Evaluation of some random effects methodology applicable to bird ringing data

KENNETH P. BURNHAM¹ & GARY C. WHITE²,

¹Colorado Cooperative Fish & Wildlife Research Unit, USGS-BRD, Colorado, USA, and
²Department of Fishery and Wildlife Biology, Colorado State University, USA
Live capture-recapture or tag recovery, so models such as Cormack-Jolly-Seber \( \{S_t, p_t\} \), or tag recovery \( \{S_t, f_t\} \).

Focus on estimable \( S_1, S_2, \ldots, S_k \), hence model \( \{S_t, \theta_t\} \).
Random effects:

$S_i$ vary about $E(S) = \mu$: $S_i = \mu + \epsilon_i$

This variation is $\sigma^2$

Can generalize, e.g., $S_i = a + b i + \epsilon$

$S_i = a + b x_i + c y_i + \epsilon$
As if $S_1, \ldots, S_k$ constitute a sample under the structural model

$(k = 5$ poor, $10$ marginal, $15 +$ preferred)
If we knew the $S_i$

$$\hat{E}(S) = \overline{S}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{k} (S_i - \overline{S})^2}{k-1}$$
We have MLEs $\hat{S}_i$ and $\text{var}(\hat{S}_i \mid S_i)$

$$\hat{S}_i = \mu + (S_i - \mu) + (\hat{S}_i - S_i)$$

$$= \mu + \delta_i + \epsilon_i$$

$\uparrow \sigma^2$  $\uparrow \text{var}(\hat{S}_i \mid S_i)$
Total variation of $\hat{S}_i$:

$$\sigma^2 + E_S[\text{var}(\hat{S}_i | S_i)]$$

Note a distinction: $E_S[\text{var}(\hat{S}_i | S_i)]$ vs. $\text{var}(\hat{S}_i | S_i)$

$\text{var}(\hat{S}_i | S_i)$ depends on sample size of animals; $\sigma^2$ does not
A simple estimator:

\[ \hat{\sigma}^2 = \frac{\sum_{i=1}^{k} (\hat{S}_i - \bar{S})^2}{k-1} - \hat{\text{var}}(\hat{S}_i \mid S_i) + \hat{\text{cov}}(\hat{S}_i, \hat{S}_j \mid S_i, S_j) \]

Called “naive” in MARK output
A concept in random effects models:

MLEs AS A SET can be improved upon

by SHRINKAGE estimators $\tilde{S}_i$,

which generally lie between $\hat{E}(S)$ and $\hat{S}_i$:

$\hat{E}(S) \quad \xrightarrow{\text{SHRINKAGE}} \quad \tilde{S}_i \quad \xrightarrow{\text{SHRINKAGE}} \quad \hat{S}_i$
For independent $\hat{S}_i$, shrinkage depends on

$$\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{E}_S[\text{var}(\hat{S}_i | S_i)]}$$

Usually use $\hat{E}_S[\text{var}(\hat{S}_i | S_i)] = \hat{\text{var}}(\hat{S}_i | S_i)$
Shrinkage estimator

$$\tilde{S}_i = \hat{E}(S) + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \text{var}(\hat{S}_i|S_i)}} \times \left[ \hat{S}_i - \hat{E}(S) \right]$$

\( \sqrt{\cdot} \) used because then

$$\hat{\sigma}^2 = \frac{k}{k-1} \sum_{i=1}^{k} (\tilde{S}_i - \overline{\tilde{S}})^2$$

and  

$$\hat{E}(S) = \overline{\tilde{S}}$$
Random effects (shrinkage) estimator model is intermediate between

\{S_t, \theta\} \quad \text{and} \quad \{S, \theta\}

Its likelihood is

\mathcal{L}(\hat{S}_1, \ldots, \hat{S}_k, \hat{\theta}(\hat{S}) \mid \text{model } \{S_t, \theta\})
\( \tilde{S} = G\hat{S} \) for a projection matrix \( G(\hat{\sigma}^2) \)

