Appendix

Bayesian Model Selection for Spatial Capture Recapture Models

Soumen Dey\textsuperscript{1}, Mohan Delampady\textsuperscript{1}, and Arjun M. Gopalaswamy \textsuperscript{*1}

\textsuperscript{1}Statistics and Mathematics Unit, Indian Statistical Institute, Bangalore 560059, India

A Steps of resampling procedure to obtain MAP estimate using MCMC draws

We describe a MAP approximation approach to compute the Gelfand-Dey estimator of the Bayes factor in Section 2.2.1 of the main text. We denote the collection of all parameters $\mu = (\mu_p, \mu_s)$, where $\mu_p$ is the collection of scalar parameters and $\mu_s$ is the collection of all latent variables. We first fix these high-dimensional variables at their MAP estimate $\hat{\mu}_s$ assuming that the posterior distribution is well summarised by this estimate. The Gelfand-Dey estimator of marginal likelihood of data $m(Y)$ is then computed as (also given as (2.7) in the main text):

$$\hat{m}_{GD}(Y) = \left[ \frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \frac{g(\mu_{p}^{(d)})}{f(Y | \mu_{p}^{(d)}, \hat{\mu}_s) \pi(\mu_{p}^{(d)})} \right]^{-1}.$$  \hspace{1cm} (A.1)

where $\{(\mu_{p}^{(d)}, \mu_{s}^{(d)}) : d = 1, \ldots, N_{\text{iter}}\}$ is a set of MCMC draws from the posterior $\pi(\mu_{p}, \mu_{s} | Y)$, $f(Y | \mu_{p}, \mu_{s})$ denotes the model likelihood and $\pi(\mu_{p})$ denotes the prior density of the parameters $\mu_p$ and $g(\mu_{p})$ denotes a tuning density. We begin with

\textsuperscript{*}arjungswamy@gmail.com (✉)
\((\mu_p^{(d_0)}, \mu_s^{(d_0)})\) as an initial estimate of \((\mu_p, \mu_s)\) where

\[
f(Y | \mu_p^{(d_0)}, \mu_s^{(d_0)}) \pi(\mu_p^{(d_0)}, \mu_s^{(d_0)}) = \max_d \{f(Y | \mu_p^{(d)}, \mu_s^{(d)}) \pi(\mu_p^{(d)}, \mu_s^{(d)})\}.\]

This estimate of posterior mode of \((\mu_p, \mu_s)\) may not be optimal since in our high-dimensional parameter setting, an MCMC sample of a practical size may not be enough to extensively explore the posterior surface. Below we describe a resampling algorithm based on MCMC draws to obtain a better MAP estimate \((\hat{\mu}_p, \hat{\mu}_s)\).

**Iteration 1.** Obtain \(d_1\) such that

\[
f(Y | \mu_p^{(d_1)}, \mu_s^{(d_0)}) \pi(\mu_p^{(d_1)}, \mu_s^{(d_0)}) = \max_d \{f(Y | \mu_p^{(d)}, \mu_s^{(d)}) \pi(\mu_p^{(d)}, \mu_s^{(d)})\}.
\]

- If \(f(Y | \mu_p^{(d_1)}, \mu_s^{(d_0)}) \pi(\mu_p^{(d_1)}, \mu_s^{(d_0)}) \leq f(Y | \mu_p^{(d_0)}, \mu_s^{(d_0)}) \pi(\mu_p^{(d_0)}, \mu_s^{(d_0)})\)
  
  \(\rightarrow\) Stop the iterations, set \((\hat{\mu}_p, \hat{\mu}_s) = (\mu_p^{(d_0)}, \mu_s^{(d_0)})\).

- Otherwise go to the next iteration.

**Iteration 2.** Obtain \(d_2\) such that

\[
f(Y | \mu_p^{(d_1)}, \mu_s^{(d_2)}) \pi(\mu_p^{(d_1)}, \mu_s^{(d_2)}) = \max_d \{f(Y | \mu_p^{(d)}, \mu_s^{(d)}) \pi(\mu_p^{(d)}, \mu_s^{(d)})\}.
\]

- If \(f(Y | \mu_p^{(d_1)}, \mu_s^{(d_2)}) \pi(\mu_p^{(d_1)}, \mu_s^{(d_2)}) \leq f(Y | \mu_p^{(d_1)}, \mu_s^{(d_0)}) \pi(\mu_p^{(d_1)}, \mu_s^{(d_0)})\)
  
  \(\rightarrow\) Stop the iterations, set \((\hat{\mu}_p, \hat{\mu}_s) = (\mu_p^{(d_1)}, \mu_s^{(d_0)})\).

- Otherwise go to the next iteration.

**Iteration 3.** Obtain \(d_3\) such that

\[
f(Y | \mu_p^{(d_3)}, \mu_s^{(d_2)}) \pi(\mu_p^{(d_3)}, \mu_s^{(d_2)}) = \max_d \{f(Y | \mu_p^{(d)}, \mu_s^{(d_2)}) \pi(\mu_p^{(d)}, \mu_s^{(d_2)})\}.
\]

- If \(f(Y | \mu_p^{(d_3)}, \mu_s^{(d_2)}) \pi(\mu_p^{(d_3)}, \mu_s^{(d_2)}) \leq f(Y | \mu_p^{(d_1)}, \mu_s^{(d_2)}) \pi(\mu_p^{(d_1)}, \mu_s^{(d_2)})\)
  
  \(\rightarrow\) Stop the iterations, set \((\hat{\mu}_p, \hat{\mu}_s) = (\mu_p^{(d_1)}, \mu_s^{(d_2)})\).

- Otherwise go to the next iteration.
**Iteration 4.** Obtain $d_4$ such that

\[
f(Y | \mu_p(d_3), \mu_s(d_4)) \pi(\mu_p(d_3), \mu_s(d_4)) = \max_d \{ f(Y | \mu_p(d_3), \mu_s(d)) \pi(\mu_p(d_3), \mu_s(d)) \}.
\]

- If \( f(Y | \mu_p(d_3), \mu_s(d_4)) \pi(\mu_p(d_3), \mu_s(d_4)) \leq f(Y | \mu_p(d_3), \mu_s(d_2)) \pi(\mu_p(d_3), \mu_s(d_2)) \)

  \( \rightarrow \) Stop the iterations, set \((\hat{\mu}_p, \hat{\mu}_s) = (\mu_p(d_3), \mu_s(d_2)).\)

- Otherwise go to the next iteration.

\[\vdots\]

\[\vdots\]

**Iteration m.** (i) If \( m \) is even, obtain $d_m$ such that

\[
f(Y | \mu_p(d_{m-1}), \mu_s(d_m)) \pi(\mu_p(d_{m-1}), \mu_s(d_m)) = \max_d \{ f(Y | \mu_p(d_{m-1}), \mu_s(d)) \pi(\mu_p(d_{m-1}), \mu_s(d)) \}.
\]

- If \( f(Y | \mu_p(d_{m-1}), \mu_s(d_m)) \pi(\mu_p(d_{m-1}), \mu_s(d_m)) \leq f(Y | \mu_p(d_{m-2}), \mu_s(d_{m-2})) \pi(\mu_p(d_{m-1}), \mu_s(d_{m-2})) \)

  \( \rightarrow \) Stop the iterations, set \((\hat{\mu}_p, \hat{\mu}_s) = (\mu_p(d_{m-1}), \mu_s(d_{m-2})).\)

- Otherwise go to the next iteration.

