}$ with $\hat{\nu}_{\mathrm{ht}}^{(\infty)}$. Results are reported for a population size of $\nu = 500$, capture proportion $\pi = 0.6$ with heterogeneity parameter $\alpha = 0.5, 1, 2$, and $K = 5, 10, 15, \infty$ in the case of equal capture efforts for the subintervals. This is a larger population size with a reasonable capture proportion, allowing a range of heterogeneity. These results are given in Table 6. Note that the reported values of K are larger than for the homogeneous cases (Tables 4 and 5).

– Table 6 –

It is seen that a greater number of subdivisions are required to achieve reasonable efficiency. The effect of discretization is worse when the population is more heterogeneous. Further, these results are for the case when the capture efforts in the sub-intervals are equal. In the case that the capture efforts are unequal, the efficiency of the discrete-time estimators is further reduced.

Following Xi, Yip & Watson (2005), the asymptotic variances of $\hat{\nu}_{\mathrm{h}}^{(K)}$ and $\hat{\nu}_{\mathrm{ht}}^{(\infty)}$ can be obtained, then the asymptotic efficiency of $\hat{\nu}_{\mathrm{h}}^{(K)}$ relative to $\hat{\nu}_{\mathrm{ht}}^{(\infty)}$. We omit the details but plot the efficiency against K, controlling π as 30%, 60% and 90% respectively; see Figure 3.

– Figure 3 –

19

It is seen much larger K is required to achieve reasonable efficiency comparing with the homogeneous cases in Figure 1, 2.

Based on the proposed continuous-time model $\mathcal{M}_{\text{ht}}$, the sample coverage estimate (see Chao et al., 1992; Lee & Chao, 1994) and the jackknife estimate (see Burnham & Overton, 1978; Rexstad & Burnham, 1991) are also applied to the corresponding discrete-time $\mathcal{M}_{\text{h}}$ for comparison. Simulation studies show that, $\hat{\nu}_{\text{ht}}^{(\infty)}$ performs better than the two, in terms of smaller bias and rmse, and with higher coverage of the 95% confidence interval. The exception is for small population with small capture proportion. For $\nu < 200$ with $\pi < 0.7$, in term of rmse, the sample coverage estimate and the jackknife estimate perform better than $\hat{\nu}_{\text{ht}}^{(\infty)}$, but $\hat{\nu}_{\text{ht}}^{(\infty)}$ has a smaller bias and higher coverage of the 95% confidence interval.

5   Discussion

The proposed likelihood-based approach provided estimator for the population size ν for models $\mathcal{M}_{\text{h}}$, $\mathcal{M}_{\text{ht}}$, $\mathcal{M}_{\text{hb}}$ and $\mathcal{M}_{\text{htb}}$. Estimates and standard errors for ν are obtained from the likelihood equations. Simulation showed that the performance is acceptable if the capture proportion π is not too small. We found a few problems in dealing with a capture-recapture situation with a small capture proportion. However, any method will have some problems in this case. The use of the likelihood method is preferred since it is more stable and the asymptotic properties can readily be derived via maximum likelihood theory.

Simulation studies and the asymptotic efficiencies demonstrate that there are considerable efficiency advantages to be gained by using continuous-time estimators, particularly when the population is heterogeneous. Whether efficiency advantages

outweigh the cost of obtaining continuous-time data is another question however. The results would be certainly better if the proposed estimator were used for genuine continuous time data.

References

[1] Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics* **6**, 701-726

[2] Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models based on counting processes*. New York: Springer-Verlag.

[3] Becker, N. G. (1984). Estimating population size from capture-recapture experiments in continuous time. *Australian Journal of Statistics* **26**,1-7.

[4] Burnham, K. P. & Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**, 625-633.

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

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

[7] Chao, A. and Lee , S.-M. (1993). Estimating population size for continuous-time capture-recapture models via sample coverage. *Biometrical Journal* **35**, 29-45.

[8] 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.

[9] Craig, C. C. (1953). On a method of estimating biological populations in the field. *Biometrika* **40**, 216-218.

[10] da Silva, C. Q., Zeh, J., Madigan, D., Laake, J., Rugh, D., Baraff, L., Koski, W. and Miller, G. (2000). Capture-recapture estimation of bowhead whale population size using photo-identification data. *J. Cetacean Research Management* **2**(1), 45-61.

