Supplementary material for "Penalized empirical likelihood estimation and EM algorithms for closed-population capture–recapture models"

Yang Liu[1], Pengfei Li[2], and Yukun Liu[1,*]

[1] KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200062, China

[2] Department of Statistics and Actuarial Science, University of Waterloo, Ontario N2L 3G1, Canada

[*] *email:* ykliu@sfs.ecnu.edu.cn

Abstract

The supplementary material consists of two sections. Section 1 presents proofs of the theorems and propositions in the main paper. Section 2 provides some additional simulation studies under the continuous-time capture–recapture models. In Section 3, the proposed penalized empirical likelihood (PEL) method is further extended to the capture-recapture models with ephemeral behavioral effect.

Contents

1 Proofs of the theorems and propositions in the main paper

This section is divided into seven subsections, in each of which we prove one of the theorems and propositions in the main paper. The following lemma plays an important role in these proofs.

Lamma 1. *Suppose that $C = O_p(N_0^{-2})$, $f(N) = -(N - \widetilde{N}_c)^2 I(N > \widetilde{N}_c)$ and $\widetilde{N}_c$ is the Chao (1987)'s lower bound. Then (a) $Cf'(N_0) = O_p(N_0^{-1})$; and (b) $Cf''(N_0) = O_p(N_0^{-3/2})$.*

Proof. It suffices to show that $\widetilde{N}_c = O_p(N_0)$, which clearly holds according to the definition of Chao (1987)'s lower bound. $\square$

1.1 Proof of Theorem 1

Define $\mathbf{Z}_k$, $\mathbf{Z}_{ik}$, and $\mathbf{z}_{ik}$ similar to $\mathbf{z}_k$ and define $\mathbf{Z}_{k0}$, $\mathbf{Z}_{ik0}$, and $\mathbf{z}_{ik0}$ similar to $\mathbf{z}_{k0}$, with $\mathbf{X}$, $\mathbf{X}_i$, and $\mathbf{x}_i$ in substitution for $\mathbf{x}$.

Theorem 1. *Let* $(N_0, \boldsymbol{\beta}_0, \alpha_0)$ *be the true value of* $(N, \boldsymbol{\beta}, \alpha)$ *where* $\alpha_0 \in (0,1)$. *Define*

$$\mathbf{W} = \begin{pmatrix} -\mathrm{V}_{11} & \mathbf{0} & -\mathrm{V}_{13} \\ \mathbf{0} & -\mathbf{V}_{22} + \mathbf{V}_{24}\mathrm{V}_{44}^{-1}\mathbf{V}_{42} & -\mathbf{V}_{23} + \mathbf{V}_{24}\mathrm{V}_{44}^{-1}\mathrm{V}_{43} \\ -\mathrm{V}_{31} & -\mathbf{V}_{32} + \mathrm{V}_{34}\mathrm{V}_{44}^{-1}\mathbf{V}_{42} & -\mathrm{V}_{33} + \mathrm{V}_{34}\mathrm{V}_{44}^{-1}\mathrm{V}_{43} \end{pmatrix}, \tag{1}$$

where $\mathrm{V}_{11} = 1 - \alpha_0^{-1}$, $\mathrm{V}_{13} = \mathrm{V}_{31} = \alpha_0^{-1}$, *and*

$$\begin{aligned} \mathbf{V}_{22} &= \mathbb{E}\left[\frac{\{\partial\phi(\mathbf{X};\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}\}^{\otimes 2}}{\{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)\}\phi(\mathbf{X};\boldsymbol{\beta}_0)} - \sum_{k=1}^{K} g(\mathbf{Z}_k;\boldsymbol{\beta}_0)\{1-g(\mathbf{Z}_k;\boldsymbol{\beta}_0)\}\mathbf{Z}_k^{\otimes 2}\right], \\ \mathbf{V}_{23} &= \mathbf{V}_{32}^{\top} = -\mathbb{E}\left\{\frac{\partial\phi(\mathbf{X};\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)}\right\}, \quad \mathbf{V}_{24} = \mathbf{V}_{42}^{\top} = (1-\alpha_0)^2\mathbf{V}_{23}, \quad \mathrm{V}_{33} = \varphi - \alpha_0^{-1}, \\ \mathrm{V}_{34} &= \mathrm{V}_{43} = (1-\alpha_0)^2\varphi, \quad \mathrm{V}_{44} = (1-\alpha_0)^4\varphi - (1-\alpha_0)^3, \quad \varphi = \mathbb{E}\{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)\}^{-1}. \end{aligned}$$

Here the expectation operator $\mathbb{E}$ *is taken with respect to the distribution of* $(\mathbf{X}^{\top}, D_{(1)}, \ldots, D_{(K)})^{\top}$. *Suppose that the matrix* $\mathbf{W}$ *is positive definite. When* $f(N) = -(N-\widetilde{N}_c)^2 I(N > \widetilde{N}_c)$ *and* $C = O_p(N_0^{-2})$, *as* $N_0 \to \infty$,

- *(a)* $\sqrt{N_0}\{\log(\widehat{N}_p/N_0), (\widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_0)^{\top}, \widehat{\alpha}_p - \alpha_0\}^{\top} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{W}^{-1})$, *where* $\xrightarrow{d}$ *stands for convergence in distribution;*

- *(b)* $R_p(N_0, \boldsymbol{\beta}_0, \alpha_0) \xrightarrow{d} \chi^2_{2+s}$ *and* $R'_p(N_0) \xrightarrow{d} \chi^2_1$, *where* s *is the size of* $\boldsymbol{\beta}$, *and* χ^2_{df} *is the chi-square distribution with* df *degree of freedom.*

Proof. Recall the profile penalized log EL function is

$$\begin{aligned}\ell_p(N, \boldsymbol{\beta}, \alpha) &= \log \binom{N}{n} + (N-n)\log(\alpha) - \sum_{i=1}^{n} \log[1 + \xi\{\phi(\mathbf{x}_i; \boldsymbol{\beta}) - \alpha\}] + Cf(N) \\ &\quad + \sum_{i=1}^{n}\sum_{k=1}^{K}[d_{ik}\log\{g(\mathbf{z}_{ik}; \boldsymbol{\beta})\} + (1-d_{ik})\log\{1 - g(\mathbf{z}_{ik}; \boldsymbol{\beta})\}],\end{aligned}$$

where $\phi(\mathbf{x}_i; \boldsymbol{\beta}) = \prod_{k=1}^{K}\{1 - g(\mathbf{z}_{ik0}; \boldsymbol{\beta})\}$ and $\xi = \xi(\boldsymbol{\beta}, \alpha)$ satisfies

$$\sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i; \boldsymbol{\beta}) - \alpha}{1 + \xi\{\phi(\mathbf{x}_i; \boldsymbol{\beta}) - \alpha\}} = 0. \tag{2}$$

Under model $\mathbf{M}_h$ and when $C = 0$ (i.e. there is no penalty), the profile log EL reduces to $\ell_s(N, \boldsymbol{\beta}_s, \alpha)$ in Section 3.2 of Liu et al. (2017). Thus, the former can be regarded as an extension of the latter to general capture probability models. Theorem 1 can be proved with similar arguments to those in the proof of Corollary 1 of Liu et al. (2017). We highlight only their differences which lie in matrices $\mathbf{V}$ and $\boldsymbol{\Sigma}$.

Note that $\xi(\widehat{\boldsymbol{\beta}}_p, \widehat{\alpha}_p)$ is the solution to equation (2) with $(\widehat{\boldsymbol{\beta}}_p, \widehat{\alpha}_p)$ in substitution for $(\boldsymbol{\beta}, \alpha)$. It can be shown that the limit value of $\xi(\widehat{\boldsymbol{\beta}}_p, \widehat{\alpha}_p)$ is $\xi_0 = -1/(1-\alpha_0)$. Define

$$\begin{aligned}\hbar(N, \boldsymbol{\beta}, \alpha, \xi) &= \log \binom{N}{n} + (N-n)\log(\alpha) - \sum_{i=1}^{n} \log[1 + \xi\{\phi(\mathbf{x}_i; \boldsymbol{\beta}) - \alpha\}] + Cf(N) \\ &\quad + \sum_{i=1}^{n}\sum_{k=1}^{K}[d_{ik}\log\{g(\mathbf{z}_{ik}; \boldsymbol{\beta})\} + (1-d_{ik})\log\{1 - g(\mathbf{z}_{ik}; \boldsymbol{\beta})\}].\end{aligned}$$

It can be seen that $\ell_p(N, \boldsymbol{\beta}, \alpha) = \hbar(N, \boldsymbol{\beta}, \alpha, \xi_*)$, where ξ_* is the solution to $\partial\hbar/\partial\xi = 0$. Denote $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2^\top, \theta_3, \theta_4)^\top$, where $\theta_1 = \sqrt{N_0}(N/N_0 - 1), \boldsymbol{\theta}_2 = \sqrt{N_0}(\boldsymbol{\beta} - \boldsymbol{\beta}_0), \theta_3 = \sqrt{N_0}(\alpha - \alpha_0)$, and $\theta_4 = \sqrt{N_0}(\xi - \xi_0)$. Define

$$\mathcal{H}(\boldsymbol{\theta}) = \hbar(N_0 + N_0^{1/2}\theta_1, \boldsymbol{\beta}_0 + N_0^{-1/2}\boldsymbol{\theta}_2, \alpha_0 + N_0^{-1/2}\theta_3, \xi_0 + N_0^{-1/2}\theta_4).$$

According to Lemma 2 in the supplementary material of Liu et al. (2017), deriving the formula of $\mathbf{V}$ is equivalent to calculating the first two derivatives of $\mathcal{H}(\boldsymbol{\theta})$. By the weak law of large numbers and the central limit theorem, we have

$$\begin{aligned}
\frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_1} &= N_0^{1/2}\left(\frac{n/N_0 - 1}{\alpha_0} + 1\right) + N_0^{1/2} C f'(N_0) + O_p(N_0^{-1/2}) \\
&= N_0^{1/2}\left(\frac{n/N_0 - 1}{\alpha_0} + 1\right) + O_p(N_0^{-1/2}), \\
\frac{\partial \mathcal{H}(\mathbf{0})}{\partial \boldsymbol{\theta}_2} &= N_0^{-1/2} \sum_{i=1}^{n} \left[\sum_{k=1}^{K} \{d_{ik} - g(\mathbf{z}_{ik}; \boldsymbol{\beta}_0)\}\mathbf{z}_{ik} + \frac{\dot{\phi}(\mathbf{x}_i; \boldsymbol{\beta}_0)}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0)}\right], \\
\frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_3} &= N_0^{-1/2} \left\{\frac{N_0 - n}{\alpha_0} - \sum_{i=1}^{n} \frac{1}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0)}\right\}, \\
\frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_4} &= -N_0^{-1/2}(1 - \alpha_0) \sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i; \boldsymbol{\beta}_0) - \alpha_0}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta}_0)},
\end{aligned}$$

where $\dot{\phi}(\mathbf{x}; \boldsymbol{\beta}_0) = \partial \phi(\mathbf{x}; \boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}$ and the first equation is from the result in Section 2.2 of the supplementary material of Liu et al. (2017). In addition,