(ii) If \( m \) is odd, obtain $d_m$ such that

\[
f(Y | \mu_p(d_m), \mu_s(d_{m-1})) \pi(\mu_p(d_m), \mu_s(d_{m-1})) = \max_d \{ f(Y | \mu_p(d), \mu_s(d_{m-1})) \pi(\mu_p(d), \mu_s(d_{m-1})) \}.
\]

- If \( f(Y | \mu_p(d_m), \mu_s(d_{m-1})) \pi(\mu_p(d_m), \mu_s(d_{m-1})) \leq f(Y | \mu_p(d_{m-2}), \mu_s(d_{m-1})) \pi(\mu_p(d_{m-2}), \mu_s(d_{m-1})) \)

  \( \rightarrow \) Stop the iterations, set \((\hat{\mu}_p, \hat{\mu}_s) = (\mu_p(d_{m-2}), \mu_s(d_{m-1})).\)

- Otherwise go to the next iteration.

This procedure is continued iteratively to eventually give us the best MAP estimate of the posterior mode \((\hat{\mu}_p, \hat{\mu}_s).\) Observe that, \((\hat{\mu}_p, \hat{\mu}_s)\) is closer to the posterior mode than \((\mu_p^{(d_0)}, \mu_s^{(d_0)})\) since

\[
f(Y | \hat{\mu}_p, \hat{\mu}_s) \pi(\hat{\mu}_p, \hat{\mu}_s) \geq f(Y | \mu_p^{(d_0)}, \mu_s^{(d_0)}) \pi(\mu_p^{(d_0)}, \mu_s^{(d_0)}).
\]
Thus the asymptotic (large posterior sample) convergence of \((\mu_p^{(d_0)}, \mu_s^{(d_0)})\) to the posterior mode ensures that our improved MAP estimate also converges to the same point.

Note that we have ensured via suitable transformations (see Appendix C) on the parameters that all the points in \(\{(\mu_p^{(a)}, \mu_s^{(b)}) : a, b = 1, \ldots, N_{\text{iter}}; a \neq b\}\) do belong to the posterior support. Indeed, we have simply used the MCMC sample itself to explore the posterior surface better, and thus have obtained an improved MAP estimate.

### B Integrated likelihoods

#### B.1 Integrated likelihood under \(M_1\)

The joint density of \(Y^* := (Y^{(1)}, Y^{(2)}) = ((y_{ijk}^{(1)}, y_{ijk}^{(2)}))\) and \(u = (u_1, \ldots, u_M)'\) under \(M_1\) is the following (also given as model likelihood (2.1) in the main text):

\[
\begin{align*}
\text{f}(Y^*, u_{\text{obs}} | \theta, \phi, \omega_0, \sigma_m, \sigma_f, u_0, z, S, L) &= \prod_{i=1}^M \left[ \theta^{n_i} (1 - \theta)^{1-u_i} \phi^{r_i} (1 - \phi)^{2m_i - y_i} \prod_{j=1}^J \eta_j(s_i, u_i)^{n_{ij}} \{1 - \eta_j(s_i, u_i)\}^{1 - \eta_j(s_i, u_i)} \right]^{z_i}. \\
&= \prod_{i: u_i \in u_{\text{obs}}} \left[ \theta^{n_i} (1 - \theta)^{1-u_i} \phi^{r_i} (1 - \phi)^{2m_i - y_i} \prod_{j=1}^J \eta_j(s_i, u_i)^{n_{ij}} \{1 - \eta_j(s_i, u_i)\}^{1 - \eta_j(s_i, u_i)} \right]^{z_i}
\end{align*}
\]

(B.1)

**Marginalisation over \(u_0\).** Each \(u_i \in u\) is assumed to have a Bernoulli distribution with parameter \(\theta\), independent of the other \(u_j\)'s in \(u\). Now we integrate (B.1) over the missing \(u_i\)'s.

\[
\begin{align*}
\text{f}(Y^*, u_{\text{obs}} | \theta, \phi, \omega_0, \sigma_m, \sigma_f, z, S, L) &= \sum_{u_0} \text{f}(Y^*, u_{\text{obs}} | \theta, \phi, \omega_0, \sigma_m, \sigma_f, u_0, z, S, L) \pi(u_0 | z, \theta) \\
&= \prod_{i: u_i \in u_{\text{obs}}} \left[ \theta^{n_i} (1 - \theta)^{1-u_i} \phi^{r_i} (1 - \phi)^{2m_i - y_i} \prod_{j=1}^J \eta_j(s_i, 1)^{n_{ij}} \{1 - \eta_j(s_i, 1)\}^{1 - \eta_j(s_i, 1)} \right]^{z_i}
\end{align*}
\]

(B.2)
We have assumed that all the individuals with recorded sex category are also detected at least once in one of the detectors (i.e. \( y_i^{(1)} + y_i^{(2)} > 0 \)). Whereas among the \( u_0 \) individuals, some may be detected and some may not.

**Marginalisation over \( z \).** Now as mentioned above, each \( z_i \), is assumed to have a Bernoulli prior with parameter \( \psi \), independent of the other \( z_j \)'s. We integrate \( \text{(B.2)} \) over \( z \),

\[
\begin{align*}
    f(Y^*, u_{\text{obs}} \mid \psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, S, L) &= \sum_z f(Y^*, u_{\text{obs}} \mid \theta, \phi, \omega_0, \sigma_m, \sigma_f, z, S, L) \pi(z \mid \psi) \\
    &= \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \theta^{u_i} (1 - \theta)^{1 - u_i} \phi^{y_i} (1 - \phi)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
    &\times \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \phi^{y_i} \left( \psi \phi \right)^{2n_i - y_i} \right] \\
\end{align*}
\]