[11] Darroch, J. N. (1958). The multiple-recapture census: I. Estimation of a closed population. *Biometrika* **45**, 343-359.

[12] Hwang, W-H, Chao, A. & Yip, Paul S. F. (2002). Continuous-time capture-recapture models with time variation and behavioural response. *Australian and New Zealand Journal of Statistics* **44**(1), 41-54.

[13] Kiefer, J. & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. The *Annals of Mathematical Statistics* **27**, 887-906.

[14] Lee, S.-M. & Chao, A (1994). Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* **50**, 88-97.

[15] Murphy, S. A. & Van der vaart A. W. (2000). On profile likelihood. *Journal of the American Statistical Association* **95**, 449-465.

[16] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945-965.

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

[18] Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics* **26**, 183-214.

[19] Pollock, K. H. (1991). Modeling capture, recapture and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present and future. *Journal of the American Statistical Association* **86**, 225-238.

[20] Rexstad, E. and Burnham, K. P. (1991). *User's Guide for Interactive Program CAPTURE*. Fort Collins: Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University.

[21] Yip, P. S. F. & Chao A. (1996). Estimating population size from capture-recapture studies via sample coverage and estimating functions. *Communications in Statistics–Stochastic Models* **12**, 17-35

[22] Yip, P. S. F., Fong, D. Y. T. & Wilson, K. (1993). Estimating population size by recapture sampling via estimating function. *Communications in Statistics–Stochastic Models* **9**, 179-193.

[23] Xi, L., Yip, P. S. F & Watson, R. (2005). Estimating population size in a continuous-time removal experiment with a known sub-population size ratio. *Environmental and Ecological Statistics*, to appear.

$$\nu = 400$$

α	π	$\mathrm{av}(\hat{\nu})$	$\mathrm{sd}(\hat{\nu})$	$\mathrm{av.se}(\hat{\nu})$	C	ℓ
1	0.60	403.4	68.4	65.5	93.8%	0
1	0.75	398.5	30.8	29.2	94.5%	0
1	0.90	399.5	11.3	11.2	94.9%	0
0.5	0.85	399.9	17.7	17.4	94.8%	0
1	0.85	399.7	16.6	16.4	95.1%	0
2	0.85	398.7	15.4	15.5	94.3%	0
4	0.85	399.6	15.7	15.1	94.0%	0
5	0.80	399.9	21.4	20.3	93.3%	4
10	0.80	402.1	18.9	20.4	87.1%	97

Table 1: Simulation results for model $\mathcal{M}_{\mathrm{ht}}$ (and $\mathcal{M}_{\mathrm{h}}$)

$$\lambda(t) = 1, \alpha = 1, \ \pi = 0.70 \text{ and } \nu = 400$$

ϕ	estimator for	$\mathrm{av}(\hat{\nu})$	$\mathrm{sd}(\hat{\nu})$	$\mathrm{av.se}(\hat{\nu})$	C	ℓ
	$\mathcal{M}_{\mathrm{h}}$	400.0	38.2	38.2	95.3%	0
1.0	$\mathcal{M}_{\mathrm{hb}}$	399.8	42.1	40.8	95.2%	0
	$\mathcal{M}_{\mathrm{htb}}$	402.6	48.3	46.0	95.1%	1
0.5	$\mathcal{M}_{\mathrm{hb}}$	404.0	54.5	52.0	94.6%	1
2.0	$\mathcal{M}_{\mathrm{hb}}$	401.6	37.9	36.2	95.2%	0

Table 2: Simulation results of applying the different estimators to models $\mathcal{M}_{\mathrm{h}}$ and $\mathcal{M}_{\mathrm{hb}}$

$$\lambda(t) = 1 + \sin 9t, \ \nu = 400$$

α	ϕ	π	$\mathrm{av}(\hat{\nu})$	$\mathrm{sd}(\hat{\nu})$	$\mathrm{av.se}(\hat{\nu})$	C	ℓ
1	0.5	0.70	415.0	65.9	64.2	97.1%	0
1	1.5	0.70	400.0	40.4	40.5	95.3%	0
1	1.5	0.90	396.9	11.0	10.9	95.5%	0
2	1.5	0.90	398.9	11.4	11.5	96.1%	0