Analogy to dimension \( k \) of \( \hat{S} \) is trace\((G)\)

AIC for RE model is standard formula with

\[ K_{re} = \dim(\theta) + \text{tr}(G) \]
Mallard adult males banded preseason in California (USA) for 42 years, 1955-1996, data from banding in years 1 to 21

Numbers banded:
1020 1228 2279 1720 1571 1769 1076 1100 1218 1350 1713 1313 1226 1348 2192 1269 1165 858 2222 1618 1037

Recoveries to year 21 for banding years 1 to 21:

76 36 30 14 8 14 5 1 0 0 1 1 0 0 0 0 0 0 0 0 0
93 79 30 18 11 11 6 4 2 1 0 0 1 0 0 1 0 0 0 0
197 122 51 26 35 27 8 14 4 1 0 0 1 0 1 0 0 0 0 0
133 74 47 30 18 7 15 5 1 4 0 2 0 0 0 0 0 0 0 0
113 78 37 20 21 9 9 4 6 2 1 0 0 1 0 0 0 0 0
117 88 45 29 23 11 10 7 10 4 7 1 0 0 0 0 0
72 41 19 24 8 8 5 4 5 6 1 0 1 0 0 0
68 33 36 20 13 9 3 9 1 4 1 0 0 0 1
74 54 24 23 14 11 6 4 5 4 0 3 1
103 42 33 13 18 10 9 5 4 1 3 3
75 63 39 24 17 16 15 5 7 4 3
85 34 40 16 20 23 9 7 3 3
60 50 32 28 16 9 12 11 8
96 45 44 22 23 15 9 13
82 74 54 39 26 22 17
96 47 37 22 16 12
62 50 31 26 15
36 26 24 18
145 99 70
124 68
85
California mallard data
Under model \{ S_t, f_t \}, a GOF test gave $\chi^2 = 280.861$, 235 df, $P = 0.0216$, over dispersion factor $\hat{c} = 1.1952$ (235 df)

Test of model \{ S, f_t \} vs. \{ S_t, f_t \}, $P = 2.1 \times 10^{-12}$
but time-saturated model adds 40 parameters
Basic random effects model

\[ \hat{S} = X\beta + \delta + \epsilon, \]

\[ \text{VC}(\hat{S}) = D = \sigma^2 I + E_S(W) \]
estimate $\hat{\beta}$, $\sigma^2$, confidence interval on $\sigma$
unconditional VC matrix for $\hat{\beta}$
shrinkage estimator, $\tilde{S}$.
conditional C.I. based on $\tilde{S}$
\[ D = \sigma^2 I + E_S(W) \]

\[ \hat{\beta} = (X'D^{-1}X)^{-1}X'D^{-1}\hat{S} \]

\[ k - r = (\hat{S} - X\hat{\beta})'D^{-1}(\hat{S} - X\hat{\beta}) \]

\[ \text{VC}(\hat{\beta}) = (X'D^{-1}X)^{-1} \]
\[ \chi^2_{df,1-\alpha/2} = (\hat{S} - \hat{X}\beta)'D^{-1}(\hat{S} - \hat{X}\beta) \]

\[ \chi^2_{df,\alpha/2} = (\hat{S} - \hat{X}\beta)'D^{-1}(\hat{S} - \hat{X}\beta) \]
Re $\tilde{S}$:

$H = \sigma D^{-1/2}$

$\tilde{S} = H(\hat{S} - X\hat{\beta}) + X\hat{\beta}$

$G = H + (I - H)AD^{-1}$

$A = X(X' D^{-1} X)^{-1} X'$

$\tilde{S} = G\hat{S}$
\[
VC(\tilde{S} \mid S) = GE_S(W)G'
\]
\[
\hat{rmse}(\tilde{S}_i \mid S) = \sqrt{\hat{\text{var}}(\tilde{S}_i \mid S) + (\tilde{S}_i - \hat{S}_i)^2}
\]
\[
\log \mathcal{L}(\tilde{S}, \tilde{\theta}) \equiv \log \mathcal{L}(\tilde{S}, \hat{\theta} (\tilde{S})) = \max_{\theta} \left[ \log \mathcal{L}(\tilde{S}, \theta) \right]
\]

