Heterogeneous Variance Adjustment in Across-Country Genetic Evaluation with Country-Specific Heritabilities

M.H. Lidauer¹, E.A. Mäntysaari¹, J. Pösö², J.-Å. Eriksson³, U.S. Nielsen⁴, G.P. Aamand⁵

¹MTT Agrifood Research Finland, ²Faba Service, Finland, ³Swedish Dairy Association, ⁴The Danish Agricultural Advisory, ⁵NAV Nordic Cattle Genetic Evaluation
Motivation

• Across-country genetic evaluation
  1. Single trait approach
  2. Multiple trait approach but $r_g$ across countries = unity
  3. Multiple trait approach with $r_g$ across countries < unity

• Approaches 1 and 2 yield one set of breeding values

→ how to ensure a homogeneous genetic variance across countries?
Motivation

- Nordic random regression TDM for Red Cattle
  - Multiple trait approach but $r_{g \text{ across countries}} = 1.0$
  - Finnish Ayrshire, Red Danish Cattle, Swedish Red Breed
  - Different variance components for each country (breed)
Motivation

- Nordic random regression TDM for Red Cattle
  - Multiple trait approach but $r_g$ across countries = 1.0
  - Finnish Ayrshire, Red Danish Cattle, Swedish Red Breed
  - Different variance components for each country (breed)

### Estimated heritabilities compiled for 305-d yields

<table>
<thead>
<tr>
<th>Trait</th>
<th>Milk</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Lactation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finnish Ayrshire</td>
<td>0.38</td>
<td>0.33</td>
<td>0.31</td>
<td>0.33</td>
<td>0.32</td>
<td>0.31</td>
</tr>
<tr>
<td>Red Danish Cattle</td>
<td>0.42</td>
<td>0.35</td>
<td>0.34</td>
<td>0.38</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Swedish Red Breed</td>
<td>0.44</td>
<td>0.33</td>
<td>0.34</td>
<td>0.43</td>
<td>0.34</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Motivation

- Nordic random regression TDM for Red Cattle
  - Multiple trait approach but $r_g$ across countries = 1.0
  - Finnish Ayrshire, Red Danish Cattle, Swedish Red Breed
  - Different variance components for each country (breed)

Estimated heritabilities compiled for 305-d yields

<table>
<thead>
<tr>
<th>Trait</th>
<th>Milk</th>
<th></th>
<th>Protein</th>
<th></th>
<th>Fat</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Lactation</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Finnish Ayrshire</td>
<td>0.38</td>
<td>0.33</td>
<td>0.31</td>
<td>0.33</td>
<td>0.32</td>
<td>0.31</td>
</tr>
<tr>
<td>Red Danish Cattle</td>
<td>0.42</td>
<td>0.35</td>
<td>0.34</td>
<td>0.38</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>Swedish Red Breed</td>
<td>0.44</td>
<td>0.33</td>
<td>0.34</td>
<td>0.43</td>
<td>0.34</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Aim

• Estimation of genetic variances using Mendelian sampling deviations

• Calibration of the heterogeneous variance adjustment method to ensure homogeneous genetic variance across countries
Estimation of genetic variance from Mendelian sampling deviations

- $\hat{\sigma}_{a_t}^2$ estimated from a group of animals (Sullivan, 1999):

$$\hat{\sigma}_{a_t}^2 = \frac{1}{n_t} \sum_{k=1}^{n_t} d_k \left[ \hat{m}_{kt}^2 + PEV(\hat{m}_{kt}) \right]$$

- $n_t$ number of animals
- $d_k$ is 2, 4/3, or 1 depending on known parents
- $\hat{m}_{kt} = EBV_{kt} - \frac{1}{2} \left( EBV_{st} + EBV_{dt} \right)$ Mendelian sampling deviation for animal $k$ and trait $t$
- $PEV(\hat{m}_{kt})$ prediction error variance for animal $k$ and trait $t$
Estimation of genetic variance from Mendelian sampling deviations

- Monte Carlo sampling for PEV (Hickey et al., 2009)
Estimation of genetic variance from Mendelian sampling deviations

- Monte Carlo sampling for $PEV$ (Hickey et al., 2009)
- Considering formulation:

$$PEV = \sigma_a^2 - \left[\text{Var}(\hat{u})/\text{Var}(u)\right]\sigma_a^2$$
Estimation of genetic variance from Mendelian sampling deviations

- Monte Carlo sampling for $PEV$ (Hickey et al., 2009)
- Considering formulation:
  \[
  PEV = \sigma_a^2 - \left[ \frac{Var(\hat{u})}{Var(u)} \right] \sigma_a^2
  \]
- $\hat{\sigma}_a^2$ can be estimated for a sufficiently large animal group:
  \[
  \hat{\sigma}_a^2[q] = \frac{1}{n_t} \sum_{k=1}^{n_t} d_k \hat{m}_{kt}^2 \left[ \sum_{k=1}^{n_t} d_k \hat{m}_{kt}^2 \right] \left[ \sum_{k=1}^{n_t} d_k \hat{m}_{kt}^2 \right]
  \]
- $\hat{m}_{kt}$: Mendelian sampling deviation from the real data
- $\tilde{m}_{kt}$: true Mendelian sampling deviation of the simulated data
- $\hat{m}_{kt}$: estimated Mendelian sampling deviation from the simulated data
Full model sampling to obtain $\tilde{m}_{ktr}$ and $\hat{m}_{ktr}$