Table 3: Simulation results for model $\mathcal{M}_{\mathrm{htb}}$

$\nu = 100$		$\pi = 0.4$			$\pi = 0.6$			$\pi = 0.8$	
K	av	rmse	eff	av	rmse	eff	av	rmse	eff
2	113.7	58.5	0.542	101.6	19.6	0.709	100.0	8.5	0.776
4	107.8	40.6	0.781	101.2	16.2	0.858	99.7	7.2	0.917
6	106.8	37.4	0.848	100.9	15.3	0.908	99.8	6.9	0.957
∞	106.4	31.7	1.000	100.6	13.9	1.000	99.8	6.6	1.000
$\nu = 200$		$\pi = 0.4$			$\pi = 0.6$			$\pi = 0.8$	
K	av	rmse	eff	av	rmse	eff	av	rmse	eff
2	212.5	68.0	0.606	200.5	26.1	0.728	199.7	11.3	0.796
4	208.4	48.6	0.848	200.4	21.6	0.880	199.7	9.9	0.909
6	207.1	46.3	0.890	200.1	20.6	0.922	199.8	9.5	0.948
∞	205.7	41.2	1.000	200.2	19.0	1.000	199.7	9.0	1.000
$\nu = 500$		$\pi = 0.4$			$\pi = 0.6$			$\pi = 0.8$	
K	av	rmse	eff	av	rmse	eff	av	rmse	eff
2	510.6	84.5	0.708	502.4	40.3	0.727	500.4	17.9	0.810
4	508.6	69.1	0.865	501.3	33.2	0.883	500.3	15.7	0.924
6	507.3	64.6	0.926	501.2	32.0	0.916	500.4	15.4	0.942
∞	506.4	59.8	1.000	500.9	29.3	1.000	500.4	14.5	1.000

Table 4: Simulation results based on 2000 repetitions with constant capture efforts for different capture occasions, comparing the discrete-time estimator $\hat{\nu}_0^{(K)}$ with the continuous-time estimator $\hat{\nu}_t^{(\infty)}$, for $\nu = 100, 200, 500$ and $\pi = 0.4, 0.6, 0.8$. eff=rmse($\hat{\nu}_t^{(\infty)}$)/rmse($\hat{\nu}_0^{(K)}$).

$\nu = 100$		$\pi = 0.4$			$\pi = 0.6$			$\pi = 0.8$	
K	av	rmse	eff	av	rmse	eff	av	rmse	eff
2	–	–	–	102.2	29.7	0.468	99.5	10.1	0.653
4	107.4	49.8	0.636	100.4	16.5	0.842	99.2	7.5	0.880
6	105.5	38.8	0.817	100.2	15.4	0.903	99.3	7.0	0.943
∞	106.4	31.7	1.000	100.6	13.9	1.000	99.8	6.6	1.000

$\nu = 200$		$\pi = 0.4$			$\pi = 0.6$			$\pi = 0.8$	
K	av	rmse	eff	av	rmse	eff	av	rmse	eff
2	214.9	81.5	0.506	200.0	30.4	0.625	199.2	13.6	0.662
4	205.7	49.6	0.831	199.6	22.1	0.860	199.2	10.0	0.900
6	205.1	47.0	0.877	199.5	21.1	0.900	199.4	9.7	0.928
∞	205.7	41.2	1.000	200.2	19.0	1.000	199.7	9.0	1.000

$\nu = 500$		$\pi = 0.4$			$\pi = 0.6$			$\pi = 0.8$	
K	av	rmse	eff	av	rmse	eff	av	rmse	eff
2	516.0	106.0	0.564	500.3	45.2	0.648	499.5	21.1	0.687
4	505.5	70.1	0.853	500.5	35.0	0.837	499.9	16.1	0.901
6	504.8	64.6	0.926	500.3	32.7	0.896	500.2	15.4	0.942
∞	506.4	59.8	1.000	500.9	29.3	1.000	500.4	14.5	1.000

Table 5: Simulation results based on 2000 repetitions with varying capture efforts for different capture occasions, comparing the discrete-time estimator $\hat{\nu}_t^{(K)}$ with the continuous-time estimator $\hat{\nu}_t^{(\infty)}$, for $\nu = 100, 200, 500$ and $\pi = 0.4, 0.6, 0.8$. eff=rmse($\hat{\nu}_t^{(\infty)}$)/rmse($\hat{\nu}_t^{(K)}$). For $\nu = 100$, $\pi = 0.4$ and $K = 2$, the results are not presented for comparison due to many failures.