(B.3)

where \( A_i(s_i, u_i) = \prod_{j=1}^{J} \eta_j(s_i, u_i)^{n_{ij}} \left( \psi + \phi \right)^{K - n_{ij}} \cdot \pi(z \mid \psi) \) is the density of Bernoulli distribution with parameter \( \psi \).
Marginalisation over $\mathbf{S}$. Each activity centre $s_i$ is assumed to have a Uniform distribution over the whole state space $\mathbf{S}$, independent of the other $s_j$'s. Now we integrate (B.3) over $\mathbf{S}$,

$$f(\mathbf{Y}, \mathbf{u}_{\text{obs}} | \psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, \mathbf{L}) = \int_{\mathbf{S}} f(\mathbf{Y}, \mathbf{u}_{\text{obs}} | \psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, \mathbf{S}, \mathbf{L}) \pi(\mathbf{S}) \, d\mathbf{S}$$

$$= \prod_{i : u_i \in \mathbf{u}_{\text{obs}}} \left[ \psi \theta^{u_i} (1 - \theta)^{1 - u_i} \phi^{y_i} (1 - \phi)^{2n_i - y_i} \int_{s_i} A_i(s_i, u_i) \pi(s_i) \, ds_i \right]$$

$$\times \prod_{i : u_i \in \mathbf{u}_{\text{obs}}} \left[ \psi \phi^{y_i} (1 - \phi)^{2n_i - y_i} \int_{s_i} \left\{ \theta A_i(s_i, 1) + (1 - \theta) A_i(s_i, 0) \right\} \pi(s_i) \, ds_i + (1 - \psi) I(y_i = 0) \right]$$

(B.4)

where $\pi(\mathbf{S})$ denotes the density of uniform distribution over the bounded state space $\mathcal{V}$.

To compute the integrals in (B.4) over each $s_i$, first we partition the region $\mathcal{V}$ into a sufficiently fine grid and then evaluate a Riemann sum, since each of those integrals cannot be expressed in closed form. Assume $\{\mathbf{s}(l) : l = 1, 2, \ldots, n_G\}$ are the pixel centres of one such fine grid over $\mathcal{V}$. Then (B.4) becomes

$$f(\mathbf{Y}, \mathbf{u}_{\text{obs}} | \psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, \mathbf{L}) = \int_{\mathbf{S}} f(\mathbf{Y}, \mathbf{u}_{\text{obs}} | \psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, \mathbf{S}, \mathbf{L}) \pi(\mathbf{S}) \, d\mathbf{S}$$

$$\approx \prod_{i : u_i \in \mathbf{u}_{\text{obs}}} \left[ \psi \theta^{u_i} (1 - \theta)^{1 - u_i} \phi^{y_i} (1 - \phi)^{2n_i - y_i} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} A_i(s(l), u_i) \right\} \right]$$

$$\times \prod_{i : u_i \in \mathbf{u}_{\text{obs}}} \left[ \psi \phi^{y_i} (1 - \phi)^{2n_i - y_i} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} \left( \theta A_i(s(l), 1) + (1 - \theta) A_i(s(l), 0) \right) \right\} + (1 - \psi) I(y_i = 0) \right]$$

$$= \prod_{i : u_i \in \mathbf{u}_{\text{obs}}} \left[ \psi \theta^{u_i} (1 - \theta)^{1 - u_i} \phi^{y_i} (1 - \phi)^{2n_i - y_i} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} \prod_{j=1}^{J} \eta_j(s(l), u_i)^{n_{ij}} \right\} \right]$$

$$\times \prod_{i : u_i \in \mathbf{u}_{\text{obs}}} \left[ \psi \phi^{y_i} (1 - \phi)^{2n_i - y_i} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} \left( \theta \prod_{j=1}^{J} \eta_j(s(l), 1)^{n_{ij}} + (1 - \eta_j(s(l), 1)) \eta_j(s(l), 1)(1 - \phi)^2 \right) \right\} + (1 - \psi) I(y_i = 0) \right].$$

(B.5)
Observe that, when $y_{i..} = 0$, $A_i(s^{(l)}, 1) = \prod_{j=1}^{J} \{(1 - \eta_j(s^{(l)}, 1)) + \eta_j(s^{(l)}, 1)(1 - \phi)^2\}^J$, 

$A_i(s^{(l)}, 0) = \prod_{j=1}^{J} \{(1 - \eta_j(s^{(l)}, 0)) + \eta_j(s^{(l)}, 0)(1 - \phi)^2\}^J$.

### B.2 Integrated likelihood under $M_2$

The joint density of $Y^* := (Y^{(1)}, Y^{(2*)}) = ((y_{ij1}, y_{ij2}))$ and $u = (u_1, \ldots, u_M)'$ under $M_2$ is the following (also given as model likelihood (2.2) in the main text):

$$ f_{R}(Y^*, u_{\text{obs}} | \theta, p_0, \sigma_m, \sigma_f, u_0, z, S, L) = \prod_{i=1}^{M} \left[ \left\{ \theta u_i (1 - \theta)^{1 - u_i} \prod_{j=1}^{J} p_j(s_i, u_i)^{y_{ij}} \right\}^z \right], $$

(B.6)

The corresponding integrated likelihood is obtained by integrating (B.6) with respect to $u_0, z, S$:

$$ f_{R}(Y^*, u_{\text{obs}} | \psi, \theta, p_0, \sigma_m, \sigma_f, L) $$

$$ = \int \sum_{S, u_0, z} f(Y^*, u_{\text{obs}} | \psi, \theta, p_0, \sigma_m, \sigma_f, u_0, z, S, L) \pi(u_0 | z, \theta) \pi(z | \psi) \pi(S) dS $$

$$ \approx \prod_{i: u_i \in u_{\text{obs}}} \left[ \psi \theta^{x_i} (1 - \theta)^{1 - x_i} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} B_i(s^{(l)}, u_i) \right\} \right] $$

$$ \times \prod_{i: u_i \in u_0} \left[ \psi \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} \left( \theta B_i(s^{(l)}, 1) + (1 - \theta) B_i(s^{(l)}, 0) \right) \right\} + (1 - \psi) \mathbb{I}(y_{i..} = 0) \right], $$

(B.7)

where $B_i(s, u) = \prod_{j=1}^{J} p_j(s, u)^{y_{ij}} (1 - p_j(s, u))^{2K - y_{ij}}$, $\pi(z | \psi)$ is the density of Bernoulli distribution with parameter $\psi$, $\pi(S)$ denotes the density of uniform distribution over the bounded state space $\mathcal{Y}$. 
B.3 Integrated likelihood under $M_3$

The joint density of $Y^* := (Y^{(1)}, Y^{(2*)}) = ((y_{ijk}^{(1)}, y_{ijk}^{(2*)}))$ under $M_3$ is the following (also given as model likelihood (2.3) in the main text):

$$f(Y^* | \phi, \omega_0, \sigma, z, S, L) = \prod_{i=1}^{M} \left\{ \phi^{y_i} (1 - \phi)^{2n_i - y_i} \prod_{j=1}^{J} \eta_j(s_i)^{n_{ij}} \{(1 - \eta_j(s_i)) + \eta_j(s_i)(1 - \phi)^2\}^{K-n_{ij}} \right\}^{z_i}. \quad (B.8)$$

