Modelling of grouped recorded data for different population structures in pig breeding

Mahmoud Shirali¹, Bjarne Nielsen², Tage Ostersen², Ole Christensen¹, Per Madsen¹, Guosheng Su¹, Just Jensen¹

¹Center for Quantitative Genetics & Genomics, Aarhus University, Denmark
²SEGES, Pig Research Centre, Denmark
Objective: To investigate the effect of population structure in the data on grouped analysis:

- Model group records with varying group sizes
- Considering multiple random effects
- Asses genetic parameter and breeding values

1) Develop methodology based on analysing simulated data
2) Applying the method to real data
Modelling pooled data with varying group size and different random effects

• Individual record:

\[ y = Xb + Z_l l + Z_c c + Z_a a + e \]

• Group record:

\[ Ty = T(Xb + Z_l l + Z_c c + Z_a a + e) \]

\[ \sigma^2_{a*} = Z^*_a A Z^*_a \sigma^2_a = \sum_{j=1}^{n_g} A_{jj} \sigma^2_a + n_g (n_g - 1) A_{jj} \sigma^2_a \]

\[ \sigma^2_{l*} = Z^*_1 Z^*_1 \sigma^2_l = \sum_{k=1}^{N_L} n_{lk}^2 \sigma^2_l \]

\[ \sigma^2_{c*} = Z^*_c Z^*_c \sigma^2_c = n_g^2 \sigma^2_c \]

\[ \sigma^2_{e*} = n_g \sigma^2_e \]

Su et al. 2018
Real data

• **Low within pen relationship (Case 1)**
  • 23824 animals for analysis in 709 sections in 15 years
  • Majority 1 littermate per pen.
  • 47868 animals in pedigree
  • Group size: *8 to 14* pigs per pen

• **Large within pen relationship (Case 2)**
  • 100285 pigs with phenotype
  • 112959 pigs with pedigree
  • Group size: *6 to 15* pigs per pen
Simulation of data

- Pedigree & data structure from low and large within pen relationship data are used
- Breeding values
  \[ a_i = \frac{1}{2} a_s + \frac{1}{2} a_d + m_i \]
- Mendelian sampling term has expected value of 0 and variance of
- No selection
- Taking into account inbreeding
- 10 replicates

<table>
<thead>
<tr>
<th></th>
<th>( \sigma^2_a )</th>
<th>( \sigma^2_c )</th>
<th>( \sigma^2_l )</th>
<th>( \sigma^2_e )</th>
<th>( h^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>simulated</td>
<td>300</td>
<td>70</td>
<td>40</td>
<td>490</td>
<td>0.33</td>
</tr>
</tbody>
</table>
Simulation: genetic parameter estimation

\[ y = Xb + Za + Wc + Sl + e \]
\[ y^* = X^*b + Z^*a + W^*c + S^*l + e^* \]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Analysis</th>
<th>( \sigma_a^2 )</th>
<th>( \sigma_c^2 )</th>
<th>( \sigma_l^2 )</th>
<th>( \sigma_e^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated</td>
<td></td>
<td>300</td>
<td>70</td>
<td>40</td>
<td>490</td>
</tr>
<tr>
<td>Case 1</td>
<td>Individual</td>
<td>298 (18)</td>
<td>71 (5)</td>
<td>44 (11)</td>
<td>490 (13)</td>
</tr>
<tr>
<td></td>
<td>Pooled</td>
<td>241 (60)</td>
<td>70 (26)</td>
<td>76 (91)</td>
<td>495 (254)</td>
</tr>
<tr>
<td>Case 2</td>
<td>Individual</td>
<td>298 (15)</td>
<td>71 (2)</td>
<td>40 (3)</td>
<td>492 (6)</td>
</tr>
<tr>
<td></td>
<td>Pooled</td>
<td>302 (43)</td>
<td>72 (7)</td>
<td>31 (16)</td>
<td>491 (75)</td>
</tr>
</tbody>
</table>
Simulation: accuracy of breeding value estimation

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Individual</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>0.67 (0.02)</td>
<td>0.34 (0.03)</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.70 (0.01)</td>
<td>0.48 (0.02)</td>
</tr>
</tbody>
</table>
Real data analysis: low within pen

Modelling feed efficiency

\[ y_{ij} = YHM_i + b_1 SBW_j + b_2 ADG_j + a_j + c_j + e_{ij} \]

\[ y_{ij}^* = YHM_i^* + b_1^* n_j^* + b_2^* SBW_j^* + b_3^* ADG_j^* + a_k^* + e_{ijk}^* \]

<table>
<thead>
<tr>
<th>Trait</th>
<th>( \sigma_c^2 )</th>
<th>( \sigma_a^2 )</th>
<th>( \sigma_e^2 )</th>
<th>Total</th>
<th>( h^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual RFI</td>
<td>6.6</td>
<td>16.7</td>
<td>73.2</td>
<td>96.5</td>
<td>0.17</td>
</tr>
<tr>
<td>Sum group RFI</td>
<td>-</td>
<td>18.8</td>
<td>93.0</td>
<td>111.8</td>
<td>0.17</td>
</tr>
</tbody>
</table>
Accuracy: low within relationship

\[ \text{Cor}(BV_{\text{individual}}, BV_{\text{group}}) = 0.36 \]
## Real data analysis: large within pen relationship

### Modelling growth

\[
\begin{align*}
\gamma_{ijk} &= YHM_i + \text{SEX}_j + a_k + c_k + l_k + e_{ijk} \\
\gamma_{ijkl}^* &= YHM_i^* + \text{SEX}_j^* + b_1 n_k^* + a_l^* + c_l^* + l_l^* + e_{ijkl}^*
\end{align*}
\]

<table>
<thead>
<tr>
<th>Trait</th>
<th>Analysis</th>
<th>(\sigma_a^2)</th>
<th>(\sigma_c^2)</th>
<th>(\sigma_l^2)</th>
<th>(\sigma_e^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADG</td>
<td>Individual</td>
<td>2691 (113)</td>
<td>694 (18)</td>
<td>424 (19)</td>
<td>4219 (51)</td>
</tr>
<tr>
<td></td>
<td>Pooled</td>
<td>2909 (370)</td>
<td>434 (88)</td>
<td>484 (74)</td>
<td>5661 (864)</td>
</tr>
</tbody>
</table>

The accuracy of breeding value estimation = 0.65
Conclusions

• Developed model for pooled data considers:
  - multiple fixed and random effects
  - varying group sizes

• Developed model can be used for genetic parameter estimation
  - With larger standard error

• A close genetic relationship within a pen is favourable for breeding value estimation

• Genetic parameter estimation with group records on feed intake and growth is feasible