$$\begin{aligned}
\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_1^2} &= 1 - \alpha_0^{-1} + N_0 C f''(N_0) + O_p(N_0^{-1/2}) = 1 - \alpha_0^{-1} + O_p(N_0^{-1/2}), \\
\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top} &= \mathbb{E}\left[\frac{\{\dot{\phi}(\mathbf{X}; \boldsymbol{\beta}_0)\}^{\otimes 2}}{\{1 - \phi(\mathbf{X}; \boldsymbol{\beta}_0)\}\phi(\mathbf{X}; \boldsymbol{\beta}_0)}\right] - \sum_{k=1}^{K} \mathbb{E}\Big[g(\mathbf{Z}_k; \boldsymbol{\beta}_0)\{1 - g(\mathbf{Z}_k; \boldsymbol{\beta}_0)\}\mathbf{Z}_k^{\otimes 2}\Big] \\
&\quad + O_p(N_0^{-1/2}), \\
\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \boldsymbol{\theta}_2 \partial \theta_3} &= \left\{\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_3 \partial \boldsymbol{\theta}_2^\top}\right\}^\top = -\mathbb{E}\frac{\dot{\phi}(\mathbf{X}; \boldsymbol{\beta}_0)}{1 - \phi(\mathbf{X}; \boldsymbol{\beta}_0)} + O_p(N_0^{-1/2}), \\
\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \boldsymbol{\theta}_2 \partial \theta_4} &= \left\{\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_4 \partial \boldsymbol{\theta}_2^\top}\right\}^\top = -(1 - \alpha_0)^2 \mathbb{E}\frac{\dot{\phi}(\mathbf{X}; \boldsymbol{\beta}_0)}{1 - \phi(\mathbf{X}; \boldsymbol{\beta}_0)} + O_p(N_0^{-1/2}), \\
\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_3^2} &= -\alpha_0^{-1} + \mathbb{E}\frac{1}{1 - \phi(\mathbf{X}; \boldsymbol{\beta}_0)} + O_p(N_0^{-1/2}), \quad \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_3 \partial \theta_1} = \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_1 \partial \theta_3} = \alpha_0^{-1}, \\
\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_4 \partial \theta_3} &= \frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_3 \partial \theta_4} = (1 - \alpha_0)^2 \mathbb{E}\frac{1}{1 - \phi(\mathbf{X}; \boldsymbol{\beta}_0)} + O_p(N_0^{-1/2}),
\end{aligned}$$

$$
\frac{\partial^2 \mathcal{H}(\mathbf{0})}{\partial \theta_4^2} = (1-\alpha_0)^4 \mathbb{E} \frac{1}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)} - (1-\alpha_0)^3 + O_p(N_0^{-1/2}).
$$

The matrix $\mathbf{V}$ is the leading term of $\partial^2 \mathcal{H}(\mathbf{0})/(\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top)$. It follows that

$$
\mathbf{V} = \begin{pmatrix} \mathrm{V}_{11} & \mathbf{0}_{1\times s} & \mathrm{V}_{13} & 0 \\ \mathbf{0}_{s\times 1} & \mathbf{V}_{22} & \mathbf{V}_{23} & \mathbf{V}_{24} \\ \mathrm{V}_{31} & \mathbf{V}_{32} & \mathrm{V}_{33} & \mathrm{V}_{34} \\ 0 & \mathbf{V}_{42} & \mathrm{V}_{43} & \mathrm{V}_{44} \end{pmatrix},
$$

where $\mathbf{0}_{s\times 1}$ is a s-dimensional zero vector and V_{ij}'s are defined in Theorem 1.

Let $\partial \mathcal{H}(\mathbf{0})/\partial \boldsymbol{\theta} = \boldsymbol{u}_n + O_p(N_0^{-1/2})$, where $\boldsymbol{u}_n = (u_{n1}, \boldsymbol{u}_{n2}^\top, u_{n3}, u_{n4})^\top$ with

$$
u_{n1} = N_0^{1/2}\left(\frac{n/N_0 - 1}{\alpha_0} + 1\right),\ \boldsymbol{u}_{n2} = \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \boldsymbol{\theta}_2},\ u_{n3} = \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_3},\ u_{n4} = \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_4}.
$$

It can be verified that $E(\boldsymbol{u}_n) = \mathbf{0}$ and

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}(u_{n1}) &= \frac{N_0}{\alpha_0^2} \cdot \frac{(1-\alpha_0)\alpha_0}{N_0} = \alpha_0^{-1} - 1,\ \mathbb{C}\mathrm{ov}(u_{n3}, u_{n1}) = -\alpha_0^{-1},\ \mathbb{C}\mathrm{ov}(u_{n4}, u_{n1}) = 0, \\
\mathbb{V}\mathrm{ar}(\boldsymbol{u}_{n2}) &= \mathbb{E}\left\{\left[\sum_{k=1}^K \{D_{(k)} - g(\mathbf{Z}_k;\boldsymbol{\beta}_0)\}\mathbf{Z}_k + \frac{\dot{\phi}(\mathbf{X};\boldsymbol{\beta}_0)}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)}\right]^{\otimes 2} I\left(\sum_{k=1}^K D_{(k)} > 0\right)\right\} \\
&= \sum_{k=1}^K \mathbb{E}\Big[g(\mathbf{Z}_k;\boldsymbol{\beta}_0)\{1-g(\mathbf{Z}_k;\boldsymbol{\beta}_0)\}\mathbf{Z}_k^{\otimes 2}\Big] - \mathbb{E}\left[\frac{\{\dot{\phi}(\mathbf{X};\boldsymbol{\beta}_0)\}^{\otimes 2}}{\{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)\}\phi(\mathbf{X};\boldsymbol{\beta}_0)}\right], \\
\mathbb{V}\mathrm{ar}(u_{n3}) &= \alpha_0^{-1} + \varphi,\ \mathbb{C}\mathrm{ov}(\boldsymbol{u}_{n2}, u_{n1}) = \mathbb{C}\mathrm{ov}(\boldsymbol{u}_{n2}, u_{n3}) = \mathbb{C}\mathrm{ov}(\boldsymbol{u}_{n2}, u_{n4}) = \mathbf{0}, \\
\mathbb{C}\mathrm{ov}(\boldsymbol{u}_{n3}, u_{n4}) &= (1-\alpha_0)^2\varphi - (1-\alpha_0), \quad \mathbb{V}\mathrm{ar}(u_{n4}) = (1-\alpha_0)^4\varphi - (1-\alpha_0)^3.
\end{aligned}
$$

By the central limit theorem, we have $\boldsymbol{u}_n \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{\Sigma})$ as $N_0 \to \infty$, where

$$\mathbf{\Sigma} = \begin{pmatrix} -\mathrm{V}_{11} & \mathbf{0}_{1\times s} & -\mathrm{V}_{13} & 0 \\ \mathbf{0}_{s\times 1} & -\mathbf{V}_{22} & \mathbf{0}_{s\times 1} & \mathbf{0}_{s\times 1} \\ -\mathrm{V}_{31} & \mathbf{0}_{1\times s} & 2\mathrm{V}_{34}(1-\alpha_0)^{-2} - \mathrm{V}_{33} & \mathrm{V}_{44}(1-\alpha_0)^{-2} \\ 0 & \mathbf{0}_{1\times s} & \mathrm{V}_{44}(1-\alpha_0)^{-2} & \mathrm{V}_{44} \end{pmatrix}.$$

Now that $\mathbf{\Sigma}$ has the same form as that in Lemma 3 of the supplementary material of Liu et al. (2017), so does the matrix $\mathbf{W}$. Using similar arguments to those in the proof of Theorem 1 of Liu et al. (2017), we have

$$\sqrt{N_0}\{\log(\widehat{N}_p/N_0), (\widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_0)^\top, \widehat{\alpha}_p - \alpha_0\}^\top = \mathbf{W}^{-1}\boldsymbol{t} + O_p(N_0^{-1/2}), \tag{3}$$

as $N_0 \to \infty$, where $\boldsymbol{t} = (t_1, \boldsymbol{t}_2, t_3)$, $t_1 = u_{n1}$, $\boldsymbol{t}_2 = \boldsymbol{u}_{n2} - \mathbf{V}_{24}\mathrm{V}_{44}^{-1}u_{n4}$, and $t_3 = u_{n3} - \mathrm{V}_{34}\mathrm{V}_{44}^{-1}u_{n4}$.

The rest of the proof is similar and omitted. This completes the proof of Theorem 1. $\square$

1.2 Proof of Proposition 2

Proposition 2. *Under the conditions in Theorem 1, as $N_0 \to \infty$,*

(a) $\widehat{\boldsymbol{\beta}}_p - \widehat{\boldsymbol{\beta}}_c = O_p(N_0^{-1})$ *and* $\widehat{N}_p - \widehat{N}_c = O_p(1)$*;*

(b) $\sqrt{N_0}(\widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, -\mathbf{V}_{22}^{-1})$ *and* $\sqrt{N_0}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, -\mathbf{V}_{22}^{-1})$*;*

(c) $N_0^{-1/2}(\widehat{N}_p - N_0) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ *and* $N_0^{-1/2}(\widehat{N}_c - N_0) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$*, where* $\sigma^2 = \varphi - 1 - \mathbf{V}_{32}\mathbf{V}_{22}^{-1}\mathbf{V}_{23}$*,* φ *and* $\mathbf{V}_{ij}$*'s are defined in Theorem 1.*

Proof. By definition, $\widehat{\boldsymbol{\beta}}_c$ must be a stationary point of of $\ell_c(\boldsymbol{\beta}) = \log\{L_c(\boldsymbol{\beta})\}$, which implies

$\mathbf{0} = \partial\ell_c(\widehat{\boldsymbol{\beta}}_c)/\partial\boldsymbol{\beta}$. Applying the first-order Taylor expansion to this equation, we have

$$\begin{aligned}\mathbf{0} &= \frac{\partial\ell_c(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}} + \frac{\partial^2\ell_c(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0) + O_p(1)\\ &= N_0^{1/2}\boldsymbol{u}_{n2} + N_0\mathbf{V}_{22}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0) + O_p(1),\end{aligned}$$

which implies

$$\sqrt{N_0}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0) = -\mathbf{V}_{22}^{-1}\boldsymbol{u}_{n2} + O_p(N_0^{-1/2}). \tag{4}$$

Denote the inverse of $\mathbf{W} = (\mathrm{W}_{ij})_{1\le i,j\le 3}$ by $(\mathrm{W}^{ij})_{1\le i,j\le 3}$. Equation (3) implies

$$\begin{aligned}N_0^{-1/2}(\widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_0) &= \mathbf{W}^{21}t_1 + \mathbf{W}^{22}\boldsymbol{t}_2 + \mathbf{W}^{23}t_3 + O_p(N_0^{-1/2})\\ &= \mathbf{W}^{21}u_{n1} + \mathbf{W}^{22}\boldsymbol{u}_{n2} + \mathbf{W}^{23}u_{n3}\\ &\quad -(\mathbf{W}^{22}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{23}\mathbf{V}_{34}\mathrm{V}_{44}^{-1})u_{n4} + O_p(N_0^{-1/2}).\end{aligned}$$

By the definition of $\boldsymbol{u}_n$, it can be verified that $u_{n4} = (1-\alpha_0)u_{n1} + (1-\alpha_0)^2u_{n3} + O_p(N_0^{-1/2})$. Using this expression, we have

$$\begin{aligned}N_0^{-1/2}(\widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_0) &= \{\mathbf{W}^{21} - (1-\alpha_0)(\mathbf{W}^{22}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{23}\mathbf{V}_{34}\mathrm{V}_{44}^{-1})\}u_{n1} + \mathbf{W}^{22}\boldsymbol{u}_{n2}\\ &\quad +\{\mathbf{W}^{23} - (1-\alpha_0)^2(\mathbf{W}^{22}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{23}\mathbf{V}_{34}\mathrm{V}_{44}^{-1})\}u_{n3} + O_p(N_0^{-1/2}).\end{aligned} \tag{5}$$