$\pi = 0.6$ $\nu = 500$		$\alpha = 0.5$			$\alpha = 1$			$\alpha = 2$	
K	av	rmse	eff	av	rmse	eff	av	rmse	eff
5	520.4	138.0	0.577	512.7	117.8	0.637	514.5	104.1	0.651
10	511.0	92.8	0.858	507.1	85.6	0.876	506.8	80.1	0.846
15	508.9	89.5	0.889	506.4	84.0	0.893	505.3	74.7	0.907
∞	506.2	79.6	1.000	504.5	75.0	1.000	503.2	67.7	1.000

Table 6: Simulation results based on 2000 repetitions with constant capture efforts for different capture occasions, comparing the discrete-time estimator based on K intervals with the continuous-time estimator ($K = \infty$), for $\nu = 500$, $\pi = 0.6$ and heterogeneity parameter $\alpha = 0.5, 1, 2$.
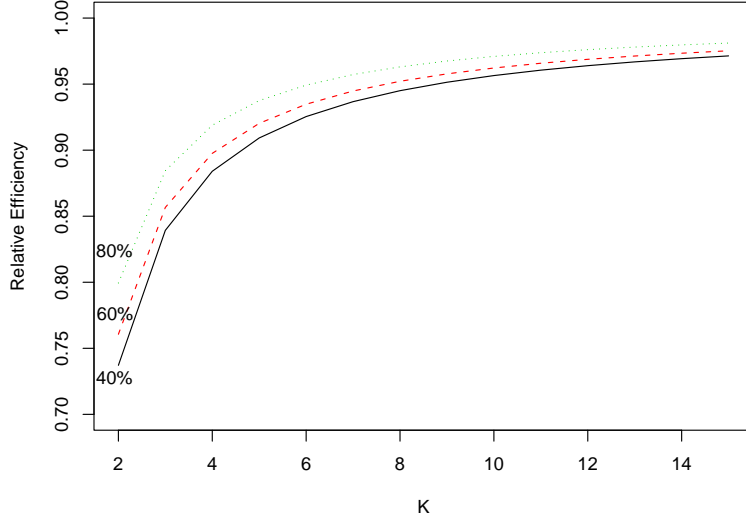
Figure 1: Asymptotic efficiency of $\hat{\nu}_0^{(K)}$ relative to $\hat{\nu}_{\mathrm{t}}^{(\infty)}$ by controlling π as 40%, 60% and 80% respectively.
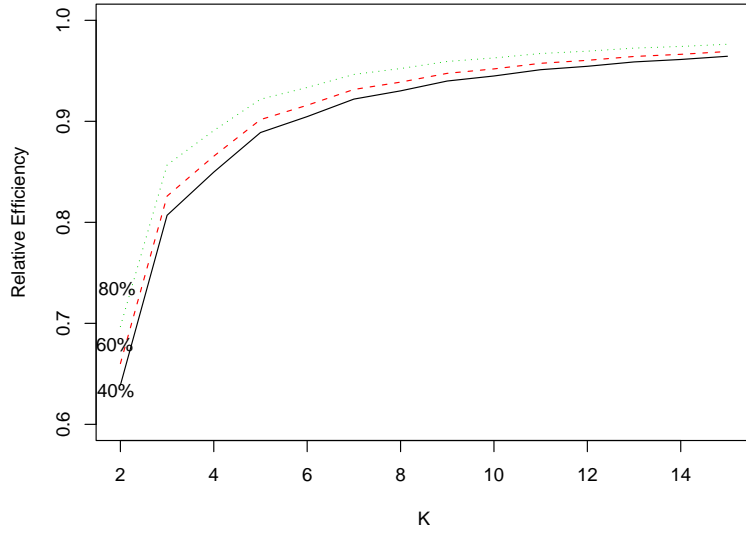

