Supplementary material for "Laplace Approximations for Capture-Recapture Models in the Presence of Individual Heterogeneity"

Appendix A: Derivatives for M_h-type models

In this appendix, we calculate the derivatives of the joint density of the complete data likelihood and associated random effects for M_h-type models with respect to the individual random effects to obtain the Laplace approximation of the (marginal) observed data likelihood. For $i = 1, \dots, n$, we let $g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)$ denote the objective function corresponding to the negative of the log of the joint density of the observed capture history for individual i, denoted $\boldsymbol{x}_i$, and associated individual random effect density, ϵ_i, given the model model parameters, $\boldsymbol{\theta}$, and individual heterogeneity variance, σ^2, i.e. $g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2) = -\log f(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)$. Similarly we let $g(\boldsymbol{x}_0, \epsilon_0|\boldsymbol{\theta}, \sigma^2)$ denote the analogous objective function for the null history, $\boldsymbol{x}_0$ (i.e. an unobserved individual). For $i = 0, 1, \dots, n$, the corresponding objective function is of the form:

$$\begin{aligned} g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2) &= -\log f(\boldsymbol{x}_i|\boldsymbol{\theta}, \epsilon_i) - \log f(\epsilon_i|\sigma^2) \\ &= -\sum_{t=1}^{T} [x_{it}\log(p_{it}) + (1-x_{it})\log(1-p_{it})] + \frac{1}{2}\log(2\pi\sigma^2) + \frac{\epsilon_i}{2\sigma^2}. \end{aligned}$$

For notational simplicity, we let $\eta_{it} = \alpha_t + \lambda S_{it} + \epsilon_i$. The capture probabilities are assumed to be logistically regressed on the covariate, so that

$$p_{it} = \frac{\exp(\eta_{it})}{1+\exp(\eta_{it})} = \frac{1}{1+\exp(-\eta_{it})}; \text{ and } 1 - p_{it} = \frac{1}{1+\exp(\eta_{it})}.$$

The first derivative of $g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)$ with respect to ϵ_i is given by,

$$\begin{aligned}
\frac{dg(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)}{d\epsilon_i} &= \sum_{t=1}^{T}\left[\frac{-x_{it}\exp(-\eta_{it})}{1+\exp(-\eta_{it})} + \frac{(1-x_{it})\exp(\eta_{it})}{1+\exp(\eta_{it})}\right] + \frac{\epsilon_i}{\sigma^2} \\
&= \sum_{t=1}^{T}\left[\frac{-x_{it}}{1+\exp(\eta_{it})} + \frac{(1-x_{it})}{1+\exp(-\eta_{it})}\right] + \frac{\epsilon_i}{\sigma^2} \\
&= \sum_{t=1}^{T}(p_{it} - x_{it}) + \frac{\epsilon_i}{\sigma^2}.
\end{aligned}$$

Similarly, the second derivative is given by:

$$\begin{aligned}
g''(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2) = \frac{d^2 g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)}{d\epsilon_i^2} &= \sum_{t=1}^{T}\left[\frac{\exp(-\eta_{it})}{\{1+\exp(-\eta_{it})\}^2}\right] + \frac{1}{\sigma^2} \\
&= \sum_{t=1}^{T}\left[p_{it}\left(1-p_{it}\right)\right] + \frac{1}{\sigma^2}.
\end{aligned}$$

Thus the negative log likelihood of the M_h-type model (given in Equation (2) of the main paper) can be estimated using the second order Laplace approximation,

$$\begin{aligned}
-\log f(\boldsymbol{x}|\boldsymbol{\theta}, \sigma^2) = &-\log(N!) + \log(N-n)! \\
&+ \sum_{i=1}^{n}\left[g(\boldsymbol{x}_i, \hat{\epsilon}_i|\boldsymbol{\theta}, \sigma^2) + \frac{1}{2}\log g''(\boldsymbol{x}_i, \hat{\epsilon}_i|\boldsymbol{\theta}, \sigma^2) - \frac{1}{2}\log(2\pi)\right] \\
&+ (N-n)\left[g(\boldsymbol{x}_0, \hat{\epsilon}_0|\boldsymbol{\theta}, \sigma^2) + \frac{1}{2}\log g''(\boldsymbol{x}_0, \hat{\epsilon}_0|\boldsymbol{\theta}, \sigma^2) - \frac{1}{2}\log(2\pi)\right].
\end{aligned}$$