Comparing Equations (4) and (5), we find that proving $\widehat{\boldsymbol{\beta}}_p - \widehat{\boldsymbol{\beta}}_c = O_p(N_0^{-1})$ is equivalent to proving the following equations:

$$\mathbf{W}^{21} - (1-\alpha_0)(\mathbf{W}^{22}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{23}\mathbf{V}_{34}\mathrm{V}_{44}^{-1}) = \mathbf{0}, \tag{6}$$

$$\mathbf{W}^{23} - (1-\alpha_0)^2(\mathbf{W}^{22}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{23}\mathbf{V}_{34}\mathrm{V}_{44}^{-1}) = \mathbf{0}, \tag{7}$$

$$\mathbf{W}^{22} = -\mathbf{V}_{22}^{-1}. \tag{8}$$

In fact, we can show that

$$\begin{aligned}
\mathbf{W}_{23} &= -\mathbf{V}_{23}+\mathbf{V}_{24}\mathrm{V}_{44}^{-1}\mathrm{V}_{43} = (1-\alpha_0)\mathbf{V}_{24}\mathrm{V}_{44}^{-1},\\
\mathrm{W}_{33} &= -\mathrm{V}_{33}+\mathrm{V}_{34}\mathrm{V}_{44}^{-1}\mathrm{V}_{43} = \frac{(1-\alpha_0)^3(\varphi-1)}{\alpha_0}\mathrm{V}_{44}^{-1}.
\end{aligned}$$

Comparing the second row of $\mathbf{W}^{-1}\cdot\mathbf{W}=\mathbf{I}$, we have

$$\begin{aligned}
\mathbf{W}^{21}(\alpha_0^{-1}-1)-\alpha_0^{-1}\mathbf{W}^{23} &= \mathbf{0},\\
-\alpha_0^{-1}\mathbf{W}^{21}+\mathbf{W}^{22}(1-\alpha_0)\mathbf{V}_{24}\mathrm{V}_{44}^{-1}+\mathbf{W}^{23}\frac{(1-\alpha_0)^3(\varphi-1)}{\alpha_0}\mathrm{V}_{44}^{-1} &= \mathbf{0},
\end{aligned}$$

$$\mathbf{W}^{22}\left(-\mathbf{V}_{22}+\mathbf{V}_{24}\mathrm{V}_{44}^{-1}\mathbf{V}_{42}\right)+\mathbf{W}^{23}(1-\alpha_0)\mathrm{V}_{44}^{-1}\mathbf{V}_{42} = \mathbf{I}. \qquad (9)$$

Simplifying the first two equations gives

$$\mathbf{W}^{23}=(1-\alpha_0)\mathbf{W}^{21}, \quad \mathbf{W}^{21}=-\mathbf{W}^{22}\mathbf{V}_{23}.$$

From these two equations, we immediately have Equations (6) and (7) and Equation (9) becomes $-\mathbf{W}^{22}\mathbf{V}_{22}=\mathbf{I}$, which is equivalent to Equaion (8).

So far, we have proven $\widehat{\boldsymbol{\beta}}_p-\widehat{\boldsymbol{\beta}}_c=O_p(N_0^{-1})$. Since $\sqrt{N_0}(\widehat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}_0)=-\mathbf{V}_{22}^{-1}\boldsymbol{u}_{n2}+O_p(N_0^{-1/2})$ and $\boldsymbol{u}_{n2}\xrightarrow{d}\mathrm{N}(\mathbf{0},-\mathbf{V}_{22})$, then we conclude that both $\sqrt{N_0}(\widehat{\boldsymbol{\beta}}_p-\boldsymbol{\beta}_0)$ and $\sqrt{N_0}(\widehat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}_0)$ converge to $\mathrm{N}(\mathbf{0},-\mathbf{V}_{22}^{-1})$ in distribution. This completes the proof of Result (b).

We now turn to Result (c). By the definition of $\widehat{N}_c$,

$$\begin{aligned}
N_0^{-1/2}(\widehat{N}_c-N_0) &= N_0^{-1/2}\sum_{i=1}^{n}\left\{\frac{1}{1-\phi(\mathbf{x}_i;\widehat{\boldsymbol{\beta}}_c)}-\frac{1}{1-\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)}\right\}\\
&\quad +N_0^{-1/2}\left\{\sum_{i=1}^{n}\frac{1}{1-\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)}-N_0\right\}
\end{aligned}$$

$$
\begin{aligned}
&= -\mathbf{V}_{32}\sqrt{N_0}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0) - (u_{n1} + u_{n3}) + O_p(N_0^{-1/2}) \\
&= \mathbf{V}_{32}\mathbf{V}_{22}^{-1}\boldsymbol{u}_{n2} - (u_{n1} + u_{n3}) + O_p(N_0^{-1/2}).
\end{aligned}
$$

In addition, it follows from (3) that

$$
\begin{aligned}
N_0^{-1/2}(\widehat{N}_p - N_0) &= \{\mathbf{W}^{11} - (1-\alpha_0)(\mathbf{W}^{12}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{13}\mathrm{V}_{34}\mathrm{V}_{44}^{-1})\}u_{n1} + \mathbf{W}^{12}\boldsymbol{u}_{n2} \\
&\quad + \{\mathbf{W}^{13} - (1-\alpha_0)^2(\mathbf{W}^{12}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{13}\mathrm{V}_{34}\mathrm{V}_{44}^{-1})\}u_{n3} + O_p(N_0^{-1/2}).
\end{aligned}
$$

Comparing the above two approximates, we find that proving $\widehat{N}_p - \widehat{N}_c = O_p(1)$ is equivalent to verifying the following equations:

$$
\begin{aligned}
\mathbf{W}^{11} - (1-\alpha_0)(\mathbf{W}^{12}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{13}\mathrm{V}_{34}\mathrm{V}_{44}^{-1}) &= -1, && (10)\\
\mathbf{W}^{13} - (1-\alpha_0)^2(\mathbf{W}^{12}\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{13}\mathrm{V}_{34}\mathrm{V}_{44}^{-1}) &= -1, && (11)\\
\mathbf{W}^{12} &= \mathbf{V}_{32}\mathbf{V}_{22}^{-1}. && (12)
\end{aligned}
$$

In fact, comparing the first row of $\mathbf{W}^{-1} \cdot \mathbf{W} = \mathbf{I}$ gives

$$
\begin{aligned}
\mathbf{W}^{11}(\alpha_0^{-1} - 1) - \alpha_0^{-1}\mathbf{W}^{13} &= 1, && (13)\\
-\alpha_0^{-1}\mathbf{W}^{11} + \mathbf{W}^{12}(1-\alpha_0)\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{13}\frac{(1-\alpha_0)^3(\varphi - 1)}{\alpha_0}\mathrm{V}_{44}^{-1} &= 0, && (14)\\
\mathbf{W}^{12}\left(-\mathbf{V}_{22} + \mathbf{V}_{24}\mathrm{V}_{44}^{-1}\mathbf{V}_{42}\right) + \mathbf{W}^{13}(1-\alpha_0)\mathrm{V}_{44}^{-1}\mathbf{V}_{42} &= 0. && (15)
\end{aligned}
$$

Then, Equation (13) plus Equation (14) implies Equation (10). Equation (10) $\times(1-\alpha_0)$ minus Equation (13) $\times\alpha_0$ gives Equation (11). Equation (13) times $(1-\alpha_0)^{-1}$ plus Equation (14) leads to

$$
\mathbf{W}^{12}(1-\alpha_0)\mathbf{V}_{24}\mathrm{V}_{44}^{-1} + \mathbf{W}^{13}(1-\alpha_0)^2\mathrm{V}_{44}^{-1} = \frac{1}{1-\alpha_0}. \qquad (16)
$$

Equation (16) $\times \mathrm{V}_{42}/(1-\alpha_0)$ plus Equation (15) gives

$$\mathbf{W}^{12}\mathbf{V}_{22} = \frac{\mathbf{V}_{42}}{(1-\alpha_0)^2} = \mathbf{V}_{32},$$

which implies Equation (12).

Therefore $\widehat{N}_p - \widehat{N}_c = O_p(1)$. Since $N_0^{-1/2}(\widehat{N}_c - N_0) = \mathbf{V}_{32}\mathbf{V}_{22}^{-1}\boldsymbol{u}_{n2} - (u_{n1}+u_{n3}) + O_p(N_0^{-1/2})$ and $\boldsymbol{u}_n \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, then we conclude that both $\sqrt{N_0}(\widehat{N}_p - N_0)$ and $\sqrt{N_0}(\widehat{N}_c - N_0)$ converge to $\mathrm{N}(0, \sigma^2)$ in distribution, where

$$\sigma^2 = \mathbb{V}\mathrm{ar}\{\mathbf{V}_{32}\mathbf{V}_{22}^{-1}\boldsymbol{u}_{n2} - (u_{n1}+u_{n3})\} = \mathbb{E}\{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)\}^{-1} - 1 - \mathbf{V}_{32}\mathbf{V}_{22}^{-1}\mathbf{V}_{23}.$$

This completes the proof of Result (c). Result (a) has been proved in our proofs of Results (b) and (c). □

1.3 Proof of Theorem 3

The penalized log EL function under discrete-time capture–recapture models is

$$\begin{aligned}\widetilde{\ell}_p(N, \boldsymbol{\beta}, \alpha, \{p_i\}) \;=\;& \log\binom{N}{n} + (N-n)\log(\alpha) + \sum_{i=1}^{n}\log(p_i) + Cf(N) \\ &+ \sum_{i=1}^{n}\sum_{k=1}^{K}[d_{ik}\log\{g(\mathbf{z}_{ik};\boldsymbol{\beta})\} + (1-d_{ik})\log\{1-g(\mathbf{z}_{ik};\boldsymbol{\beta})\}]. \quad (17)\end{aligned}$$

Theorem 3. *With discrete-time capture–recapture models, the EM algorithm proposed for the PEL method has following properties.*

(a) When N is fixed, the penalized log EL is nondecreasing after each EM iteration.

(b) When N is unknown, the penalized log EL is nondecreasing after each EM iteration.

(c) When N is unknown, the sequence of EM iterations $(N^{(r)}, \boldsymbol{\beta}^{(r)}, \alpha^{(r)})$ converges to a local maximum PEL estimator $(\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p, \widehat{\alpha}_p)$.

1.3.1 Proof of Result (a)

Proof. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$ and $\boldsymbol{\psi} = (\boldsymbol{\theta}^\top, p_1, \ldots, p_n)^\top$ with $\boldsymbol{\theta}^{(r)}$ and $\boldsymbol{\psi}^{(r)}$ being their values at the rth iteration ($r = 0, 1, \ldots$). When N is fixed, the penalized log EL function of $\boldsymbol{\psi}$ is

$$\begin{aligned}\widetilde{\ell}_p(\boldsymbol{\psi}) &= \log\binom{N}{n} + (N-n)\log(\alpha) + \sum_{i=1}^n \log(p_i) + Cf(N) \\ &\quad + \sum_{i=1}^n \sum_{k=1}^K [d_{ik}\log\{g(\mathbf{z}_{ik}; \boldsymbol{\beta})\} + (1-d_{ik})\log\{1-g(\mathbf{z}_{ik}; \boldsymbol{\beta})\}].\end{aligned}$$

To prove Result (a), it suffices to prove $\widetilde{\ell}_p(\boldsymbol{\psi}^{(r+1)}) - \widetilde{\ell}_p(\boldsymbol{\psi}^{(r)}) \geq 0$ for any $r = 0, 1, \ldots$. To this end, we define $\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}^{(r)}) = \mathcal{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)}) - \widetilde{\ell}_p(\boldsymbol{\psi})$. Since $\mathcal{Q}(\boldsymbol{\psi}^{(r+1)}|\boldsymbol{\psi}^{(r)}) \geq \mathcal{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)})$ for all $\boldsymbol{\psi}$, we shall have $\widetilde{\ell}_p(\boldsymbol{\psi}^{(r)}) \leq \widetilde{\ell}_p(\boldsymbol{\psi}^{(r+1)})$ if we can prove $\kappa(\boldsymbol{\psi}^{(r+1)}, \boldsymbol{\psi}^{(r)}) \leq \kappa(\boldsymbol{\psi}^{(r)}, \boldsymbol{\psi}^{(r)})$.

