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

M.H. Lidauer[1], E.A. Mäntysaari[1], J. Pösö[2], J.-Å. Eriksson[3], U.S. Nielsen[4], G.P. Aamand[5]

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

MTT faba svenskmjölk SWEDISH DAIRY ASSOCIATION DANSK·KVÆG Nordisk Avlsværdivurdering

Motivation

- Across-country genetic evaluation
  1. Single trait approach
  2. Multiple trait approach but $r_{g\ \text{across countries}} = \text{unity}$
  3. Multiple trait approach with $r_{g\ \text{across countries}} < \text{unity}$
- 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 $r_{g \text{ across countries}} = 1.0$
  - Finnish Ayrshire, Red Danish Cattle, Swedish Red Breed
  - Different variance components for each country (breed)

Estimated heritabilities compiled for 305-d yields

Trait	Milk			Protein			Fat		
Lactation	1	2	3	1	2	3	1	2	3
Finnish Ayrshire	**0.38**	**0.33**	**0.31**	**0.33**	**0.32**	**0.31**	**0.35**	**0.34**	**0.33**
Red Danish Cattle	**0.42**	**0.35**	**0.34**	**0.38**	**0.35**	**0.35**	**0.39**	**0.35**	**0.34**
Swedish Red Breed	**0.44**	**0.33**	**0.34**	**0.43**	**0.34**	**0.35**	**0.43**	**0.34**	**0.37**

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)

Estimated heritabilities compiled for 305-d yields

Trait	Milk			Protein			Fat		
Lactation	1	2	3	1	2	3	1	2	3
Finnish Ayrshire	0.38	0.33	0.31	0.33	0.32	0.31	0.35	0.34	0.33
Red Danish Cattle	0.42	0.35	0.34	0.38	0.35	0.35	0.39	0.35	0.34
Swedish Red Breed	0.44	0.33	0.34	0.43	0.34	0.35	0.43	0.34	0.37

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

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

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

  - n_t number of animals
  - d_k is 2, 4/3, or 1 depending on known parents
  - $\hat{m}_{kt} = EBV_{kt} - ½\left(EBV_{st} + EBV_{dt}\right)$ 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:

$$PEV = \sigma_a^2 - \left[Var(\hat{u}) / Var(u) \right] \sigma_a^2$$

Estimation of genetic variance from Mendelian sampling deviations

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

$$PEV = \sigma_a^2 - \left[Var(\hat{u})/Var(u)\right]\sigma_a^2$$

- $\hat{\sigma}_{a_t}^2$ can be estimated for a sufficiently large animal group:

$$\hat{\sigma}_{a_t}^{2[q]} = \frac{1}{n_t}\sum_{k=1}^{n_t} d_k \hat{m}_{kt}^{2[q]} \left[\sum_{k=1}^{n_t} d_k \tilde{m}_{kt}^2 \Big/ \sum_{k=1}^{n_t} d_k \hat{\tilde{m}}_{kt}^2 \right]$$

  - $\hat{m}_{kt}$ Mendelian sampling deviation from the real data
  - $\tilde{m}_{kt}$ true Mendelian sampling deviation of the simulated data
  - $\hat{\tilde{m}}_{kt}$ estimated Mendelian sampling deviation from the simulated data

Full model sampling to obtain $\tilde{m}_{ktr}$ and $\hat{\tilde{m}}_{ktr}$

- 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 $\hat{\tilde{m}}_{ktr}$

- 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:

$$\mathbf{y}_{ti}\lambda_{ti} = \mathbf{X}_{ti}\mathbf{b}_t + \mathbf{T}_{ti}\mathbf{h}_t + \mathbf{Z}_{ti}\mathbf{a} + \mathbf{U}_{ti}\mathbf{p} + \mathbf{V}_{ti}\mathbf{w} + \mathbf{e}_{ti}$$

$\mathbf{y}_{ti}$	observations of trait *t* in stratum *i*
λ_{ti}	multiplicative adjustment factor for stratum *i*
$\mathbf{b}_t, \mathbf{h}_t$	vector of fixed effects for trait *t*
$\mathbf{a}, \mathbf{p}, \mathbf{w}$	add. genetic and non-add. genetic animal effects
$\mathbf{e}_{ti}$	random residuals