After integrating the above likelihood (B.8) over $z$ and $S$ with respect to their priors, we obtain

$$f(Y^* | \phi, \omega_0, \sigma, L) = \int_S \sum_z f(Y^* | \phi, \omega_0, \sigma, z, S, L) \pi(z | \psi) \pi(S) dS$$

$$\approx \prod_{i:y_{i*} > 0} \left[ \psi \phi^{y_{i*}} (1 - \phi)^{2n_{i*} - y_{i*}} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} \prod_{j=1}^{J} \eta_j(s^{(l)})^{n_{ij}} \{(1 - \eta_j(s^{(l)})) + \eta_j(s^{(l)})(1 - \phi)^2\}^{K-n_{ij}} \right\} \right]$$

$$\times \prod_{i:y_{i*} = 0} \left[ \psi \phi^{y_{i*}} (1 - \phi)^{2n_{i*} - y_{i*}} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} \prod_{j=1}^{J} \eta_j(s^{(l)})^{n_{ij}} \{(1 - \eta_j(s^{(l)})) + \eta_j(s^{(l)})(1 - \phi)^2\}^{K-n_{ij}} \right\} \right]$$

$$\times (1 - \psi) I(y_{i*} = 0)$$

$$= \prod_{i:y_{i*} > 0} \left[ \psi \phi^{y_{i*}} (1 - \phi)^{2n_{i*} - y_{i*}} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} A_l(s^{(l)}) \right\} \right]$$

$$\times \prod_{i:y_{i*} = 0} \left[ \psi \phi^{y_{i*}} (1 - \phi)^{2n_{i*} - y_{i*}} \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} A_l(s^{(l)}) \right\} + (1 - \psi) \right], \quad (B.9)$$

where $A_l(s) = \prod_{j=1}^{J} \eta_j(s)^{n_{ij}} \{(1 - \eta_j(s)) + \eta_j(s)(1 - \phi)^2\}^{K-n_{ij}}$, $\pi(z | \psi)$ is the density of Bernoulli distribution with parameter $\psi$, $\pi(S)$ denotes the density of uniform distribution over the bounded state space $V$. 

8
B.4 Integrated likelihood under $M_4$

The joint density of $Y^* := (Y^{(1)}, Y^{(2)}) = ((y_{ijk}^{(1)}, y_{ijk}^{(2)}))$ under $M_4$ is the following (also given as model likelihood (2.4) in the main text):

$$f_R(Y^* | p_0, \sigma, z, S, L) = \prod_{i=1}^M \prod_{j=1}^J \left\{ p_j(s_i)^{y_{ij}} (1 - p_j(s_i))^{2K - y_{ij}} \right\}^{\gamma_i}, \quad (B.10)$$

The integrated likelihood in the absence of individual covariate on sex category $u$ is obtained by integrating (B.10) with respect to $z, S$, and is given by

$$f_R(Y^* | \psi, p_0, \sigma, L) = \int \sum_{z, S} f_R(Y^* | \psi, p_0, \sigma, z, S, L) \pi(z | \psi) \pi(S) dS \approx \prod_{i : y_{ij} > 0} \left[ \psi \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} B_l(s^{(l)}) \right\} \right] \times \prod_{i : y_{ij} = 0} \left[ \psi \left\{ \frac{1}{n_G} \sum_{l=1}^{n_G} B_l(s^{(l)}) \right\} + (1 - \psi) \right], \quad (B.11)$$

where $B_l(s) = \prod_{j=1}^J p_j(s)^{y_{ij}} (1 - p_j(s))^{2K - y_{ij}}$, $\pi(z | \psi)$ is the density of Bernoulli distribution with parameter $\psi$ and $\pi(S)$ denotes the density of uniform distribution over the bounded state space $V$.

C Transformed prior distributions and tuning densities

Under our model settings, $\mu$ denotes the collection of all parameters and latent variables for each model as a generic notation. Specifically, we let $\mu = (\mu_p, \mu_s)$, where $\mu_p$ is the collection of scalar parameters and $\mu_s$ is the collection of all latent variables. As indicated in Section 3.1 of the main text, first we transform the components of $\mu_p$ so that the transformed parameter space is the entire Euclidean space and also to get a better mixing during the MCMC run. The transformations on the components of $\mu_p$, i.e., the parameters $\psi, \theta, \phi, \omega_0, p_0, \sigma, \sigma_m, \sigma_f$ are as follows: $\psi^* = \logit(\psi), \theta^* = \logit(\theta), \phi^* = \logit(\phi), \omega_0^* = \logit(\omega_0), p_0^* = \logit(p_0), \sigma^* = \log(\sigma), \sigma_m^* = \log(\sigma_m), \sigma_f^* = \log(\sigma_f)$.

We denote the set of transformed parameters as $\mu_p^* = h(\mu_p)$ and the transformed prior
by $\pi^{\ast}(\mu_p^{\ast})$ with domain as the Euclidean space.

**Transformed prior distributions.**

The prior distribution on the parameters $\psi$, $\theta$, $\phi$, $\omega_0$, $p_0$, $\sigma$, $\sigma_m$ and $\sigma_f$ are given in Table 1. The transformed prior on $\psi^{\ast}$, $\theta^{\ast}$, $\phi^{\ast}$, $\omega_0^{\ast}$, $p_0^{\ast}$, $\sigma^{\ast}$, $\sigma_m^{\ast}$ and $\sigma_f^{\ast}$ are

\[
\begin{align*}
\pi^{\ast}(\psi^{\ast}) &= \pi\left(h^{-1}(\psi^{\ast})\right) | J | = \frac{\exp(\psi^{\ast})}{(1 + \exp(\psi^{\ast}))^2}, \\
\pi^{\ast}(\theta^{\ast}) &= \pi\left(h^{-1}(\theta^{\ast})\right) | J | = \frac{\exp(\theta^{\ast})}{(1 + \exp(\theta^{\ast}))^2}, \\
\pi^{\ast}(\phi^{\ast}) &= \pi\left(h^{-1}(\phi^{\ast})\right) | J | = \frac{\exp(\phi^{\ast})}{(1 + \exp(\phi^{\ast}))^2}, \\
\pi^{\ast}(\omega_0^{\ast}) &= \pi\left(h^{-1}(\omega_0^{\ast})\right) | J | = \frac{\exp(\omega_0^{\ast})}{(1 + \exp(\omega_0^{\ast}))^2}, \\
\pi^{\ast}(p_0^{\ast}) &= \pi\left(h^{-1}(p_0^{\ast})\right) | J | = \frac{\exp(p_0^{\ast})}{(1 + \exp(p_0^{\ast}))^2}, \\
\pi^{\ast}(\sigma^{\ast}) &= \pi\left(h^{-1}(\sigma^{\ast})\right) | J | = \frac{\exp(\sigma^{\ast})}{R}, \\
\pi^{\ast}(\sigma_m^{\ast}) &= \pi\left(h^{-1}(\sigma_m^{\ast})\right) | J | = \frac{\exp(\sigma_m^{\ast})}{R}, \\
\pi^{\ast}(\sigma_f^{\ast}) &= \pi\left(h^{-1}(\sigma_f^{\ast})\right) | J | = \frac{\exp(\sigma_f^{\ast})}{R},
\end{align*}
\]

where $| J |$ denotes the Jacobian of transformation and $h^{-1}(\cdot)$ denotes the inverse transformation of the parameter.