It can be verified that

$$\begin{aligned}\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}^{(r)}) &= \sum_{i=1}^n \left[w_i^{(r)} \log\{\phi(\mathbf{x}_i; \boldsymbol{\beta})p_i\}\right] - \log\binom{N}{n} - (N-n)\log(\alpha) - Cf(N) \\ &= \sum_{i=1}^n \left[w_i^{(r)} \log\left\{\frac{\phi(\mathbf{x}_i; \boldsymbol{\beta})p_i}{\alpha}\right\}\right] - \log\binom{N}{n} - Cf(N) \\ &= \sum_{i=1}^n \sum_{j=n+1}^N \left(\mathbb{E}[I(\mathbf{X}_j = \mathbf{x}_i)|D_j = 0, \boldsymbol{\psi} = \boldsymbol{\psi}^{(r)}] \log\left\{\frac{\phi(\mathbf{x}_i; \boldsymbol{\beta})p_i}{\alpha}\right\}\right) \\ &\qquad - \log\binom{N}{n} - Cf(N) \\ &= \mathbb{E}_{\boldsymbol{\psi}^{(r)}}\left(\log\left[\prod_{j=n+1}^N \prod_{i=1}^n \left\{\frac{\phi(\mathbf{x}_i; \boldsymbol{\beta})p_i}{\alpha}\right\}^{I(\mathbf{X}_j = \mathbf{x}_i)}\right] \Big| D_{n+1} = \ldots = D_N = 0\right) \\ &\qquad - \log\binom{N}{n} - Cf(N) \\ &= \mathbb{E}_{\boldsymbol{\psi}^{(r)}}[\log\{k(\mathbf{X}^*; \boldsymbol{\psi})\}|D_{n+1} = \ldots = D_N = 0] - \log\binom{N}{n} - Cf(N),\end{aligned}$$

where $\mathbf{X}^* = (\mathbf{X}_{n+1}^\top, \ldots, \mathbf{X}_N^\top)^\top$, $\mathbb{E}_{\boldsymbol{\psi}^{(r)}}$ is the expectation operator with respect to $k(\mathbf{x}; \boldsymbol{\psi}^{(r)})$, and $k(\mathbf{X}^*; \boldsymbol{\psi})$ is the conditional probability function of $\mathbf{X}^*$ given $D_{n+1} = \ldots = D_N = 0$.

When N is fixed, it follows from Jensen's inequality that

$$\begin{aligned}
\kappa(\boldsymbol{\psi}^{(r+1)}, \boldsymbol{\psi}^{(r)}) - \kappa(\boldsymbol{\psi}^{(r)}, \boldsymbol{\psi}^{(r)}) &= \mathbb{E}_{\boldsymbol{\psi}^{(r)}}\left[\log\left\{\frac{k(\mathbf{X}^*; \boldsymbol{\psi}^{(r+1)})}{k(\mathbf{X}^*; \boldsymbol{\psi}^{(r)})}\right\}\Big| D_{n+1} = \ldots = D_N = 0\right] \\
&\leq \log\left[\mathbb{E}_{\boldsymbol{\psi}^{(r)}}\left\{\frac{k(\mathbf{X}^*; \boldsymbol{\psi}^{(r+1)})}{k(\mathbf{X}^*; \boldsymbol{\psi}^{(r)})}\Big| D_{n+1} = \ldots = D_N = 0\right\}\right] \\
&= 0.
\end{aligned}$$

□

1.3.2 Proof of Result (b)

Proof. When N is unknown, we redefine $\boldsymbol{\theta} = (N, \boldsymbol{\beta}^\top, \alpha)^\top$ and $\boldsymbol{\psi} = (\boldsymbol{\theta}^\top, p_1, \ldots, p_n)^\top$ with $\boldsymbol{\theta}^{(r)}$ and $\boldsymbol{\psi}^{(r)}$ being their values at the rth iteration ($r = 0, 1, \ldots$). Now,

$$\begin{aligned}
\kappa(\boldsymbol{\psi}, \boldsymbol{\psi}^{(r)}) &:= \mathcal{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)}) - \ell(\boldsymbol{\psi}) \\
&= \sum_{i=1}^{n}[w_i^{(r)} \log\{\phi(\mathbf{x}_i; \boldsymbol{\beta})p_i\}] - \log\binom{N}{n} - (N-n)\log(\alpha) - Cf(N) \\
&= \sum_{i=1}^{n}\left[w_i^{(r)} \log\left\{\frac{\phi(\mathbf{x}_i; \boldsymbol{\beta})p_i}{\alpha}\right\}\right] - \log\binom{N}{n} - (N-N^{(r)})\log(\alpha) - Cf(N) \\
&= \mathbb{E}_{\boldsymbol{\psi}^{(r)}}[\log\{k(\mathbf{X}^*; \boldsymbol{\psi})\}|D_{n+1} = \ldots = D_{N^{(r)}} = 0] \\
&\quad - \log\binom{N}{n} - (N-N^{(r)})\log(\alpha) - Cf(N),
\end{aligned}$$

where $\mathbf{X}^* = (\mathbf{X}_{n+1}^\top, \ldots, \mathbf{X}_{N^{(r)}}^\top)^\top$ and $k(\mathbf{X}^*; \boldsymbol{\psi})$ is the conditional probability function of $\mathbf{X}^*$ given $D_{n+1} = \ldots = D_{N^{(r)}} = 0$.

It follows from Jensen's inequality that

$$\begin{aligned}
& \kappa(\boldsymbol{\psi}^{(r+1)}, \boldsymbol{\psi}^{(r)}) - \kappa(\boldsymbol{\psi}^{(r)}, \boldsymbol{\psi}^{(r)}) \\
= \; & \mathbb{E}_{\boldsymbol{\psi}^{(r)}}\left[\log\left\{\frac{k(\mathbf{X}^*; \boldsymbol{\psi}^{(r+1)})}{k(\mathbf{X}^*; \boldsymbol{\psi}^{(r)})}\right\} \Big| D_{n+1} = \ldots = D_{N^{(r)}} = 0\right] \\
& + \log\binom{N^{(r)}}{n} - \log\binom{N^{(r+1)}}{n} - (N^{(r+1)} - N^{(r)})\log(\alpha^{(r+1)}) \\
& + C\{f(N^{(r)}) - f(N^{(r+1)})\} \\
\leq \; & \log\binom{N^{(r)}}{n} + (N^{(r)} - n)\log(\alpha^{(r+1)}) + Cf(N^{(r)}) \\
& - \left\{\log\binom{N^{(r+1)}}{n} + (N^{(r+1)} - n)\log(\alpha^{(r+1)}) + Cf(N^{(r+1)})\right\} \\
\leq \; & 0,
\end{aligned}$$

where the last inequality holds because $N^{(r+1)}$ maximizes the function $\log\binom{N}{n} + (N - n)\log(\alpha^{(r+1)}) + Cf(N)$. Then,

$$\begin{aligned}
\widetilde{\ell}_p(\boldsymbol{\psi}^{(r)}) &= \mathcal{Q}(\boldsymbol{\psi}^{(r)}|\boldsymbol{\psi}^{(r)}) - \kappa(\boldsymbol{\psi}^{(r)}, \boldsymbol{\psi}^{(r)}) \\
&\leq \mathcal{Q}(\boldsymbol{\psi}^{(r+1)}|\boldsymbol{\psi}^{(r)}) - \kappa(\boldsymbol{\psi}^{(r+1)}, \boldsymbol{\psi}^{(r)}) \\
&= \widetilde{\ell}_p(\boldsymbol{\psi}^{(r+1)}).
\end{aligned}$$

□

1.3.3 Proof of Result (c)

Proof. Let $\boldsymbol{\theta} = (N, \boldsymbol{\beta}^\top, \alpha)^\top$ and $\boldsymbol{\psi} = (\boldsymbol{\theta}^\top, p_1, \ldots, p_n)^\top$ with $\boldsymbol{\theta}^{(r)}$ and $\boldsymbol{\psi}^{(r)}$ being their values at the rth iteration ($r = 0, 1, \ldots$). To prove Result (c), we first note that $\widehat{\boldsymbol{\theta}}_p = (\widehat{N}_p, \widehat{\boldsymbol{\beta}}_p^\top, \widehat{\alpha}_p)^\top$

is the solution of

$$\begin{aligned} S_1(N,n) - \log(\alpha) - 2C(N-\widetilde{N}_c)I(N>\widetilde{N}_c) &= 0, & (18)\\ \frac{N-n}{\alpha} + \sum_{i=1}^{n} \frac{\xi}{1+\xi\{\phi(\mathbf{x}_i;\boldsymbol{\beta})-\alpha\}} &= 0, & (19)\\ \sum_{i=1}^{n}\sum_{k=1}^{K}\{d_{ik}-g(\mathbf{z}_{ik};\boldsymbol{\beta})\}\mathbf{z}_{ik} - \sum_{i=1}^{n} \frac{\xi\dot{\phi}(\mathbf{x}_i;\boldsymbol{\beta})}{1+\xi\{\phi(\mathbf{x}_i;\boldsymbol{\beta})-\alpha\}} &= \mathbf{0}, & (20)\\ \sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i;\boldsymbol{\beta})-\alpha}{1+\xi\{\phi(\mathbf{x}_i;\boldsymbol{\beta})-\alpha\}} &= 0, & (21) \end{aligned}$$

where $S_1(N,n) = \partial \log\{\Gamma(N+1)\}/\partial N - \partial\log\{\Gamma(N-n+1)\}/\partial N$. Combining (19) with (21) gives $\xi = -(N-n)/(n\alpha)$.

Recall the conditional expectaton of the complete-data log-likelihood $\mathcal{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)}) = \ell_1(\boldsymbol{\beta}) + \ell_2(p_1,\ldots,p_n)$, where

$$\ell_1(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left(\sum_{k=1}^{K}[d_{ik}\log\{g(\mathbf{z}_{ik};\boldsymbol{\beta})\} + (1-d_{ik})\log\{1-g(\mathbf{z}_{ik};\boldsymbol{\beta})\}] + w_i^{(r)}\log\{\phi(\mathbf{x}_i;\boldsymbol{\beta})\}\right),$$

$$\ell_2(p_1,\ldots,p_n) = \sum_{i=1}^{n}(w_i^{(r)}+1)\log(p_i), \quad w_i^{(r)} = (N^{(r)}-n)\phi(\mathbf{x}_i;\boldsymbol{\beta}^{(r)})p_i^{(r)}/\alpha^{(r)}.$$

Maximizing $\mathcal{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)})$ or $\ell_1(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ gives $\boldsymbol{\beta}^{(r+1)}$, which satisfies

$$\frac{\partial \ell_1(\boldsymbol{\beta}^{(r+1)})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}\sum_{k=1}^{K}\{d_{ik}-g(\mathbf{z}_{ik};\boldsymbol{\beta}^{(r+1)})\}\mathbf{z}_{ik} + \sum_{i=1}^{n}\frac{w_i^{(r)}\dot{\phi}(\mathbf{x}_i;\boldsymbol{\beta}^{(r+1)})}{\phi(\mathbf{x}_i;\boldsymbol{\beta}^{(r+1)})} = \mathbf{0}. \qquad (22)$$

Under the constraint $\sum_{i=1}^{n} p_i = 1$, maximizing $\mathcal{Q}(\boldsymbol{\psi}|\boldsymbol{\psi}^{(r)})$ or $\ell_2(p_1,\ldots,p_n)$ with respect to

p_i's gives the $(r+1)$th iteration of p_i's, namely

$$p_i^{(r+1)} = \frac{w_i^{(r)} + 1}{n + \sum_{i=1}^n w_i^{(r)}} = \frac{w_i^{(r)} + 1}{N^{(r)}}, \quad i = 1, \ldots, n. \tag{23}$$

Note that $\alpha^{(r+1)}$ and $N^{(r+1)}$ satisfy

$$\alpha^{(r+1)} = \sum_{i=1}^n \phi(\mathbf{x}_i; \boldsymbol{\beta}^{(r+1)}) p_i^{(r+1)}, \tag{24}$$

$$S_1(N^{(r+1)}, n) - \log(\alpha^{(r+1)}) - 2C(N^{(r+1)} - \widetilde{N}_c) I(N^{(r+1)} > \widetilde{N}_c) = 0. \tag{25}$$