\(\tilde{S}\) is fixed, and essentially “contains” \(\hat{\sigma}^2\)

\[
\text{AICc} = -2 \log \mathcal{L}(\tilde{S}, \tilde{\theta}) + 2K_{re} + 2 \frac{K_{re}(K_{re}+1)}{n+K_{re}-1}
\]

\[
K_{re} = \text{tr}(G) + \ell,
\]
MONTE CARLO SIMULATION STUDY
Single group CJS capture-recapture data

Design:

- capture occasions \((t = k + 2)\), \(4\) levels \((7, 15, 23, 31)\),
- new releases \((u)\) on each occasion, \(2\) levels \((100, 400)\),
- constant \(p\) on each occasion, \(2\) levels \((0.6, 0.8)\),
- mean survival probability, \(E(S)\), \(2\) levels \((0.6, 0.8)\),
- process variation, \(\sigma\), \(4\) levels \((0, 0.025, 0.05, 0.1)\)

Initially more extensive
128 = 4 × 2 × 2 × 2 × 4) cases (design points)
500 independent data sets each case (64,000 total)

If \( \sigma > 0 \), then \( S_1, \ldots, S_k \) have a
beta distribution, mean \( E(S) \), variance \( \sigma^2 \)

Used SAS and MARK
Q1a  What is the bias of (signed) $\hat{\sigma}^2$

<table>
<thead>
<tr>
<th>t</th>
<th>(0)</th>
<th>(0.025)</th>
<th>(0.05)</th>
<th>(0.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0.0001078</td>
<td>0.000761</td>
<td>0.00249</td>
<td>0.01015</td>
</tr>
<tr>
<td>15</td>
<td>0.0000446</td>
<td>0.000620</td>
<td>0.00253</td>
<td>0.01002</td>
</tr>
<tr>
<td>23</td>
<td>0.0000299</td>
<td>0.000637</td>
<td>0.00251</td>
<td>0.00999</td>
</tr>
<tr>
<td>31</td>
<td>0.0000273</td>
<td>0.000645</td>
<td>0.00251</td>
<td>0.00994</td>
</tr>
<tr>
<td>mean</td>
<td>0.0000524</td>
<td>0.000666</td>
<td>0.00251</td>
<td>0.01002</td>
</tr>
</tbody>
</table>
Q1b  What is the bias of $\hat{\sigma}$

<table>
<thead>
<tr>
<th>$t$</th>
<th>0</th>
<th>0.025</th>
<th>0.05</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0.00916</td>
<td>0.0275</td>
<td>0.0499</td>
<td>0.1007</td>
</tr>
<tr>
<td>15</td>
<td>0.00591</td>
<td>0.0249</td>
<td>0.0502</td>
<td>0.1001</td>
</tr>
<tr>
<td>23</td>
<td>0.00467</td>
<td>0.0252</td>
<td>0.0501</td>
<td>0.0999</td>
</tr>
<tr>
<td>31</td>
<td>0.00462</td>
<td>0.0254</td>
<td>0.0501</td>
<td>0.0997</td>
</tr>
</tbody>
</table>

mean 0.00609 0.0257 0.0501 0.1001
Q2  Relative frequency of $\hat{\sigma}^2 < 0$?

<table>
<thead>
<tr>
<th>t</th>
<th>0</th>
<th>0.025</th>
<th>0.05</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0.592</td>
<td>0.339</td>
<td>0.159</td>
<td>0.026</td>
</tr>
<tr>
<td>15</td>
<td>0.516</td>
<td>0.144</td>
<td>0.022</td>
<td>0.001</td>
</tr>
<tr>
<td>23</td>
<td>0.510</td>
<td>0.076</td>
<td>0.004</td>
<td>0.000</td>
</tr>
<tr>
<td>31</td>
<td>0.484</td>
<td>0.047</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>mean</td>
<td>0.526</td>
<td>0.152</td>
<td>0.047</td>
<td>0.007</td>
</tr>
</tbody>
</table>
Q3 C. I. coverage on signed $\sigma^2$?