Full model sampling to obtain $\tilde{m}_{ktr}$ and $\hat{\tilde{m}}_{ktr}$

- Following García-Cortés et al. (1992)
  - $\tilde{\mathbf{b}}_t = \mathbf{0}, \tilde{\mathbf{h}}_t = \mathbf{0}$
  - $\tilde{\mathbf{a}} = (\mathbf{L} \otimes \mathbf{T}_{\mathrm{a}})\mathbf{x}_{\mathrm{n_a t_a}}$
  - $\tilde{\mathbf{p}} = (\mathbf{I}_{\mathrm{n_p}} \otimes \mathbf{T}_{\mathrm{p}})\mathbf{x}_{\mathrm{n_p t_p}}$, $\tilde{\mathbf{w}} = (\mathbf{I}_{\mathrm{n_w}} \otimes \mathbf{T}_{\mathrm{w}})\mathbf{x}_{\mathrm{n_w t_w}}$
  - $\tilde{\mathbf{e}}_{\mathrm{j}} = \mathbf{P}_{\mathrm{j}} \mathbf{T}_{\mathrm{r}} \mathbf{x}_{\mathrm{t_r}}$

    where $\mathbf{L}, \mathbf{T}_{\mathrm{a}}, \mathbf{T}_{\mathrm{p}}, \mathbf{T}_{\mathrm{w}}, \mathbf{T}_{\mathrm{r}}$ are Cholesky decompositions of A and of the corresponding VCV matrices,
    and $\mathbf{x}_{\mathrm{n}} \sim N(\mathbf{0}, \mathbf{I}_{\mathrm{n}})$ are random samples from stand.N.D.
  - $\tilde{\lambda}_{ti} = 1$

Full model sampling to obtain $\tilde{m}_{ktr}$ and $\hat{\tilde{m}}_{ktr}$

- Following García-Cortés et al. (1992)
  - $\tilde{\mathbf{b}}_t = \mathbf{0}, \tilde{\mathbf{h}}_t = \mathbf{0}$
  - $\tilde{\mathbf{a}} = (\mathbf{L} \otimes \mathbf{T}_{\mathrm{a}})\mathbf{x}_{\mathrm{n_a t_a}}$
  - $\tilde{\mathbf{p}} = (\mathbf{I}_{\mathrm{n_p}} \otimes \mathbf{T}_{\mathrm{p}})\mathbf{x}_{\mathrm{n_p t_p}}$, $\tilde{\mathbf{w}} = (\mathbf{I}_{\mathrm{n_w}} \otimes \mathbf{T}_{\mathrm{w}})\mathbf{x}_{\mathrm{n_w t_w}}$
  - $\tilde{\mathbf{e}}_{\mathrm{j}} = \mathbf{P}_{\mathrm{j}}\mathbf{T}_{\mathrm{r}}\,\mathbf{x}_{\mathrm{t_r}}$

    where $\mathbf{L}, \mathbf{T}_{\mathrm{a}}, \mathbf{T}_{\mathrm{p}}, \mathbf{T}_{\mathrm{w}}, \mathbf{T}_{\mathrm{r}}$ are Cholesky decompositions of A and of the corresponding VCV matrices,
    and $\mathbf{x}_{\mathrm{n}} \sim N(\mathbf{0}, \mathbf{I}_{\mathrm{n}})$ are random samples from stand.N.D.
  - $\tilde{\lambda}_{ti} = 1$
- … yields: $\tilde{\mathbf{y}}_{ti}\tilde{\lambda}_{ti} = \mathbf{X}_{ti}\tilde{\mathbf{b}}_t + \mathbf{T}_{ti}\tilde{\mathbf{h}}_t + \mathbf{Z}_{ti}\tilde{\mathbf{a}} + \mathbf{U}_{ti}\tilde{\mathbf{p}} + \mathbf{V}_{ti}\tilde{\mathbf{w}} + \tilde{\mathbf{e}}_{ti}$

Heterogeneity of genetic variance across countries

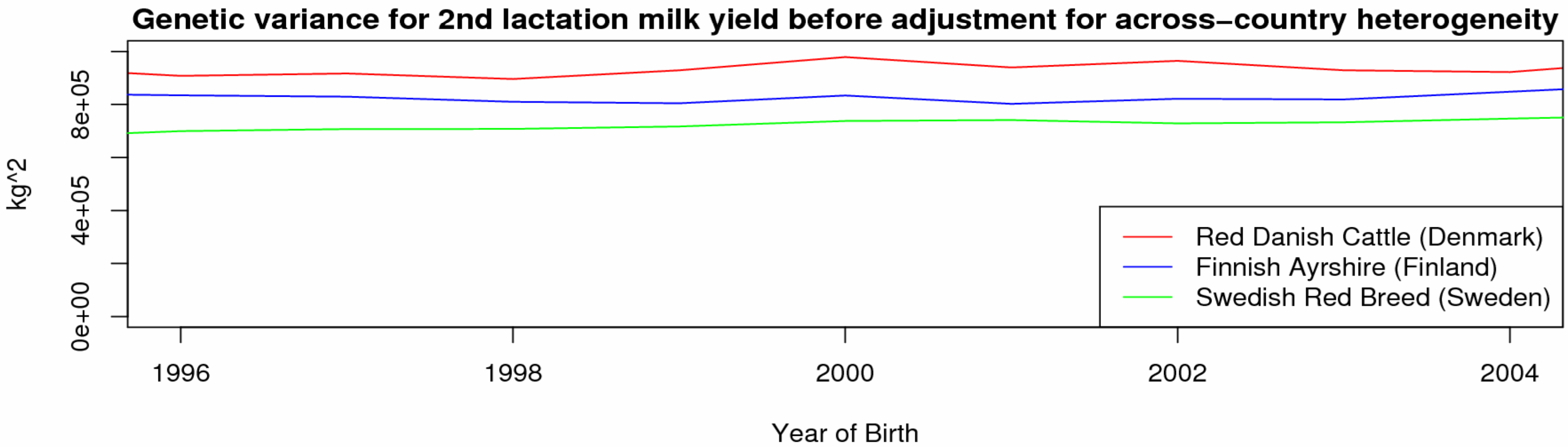
- Estimation of genetic variances
  - Applied to evaluation model which accounts for heterogeneous variance within traits only
  - For each cow birth year group