Without loss of generality, we assume that $\boldsymbol{\psi}^{(r+1)}$ and $\boldsymbol{\psi}^{(r)}$ converge to $\boldsymbol{\psi}$. After substituting $w_i^{(r)} = (N^{(r)} - n)\phi(\mathbf{x}_i; \boldsymbol{\beta}^{(r)}) p_i^{(r)} / \alpha^{(r)}$ and $\boldsymbol{\psi}^{(r+1)} = \boldsymbol{\psi}^{(r)} = \boldsymbol{\psi}$ into Equations (22)–(25), it can be verified that

$$\begin{aligned}
(23) \quad &\Rightarrow \quad p_i = \frac{\alpha}{\alpha N - (N-n)\phi(\mathbf{x}_i; \boldsymbol{\beta})}, \quad i = 1, 2, \ldots, n, \\
(25) \quad &\Rightarrow \quad S_1(N, n) - \log(\alpha) - 2C(N - \widetilde{N}_c) I(N > \widetilde{N}_c) = 0, \\
\sum_{i=1}^n p_i = 1 \quad &\Rightarrow \quad \sum_{i=1}^n \frac{1}{\alpha N - (N-n)\phi(\mathbf{x}_i; \boldsymbol{\beta})} = \frac{1}{\alpha}, \\
(22) \quad &\Rightarrow \quad \sum_{i=1}^n \sum_{k=1}^K \{d_{ik} - g(\mathbf{z}_{ik}; \boldsymbol{\beta})\} \mathbf{z}_{ik} + \sum_{i=1}^n \frac{(N-n)\dot{\phi}(\mathbf{x}_i; \boldsymbol{\beta})}{\alpha N - (N-n)\phi(\mathbf{x}_i; \boldsymbol{\beta})} = \mathbf{0}, \\
(24) \quad &\Rightarrow \quad \sum_{i=1}^n \frac{\phi(\mathbf{x}_i; \boldsymbol{\beta})}{\alpha N - (N-n)\phi(\mathbf{x}_i; \boldsymbol{\beta})} = 1.
\end{aligned}$$

The last four equations agree with Equations (18)–(21), respectively. This guarantees that the sequence $\{\boldsymbol{\psi}^{(r)} : r = 1, 2, \ldots, \}$ converges to a local maximum PEL estimator. □

1.4 Proof of Proposition 4

Proposition 4. *In the framework of Remark 2 in the main paper, the sequence of EM iterations* $\{(N^{(r)}, \boldsymbol{\beta}^{(r)}) : r = 1, 2\}$ *converge to a local maximum CL estimator* $(\widehat{N}_c, \widehat{\boldsymbol{\beta}}_c)$.

Proof. By definition, the maximum CL estimator $\widehat{\boldsymbol{\theta}}_c = (\widehat{N}_c, \widehat{\boldsymbol{\beta}}_c^\top)^\top$ is the solution of

$$N - \sum_{i=1}^{n} \frac{1}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta})} = 0, \tag{26}$$

$$\sum_{i=1}^{n} \sum_{k=1}^{K} \{d_{ik} - g(\mathbf{z}_{ik}; \boldsymbol{\beta})\}\mathbf{z}_{ik} - \sum_{i=1}^{n} \frac{\dot{\phi}(\mathbf{x}_i; \boldsymbol{\beta})}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta})\}} = \mathbf{0}. \tag{27}$$

The remaining proof is similar to that of Result (c) in Theorem 3. For the CL method, the iterations $\{(N^{(r)}, \boldsymbol{\beta}^{(r)}, \alpha^{(r)}, p_i^{(r)}) : r = 0, 1, \ldots\}$ satisfy Equations (22), (23), (24), and

$$N^{(r+1)} = \frac{n}{1 - \alpha^{(r+1)}}. \tag{28}$$

Without loss of generality, we assume that $\boldsymbol{\psi}^{(r+1)}$ and $\boldsymbol{\psi}^{(r)}$ converge to $\boldsymbol{\psi}$. After substituting $w_i^{(r)} = (N^{(r)} - n)\phi(\mathbf{x}_i; \boldsymbol{\beta}^{(r)})p_i^{(r)}/\alpha^{(r)}$ and $\boldsymbol{\psi}^{(r+1)} = \boldsymbol{\psi}^{(r)} = \boldsymbol{\psi}$ into Equations (22)–(24) and (28), we have

$$\begin{aligned}
(23)\ \&\ (28) \quad &\Rightarrow \quad p_i = \frac{1}{N\{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta})\}}, \quad i = 1, 2, \ldots, n, \\
\sum_{i=1}^{n} p_i = 1 \quad &\Rightarrow \quad \sum_{i=1}^{n} \frac{1}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta})} = N, \\
(22) \quad &\Rightarrow \quad \sum_{i=1}^{n} \sum_{k=1}^{K} \{d_{ik} - g(\mathbf{z}_{ik}; \boldsymbol{\beta})\}\mathbf{z}_{ik} + \sum_{i=1}^{n} \frac{\dot{\phi}(\mathbf{x}_i; \boldsymbol{\beta})}{1 - \phi(\mathbf{x}_i; \boldsymbol{\beta})} = \mathbf{0}.
\end{aligned}$$

Note that the last two equations respectively agree with Equations (26)–(27). This means that the EM iterations converge to a local maximum CL estimator. □

1.5 Proof of Theorem 5

Theorem 5. *Let* $(N_0, \boldsymbol{\beta}_0, \eta_0, \alpha_0)$ *with* $\alpha_0 \in (0,1)$ *be the true value of* $(N, \boldsymbol{\beta}, \eta, \alpha)$. *Define*

$$\mathbf{W} = \begin{pmatrix} -\mathrm{V}_{11} & \mathbf{0} & -\mathrm{V}_{14} \\ \mathbf{0} & -\mathbf{V}_{\Theta\Theta} + \mathbf{V}_{\Theta 5}\mathrm{V}_{55}^{-1}\mathbf{V}_{5\Theta} & -\mathbf{V}_{\Theta 4} + \mathbf{V}_{\Theta 5}\mathrm{V}_{55}^{-1}\mathrm{V}_{54} \\ -\mathrm{V}_{41} & -\mathbf{V}_{4\Theta} + \mathrm{V}_{45}\mathrm{V}_{55}^{-1}\mathbf{V}_{5\Theta} & -\mathrm{V}_{44} + \mathrm{V}_{45}\mathrm{V}_{55}^{-1}\mathrm{V}_{54} \end{pmatrix}, \tag{29}$$

where $\mathrm{V}_{11} = 1 - \alpha_0^{-1}$, $\mathrm{V}_{14} = \mathrm{V}_{41} = \alpha_0^{-1}$, $\mathbf{V}_{\Theta\Theta} = (\mathbf{V}_{ij})_{2\le i,j\le 3}$, $\mathbf{V}_{4\Theta} = \mathbf{V}_{\Theta 4}^\top = (\mathbf{V}_{42}, 0)$, $\mathbf{V}_{5\Theta} = \mathbf{V}_{\Theta 5}^\top = (\mathbf{V}_{52}, 0)$, $\varphi = \mathbb{E}\{1 - \phi(\mathbf{X};\boldsymbol{\beta}_0)\}^{-1}$, *and*

$$\begin{aligned}
\mathbf{V}_{22} &= \mathbb{E}\left[\frac{\tau^2 e^{2\boldsymbol{\beta}_0^\top \mathbf{X}}\phi(\mathbf{X};\boldsymbol{\beta}_0)}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)} + (e^{\eta_0}-1)\{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)\} - \tau e^{\beta_0^\top \mathbf{X}+\eta_0}\right]\mathbf{X}^{\otimes 2}, \\
\mathbf{V}_{23} &= \mathbf{V}_{32}^\top = \mathbb{E}\left[e^{\eta_0}\{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)\} - \tau e^{\beta_0^\top \mathbf{X}+\eta_0}\right]\mathbf{X}, \quad \mathbf{V}_{24} = \mathbf{V}_{42}^\top = \mathbb{E}\left\{\frac{\tau e^{\boldsymbol{\beta}_0^\top \mathbf{X}}\phi(\mathbf{X};\boldsymbol{\beta}_0)\mathbf{X}}{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)}\right\}, \\
\mathbf{V}_{25} &= \mathbf{V}_{52}^\top = (1-\alpha_0)^2\mathbf{V}_{24}, \quad \mathrm{V}_{33} = \mathbb{E}\left[e^{\eta_0}\{1-\phi(\mathbf{X};\boldsymbol{\beta}_0)\} - \tau e^{\beta_0^\top \mathbf{X}+\eta_0}\right], \\
\mathrm{V}_{44} &= -\alpha_0^{-1} + \varphi, \quad \mathrm{V}_{45} = \mathrm{V}_{54} = (1-\alpha_0)^2\varphi, \quad \mathrm{V}_{55} = (1-\alpha_0)^4\varphi - (1-\alpha_0)^3.
\end{aligned}$$

Suppose that the matrix $\mathbf{W}$ *is positive definite. When* $f(N) = -(N - \widetilde{N}_c)^2 I(N > \widetilde{N}_c)$ *and* $C = O_p(N_0^{-2})$, *as* $N_0 \to \infty$,

- *(a)* $\sqrt{N_0}\{\log(\widehat{N}_p/N_0), (\widehat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_0)^\top, \widehat{\eta}_p - \eta_0, \widehat{\alpha}_p - \alpha_0\}^\top \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{W}^{-1})$;
- *(b)* $R_p(N_0, \boldsymbol{\beta}_0, \eta_0, \alpha_0) \xrightarrow{d} \chi^2_{3+s}$ *and* $R'_p(N_0) \xrightarrow{d} \chi^2_1$ *where* s *is the dimension of* $\boldsymbol{\beta}$.

Proof. The profile penalized log EL under continuous-time model M_{hb} is

$$\begin{aligned}
\ell_p(N, \boldsymbol{\beta}, \phi, \alpha) &= \log\binom{N}{n} + (N-n)\log(\alpha) - \sum_{i=1}^n \log[1 + \xi\{\phi(\mathbf{x}_i;\boldsymbol{\beta}) - \alpha\}] + Cf(N) \\
&\quad + \sum_{i=1}^n \left[m_i\boldsymbol{\beta}^\top\mathbf{x}_i + (m_i - 1)\eta - \{\tau e^\eta + t_{i1}(1-e^\eta)\}\, e^{\boldsymbol{\beta}^\top\mathbf{x}_i}\right],
\end{aligned}$$

where $\phi(\mathbf{x};\boldsymbol{\beta}) = \exp(-\tau e^{\boldsymbol{\beta}^\top \mathbf{x}})$ and $\xi = \xi(\boldsymbol{\beta}, \alpha)$ satisfies

$$\sum_{i=1}^{n} \frac{\phi(\mathbf{x}_i;\boldsymbol{\beta}) - \alpha}{1 + \xi\{\phi(\mathbf{x}_i;\boldsymbol{\beta}) - \alpha\}} = 0. \tag{30}$$

It can be verified that $\xi_0 = -1/(1-\alpha_0)$ is the limit of $\xi(\widehat{\boldsymbol{\beta}}_p, \widehat{\alpha}_p)$ which is the solution of (30) with $(\widehat{\boldsymbol{\beta}}_p, \widehat{\alpha}_p)$ in place of $(\boldsymbol{\beta}, \alpha)$. Define

$$\begin{aligned} \hbar(N, \boldsymbol{\beta}, \eta, \alpha, \xi) &= \log \binom{N}{n} + (N-n)\log(\alpha) - \sum_{i=1}^{n} \log[1 + \xi\{\phi(\mathbf{x}_i;\boldsymbol{\beta}) - \alpha\}] + Cf(N) \\ &\quad + \sum_{i=1}^{n} [m_i \boldsymbol{\beta}^\top \mathbf{x}_i + (m_i - 1)\eta - \{\tau e^{\eta} + t_{i1}(1 - e^{\eta})\} e^{\boldsymbol{\beta}^\top \mathbf{x}_i}]. \end{aligned}$$