To obtain the closed form of the fourth order Laplace approximation, we require higher order derivatives, in terms of the third and the fourth derivatives, with respect to ϵ_i. For $i = 0, \dots, n$, the third derivative of $g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)$ is given by:

$$\begin{aligned} g^{(3)}(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2) = \frac{d^3 g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)}{d\epsilon_i^3} &= \sum_{t=1}^{T} \left[\frac{\exp(\eta_{it})}{\{1+\exp(\eta_{it})\}^3} - \frac{\exp(-\eta_{it})}{\{1+\exp(-\eta_{it})\}^3} \right] \\ &= \sum_{t=1}^{T} \left[p_{it} \left(1-p_{it}\right)^2 - p_{it}^2 \left(1-p_{it}\right) \right]. \end{aligned}$$

Similarly, the fourth derivative of $g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)$ is given by:

$$\begin{aligned} g^{(4)}(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2) = \frac{d^4 g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma^2)}{d\epsilon_i^4} &= \sum_{t=1}^{T} \left[\frac{\exp(\eta_{it})}{\{1+\exp(\eta_{it})\}^4} - \frac{4\exp(2\eta_{it})}{\{1+\exp(\eta_{it})\}^4} + \frac{\exp(-\eta_{it})}{\{1+\exp(-\eta_{it})\}^4} \right] \\ &= \sum_{t=1}^{T} \left[p_{it}(1-p_{it})^3 - 4p_{it}^2(1-p_{it})^2 + p_{it}^3(1-p_{it}) \right]. \end{aligned}$$

These analytic expressions can be substituted into the higher order Laplace approximation given in Equation (5) of the main paper. The objective function including the inner optimisation is coded in C++ utilizing the TMB library. The estimation of the model parameters is obtained by subsequently minimizing the negative log likelihood which is evaluated in TMB at given parameter values (for $\boldsymbol{\theta}$ and σ^2) using standard optimisation routines in `R`.

Appendix B: Simulation study

We provide further details of the simulation study described in Section 4 of the paper.

Closed M_h-type models

We consider two simulation studies. The first investigates the performance of the Laplace approximation for the four individual heterogeneity models, given a set of parameter values;

while the second focuses on the individual heterogeneity variance term, σ^2, given model M_h. For each dataset simulated we fit the generating model using both the second-order (LA2) and fourth-order (LA4) Laplace approximations and compare with a GHQ approach using 50 quadrature points as a "gold standard". For each method we calculate the MLE and 95% confidence intervals using a non-parametric bootstrap approach (see King and McCrea, 2019). In particular, to compute the associated 95% confidence intervals, we simulate B bootstrap replicates of the data, such that each simulated dataset is of the same size as the original data (i.e. same number of observed individuals). Each bootstrap dataset is simulated by randomly drawing with replacement each observed capture history with equal probability. We fit the model to each bootstrap dataset and obtain the associated MLEs of the model parameters. We add the original MLEs of the parameters to this set of values, so that we have $B + 1$ parameter values corresponding to the MLEs of the parameters from the original and bootstrap datasets. Finally, the 95% confidence interval for each parameter is calculated as the associated lower and upper 2.5% quantile values of the $B + 1$ values. In practice we used $B = 999$.

For the first simulation study we set the total population size to be $N = 100$ individuals with $T = 6$ capture occasions and consider the four different individual heterogeneity models: M_h, M_{th}, M_{bh} and M_{tbh}. The parameters specified for the parameters were motivated by the snowshore hare dataset (see for example, Cormack (1989); Baillargeon and Rivest (2007)), after fitting the given models to the data. For the time-invariant models (M_h and M_{bh}) we set $\alpha_t = \alpha = -1$ for $t = 1, \ldots, T$. For the time-varying models, for M_{th} we set $\boldsymbol{\alpha} = \{\alpha_t : t = 1, \ldots, T\} = \{-1.75, -0.91, -1.44, -1.03, -1.22, -0.67\}$; and for M_{tbh}, $\boldsymbol{\alpha} = \{-1.49, -0.29, -0.44, 0.15, 0.10, 0.81)\}$, based on the parameter estimates from fitting the models to the data. For the models with a behavioural effect (M_{bh} and M_{tbh}), we specify a "trap happy" response, with $\lambda = 0.75$. Finally, for the individual heterogeneity component we set $\sigma^2 = 0.75^2$. For models M_h and M_{bh} the probability of an individual not being observed within the study is 0.8; for models M_{th} this probability is 0.77; and for model M_{tbh}, 0.95. 1000 datasets were simulated for each model.