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%



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:

$$n_{ti}\hat{\sigma}^2_{e_t} = \mathbf{y}^{\mathrm{T}}_{ti}\lambda_{ti}\hat{\mathbf{e}}_{ti}$$

where $\hat{\sigma}^2_{e_t}$ 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:

$$n_{ti}\hat{\sigma}^2_{e_t} = \mathbf{y}^{\mathrm{T}}_{ti}\lambda_{ti}\hat{\mathbf{e}}_{ti}$$

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

modified condition

$$n_{ti}\hat{\sigma}^2_{e_t}\alpha_t = \mathbf{y}^{\mathrm{T}}_{ti}\lambda_{ti}\hat{\mathbf{e}}_{ti}$$

where α_t is a calibration factor for trait t

Calibration of heterogeneous variance adjustment method

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

    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 α_t
  - initialize q=1, $\alpha_t^{[q]} = 1.0$
  - Solve multiplicative random regression TDM
  - Estimate genetic variances $\hat{\sigma}_{a_t}^{2[q]}$
  - Update calibration factors $\alpha_{t.c}^{[q+1]} = \alpha_{t.c}^{[q]} \hat{\sigma}_{a_{t.BASE}}^{2[q]} \Big/ \hat{\sigma}_{a_{t.c}}^{2[q]}$

    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

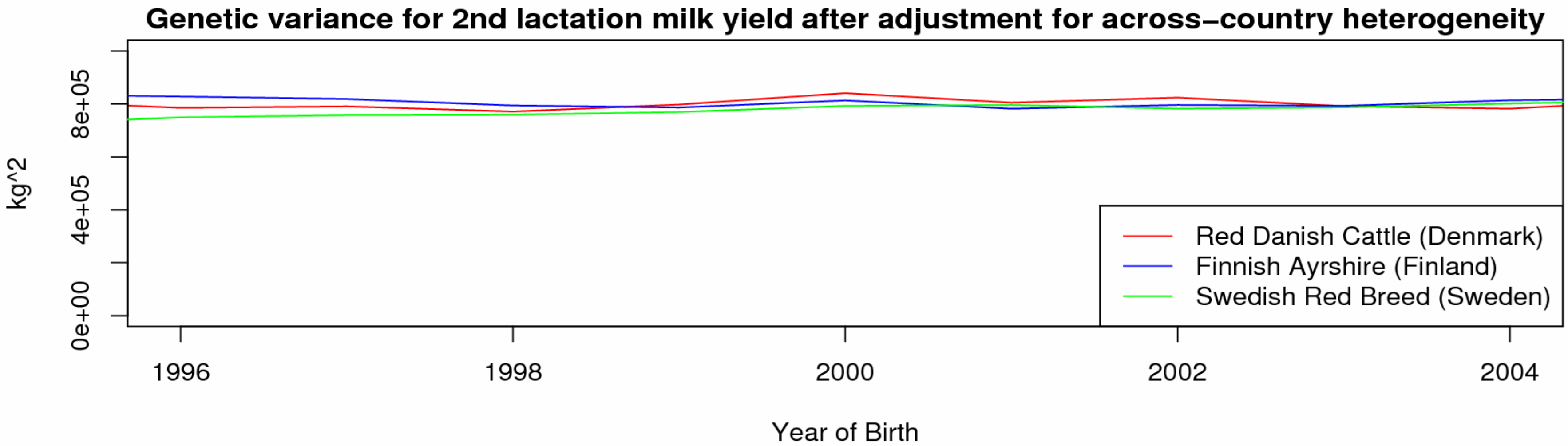
- One data sample was sufficient for sampling the mean prediction error variances for the three cow groups
- Calibration factors converged after six calibration cycles

  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)

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

Results

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

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

- 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

THANK YOU

Acknowledgement

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

faba

svenskmjölk
SWEDISH DAIRY ASSOCIATION

DANSK·KVÆG

Nordisk Avlsværdivurdering

MTT