It can be seen that $\ell_p(N, \boldsymbol{\beta}, \phi, \alpha) = \hbar(N, \boldsymbol{\beta}, \phi, \alpha, \xi_*)$, where ξ_* is the solution to $\partial\hbar/\partial\xi = 0$. Denote $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2^\top, \theta_3, \theta_4, \theta_5)^\top$, where $\theta_1 = \sqrt{N_0}(N/N_0 - 1)$, $\boldsymbol{\theta}_2 = \sqrt{N_0}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, $\theta_3 = \sqrt{N_0}(\eta - \eta_0)$, $\theta_4 = \sqrt{N_0}(\alpha - \alpha_0)$, and $\theta_5 = \sqrt{N_0}(\xi - \xi_0)$. Define

$$\mathcal{H}(\boldsymbol{\theta}) = \hbar(N_0 + N_0^{1/2}\theta_1, \boldsymbol{\beta}_0 + N_0^{-1/2}\boldsymbol{\theta}_2, \eta_0 + N_0^{-1/2}\theta_3, \alpha_0 + N_0^{-1/2}\theta_4, \xi_0 + N_0^{-1/2}\theta_5).$$

As clarified in the proof of Theorem 1, the proof can be done if we derive the formulae of $\mathbf{V}$ and $\boldsymbol{\Sigma}$. It follows from the law of large numbers and the central limit theorem that

$$\begin{aligned} \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_1} &= N_0^{1/2}\left(\frac{n/N_0 - 1}{\alpha_0} + 1\right) + O_p(N_0^{-1/2}), \\ \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \boldsymbol{\theta}_2} &= N_0^{-1/2} \sum_{i=1}^{n} \left[m_i - \tau\gamma_{i0} e^{\eta_0} + t_{i1}(e^{\eta_0} - 1)\gamma_{i0} - \frac{\tau\gamma_{i0}\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)}{1 - \phi(\mathbf{x}_i;\boldsymbol{\beta}_0)} \right] \mathbf{x}_i, \\ \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_3} &= N_0^{-1/2} \sum_{i=1}^{n} \{m_i - 1 - (\tau - t_{i1})\gamma_{i0} e^{\eta_0}\}, \\ \frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_4} &= N_0^{-1/2} \left\{ \frac{N_0 - n}{\alpha_0} - \sum_{i=1}^{n} \frac{1}{1 - \phi(\mathbf{x}_i;\boldsymbol{\beta}_0)} \right\}, \end{aligned}$$

$$
\frac{\partial \mathcal{H}(\mathbf{0})}{\partial \theta_5} = -N_0^{-1/2}(1-\alpha_0)\sum_{i=1}^{n}\frac{\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)-\alpha_0}{1-\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)},
$$

where $\gamma_{i0} = e^{\boldsymbol{\beta}_0^\top \mathbf{x}_i}$ and $\phi(\mathbf{x}_i;\boldsymbol{\beta}_0) = e^{-\tau\gamma_{i0}}$ for $i = 1,\ldots,n$. Let $\partial\mathcal{H}(\mathbf{0})/\partial\boldsymbol{\theta} = \boldsymbol{u}_n + O_p(N_0^{-1/2})$, where $\boldsymbol{u}_n = (u_{n1}, \boldsymbol{u}_{n2}^\top, u_{n3}, u_{n4}, u_{n5})^\top$ and

$$
u_{n1} = N_0^{1/2}\left(\frac{n/N_0 - 1}{\alpha_0} + 1\right),\ \boldsymbol{u}_{n2} = \frac{\partial\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2},\ u_{nj} = \frac{\partial\mathcal{H}(\mathbf{0})}{\partial\theta_j},\ j = 3,4,5.
$$

For the second partial derivative of $\mathcal{H}$ at $\boldsymbol{\theta} = \mathbf{0}$, we have

$$
\begin{aligned}
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_1^2} &= 1-\alpha_0^{-1} + O_p(N_0^{-1/2}), \quad \frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_4\partial\theta_1} = \frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_1\partial\theta_4} = \alpha_0^{-1}, \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top} &= N_0^{-1}\sum_{i=1}^{n}\left[-\{\tau e^{\eta_0} + t_{i1}(1-e^{\eta_0})\}\gamma_{i0} - \frac{e^{-\tau\gamma_{i0}}\tau\gamma_{i0}(1-\tau\gamma_{i0})}{1-e^{-\tau\gamma_{i0}}}\right. \\
&\quad \left. + \frac{\tau^2\gamma_{i0}^2 e^{-2\tau\gamma_{i0}}}{(1-e^{-\tau\gamma_{i0}})^2}\right]\mathbf{x}_i^{\otimes 2}, \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2\partial\theta_3} &= \left\{\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_3\partial\boldsymbol{\theta}_2^\top}\right\}^\top = -N_0^{-1}\sum_{i=1}^{n}\left\{(\tau - t_{i1})e^{\eta_0}\gamma_{i0}\mathbf{x}_i\right\}, \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2\partial\theta_4} &= \left\{\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_4\partial\boldsymbol{\theta}_2^\top}\right\}^\top = N_0^{-1}\sum_{i=1}^{n}\frac{\tau\gamma_{i0}\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)\mathbf{x}_i}{\{1-\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)\}^2}, \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2\partial\theta_5} &= \left\{\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_5\partial\boldsymbol{\theta}_2^\top}\right\}^\top = (1-\alpha_0)^2 N_0^{-1}\sum_{i=1}^{n}\frac{\tau\gamma_{i0}\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)\mathbf{x}_i}{\{1-\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)\}^2}, \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_3^2} &= -N_0^{-1}\sum_{i=1}^{n}\{(\tau - t_{i1})e^{\eta_0}\gamma_{i0}\}, \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_4^2} &= N_0^{-1}\left[-\frac{N_0 - n}{\alpha_0^2} + \sum_{i=1}^{n}\frac{1}{\{1-\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)\}^2}\right], \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_4\partial\theta_5} &= \frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_5\partial\theta_4} = (1-\alpha_0)^2 N_0^{-1}\sum_{i=1}^{n}\frac{1}{\{1-\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)\}^2}, \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_5^2} &= (1-\alpha_0)^2 N_0^{-1}\sum_{i=1}^{n}\frac{\{\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)-\alpha_0\}^2}{\{1-\phi(\mathbf{x}_i;\boldsymbol{\beta}_0)\}^2}.
\end{aligned}
$$

In the process of deriving $\mathbf{V}$ and $\mathbf{\Sigma}$, we need to calculate the limit of $\partial^2\mathcal{H}(\mathbf{0})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ and the mean and variance of $\boldsymbol{u}_n$. The following lemma eases much of the calculation burden in our proofs.

Lamma 2. *Define $\gamma_0 = e^{\boldsymbol{\beta}_0^\top\mathbf{X}}$ and let T_1 denote the time of an individual being first captured. Given the covariate $\mathbf{X}$, the following conditional expectations hold:*

$$
\begin{aligned}
\mathbb{E}\{I(\mathcal{N}(\tau)>0)|\mathbf{X}\} &= 1-e^{-\tau\gamma_0}, \quad \mathbb{E}\{\mathcal{N}(\tau)|\mathbf{X}\} = 1-e^{-\tau\gamma_0}+\tau e^{\eta_0}\gamma_0 - e^{\eta_0}+e^{\eta_0-\tau\gamma_0},\\
\mathbb{E}\{T_1 I(\mathcal{N}(\tau)>0)|\mathbf{X}\} &= 1/\gamma_0 - \tau e^{-\tau\gamma_0} - e^{-\tau\gamma_0}/\gamma_0,\\
\mathbb{E}\{\mathcal{N}^2(\tau)|\mathbf{X}\} &= 1-e^{-\tau\gamma_0}+3\tau\gamma_0 e^{\eta_0}-3e^{\eta_0}+3e^{\eta_0-\tau\gamma_0}+\tau^2\gamma_0^2 e^{2\eta_0}\\
&\quad -2\tau\gamma_0 e^{2\eta_0}+2e^{2\eta_0}-2e^{2\eta_0-\tau\gamma_0},\\
\mathbb{E}\{T_1\mathcal{N}(\tau)|\mathbf{X}\} &= \tau e^{\eta_0}-\tau e^{-\tau\gamma_0}+\tau e^{\eta_0-\tau\gamma_0}+1/\gamma_0-e^{-\tau\gamma_0}/\gamma_0\\
&\quad -2e^{\eta_0}/\gamma_0+2e^{\eta_0-\tau\gamma_0}/\gamma_0,\\
\mathbb{E}\{T_1^2 I(\mathcal{N}(\tau)>0)|\mathbf{X}\} &= 2/\gamma_0^2-2e^{-\tau\gamma_0}/\gamma_0^2-2\tau e^{-\tau\gamma_0}/\gamma_0-\tau^2 e^{-\tau\gamma_0}.
\end{aligned}
$$

Proof. The first equation is derived from the fact that $\mathcal{N}(\tau)$ follows a Poisson distribution $\text{Poi}(\tau\gamma_0)$ given $\mathbf{X}$. By the property of conditional expectation and the integration by parts, we have

$$
\begin{aligned}
\mathbb{E}\{\mathcal{N}(\tau)|\mathbf{X}\} &= \mathbb{E}[\mathbb{E}\{\mathcal{N}(\tau)|T_1,\mathbf{X}\}|\mathbf{X}] = \int_0^\tau \{1+(\tau-t)\gamma_0 e^{\eta_0}\}\gamma_0 e^{-\gamma_0 t}\mathrm{d}t\\
&= 1-e^{-\tau\gamma_0}+\tau e^{\eta_0}\gamma_0-e^{\eta_0}+e^{\eta_0-\tau\gamma_0},\\
\mathbb{E}\{T_1 I(\mathcal{N}(\tau)>0)|\mathbf{X}\} &= \mathbb{E}[T_1\mathbb{E}\{I(\mathcal{N}(\tau)>0)|T_1,\mathbf{X}\}|\mathbf{X}] = \int_0^\tau t\gamma_0 e^{-\gamma_0 t}\mathrm{d}t\\
&= 1/\gamma_0-\tau e^{-\tau\gamma_0}-e^{-\tau\gamma_0}/\gamma_0,\\
\mathbb{E}\{\mathcal{N}^2(\tau)|\mathbf{X}\} &= \mathbb{E}[\mathbb{E}\{\mathcal{N}^2(\tau)|T_1,\mathbf{X}\}|\mathbf{X}]\\
&= \int_0^\tau [\mathbb{V}\text{ar}\{\mathcal{N}(\tau)|T_1=t,\mathbf{X}\}+\mathbb{E}^2\{\mathcal{N}(\tau)|T_1=t,\mathbf{X}\}]\gamma_0 e^{-\gamma_0 t}\mathrm{d}t\\
&= \int_0^\tau [(\tau-t)\gamma_0 e^{\eta_0}+\{1+(\tau-t)\gamma_0 e^{\eta_0}\}^2]\gamma_0 e^{-\gamma_0 t}\mathrm{d}t
\end{aligned}
$$

$$
\begin{aligned}
&= 1 - e^{-\tau\gamma_0} + 3\tau\gamma_0 e^{\eta_0} - 3e^{\eta_0} + 3e^{\eta_0 - \tau\gamma_0} + \tau^2\gamma_0^2 e^{2\eta_0} \\
&\quad -2\tau\gamma_0 e^{2\eta_0} + 2e^{2\eta_0} - 2e^{2\eta_0 - \tau\gamma_0}, \\
\mathbb{E}\{T_1\mathcal{N}(\tau)|\mathbf{X}\} &= \int_0^\tau \mathbb{E}\{\mathcal{N}(\tau)|T_1 = t\} t\gamma_0 e^{-\gamma_0 t}\mathrm{d}t \\
&= \int_0^\tau \{1 + (\tau - t)\gamma_0 e^{\eta_0}\} t\gamma_0 e^{-\gamma_0 t}\mathrm{d}t \\
&= \tau e^{\eta_0} - \tau e^{-\tau\gamma_0} + \tau e^{\eta_0 - \tau\gamma_0} + 1/\gamma_0 - e^{-\tau\gamma_0}/\gamma_0 \\
&\quad -2e^{\eta_0}/\gamma_0 + 2e^{\eta_0 - \tau\gamma_0}/\gamma_0, \\
\mathbb{E}\{T_1^2 I(\mathcal{N}(\tau) > 0)|\mathbf{X}\} &= \int_0^\tau t^2\gamma_0 e^{-\gamma_0 t}\mathrm{d}t = 2/\gamma_0^2 - 2e^{-\tau\gamma_0}/\gamma_0^2 - 2\tau e^{-\tau\gamma_0}/\gamma_0 - \tau^2 e^{-\tau\gamma_0}.
\end{aligned}
$$