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>%coverage</th>
<th>$t$</th>
<th>%coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>94.6</td>
<td>7</td>
<td>94.9</td>
</tr>
<tr>
<td>0.025</td>
<td>95.0</td>
<td>15</td>
<td>94.9</td>
</tr>
<tr>
<td>0.050</td>
<td>95.1</td>
<td>23</td>
<td>94.8</td>
</tr>
<tr>
<td>0.100</td>
<td>94.7</td>
<td>31</td>
<td>94.8</td>
</tr>
</tbody>
</table>

$\sigma^2$ above C.I.: 2.53%
$\sigma^2$ below C.I.: 2.63%
Q4  C.I. coverage on $S_i$, focus on average over $k$

$\hat{S}_i \pm 1.96 \hat{se}(\hat{S}_i \mid S)$

$\tilde{S}_i \pm 1.96 \hat{rmse}(\tilde{S}_i \mid S)$
Global coverage:

95.5% for the shrinkage estimator

95.0% for the MLE
Coverage based on $\tilde{S}$

<table>
<thead>
<tr>
<th>t</th>
<th>0</th>
<th>0.025</th>
<th>0.05</th>
<th>0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>97.8</td>
<td>93.1</td>
<td>92.1</td>
<td>93.7</td>
</tr>
<tr>
<td>15</td>
<td>99.5</td>
<td>94.0</td>
<td>94.3</td>
<td>94.8</td>
</tr>
<tr>
<td>23</td>
<td>99.7</td>
<td>94.8</td>
<td>94.7</td>
<td>94.7</td>
</tr>
<tr>
<td>31</td>
<td>99.8</td>
<td>95.3</td>
<td>95.0</td>
<td>94.8</td>
</tr>
<tr>
<td>mean</td>
<td>99.2</td>
<td>94.3</td>
<td>94.0</td>
<td>94.5</td>
</tr>
</tbody>
</table>
Q5 What are the relative lengths of these two confidence intervals?

\[
\text{grand average length ratio} = \frac{\sum_{i=1}^{k} \text{mse}(\tilde{S}_i|S)}{\sum_{i=1}^{k} \text{se}(\hat{S}_i|S)}
\]
<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>confidence intervals ratio, formula (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>0.739</td>
</tr>
<tr>
<td>0.025</td>
<td>0.799</td>
</tr>
<tr>
<td>0.050</td>
<td>0.879</td>
</tr>
<tr>
<td>0.100</td>
<td>0.949</td>
</tr>
</tbody>
</table>
Looked various MSE ratios, like

\[
\frac{\sum_{i=1}^{k} \text{MSE}(\hat{S}_i)}{\sum_{i=1}^{k} \text{MSE}(\hat{S}_i)}
\]

Shrinkage (empirical Bayes) beats MLE every time
Q8  What is the performance here of AIC?
    3 models fit \( \hat{S}(\cdot), \tilde{S}(t), \hat{S}(t) \)

Did it work?  YES
Our most striking finding: AIC for the random effects model was always smaller than AIC for the “parent" fixed effects model, \( \{S_t, p_t\} \)