Table 1 provides the average relative bias (RB),95% coverage probabilities (CP) and the average width of the 95% confidence intervals (width) for the 1000 simulated datasets for the two parameters of interest, population size, N, and individual heterogeneity standard deviation, σ, for each model and model-fitting approach. In general, across all approaches and models, the MLEs of the parameters, N and σ, appear to be consistently slightly negatively biased for all models. This bias is, however, consistently less for the LA4 and HQ approaches, relative to LA2. Given this larger bias, it is perhaps unsurprising that the associated coverage probabilities for LA2 are also consistently lower. This suggests that the second-order Taylor series expansion in the standard Laplace approximation is not sufficient to approximate the integral in the observed data likelihood of the M_h-type models. However adding in the higher-order Laplace approximation terms does improve the performance of the algorithm, with very similar performance between LA4 and GHQ in terms of both relative bias (though the bias appears to be very slightly less using the Laplace approximation) and coverage probabilities.

For the second simulation study, we consider model M_h and investigate the performance of the Laplace approximations and GHQ for differing values of the individual heterogeneity variance, σ^2. Increasing the individual heterogeneity increases the variability of the survival probabilities of individuals in the study and thus is likely to be increasing challenging, see White and Cooch (2017) for further discussion. Motivated by the golf tees data (see Section 5.1 of the main paper), we set $N = 250, T = 8$ and $\alpha = -1.5$ for the simulation study. We then consider a range of values for the individual heterogeneity variance, such that $\sigma = 1, 1.5, 2$. For each σ value we simulate 1000 datasets.

[Table 1 about here.]

Table 2 provides the average relative bias, the 95% coverage probabilities and mean 95%

confidence intervals width for N and σ for the different model-fitting approaches. As for the previous simulation study, the LA2 approach has the poorest performance; while LA4 and GHQ perform better and have similar relative biases and coverage probabilities. However, interestingly despite these similar coverage probabilities, for GHQ the width of the confidence interval increases as σ increases, due to the long tails for the upper bound; while this relationship is not present for LA4 with similar width confidence intervals across the different values of σ. These findings are consistent with the real data golf tees example in Section 5.1, where the 95% non-parametric confidence intervals of the quadrature approach are substantially wider for each of the four individual heterogeneity models. Finally we note that both the standard error of σ and relative bias of N both increase as σ increases for all the model-fitting approaches.

[Table 2 about here.]

CJS model with missing continuous covariates

We consider the CJS model where we specify the survival probability as a function of a single individual covariate for 2 different covariate models:

- Model 1: $y_{i,t+1}|y_{it} \sim N(y_{it}, \sigma_y^2)$ (a simple random walk);
- Model 2: $y_{i,t+1}|y_{it} \sim N(y_{it} + \mu_t, \sigma_y^2)$ (a random walk with additional temporal effects).

We consider a range of scenarios motivated by the real meadow vole data considered in Section 5.2: (i) we initially set $n = 200$ and consider a constant capture probability for for two different regimes ($p = 0.5, 0.75$) for studies of length $T = 4, 6$; (ii) to investigate the sample size, we then set $T = 4$ and repeat the simulation study but increase n to 400 (with same constant capture probabilities as before). For all studies we set $\sigma_y = 1.2$. The initial covariate value for each individual at the time of initial capture is simulated from a Normal distribution with mean of 15 and standard deviation of 2, i.e. $y_{if_i} \sim \mathrm{N}(15, 2^2)$ with f_i randomly sampled from $\{1, \ldots, T -$

$1\}$. For covariate model 2 we simulate $\mu_t \sim \mathrm{N}(0, 1)$ for $t = 1, \ldots, T-1$, independently for each simulated dataset. The survival probability was specified as a logistic regression on the individual covariate with regression parameters $\beta_0 = -3.0$ and $\beta_1 = 0.2$. For each model and parameter combination we simulated 1000 datasets.

For each dataset, we fit the Laplace approximation and compared with an HMM-approximation (Langrock and King, 2013). Tables 3 and 4 provides the corresponding averaged relative biases (RB), 95% coverage probabilities (CP) and the mean 95% confidence interval widths of the regression coefficients, β_0 and β_1 across the generated datasets. Overall, both the Laplace and HMM approximations perform well with all coverage probabilities $> 94\%$ and small relative biases.

[Table 3 about here.]

[Table 4 about here.]

Computational comparisons

We compare the computational time of the Laplace approximations with other competing methods (GHQ for closed populations; and HMM approximations for open populations) for the simulation studies. Each algorithm was run on a 1.70 GHz Intel Core CPU i5-8350U computer with Windows 10. All methods were implemented in the TMB package for comparability.