□

Applying Lemma 2 and the central limit theorem, it can be verified that

$$
\begin{aligned}
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2\partial\boldsymbol{\theta}_2^\top} &= \mathbb{E}\left(\frac{\tau^2\gamma_0^2 e^{-2\tau\gamma_0}}{1 - e^{-\tau\gamma_0}} + \tau^2\gamma_0^2 e^{-\tau\gamma_0} - 1 - \tau\gamma_0 e^{\eta_0} - e^{\eta_0 - \tau\gamma_0} + e^{\eta_0} + e^{-\tau\gamma_0}\right)\mathbf{X} \\
&\quad + O_p(N_0^{-1/2}), \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2\partial\theta_3} &= \left\{\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_3\partial\boldsymbol{\theta}_2^\top}\right\}^\top = -\mathbb{E}\left(\tau\gamma_0 e^{\eta_0} + e^{\eta_0 - \tau\gamma_0} - e^{\eta_0}\right)\mathbf{X} + O_p(N_0^{-1/2}), \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2\partial\theta_4} &= \left\{\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_4\partial\boldsymbol{\theta}_2^\top}\right\}^\top = -\mathbb{E}\frac{\dot\phi(\mathbf{X};\boldsymbol{\beta}_0)}{1 - \phi(\mathbf{X};\boldsymbol{\beta}_0)} + O_p(N_0^{-1/2}), \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\boldsymbol{\theta}_2\partial\theta_5} &= \left\{\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_5\partial\boldsymbol{\theta}_2^\top}\right\}^\top = -(1 - \alpha_0)^2\mathbb{E}\frac{\dot\phi(\mathbf{X};\boldsymbol{\beta}_0)}{1 - \phi(\mathbf{X};\boldsymbol{\beta}_0)} + O_p(N_0^{-1/2}) \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_3^2} &= -\mathbb{E}(\tau\gamma_0 e^{\eta_0} + e^{\eta_0 - \tau\gamma_0} - e^{\eta_0}) + O_p(N_0^{-1/2}), \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_4^2} &= -\alpha_0^{-1} + \mathbb{E}\frac{1}{1 - \phi(\mathbf{X};\boldsymbol{\beta}_0)} + O_p(N_0^{-1/2}), \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_4\partial\theta_5} &= \frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_5\partial\theta_4} = (1 - \alpha_0)^2\mathbb{E}\frac{1}{1 - \phi(\mathbf{X};\boldsymbol{\beta}_0)} + O_p(N_0^{-1/2}), \\
\frac{\partial^2\mathcal{H}(\mathbf{0})}{\partial\theta_5^2} &= (1 - \alpha_0)^4\mathbb{E}\frac{1}{1 - \phi(\mathbf{X};\boldsymbol{\beta}_0)} - (1 - \alpha_0)^3 + O_p(N_0^{-1/2}).
\end{aligned}
$$

The matrix $\mathbf{V}$, which is the leading term of $\partial^2 \mathcal{H}(\mathbf{0})/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top)$, is

$$\mathbf{V} = \begin{pmatrix} \mathrm{V}_{11} & \mathbf{0}_{1\times s} & 0 & \mathrm{V}_{14} & 0 \\ \mathbf{0}_{s\times 1} & \mathbf{V}_{22} & \mathbf{V}_{23} & \mathbf{V}_{24} & \mathbf{V}_{25} \\ 0 & \mathbf{V}_{32} & \mathrm{V}_{33} & 0 & 0 \\ \mathrm{V}_{41} & \mathbf{V}_{42} & 0 & \mathrm{V}_{44} & \mathrm{V}_{45} \\ 0 & \mathbf{V}_{52} & 0 & \mathrm{V}_{54} & \mathrm{V}_{55} \end{pmatrix},$$

where V_{ij}'s are defined in Theorem 5. Using Lemma 2, it can be verified that $\mathbb{E}(\boldsymbol{u}_n) = \mathbf{0}$ and

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(u_{n1}) &= \frac{N_0}{\alpha_0^2} \cdot \frac{(1-\alpha_0)\alpha_0}{N_0} = \alpha_0^{-1} - 1, \quad \mathbb{C}\mathrm{ov}(u_{n4}, u_{n1}) = -\alpha_0^{-1}, \\
\mathbb{V}\mathrm{ar}(\boldsymbol{u}_{n2}) &= \mathbb{E}\left[1 - e^{-\tau\gamma_0} + \tau\gamma_0 e^{\eta_0} - e^{\eta_0} + e^{\eta_0 - \tau\gamma_0} - \tau^2\gamma_0^2 e^{-\tau\gamma_0} - \frac{\tau^2\gamma_0^2 e^{-2\tau\gamma_0}}{1 - e^{-\tau\gamma_0}}\right] \mathbf{X}^{\otimes 2} \\
&= -\mathbf{V}_{22}, \qquad \mathbb{C}\mathrm{ov}(u_1, u_3) = \mathbb{C}\mathrm{ov}(u_1, u_5) = 0, \\
\mathbb{C}\mathrm{ov}(\boldsymbol{u}_{n2}, u_{n3}) &= \mathbb{E}\{(\tau\gamma_0 e^{\eta_0} + e^{\eta_0 - \tau\gamma_0} - e^{\eta_0})\mathbf{X}\} = -\mathbf{V}_{23}, \\
\mathbb{C}\mathrm{ov}(\boldsymbol{u}_{n2}, u_{n1}) &= \mathbb{C}\mathrm{ov}(\boldsymbol{u}_{n2}, u_{n4}) = \mathbb{C}\mathrm{ov}(\boldsymbol{u}_{n2}, u_{n5}) = \mathbf{0}_{s\times 1}, \\
\mathbb{V}\mathrm{ar}(u_{n3}) &= \mathbb{E}\left(\tau\gamma_0 e^{\eta_0} + e^{\eta_0 - \tau\gamma_0} - e^{\eta_0}\right) = -\mathrm{V}_{33}, \\
\mathbb{C}\mathrm{ov}(u_{n3}, u_{n4}) &= \mathbb{C}\mathrm{ov}(u_{n3}, u_{n5}) = 0, \quad \mathbb{V}\mathrm{ar}(u_{n4}) = \alpha_0^{-1} + \varphi, \\
\mathbb{C}\mathrm{ov}(u_{n4}, u_{n5}) &= (1-\alpha_0)^2\varphi - (1-\alpha_0), \quad \mathbb{V}\mathrm{ar}(u_{n5}) = (1-\alpha_0)^4\varphi - (1-\alpha_0)^3.
\end{aligned}$$

By the central limit theorem, we have $\boldsymbol{u}_n \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ as $N_0 \to \infty$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} -\mathrm{V}_{11} & \mathbf{0}_{1\times s} & 0 & -\mathrm{V}_{14} & 0 \\ \mathbf{0}_{s\times 1} & -\mathbf{V}_{22} & -\mathrm{V}_{23} & \mathbf{0}_{s\times 1} & \mathbf{0}_{s\times 1} \\ 0 & -\mathrm{V}_{23} & -\mathrm{V}_{33} & 0 & 0 \\ -\mathrm{V}_{41} & \mathbf{0}_{1\times s} & 0 & 2\mathrm{V}_{45}(1-\alpha_0)^{-2} - \mathrm{V}_{44} & \mathrm{V}_{55}(1-\alpha_0)^{-2} \\ 0 & \mathbf{0}_{1\times s} & 0 & \mathrm{V}_{55}(1-\alpha_0)^{-2} & \mathrm{V}_{55} \end{pmatrix}.$$

Note that $\boldsymbol{\Sigma}$ has the same form as that in Lemma 3 of the supplementary material of Liu et al. (2017), so does the matrix $\mathbf{W}$. This completes the proof of Theorem 5. $\square$

1.6 Proof of Proposition 6

Proposition 6. *Define* $\Theta = (\boldsymbol{\beta}^\top, \eta)^\top$, $\widehat{\Theta}_p = (\widehat{\boldsymbol{\beta}}_p^\top, \widehat{\eta}_p)^\top$, *and* $\widehat{\Theta}_c = (\widehat{\boldsymbol{\beta}}_c^\top, \widehat{\eta}_c)^\top$. *Under the conditions in Theorem 5, as* $N_0 \to \infty$,

(a) $\widehat{\Theta}_p - \widehat{\Theta}_c = O_p(N_0^{-1})$ *and* $\widehat{N}_p - \widehat{N}_c = O_p(1)$;

(b) $\sqrt{N_0}(\widehat{\Theta}_p - \Theta_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, -\mathbf{V}_{\Theta\Theta}^{-1})$ *and* $\sqrt{N_0}(\widehat{\Theta}_p - \Theta_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, -\mathbf{V}_{\Theta\Theta}^{-1})$;

(c) $N_0^{-1/2}(\widehat{N}_p - N_0) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ *and* $N_0^{-1/2}(\widehat{N}_c - N_0) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$, *where* $\sigma^2 = \varphi - 1 - \mathbf{V}_{4\Theta}\mathbf{V}_{\Theta\Theta}^{-1}\mathbf{V}_{\Theta 4}$.

Proof. Since the matrix $\mathbf{W}$ defined in (29) of Theorem 5 has the same form as (1) in Theorem 1, the proof of this proposition is similar to that of Proposition 2 and hence omitted. $\square$

1.7 Proof of Theorem 7

Theorem 7. *Under continuous-time capture–recapture models, the EM algorithm proposed guarantees that the penalized log EL* $\widetilde{\ell}_p(N, \boldsymbol{\beta}, \eta, \alpha, \{p_i\})$ *increases after each EM iteration.*

Proof. The proof is the same as that of Theorem 3 (a) except that $\phi(x; \beta) = \exp\{-\tau \exp(\mathbf{x}^\top \beta)\}$ in this case.

$\square$

2 Simulations for continuous-time models

In this subsection, we investigate the finite-sample performance of the proposed PEL and the EM algorithm under continuous-time capture–recapture models. This section can be seen as a complement to the simulation studies in the main paper. Here, we focus on the capture intensity model M_{hb} and generate data from the following two scenarios:

(D) Set $N_0 = 200$. Consider the covariate $\mathbf{X} = (1, X)^\top$, where $X \sim \mathrm{N}(0, 1)$. Let $\boldsymbol{\beta}_0 = (-0.7, 1.2)^\top$ and $\eta_0 = 0.5$.

(E) Same as (D) except that $\eta_0 = -0.5$.

In each scenario, we set the endpoint to $\tau = 0.4$, 0.6, and 1.0, where the corresponding capture probability varies from 26% to 45%. For each simulated data set, we deploy the proposed EM algorithm to implement the CL, EL, and PEL estimation methods under the model M_{hb}. We compare their performances based on 500 simulated samples.