“Always:" in all 64,000 simulated trials
Table 6. The proportion of simulation trials in which AIC selected the random effects model rather than the constant $S$ model.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>0</th>
<th>0.025</th>
<th>0.05</th>
<th>0.10</th>
<th>0.25</th>
<th>0.50</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>100</td>
<td>0.303</td>
<td>0.445</td>
<td>0.653</td>
<td>0.914</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>400</td>
<td>0.374</td>
<td>0.760</td>
<td>0.911</td>
<td>0.992</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>0.228</td>
<td>0.550</td>
<td>0.892</td>
<td>0.994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>400</td>
<td>0.350</td>
<td>0.892</td>
<td>0.996</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>100</td>
<td>0.186</td>
<td>0.632</td>
<td>0.948</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>400</td>
<td>0.255</td>
<td>0.954</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>100</td>
<td>0.158</td>
<td>0.677</td>
<td>0.967</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>400</td>
<td>0.201</td>
<td>0.976</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>100</td>
<td>0.219</td>
<td>0.576</td>
<td>0.865</td>
<td>0.977</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>400</td>
<td>0.295</td>
<td>0.896</td>
<td>0.977</td>
<td>0.998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td>0.257</td>
<td>0.736</td>
<td>0.921</td>
<td>0.987</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7  Average $\Delta$AIC results, over 500 trials, for some cases where $\Delta$AIC(.) $\neq$ $\Delta$AIC(t) but $\Delta$AIC(RE) is substantial lower.

<table>
<thead>
<tr>
<th>t</th>
<th>E(S)</th>
<th>$\sigma$</th>
<th>p</th>
<th>u</th>
<th>$\Delta$AIC(.)</th>
<th>$\Delta$AIC(RE)</th>
<th>$\Delta$AIC(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.6</td>
<td>0.025</td>
<td>0.6</td>
<td>400</td>
<td>5.33</td>
<td>0.084</td>
<td>7.24</td>
</tr>
<tr>
<td>15</td>
<td>0.6</td>
<td>0.050</td>
<td>0.6</td>
<td>100</td>
<td>5.19</td>
<td>0.095</td>
<td>7.26</td>
</tr>
<tr>
<td>15</td>
<td>0.8</td>
<td>0.025</td>
<td>0.8</td>
<td>100</td>
<td>5.41</td>
<td>0.097</td>
<td>7.05</td>
</tr>
<tr>
<td>23</td>
<td>0.6</td>
<td>0.025</td>
<td>0.6</td>
<td>400</td>
<td>9.62</td>
<td>0.070</td>
<td>10.29</td>
</tr>
<tr>
<td>23</td>
<td>0.6</td>
<td>0.050</td>
<td>0.6</td>
<td>100</td>
<td>8.95</td>
<td>0.124</td>
<td>10.79</td>
</tr>
<tr>
<td>23</td>
<td>0.8</td>
<td>0.025</td>
<td>0.8</td>
<td>100</td>
<td>10.39</td>
<td>0.103</td>
<td>9.91</td>
</tr>
<tr>
<td>31</td>
<td>0.6</td>
<td>0.025</td>
<td>0.6</td>
<td>400</td>
<td>13.82</td>
<td>0.064</td>
<td>13.71</td>
</tr>
<tr>
<td>31</td>
<td>0.6</td>
<td>0.050</td>
<td>0.6</td>
<td>100</td>
<td>12.64</td>
<td>0.131</td>
<td>14.45</td>
</tr>
<tr>
<td>31</td>
<td>0.8</td>
<td>0.025</td>
<td>0.8</td>
<td>100</td>
<td>15.13</td>
<td>0.077</td>
<td>12.93</td>
</tr>
</tbody>
</table>
Basic message:

The random effects inference methods in MARK worked very well in this idealized, small simulation study
Some practicalities and concerns

Over dispersion happens; adjust $\hat{W}$ by overdispersion factor $c$

MARK adjusts model-produced $\hat{W}$ to be $\hat{c}\hat{W}$, then this sampling VC matrix is used in above formulae

Critical: reliable $\hat{c}$; this also ties to reliable GOF testing
Practicalities and concerns (more)

Re $\hat{\sigma}^2$, bad things happen if any $\hat{S}_i$ is at, or too near, a bound of 1; you get $\hat{\text{var}}(\hat{S}_i | S_i) = 0$, or badly biased low.
Hence, use identity link (often does not matter)

In $\{S_t, \theta_t\}$ do not constrain $\theta$
Use $\{S_t, f_i\}$ not $\{S_t, r_i\}$ for tag recoveries