- Nordic Red Cattle yield evaluation data
  - 68 million test-day records on milk, protein and fat
  - 4.3 million animals
Full model sampling to obtain $\tilde{m}_{ktr}$ and $\hat{m}_{ktr}$

- Nordic Red Cattle yield evaluation data
  - 68 million test-day records on milk, protein and fat
  - 4.3 million animals

- Multiplicative reduced rank random regression TDM:

\[
y_{ti} \lambda_{ti} = X_{ti} b_t + T_{ti} h_t + Z_{ti} a + U_{ti} p + V_{ti} w + e_{ti}
\]

- $y_{ti}$ observations of trait $t$ in stratum $i$
- $\lambda_{ti}$ multiplicative adjustment factor for stratum $i$
- $b_t, h_t$ vector of fixed effects for trait $t$
- $a, p, w$ add. genetic and non-add. genetic animal effects
- $e_{ti}$ random residuals
Full model sampling to obtain $\tilde{m}_{ktr}$ and $\hat{m}_{ktr}$

- Following García-Cortés et al. (1992)
  - $\tilde{b}_t = 0, \tilde{h}_t = 0$
  - $\tilde{a} = (L \otimes T_a) x_{na} t_a$
  - $\tilde{p} = (I_{np} \otimes T_p) x_{np} t_p$, $\tilde{w} = (I_{nw} \otimes T_w) x_{nw} t_w$
  - $\tilde{e}_j = P_j T_r x_{tr}$

where $L, T_a, T_p, T_w, T_r$ are Cholesky decompositions of $A$ and of the corresponding VCV matrices, and $x_n \sim N(0, I_n)$ are random samples from stand.N.D.

- $\hat{\lambda}_{ti} = 1$
Full model sampling to obtain $\tilde{m}_{ktr}$ and $\hat{m}_{ktr}$

- Following García-Cortés et al. (1992)
  - $\tilde{b}_t = 0, \tilde{h}_t = 0$
  - $\tilde{a} = (L \otimes T_a) x_{n_a t_a}$
  - $\tilde{p} = (I_{n_p} \otimes T_p) x_{n_p t_p}$, $\tilde{w} = (I_{n_w} \otimes T_w) x_{n_w t_w}$
  - $\tilde{e}_j = P_j T_r x_{t_r}$

  where $L, T_a, T_p, T_w, T_r$ are Cholesky decompositions of $A$ and of the corresponding VCV matrices, and $x_n \sim N(0, I_n)$ are random samples from stand.N.D.

- $\tilde{\lambda}_{ti} = 1$

- ... yields: $\tilde{y}_{ti} \tilde{\lambda}_{ti} = X_{ti} \tilde{b}_t + T_{ti} \tilde{h}_t + Z_{ti} \tilde{a} + U_{ti} \tilde{p} + V_{ti} \tilde{w} + \tilde{e}_{ti}$
Heterogeneity of genetic variance across countries

• Estimation of genetic variances
  • Applied to evaluation model which accounts for heterogeneous variance within traits only
  • For each cow birth year group
Heterogeneity of genetic variance across countries

- Estimation of genetic variances
  - Applied to evaluation model which accounts for heterogeneous variance within traits only
  - For each cow birth year group
- Result
  - Genetic variances differ between countries up to 30%
Heterogeneity of genetic variance across countries

• Estimation of genetic variances
  • Applied to evaluation model which accounts for heterogeneous variance within traits only
  • For each cow birth year group

• Result
  • Genetic variances differ between countries up to 30%

---

Genetic variance for 2nd lactation milk yield before adjustment for across-country heterogeneity

- Red Danish Cattle (Denmark)
- Finnish Ayrshire (Finland)
- Swedish Red Breed (Sweden)

- Year of Birth
  - 1996
  - 1998
  - 2000
  - 2002
  - 2004
Calibration of heterogeneous variance adjustment method

- Multiplicative mixed model approach (Meuwissen et. al., 1996)
  - Scales all effects in the model in proportionality to the residual variance
  - Converges to a set of solutions, which fulfill:
    \[ n_{ti} \hat{\sigma}_{e_t}^2 = y_{ti}^T \lambda_{ti} \hat{e}_{ti} \]
    where \( \hat{\sigma}_{e_t}^2 \) is the residual variance for trait \( t \)
Calibration of heterogeneous variance adjustment method