**Tuning density.**

We have carried out the computation of Gelfand-Dey estimator using different choices of $g(\mu_p^{\ast})$ to assess the sensitivity of the resulting estimates (see Section 2.2.1). First, density of the multivariate normal distribution $\mathcal{N}(\hat{\mu}_p^{\ast}, \hat{\Sigma})$ with mean $\hat{\mu}_p^{\ast}$ and covariance matrix $\hat{\Sigma}$ is chosen. $\hat{\mu}_p^{\ast}$ and $\hat{\Sigma}$ are computed by using the MCMC draws $\{\mu_p^{\ast,d} : d = 1, \ldots, N_{\text{iter}}\}$. The mean parameter $\hat{\mu}_p^{\ast} = (\hat{\mu}_{p1}^{\ast}, \hat{\mu}_{p2}^{\ast}, \ldots)$ has components $\hat{\mu}_{pi}^{\ast} = \frac{1}{N_{\text{iter}}} \sum_{d=1}^{N_{\text{iter}}} \mu_{pi}^{\ast,d}$. The entries of the covariance matrix $\hat{\Sigma}$ can be obtained by using the MCMC sample as

\[
\hat{\Sigma}_{ii} = \frac{1}{N_{\text{iter}} - 1} \sum_{d=1}^{N_{\text{iter}}} (\mu_{pi}^{\ast,d} - \hat{\mu}_{pi}^{\ast})^2, \quad \hat{\Sigma}_{ij} = \frac{1}{N_{\text{iter}} - 1} \sum_{d=1}^{N_{\text{iter}}} (\mu_{pi}^{\ast,d} - \hat{\mu}_{pi}^{\ast})(\mu_{pj}^{\ast,d} - \hat{\mu}_{pj}^{\ast}).
\]

Additionally, we have used multivariate-$t$ densities for $g(\mu_p^{\ast})$ with mean parameter $\mu_p^{\ast}$ and scale matrix $\hat{\Sigma}$ using different degrees of freedom: 10, 100, 500, 1000, 10000. A truncated normal density has also been tried for $g(\mu_p^{\ast})$ following the suggestion of Geweke (1999). The parameters of the truncated normal density $\tilde{\mu}_p^{\ast}, \tilde{\Sigma}$ are obtained as above.
Table 1: Prior distributions and transformations of model parameters and latent variables that are used for the MCMC algorithms. Here $|J|$ denotes the Jacobian of transformation and $h^{-1}(\cdot)$ denotes the inverse transformation of the parameter.

| Parameter | Prior | Transformed parameter | Transformed prior pdf | Corresponding models |
|-----------|-------|-----------------------|----------------------|---------------------|
| $\phi$    | Uniform(0, 1) | $\phi^* = \text{logit}(\phi)$ | $\frac{\exp(\phi^*)}{(1 + \exp(\phi^*))^2}$ | $M_1, M_3$ |
| $\omega_0$| Uniform(0, 1) | $\omega_0^* = \text{logit}(\omega_0)$ | $\frac{\exp(\omega_0^*)}{(1 + \exp(\omega_0^*))^2}$ | $M_1, M_3$ |
| $p_0$     | Uniform(0, 1) | $p_0^* = \text{logit}(p_0)$ | $\frac{\exp(p_0^*)}{(1 + \exp(p_0^*))^2}$ | $M_2, M_4$ |
| $\sigma$  | Uniform(0, $R$) | $\sigma^* = \log(\sigma)$ | $\frac{\exp(\sigma^*)}{R}$ | $M_3, M_4$ |
| $\sigma_m$| Uniform(0, $R$) | $\sigma_m^* = \log(\sigma_m)$ | $\frac{\exp(\sigma_m^*)}{R}$ | $M_1, M_2$ |
| $\sigma_f$| Uniform(0, $R$) | $\sigma_f^* = \log(\sigma_f)$ | $\frac{\exp(\sigma_f^*)}{R}$ | $M_1, M_2$ |
| $u_i$     | Bernoulli(\(\theta\)) | | | $M_1, M_2$ |
| $\theta$  | Uniform(0, 1) | $\theta^* = \text{logit}(\theta)$ | $\frac{\exp(\theta^*)}{(1 + \exp(\theta^*))^2}$ | $M_1, M_2$ |
| $\psi$    | Bernoulli(\(\psi\)) | | | $M_1, M_2, M_3, M_4$ |
| $L$       | Uniform distribution over the permutation space of \(\{1, \ldots, M\}\) | | | $M_1, M_2, M_3, M_4$ |
| $s_i = (s_{i1}, s_{i2})'$ | Uniform($\mathcal{V}$) | | | $M_1, M_2, M_3, M_4$ |

Then for some $\alpha \in (0, 1)$, we define the set $\hat{\Theta}$ as follows:

$$\hat{\Theta} = \left\{ \mu_p^* : (\mu_p^* - \hat{\mu}_p^*)' \hat{\Sigma}^{-1} (\mu_p^* - \hat{\mu}_p^*) \leq \chi_r^2(\alpha) \right\}$$

and we take

$$g(\mu_p^*) = \alpha^{-1}(2\pi)^{-r/2} |\hat{\Sigma}|^{-1/2} \exp \left[ (\mu_p^* - \hat{\mu}_p^*)' \hat{\Sigma}^{-1} (\mu_p^* - \hat{\mu}_p^*) \right] I(\mu_p^* \in \hat{\Theta}),$$

where $\chi_r^2(\alpha)$ denotes 100 \(\alpha\)% quantile of $\chi_r^2$ distribution with \(r\) denoting the dimension of $\mu_p^*$. We have carried out the computations for three values of $\alpha$: 0.9, 0.95 and 0.99.

Note that, in the integrated likelihood (IL) approximation approach for computing Bayes factors, we assume $g(\mu_p^*, L) = g_1(\mu_p^*) g_2(L)$. We take $g_1(\mu_p^*)$ from the above choices in MAP approximation approach and take the prior density $\pi(L)$ as the tuning density $g_2(L)$. 