For the closed population simulation study, we consider model M_h. The Laplace approximations (LA2 and LA4) are nearly twice as fast as the GHQ approach (using 50 quadrature points) in evaluating the log-likelihood function. For more complex models, the difference in computational speed is even greater. For example, for model M_{th}, the differential in computation speed increases to 8-10 times faster for the the Laplace approximations compared to GHQ. Finally, we note that both approximations are "fast" in terms of absolute computational time for these models, with the maximisation of the likelihood to obtain the MLEs of the parameters in the order of seconds.

For the more computationally challenging CJS model with individual continuous covariates, we focus on the computational times associated with model 2. For example, to fit this model for a moderate sized dataset where $n = 500$ and $T = 10$, the Laplace approximation took an average of 1.55 seconds to evaluate the observed data likelihood function; while the HMM approximation using 20 intervals took an average of 70.13 seconds (this is approximately 45 times slower than the Laplace approximation). The computational time of the HMM approach is dependent on the number of intervals used, and this relationship is non-linear. For example, if we increase the number of intervals from 20 to 30 the computational time of the HMM more than doubles, such that it is over 100 times slower than the Laplace approximation.

The Laplace approximation is consistently faster than the considered "gold-standard" approaches (GHQ for M_h-type models; and HMM approximation for the CJS model with individual time-varying continuous covariates). Further, the Laplace approximation requires no tuning parameters, while these alternatives do require some specification (e.g. number of quadrature points or intervals), with an associated trade-off between the computational time and accuracy.

References

Baillargeon, S. and L.-P. Rivest (2007). Rcapture: Loglinear models for capture-recapture in r. *Journal of Statistical Software 19*(5), 1–31.

Cormack, R. M. (1989). Log-linear models for capture-recapture. *Biometrics 45*, 395.

King, R. and R. McCrea (2019). Capture–recapture Methods and Models: Estimating Population Size. In *Handbook of Statistics*, Volume 40, pp. 33–83. Elsevier.

Langrock, R. and R. King (2013). Maximum likelihood estimation of mark–recapture–recovery models in the presence of continuous covariates. *The Annals of Applied Statistics 7*, 1709–1732.

White, G. C. and E. G. Cooch (2017). Population abundance estimation with heterogeneous encounter probabilities using numerical integration: Individual heterogeneity and abundance estimation. *The Journal of Wildlife Management 81*, 322–336.

Table 1: Simulation results in terms of averaged relative bias (RB) and 95% coverage probabilities (CP) for 1000 simulated datasets for the different individual heterogeneity models fitted via Laplace approximations, second-order (LA2) and fourth-order (LA4), and Gauss-Hermite quadrature (GHQ).

Models	Methods	N			σ		
		RB	CI	Width	RB	CI	Width
M_h	LA2	-0.023	0.838	53.569	-0.205	0.833	1.093
	LA4	-0.004	0.884	36.689	-0.098	0.914	1.108
	GHQ	-0.004	0.882	47.833	-0.103	0.905	1.146
M_{th}	LA2	-0.003	0.854	75.585	-0.177	0.861	1.184
	LA4	0.002	0.893	45.736	-0.081	0.916	1.146
	GHQ	0.005	0.894	63.262	-0.084	0.910	1.221
M_{bh}	LA2	0.033	0.898	124.626	-0.142	0.875	1.349
	LA4	0.001	0.921	53.521	-0.075	0.930	1.086
	GHQ	0.005	0.920	66.280	-0.078	0.930	1.119
M_{tbh}	LA2	-0.010	0.859	35.314	-0.144	0.911	0.949
	LA4	-0.007	0.875	47.026	-0.039	0.957	0.994
	GHQ	-0.007	0.873	41.516	-0.044	0.950	0.986

Table 2: Simulation results in terms of averaged relative bias (RB) and 95% coverage probabilities (CP) for 1000 simulated datasets from model M_h for values of σ = 1, 1.5, 2 and fitted via Laplace approximations, second-order (LA2) and fourth-order (LA4), and Gauss-Hermite quadrature (GHQ).