Figure 2: Asymptotic efficiency of $\hat{\nu}_{\mathrm{t}}^{(K)}$ relative to $\hat{\nu}_t^{(\infty)}$ for the settings in Table 5.
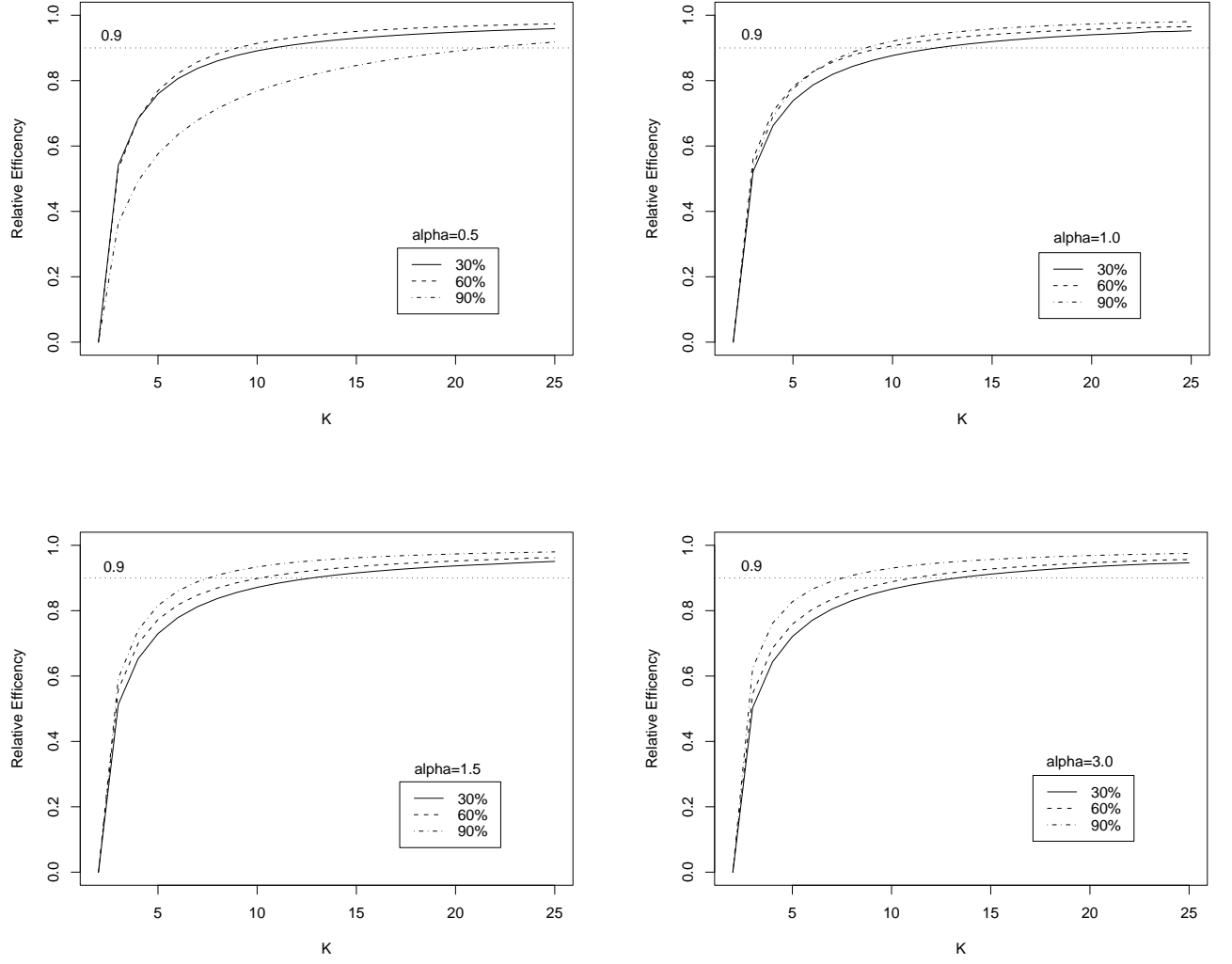
Figure 3: Asymptotic efficiency of $\hat{\nu}_{\mathrm{h}}^{(K)}$ relative to $\hat{\nu}_{\mathrm{ht}}^{(\infty)}$ for $\alpha = 0.5$, 1, 1.5 and 3.