Table 1 presents the simulated root mean square errors (RMSEs) of the maximum CL, EL, and PEL estimators of N. It is clear that the maximum PEL estimator $\widehat{N}_p$ always has the smallest RMSE, followed by the maximum EL estimator $\widehat{N}_e$, and the maximum CL estimator $\widehat{N}_c$ has the largest RMSE in most cases. This demonstrate the advantage of the PEL method or adding penalty on EL. In Scenario E with $\tau = 0.4$, $\widehat{N}_c$ has a smaller RMSE than $\widehat{N}_e$, which seems controversial to the findings in Liu et al. (2018). A possible reason for this phenomenon is that the EM algorithm might stop far away from the "true" estimates when the CL or EL function is large.

Table 1: Simulation results under continuous-time capture–recapture models.

	Scenario D						Scenario E					
	RMSE			Level: 95%			RMSE			Level: 95%		
τ	$\widehat{N}_c$	$\widehat{N}_e$	$\widehat{N}_p$	$\mathcal{I}_c$	$\mathcal{I}_e$	$\mathcal{I}_p$	$\widehat{N}_c$	$\widehat{N}_e$	$\widehat{N}_p$	$\mathcal{I}_c$	$\mathcal{I}_e$	$\mathcal{I}_p$
0.4	578.7	429.5	114.0	86.4	94.4	93.4	3677.1	4391.6	216.7	85.4	94.6	94.8
0.6	110.5	107.8	77.9	90.6	94.8	94.8	175.9	170.7	131.2	88.2	93.4	93.8
1.0	63.5	62.6	54.8	91.2	95.0	94.4	70.4	68.0	66.6	90.8	94.4	94.4

Next, we investigate the performance of the CL ($\mathcal{I}_c$), EL ($\mathcal{I}_e$) and PEL ($\mathcal{I}_p$) confidence intervals at the 95% level. Table 1 presents their coverage probabilities. We can see that the

EL and PEL confidence intervals have comparable coverage probabilities, which are closer to 95% than those of the Wald-type confidence interval. Figure 1 displays the boxplots of the logarithm of interval widths. There are many extremely long widths of $\mathcal{I}_c$ and $\mathcal{I}_e$ especially in the low capture probability case. In Scenario D (E) with $\tau = 0.4$, 29% (32%) of the upper limits of $\mathcal{I}_e$ are greater than 200 $\times n$, which is undesirable. In contrast, there are no such case for the proposed PEL interval $\mathcal{I}_p$. This indicates that the PEL can successfully overcome the unstability of the EL and produce much better interval estimates.

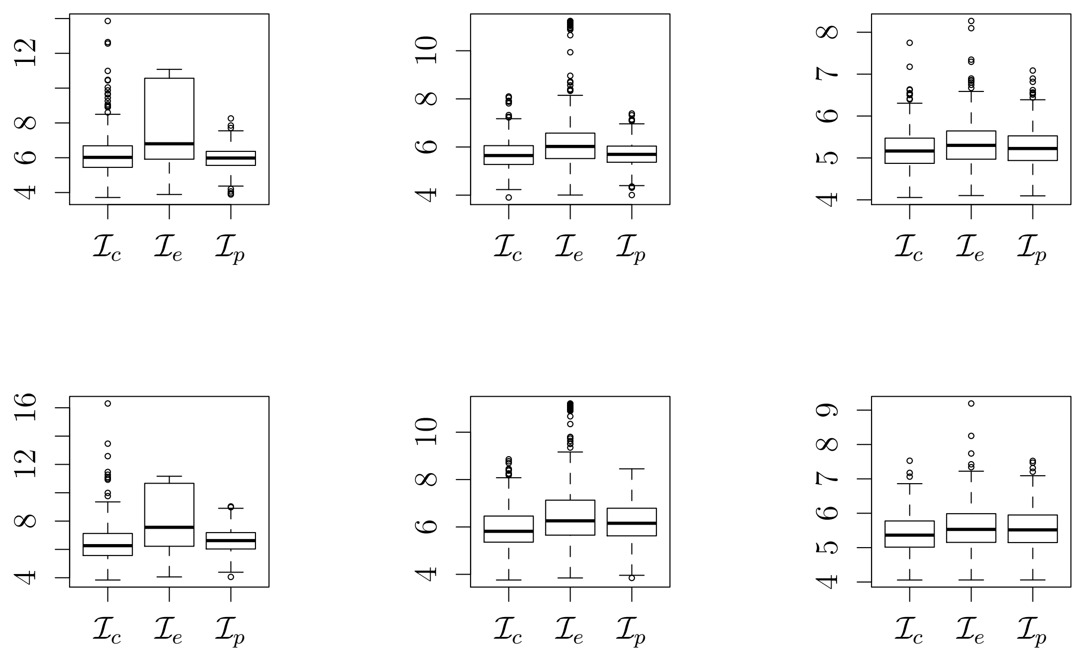

Figure 1: Boxplots of the logarithm of lengths of the CL ($\mathcal{I}_c$), EL ($\mathcal{I}_e$) and PEL ($\mathcal{I}_p$) confidence intervals at the 95% level when $\tau = 0.4$ (column 1), 0.6 (column 2) and 1.0 (column 3) in Scenarios D (row 1) and E (row 2).

3 Extension to the ephemeral behavioral models

In the main paper, we have considered the behavioral effect with enduring (long-term) memory. That is, after an individual is captured, the individual has a long memory of its first-capture experience and the effect lasts in the remaining period of the experiment, leading to a higher (trap-happy) or lower (trap-shy) capture probability for all subsequent recaptures.

For individuals with short-term memory, the capture probability of an individual may be ephemeral and depends on whether or not it was caught on the most recent occasion (Yang and Chao, 2005). In this section, we take into account this kind of ephemeral behavioral effect and consider a general model M_{thbc_1}, where the subscripts h, t, b, and c_1 represent individual heterogeneity, time, enduring behavioral effect, and ephemeral behavioral effect, respectively. Similarly, we may consider M_{hc_1}, M_{htc_1}, and M_{hbc_1} as special cases.

Under model M_{hc_1}, M_{htc_1}, M_{hbc_1}, and M_{htbc_1}, we assume that the conditional probability of the individual being captured on the kth occasion is

$$P(D_{(k)} = 1|\mathbf{X} = \mathbf{x}, D_{(1)} = d_{(1)}, \ldots, D_{(k-1)} = d_{(k-1)}) = \frac{\exp(\boldsymbol{\beta}^{\top}\mathbf{z}_k + \gamma d_{(k-1)})}{1 + \exp(\boldsymbol{\beta}^{\top}\mathbf{z}_k + \gamma d_{(k-1)})}$$

for $k = 1, \ldots, K$, where the unknown parameter γ measures the ephemeral behavioral effect, $\boldsymbol{\beta}$ and $\mathbf{z}_k$ are same as those under model M_h, M_{ht}, M_{hb}, and M_{htb} defined in Table 1 of the main paper.

With similar arguments, we can show that Theorems 1 and 3 and Propositions 2 and 4 also hold under the general models involving the ephemeral behavioral effect. This means that the PEL estimation approach and the EM algorithm proposed in Sections 2 and 3 of the main paper are applicable to these general models.

3.1 Simulation studies

We carry out simulations to investigate the proposed method and algorithms under models M_{hc_1} and M_{hbc_1}. We set $N_0 = 200$ and $K = 6$, and generate the capture–recapture data from the following two scenarios:

(F) Let the covariate vector be $\mathbf{X} = (X_1, X_2)^\top$, where $X_1 \sim \mathrm{N}(0,1)$ and $X_2 \sim \mathrm{Bi}(1, 0.5)$. Consider the capture probability model M_{hc_1} with $\boldsymbol{\beta} = (0.1, -2.5, -0.15)^\top$ and $\gamma = 0.8$.

(G) Same as Scenario (F) except that model M_{hbc_1} is considered with $\boldsymbol{\beta} = (0.1, -2.5, -0.15, 0.3)^\top$ and $\gamma = 0.5$.

All simulation results were obtained based on 500 repetitions. Table 2 presents the root mean square errors (RMSEs) of the point estimators and the coverage probabilities and widths of the interval estimators for the abundance N based on the CL, EL and PEL methods. From Table 2, we have similar findings to those in Section 5 of the main paper. When model M_{hc_1} or M_{hbc_1} is correctly specified, in terms of RMSE, the PEL estimator performs the best, the EL estimator $\widehat{N}_e$ second best, and the CL estimator worst. For interval estimation, both the PEL and EL interval estimators have much better coverage accuracy than the CL-based Wald-type interval estimator, which has severe undercoverage. Although the PEL and EL intervals have close coverage probabilities, the PEL interval has much shorter widths. In summary, the PEL estimation method has the best overall performance among the CL, EL and PEL methods.

Table 2: Root mean square errors (RMSEs) of the CL, EL, and PEL abundance estimators, coverage probabilities (unit: %) and average widths of the CL, EL and PEL confidence intervals at the 95% level.

	RMSE			Coverage probability			Interval width		
Scenario	$\widehat{N}_c$	$\widehat{N}_e$	$\widehat{N}_p$	$\mathcal{I}_c$	$\mathcal{I}_e$	$\mathcal{I}_p$	$\mathcal{I}_c$	$\mathcal{I}_e$	$\mathcal{I}_p$
D	28	24	22	90.00	92.28	92.22	97	111	93
E	32	28	24	89.20	93.60	93.80	111	129	100

3.2 Analysis of the Black bear data

For illustration, we reanalyze the black bear data in Section 6 of the main paper. Besides enduring behavioral effect, we also take ephemeral behavioral effect into account and use the proposed EM algorithm to implement the CL, EL, and PEL estimation methods for the population size under models M_{hc_1}, M_{hbc_1}, M_{htc_1}, and M_{htbc_1}. Here, M_{htc_1} and M_{htbc_1} also incorporate the time effect on the capture probability.

Table 3 reports the analysis results of the CL, EL and PEL methods under these four models. No matter which model is assumed, the EL and PEL estimates always have smaller standard errors than the CL estimates. Under the most general model M_{htbc_1}, the CL and EL intervals have extremely large upper limits and widths. The upper limit of the EL interval is even not available. In contrast, the PEL ratio confidence interval is always desirable.

Table 3: Analysis results of the black bear data.†

Method	Est.	SE	CI	AIC	Est.	SE	CI	AIC
	Model M_{hc_1}				Model M_{hbc_1}			
CL	57	4.99	[48, 67]	446.07	67	16.87	[34, 101]	446.79
EL	56	4.83	[49, 71]	815.93	63	13.18	[49, 182]	817.12
PEL	56	4.83	[49, 71]	815.93	63	13.18	[49, 150]	817.12
	Model M_{htc_1}				Model M_{htbc_1}			
CL	57	5.20	[46, 67]	446.46	139	254.87	[0, 639]	446.65
EL	55	5.00	[49, 69]	816.25	73	38.41	[48, –]‡	817.59
PEL	55	5.00	[49, 69]	816.25	72	35.82	[48, 269]	817.60

† Est.: point estimate, SE: standard error, and CI: confidence interval at the 95% confidence level.
‡ –: A number greater than 10^9.

To determine the type of behavioral effect, we conduct EL ratio tests for two hypotheses, M_{hc_1} versus M_{hbc_1} and M_{hb} versus M_{hbc_1}. The resulting p-values are 37% and 0.2‰, respectively, indicating that the ephemeral behavioral effect is present while the enduring behavioral effect is not at the 5% significance level. We also report the Akaive information criterion (AIC) values for the four models under study. Based on the EL and PEL methods, model M_{hc_1} has the smallest AIC and hence can be regarded as the most suitable model for the Black bear data.

References

Chao, A. (1987). Estimating the population size for capture–recapture data with unequal catchability. *Biometrics 43*(4), 783–791.

Liu, Y., P. Li, and J. Qin (2017). Maximum empirical likelihood estimation for abundance in a closed population from capture–recapture data. *Biometrika 104*(3), 527–543.

Liu, Y., Y. Liu, P. Li, and J. Qin (2018). Full likelihood inference for abundance from continuous time capture–recapture data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(5), 995–1014.

Yang, H.-C. and A. Chao (2005). Modeling animals' behavioral response by Markov chain models for capture–recapture experiments. *Biometrics 61*(4), 1010–1017.