σ	Methods	N			σ		
		RB	CP	Width	RB	CP	Width
2.0	LA2	-0.053	0.883	84.145	-0.100	0.853	0.922
	LA4	0.010	0.929	75.001	0.008	0.952	1.007
	GHQ	0.017	0.929	158.29	-0.001	0.944	1.315
1.5	LA2	-0.034	0.879	93.322	-0.087	0.873	0.794
	LA4	-0.003	0.928	79.066	-0.013	0.935	0.768
	GHQ	-0.003	0.928	127.852	-0.019	0.933	0.942
1.0	LA2	-0.021	0.868	95.648	-0.084	0.869	0.682
	LA4	-0.002	0.903	76.665	-0.016	0.932	0.642
	GHQ	-0.002	0.903	85.726	-0.020	0.929	0.668

Table 3: Simulation results for CJS model with $n = 200$ considering varying values of $T = 4, 6$, $p = 0.5, 0.75$ for the two different covariate models. Presented are the relative bias (RB), 95% coverage probabilities (CP) and the mean 95% confidence interval widths (Width) of the regression parameters for the survival probabilities averaged over 1000 simulated datasets using the Laplace approximation and a hidden Markov model (HMM) approximation respectively. Model 1 corresponds to a simple random walk on the continuous covariate; model 2 to a random walk with additional temporal effects.

(a) Model 1

	p	Methods	β_0			β_1		
			RB	CP	Width	RB	CP	Width
$T = 4$	0.50	Laplace	0.060	0.955	5.585	0.064	0.956	0.378
		HMM	0.062	0.955	5.530	0.064	0.956	0.373
	0.75	Laplace	0.009	0.961	4.269	0.008	0.956	0.282
		HMM	0.010	0.961	4.266	0.009	0.956	0.282
$T = 6$	0.50	Laplace	0.036	0.962	4.485	0.031	0.962	0.293
		HMM	0.041	0.962	4.479	0.035	0.960	0.293
	0.75	Laplace	0.016	0.961	3.733	0.016	0.960	0.243
		HMM	0.017	0.961	3.731	0.016	0.960	0.244

(b) Model 2

	p	Methods	β_0			β_1		
			RB	CP	Width	RB	CP	Width
$T = 4$	0.50	Laplace	0.051	0.968	5.455	0.049	0.967	0.375
		HMM	0.084	0.966	5.545	0.098	0.969	0.384
	0.75	Laplace	0.020	0.959	4.237	0.018	0.959	0.284
		HMM	0.027	0.961	4.245	0.026	0.958	0.282
$T = 6$	0.50	Laplace	0.054	0.957	4.503	0.048	0.956	0.296
		HMM	0.078	0.959	4.513	0.074	0.956	0.298
	0.75	Laplace	0.036	0.943	3.732	0.036	0.945	0.245
		HMM	0.042	0.943	3.734	0.042	0.946	0.245

Table 4: Simulation results for CJS model with $T = 4$ considering varying values of $n = 200, 400$, $p = 0.5, 0.75$ for the two different covariate models. Presented are the relative bias (RB), 95% coverage probabilities (CP) and the mean 95% confidence interval widths (Width) of the regression parameters for the survival probabilities averaged over 1000 simulated datasets for the Laplace approximation and a hidden Markov model (HMM) approximation respectively. Model 1 corresponds to a simple random walk on the continuous covariate; model 2 to a random walk with additional temporal effects.

(a) Model 1

	p	Methods	β_0			β_1		
			RB	CP	Width	RB	CP	Width
$n = 200$	0.50	Laplace	0.060	0.955	5.585	0.064	0.956	0.378
		HMM	0.062	0.955	5.530	0.064	0.956	0.373
	0.75	Laplace	0.009	0.961	4.269	0.008	0.956	0.282
		HMM	0.010	0.961	4.266	0.009	0.956	0.282
$n = 400$	0.50	Laplace	0.028	0.958	3.737	0.030	0.957	0.250
		HMM	0.030	0.958	3.730	0.033	0.956	0.250
	0.75	Laplace	0.003	0.943	2.981	0.003	0.941	0.197
		HMM	0.004	0.943	2.980	0.004	0.941	0.197

(b) Model 2

	p	Methods	β_0			β_1		
			RB	CP	Width	RB	CP	Width
$n = 200$	0.50	Laplace	0.051	0.968	5.455	0.049	0.967	0.375
		HMM	0.084	0.966	5.545	0.098	0.969	0.384
	0.75	Laplace	0.020	0.959	4.237	0.018	0.959	0.284
		HMM	0.027	0.961	4.245	0.026	0.958	0.282
$n = 400$	0.50	Laplace	0.024	0.958	3.713	0.023	0.956	0.254
		HMM	0.050	0.960	3.746	0.053	0.956	0.258
	0.75	Laplace	0.016	0.949	2.971	0.016	0.949	0.199
		HMM	0.022	0.948	2.966	0.023	0.950	0.200