128
Steps of MCMC sampling

The joint posterior density of the parameter $\mu := \{\psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, u_0, z, S, L\}$ under a model $M_i$ is obtained by using the corresponding model likelihood and the prior distributions of the parameters and latent variables, $i = 1, 2, 3, 4$. For instance, the joint posterior density of $\mu$ under $M_1$ is given below (using the model likelihood (??) from the main text and the prior distributions mentioned in Table 1):

$$
\pi(\mu \mid Y^*, u_{\text{obs}}) \propto f(Y^*, u_{\text{obs}} \mid \mu) \pi(z \mid \psi) \pi(u_0 \mid z, \theta) \pi(\psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, S, L)
= \prod_{i=1}^{M} \left[ \left\{ \psi^i \theta^{n_i} (1 - \theta)^{1 - u_i} \phi^{y_i - (1 - \phi)^{2n_i - y_i}} \prod_{j=1}^{J} \eta_j(s_i, u_i)^{n_{ij}} \right\}^z_i (1 - \psi)^{1 - z_i} \right] \times \pi(\psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, S, L),
$$

(D.1)

where $y_{ij} = y_{ij}^{(1)} + y_{ij}^{(2*)}$, $y_{i..} = y_{i..}^{(1)} + y_{i..}^{(2*)}$, $n_{ij} = \sum_{k=1}^{K} I(y_{ijk}^{(1)} + y_{ijk}^{(2*)} > 0)$, $n_{i..} = \sum_{j=1}^{J} n_{ij}$ and $\pi(\psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, S, L)$ is the joint prior density for the parameters $\psi, \theta, \phi, \omega_0, \sigma_m, \sigma_f, S, L$.

The other posterior densities can be obtained similarly.

Metropolis-within-Gibbs is easy to implement and is quite flexible in this modelling set up. Since parameters are updated sequentially, it is easy for investigators to follow and even modify specifications such as jumping probabilities etc. Here, Metropolis within Gibbs sampling is performed to obtain MCMC chains for the components of $\mu$ (viz., $\psi$, $\theta$, $\phi$, $\omega_0$, $\sigma_m$, $\sigma_f$, $u_0$, $z$, $S$, $L$ for model $M_1$). The transformations and the assumed prior distributions are given in Table 1. Next we give the steps of each iteration of the MCMC analysis.

1. **Initialisation.** Initialise each of the components of $\mu$. While initialising $L$, we should take care of the fact that no mismatch takes place with respect to the sexes, i.e., we should avoid linking male (or female) individual from detector 2 to female (or male) individual from detector 1.
2. **Augmentation.** Augment all-zero detection histories to $Y^{(1)}_{\text{obs}}$ and $Y^{(2)}_{\text{obs}}$ for making them of dimension $M \times J \times K$. We denote them $Y^{(1)}$ and $Y^{(2)}$.

3. **Re-ordering.** Given the matching vector $L$, we re-order the rows of $Y^{(2)}$ so that they are matched properly with $Y^{(1)}$. We denote this ordered right side data set as $Y^{(2*)}$.

4. **Update the movement scale parameter(s).**

   (a) Under $M_1$ or $M_2$, we have $\sigma^*_m$ and $\sigma^*_f$ as the movement scale parameters and are updated under each of these two models similarly. For model $M_1$, the full conditional of $\sigma^*_m$ is obtained from (D.1) and is of a non-standard form

   $\pi(\sigma^*_m | Y^*, \mu_{\text{rest}}) \propto \pi(\sigma^*_m) \prod_{i:u_i=1}^J \prod_{j=1}^K \eta_j(s_i, u_i) \frac{z_i}{z_i n_{ij}} \left\{ (1 - \eta_j(s_i)) + \left( 1 - \frac{\exp(\phi^*)}{1 + \exp(\phi^*)} \right)^2 \eta_j(s_i) \right\}^{z_i(K-n_{ij})}$,

   where $\mu_{\text{rest}}$ denotes all the remaining parameters other than $\sigma^*_m$. Here a normal proposal is assumed for $\sigma^*_m$ and is updated using a Metropolis algorithm. The parameters of this proposal distribution at $t^{th}$ iteration are $\sigma^*_m(t-1)$ and $\tau^2_m$,

   where $\sigma^*_m(t-1)$ is the $(t-1)^{th}$ element in the Markov chain of $\sigma^*_m$. The movement scale parameter for female individuals $\sigma^*_f$ is updated in a similar manner.

   (b) Under $M_3$ or $M_4$, we have $\sigma^*$ as the only movement scale parameters and is updated under each of these two models similarly. For model $M_3$, the full conditional of $\sigma^*$ is obtained from the corresponding joint posterior density of $\mu$ and is of a non-standard form

   $\pi(\sigma^* | Y^*, \mu_{\text{rest}}) \propto \pi(\sigma^*) \prod_{i=1}^J \prod_{j=1}^K \eta_j(s_i) \frac{\eta_j(s_i)}{z_i n_{ij}} \left\{ (1 - \eta_j(s_i)) + \left( 1 - \frac{\exp(\phi^*)}{1 + \exp(\phi^*)} \right)^2 \eta_j(s_i) \right\}^{z_i(K-n_{ij})}$,

   where $\mu_{\text{rest}}$ denotes all the remaining parameters other than $\sigma^*$. Here a normal proposal is assumed for $\sigma^*$ and is updated using a Metropolis algorithm. The parameters of this proposal distribution at $t^{th}$ iteration are $\sigma^*(t-1)$ and $\tau^2$,

   where $\sigma^*(t-1)$ is the $(t-1)^{th}$ element in the Markov chain of $\sigma^*$.

5. **Update each $u_i \in u_0$ (only for $M_1$ and $M_2$).** When $z_i = 1$, the full conditional
distribution of \( u_i (\in \mathbf{u}_0) \) is the Bernoulli distribution with parameter \( \theta_{0i} \). Under model \( M_1 \), \( \theta_{0i} \) takes the following form
\[
\theta_{0i} = \frac{\theta A_i(s_i, 1)}{\theta A_i(s_i, 1) + (1 - \theta) A_i(s_i, 0)}
\]
where \( A_i(s_i, u_i) = \prod_{j=1}^{J} \eta_j(s_i, u_i)^{n_{ij}} \left\{ (1 - \eta_j(s_i, u_i)) + \left( 1 - \frac{\exp(\phi^s)}{1 + \exp(\phi^s)} \right)^2 \eta_j(s_i, u_i) \right\}^{K-n_{ij}} \).

Note that the full conditional of \( u_i \) is independent of any other \( u_a \)'s \((a \neq i)\), \( u_i, u_a \in \mathbf{u} \). When \( u_i \notin \mathbf{u}_0 \) or \( z_i = 0 \), we don’t require to update \( u_i \). The latent variable \( u_i \) is updated in a similar manner under \( M_2 \).

6. **Update \( \theta \) (only for \( M_1 \) and \( M_2 \)).** The full conditional distribution of \( \theta \) is \( \text{Beta} \left( \sum_{i=1}^{M} z_i u_i + 1, \sum_{i=1}^{M} z_i (1 - u_i) + 1 \right) \) under both \( M_1 \) and \( M_2 \).