• Multiplicative mixed model approach (Meuwissen et. al., 1996)
  • Scales all effects in the model in proportionality to the residual variance
  • Converges to a set of solutions, which fulfill:

\[ n_{ti} \hat{\sigma}_{e_t}^2 = y_{ti}^T \lambda_{ti} \hat{e}_{ti} \]

where \( \hat{\sigma}_{e_t}^2 \) is the residual variance for trait \( t \)

modified condition

\[ n_{ti} \hat{\sigma}_{e_t}^2 \alpha_t = y_{ti}^T \lambda_{ti} \hat{e}_{ti} \]

where \( \alpha_t \) is a calibration factor for trait \( t \)
Calibration of heterogeneous variance adjustment method

- Iterative procedure to obtain $\boldsymbol{\alpha}_t$
  - initialize $q=1$, $\alpha_t^{[q]} = 1.0$
  - Solve multiplicative random regression TDM
  - Estimate genetic variances $\hat{\sigma}^2_a^{[q]}$
  - Update calibration factors
    
    $\alpha^{[q+1]}_t = \frac{\alpha^{[q]}_t \hat{\sigma}^2_a^{[q]} / \hat{\sigma}^2_a^{[q]}_{t \text{BASE}}}{\hat{\sigma}^2_a^{[q]}_{t,c}}$

  where $c$ is either Finland or Denmark and BASE is Sweden

- Repeat until differences in genetic SDs < +/- 1%
Calibration of heterogeneous variance adjustment method

- Iterative procedure to obtain $\alpha_t$
  - Initialize $q=1$, $\alpha_t^{[q]} = 1.0$
  - Solve multiplicative random regression TDM
  - Estimate genetic variances $\hat{\sigma}^2[q]$
  - Update calibration factors
    
    $\alpha_{t,c}^{[q+1]} = \alpha_{t,c}^{[q]} \hat{\sigma}^2[q] \hat{\sigma}_{a_{t,BASE}} / \hat{\sigma}_{a_{t,c}}$

  where $c$ is either Finland or Denmark and BASE is Sweden
  - Repeat until differences in genetic SDs < +/- 1%

- Animal groups for estimation of genetic variances
  - All cows born in 2002 to 2004
    - Finnish Ayrshire: 180 573
    - Red Danish Cattle: 50 067
    - Swedish Red Breed: 159 961
Results

• One data sample was sufficient for sampling the mean prediction error variances for the three cow groups

• Calibration factors converged after six calibration cycles

  Range of calibration factors: 0.75 … 1.38
Results

- One data sample was sufficient for sampling the mean prediction error variances for the three cow groups.
- Calibration factors converged after six calibration cycles.
  
  Range of calibration factors: 0.75 … 1.38
### Results

Genetic standard deviations by trait and country (in kg for 305d yields)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Milk</th>
<th>Protein</th>
<th>Fat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Lactation</td>
<td>742</td>
<td>895</td>
<td>936</td>
</tr>
<tr>
<td>Finnish Ayrshire</td>
<td>739</td>
<td>894</td>
<td>933</td>
</tr>
<tr>
<td>Red Danish Cattle</td>
<td>740</td>
<td>889</td>
<td>931</td>
</tr>
</tbody>
</table>
Results

• Genetic standard deviations by trait and country (in kg for 305d yields)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Milk 1</th>
<th>Milk 2</th>
<th>Milk 3</th>
<th>Protein 1</th>
<th>Protein 2</th>
<th>Protein 3</th>
<th>Fat 1</th>
<th>Fat 2</th>
<th>Fat 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactation</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Finnish Ayrshire</td>
<td>742</td>
<td>895</td>
<td>936</td>
<td>21.0</td>
<td>28.6</td>
<td>30.5</td>
<td>27.4</td>
<td>36.2</td>
<td>40.2</td>
</tr>
<tr>
<td>Red Danish Cattle</td>
<td>739</td>
<td>894</td>
<td>933</td>
<td>21.2</td>
<td>28.6</td>
<td>30.1</td>
<td>27.6</td>
<td>36.4</td>
<td>40.1</td>
</tr>
<tr>
<td>Swedish Red Breed</td>
<td>740</td>
<td>889</td>
<td>931</td>
<td>21.3</td>
<td>28.3</td>
<td>30.0</td>
<td>27.7</td>
<td>36.0</td>
<td>39.8</td>
</tr>
</tbody>
</table>

• Correlations between EBVs with and without across-country calibration

  • Across all cows born 2002 – 2007: 0.9973 – 0.9990
  • Finnish Ayrshire: 0.9959 – 0.9988
  • Red Danish Cattle: 0.9987 – 0.9995
  • Swedish Red Breed: 0.9998 – 0.9999
Conclusions

- Estimation of genetic variance from Mendelian sampling deviations is useful for model development and validation.
- Monte Carlo sampling for PEV requires one replicate only (one additional BLUP run) when mean PEV are needed.
- Calibration procedure yielded homogeneous genetic variances across countries.
- And it is applicable for any heterogeneous variance adjustment method which scales the observations.
Acknowledgement

Nordic Cattle Genetic Evaluation NAV for co-financing the work and for providing the data

THANK YOU