7. **Update each \( z_i \).** For all the detected individuals, i.e., \( i \)'s such that \( y_{i.}^{(1)} + y_{i.}^{(2 \ast)} > 0 \), the full conditional of \( z_i \) has a degenerate distribution at the value 1, denoted by
\[
\pi(z_i | y_{i.}^{(1)}, y_{i.}^{(2 \ast)}, y_{i.}^{(1)} + y_{i.}^{(2 \ast)} > 0, \mu_{\text{rest}}) = I(z_i = 1)
\]
where \( \mu_{\text{rest}} \) denotes all the remaining parameters other than \( z_i \). For the undetected individuals with all zero detection histories (i.e., \( y_{i.}^{(1)} = 0 \) and \( y_{i.}^{(2 \ast)} = 0 \)), the full conditional distribution of \( z_i \) is Bernoulli with parameter \( \psi_{i0} \). Under \( M_1 \), \( \psi_{i0} \) has the following expression
\[
\psi_{i0} = \frac{\psi \theta^{u_i} (1 - \theta)^{1-u_i} \prod_{j=1}^{J} \left\{ (1 - \eta_j(s_i, u_i)) + \left( 1 - \frac{\exp(\phi^s)}{1 + \exp(\phi^s)} \right)^2 \eta_j(s_i, u_i) \right\}^K}{\psi \theta^{u_i} (1 - \theta)^{1-u_i} \prod_{j=1}^{J} \left\{ (1 - \eta_j(s_i, u_i)) + \left( 1 - \frac{\exp(\phi^s)}{1 + \exp(\phi^s)} \right)^2 \eta_j(s_i, u_i) \right\}^K + (1 - \psi)}.
\]
Note that the full conditional distribution of \( z_i \) is independent of any other \( z_j \)'s \((j \neq i)\). The full conditional distribution of \( z_i \) can be obtained similarly for the other models.

8. **Update \( \psi \).** The full conditional distribution of \( \psi \) is \( \text{Beta} \left( \sum_{i=1}^{M} z_i + 1, M - \sum_{i=1}^{M} z_i + 1 \right) \).

9. **Update each \( s_i \).** For model \( M_1 \), the full conditional distribution for \( s_i \) is of a
non-standard form with density

$$
\pi(s_i \mid Y^*, \mu_{\text{rest}}) \propto \pi(s_i) \prod_{j=1}^{J} \eta_j(s_i, u_i)^{z_{i,n_{ij}}} \left\{ \left( 1 - \eta_j(s_i, u_i) \right) + \left( 1 - \frac{\exp(\phi^*)}{1 + \exp(\phi^*)} \right)^2 \eta_j(s_i, u_i) \right\}^{2(K-n_{ij})},
$$

where $\mu_{\text{rest}}$ denotes all the remaining parameters other than $s_i$. Each $s_i$ can be simulated using a random walk Metropolis algorithm. A normal proposal is assumed for $s_i$. The parameters of this proposal distribution at $t^{th}$ iteration are $s_i^{(t-1)}$ and $\tau_s^2$, where $s_i^{(t-1)}$ is the $(t-1)^{th}$ element in the Markov chain of $s_i$. Note that the full conditional distribution of $s_i$ is independent of any other $s_j$'s ($j \neq i$). Each $s_i$ can be updated similarly for other models.

10. Update $L$.

(a) We assume a Uniform prior over the permutation space of $L$. Observe that, each individual from detector 2 (ID2) can be linked with only one from the list of the individuals from detector 1 (ID1s) (of the same sex). The technique to update $L$ is given below.

(b) We start by selecting one ID2, say $i$ and get $L_i = l$ (say), the ID1 (and also the corresponding $s_l$) to whom it was linked in the previous iteration. The activity centres $S$ and data on sex $u$ are in the same order as ID1’s.

(c) Then we find the set of (real) individual indices of the same sex as ID1 $l$ which is also the candidate set to select a new ID1 for ID2 $i$. We then calculate the weights proportional to $b_{l,a} = \exp\left(-\frac{1}{2\sigma_L^2} \|s_l - s_a\|^2\right)$ for each individual in this set, where $s_a$ is the the location of the activity centre for ID1 $a$ and $\sigma_L$ is a tuning parameter. A good choice for $\sigma_L$ can be $\sigma_m$ or $\sigma_f$ based on whether $u_l$ is 1 or 0. We randomly draw one ID1 from this candidate set, say $l^*$, according to these weights.

(d) This is the swapping step. We find the ID2 who is linked with ID1 $l^*$ and denote it by $i^*$. Now we link ID2 $i^*$ with ID1 $l$ and link ID2 $i$ with ID1 $l^*$. Consequently, we have $L_{i^*} = l$ and $L_i = l^*$. We also find the set of (real) activity centres with same sex as ID1 $j^*$ and calculate the weights $b^*_{l^*,a} = \ldots$
exp\left(-\frac{1}{2\sigma_L^2}\|s_{l^*} - s_a\|^2\right) \text{ for each individual in this set.}

**ID1**  \hspace{1cm} **ID2**

\[
l \quad \quad \quad \quad \quad \quad \quad i \quad \quad \quad \quad \quad \quad \quad L_i = l^*
\]

\[
l^* \quad \quad \quad \quad \quad \quad \quad i^* \quad \quad \quad \quad \quad \quad \quad L_{i^*} = l
\]

(e) After the above swapping we obtain a candidate matching vector \(L_{\text{cand}}\). Observe that the linking and swapping only takes place between individuals of the same sex so that mismatch does not occur. Then we update \(L\) using the Metropolis-Hastings acceptance ratio (\(\rho\))

\[
\rho = \frac{\pi(L_{\text{cand}} \mid Y^*, u_{\text{obs}}, L_{\text{rest}}) (b_{l^*, l}^{\text{back}} / \sum_a b_{l^*, a}^{\text{back}})}{\pi(L \mid Y^*, u_{\text{obs}}, L, \mu_{\text{rest}}) (b_{l, l^*} / \sum_a b_{l, a})}
\]

where \(\mu_{\text{rest}}\) denotes all the remaining parameters other than \(L\) and \(\pi(L)\) denotes the prior density of \(L\).

11. **Update \(\phi\) (only for \(M_1\) and \(M_3\)).** The full conditional of \(\phi^*\) is obtained from (D.1) and is of a non-standard form

\[
\pi(\phi^* \mid Y^*, \mu_{\text{rest}}) \propto \pi(\phi^*) \prod_{i=1}^{M} \left( \frac{\exp(\phi^*)}{1 + \exp(\phi^*)} \right)^{z_i y_{i^*}} \left( \frac{1 - \exp(\phi^*)}{1 + \exp(\phi^*)} \right)^{z_i (2n_i - y_{i^*})} \\
\times \prod_{j=1}^{J} \left\{ (1 - \eta_j(s_i, u_i)) + \left( 1 - \frac{\exp(\phi^*)}{1 + \exp(\phi^*)} \right)^2 \frac{\eta_j(s_i, u_i)}{\eta_j(s_i, u_i)} \right\}^{z_i (K - n_{ij})},
\]

where \(y_i = y_{i^*}^{(1)} + y_{i^*}^{(2)}\) for each \(i\) and \(\mu_{\text{rest}}\) denotes all the remaining parameters other than \(\phi^*\). The parameter \(\phi^*\) can be simulated using a random walk Metropolis algorithm. A normal proposal is assumed for \(\phi^*\). The parameters of this proposal distribution at \(t\)th iteration are \(\phi^{*(t-1)}\) and \(\tau_\phi^2\), where \(\phi^{*(t-1)}\) is the \((t - 1)\)th element in the Markov chain of \(\phi^*\).

12. **Update \(\omega_0^*\) (only for \(M_1\) and \(M_3\)).** The full conditional of \(\omega_0^*\) is obtained from
(D.1) and is of a non-standard form

\[
\pi(\omega_0^* | Y^*, \mu_{\text{rest}}) \propto \pi(\omega_0^*) \prod_i \prod_j \left( \frac{\exp(\omega_0^*)}{1 + \exp(\omega_0^*)} \right)^{z_{n_i j}} \left\{ (1 - \eta_j(s_i, u_i)) + \left( 1 - \frac{\exp(\phi^*)}{1 + \exp(\phi^*)} \right)^2 \eta_j(s_i, u_i) \right\}^{z_i(K-n_{i j})},
\]

where \( \mu_{\text{rest}} \) denotes all the remaining parameters other than \( \omega_0^* \). The parameter \( \omega_0^* \) can be updated using a random walk Metropolis algorithm. A normal proposal is assumed for \( \omega_0^* \). The parameters of this proposal distribution at \( t \)th iteration are \( \omega_0^{*(t-1)} \) and \( \tau_\phi^2 \), where \( \omega_0^{*(t-1)} \) is the \( (t-1) \)th element in the Markov chain of \( \omega_0^* \).

13. **Update** \( p_0^* \) (only for \( M_2 \) and \( M_4 \)). Under \( M_2 \), the full conditional density of \( p_0^* \) can be expressed as the following

\[
\pi(p_0^* | Y^*, \mu_{\text{rest}}) \propto \pi^*(p_0^*) \prod_i \prod_j \left( \frac{\exp(p_0^*)}{1 + \exp(p_0^*)} \right)^{z_{i y_{ij}}} \left( 1 - p_j(s_i, u_i) \right)^{z_i(2K-y_{ij})}.
\]

where \( \mu_{\text{rest}} \) denotes all the remaining parameters other than \( p_0^* \). The full conditional of \( p_0^* \) is of a non-standard form and can be updated using a random walk Metropolis algorithm. A normal proposal is assumed for \( p_0^* \). The parameters of this proposal distribution at \( t \)th iteration are \( p_0^{*(t-1)} \) and \( \tau_\phi^2 \), where \( p_0^{*(t-1)} \) is the \( (t-1) \)th element in the Markov chain of \( p_0^* \). The parameter \( p_0^* \) can be updated similarly for \( M_4 \).

14. Repeat the above steps (1) - (13) until convergence of each parameter.
E Results of the simulation study

(a) GD - MAP Multivariate normal

(b) GD - MAP Multivariate $t$, df = 10

(c) GD - MAP Multivariate $t$, df = 100

(d) GD - MAP Multivariate $t$, df = 500

(e) GD - MAP Multivariate $t$, df = 1000

(f) GD - MAP Multivariate $t$, df = 10000

Figure 1: The proportion of times Bayes factor favours any particular model using the MAP approximation approach. Plot (a): Gelfand-Dey estimator with a multivariate normal density for $g$. Plots (b)-(f) : Gelfand-Dey estimator with five different choices for $g$, viz., densities of multivariate-$t$ distribution with degrees of freedom 10, 100, 500, 1000, 10000 respectively.
Figure 2: The proportion of times Bayes factor favours any particular model using the MAP approximation approach. Plots (a) - (c) are obtained by computing Gelfand-Dey estimator with three different choices for $g$, viz., truncated normal density with $\alpha = 0.9, 0.95, 0.99$. 
Figure 3: The proportion of times Bayes factor favours any particular model using the integrated likelihood approximation (IL) approach. Plot (a): Gelfand-Dey estimator with a multivariate normal density for $g$. Plots (b)-(f): Gelfand-Dey estimator with five different choices for $g$, viz., densities of multivariate-$t$ distribution with degrees of freedom 10, 100, 500, 1000, 10000 respectively.
Figure 4: The proportion of times Bayes factor favours any particular model using the integrated likelihood approximation approach. Plots (a) - (c) : Gelfand-Dey estimator with three different choices for $g$, viz., truncated normal density with $\alpha = 0.9, 0.95, 0.99$.

Figure 5: The proportion of times Bayes factor (using the harmonic mean estimator) favours any particular model.
Figure 6: The proportion of time a model selection method favours any particular model. Plots (a)-(f) correspond to WAIC1, WAIC2, WAIC3, DIC1, DIC2 and posterior predictive loss respectively.
Figure 7: Plot of average RMSE estimates of $N$, $\psi$, $N_{Male}$, $\theta$ and $\phi$ over different simulation scenarios.
Figure 8: Plot of average RMSE estimates of $\omega_0$, $p_0$, $\sigma_m$, $\sigma_f$ and $\sigma$ over different simulation scenarios.
Figure 9: Scatter plot of the parameters obtained from MCMC chains of a simulated data analysis of $M_1$ corresponding to scenario 1. Here $r$ denotes the correlation coefficient and is computed by using the MCMC chains of the respective parameters.
Figure 10: Scatter plot of the parameters obtained from MCMC chains of a simulated
data analysis of $M_1$ corresponding to scenario 2. Here $r$ denotes the correlation coefficient
and is computed by using the MCMC chains of the respective parameters.
Figure 11: Scatter plot of the parameters obtained from MCMC chains of a simulated data analysis of $M_1$ corresponding to scenario 3. Here $r$ denotes the correlation coefficient and is computed by using the MCMC chains of the respective parameters.
Figure 12: Scatter plot of the parameters obtained from MCMC chains of a simulated data analysis of $M_1$ corresponding to scenario 7. Here $r$ denotes the correlation coefficient and is computed by using the MCMC chains of the respective parameters.
Figure 13: Scatter plot of the parameters obtained from MCMC chains of a simulated data analysis of $M_1$ corresponding to scenario 9. Here $r$ denotes the correlation coefficient and is computed by using the MCMC chains of the respective parameters.
Figure 14: Scatter plot of the parameters obtained from MCMC chains of a simulated data analysis of $M_1$ corresponding to scenario 12. Here $r$ denotes the correlation coefficient and is computed by using the MCMC chains of the respective parameters.

References

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: inference, development, and communication. *Econometric Reviews*, 18(1):1–73.